UNIVERSITY OF CALIFORNIA SAN DIEGO

High Dimensional Expanders in Analysis and Computation

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Computer Science

by

Nathaniel Max Klevit Hopkins

Committee in charge:

Professor Daniel Kane, Co-Chair
Professor Shachar Lovett, Co-Chair
Professor Russell Impagliazzo
Professor Arya Mazumdar

2024

The Dissertation of Nathaniel Max Klevit Hopkins is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

To my parents, Paul and Sarah Hopkins, without whose guidance, kindness, and love I could never be where I am today.

EPIGRAPH

If I know only one thing, it's that everything that I see

Of the world outside is so inconceivable, often I can barely can speak

Robin Pecknold

TABLE OF CONTENTS

LIST OF ALGORITHMS

ACKNOWLEDGEMENTS

There are far too many people to whom I owe thanks to list in this short section, but I will try to do justice to those who most strongly impacted my time in graduate school and the completion of this dissertation.

First, I thank my wonderful advisors, Daniel Kane and Shachar Lovett. I am eternally grateful for their willingness to support me in the exploration of fields outside their expertise, and their incredible capacity for mentorship regardless of topic. Thank you as well to my committee members Russell Impagliazzo and Arya Mazumdar for the many insights they shared in my PhD.

I am deeply thankful as well to Tali Kaufman, without whom this dissertation would not exist. I thank her both for her generous mentorship of a neophyte in world of expanders and HDX, and for her patience, persistence, and deeply insightful questions, all of which played a critical role in my development as a researcher.

I thank Irit Dinur for introducing me to the field of high dimensional expansion and inspiring me to work on the topic. I thank her as well for her generosity in welcoming me to participate in multiple HDX workshops before I had any relevant work to show for it, and later for hosting me for a year at Weizmann where a substantial portion of this dissertation was completed. Our discussions greatly impacted my thinking in the area and the later work in this dissertation.

I have also benefited greatly from working with many wonderful collaborators throughout my PhD: Mitali Bafna, Robi Bhattacharjee, Daniel Beaglehole, Omri Ben-Eliezer, Mark Bun, Kamalika Chaudhuri, Yotam Dikstein, Marco Gaboardi, Jason Gaitonde, Suprovat Ghoshal, Russell Impagliazzo, Akash Kumar, Rex Lei, Ting-Chun Lin, Sihan Liu, Gaurav Mahajan, Tushant Mittal, Shay Moran, Michal Moshkovitz, Toni Pitassi, Arka Ray, Satchit Sivakumar, Jessica Sorrell, Madhur Tulsiani, Chutong Yang, Christopher Ye, Hantao Yu, and Ruizhe Zhang. Thank you for sharing your insights in and outside the field. I look forward to many more to come.

Thank you as well to my 4232 labmates, both the old guard Nicholas Genise, Kaave Hosseini, Sankeerth Rao, and Jiapeng Zhang from whose mentorship I frequently benefited, and to the new Daniel Beaglehole, Farzan Byramji, Sihan Liu, Jack Morris, and Anthony Ostuni for making the office a pleasure to work in. Thank you as well to my other friends and colleagues both at UCSD and the community at large, Vedat Alev, Marco Carmosino, Yuval Filmus, Tom Gur, Siqi Liu, Sidhanth Mohanty, Chinmay Nirkhe, Audrey Randall, Mihir Sathe, Mark Schultz, and June Vuong among many others, for making grad school a welcoming and fun experience. I am especially thankful to Yotam Dikstein and Gil Melnik for their hospitality during my time in Israel.

Thank you Jiahao Sun for sharing with me your unbridled joy, and reminding me research is not the only facet of a good life—sometimes a rare Pokemon is important too.

And thanks to my family, to whom I owe it all.

Chapter 1, in full, is based on the material as it appears in Symposium on Theory of Computing 2022. Bafna, Mitali; Hopkins, Max; Kaufman, Tali; Lovett, Shachar. "Hypercontractivity on High Dimensional Expanders". The dissertation author was a primary investigator and author of this material.

Chapter 4, in full, is based on the material as it appears in Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques 2022. Gaitonde, Jason; Hopkins, Max; Kaufman, Tali; Lovett, Shachar; Zhang; Ruizhe. "Eigenstripping, Spectral Decay, and Edge-Expansion on Posets". The dissertation author was a primary investigator and author of this material.

Chapter 5, in full, has been submitted for publication of the material as it may appear in Foundations of Computer Science 2024. Dikstein, Yotam; Hopkins, Max. "Chernoff Bounds and Reverse Hypercontractivity on HDX". The dissertation author was a primary investigator and author of this material.

Chapter 6, in full, is based on the material as it appears in Foundations of Computer Science 2022. Hopkins, Max; Lin, Ting-Chun. "Explicit Lower Bounds Against $\Omega(n)$-

Rounds of Sum-of-Squares". The dissertation author was a primary investigator and author of this material.

Chapter 8, in full, is based on the material as it appears in the Symposium on Discrete Algorithms 2023. Beaglehole, Daniel; Hopkins, Max; Kane, Daniel; Liu, Sihan; Lovett, Shachar. "Sampling Equilibria: Fast No-Regret Learning in Structured Games". The dissertation author was a primary investigator and author of this material.

Chapter 7, in full, is based on the material as it appears in the Symposium on Discrete Algorithms 2022. Bafna, Mitali; Hopkins, Max; Kaufman, Tali; Lovett, Shachar. "High Dimensional Expanders: Eigenstripping, Pseudorandomness, and Unique Games". The dissertation author was a primary investigator and author of this material.

VITA

2018	Bachelor of Arts, Mathematics, Harvard University
2022	Master of Science, Computer Science, University of California San Diego
2024	Doctor of Philosophy, Computer Science, University of California San Diego

PUBLICATIONS

- Max Hopkins, Russell Impagliazzo, Daniel Kane, Sihan Liu, Chris Ye. *Replicability in High Dimensional Statistics.* Pre-Print [201]
- Yotam Dikstein, Max Hopkins. *Chernoff Bounds and Reverse Hypercontractivity on HDX.* Pre-Print [114]
- Mark Bun, Marco Gaboardi, Max Hopkins, Russell Impagliazzo, Rex Lei, Toniann Pitassi, Satchit Sivakumar, Jessica Sorrell. *Stability Is Stable: Connections between Replicability, Privacy, and Adaptive Generalization.* Symposium on Theory of Computing (STOC 2023) [84]
- Robi Bhattacharjee, Max Hopkins, Akash Kumar, Hantao Yu, Kamalika Chaudhuri. *Robust Empirical Risk Minimization with Tolerance.* Algorithmic Learning Theory (ALT 2023) [64]
- Max Hopkins, Daniel Kane, Shachar Lovett, Gaurav Mahajan. *Do PAC-Leaners Learn the Marginal Distribution?* Pre-Print [207]
- Daniel Beaglehole, Max Hopkins, Daniel Kane, Sihan Liu, Shachar Lovett, *Sampling Equilibria: Fast No-Regret Learning in Structured Games.* Symposium on Discrete Algorithms (SODA 2023) [52]
- Omri Ben-Eliezer, Max Hopkins, Chutong Yang, Hantao Yu. *Active Learning Polynomial Threshold Functions.* Neural Information Processing Systems (NeurIPS 2022) [58]
- Max Hopkins, Ting-Chun Lin. *Explicit Lower Bounds Against $\Omega(n)$ Rounds of Sum of Squares.* Foundations of Computer Science (FOCS 2022) [208]
- Jason Gaitonde, Max Hopkins, Tali Kaufman, Shachar Lovett, Ruizhe Zhang. *Eigenstripping, Spectral Decay, and Edge-Expansion on Posets.* Approximation, Randomization, and Combinatorial Optimization (RANDOM 2022) [162]
- Max Hopkins, Daniel Kane, Shachar Lovett, Gaurav Mahajan. *Realizable Learning is All You Need.* Conference on Learning Theory (COLT 2022) [206]

– Mitali Bafna, Max Hopkins, Tali Kaufman, Shachar Lovett. *Hypercontractivity on High Dimensional Expanders*. Symposium on Theory of Computing (STOC 2022) [39]

– Mitali Bafna, Max Hopkins, Tali Kaufman, Shachar Lovett. *High Dimensional Expanders: Eigenstripping, Pseudorandomness, and Unique Games*. Symposium on Discrete Algorithms (SODA 2022) [38]

– Max Hopkins, Daniel Kane, Shachar Lovett, Michal Moshkovitz. *Bounded Memory Active Learning through Enriched Queries*. Conference on Learning Theory (COLT 2021) [205]

– Max Hopkins, Daniel Kane, Shachar Lovett. *The Power of Comparisons for Actively Learning Linear Separators*. Neural Information Processing Systems (NeurIPS 2020) [202]

– Max Hopkins, Daniel Kane, Shachar Lovett, Gaurav Mahajan. *Point Location and Active Learning: Learning Halfspaces Almost Optimally*. Foundations of Computer Science (FOCS 2020) [204]

– Max Hopkins, Daniel Kane, Shachar Lovett, Gaurav Mahajan. *Noise-tolerant, Reliable Active Classification with Comparison Queries*. Conference on Learning Theory (COLT 2020) [203]

– Sebastian Wagner-Carena, Max Hopkins, Ana Rivero, Cora Dvorkin. *A Novel CMB Component Separation Method: Hierarchical Generalized Morphological Component Analysis*. Monthly Notices of the Royal Astronomical Society (MNRAS 2020) [360]

– Thomas Browning, Max Hopkins, Zander Kelley. *Doppelgangers: the Ur-operation and posets of bounded height*. Formal Power Series and Algebraic Combinatorics (FPSAC 2018) [83]

ABSTRACT OF THE DISSERTATION

High Dimensional Expanders in Analysis and Computation

by

Nathaniel Max Klevit Hopkins

Doctor of Philosophy in Computer Science

University of California San Diego, 2024

Professor Daniel Kane, Co-Chair
Professor Shachar Lovett, Co-Chair

High dimensional expanders (HDX) are a nascent generalization of expander graphs (sparse yet robustly connected networks that play a core role in the theory of computation) to high dimensional domains. Despite recent breakthrough use in sampling [239, 25] and local testability [315, 117], relatively little is known about HDX, their properties, and their broader position in theoretical computer science. In this dissertation, we develop the role of high dimensional expanders in computation through the interplay of Boolean analysis, concentration of measure, approximation algorithms, and hardness of approximation.

In the first half of this dissertation, we develop a robust theory of Fourier and

probabilistic analysis on HDX. This includes generalizations of standard tools of theoretical computer science such as the Fourier decomposition, hypercontractivity, and Chernoff bounds, as well as more application-focused techniques such as symmetrization, reverse hypercontractivity, and concentration of high degree functions. In many cases, our results give the first *sparse* domains satisfying such notions, a critical consideration in application where density or 'degree' controls the cost associated with their use.

In the second half of this dissertation, we give applications of these ideas to algorithms, complexity, and mathematics. Algorithmically, we show the local structure of high dimensional expanders can be exploited to build fast approximation algorithms for unique games, and explore implications of fast approximate sampling algorithms on HDX to massive multiplayer matrix games. In mathematics, we show high dimensional expanders have optimal geometric overlap, extend a variant of the Frankl–Rödl theorem to HDX, and prove new degree lower bounds for certain HDX. Finally in complexity we leverage new topological HDX to construct optimally hard explicit constraint satisfaction problems for Sum-of-Squares (a powerful optimization paradigm), and prove spectral HDX satisfy an optimal (local) agreement testing theorem and an optimal global tester under stronger ℓ_∞-type assumptions, a stepping stone towards improved low-soundness probabilistically checkable proofs and hardness of approximation.

Introduction

Expander graphs are an immensely useful tool in the theory of computation. Introduced in the 60s [51, 320], these sparse but robustly connected graphs remain a highly active area in mathematics and computer science, with decades of breakthrough applications across network design [51], error correction [345], de-randomization [329], approximation algorithms [228], hardness of approximation [116], and much more (see e.g. [198] and citations therein). Today, it's no exaggeration to say it's difficult to find a sub-field of theoretical computer science that *hasn't* been strongly impacted by expanders.

Still, expander graphs are not one-size-fits-all. The issue, frequently, is that many core problems in computer science are *high dimensional* (say k-CSPs, learning and data analysis, distribution sampling, error correction...). Expanders, in contrast, are *1-dimensional* objects; they may give partial insights to these problems, but typically fail to capture the full picture.

Our running example of such behavior (and key motivation in this dissertation) is the *PCP Theorem* [31, 30], a breakthrough result in complexity theory showing many basic NP-hard problems like 3-SAT are *hard to approximate*.[1] Famously, the PCP Theorem can be proved using expanders [116], but the latter's low-dimensionality is also a known barrier to proving *optimal* in-approximability this way [69, 110]. Instead, modern PCPs overcome this barrier by composing expander-based constructions with high dimensional domains like the hypercube [192], complete hypergraph [327, 128], polynomials [300], or

[1] In other words, not only is it NP-hard to tell whether a 3-SAT instance is satisfiable, we cannot even distinguish between fully satisfiable instances and instances in which every potential solution fails a constant fraction of the constraints!

the Grassmann (subspaces) [129, 126, 255, 297].

Unfortunately, in many key settings, this approach is too costly: because of their high density (size), their use massively blows up the size of the problem at hand, ruining any chance at NP-reduction. Perhaps the most natural example is the parallel repetition theorem [327], which 'amplifies' a PCP by repeating it many times in parallel. If we start with a problem of size n and repeat d times, we end up with a problem of size n^d — too large to use in NP-reductions when $d \geq \omega_n(1)$. This raises a natural question, if density is truly the problem, could we replace these high dimensional domains with *sparsified* variants just as expanders replace the complete graph?

This brings us to the theory of *high dimensional expanders* (HDX). Developed jointly within topology [163, 283, 182], algebra [290], and computer science [141, 234, 124, 309], HDX are a family of generalizations of expanders to high dimensional domains like hypergraphs, posets, and chain complexes with great initial success capturing high dimensional problems in approximate sampling [239, 25, 24] and error correction [117, 315]. Nevertheless, compared to the well-developed field of graph expansion, little is known about HDX and their position in computation, especially with respect hardness of approximation where current progress, while exciting, is still fairly preliminary [124, 119, 40, 113].

In this dissertation, we develop a series of powerful analytic and probabilistic tools on HDX that form the core behind the classical theory of hardness of approximation and PCPs, including hypercontractivity and strong concentration of measure. We then give several applications of this machinery to the theory of computation and study of HDX themselves, including (among others) new hardness of approximation results within the Sum-of-Squares hierarchy, new optimal agreement tests (a core component of PCPs), efficient algorithms for unique games and high dimensional online learning problems, a proof that HDX are optimal geometric expanders, and new degree lower bounds for HDX.

Chapter 0: High Dimensional Expanders (A Probabilistic View).

What is a high dimensional expander? At the moment, there are two main variants of HDX: spectral and topological. We will eventually discuss both definitions in depth, but for the sake of this overview focus on the simplest setting: spectral expansion of partite hypergraphs. We emphasize this section *does not aim to provide a holistic view of HDX*. Rather, we give a simplified introduction focusing on one useful interpretation of spectral HDX as probabilistic objects that plays an important role in what's to come.

Recall a *bi-partite graph* $G = (L, R, E)$ consists of two sets of vertices L and R, and a set of edges (pairs) $E \subset L \times R$ passing between them. For simplicity, let's assume $L = R = [n]$. Then the graph G, defined by its edges, is really just a subset $G \subset [n]^2$. Assuming for simplicity G is also d-regular, the normalized adjacency matrix is given by

$$A_G(u, v) \coloneqq \begin{cases} \frac{1}{d} & (u, v) \in E \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to check that A has a spectral decomposition and its eigenvalues satisfy

$$1 = \lambda_1 \geq \ldots \geq \lambda_n = -1.$$

G is said to be a λ-spectral expander if $\lambda_2 \leq \lambda$.

We can also view bi-partite graphs as *two dimensional distributions*. In other words, picking a random edge from G generates a random variable (X_1, X_2) supported on $[n]^2$. By the variational characterization, an equivalent way to formulate spectral expansion is to say G is a spectral expander if the variables (X_1, X_2) are *approximately independent*. In other words, if for any $f : L \to \mathbb{R}$ and $g : R \to \mathbb{R}$, the expectation of the product fg is

close to the product of their expectations:

$$\mathbb{E}[fg] \in \mathbb{E}[f]\mathbb{E}[g] \pm \lambda\sqrt{\mathrm{Var}(f)\mathrm{Var}(g)}. \tag{1}$$

Keeping this view in mind, a *d-partite hypergraph* is a subset $X \subset [n]^d$. Similarly, X can be viewed as a *d-dimensional distribution* $(X_1, \ldots, X_d)$. Informally, we'll call X a *spectral HDX* if all d variables act 'approximately independent'. In other words, if X acts like a product space.

Formally, there are several reasonable generalizations of Equation (1) we could consider in this sense. Perhaps the first that comes to mind is simply to require a d-dimensional variant, e.g. something of the form $\forall\{f_i : [n] \to \mathbb{R}\}_{i\in[d]}$:

$$\mathbb{E}\left[\prod_{i=1}^{d} f_i(x_i)\right] \approx \prod_{i=1}^{d} \mathbb{E}[f_i]. \tag{2}$$

Or, one could instead require that every pair of variables satisfies Equation (1), a type of approximate pairwise independence. Both of these are useful notions, but miss a critical piece of the picture: the closure of products *under restriction*.

Consider a true product $X = [n]^k$. If we fix the value of X_1 (the first coordinate), it has no effect on the remaining values $(X_2, \ldots, X_k)$. Namely, the corresponding marginal, even conditioned on X_1, is *still a product*. Once we move to *sparse* subsets of $[n]^k$, it is of course no longer true that X_1 has *no* effect on $(X_2, \ldots, X_k)$, but we might hope that the resulting marginal is still 'product-like'. Combining this with our prior intuition leads to the following natural definition of a spectral HDX:

Definition 0.0.1 (Spectral High Dimensional Expanders)**.** A d-partite hypergraph X is called a λ-HDX if for all $i, j \in [d]$:

1. (X_i, X_j) is a λ-spectral expander

2. This remains true under all (valid) conditionings[2] $X_S = y_S$ for $i, j \notin S$.

Equivalently, one could require a variant of Equation (2) for all conditioned marginals.[3]

Incredibly, *bounded degree* constructions of high dimensional expanders exist for any $\lambda > 0$ [290, 238, 108], meaning every vertex sits in only $O_{d,\lambda}(1)$ hyperedges, rather than n^{d-1} hyperedges as in $X = [n]^d$. We refer the reader to [190] for an exposition on one such construction. In this dissertation, we focus instead on properties and applications of objects satisfying Definition 0.0.1 without worrying too deeply about their construction.

Before moving on, we remark that Definition 0.0.1 is not quite the standard notion of a spectral HDX, called *local-spectral expanders* [124, 309], which are instead typically defined by viewing X as a simplicial complex and analyzing the spectrum of local components called *links*.[4] Definition 0.0.1 and the viewpoint of HDX as approximately independent variables was introduced independently by Gur, Lifshitz, and Liu [187] in the partite case as the notion of a 'λ-product', and concurrently by Bafna, Hopkins, Kaufman, and Lovett [39] in the non-partite case as the notion of 'average independence' (see Chapter 1). The definition is known to be equivalent to the standard notion up to small changes in parameters [309, 109, 12].

Chapters 1-4: Boolean Function Analysis.

Having introduced our initial view of spectral HDX as sparsifying high dimensional domains like products, we arrive at a natural follow-up question: do HDX actually satisfy the core properties of these domains used in application? It seems reasonable a spectral sparsification would maintain spectral properties, what about techniques beyond spectral methods? In the context of hardness of approximation, there is a clear place to start: *boolean function analysis*, the core of all modern PCPs.

[2] In other words, y_S should actually appear in the support of X_S.

[3] It is not immediately obvious these are equivalent (up to slight changes in parameters). It follows from Oppenheim's Trickling-Down Theorem [309] and the high dimensioinal expander-mixing lemma [109, 114].

[4] In our probabilistic notation, a link simply corresponds to a conditioned marginal, e.g. $(X_S | X_{\bar{S}} = x_{\bar{S}})$. Local-spectral expansion requires the graph underlying all such marginals to be an expander.

But first, let's take a step back. The story of modern boolean function analysis starts in 1988 in the seminal work of Kahn, Kalai, and Linial (KKL) [226]. KKL, based on earlier work in social choice [59], were interested in understanding the structure of (balanced) boolean functions $f : \mathbb{F}_2^d \to \mathbb{F}_2$ with *low total influence*, a measure of the sensitivity of f to flipping individual coordinates:

$$\mathbf{I}[f] = \sum_{i=1}^{d} \mathbf{I}_i[f], \qquad \mathbf{I}_i[f] = \mathop{\mathbb{P}}_{x}[f(x) \neq f(x + e_i)].$$

What sort of functions have low influence? The most obvious example of a balanced function with low total influence is a *dictator function* $\mathbf{1}_i(x) \coloneqq x_i$. This has total influence 1, all of which is concentrated on the ith coordinate. KKL showed a type of converse to this fact: *any* low influence function must have some coordinate with much higher influence than the average. Friedgut [157] extended this result to a beautiful characterization showing any low influence function must essentially be a *junta*, a generalization of dictators that allows the function to depend on a constant number of coordiantes:

Theorem 0.0.2 (Friedgut's Theorem)**.** *If $I(f) \leq K$, then f is ε-close to a $2^{O(K/\varepsilon)}$-junta.*

Perhaps even more important than the statements of KKL and Friedgut's theorem is *how* they were proved: *Fourier analysis*. Recall on the hypercube that any function $f : \mathbb{F}_2^d \to \mathbb{R}$ can be expressed uniquely as the sum of monomials (the 'Fourier basis'):

$$f = \sum_{S \subseteq [d]} \hat{f}(S)\chi_S, \qquad f^{\leq i} = \sum_{|S| \leq i} \hat{f}(S)\chi_S$$

where $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$ are the characters of $\mathbb{F}_2^d$. Coarsely, the Fourier decomposition divides f into components by degree ($f^{\leq i}$ above is the degree at most i part of f). KKL proved their result by considering the structure of low degree functions, in particular, via a powerful tool called *hypercontractivity*. The most basic form of hypercontractivity, the

Bonami lemma [71], states low degree functions are smooth, i.e. for any $f : \mathbb{F}_2^d \to \mathbb{R}$

$$\|f^{\leq i}\|_4 \leq 2^{O(i)}\|f^{\leq i}\|_2,$$

This simple statement and its implications (the KKL Theorem [226], Junta theorem [157], Majority is stablest [302]...) are by now some of the most broadly used tools in analysis and hardness of approximation, not to mention a number of other areas such as analysis of Markov chains [107], learning [264], and algorithms [37].

The hypercube, of course, is not always the right tool for the job, and many modern problems in analysis and hardness of approximation require analogs of Bonami *beyond the cube* (see e.g. [158, 46, 227, 126, 255, 245, 280, 248, 297] among others). Here the story takes a turn: even simple extended domains such as the p-biased cube[5] actually fail standard hypercontractivity. Consider again the dictator function $\mathbf{1}_i$. On the p-biased cube, we have

$$\mathbb{E}[\mathbf{1}_i(x)^4] = p \gg \mathbb{E}[\mathbf{1}_i(x)^2]^2 = p^2.$$

Since $\mathbf{1}_i$ is a degree-1 function, the Bonami lemma cannot hold for small p. The issue is that the dictator $\mathbf{1}_i$ is *local*—while its overall density is p, upon restricting to the ith coordinate it becomes the all 1's function. Building on a long line of work studying hypercontractivity in extended domains [78, 349, 160, 157, 252], O'Donnell and Zhao [364] and Keevash, Lifhistz, Long, and Minzer [245, 248] proved this type of local function is in fact the *only* barrier to the Bonami lemma on products. Namely for any $f : [n]^d \to \mathbb{R}$, they show

$$\|f^{\leq i}\|_4^4 \leq 2^{O(i)}\|f^{\leq i}\|_2^2 \max_{|S|\leq i, x_S}\{\|f|_{x_S}\|_2^2\}. \tag{3}$$

where $\|f|_{x_S}\|_2^2 = \mathbb{E}_y[f^2(y) \mid y_S = x_S]$ and $f^{\leq i}$ now corresponds to the low degree Efron-Stein

[5]The p-biased cube is the distribution over $\mathbb{F}_2^d$ in which each bit is 1 independently with probability p.

components, a generalization of classical Fourier analysis to products:

$$f^{\leq i} = \sum_{|S| \leq i} f^{=S}, \qquad f^{=S}(y) = \sum_{T \subseteq S} (-1)^{|S \setminus T|} \mathbb{E}_x[f(x) \mid x_S = y_S].$$

Equation (3) is often called *global hypercontractivity*, since it implies any 'global' function f such that $\|f|_{x_S}\|_2^2 \leq 2^{O(i)}\|f\|_2^2$ satisfies the standard Bonami lemma. On the cube, one can check *all* functions are global, so global hypercontractivity is a true generalization of the standard notion. A simple corollary of Equation (3) is a spiritual variant of the KKL or junta theorem, any function $f : [n]^d \to \mathbb{F}_2$ with total influence $\mathbf{I}[f] \leq K$ is local in that it must have a restriction with constant density:

$$\exists |S| \leq O(K), x_S : \qquad \mathbb{E}[f|_{x_S}] \geq 2^{-O(K)}.$$

Global hypercontractivity on extended domains has become an incredibly powerful tool (leading for instance to the proof of the 2-to-2 Games Conjecture [124, 47, 255]), but prior to this point, all known examples were for *dense* settings (e.g. the slice [252], symmetric group [147], or Grassmann [255]). In Chapter 1, we prove the first hypercontractive inequality for sparse domains. In particular, building on an initial spectral theory in [111, 239, 38], we introduce a new Fourier decomposition for two-sided local-spectral HDX based on Efron-Stein, and prove it satisfies global hypercontractivity. We prove a corresponding characterization of (sparse) low influence functions generalizing the product case, and show our bound is tight. A critical component of our proof is (an un-ordered variant of) the new viewpoint of spectral HDX as 'approximately independent random variables' introduced in the prior section, called 'average independence' in Chapter 1.

The bounds we prove in Chapter 1 are not without room for improvement. First, they only hold under the strict notion of two-sided local-spectral expansion. Gur, Lifshitz, and Liu [187] independently proved a similar hypercontractive inequality for one-sided

partite HDX (in some sense a more general case), but paid for this generality in sub-optimal dependence on degree, a critical parameter in application. Second, both the results of Chapter 1 and [187] are only meaningful for certain regimes of functions, partially due to the notion of global hypercontractivity itself (which in some sense requires sparsity of f to be meaningful), and partly due to costly error terms stemming from sub-optimal handling of high norms.

In Chapter 2, we resolve these problems by extending a powerful analytic technique called *symmetrization* to HDX. One natural approach to studying functions on products is to try to reduce their analysis to the cube. Bourgain [77] realized this could be done by 'symmetrizing' a function $f : [n]^d \to \mathbb{R}$ by essentially convolving it with a random boolean string. Define the symmetrization of f as $\tilde{f}(r, x) : \mathbb{F}_2^d \times [n]^d \to \mathbb{R}$ such that

$$\tilde{f}(r, x) = \sum_{S \subseteq [d]} r_S f^{=S}(x),$$

where $r_S = (-1)^{\sum_{i \in S} r_i}$. Bourgain proved that up to the application of a little noise (applied through the standard *noise operator* T_ρ, discussed further in the next chapter), f and $\tilde{f}$ have essentially the same behavior under high norms:

$$\|T_{\frac{1}{2}} f(x)\|_q \leq \|\tilde{f}(r, x)\|_q \leq \|T_{c_q} f(x)\|_q.$$

We extend Bourgain's theorem to partite high dimensional expanders. Adapting methods of O'Donnell and Zhao [364], this gives an elementary proof of optimal global hypercontractivity for partite HDX. Adapting work of Bourgain [158], we also give a 'booster theorem' for HDX, stating any balanced low influence function must have a restriction on which it *deviates* substantially from the mean, better capturing the structural properties of highly balanced functions with low influence.

Our proof of the symmetrization theorem is based on two elementary new ideas in

the theory of high dimensional expansion of independent interest. First, we introduce the notion of 'q-norm HDX', which bound the distance of marginals (X_i, X_j) from independent in q rather than spectral norm. We observe every strong enough spectral HDX is a q-norm HDX, allowing us avoid high norm error terms lost in Chapter 1. Second, we introduce a simple method of coordinate-wise analysis on HDX which breaks high dimensional operators into coordinate-wise components and analyzes them as 1-*dimensional* operator on the marginals of X. This allows for application of standard tricks such as the replacement method, greatly simplifying prior analysis.

Finally, in Chapters 3 and 4, we study the structure of low influence functions on more general families of high dimensional expanders. In the former, we study what can be said under substantially weaker spectral assumptions on X, where no Fourier decompositions are known. Building on tools of Gotlib and Kaufman [176], we prove upper and lower bounds on the locality of low influence functions that scale with the underlying spectral expansion of X. Our results give the first structure theorems for combinatorial HDX such as (dense) clique complexes and product-complexes [170].

In the latter, we move away from hypergraphs and study Fourier analysis on more general *ranked posets* like the Grassmann. Here we give a tight spectral theory and focus on how the *regularity structure* of the underlying poset affects the resulting object's eigenvalues, explaining for instance why the Johnson graphs have *linear* eigenvalue decay, while the Grassmann has *exponential* decay, a critical parameter in application [126].[6] Unfortunately, we note in both cases the characterizations we give are likely too weak to be used in the context of hardness of approximation.

Chapter 5(a): Concentration of Measure.

Another powerful (and arguably far more ubiquitous) property of product spaces is their *concentration of measure* [75]. Somewhat counter-intuitively, here it will actually be

[6]We note these results were obtained independently in an updated version of [111].

simpler to look at the *non-partite* setting where $X \subset \binom{[n]}{d}$ rather than $[n]^d$.

Given a d-uniform hypergraph X, what does it mean for X to satisfy a concentration bound? Perhaps the simplest interpretation is to look at how well X *samples* a subset $A \subset [n]$ of its vertices. In other words, over worst-case choice of A, what is the probability that the density of A inside a random hyperedge is off by more than ε from the mean?

$$\mathbb{P}_{(x_1,\dots,x_d)\in X}\left[\left|\frac{1}{d}\sum_{i=1}^{d} 1_A(x_i) - \mu_A\right| \geq \varepsilon\right] \overset{?}{\leq} \beta(\varepsilon, d).$$

When $X = \binom{[n]}{k}$, the reader might recognize this as the classical *Chernoff-Hoeffding* bound,[7] easily the most widely-used tail bound in theoretical computer science:

$$\mathbb{P}_{(x_1,\dots,x_d)\in\binom{[n]}{d}}\left[\left|\frac{1}{d}\sum_{i=1}^{d} 1_A(x_i) - \mu_A\right| \geq \varepsilon\right] \lesssim e^{-\varepsilon^2 d}.$$

Thus, if HDX truly 'model' the complete hypergraph, a reasonable first goal would be to show they satisfy a Chernoff bound.

In fact, the question of what hypergraph families (or in light of our view of hypergraphs as distributions, what families of d-dimensional distributions) satisfy Chernoff is extremely well studied [337, 338, 339, 133, 313, 318, 75, 76, 214, 317, 273, 232, 22], and while almost all general conditions (e.g. those based on entropy or ℓ_∞-behavior) require X to be *dense*, sparse constructions called *sampler graphs* or *seeded extractors* have actually been known since the 90s,[8] and are a core tool in pseudorandomness, derandomization, and cryptography [167]. This raises a natural question: if we already know optimal sampler graphs, why use HDX?

The reason is that modern applications in complexity tend to require concentration against *broader classes of functions*. Above, we restricted our attention to sampling functions sitting on the *vertices* of X. In some sense, these correspond to *degree-1* or *linear*

[7] Formally, without replacement.

[8] In fact, they can even be constructed via expander graphs [165]!

functions. Applications in pseudorandomness and PCPs often require strong concentration bounds against *degree-i* functions. In our setting, this corresponds to looking at functions that sit on 'i-sets' of X.[9] In particular, given a bounded function $f : X(i) \to [0,1]$,[10] we'd like to bound the probability the conditional density of f in a hyperedge is far from its expectation:

$$\mathop{\mathbb{P}}_{s\in\binom{[n]}{d}}\left[\left|\mathop{\mathbb{E}}_{t\subset s}[f(t)] - \mathbb{E}[f]\right| \geq \varepsilon\right] \stackrel{?}{\leq} \beta(\varepsilon, i, d).$$

Concentration for degree-i functions was first considered explicitly by Impagliazzo, Kabanets, and Wigderson (IKW) [215] for the complete complex $X = \binom{[n]}{d}$ as a core component in the construction of low soundness PCPs. For n sufficiently large, they proved the following generalization of Chernoff-Hoeffding to this setting

$$\mathop{\mathbb{P}}_{s\in\binom{[n]}{d}}\left[\left|\mathop{\mathbb{E}}_{t\subset s}[f(t)] - \mathbb{E}[f]\right| \geq \varepsilon\right] \lesssim e^{-\varepsilon^2\frac{d}{i}}. \tag{4}$$

Roughly speaking, the dependence $\frac{d}{i}$ corresponds to viewing a d-set in X as $\frac{d}{i}$ independent i-sets, and applying Chernoff to this system. In Section 5.11, we show (under weak assumptions) that this bound is indeed tight.

In fact, driven by the same desire to sparsify the complete complex for better hardness amplification, IKW actually also proved a variant of this bound for the *Grassmann* (the complex of subspaces). The details are too involved for this overview, but it suffices to say the resulting complex is *polynomial size* (independent of dimension), rather than size $\binom{n}{d}$, but comes with the catch that the associated tail bound is only inverse polynomial:

$$\mathop{\mathbb{P}}_{W\in Gr}\left[\left|\mathop{\mathbb{E}}_{V\subset W}[f(V)] - \mathbb{E}[f]\right| \geq \varepsilon\right] \lesssim O_i\left(\frac{1}{\sqrt{\varepsilon d}}\right).$$

[9] Note our choice of the word 'degree' is not just a moral connection; there is a literal sense in which these functions correspond to (bounded) degree-i terms in the Fourier decomposition of the complete complex and strong enough HDX.

[10] Formally $X(i) \subset \binom{[n]}{i}$ is the downward closure of X, the family of all i-sets that live in some original hyperedge.

As a result, while it is possible to perform hardness amplification [215, 129] via either of these bounds, the former exponentially blows up the *size* of the instance (as in parallel repetition), while the latter maintains polynomial size but exponentially blows up the *alphabet* due to its poor tail. *Are there hypergraphs that achieve the best of both worlds?*

This brings us back to the study of (non-partite) high dimensional expanders, which as sparse models of $\binom{[n]}{d}$, are a natural candidate for this problem. In fact, concentration of degree i functions was in some sense the main motivation behind Dinur and Kaufman's [124] original formalization of local-spectral expanders (though they called the property *double sampling*). Using spectral techniques, Dinur and Kaufman proved the following 'Chebyshev-type' tail bound for sampling degree-i functions on HDX:

$$\mathop{\mathbb{P}}_{s\in X}\left[\left|\mathop{\mathbb{E}}_{t\subset s}[f(t)]-\mathbb{E}[f]\right|\geq\varepsilon\right]\lesssim\frac{i}{\varepsilon^2 d}.$$

Their result, while hugely impactful, left open whether a stronger Chernoff-type bound could hold for HDX (or indeed for any sparse system at all).

In Chapter 5, we resolve this problem: HDX indeed satisfy optimal concentration of measure for degree-i functions. Namely for any $f : X(i) \to \mathbb{R}$:

$$\mathop{\mathbb{P}}_{s\in X}\left[\left|\mathop{\mathbb{E}}_{t\subset s}[f(t)]-\mathbb{E}[f]\right|\geq\varepsilon\right]\lesssim e^{-\varepsilon^2\frac{d}{i}}. \tag{5}$$

Combined with standard bounded degree constructions of high dimensional expanders [290, 238], this gives the first sparse hypergraph families with strong concentration for degree-i functions. Indeed to the best of our knowledge, no such objects asymptotically better than the complete complex itself were known prior to this bound.[11]

Equation (5) is powerful for smaller i, but our corresponding lower bounds show

[11]Note degree-i concentration, while a core component in PCPs, does not alone imply improved hardness amplification. This requires two other components: an agreement test, and a PCP embedding. We will discuss agreement testing in the next section overview. PCP embedding, that is 'embedding' an initial hard problem (base PCP) into the skeleton of an HDX to amplify, remains a major open problem.

one cannot hope for strong concentration in the regime where $i = \Theta(d)$. Unfortunately, the latter is actually a critical regime in application. Indeed IKW's PCPs based on Equation (4) suffered sub-optimal soundness due to exactly this issue: by setting $i = \sqrt{d}$, they could only amplify hardness to 1 vs. $\exp(-\sqrt{d})$, quasipolynomially far from the optimal bound 1 vs. $\exp(-d)$. This issue was subsequently fixed by Dinur and Livni-Navon [128], who observed that concentration in this regime can be replaced with a powerful analytic inequality from boolean function analysis called *reverse hypercontractivity*.

Reverse hypercontractivity is a property of the *noise operator* T_ρ, a classic smoothing operation in boolean analysis which given a hyperedge $s \in X$, samples a new correlated hyperedge s' by re-sampling each vertex in s with probability $1 - \rho$. A '*ρ-correlated pair*' $(s, s') \sim T_\rho$ is generated by drawing $s \in X$ uniformly at random, then drawing s' via the noise operator applied to s. Reverse hypercontractivity bounds the probability that a ρ-correlated pair passes between any two sets of hyperedges $A \subset X$ and $B \subset X$ by (a polynomial of) the product of their measure:

$$\mathop{\mathbb{P}}_{s,s' \sim T_\rho}[s \in A, s' \in B] \geq \mathbb{P}[A]^{O(1)}\, \mathbb{P}[B]^{O(1)}. \tag{6}$$

Note that if s and s' were totally independent, the above would of course be exactly $\mathbb{P}[A]\,\mathbb{P}[B]$. Thus reverse hypercontractivity also promises some sort of weak 'independence' type property, albeit of a different sort than discussed in the prior sections.

The reader may now reasonably wonder two things: why is this called reverse hypercontractivity, and what does this have anything to do with sampling? For the first, we refer the reader to the treatment in [308], and just note that the above can be re-phrased as a typical 'two-function hypercontractive form' as

$$\langle 1_A, T_\rho 1_B \rangle \geq \mathbb{P}[A]^{O(1)}\, \mathbb{P}[B]^{O(1)}.$$

While it is not immediately clear from our treatment of the Bonami Lemma above, standard hypercontractivity can actually be rephrased as a matching upper bound

$$\langle 1_A, T_\rho 1_B \rangle \leq \mathbb{P}[A]^{O(1)} \mathbb{P}[B]^{O(1)},$$

so Equation (6) indeed 'reverses' the standard hypercontractive inequality.

With regards to the second question, the key is to view our ρ-correlated pair (s, s') as generated through a slightly different random process. Let $j \sim \text{Bin}(\rho, d)$ denote the number of vertices *fixed* by the noise operator. Instead of drawing s uniformly, removing each vertex with probability $1-\rho$ to reach the 'fixed set' $t \subset s$, then re-sampling $s' \supset t$, we can instead samples (s, s') by *first* drawing t uniformly from $X(j)$, then *independently* $s, s' \supset t$. The probability that $s \in A$ and $s' \in B$, *given* t, is exactly the product of A and B's conditional measure within t, so

$$\mathop{\mathbb{P}}_{(t,s,s')}[s \in A, s' \in B] = \mathop{\mathbb{E}}_{t \sim X(j)} \left[\mathbb{P}[A|t] \, \mathbb{P}[B|t] \right].$$

In other words, reverse hypercontractivity holds exactly when a 'typical' j-set t for $j \approx (1-\rho)d$ 'sees' a good proportion of the sets A and B.

Prior to this work, reverse hypercontractivity was only known for hypergraphs satisfying a 'modified log sobolev inequality' [304], an entropy-based inequality that inherently relies on density of X. By instead leveraging the above viewpoint of reverse hypercontractivity as a form of sampling, we prove any hypergraph X with optimal concentration for degree-i functions in all links[12] is reverse hypercontractive.[13] Since we already proved above spectral HDX satisfy this condition, we get the first sparse families

[12]Formally we never defined this in the non-partite case. The link of an i-set s can either be viewed directly as the hypergraph $X_s\{t \in X(d-i) : s \cup t \in X\}$, or probabilistically by conditioning on $(X_1, \ldots, X_d)$ containing s, then marginalizing to the remaining unset variables.

[13]Note we cannot just directly apply concentration of X, since the bound for d vs. $(1-\rho)d$ is essentially trivial. Instead one draws the subset t in small parts $t = t_1 \cup \ldots \cup t_m$, and applies concentration of t_j at each step conditional on the prior t_i.

of reverse hypercontractive hypergraphs.

Finally, before we move on to applications of concentration, it is worth discussing what 'strong enough' high dimensional expansion really means in the above context. HDX are known to go through a *phase transition* at a certain expansion parameter we call the '*Trickling-Down (TD)-Threshold*', in particular when the (worst-case conditioning of) the marginals (X_i, X_j) has expansion $\frac{1}{d}$. Any complex X beating this bound immediately exhibits 'local-to-global' structure—in other words it is possible to infer *global* properties such as expansion of the skeleton of X [309] and fast-mixing of certain random walks [234, 239, 11] just from spectral behavior of the marginals. On the other hand, *at* the TD-threshold, no such structure can be inferred.[14]

With this in mind, for $\lambda \in [0, 1]$ we call a hypergraph X 'λ-TD' if all such marginals have expansion $\frac{\lambda}{d}$ (or better). We prove that any λ-TD complex satisfies concentration

$$\mathop{\mathbb{P}}_{s \in X}\left[\left|\mathop{\mathbb{E}}_{t \subset s}[f(t)] - \mathbb{E}[f]\right| \geq \varepsilon\right] \lesssim e^{-(1-\lambda)\frac{\sqrt{d}}{i}}$$

for degree-i functions. As $\lambda \to 1$, (the TD-Threshold), this bound becomes trivial. This is necessary: there exist hypergraphs at the threshold with arbitrarily poor concentration.

More generally, by directly bounding the moment generating function as a function of X's underlying expansion, we give an argument that interpolates between the above regime, called *sub-exponential* concentration, and full Chernoff (*sub-gaussian* concentration) when $\lambda \approx 2^{-d}$. In any regime, it is possible to recover optimal concentration by taking an appropriate 'skeleton' of X. E.g. in the weakest case, considering the family of k-sets of X for $k \leq \sqrt{d}$ will recover optimal (sub-gaussian) concentration at that level.

It is a critical open question whether λ-TD complexes (or some variant thereof) satisfy sub-gaussian concentration at their *top* level. This would have implications, for

[14] To our knowledge this statement does not appear explicitly in the literature, but it is fairly easy to infer from [170] and could reasonably be considered 'folklore'.

instance, for lower bounding the best *degree* of an HDX as we discuss below, and we expect may eventually be important in the construction of improved low soundness PCPs.

Chapters 5(b): Applications I (Agreement Tests).

Much of the motivation in studying global and reverse hypercontractivity comes from the study of *agreement testing*. Agreement tests are a core component of PCPs generalizing standard property tests like the line vs. plane test. For the sake of this overview, we consider a somewhat simplified setup. Given a hypergraph $X \subset \binom{[n]}{d}$, imagine we are given a family of 'local' assignments for each hyperedge mapping its vertices to $\mathbb{F}_2$:

$$\mathcal{F} \coloneqq \{f_s : s \to \mathbb{F}_2\}_{s \in X}.$$

We'd like to test whether these local assignments are actually an encoding of a *global* assignment $g : [n] \to \mathbb{F}_2$ on the vertices. In other words, is it the case that for most $s \in X$:

$$f_s \overset{?}{\approx} g|_s.$$

For context, this type of scenario arises naturally in hardness amplification. Think of the assignment to g as a solution to the original problem we'd like to amplify, and $\mathcal{F}$ is a solution to the (possibly derandomized) 'repeated' problem. In order to show the lifted problem is hard, we need to argue that a good solution $\mathcal{F}$ actually corresponds to an honest solution to the original problem. In other words, in the PCP context, we want to make sure the provers can't 'cheat' by correlating their answers across the hyperedge rather than 'independently' solving the problem d times.

Roughly speaking, an agreement test is any procedure which queries $O(1)$ local functions in $\mathcal{F}$ and outputs 'pass' or 'fail' constraint to the following guarantees:

1. **Completeness**: Any true global $\mathcal{F}$ passes with probability 1

2. **Soundness**: If $\mathcal{F}$ passes with good probability, $\mathcal{F}$ should be 'approximately' global

Formally, there are two main regimes of soundness in which agreement tests are studied: the 99% (or 'high acceptance' regime), where we assume the test passes with high probability, and the challenging 1% regime in which global structure should be inferred even when the test passes with only *non-trivial* probability:

$$\mathbb{P}[\text{Test Passes}] \geq \delta \implies \exists g, \mathbb{P}[f_s \approx g|_s] \geq \text{poly}(\delta).$$

Both regimes have been studied extensively in the literature (see e.g. [31, 168, 130, 116, 120, 212, 115, 299, 124, 109, 236, 41, 40, 110, 113, 114] among others). For now, we focus on the 1%-regime where such tests for the complete hypergraph and Grassmann lead to low-soundness PCPs [215, 129, 128, 299, 297] and the 2-to-2 Games Conjecture [126, 255].

For what values of δ can we hope to prove soundness of X? To see this, it is easiest to first consider what a typical agreement test actually looks like. Recall our goal is to check whether the local assignments $\{f_s\}$ come from the same global function. The only real strategy to do this is to randomly query a few hyperedges, and check whether they agree on their intersection. For instance, the simplest such test is just to draw two hyperedges intersecting in half their vertices. This is called the *V-Test*:

1. Randomly pick $s, s' \in X$ such that $|s \cap s'| = d/2$
2. Accept if $f_s|(s \cap s') = f_{s'}(s \cap s')$

One might reasonably come up with more involved ways to draw s and s' (or draw more than 2), but all agreement tests boil down to this type of check.

Let's now consider the probability a *random* function passes this test. No matter how we draw the queried hyperedges $\{s_i\}$, they are going to intersect on at most $O(d)$ vertices. A random assignment (which has essentially no correlation with any global function), will pass any such test with probability about $\exp(-O(d))$, since every individual intersection

has a 1/2 probability of passing. This tells us the best we can possible hope for is to prove soundness in the regime where $\delta \gg \exp(-O(d))$. In other words, if we pass with probability *better than random*, we'd like to infer global structure.

As a brief aside, we note the particular dependence $\exp(-O(d))$ is also critical from the standpoint of PCPs. This is because our *alphabet* is boolean functions over d-sets, and is therefore of size 2^d. Achieving inverse exponential soundness ensures the alphabet and soundness remain *polynomially related*. This is a necessary component, for instance, in the famous *sliding scale conjecture* [57], which asks for PCPs with inverse polynomial soundness and polynomial alphabet.

Let's now consider the basic setting $X = \binom{[n]}{d}$. While the 2-query 'V-test' discussed above hits a known barrier at $\frac{1}{k}$-soundness [120], IKW [215] realized this could be circumvented using the *Z-test*, a procedure which draws a third query s'' intersecting with s', and tests agreement on both pairs (s, s') and (s', s''). As discussed in the previous section, IKW's analysis of the Z-test relied on Equation (4) and therefore only achieved $\exp(-\sqrt{d})$ soundness, but this was improved to the optimal $\exp(-O(d))$ bound by Dinur and Livni-Navon [128] using reverse hypercontractivity.

While these tests can be used to construct matching low-soundness PCPs, they cannot achieve *sub-constant* soundness due to the size $|X| = n^d$ (again, setting $d \geq \omega(1)$ results in superpolynomial size). This motivates the study of *derandomized testers*. Can we construct *polynomial* or even *linear* size hypergraphs X satisfying optimal soundness? Currently, the best known derandomized test is IKW's Z-test for the Grassmann (subspace complex). As discussed in the previous section, it has *inverse polynomial* instead of inverse exponential soundness due to its poor concentration of measure.

High dimensional expanders are of course a natural candidate for such a test (indeed a 1% test for HDX was conjectured in Dinur and Kaufman's original work showing an HDX tester for the 99%-regime [124]). In Section 5.8, we make progress on this question on two fronts. First, we prove an optimal *local* soundness theorem for the V-test on any spectral

HDX. Roughly speaking, the statement says that if the V-test passes with probability $\delta \gg \exp(-O(d))$, then the family $\mathcal{F}$ is global on many of X's $\frac{d}{2}$-links. This is the typical core lemma used to prove exponential soundness of the Z-test [212, 128] — the third query is used to patch together these local solutions.

Unfortunately, there are known obstructions to this patching step on sparse complexes, and it is not in general possible on all spectral HDX [110, 41]. To circumvent this issue, we introduce a stronger assumption we call 'globalness'. In particular, we require that for most pairs (t, t') of $d/2$-sets in X, their union $t \cup t'$ is a d-set (full hyperedge) in the original hypergraph. Under this assumption, we complete the 'patching' argument and show the Z-test has inverse exponential soundness. We give many examples of such complexes beyond $\binom{[n]}{d}$, such as the full linear matroid and skeletons of many spin-systems studied in approximate sampling (see e.g. [232]). Unfortunately, we are not aware of any global complexes that are polynomial size. It is possible such objects exist, but they would have to look quite different from current HDX constructions.

Finally, we remark that quite recently, leveraging several of the concentration tools developed in Chapter 5, Dikstein, Dinur, and Lubotkzy [110, 113] constructed the first bounded-degree testers in the 1% regime, with soundness $\frac{1}{\log(d)}$ (similar results were given independently by Bafna, Lifshitz, and Minzer [41, 40] using different tools). It remains an open problem whether the tools they develop can be combined with our methods above to give sparse (or even bounded degree) testers with inverse exponential soundness.

Chapter 6: Applications II (Sum-of-Squares).

Even if we could prove an agreement test with optimal soundness, it would only lead to new PCPs if we could also 'embed' a base hard problem into the HDX to amplify. While it is a major open problem to give such an embedding, interestingly there do exist examples of 'hard problems' derived from the structure of HDX, just not in a traditional NP-sense. In 2020, Dinur, Filmus, Harsha, and Tulsiani (DFHT) [119] gave an explicit

construction of a hard family of 3-XOR instances sitting on the triangles of the infamous Ramanunan complexes [290], Lubotzky, Samuels, and Vishne's seminal construction of high dimensional expanders. However, instead of working in the classical NP-setting, DFHT showed a lower bound for these instances against a popular optimization framework known as the *Sum-of-Squares Semi-definite Programming Hierarchy* (SoS).

The SoS hierarchy is the most powerful known algorithmic framework for solving constraint satisfaction problems (indeed, famously, it is optimal for CSPs under the unique games conjecture [323]). Roughly speaking, for an n-variable CSP, the SoS hierarchy is broken into levels $t \in [n]$, each of which is an SDP relaxation of the original problem that 'sees' t variables at once and correspondingly runs in time $n^{O(t)}$. A family of CSPs $\{\Phi_n\}$ is 'hard' for SoS at the t-th level if there exists a constant $\alpha \in (0, 1)$ such that the SDP Value of the t-th relaxation of every instance is 1, but the true value (the maximum number of satisfiable constraints) is at most α. Such a bound, called an 'integrality gap', witnesses the fact that the SoS relaxation is *not* a $\frac{1}{\alpha}$-approximation algorithm for the problem.

In fact, constructing *explicit* examples of CSPs with a constant integrality gap for Sum-of-Squares was a relatively long-standing open problem in the area. Due to its power as an algorithmic paradigm, and the fact that we may prove *unconditional* lower bounds in the framework without relying on P vs. NP, there is a substantial amount of interest in understanding what types of problems are hard for SoS. The classical answer, dating back to Grigoriev [179] in the late 90s, is that *random* CSPs are hard for Sum-of-Squares, even fooling $\Omega(n)$-levels of the paradigm (this is asymptotically optimal, since n levels solves any CSP). While there has been a great deal of progress on the topic since Grigoriev's result (see e.g. [340, 354, 45, 90, 269] and citations therein), randomness has always played a key role. Constructing an explicit instance has typically been seen as a challenging task since it requires giving a short proof of unsatisfiability not captured by the powerful SoS proof system—no candidates were even known for such a task before DFHT [119].

Unfortunately, DFHT's instances, while explicit, actually don't solve this problem:

the issue is that while the instances are hard, they only fool $\sqrt{\log(n)}$ levels of SoS. While certainly still interesting, from an SoS viewpoint this was somewhat disappointing since it is easy to construct hard explicit instances against $\tilde{\Omega}(\log(n))$ levels simply by brute forcing over $\log(n)$ size instances [322, 355]. In Chapter 6, we resolve this issue by adapting DFHT's argument to a new form of non-simplicial high dimensional expanders we call 'small-set' HDX. Leveraging the breakthrough construction of quantum LDPC codes of [314, 277], we show explicit constructions of small-set HDX exist, therefore giving the first explicit XOR instances hard for $\Omega(n)$-levels of SoS, essentially matching the hardness of random instances.

We will not go into exact details about small-set high dimensional expansion here, but since we won't otherwise cover toplogical expansion in this overview we briefly discuss the idea and where it fits into the picture of SoS lower bounds. In fact, it has been known since Grigoriev's works [178, 179] that SoS bounds are closely related to graph expansion. Recall a XOR instance consists of n variables $\{x_i\}$ and m clauses $\{C_j\}$ of the form

$$C_j \coloneqq \{x_i + x_j + x_k = \beta_j\}$$

for $\beta_j \in \mathbb{F}_2$. We can view Φ as a pair of a bi-partite *constraint graph* (whose left-hand vertices are the variables, and right-hand side are the clauses connected in the obvious way) and a *target function* β on the clauses giving the literals. Grigoriev, and later independently Schoenebeck [340], proved that any instance Φ whose constraint graph is an expander in the sense that every (not too large) righthand neighborhood has many left neighbors (mod 2) is complete against $\Omega(n)$-levels of SoS. That is, regardless of target, the SDP Value of such an instance will be 1.

Since we have explicit constructions of such expanders [17], the difficulty in giving an explicit bound actually lies in finding an unsatisfiable target function. To do this, DFHT force the constraint graph to sit on a high dimensional complex, adding a corresponding

3rd layer that gives a series of parity checks between the constraints, allowing them to find an unsatisfiable target explicitly via linear algebra. This approach, however, presents a problem: by enforcing high dimensional structure, it turns out the original constraint graph can no longer be a small-set expander! In particular, high dimensionality introduces canonical families of non-expanding sets called 'boundaries' and 'co-boundaries' on the middle layer (these are basically the neighborhoods of vertices on the 1st and 3rd layer). The key, similar in some sense to global hypercontractivity from Chapters 1-4, is to show these local functions are the *only* obstruction to expansion.[15] We call complexes satisfying such a property small-set HDX.

Finally, we briefly mention some prior and subsequent work to the above. The discussed notion of expansion is a variant of *(co)-boundary expansion*, a classical notion of topological HDX introduced by Linial and Meshulam [283], and Gromov [182]. Unlike the standard version, it is critical in our context that we only require the notion for small sets,[16] and that we have expansion in both directions. A similar notion to the former (expansion for small sets) was also considered in the co-boundary direction for simplicial complexes in [230, 141, 237]. Since the publication of our work, small-set HDX have played a key role in building efficiently decodable quantum codes [123, 278, 279], and in the resolution of the NLTS conjecture [27]. Recently, Golowich and Kaufman [173] gave a variant of our construction which is *strongly* explicit, meaning the constraints and target function can be locally computed in polylogarithmic time.

Chapter 5(c): Applications III (Geometry and Combinatorics).

Moving away from hardness of approximation, in Chapter 5 we also give a number of further applications of our probabilistic machinery in analysis, geometry, combinatorics,

[15] In an SoS sense, this works because we carefully choose the target function to avoid any contradiction over these local boundaries. Interestingly, a random target in this case would actually be *easily* refuted by SoS, since the local sets would give SoS a short witness of unsatisfiability.

[16] For reasons we have not explained here, we need our complex to have non-trivial co-homology. True co-boundary expansion implies the vanishing of cohomology, so we cannot use complexes satisfying this notion.

and coding theory. Here we overview two of these applications of particular interest to the study of HDX, geometric overlap and degree lower bounds. We refer the reader to Section 5.2.6 for an overview of the remaining applications.

The *Geometric overlap property*, due to Gromov [182], is one of the earliest notions of high dimensional expansion. The idea is based on the following famous theorem of Boros and Füredi [74] and Bárány [50]. Pick any n points in $\mathbb{R}^{d+1}$ and draw every d-simplex between them. There must exist a point $q \in \mathbb{R}^{d+1}$ covered by a *constant* fraction of the simplices (here the constant depends on d, but not n). Instead of viewing this as a geometric property of $\mathbb{R}^{d+1}$, Gromov viewed the result as a property of the hypergraph $\binom{[n]}{d+1}$. In particular, any affine embedding of $\binom{[n]}{d+1}$ into $\mathbb{R}^d$ must have a point covered by a constant fraction c of its vertices. Hypergraphs satisfying this notion are now called 'c-geometric expanders' [288], or are said to have 'c-geometric overlap' [153].

Gromov [182] asked whether there exist *bounded degree* geometric expanders of every dimension. Before answering, it's worth discussing what this problem actually has to do with 'expanders' as we've defined them in the first place. In 1-dimension, it is fairly clear Gromov's problem (indeed even the stricter version called *topological* overlap[17]) is solved by any bounded degree family of expanders. If we embed an expander $G = (V, E)$ into the real line, we can simply take the median vertex as our point. Since roughly half the points of the graph are on each side, edge expansion[18] promises that a constant fraction of edges must go between the two sides (and therefore contain the median point as desired).

Given the above, it is reasonable to conjecture Gromov's problem should be resolved by some variant of high dimensional expander. Indeed, this was shown not long after by Fox, Gromov, Lafforgue, Naor, and Pach [153] who gave both optimal *randomized* geometric expanders, and an *explicit* family with constant (but sub-optimal) overlap

[17]Topological overlap relaxes the embedding to be continuous. It is a much more challenging notion, and is known only for a few bounded degree complexes [141, 240].

[18]A graph is an edge expander if every cut has proportionally many edges passing between it. By Cheeger's inequality [20], spectral and edge expansion are equivalent up to polynomial factors.

based on the Ramanujan complexes. In the following years, the connections between geometric overlap and high dimensional expanders were further developed in [316, 140, 309], eventually resulting in a proof that all sufficiently strong spectral HDX have constant geometric overlap. Nevertheless, the question of whether such constructions (or indeed *any* explicit construction) could achieve optimal overlap remained.

Leveraging a variant of high dimensional expander mixing (closely related to the Chernoff bounds discussed in the prior section), we resolve this problem, proving that every sufficiently strong spectral HDX has near-optimal geometric overlap. The idea is via a reduction to the complete complex, using tools in [153] to partition the vertices of X into parts that optimally cover some point q 'pretending' X were complete. Mixing then implies the fraction of simplices between parts in X approaches that of the complete complex, giving the desired covering of q.

We now turn our attention to the second topic: *what is the optimal degree of a high dimensional expander?* As we've discussed in the previous sections, degree, i.e. the number of hyperedges touching each vertex, is a critical parameter in application, controlling the blow-up incurred using the hypergraph as a gadget. On expander graphs, degree is a classical and well understood question. The Alon-Boppana theorem [307], which roughly states any family of λ-expanders have degree at least $\frac{2}{\lambda^2}$, is a foundational result of the field; graphs meeting this bound are called *Ramanujan*, the gold standard of expander constructions. Despite its central role, our understanding of degree in high dimensions is extremely poor. The best known constructions of λ-TD complexes have degree $\lambda^{\Theta(d^2)}$. Is this optimal? Could $\lambda^{o(d^2)}$ be achieved?

We take the first step toward answering this question. We prove that any sufficiently regular complex X which satisfies a Chernoff bound in all links must have degree $2^{\Omega(d^2)}$. Unfortunately, this does not resolve the above question, since as discussed in the prior section we can only prove sub-exponential concentration for λ-TD complexes, which instead implies a lower bound of $2^{\Omega(d)}$. However, we do get strong lower bounds for *skeletons*

of such objects. This is in contrast to product-based constructions of HDX [170] at the TD-Threshold which have degree $2^{O(k)}$ at every level $k \leq d$. Our bound implies the k-skeleton of any λ-TD complex must be at least $2^{\Omega(k^2)}$ when k is sufficiently smaller than d, exhibiting another strong phrase transition at the TD-Threshold.

Chapters 7-8: Applications IV (Algorithms).

Most of this dissertation focuses on developing tools and proving results related to understanding the *limitations* of computation. However, hardness and algorithms are two sides of the same coin, and it is frequently the case that objects used or developed for one turn out to be useful for the other. In this section, we overview two algorithmic results based on high dimensional expanders: an approximation algorithm for unique games defined on HDX, and efficient no-regret learning strategies for matrix games.

Unique games are a simple class of 2-CSPs that play a central role in hardness of approximation [250]. The past twenty years has seen a great deal of work both towards attempted proofs of NP-hardness of unique games (the so-called unique games conjecture) [126, 256, 252, 47, 255], and towards ruling out candidate hard instances or NP-hardness altogether by constructing algorithms for the problem [29, 293, 28, 48, 2, 37, 42]. One of the most classical and widely used results in the algorithmic side of unique games is that they are easy on expander graphs [29, 293]. In Chapter 7, building on work of Bafna, Barak, Kothari, Schramm, and Steurer [37], we extend this result HDX.

At outset, it is not really clear exactly what this means. High dimensional expanders are d-uniform hypergraphs; how does one define a 2-CSP on such an object? The answer is a subject we have touched on a few times in this overview, but never discussed explicitly, *high order random walks* [234]. Just as graphs are closely intertwined with the random walk along their edges, high dimensional expanders are closely tied to a series of correponding high order random walks on their hyperedges. One example of such a walk is the noise operator discussed in the prior sections. This generates a graph whose vertices are the

hyperedges of X, and whose edges are given by ρ-correlated pairs. Another example of a high order random walk is the 'down-up' walk, first studied on HDX by Kaufman and Mass [234], which moves between two hyperedges by removing a single random vertex, and re-sampling to reach a new hyperedge. Similarly, one could consider the 'up-down' walk on the $(d-1)$st level of X that first adds a vertex to reach a d-hyperedge, then removes one uniformly (note this indeed generalizes the lazy random walk on a graph).

We give an algorithm for approximating affine unique games on a variety of high order walks on HDX, including all of the given examples above. The run-time of our algorithm depends on the *spectral behavior* of the underlying random walk. We show that on a strong enough HDX, the spectrum of any high order walk is highly concentrated inside $k+1$ strips, each corresponding to a 'level' of the complex. Our guarantees then scale with the 'strip' threshold rank, the number of strips with large eigenvalues, generalizing in some sense prior algorithms for unique games based on standard threshold rank (the total number of large eigenvalues) [48]. Note that this is $k+1$ in the worst case (whereas the standard threshold rank would be $\text{poly}(n)$), so for constant dimensional HDX our algorithms run in polynomial time and give a constant approximation factor. Recently, Bafna and Minzer [42] removed the dimension dependence for some special cases such as the noise operator using our hypercontractive inequality from Chapter 1.

Moving away from the setting of CSPs, in Chapter 8 we study how the structure of high dimensional expansion can allow for efficient algorithms over exponential size structured online learning problems. In particular, we study a few classical problems in the setting of *matrix games*. A matrix game is a standard object in game theory and economics where 2 (or more) players P_1 and P_2 have access to a set of actions A_1 and A_2 and a reward matrix over $A_1 \times A_2$. In each round of the game, the players choose an action based only on prior rounds' information and receive the corresponding reward. The goal is to build efficient algorithms that minimize the player's *regret*, the difference from your actual reward to the reward of the best (single) possible action.

In many settings of interest, the action spaces A_i are really exponential in the problem parameters. A simple example of this is a *security game*. Player 1, the defender, has n servers/checkpoints/etc they need to defend, but can only pick d servers to defend at once. Player 2, the attacker, can pick any server to attack, and wins if Player 1 failed to defend this server (the servers may have different rewards or costs to attack). Strategies for this game are actually implemented in practice, e.g. at airports, to decide where to send security agents [350].

In the described game, it is clear that Player 1's action space is exactly the complete hypergraph $\binom{[n]}{d}$, and of size n^d. Is it possible nevertheless for the player to find a no-regret strategy in $\text{poly}(n, d)$, or even $\text{poly}(d, \log(n))$ time? There are a couple classic 'no-regret' strategies Player 1 could try to employ. Perhaps the most popular is the 'multiplicative weights update' (MWU) algorithm [285] and its variants. The idea of multiplicative weights is to play a *randomized* strategy weighted by the prior cost of each action. In particular, we build a distribution over actions by exponentially weighting them by their loss in the previous rounds. In the security game, the loss for a hyperedge is a weighted sum over its vertices, so the resulting distribution is of the form

$$\mathbb{P}[(v_1, \ldots, v_d)] \propto \prod_{i=1}^{d} e^{-L_i}$$

where L_i is the historic loss of vertex v_i. Of course, this distribution is still over an exponential size space; how can Player 1 play from this strategy efficiently?

The trick is to realize that while Player 1 may not be able to *exactly* sample from this distribution, they can do so *approximately* while still maintaining low regret. In particular, Player 1 can approximately sample from the MWU distribution by using the high order 'down-up' walk described above. Starting at an arbitrary hyperedge, they repeatedly apply the process of removing a random vertex, and re-sampling a new server conditioning on the above probabilities, which can be done in linear time (or even in

polylog(n) time in special cases). It is known that the down-up walk on the complete hypergraph mixes in poly($d, \log(n)$) steps given a distribution weighted in the above sense (as a product across its vertices, sometimes called a 'tilt' or 'external field').

In fact, this is actually possible for a broader family of hypergraphs, under a strong notion of high dimensional expansion called spectral independence under external fields [13, 22]. Thus any matrix game over such a space whose rewards have a *linear* structure as above has an efficient no-regret algorithm. This includes, for instance, matroid bases and classic games in the literature such as matroid congestion. Moreover, since no-regret algorithms are actually known to converge to approximate equilibria of the underlying game, one can use these algorithms to study the Nash (or Coarse Correlated Equilibria in the many player setting) of such games. In many settings, our HDX-based method results in the best known strategies for both no-regret learning and equilibrium computation for these classical problems.

Chapter 1

Hypercontractivity on HDX

1.1 Introduction

Introduced over 50 years ago today, *hypercontractivity* remains one of the most powerful tools in the analysis of boolean functions. Originally used to prove numerous landmark results on the discrete hypercube such as the KKL Theorem [226] and Majority is Stablest [302], the study of hypercontractivity has since seen a resurgence on extended domains such as the p-biased cube [245], slice [252], and Grassmannian [255]. Fascinatingly, these regimes all share a common thread: while hypercontractivity doesn't hold in general, it is satisfied for certain classes of *pseudorandom functions*. This recently discovered phenomenon has led to a slew of breakthroughs, most famously including the resolution of Khot's 2-2 Games Conjecture [255]. Unfortunately, the scope of these results is currently restricted, as all known proof techniques rely on product structure or other strong symmetries, and no unifying theory is known to exist.

In this work we take the first substantive step towards solving this issue with the introduction of a new theory of hypercontractivity for the general class of *high dimensional expanders* (HDX). HDX are a family of expanding complexes that have seen an explosion of work in recent years, leading to major breakthroughs across a number of areas including (among others) the recent construction of c3-LTCs and qLDPC codes [117, 315], and efficient approximate sampling for many important systems (e.g. for matroid bases

[25], independent sets [24], Ising models [22], and more). Our results lead to a new understanding of the structure of boolean functions on HDX, including a tight analog of the KKL Theorem, and a characterization of non-expanding sets similar to that used in the proof of 2-2 Games [255]. Proving such results previously seemed out of reach since HDX are very far from products, asymmetric, and can be quite sparse. To handle these challenges, we introduce a new set of tools including a new explicit Fourier decomposition and a local-to-global method for analyzing higher order moments. Interestingly, unlike previous ℓ_2-based techniques which apply equally across all types of expanding complexes, our methods rely crucially on the underlying HDX structure being *simplicial*. This suggests a new stratification of spectral HDX based upon their behavior *beyond the second moment*.

1.1.1 Contributions

Before jumping into a more detailed breakdown of our results, we start by giving an informal overview of our main contributions within the broader context of classical Fourier analysis and the theory of high dimensional expanders.

Classical Fourier Analysis:.

Classical Fourier Analysis on the discrete hypercube focuses on analyzing functions $f : \{0,1\}^n \to \mathbb{R}$ through their *Fourier Expansion*, a decomposition that breaks f into a series of orthogonal "level functions," each corresponding to the projection of f onto a certain eigenspace of the (noisy) hypercube graph.[1] At a basic level, a function's Fourier decomposition gives a nice method for understanding its *second moment*, since orthogonality allows one to move between this and the standard basis freely (a result usually known as *Parseval's Theorem*). On the other hand, in computer science, we are usually interested in analyzing the special class of *boolean functions* $f : \{0,1\}^n \to \{0,1\}$. These functions exhibit rich structure that Parseval's Theorem isn't equipped to capture—to understand them, we usually need to look beyond the second moment.

[1] More generally, these are the eigenspaces of the Hamming scheme.

Hypercontractivity, introduced in 1970 by Bonami [71] (and later independently by Beckner [54] and Gross [183]), is exactly the tool for the job. In its simplest form, hypercontractivity boils down to the statement that the fourth moment of low levels of the Fourier decomposition should behave nicely. Namely that the ith level of a boolean function f, denoted f_i, should satisfy:

$$\|f_i\|_4 \leq 2^{O(i)}\|f_i\|_2. \tag{1.1}$$

This deceptively simple observation, known in the above form as "Bonami's Lemma" [71], led to many landmark results including the KKL Theorem [226], noise-sensitivity of sparse functions [226], Friedgut's Junta Theorem [157], and Majority is Stablest [302]. What's more, hypercontractivity (and its resulting applications) actually extend beyond the hypercube. After KKL's seminal work, many authors studied extensions and applications of hypercontractivity [78, 349, 160, 157], but it wasn't until recently that tight analogues of Equation (1.1) were developed for general product spaces [245] (generalizing work of Bourgain [158] and Hatami [194]) as well as for other structured domains such as the symmetric group [147] and Grassmannian [255]. These extended domains differ from the hypercube in that they are only hypercontractive for special classes of pseudorandom functions, but are nevertheless responsible for an impressive set of applications including analogues of classical results, a variety of sharp threshold theorems [158, 245, 280, 247], and perhaps most famously the proof of the 2-2 Games Conjecture [253, 126, 125, 47, 252, 255]. Unfortunately, despite the stark similarities between these settings, no unified theory explaining the phenomenon exists. Further, all known techniques rely heavily on product structure or other strong forms of symmetry, which makes it difficult to approach the problem in more general settings.

Fourier Analysis on HDX:.

High dimensional expanders (HDX) are a class of robustly connected complexes that have seen an incredible amount of development and application throughout theoretical computer science in the past few years, most famously in coding theory [117, 315, 142, 221, 241, 244, 112, 220, 121] and approximate sampling [25, 11, 24, 95, 96, 94, 146, 218, 286, 66], but also in agreement testing [124, 109, 236], CSP-approximation [9, 38], and (implicitly) hardness of approximation [252, 255]. In this work, we study a central notion of high dimensional expansion called *two-sided local-spectral expansion*, originally developed by Dinur and Kaufman [124] to build sparse agreement testers. For simplicity, we'll often refer to these objects just as *local-spectral expanders*, but the reader should be aware we always refer to the two-sided variant, not the weaker one-sided variant commonly used in approximate sampling.

Interestingly, local-spectral expanders are actually known to admit a (nascent) theory of Fourier analysis [111, 239]. Initial works in this area have focused on the development and application of Fourier Decompositions and Parseval's Theorem, and while the existing theory does have a few interesting applications (e.g. an FKN theorem for HDX [161, 111], efficient CSP-approximation [9, 38]), it is subject to the same limitations as original second moment methods on the hypercube: they simply don't capture the richer structure of boolean functions. Let's consider a concrete and important example: the expansion of pseudorandom sets (an analog of "sparse functions are noise-sensitive" on the hypercube).[2] Traditionally proved via hypercontractivity, a variant of this result on the Grassmannian recently led to the resolution of the 2-2 Games Conjecture [255]. On the other hand, Bafna, Hopkins, Kaufman, and Lovett [38] showed that second moment methods cannot recover such a result. While they are able to recover some sort of characterization with these techniques, it necessarily decays as the dimension grows to

[2]The connection lies in the fact that the noise-sensitivity result can equivalently be phrased as saying that small sets on the noisy hypercube are expanding.

infinity, becoming trivial in the regime useful for hardness of approximation—if we want to do better, it appears we need a theory of hypercontractivity.

This is easier said than done: local-spectral expanders look nothing like any object previously known to satisfy hypercontractivity. They can be sparse, asymmetric, and very far from products. Moreover, there are no known techniques for analyzing local-spectral expanders beyond the second moment.[3] Even DDFH and KO's Fourier decompositions are intrinsically tied to second moment methods, since they are defined by linear algebraic manipulation of the standard inner product. Surprisingly, it turns out that these barriers are not inherent, and can be removed with the introduction of just two new tools: a *combinatorial* Fourier decomposition for HDX, and a new local-to-global method to replace reliance on product structure in the analysis of higher moments.

Our new decomposition is the natural analog of the standard Fourier decomposition on product spaces (often called the "orthogonal" or "Efron-Stein" decomposition). It is equivalent to old decompositions in an ℓ_2-sense (and therefore shares all relevant ℓ_2-based properties), but comes with a number of additional benefits: it has simple explicit and recursive forms, and it behaves nicely under restriction. This allows us to bring to bear much of the power of more traditional Fourier-analytic machinery, which often relies on these same properties. Historically, however, applying this machinery in a useful fashion has also required the underlying object to be a product, or to satisfy some other strong symmetry. Our second key observation is that while individual variables in a local-spectral expander may be highly correlated, they look independent *on average*. More concretely, this means that in the analysis of *expectations* (such as a higher moment), we are free to treat the underlying variables as independent even if they actually exhibit a very high level of correlation.

[3]We note that recent works in the sampling literature have considered entropic notions of high dimensional expansion, but the underlying assumptions are much stronger than local-spectral expansion.

Hypercontractivity on HDX:.

Leveraging these tools, we build a theory of hypercontractivity on HDX. Concretely, we prove that Equation (1.1) holds on local-spectral expanders for an appropriate notion of pseudorandom functions—ones that are not concentrated in any local restrictions on the complex.[4] Combined with BHKL's recent spectral analysis of *higher order random walks* (which, for the moment, we'll think of as analogues of the noisy hypercube graphs or Hamming scheme), this leads to the resolution of a number of open questions in boolean function analysis. To start, we provide a tight characterization of (edge) expansion on higher order random walks, which, unlike previous methods [38], *does not decay with dimension*. This matches the version of the result on the Grassmannian which led to the resolution of the 2-2 Games Conjecture [255], and opens yet another avenue towards the use of HDX in hardness of approximation. We also introduce natural analogues[5] of two classic Fourier-analytic notions: *influence* and the *noise operator*. Combining these with the above recovers tight variants of both the KKL Theorem and noise-sensitivity of sparse (or in this case pseudorandom) functions.

Beyond these concrete applications, hypercontractivity on HDX also has interesting implications in the broader context of discrete Fourier analysis and high dimensional expansion. For the former, our result gives the first general class of hypercontractive objects beyond products, and combined with bounded degree constructions [290, 238], the first example of hypercontractivity over any sparse object at all.[6] For the latter, our result suggests a new stratification among notions of local-spectral expansion. This requires some additional explanation. While local-spectral expanders were originally introduced only

[4]In the high dimensional expansion literature, these restrictions are known as *links*.

[5]When applied to the embedding of the hypercube into a simplicial complex, these definitions return the standard notions.

[6]Formally, it is more accurate to say 'locally sparse' or 'bounded degree' here. While previous settings such as highly imbalanced products may be sparse in the sense that most of their weight is concentrated on relatively few faces, they are not sparse in the much stronger sense of a bounded-degree HDX. The former, for instance, will always have some very dense restrictions, whereas every vertex in the latter sees only a tiny fraction of the full complex.

over simplicial complexes, they were quickly extended to more general settings such as the Grassmannian, or even to general ranked posets [111]. While these classes of local-spectral expanders are essentially equivalent in an ℓ_2 sense [111, 38, 243], our analysis of the fourth moment crucially relies on simplicial structure. We conjecture that this is an inherent rather than technical barrier: only special classes of underlying objects (e.g. Grassmannian, simplicial complexes) satisfy hypercontractivity, and thereby lead to the strongest known form of spectral high dimensional expanders.

1.2 Background

Before stating our results more formally, we give a quick overview of the theory of local-spectral expanders and higher order random walks. Local-spectral expansion is a robust notion of connectivity on weighted hypergraphs introduced by Dinur and Kaufman [124] in the context of agreement testing. As is standard in the area, we will view d-uniform hypergraphs $H \subseteq \binom{[n]}{d}$ as (pure) **simplicial complexes**:

$$X_H = X(0) \cup \ldots \cup X(d),$$

where $X(d) = H$, $X(i) \subseteq \binom{[n]}{i}$ is given by downward closure, and $X(0) = \emptyset$. We note that this notation is off by one from much of the HDX literature which considers $X(i) \subseteq \binom{[n]}{i+1}$. This notation is standard in the topological literature (where an i-simplex indeed as $i+1$ points), but is less natural for our purely combinatorial work.

Most work on high dimensional expansion is based on the *local-to-global paradigm*, in which local properties of a complex are lifted to a desired global property (e.g. mixing or agreement testing). The main local structure of interest are called **links**. For every "i-face" $\tau \in X(i)$, the link of τ is the subcomplex obtained by restriction to faces including τ:

$$X_\tau = \{\sigma \ : \ \sigma \cap \tau = \emptyset, \ \sigma \cup \tau \in X\}.$$

A simplicial complex is said to be a **γ-local-spectral expander** if (the graph underlying) every link is a γ-spectral expander.[7]

Higher order random walks are an analog of the standard walk on expander graphs that moves between two vertices via an edge. Kaufman and Mass [234] observed that this process can be applied at any level of a simplicial complex: one could move between edges via a triangle, or triangles via a pyramid. Formally, these walks are defined as a composition of **averaging operators**, objects that have become ubiquitous tools in the study of high dimensional expanders. Denote the space of functions $\{f : X(k) \to \mathbb{R}\}$ as C_k. For a function $f \in C_k$, the (level k) **Up** and **Down operators** lift and lower f to level $k+1$ and $k-1$ respectively by averaging:

$$U_k f(\tau) = \mathop{\mathbb{E}}_{\sigma \subset \tau}[f(\sigma)],$$
$$D_k f(\tau) = \mathop{\mathbb{E}}_{\sigma \supset \tau}[f(\sigma)].$$

It will often be useful to compose the down or up operators multiple times to move between levels k and i, we denote this by $D_i^k = D_i \circ \ldots \circ D_k$ and $U_i^k = U_k \circ \ldots \circ U_i$. Informally, HD-walks are simply affine combinations of composed averaging operators. For instance, the basic composition $N_k^i = U_k^{k+i} D_k^{k+i}$, called a **canonical walk**, is the random process which moves between two k-faces via a shared $(k+i)$-face.

1.3 Results

We now move to an informal description of our results. We view our work as having three main contributions. First, we introduce and develop a new theory of Fourier analysis on high dimensional expanders. This includes a new explicit Fourier decomposition, as well as a number of natural generalizations of Fourier-analytic ideas such as influence

[7] A graph is a γ-spectral expander if the second largest eigenvalue of its weighted adjacency matrix (also called the random walk matrix) is at most γ in absolute value.

and the noise operator to simplicial complexes. Second, we prove that our Fourier-analytic decomposition satisfies a hypercontractive inequality for the special subclass of *pseudorandom functions*, and use this fact to characterize the small set expansion of HD-walks and give a version of Bourgain's Theorem (an analog of KKL on product spaces) on HDX. Finally, en route to our hypercontractivity theorem, we introduce a new method of localization on high dimensional expanders of independent interest that enables local-to-global analysis of higher order moments.

1.3.1 The Bottom-Up Decomposition

We start with a discussion of our new explicit Fourier-analytic decomposition. All previously known Fourier bases on local-spectral expanders [239, 111] are linear algebraic in nature, and have no known closed form. While these decompositions certainly have their place and are sufficient for a number of interesting applications [111, 9, 38], they often fall short when finer-grained calculation is required. To alleviate this issue, we introduce a new combinatorial decomposition on simplicial complexes that is an analog of the classic orthogonal (sometimes called Efron-Stein) decomposition on product spaces.

Definition 1.3.1 (Bottom-Up Decomposition)**.** Let X be a d-dimensional pure simplicial complex and $f \in C_k$ any function. For all $0 \leq i \leq k$ and $\tau \in X(i)$, define the ith level function(s) to be:

$$g_{\uparrow i}(\tau) = \sum_{\sigma \subseteq \tau} (-1)^{|\tau \setminus \sigma|} \mathop{\mathbb{E}}_{X_\sigma}[f], \quad f_{\uparrow i} = \binom{k}{i} U_i^k g_{\uparrow i}.$$

One can check that $f = \sum_{i=0}^{k} f_{\uparrow i}$ (see Theorem 1.7.2).

Here, $g_{\uparrow i}(\tau)$ should be thought of as the contribution to f coming from τ (where contributions from $\sigma \subsetneq \tau$ have been removed by inclusion/exclusion). The Fourier level $f_{\uparrow i}$ is then defined by summing over these contributions. It is worth noting that the

Bottom-Up Decomposition also has a simple recursive form (see Theorem 1.7.2):

$$g_{\uparrow i} = D_i^k f - \sum_{j=0}^{i-1} \binom{i}{j} U_j^i g_{\uparrow j}.$$

In fact, it should be noted that while the consideration of this basis is new over general simplicial complexes, the above recursive form was first studied for the special case of the complete complex by [252]. There, the authors took advantage of the complex's near-product structure to show that the decomposition gives an (approximate) Fourier basis close to the eigendecomposition of f with respect to the well-studied Johnson graphs. We prove that the assumption of near-product structure is actually unnecessarily strong—it is enough for the underlying complex to be sufficiently expanding.

Theorem 1.3.2 (Bottom-Up Properties (Informal Lemma 1.7.6+Theorem 1.7.8))**.** *Let X be a two-sided γ-local-spectral expander, and M an HD-walk. Then for any $f \in C_k$, and $0 \le i < j \le k$:*

1. $\langle f_{\uparrow i}, f_{\uparrow j} \rangle \approx 0$
2. $\|f\|_2^2 \approx \sum_{i=0}^{k} \|f_{\uparrow i}\|_2^2$
3. $\exists \lambda_i \text{ s.t. } M f_{\uparrow i} \approx \lambda_i f_{\uparrow i}$

Theorem 1.3.2 is similar to an analogous result for the HD-Level-Set Decomposition in [111, Theorem 1.3]. We will cover their definition in greater detail in Section 1.7. For the moment, it suffices to note that their decomposition also breaks f into $k+1$ Fourier levels, which we similarly denote by $f = \sum_{i=0}^{k} f_{\downarrow i}$. It turns out that the similarities between the HD-Level-Set and Bottom-Up Decompositions are no accident—the two decompositions are actually close in ℓ_2-norm.

Theorem 1.3.3 (Bottom-Up Approximates HD-Level-Set (Theorem 1.7.8))**.** *Let X be a two-sided γ-local-spectral expander and $f \in C_k$. Then the Bottom-Up and HD-Level-Set*

Decomposition are close in ℓ_2-norm:

$$\|f_{\uparrow i} - f_{\downarrow i}\|_2^2 \leq 2^{O(k)}\gamma\|f\|_2^2.$$

Similarly,

$$|\langle f_{\uparrow i}, f_{\uparrow i}\rangle - \langle f_{\downarrow i}, f_{\downarrow i}\rangle| \leq 2^{O(k)}\gamma\|f\|_2^2.$$

The main advantage of the Bottom-Up Decomposition then lies in its simple explicit and recursive forms. In Section 1.7, we will see how these properties are useful for analyzing finer-grained structure like restriction that are often key to classical Fourier-analytic arguments. It is unknown how to analyze such properties for prior linear algebraic decompositions, and determining whether the latter share similar structure at this level remains an interesting open problem.

1.3.2 Hypercontractivity

Now that we have introduced our relevant Fourier-analytic decomposition, we turn our attention to the study of hypercontractivity. Hypercontractivity is one of the most powerful tools in boolean function analysis and is crucial to proving many of area's key results (e.g. KKL [226], FKN [161], Majority is Stablest [302], sharp threshold theorems [158], etc.). Informally, hypercontractivity can be thought of as a niceness condition on "low-degree" functions. We'll start by considering a simple variant often called the Bonami or Bonami-Beckner lemma, [71] which states that a "degree-i" function p should satisfy:

$$\|p\|_4 \leq 2^{O(i)}\|p\|_2.$$

Classically, we might think of p as being a degree-i polynomial, corresponding to the ith Fourier level of a boolean function. The corresponding statement in our context is therefore that the ith level of the Bottom-Up Decomposition should satisfy an analogous

inequality:

$$\|f_{\uparrow i}\|_4 \leq 2^{O(i)}\|f_{\uparrow i}\|_2. \tag{1.2}$$

Unfortunately, it is well known that Equation (1.2) cannot hold in our setting, even over the complete complex. However, it is possible that the inequality could hold for *natural subclasses* of functions. Indeed, such a phenomenon is known to occur on general product distributions [245], where **pseudorandom** functions satisfy a form of Equation (1.2).

Definition 1.3.4 (Pseudorandomness)**.** Let X be a simplicial complex and $f \in C_k$. We say f is (ε, i)-pseudorandom if it is sparse in every i-link in the following two senses:

1. For all $\tau \in X(i)$:
$$\left|\mathop{\mathbb{E}}_{X_\tau}[f]\right| \leq \varepsilon\|f\|_\infty$$
2. For all $\tau \in X(i)$:
$$\langle f|_\tau, f|_\tau\rangle \leq \varepsilon\|f\|_\infty^2$$

While the use of $\|f\|_\infty$ here may initially seem unnatural, it is in fact the appropriate scaling factor on a bounded-degree complex (at least up to constants). Namely since restrictions are of constant size, doubling the largest value in f leads to a $(1+\delta)$ multiplicative increase in density on links including that face for some constant $\delta > 0$.

In applications, we will often only care about non-negative functions, in which case the second condition can be removed completely (as it is implied by the first). We note that functions satisfying Definition 1.3.4 are also sometimes called *global* since they are not concentrated in any local structure [245, 280]. We call them pseudorandom in keeping with prior literature on the Johnson and Grassmann graphs [252, 255], and because they cannot be distinguished from an (ε-sparse) random function by examining density inside links. Finally, note that Definition 1.3.4 requires f to be sparse. We conjecture that our results should hold in the dense regime as well, and discuss this further in Section 1.4.

Hypercontractivity for restricted subclasses is still a very powerful tool. Keevash, Lifshitz, Long, and Minzer's result [245], for instance, led to the resolution of Majority is Stablest in the p-biased setting [280], and the resolution of several conjectures in extremal combinatorics as well [247]. While previous results to this effect were restricted by their reliance on product structure or strong symmetry, we show such assumptions are not necessary and prove an analogous form of hypercontractivity for HDX.

Theorem 1.3.5 (Hypercontractivity on HDX (Informal Theorem 1.8.1))**.** *Let X be a sufficiently strong two-sided γ-local-spectral expander and $f \in C_k$ an (ϵ, i)-pseudorandom function. Then the following hypercontractive inequality holds:*

$$\mathbb{E}[f_{\uparrow i}^4] \leq 2^{O(i)}\epsilon\,\mathbb{E}[f_{\uparrow i}^2]\|f\|_\infty^2 + c_k\gamma^{1/2}\varepsilon\|f\|_2^2\|f\|_\infty^2,$$

where $c_k \leq \min\{2^{O(k)}, k^{O(i)}\}$.

Crucially Theorem 1.3.5 is independent of k for small enough γ. This means our bounds remain meaningful even when k grows large (roughly speaking, one should think of the bound as being non-trivial in the regime where $k \ll \log(|X(1)|)$.[8] This was a crucial property in the analogous result on the Grassmann in the proof of the 2-2 Games Conjecture [255].

Our overall framework for proving Theorem 1.3.5 roughly follows Khot, Minzer, Moshkovitz, and Safra's [252] strategy for the complete complex. However, even with analogous results for the Bottom-Up Decomposition in hand, most of their techniques fail in our setting due to local-spectral expanders' distinct lack of product structure. In fact, Theorem 1.3.5 gives the first general class of hypercontractive objects beyond product spaces, and combined with known bounded degree constructions of local-spectral expanders [290, 238], the first example over any *sparse* domain at all. In Section 1.3.4, we'll discuss

[8] In reality, there is a more subtle trade-off here between the expansion parameters, degree, and dimension of the complex. The stated relation is for the complete complex where one optimizes expansion at the cost of degree.

how we tackle these traditionally hard-to-handle structures with the introduction of a new notion of average-case independence that relates closely to local-spectral expansion. Our method actually allows for analysis well beyond the 4th moment, and can also be used to extend Theorem 1.3.5 to 2-to-2q hypercontractivity (where the 4-norm is replaced by a higher $2q$-norm). We focus on the 2-to-4 case in this work for simplicity.

Before moving on to applications of Theorem 1.3.5, it is worth discussing another typical form of hypercontractivity and how it translates to the setting of simplicial complexes. Hypercontractivity is frequently expressed in terms of an object called the **noise operator**. On the hypercube, the noise operator T_ρ acts as an averaging process on boolean strings which replaces each coordinate with a random bit with probability $1-\rho$. In this context, hypercontractivity states that T_ρ should act as a *smoothing operator* in the following sense:

$$\|T_\rho f\|_4 \leq \|f\|_2 \tag{1.3}$$

for some constant ρ. Despite the fact that coordinates do not exist on a simplicial complex, there is still a natural analog of T_ρ where each vertex in a k-face is removed with probability $1-\rho$, and is then re-randomized over relevant k-faces. We formalize this procedure in terms of the averaging operators.

Definition 1.3.6 (Noise Operator)**.** Let X be a d-dimensional pure simplicial complex. The noise operator $T_\rho^k(X) : C_k \to C_k$ at level $k \leq d$ of the complex is:

$$T_\rho^k(X) = \sum_{i=0}^{k} \binom{k}{i} (1-\rho)^i \rho^{k-i} U_{k-i}^k D_{k-i}^k.$$

We write just T_ρ when clear from context.

When applied to the hypercube complex,[9] this natural analog returns exactly the standard boolean noise operator T_ρ. Combining standard arguments with the spectral

[9]The hypercube complex has vertex set $[n] \times \{0,1\}$, where the first entry stands for a coordinate and the second entry a value. The top level $X(n)$ consists of all binary strings and is exactly the hypercube.

properties of the Bottom-Up Decomposition, we can also prove a variant of Equation (1.3) for pseudorandom functions on HDX. To state this result, it will be useful to have a notion of degree: as on the hypercube, we say the degree of a function f is the largest i such that $f_{\uparrow i}$ is non-zero.

Corollary 1.3.7 (Informal Proposition 1.10.15)**.** *Let X be a sufficiently strong two-sided γ-local-spectral expander and $f \in C_k$ a degree i, (δ, i)-pseudorandom function for $\delta \leq \varepsilon \|f\|_2^2/\|f\|_\infty^2$. Then for some constant $\rho = \Theta(1)$:*

$$\|T_\rho f\|_4 \leq \epsilon^{1/4}\|f\|_2.$$

1.3.3 Applications

A classical application of hypercontractivity is to give what is known as a "level-i inequality" that bounds low-level weight of a boolean function. We can use Theorem 1.3.5 to give an analog on HDX for pseudorandom functions.

Theorem 1.3.8 (Level-i inequality (Informal Theorem 1.9.4))**.** *Let X be a two-sided γ-local-spectral expander with γ sufficiently small and $f \in C_k$ an (ϵ, i)-pseudorandom boolean function of density α. Then the weight on $f_{\uparrow i}$ is bounded by:*

$$\langle f_{\uparrow i}, f_{\uparrow i}\rangle \leq 2^{O(i)}\epsilon^{1/3}\alpha.$$

Level-i inequalities have a plethora of applications in boolean Fourier analysis. We'll look at the analog of two classical applications: one to small-set expansion, and the other to the structure of functions with low influence. Starting with the former, let's recall the basic definition of edge-expansion.

Definition 1.3.9. Let M be a walk on the kth level of a simplicial complex X. The (edge) expansion of a subset $S \subseteq X(k)$ is the average probability of leaving S in a single step of

the walk:

$$\Phi(S) = \mathop{\mathbb{E}}_{v\sim S}[M(v, X(k) \setminus S)],$$

where $M(v, X(k) \setminus S)$ is the probability the walk leaves S starting from v.

Informally, a walk is called a *small-set expander* if all small subsets expand. Traditionally, the level-i inequality on the discrete hypercube is used to show that the noisy hypercube graph is a small-set expander. The analogous result on simplicial complexes, however, isn't true: HD-walks (which generalize graphs like the noisy hypercube) have well-known examples of small non-expanding sets: *links* [252, 38]. Using Theorem 1.3.8, we can prove a converse to this result: *any* non-expanding set must be concentrated in a link.

Theorem 1.3.10 (Characterizing non-expansion on HD-walks (Informal Theorem 1.9.3))**.** *For every $0 < \delta < 1$, there exists some $\varepsilon > 0$ and $r \in \mathbb{N}$ such that for all large enough k the following holds. For any HD-walk*[10] *on a sufficiently strong two-sided local-spectral expander X and any subset $S \subseteq X(k)$, if S has expansion at most $\Phi(S) \leq \delta$, then S is concentrated in a low-level link:*

$$\exists i \leq r, s \in X(i) : \frac{|X_s \cap S|}{|X_s|} \geq \varepsilon$$

Expansion is also closely related to a well-studied Fourier-analytic quantity called **total influence**. On the boolean hypercube, the total influence of a function measures its total variability across each coordinate:

$$I[f] = \sum_{i=1}^{n} \mathop{\mathbb{P}}_{x\sim\{0,1\}^n}[f(x) = f(x \oplus e_i)]$$

[10]Formally, this statement only holds for HD-walks such as $N_k^{\Theta(k)}$ which exhibit sufficiently fast eigenvalue decay. We give a more general formulation in the main body that holds for all HD-walks (see Theorem 1.9.3).

where e_i is the ith standard basis vector. One of the most celebrated results in the analysis of boolean functions is the KKL Theorem [226], which states that any function with low total influence must have an influential coordinate. In domains beyond the hypercube (such as product spaces), total influence is usually instead written equivalently as:

$$I[f] = \langle f, Lf \rangle$$

where L is the (un-normalized) **Laplacian operator** (see Section 1.10 for more details). While the KKL Theorem does not hold over arbitrary product spaces,[11] a useful analog known as "Bourgain's Sharp Threshold Theorem" [158, Appendix] does. Bourgain's Theorem states that if a boolean function has small total influence, there must exist a link (on the hypercube a subcube) in which the function is much denser than expected.

We prove an analogous result for HDX. The Laplacian formulation of total influence has a natural generalization on simplicial complexes:

$$I_X[f] = \langle f, k(I - U_{k-1}D_k)f \rangle$$

that returns the standard definition over the hypercube complex (see Section 1.10). Using Theorem 1.3.8, we prove that any function with low total influence must be concentrated in a link.

Theorem 1.3.11 (Bourgain's Theorem for HDX (Informal Theorem 1.10.5))**.** *Let X be a sufficiently strong two-sided γ-local-spectral expander, and $f \in C_k$ a boolean function. Then for any $0 \leq K \leq k$, if $I[f] \leq K\,Var(f)$, there exists $i \leq K$ and an i-face τ such that the link of τ is dense:*

$$\mathop{\mathbb{E}}_{X_\tau}[f] \geq 2^{-O(K)}.$$

[11] More accurately, it does hold but decays with the minimum probability of any marginal, becoming trivial e.g. for the p-biased cube for small enough p.

Note that Theorem 1.3.11 is actually a bit weaker than Bourgain's Theorem in the sense that it only promises a link that is much denser than average when the function f is sparse. We conjecture that this result should hold in the dense regime as well (see Section 1.4 for details). On the other hand, unlike Bourgain's Theorem (which has a density increase of $2^{-O(K^2)}$ rather than $2^{-O(K)}$ for general functions), our result is tight.[12]

Proposition 1.3.12 (Bourgain's Theorem Lower Bound (Informal Proposition 1.10.6))**.** *Let $c \geq 1$ be any constant and $K > 1$ an integer. For all $K \ll k \ll n$, there exists a Boolean function $f \in C_k$ on the k-dimensional complete complex on n vertices satisfying:*

1. *The influence of f is small:*

$$I[f] \leq K\,Var(f).$$

2. *For every $i \leq cK$, all i-links are sparse:*

$$\forall i \leq cK, \tau \in X(i) : \underset{X_\tau}{\mathbb{E}}[f] \leq 2^{-\Omega(K)}.$$

1.3.4 Localization (Average Independence)

Our hypercontractive inequality is derived from a new method of localization on high dimensional expanders of independent interest. Localization itself is of course not new—indeed such techniques have recently become synonymous with HDX. However, most prior work in the literature focuses on the localization of *second moments*, whereas hypercontractivity requires the analysis of *higher moments*. Traditionally, analysis beyond the second moment is difficult on HDX due to an inherent lack of product structure. We show that this can often be circumvented by a new method of decorrelating variables.

Theorem 1.3.13. *Let X be a d-dimensional two-sided γ-local-spectral expander and $f \in C_k$. Then for any $j \leq d - k$ and $\tau \in X(j)$, the global and localized expectation of f*

[12] A similar tight version of Bourgain's Theorem for sparse functions on the p-biased cube was proved by [245].

over X_τ differ by an operator with small spectral norm:

$$\mathop{\mathbb{E}}_{X_\tau(k)}[f] - \mathop{\mathbb{E}}_{X(k)}[f] = \Gamma f(\tau)$$

where $\Gamma : C_k \to C_j$ satisfies $||\Gamma|| \leq O_{k,j}(\gamma)$.

We emphasize that the first expectation in this definition is given by *localizing* rather than *restricting* f. In other words we are averaging over k-faces in the link X_τ (which are $(k+j)$-faces in the original complex) rather than over k-faces in the original complex X that contain τ. Similar localization strategies to the above were also considered in the context of topological expansion by Kaufman and Mass [237].

Theorem 1.3.13 should really be thought of as saying that, on average, f can be decorrelated from "irrelevant" j-faces that don't appear in the input. This is particularly useful when analyzing objects like HDX with high correlation. To understand the technique a bit more concretely, let's look at a basic example application.

Let X be a γ-local-spectral expander. We will often be interested in analyzing certain expected products on X. For instructive purposes, let's take a look at an example of such a product with just two instances of some $g \in C_2$:

$$\mathop{\mathbb{E}}_{a\sim X(1)}\mathop{\mathbb{E}}_{b\sim X_a(1)}\mathop{\mathbb{E}}_{c\sim X_{ab}(1)}\big[g(a,b)g(a,c)\big] = \mathop{\mathbb{E}}_{a\sim X(1)}\mathop{\mathbb{E}}_{b\sim X_a(1)}\big[g(a,b)\mathop{\mathbb{E}}_{c\sim X_{ab}(1)}[g(a,c)]\big]. \tag{1.4}$$

Notice that if we were working over a product space, the distribution of $c \sim X_{ab}(1)$ would be the same as the distribution of $c \sim X_a(1)$. This allows us to significantly simplify the above:

$$\begin{aligned}\mathop{\mathbb{E}}_{a\sim X(1)}\mathop{\mathbb{E}}_{b\sim X_a(1)}\big[g(a,b)\mathop{\mathbb{E}}_{c\sim X_{ab}(1)}[g(a,c)]\big] &= \mathop{\mathbb{E}}_{a\sim X(1)}\left[\mathop{\mathbb{E}}_{b\sim X_a(1)}[g(a,b)]\mathop{\mathbb{E}}_{c\sim X_a(1)}[g(a,c)]\right] \\ &= \mathop{\mathbb{E}}_{a\sim X(1)}\left[\mathop{\mathbb{E}}_{b\sim X_a(1)}[g(a,b)]^2\right].\end{aligned}$$

On the other hand in an HDX (especially one of bounded degree), this could be far from true since b and c can be highly correlated. Theorem 1.3.13 provides a simple technique for circumventing this issue. Let $g|_a$ be the restriction of g to a, that is $g|_a(b) = g(a,b)$. Theorem 1.3.13 promises that

$$\mathop{\mathbb{E}}_{c\sim X_{ab}(1)}[g(a,c)] = \mathop{\mathbb{E}}_{c\sim X_a(1)}[g(a,c)] + \Gamma g|_a(b),$$

where $\|\Gamma\| \leq O(\gamma)$. This allows us to recover the same form as above up to $O(\gamma)$ error:

$$\begin{aligned}\mathop{\mathbb{E}}_{a\sim X(1)}\left[\mathop{\mathbb{E}}_{b\sim X_a(1)}[g(a,b)]\mathop{\mathbb{E}}_{c\sim X_{ab}(1)}[g(a,c)]\right] &= \mathop{\mathbb{E}}_{a\sim X(1)}\left[\mathop{\mathbb{E}}_{b\sim X_a(1)}[g(a,b)]^2\right] \\ &\quad + \mathop{\mathbb{E}}_{a\sim X(1)}\left[\mathop{\mathbb{E}}_{b\sim X_a(1)}[g(a,b)\cdot\Gamma g(b)]\right] \\ &\leq \mathop{\mathbb{E}}_{a\sim X(1)}\left[\mathop{\mathbb{E}}_{b\sim X_a(1)}[g(a,b)]^2\right] + O_g(\gamma),\end{aligned}$$

where we have ignored some terms in g for simplicity and the last step follows from an application of Cauchy-Schwarz and the spectral norm (see Section 1.8 for details).

We emphasize that while Equation (1.4) in particular could also have been analyzed through a more direct application of the swap walk, such techniques fail when additional copies of g are added. Since there are j copies of g in analysis of the jth moment, this means the traditional HDX tool kit cannot go beyond the second moment. On the other hand, our technique is applied individually to each copy of g, so it is essentially irrelevant how many times it appears in the product.

1.4 Discussion

Before getting into the details and formalization of the above, we take a moment to give a more careful treatment of some interesting open problems and related work.

1.4.1 Open Problems

Hypercontractivity, both on the cube and on extended domains, has led to an astounding number of applications since its introduction some 50 years ago. We recover just a small sample of these classical applications in our work, and believe the theory will give rise to further results in the analysis of boolean functions. However, rather than surveying a list of classical results one might wish to extend (we refer the reader to O'Donnell's book [308] for this), we'll instead focus on three open problems we feel are most directly raised by our work.

Perhaps the most obvious direction left open is to extend hypercontractivity to the *dense regime*. While our definition of pseudorandomness implicitly assumes the underlying function is sparse, we conjecture that all of our results should hold under a weaker notion of pseudorandomness that drops this assumption.

Definition 1.4.1 (Pseudorandomness (Dense Regime))**.** Let X be a simplicial complex and $f \in C_k$ a boolean function. We say f is (ε, i)-pseudorandom if its local and global average are close on every i-link:

$$\forall \tau \in X(i) : \left| \mathop{\mathbb{E}}_{X_\tau}[f] - \mathbb{E}[f] \right| \leq \varepsilon.$$

While the stronger notion we use in this work is certainly sufficient for some applications (e.g. characterizing expansion, noise-sensitivity) and is line with previous work [252, 255, 245], it does seem to fall short in other areas. A good example of this is our variant of Bourgain's Theorem. While our version only promises the existence of a dense link, the original result on product spaces actually promises a link with *higher than average density* (albeit by a factor of $2^{-O(K^2)}$ instead of $2^{-O(K)}$), which could be recovered by proving hypercontractivity for the above definition. More generally, proving hypercontractivity for this dense variant opens the door to a broader spectrum of applications than the sparse

regime alone can handle.

The second problem we'd like to discuss is more focused on the theory of high dimensional expanders itself. As mentioned in the introduction, local-spectral expansion can be extended well beyond simplicial complexes to many natural poset structures including the Grassmann poset [111, 243], where hypercontractivity was crucial to resolving the 2-2 Games Conjecture [255]. The spectral and ℓ_2-structure of these expanding posets (eposets) is well understood [111, 9, 38], and essentially has no dependence on the underlying poset structure.[13] In stark contrast, our results break down over general eposets at several key points. In fact, it seems likely that the Bottom-Up Decomposition is not even a Fourier basis (fails to satisfy Theorem 1.3.2) over general eposets, since the proof relies heavily on simplicial structure (see Lemma 1.7.5). On the other hand, variants of hypercontractivity are known for some special eposets such as the Grassmann poset. The key difference in these cases is that the definition of pseudorandomness necessarily changes. This raises a natural question: do all eposets satisfy hypercontractivity for some notion of pseudorandomness, or are structures like the Grassman poset and simplicial complexes "special"? We conjecture that the latter is the case, and that these objects represent a new, stronger class of spectral high dimensional expanders.

Our third proposed problem is not raised quite as directly by this work, but is hard to ignore in light of recent breakthroughs in approximate sampling via HDX [25, 11, 24, 95, 96, 94, 146, 218, 286, 66]. Hypercontractivity is classically connected to the *Log-Sobolev inequality*, which gives strong control over the mixing time of its associated random walk. Applied to the hypercube, for instance, this connection improves the standard spectral mixing bound from $O(n^2)$ to the optimal $\Theta(n \log(n))$ [107]. Recent analysis of entropic notions of high dimensional expansion and a *modified Log-Sobolev inequality* have led to a slew of analogous improvements on important sampling problems [96, 66, 22]. These results, however, usually only apply to dense objects and need stronger assumptions.

[13] Different poset parameters result in different eigenvalues, but the structure is otherwise the same.

Given these connections, it is natural to ask whether our theory of hypercontractivity can improve mixing times for general local-spectral expanders in some analogous fashion.

1.4.2 Related Work

Hypercontractivity on Extended Domains:.

Nearly 20 years after its introduction, Kahn, Kalai, and Linial [226] revolutionized the study of boolean functions with hypercontractivity. Not long after, a significant interest grew in the development and application of hypercontractivity beyond the hypercube, with a particular focus on product distributions and especially the p-biased hypercube [78, 349, 160, 157]. These works offered a general theory of hypercontractivity for such domains, but their strength depended on the underlying distributions in the product space. This issue was addressed to an extent in work of Friedgut and Bourgain [158], and later Hatami [194], who showed analogues to the KKL theorem in product spaces for certain pseudorandom functions, but it was not until the recent work of Keevash, Lifshitz, Long, and Minzer [245] (and independently O'Donnell and Zhao [364]) that a true hypercontractive inequality was developed in this setting. This offered the missing piece for a number of classical applications including a tight variant of the KKL Theorem (for monotone functions) [245], Majority is Stablest [280], as well as a number of interesting applications to extremal combinatorics [247].

Another line of work has examined hypercontractivity on what are often called "exotic" domains: specific objects beyond products such as the slice [252], multislice [148, 335], Grassmannian [255] (or similarly the degree-two short code [47]), and symmetric group [147]. Like KLLM's improved result for product distributions, in unbalanced settings these examples are only hypercontractive for pseudorandom functions.[14] The main application of this line of work has been to agreement testing and hardness of approximation. In particular, hypercontractivity for the Grassmannian was used to prove

[14]We note that higher degrees of the short code are also hypercontractive, but only on low Fourier levels for general functions [46].

the soundness of an agreement tester in the "1% regime" needed for the proof of the 2-2 Games Conjecture [253, 126, 125, 47, 252, 255]. It is worth noting that agreement testing theorems are also known for local-spectral expanders [124, 109, 236] (indeed the objects were originally introduced in this context). These results, however, lie in the "99% regime," so it is interesting to ask whether our theory of hypercontractivity can be used to build a bounded degree agreement tester in the more difficult 1% regime.

Finally, we should note that our overarching proof structure for hypercontractivity builds on KMMS' work on the slice (i.e. the complete complex). Their techniques, however, rely heavily on the fact that the slice is close in ℓ_1-distance to a product. This is far from true on local-spectral expanders, especially those of bounded degree which may essentially be as far as possible from products. As previously discussed, this lack of structure is a challenging barrier broken for the first time in this work (and independently in [187]).

Fourier Analysis on HDX:.

Fourier analysis on HDX was originally studied by Diksein, Dinur, Filmus, and Harsha [111], who introduced the HD-Level-Set Decomposition, analyzed its spectral properties, and used it to prove an FKN Theorem for HDX. A similar decomposition was also proposed around the same time by Kaufman and Oppenheim [239], though their work was more focused on understanding the spectral structure of higher order random walks than on developing a theory of Fourier analysis. In the years since, the HD-Level-Set Decomposition has seen some further development [9, 242, 38], and the nascent theory has helped build efficient approximation algorithms for certain k-CSPs [9] and unique games [38], but the restriction to second moment methods seems to have limited its use otherwise. Towards breaking this same barrier, Gur, Lifshitz, and Liu [187] have also (independently) developed a similar theory of hypercontractivity on local-spectral expanders. While their work certainly shares some connections to ours, its main proof techniques differ substantially and we believe the two works are of independent interest.

1.4.3 Roadmap

Having concluded introductory discussion of our work, we lay out a brief roadmap for the rest of the paper. In Section 1.5 we give preliminaries and formally define local-spectral expansion and higher order random walks. In Section 1.6 we discuss our new local-to-global method for higher moments that allows us to move beyond product distributions. In Section 1.7 we discuss our new explicit Fourier Decomposition, its basic properties, and behavior under restriction. In Section 1.8 we prove hypercontractivity for pseudorandom functions (Theorem 1.3.5). In Section 1.9 we apply this result to characterize edge expansion in HD-walks (Theorem 1.3.10). Finally in Section 1.10 we introduce analogues of classic Fourier analytic notions such as influence and the noise operator and use them to prove both a KKL Theorem (Theorem 1.3.11) and noise-sensitivity of pseudorandom functions.

1.5 Preliminaries

Before moving into proofs and further discussion of our main results, we take a moment to cover the theory of local-spectral expanders and higher order random walks in more detail.

1.5.1 Simplicial Complexes

Our main objects of interest in this work are a family of expanding hypergraphs known as **local-spectral expanders**. In this context, it will be useful to think of d-uniform hypergraphs as objects called **pure simplicial complexes**.

Definition 1.5.1 (Weighted, Pure Simplicial Complex)**.** A d-dimensional, pure simplicial complex $X = X(0) \cup \ldots \cup X(d)$ on n vertices is the downward closure of a hypergraph $X(d) = \binom{[n]}{d}$ where

$$X(i) = \left\{ s \in \binom{[n]}{i} \;\middle|\; \exists t \in X(d), s \subseteq t \right\}.$$

We call the elements of $X(i)$ i-faces. A weighted pure simplicial complex (X, Π) is a simplicial complex X endowed with a distribution Π over $X(d)$. This induces a distribution over each $X(i)$ by downward closure:

$$\pi_i(x) = \frac{1}{i+1} \sum_{y \in X(i+1): y \supset x} \pi_{i+1}(y), \tag{1.5}$$

where $\pi_d = \Pi$.

Weighted pure simplicial complexes are equivalent to weighted hypergraphs, and we will adopt the former viewpoint throughout the rest of this work. We note that our definition of dimension is off by one from some of the literature which adopts the convention that an i-face has $i+1$ vertices. While this is natural from a topological viewpoint, it makes less sense in our combinatorial context.

Weighted simplicial complexes also come equipped with a natural set of inner products. Recall that $C_i = C_i(X)$ denotes the space of functions $f : X(i) \to \mathbb{R}$. The distribution $\Pi = (\pi_d, \ldots, \pi_0)$ induces a natural inner product on each level:

$$\forall f, g \in C_i : \quad \langle f, g \rangle_{X(i)} = \mathop{\mathbb{E}}_{\tau \sim \pi_i} [f(\tau) g(\tau)].$$

When clear from context, we drop $X(i)$ from the notation. Just like on the hypercube, these associated products are a core component of function analysis and the development of Fourier analysis on HDX.

1.5.2 Local Spectral Expansion

In this work we focus on a recent spectral notion of high dimensional expansion called *two-sided local-spectral expansion* introduced by Dinur and Kaufman [124]. The definition hinges crucially on a form of local structure in simplicial complexes called **links**.

Definition 1.5.2 (Link)**.** Let (X, Π) be a d-dimensional weighted, pure simplicial complex. The **link** of an i-face $s \in X(i)$ is a $(d-i)$-dimensional pure simplicial complex given by the restriction of X to faces containing s, that is:

$$X_s = \{t \setminus s \in X \mid t \supseteq s\}.$$

We call X_s an i**-link**. Throughout the rest of the paper, X_s will always refer to its weighted version (X_s, Π_s) where Π_s is induced by the original distribution Π by normalizing over top level faces of X_s.

When analyzing a particular level k of the complex, we will often abuse notation and write X_s to mean the set of k-faces in X which contain s when clear from context.

Much of the high dimensional expansion literature centers around what is called the **local-to-global paradigm**, where properties on links are lifted to global properties on a complex. Local-spectral expansion can be seen as a definitional formalization of this notion: a complex is said to be expanding if all its local parts are expanding.

Definition 1.5.3 (Local-spectral expansion [124])**.** A weighted, pure simplicial complex (X, Π) is a two-sided γ-local-spectral expander if for every $i \leq d-2$ and every face $s \in X(i)$, the underlying graph[15] of X_s is a two-sided γ-spectral expander.[16]

1.5.3 Higher Order Random Walks

Just like expander graphs are inextricably tied to their underlying random walks, local-spectral expanders are similarly connected to an analogous set of random processes known as higher order random walks (HD-walks). In Section 1.2, we discussed one example of these objects called the canonical walks that move between k-faces via a shared $(k+i)$-face, and saw that these could be defined by the averaging operators. We'll now extend

[15]The underlying graph of a complex X is $G = (V = X(1), E = X(2))$.
[16]A weighted graph is a two-sided γ-spectral expander if $\max\{|\lambda_2|, |\lambda_n|\} \leq \gamma$.

these definitions to the more general setting of weighted simplicial complexes and, as well as define HD-walks in full generality. We'll start with the weighted averaging operators.

Definition 1.5.4 (Averaging Operators)**.** Let (X, Π) be a d-dimensional weighted, pure simplicial complex. For every $0 \leq k < d$, the **Up Operator** U_k lifts functions from C_k to C_{k+1} by averaging:

$$\forall \tau \in X(k+1) : U_k f(\tau) = \frac{1}{k+1} \sum_{\sigma \in X(k) : \sigma \subset \tau} f(\sigma).$$

Similarly, the **Down Operator** lowers functions from C_{k+1} to C_k by averaging:

$$\forall \tau \in X(k) : D_{k+1} f(\tau) = \frac{1}{\pi_{k+1}(X_\tau)} \sum_{\sigma \in X_\tau} \pi_{k+1}(\sigma) f(\sigma),$$

where $\pi_{k+1}(X_\tau) = \sum_{\sigma \in X_\tau} \pi_{k+1}(\sigma)$, and the sum is over $k+1$ faces of X containing τ.

It is worth noting that the averaging operators are *adjoint* with respect to the associated inner products mentioned in the previous section, that is for any $f \in X(i)$ and $g \in X(i-1)$:

$$\langle f, U_i g \rangle_{X(i)} = \langle D_i f, g \rangle_{X(i-1)} = \mathop{\mathbb{E}}_{(\sigma,\tau) \sim (\pi_i, \pi_{i-1})} [f(\sigma) g(\tau)].$$

This means that basic combinations of the operators such as the canonical walks discussed in Section 1.2 are *self-adjoint* and therefore have a spectral decomposition.

Let's now formalize the notion of higher order random walks. We'll start with a basic version called **pure walks** that are simply a composition of the averaging operators.

Definition 1.5.5 (Pure Walk [9])**.** Given a weighted, pure simplicial complex (X, Π), a k-dimensional pure walk $Y : C_k \to C_k$ on (X, Π) of height $h(Y)$ is a composition:

$$Y = Z_{2h(Y)} \circ \cdots \circ Z_1,$$

where each Z_i is a copy of D or U.

For the moment we won't force these walks to be self adjoint, but as we noted basic examples such as $N_k^i = D_k^{k+i} U_k^{k+i}$ do satisfy this constraint.

We define general higher order random walks to be any linear combinations of pure walks which is stochastic and self-adjoint.

Definition 1.5.6 (HD-walk [9])**.** Let (X, Π) be a pure, weighted simplicial complex, and $\mathcal{Y}$ a family of pure walks $Y : C_k \to C_k$ on (X, Π). We call a linear combination

$$M = \sum_{Y \in \mathcal{Y}} \alpha_Y Y$$

a k-dimensional HD-walk on (X, Π) as long as it is stochastic and self-adjoint. We call $w(M) \coloneqq \sum |\alpha_Y|$ the weight of M, and $h(M) = \max\{h(Y)\}$ its height.

1.5.4 Rectangular Swap Walks

Definition 1.5.6 only captures walks which stay on some fixed level of the complex. While these are certainly our main object of study, it turns out that in analysis it is often useful to consider *rectangular walks* which move between levels of the complex. We will be particularly interested in a rectangular walk introduced independently by Alev, Jeronimo, and Tusiani [9], and Dikstein and Dinur [109] called the *swap walk*. Informally, the swap walk from $X(i)$ to $X(j)$ moves from an i-face τ to a j-face σ through a shared $(i+j)$-face, but *swaps out all original elements in* τ. In other words, the intersection between τ and σ must be empty, and the shared $(i+j)$-face is exactly $\tau \cup \sigma$. To formalize this, it is useful to first introduce the more basic rectangular walk moving between $X(i)$ and $X(j)$ with no such restrictions.

Definition 1.5.7 (Rectangular Canonical Walks)**.** Let (X, Π) be a d-dimensional, pure, weighted simplicial complex. For any $i + j \leq d$, the rectangular canonical walk $N_{i,j}$ is the

natural operator moving between $X(i)$ and $X(j)$ through $X(i+j)$:

$$N_{i,j} = D_i^{i+j} U_j^{i+j}.$$

Swap walks are then defined by forcing the down steps in a canonical walk to remove only vertices in the initial face.

Definition 1.5.8 (Rectangular Swap Walks)**.** Let (X, Π) be a d-dimensional, pure, weighted simplicial complex. For any $i + j \leq d$, the rectangular swap walk $S_{i,j}$ is the (normalized) restriction of $N_{i,j}$ to pairs $(\tau, \sigma) \in X(i) \times X(j)$ such that $|\tau \cap \sigma| = 0$.

Swap walks appear naturally in a number of areas, including agreement testing [109], coding theory [221], and approximation algorithms [9, 38] and were well studied even before their formal introduction on HDX. On the complete complex, for instance, rectangular swap walks are exactly the bipartite Kneser graphs. Swap walks are particularly useful in these contexts because unlike their canonical counterpart canonical walks, they are actually great expanders.

Theorem 1.5.9 (Theorem 7.1 [109])**.** *Let (X, Π) be a d-dimensional two-sided γ-local-spectral expander. Then for any $i + j \leq d$, the spectral expansion of $S_{i,j}$ is at most:*

$$\lambda(S_{i,j}) \leq ij\gamma,$$

where $\lambda(S_{i,j})$ is the second largest singular value of $S_{i,j}$.

We note that this result was concurrently proved by AJT [9], albeit with a quantitatively worse bound.

1.6 Localization Beyond the Second Moment

Localization is one of the (if not the) most important technique in the analysis of high dimensional expanders. Classic results, often grouped together under the name *Garland's method*, show how global functions on simplicial complexes can be broken down into an average over local parts. There are two forms of Garland's method that will be relevant to our work. The first handles *restrictions* of a function f in C_k to any $\tau \in X(i)$, s.t. $f|_\tau \in C_{k-i}(X_\tau)$ satisfies:

$$\forall \sigma \in X_\tau(k-i) : f|_\tau(\sigma) = f(\tau \cup \sigma).$$

Lemma 1.6.1 (Garland's method (restrictions) [239])**.** *Let (X, Π) be a weighted, pure simplicial complex, and $f \in C_k$. Then for any $i \leq k$, $\|f\|^2$ is equal to its average second moment restricted to i-links:*

$$\langle f, f \rangle = \mathop{\mathbb{E}}_{\tau \in X(i)} [\langle f|_\tau, f|_\tau \rangle].$$

The second form of interest handles *localizations* of f to $\tau \in X(i)$, where $f_\tau \in C_k(X_\tau)$ lifts f from $X(k)$ to $X_\tau(k)$:

$$\forall \sigma \in X_\tau(k) : f_\tau(\sigma) = f(\sigma).$$

Lemma 1.6.2 (Garland's method (localizations) [309])**.** *Let (X, Π) be a d-dimensional, weighted, pure simplicial complex, and $f \in C_k$. Then for any $k + i \leq d$, $\|f\|_2^2$ is equal to*

its average second moment localized to i-links:

$$\langle f, f \rangle = \mathop{\mathbb{E}}_{\tau \in X(i)} [\langle f_\tau, f_\tau \rangle].$$

Garland's method will play an important role in the analysis of Theorem 1.3.5, but the results are generally only useful once a problem has been reduced to analyzing second moments. Since we are mainly interested in hypercontractivity and analyzing higher moments, Garland's method alone won't be sufficient.

To this end, we introduce a new technique for analyzing higher moments on two-sided local-spectral expanders. At its core, the strategy relies on a deceptively simple observation: the difference between the global expectation of f and its localized expectation over links is exactly given by an application of the swap walk minus its stationary operator. We note that a related application of the swap walk was recently implemented in the work on group independent cosystolic expansion by [237].

Lemma 1.6.3. *Let (X, Π) be a d-dimensional pure, weighted simplicial complex and $f \in C_i$. Then for any $v \in X(j)$ such that $i + j \leq d$, we have:*

$$\mathop{\mathbb{E}}_{X_v}[f_v] - \mathbb{E}[f] = (S_{j,i} - U_0^j D_0^i) f(v).$$

Proof. This is essentially immediate from expanding the left-hand side. We have:

$$\begin{aligned}
\mathop{\mathbb{E}}_{X_v}[f_v] - \mathbb{E}[f] &= \sum_{w \in X(i)} \pi_{v,i}(w) f(w) - \sum_{w \in X(i)} \pi_i(w) f(w) \\
&= \sum_{w \in X(i)} \left(\pi_{v,i}(w) - \pi_i(w)\right) f(w) \\
&= (S_{j,i} - U_0^j D_0^i) f(v),
\end{aligned}$$

where $\pi_{v,i}(w) = 0$ for any $w \notin X_v(i)$. □

We typically refer to moving from $\mathop{\mathbb{E}}_{X_v}[f_v]$ to $\mathbb{E}[f]$ in this manner as *de-correlating*. Lemma 1.6.3 is particularly powerful on two-sided local-spectral expanders, since the spectral norm of $||S_{j,i} - U_0^j D_0^i||$ is small on every link by Theorem 1.5.9.

Corollary 1.6.4. *Let (X, Π) be a d-dimensional two-sided γ-local-spectral expander, $f \in C_i$, and $\tau \in X(\ell)$ for any $\ell < i$. Then for any $v \in X_\tau(j)$ such that $i + j - \ell \leq d$, the global and localized expectations of f differ by:*

$$\mathop{\mathbb{E}}_{X_{\tau \cup v}}[f|_\tau] - \mathop{\mathbb{E}}_{X_\tau}[f|_\tau] = \Gamma f|_\tau(v)$$

where $\Gamma : C_{i-\ell}(X_\tau) \to C_j(X_\tau)$ is an operator with spectral norm at most $||\Gamma|| \leq (i-\ell)j\gamma$.

Proof. Applying Lemma 1.6.3 to the restricted function $f|_\tau$ on X_τ (which is also a two-sided γ-local-spectral expander), we have that the left-hand side is exactly given by $(S_{j,i-\ell} - U_0^j D_0^{i-\ell})_\tau f|_\tau(v)$. Since $U_0^j D_0^{i-\ell}$ is the stationary operator of $S_{j,i-\ell}$, the spectral norm

$$\|S_{j,i-\ell} - U_0^j D_0^{i-\ell}\|$$

is exactly the second largest singular value of $S_{j,i-\ell}$. As discussed in Theorem 1.5.9, Dikstein and Dinur [109, Theorem 7.1] proved that this quantity is at most $(i-\ell)j\gamma$ on any two-sided γ-local-spectral expander. □

1.7 The Bottom-Up Decomposition

In this section, we introduce the Bottom-Up Decomposition, an explicit combinatorial decomposition on simplicial complexes which approximates Dikstein, Dinur, Filmus, and Harsha's HD-Level-Set Decomposition [111]. This is particularly useful since the latter decomposition essentially corresponds to the eigenspaces of HD-walks [111, 9, 38]. It will be convenient to first introduce an equivalent recursive form.

Definition 1.7.1 (Level Functions (Recursive Form))**.** Let (X, Π) be a d-dimensional pure simplicial complex and $f \in C_k$ any function. The ith level function of the Bottom-Up Decomposition is given by

$$g_{\uparrow i} = D_i^k f - \sum_{j=0}^{i-1} \binom{i}{j} U_j^i g_{\uparrow j}$$

The Bottom-Up Decomposition is given by lifting the level functions via the up operator.

Theorem 1.7.2 (Bottom-Up Decomposition (Explicit Form))**.** *Let (X, Π) be a d dimensional pure simplicial complex and $f \in C_k$ any function. Let $f_{\uparrow i} = \binom{k}{i} U_i^k g_{\uparrow i}$ be the lift of the ith level function to C_k. Then the following statements hold:*

1. *The lifted level functions give a decomposition of f:*

$$f = \sum_{i=0}^{k} f_{\uparrow i}$$

2. *The lifted level functions have the following explicit form:*

$$f_{\uparrow i} = \binom{k}{i} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} U_k^i D_j^k f,$$

or equivalently for all $\tau \in X(i)$ and $T \in X(k)$:

$$g_{\uparrow i}(\tau) = \sum_{\sigma \subseteq \tau} (-1)^{|\tau \setminus \sigma|} \mathop{\mathbb{E}}_{X_\sigma}[f], \quad f_{\uparrow i}(T) = \sum_{\sigma \in X(i) : \sigma \subset T} g_{\uparrow i}(\sigma)$$

Proof. (1) can be proved directly by the explicit form given in (2):

$$\begin{aligned} \sum_{i=0}^{k} f_{\uparrow i} &= \sum_{i=0}^{k} \binom{k}{i} U_i^k \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} U_j^i D_j^k f \\ &= \sum_{j=0}^{k} \left(\sum_{i=j}^{k} (-1)^{i-j} \binom{k}{i} \binom{i}{j} \right) U_j^k D_j^k f \end{aligned}$$

$$= U_k^k D_k^k f$$
$$= f.$$

where we've used the fact that:

$$\sum_{i=j}^{k} (-1)^{i-j} \binom{k}{i} \binom{i}{j} = \delta_{jk}.$$

It is left to prove (2). We proceed by induction. Note that the equality clearly holds for $i = 0$, where both sides are simply the global expectation $\mathbb{E}[f]$. Now assume by induction that the equivalence holds up to $i - 1$. We then have:

$$\begin{aligned}
g_{\uparrow i} &= D_i^k f - \sum_{j=0}^{i-1} \binom{i}{j} U_j^i g_{\uparrow j} \\
&= D_i^k f - \sum_{j=0}^{i-1} \binom{i}{j} U_j^i \sum_{\ell=0}^{j} (-1)^{j-\ell} \binom{j}{\ell} U_\ell^j D_\ell^k f \\
&= D_i^k f - \sum_{\ell=0}^{i-1} \left(\sum_{j=\ell}^{i-1} (-1)^{j-\ell} \binom{i}{j} \binom{j}{\ell} \right) U_\ell^i D_\ell^k f \\
&= D_i^k f - \sum_{\ell=0}^{i-1} (-1)^{i-1-\ell} \binom{i}{\ell} U_\ell^i D_\ell^k f \\
&= \sum_{\ell=0}^{i} (-1)^{i-\ell} \binom{i}{\ell} U_\ell^i D_\ell^k f.
\end{aligned}$$

The explicit form of $f_{\uparrow i}$ then follows simply by applying $\binom{k}{i} U_i^k$ to $g_{\uparrow i}$, and the equivalent form is immediate from the definition of the down and up operators. □

1.7.1 Bottom-Up vs. HD-Level-Set

Dikstein, Dinur, Filmus, and Harsha's HD-Level-Set Decomposition [111] is an elegant linear-algebraic decomposition for functions on local-spectral expanders. Like the Bottom-Up Decomposition, it breaks $f \in C_k$ down into $k + 1$ Fourier levels, but differs in

that it does so in a *top-down* fashion.

Theorem 1.7.3 (HD-Level-Set Decomposition, Theorem 8.2 [111])**.** *Let (X, Π) be a d dimensional two-sided γ-local-spectral expander, $\gamma < \frac{1}{d}$, $0 \leq k \leq d$, and let:*

$$H^0 = C_0, H^i = Ker(D_i), V_k^i = U_i^k H^i.$$

Then:

$$C_k = V_k^0 \oplus \ldots \oplus V_k^k.$$

In other words, every $f \in C_k$ has a unique decomposition $f = \sum f_{\downarrow i}$ such that $f_{\downarrow i} = U_i^k g_{\downarrow i}$ for $g_{\downarrow i} \in Ker(D_i)$.

While the HD-Level-Set Decomposition is certainly useful in its own right, it has no known explicit form. This can make the analysis of standard Fourier-analytic techniques like restriction difficult, and hampers the analysis of higher moments. We will show that the Bottom-Up Decomposition provides an explicit approximation of the HD-Level-Set Decomposition that circumvents these issues while maintaining the latter's useful properties.

Before jumping into the details, however, it is worth reviewing an elegant technical tool of [111] that will be crucial for our analysis. They prove that local-spectral expansion is equivalent to a global notion of spectral expansion on complexes that relates the upper and lower walks.

Theorem 1.7.4 (DDFH Claim 8.8)**.** *Let (X, Π) be a d-dimensional two-sided γ-local-spectral expander. Then for any $1 \leq i \leq j \leq d$:*

$$D_i U_j^i = \frac{j}{i} U_{j-1}^{i-1} D_j + \frac{i-j}{i} U_j^{i-1} + E_{i,j}, \tag{1.6}$$

where $\|E_{i,j}\| \leq (i-j)\gamma$.

It is worth noting that this result (and the HD-Level-Set Decomposition) hold more generally for any "expanding poset"—the difference lies in the exact coefficients in the above relation.

The crucial observation for proving that the Bottom-Up Decomposition is a Fourier basis (that explicitly approximates the HD-Level-Set Decomposition) is that while $g_{\uparrow i}$ may not lie directly in $\mathrm{Ker}(D_i)$ like $g_{\downarrow i}$, it is fairly close to doing so. Proving this actually relies crucially on the exact coefficients in Equation (1.6) which correspond to working over a simplicial complex. As a result, it is not clear the Bottom-Up Decomposition is a Fourier basis at all for general expanding posets.

Lemma 1.7.5. *Let (X, Π) be a two-sided γ-local-spectral expander, $f \in C_k$, and $g_{\uparrow i}$ be given as in the Bottom-Up Decomposition. Then:*

$$\|D_i g_{\uparrow i}\|_2 \leq 2^{O(i)}\gamma \|D_i^k f\|_2.$$

Proof. The result follows from directly expanding $D_i g_{\uparrow i}$:

$$\begin{aligned} D_i g_{\uparrow i} &= \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} D_i U_j^i D_j^k f \\ &= \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} \left(\frac{j}{i} U_{j-1}^{i-1} D_j + \frac{i-j}{i} U_j^{i-1} + E_{i,j} \right) D_j^k f. \end{aligned}$$

The key is then to notice that the main terms cancel. That is, setting $c_{i,j} = (-1)^{i-j}\binom{i}{j}$ we have:

$$\sum_{j=0}^{i} c_{i,j} \left(\frac{j}{i} U_{j-1}^{i-1} D_j + \frac{i-j}{i} U_j^{i-1} \right) D_j^k f = \sum_{j=0}^{i-1} \left(\frac{j+1}{i} c_{i,j+1} + \frac{i-j}{i} c_{i,j} \right) U_j^{i-1} D_j^k f = 0.$$

Finally by the triangle inequality we have:

$$\|D_i g_{\uparrow i}\|_2 \leq \sum_{j=0}^{i} \binom{i}{j} \|E_{i,j} D_j^k f\|_2 \leq \sum_{j=0}^{i} \binom{i}{j} (i-j)\gamma \|D_j^k f\|_2$$
$$\leq O(i 2^i \gamma) \|D_i^k f\|_2$$

where we have used the fact that D contracts ℓ_2-norm in the final step. □

We will also need approximate orthogonality of both decompositions.

Lemma 1.7.6. *Let (X, Π) be a two-sided γ-local-spectral expander. Then the following three approximate orthogonality relations hold for all $i \neq j$:*

$$|\langle f_{\downarrow i}, f_{\downarrow j} \rangle| \leq 2^{O(k)} \gamma \|f\|_2^2 \tag{1.7}$$

$$|\langle f_{\downarrow i}, f_{\uparrow j} \rangle| \leq 2^{O(k)} \gamma \|f\|_2^2 \tag{1.8}$$

$$|\langle f_{\uparrow i}, f_{\uparrow j} \rangle| \leq \min\{k^{O(i+j)}, 2^{O(k)}\} \gamma \|f\|_2^2 \tag{1.9}$$

Proof. All relations follow from Equation (1.6) and Lemma 1.7.5. The first relation is proved in [111]. The latter relations follow similarly, but we give the third for completeness. In particular, assuming $i > j$ we have:

$$\begin{aligned}
|\langle f_{\uparrow i}, f_{\uparrow j} \rangle| &= \left| \binom{k}{i} \binom{k}{j} \langle U_i^k g_{\uparrow i}, U_j^k g_{\uparrow j} \rangle \right| \\
&= \binom{k}{i} \binom{k}{j} \left| \langle D_{i-1}^k U_i^k g_{\uparrow i}, U_j^{i-1} g_{\uparrow j} \rangle \right| \\
&\leq \binom{k}{i} \binom{k}{j} \|D_{i-1}^k U_i^k g_{\uparrow i}\|_2 \|U_j^{i-1} g_{\uparrow j}\|_2 \\
&\leq \binom{k}{i} \binom{k}{j} \|D_{i-1}^k U_i^k g_{\uparrow i}\|_2 \|g_{\uparrow j}\|_2
\end{aligned}$$

where we have applied Cauchy-Schwarz and used the fact that averaging operators contract ℓ_2-norm. We now separately bound both norms. The second is the simpler of the two,

and we claim it is at most $2^{O(i)}\|f\|$. In fact a more general claim holds.

Claim 1.7.7. For any ℓ_p-norm, we have:

$$\|g_{\uparrow i}\|_p \leq 2^i \|D_i^k f\|_p$$

Proof. This follows from direct expansion of the ℓ_p-norm:

$$\begin{aligned}\|g_{\uparrow i}\|_p &\leq \sum_{j=0}^{i} \binom{i}{j} \|U_j^i D_j^k f\|_p \\ &\leq \sum_{j=0}^{i} \binom{i}{j} \|D_i^k f\|_p \\ &= 2^i \|D_i^k f\|_p\end{aligned}$$

where we have applied the triangle inequality and the fact that U and D contract p-norms. □

Plugging this back into the above, we have:

$$\langle f_{\uparrow i}, f_{\uparrow j}\rangle \leq 2^i \binom{k}{i}\binom{k}{j} \|f\|_2 \|D_{i-1}^k U_i^k g_{\uparrow i}\|_2.$$

To complete the proof, it is therefore enough to argue that $\|D_{i-1}^k U_i^k g_{\uparrow i}\|_2$ is at most $c_k\gamma\|f\|$ for some small enough c_k. This follows from $k-i$ repeated applications of Equation (1.6) (one for each instance of the up operator). Informally, each application of Equation (1.6) incurs two error terms, one stemming from the matrix $E_{i,j}$, and the other from $\frac{j}{i}U_{j-1}^{i-1}D_i$ applied to $g_{\uparrow i}$, which we know is small by Lemma 1.7.5. The final remaining term is then proportional to the original term, but with one DU pair removed. For instance, for the first application we have:

$$\|D_{i-1}^k U_i^k g_{\uparrow i}\| = \|D_{i-1}^{k-1}(D_k U_i^k) g_{\uparrow i}, U_j^{i-1} g_{\uparrow j}\|$$

$$
\begin{aligned}
&= \|D_{i-1}^{k-1}\left(\frac{i}{k}U_{i-1}^{k-1}D_i + \frac{k-i}{k}U_i^{k-1} + E_{k,i}\right)g_{\uparrow i}\| \\
&= \frac{k-i}{k}\|D_{i-1}^{k-1}U_i^{k-1}g_{\uparrow i}\| + \frac{i}{k}\|D_{i-1}^{k-1}U_{i-1}^{k-1}D_i g_{\uparrow i}\| + \|D_{i-1}^{k-1}E_{k,i}g_{\uparrow i}\| \\
&\leq \frac{k-i}{k}\|D_{i-1}^{k-1}U_i^{k-1}g_{\uparrow i}\| + 2^{O(i)}\gamma\|f\| + k\gamma\|g_{\uparrow i}\| \\
&\leq \frac{k-i}{k}\|D_{i-1}^{k-1}U_i^{k-1}g_{\uparrow i}\| + k2^{O(i)}\gamma\|f\|
\end{aligned}
$$

where we have used the facts that by Equation (1.6) and Lemma 1.7.5, $\|E_{k,i}\| \leq k\gamma$, and $\|D_i g_{\uparrow i}\| \leq 2^{O(i)}\gamma\|f\|$. A basic inductive argument then implies that $\|D_{i-1}^k U_i^k g_{\uparrow i}\| \leq 2^{O(i)}k^2\gamma\|f\|$, which completes the proof. For a more formal induction following exactly the same strategy, see [111, 38]. □

With these lemmas in hand, proving that the Bottom-Up Decomposition ℓ_2-approximates the HD-Level-Set Decomposition is elementary.

Theorem 1.7.8. *Let (X, Π) be a two-sided γ-local-spectral expander and $f \in C_k$. Then the Bottom-Up and HD-Level-Set Decomposition are close in ℓ_2-norm:*

$$\|f_{\uparrow i} - f_{\downarrow i}\|_2^2 \leq 2^{O(k)}\gamma\|f\|_2^2.$$

Similarly:

$$\left|\|f_{\uparrow i}\|^2 - \|f_{\downarrow i}\|^2\right| \leq 2^{O(k)}\gamma\|f\|_2^2$$

Proof. By Lemma 1.7.6, we have:

$$
\begin{aligned}
\langle f_{\downarrow i} - f_{\uparrow i}, f_{\downarrow i}\rangle &= \langle f - f_{\uparrow i}, f_{\downarrow i}\rangle \pm 2^{O(k)}\gamma\|f\|_2^2 \\
&= \pm 2^{O(k)}\gamma\|f\|_2^2
\end{aligned}
$$

Similarly:

$$\langle f_{\downarrow i} - f_{\uparrow i}, f_{\uparrow i} \rangle = \pm 2^{O(k)} \gamma \|f\|_2^2$$

Therefore:

$$\langle f_{\downarrow i} - f_{\uparrow i}, f_{\downarrow i} - f_{\uparrow i} \rangle \leq 2^{O(k)} \gamma \|f\|_2^2$$

as desired. To prove the second inequality, note that:

$$\begin{aligned} |\langle f_{\uparrow i}, f_{\uparrow i} \rangle - \langle f_{\downarrow i}, f_{\downarrow i} \rangle| &= |\langle f_{\uparrow i} - f_{\downarrow i}, f_{\uparrow i} - f_{\downarrow i} \rangle + 2\langle f_{\downarrow i}, f_{\uparrow i} - f_{\downarrow i} \rangle| \\ &\leq 2^{O(k)} \gamma \|f\|^2 \end{aligned}$$

by our previous observations. □

Note that since the HD-Level-Set Decomposition satisfies the Fourier-anatlyic properties in Theorem 1.3.2 [111, 9, 38], Theorem 1.7.8 implies that the Bottom-Up Decomposition does as well.

1.7.2 Properties of the Bottom-Up Decomposition

Our proof of hypercontractivity (Theorem 1.3.5) relies on a number of important structural properties of and relations between g_i and f_i. The first (and most basic) of these is the analog of a classic result for the HD-Level-Set Decomposition relating to ℓ_2-norms of g_i and f_i.

Lemma 1.7.9. *Let (X, Π) be a two-sided γ-local-spectral expander and $f \in C_k$. Then:*

$$\langle g_{\uparrow i}, g_{\uparrow i} \rangle = \frac{1}{\binom{k}{i}} \langle f_{\uparrow i}, f_{\uparrow i} \rangle \pm c_{k,i} \gamma \|D_i^k f\|_2^2$$

where $c_{k,i} \leq k^{O(i)}$.

Proof. The proof is essentially the same as for the HD-Level-Set Decomposition and as our analysis above, though we repeat the idea for completeness. The key is again to apply Equation (1.6). In particular, recall that:

$$\begin{aligned}\langle f_{\uparrow i}, f_{\uparrow i}\rangle &= \binom{k}{i}^2 \langle U_i^k g_{\uparrow i}, U_i^k g_{\uparrow i}\rangle \\ &= \binom{k}{i}^2 \langle g_{\uparrow i}, D_i^k U_i^k g_{\uparrow i}\rangle\end{aligned}$$

by adjointness of D and U [111]. The proof is then essentially the same as Lemma 1.7.6. Repeated application of Equation (1.6) gives an error term of $O(k^2)\gamma\|D_i^k f\|_2$. The only difference is that the main term no longer has an extra occurrence of D at the end. Thus instead of becoming another error term, the main term becomes:

$$\binom{k}{i}^2 \left(\prod_{j=0}^{k-i-1} \frac{k-j-i}{k-j}\right) \langle g_{\uparrow i}, g_{\uparrow i}\rangle = \binom{k}{i} \langle g_{\uparrow i}, g_{\uparrow i}\rangle$$

which gives the result. □

We now cover a few important bounds on g_i for *pseudorandom functions*.

Definition 1.7.10 (Pseudorandom)**.** Let (X, Π) be a simplicial complex. We say that $f \in C_k$ is (ϵ, i)-pseudorandom if it is sparse across all i-links in two senses:

1. For all $\tau \in X(i)$:
$$\left|\mathop{\mathbb{E}}_{X_\tau}[f]\right| \leq \varepsilon\|f\|_\infty$$
2. For all $\tau \in X(i)$:
$$\langle f|_\tau, f|_\tau\rangle \leq \varepsilon\|f\|_\infty^2$$

We note that if f is non-negative, the former condition implies the latter:

$$\langle f|_\tau, f|_\tau\rangle \leq \|f|_\tau\|_1\|f\|_\infty$$

$$= \mathbb{E}[f|_\tau]\|f\|_\infty$$
$$\leq \varepsilon\|f\|_\infty^2.$$

It is also worth noting that any (ϵ, i)-pseudorandom function is automatically (ϵ, j) pseudorandom for $j \leq i$.

We now cover the first property of the Bottom-Up Decomposition that does not follow from standard HDX analysis, the behavior of level functions under restriction. Analysis of restrictions is a classic Fourier analytic tool, and the fact that our decomposition behaves nicely under restriction is a major advantage over previous decompositions which have no clear local structure in this sense. For this particular work, we'll mostly be interested in the following bound on the ℓ_2-norm of restrictions.

Proposition 1.7.11. *Let (X, Π) be a two-sided γ-local-spectral expander and $f \in C_k$ an (ϵ, i)-pseudorandom function. Then for any $j \leq i \leq k$ and $\tau \in X(j)$:*

$$\langle g_{\uparrow i}|_\tau, g_{\uparrow i}|_\tau \rangle \leq \left(\frac{\epsilon}{\binom{k-j}{i-j}} + c_{k,i}\gamma \right) \|f\|_\infty^2,$$

where $c_{k,i} \leq k^{O(i)}$.

Proving this, however, requires a more general understanding of the Bottom-Up Decomposition under restriction. They key observation is that the restriction of our level functions is closely related to the level functions of the restriction. More formally, for any $\tau \in X(j)$ let $g_{\uparrow \ell}^{(\tau)}$ denote the Bottom-Up Decomposition of $f|_\tau$. Then following relation between $g_i|_\tau$ and the $g_{\uparrow \ell}^{(\tau)}$ holds.

Lemma 1.7.12. *Let (X, Π) be a pure, weighted simplicial complex, and $f \in C_k$. Then for any $j \leq i \leq k$ and $\tau \in X(j)$:*

$$g_{\uparrow i}|_\tau = \sum_{\sigma \subseteq \tau} (-1)^{|\sigma|} g_{\uparrow i-j}^{(\tau \setminus \sigma)}.$$

Proof. This follows almost immediately from directly expanding the definition of $g_{\uparrow i}|_\tau$. In particular, recall that for all $I \in X(i-j)$, we have by Theorem 1.7.2:

$$g_{\uparrow i}|_\tau(I) = \sum_{T \subseteq I \cup \tau} (-1)^{|(\tau \cup I) \setminus T|} \mathop{\mathbb{E}}_{X_T}[f].$$

The trick is to notice that we can divide up this sum over $T \subseteq I \cup \tau$ by T's intersection with τ. It will be convenient to phrase this in the following way. Let $\mathscr{T}$ denote the set of all sub-faces $T \subseteq I \cup \tau$, and for each $\sigma \subset \tau$, let $\mathscr{T}_\sigma$ be the set of sub-faces $T \subset I \cup \tau$ such that $T \cap \tau = \tau \setminus \sigma$. Notice that for any $\sigma \neq \sigma'$, $\mathscr{T}_\sigma$ and $\mathscr{T}_{\sigma'}$ are disjoint, and that the union of these families is exactly $\mathscr{T}$. Together, this means that we can break up the above sum by first summing over σ, and then every $T \in \mathscr{T}_\sigma$:

$$g_{\uparrow i}|_\tau(I) = \sum_{\sigma \subseteq \tau} \left(\sum_{T \in \mathscr{T}_\sigma} (-1)^{|(I \cup \tau) \setminus T|} \mathop{\mathbb{E}}_{X_T}[f] \right)$$

By definition, every $T \in \mathscr{T}_\sigma$ can be written as $T' \cup (\tau \setminus \sigma)$. Plugging this into the above gives the result:

$$\begin{aligned}
\sum_{\sigma \subseteq \tau} \left(\sum_{T \in \mathscr{T}_\sigma} (-1)^{|(I \cup \tau) \setminus T|} \mathop{\mathbb{E}}_{X_T}[f] \right) &= \sum_{\sigma \subseteq \tau} \left(\sum_{T' \cup (\tau \setminus \sigma) \in \mathscr{T}_\sigma} (-1)^{|(I \cup \tau) \setminus (T' \cup (\tau \setminus \sigma))|} \mathop{\mathbb{E}}_{X_{T' \cup (\tau \setminus \sigma)}}[f] \right) \\
&= \sum_{\sigma \subseteq \tau} (-1)^{|\sigma|} \left(\sum_{T' \cup (\tau \setminus \sigma) \in \mathscr{T}_\sigma} (-1)^{|I \setminus T'|} \mathop{\mathbb{E}}_{X_{T' \cup (\tau \setminus \sigma)}}[f] \right) \\
&= \sum_{\sigma \subseteq \tau} (-1)^{|\sigma|} g^{(\tau \setminus \sigma)}_{\uparrow i-j}(I),
\end{aligned}$$

where the final step comes from the fact that the inner summation over $\mathscr{T}_\sigma$ is equivalent to summing over all T' in the link of $\tau \setminus \sigma$. □

We note that the same result was known to hold for the Bottom-Up Decomposition over the complete complex [252], who proved the result by induction using the recursive

form of the decomposition. The same strategy will work for general simplicial complexes, but we find using the explicit form as above to be a bit simpler.

With this in hand, proving Proposition 1.7.11 is fairly elementary and follows similarly to its analogous statement for the complete complex (see [252, Corollary 3.4]).

Proof of Proposition 1.7.11. An application of Lemma 1.7.12 and Cauchy-Schwarz implies that:

$$\begin{aligned} g_{\uparrow i}|_\tau(T)^2 &= \left(\sum_{\sigma\subseteq\tau}(-1)^{|\sigma|}g_{\uparrow i-j}^{(\tau\setminus\sigma)}(T)\right)^2 \\ &\leq 2^{O(i)}\sum_{\sigma\subseteq\tau}g_{\uparrow i-j}^{(\tau\setminus\sigma)}(T)^2 \end{aligned}$$

Then applying Lemma 1.7.9 gives:

$$\begin{aligned} \langle g_{\uparrow i}|_\tau, g_{\uparrow i}|_\tau\rangle &\leq 2^{O(i)}\sum_{\sigma\subseteq\tau}\langle g_{\uparrow i-j}^{(\tau\setminus\sigma)}, g_{\uparrow i-j}^{(\tau\setminus\sigma)}\rangle \\ &\leq 2^{O(i)}\sum_{\sigma\subseteq\tau}\frac{\langle f|_{(\tau\setminus\sigma)}, f|_{(\tau\setminus\sigma)}\rangle}{\binom{k-j+|\sigma|}{i-j}} + c_1\gamma\|f\|_\infty^2 \\ &\leq 2^{O(i)}\sum_{\sigma\subseteq\tau}\frac{\varepsilon\|f\|_\infty^2}{\binom{k-j+|\sigma|}{i-j}} + c_1\gamma\|f\|_\infty^2 \\ &\leq 2^{O(i)}\frac{\varepsilon\|f\|_\infty^2}{\binom{k-j}{i-j}} + c_2\gamma\|f\|_\infty^2 \end{aligned}$$

where $c_1, c_2 \leq k^{O(i)}$. □

Finally, it will also be useful to bound the infinity norm of $g_{\uparrow i}$ as well. Our final property shows that $\|g_{\uparrow i}\|_\infty$ is particularly small when f is pseudorandom.

Lemma 1.7.13. *Let (X,Π) be a two-sided γ-local-spectral expander and $f \in C_k$ be any (ϵ, i)-pseudorandom function satisfying $\mathbb{E}[f] \leq \epsilon\|f\|_\infty$. Then the infinity norm of $g_{\uparrow i}$ is small:*

$$\|g_{\uparrow i}\|_\infty \leq 2^i\epsilon\|f\|_\infty$$

Proof. This is immediate from combining the explicit form of $g_{\uparrow i}$ with (ϵ, i) pseudorandomness.

$$\begin{aligned}|g_{\uparrow i}(w)| &= \left|\sum_{j=0}^{i}(-1)^{i-j}\binom{i}{j}U_j^i D_j^k f(w)\right| \\ &\leq \sum_{j=0}^{i}\binom{i}{j}\epsilon\|f\|_\infty \\ &= 2^i\epsilon\|f\|_\infty\end{aligned}$$

where we have used the observation that since f is (ϵ, i)-pseudorandom, for all $w \in X(i)$:

$$|U_j^i D_j^k f(w)| \leq \epsilon\|f\|_\infty$$

□

1.8 Hypercontractivity on HDX

In this section, we prove a hypercontractivity theorem for the Bottom-Up Decomposition on two-sided local-spectral expanders. Since we will work only with the Bottom-Up Decomposition in this section, we drop the $\uparrow$ for simplicity and simply write $f = \sum f_i$ for $f_i = \binom{k}{i}U_i^k g_i$ and $g_i = g_{\uparrow i}$ as defined in the Bottom-Up Decomposition.

Theorem 1.8.1. *Let (X, Π) be a two-sided γ-local-spectral expander with $\gamma \leq k^{-\Omega(i)}$, and $f \in C_k$ an (ϵ, i)-pseudorandom function. If $f = f_0 + \ldots + f_k$ is the Bottom-Up Decomposition of f, then:*

$$\mathbb{E}[f_i^4] \leq 2^{O(i)}\epsilon\,\mathbb{E}[f_i^2]\|f\|_\infty^2 + c_{k,i}\varepsilon\gamma^{1/2}\|D_i^k f\|_2^2\|f\|_\infty^2$$

where $c_{k,i} \leq k^{O(i)}$.[17]

[17]Note that this can be improved to $c_{k,i} \leq \max\left\{2^{O(i)}, \binom{k}{i}^{O(1)}\right\}$, but since we generally consider the

For simplicity of notation, we note it is sufficient to prove the result assuming $\|f\|_\infty = 1$. Given a general function f, applying this to $\frac{f}{\|f\|_\infty}$ gives the general form in Theorem 1.8.1. Keeping this in mind, we'll start by laying out our general strategy for analyzing the fourth moment. Let $[\tau]_i = \{a \subseteq \tau : a \in X(i)\}$, and note that $f_i(\tau) = \sum_{a \in [\tau]_i} g_i(a)$. Using this notation, we can expand out the 4th moment of f_i:

$$\mathbb{E}[f_i^4] = \mathop{\mathbb{E}}_{\tau \in X(k)} \sum_{a,b,c,d \in [\tau]_i} g_i(a)g_i(b)g_i(c)g_i(d) = \sum_{a,b,c,d \in X(i)} \pi_k(X_{a \cup b \cup c \cup d}) g_i(a)g_i(b)g_i(c)g_i(d),$$

where the indices a, b, c, d are *ordered*. We can further simplify this by grouping the terms by size of $a \cup b \cup c \cup d$:

$$\mathbb{E}[f_i^4] = \sum_{\ell=i}^{4i} \binom{k}{\ell} \sum_{e \in X(\ell)} \pi_\ell(e) \sum_{a,b,c,d \in X(i) : a \cup b \cup c \cup d = e} g_i(a)g_i(b)g_i(c)g_i(d).$$

Analyzing the RHS directly is difficult, so taking after [252], we will partition the term even further by summing over fixed **intersection patterns** of a, b, c, and d (an intersection pattern fixes the intersection size of every subset of $\{a, b, c, d\}$). Denote the set of such patterns where $|a \cup b \cup c \cup d| = \ell$ by Σ_ℓ, and for any $e \in X(\ell)$, and $\sigma \in \Sigma_\ell$, let $\sigma(e)$ denote all tuples (a, b, c, d) such that $a \cup b \cup c \cup d = e$, $(a, b, c, d) \in \sigma$. We may now write:

$$\mathbb{E}[f_i^4] = \sum_{\ell=i}^{4i} \binom{k}{\ell} \sum_{\sigma \in \Sigma_\ell} \sum_{e \in X(\ell)} \pi_\ell(e) \sum_{a,b,c,d \in \sigma(e)} g_i(a)g_i(b)g_i(c)g_i(d).$$

We make one final simplification of the above before moving to analysis. Let $x_1, \ldots, x_\ell$ be random variables which take on vertex values in the complex. For each intersection pattern $\sigma \in \Sigma_\ell$, let $I_1^\sigma, \ldots, I_4^\sigma$ be size-i subsets of $\{x_1, \ldots, x_\ell\}$ whose union is $\{x_1, \ldots, x_\ell\}$ and which satisfy the intersection pattern σ. Then we can simplify the above as the following

regime of $i \ll k$ we use $k^{O(i)}$ throughout for simplicity.

expectation over the x_i, that is $\mathbb{E}[f_i^4]$ is exactly

$$\sum_{\ell=i}^{4i} \binom{k}{\ell} \sum_{\sigma \in \Sigma_\ell} \beta(\sigma) \mathop{\mathbb{E}}_{x_1 \in X(1)} \left[\mathop{\mathbb{E}}_{x_2 \in X_{x_1}(1)} \cdots \left[\mathop{\mathbb{E}}_{x_\ell \in X_{x_1,\ldots,x_{\ell-1}}(1)} [g_i(I_1^\sigma) g_i(I_2^\sigma) g_i(I_3^\sigma) g_i(I_4^\sigma)] \right] \right].$$

where $\beta(\sigma) \leq 2^{O(i)}$ is a parameter dependent on the intersection pattern that accounts for the new normalization of terms in the nested expectation. For simplicity of notation, we will instead write the right-hand side as:

$$\mathbb{E}[f_i^4] = \sum_{\ell=i}^{4i} \binom{k}{\ell} \sum_{\sigma \in \Sigma_\ell} \beta(\sigma) \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell} [g_i(I_1^\sigma) g_i(I_2^\sigma) g_i(I_3^\sigma) g_i(I_4^\sigma)]$$

where it is understood that $\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}$ is a shorthand for the nested expectation

$$\mathop{\mathbb{E}}_{x_1 \in X(1)} \mathop{\mathbb{E}}_{x_2 \in X_{x_1}(1)} \cdots \mathop{\mathbb{E}}_{x_\ell \in X_{x_1,\ldots,x_{\ell-1}}(1)}.$$

We will use this convention throughout the rest of the proof, as the nested notation is cumbersome to write otherwise.

Our goal is now to upper bound this sum to get a hypercontractive inequality. We do this by bounding each sign pattern independently.

Claim 1.8.2. For every sign pattern σ, the corresponding expectation is bounded by:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell} [g_i(I_1^\sigma) g_i(I_2^\sigma) g_i(I_3^\sigma) g_i(I_4^\sigma)] \leq \left(\frac{i}{k}\right)^\ell 2^{O(i)} \epsilon \, \mathbb{E}[f_i^2] + c_{k,i} \varepsilon \gamma^{1/2} \|D_i^k f\|_2^2$$

where $c_{k,i} \leq k^{O(i)}$.

Before jumping into the proof of Claim 1.8.2, let's show how it can be used to prove Theorem 1.8.1.

Proof of Theorem 1.8.1. Recall it is sufficient to prove the result assuming $\|f\|_\infty = 1$. As

discussed earlier in the section, expanding the 4th moment gives the following relation:

$$\mathbb{E}[f_i^4] = \sum_{\ell=i}^{4i} \binom{k}{\ell} \sum_{\sigma} \beta(\sigma) \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell} \left[g_i(I_1^\sigma) g_i(I_2^\sigma) g_i(I_3^\sigma) g_i(I_4^\sigma) \right].$$

Applying Claim 1.8.2 to the righthand side gives:

$$\begin{aligned}
\mathbb{E}[f_i^4] &\leq \sum_{\ell=i}^{4i} \binom{k}{\ell} \sum_{\sigma \in \Sigma_\ell} \beta(\sigma) \left(\left(\frac{i}{k} \right)^\ell 2^{O(i)} \epsilon \, \mathbb{E}[f_i^2] + c_{k,i} \varepsilon \gamma^{1/2} \|D_i^k f\|_2^2 \right) \\
&\leq \sum_{\ell=i}^{4i} \left(\frac{ek}{\ell} \right)^\ell \left(\frac{i}{k} \right)^\ell \left(\sum_{\sigma \in \Sigma_\ell} \beta(\sigma) \right) \left(2^{O(i)} \epsilon \, \mathbb{E}[f_i^2] + c_1 \varepsilon \gamma^{1/2} \|D_i^k f\|_2^2 \right) \\
&\leq 2^{O(i)} \epsilon \, \mathbb{E}[f_i^2] + c_2 \varepsilon \gamma^{1/2} \|D_i^k f\|_2^2
\end{aligned}$$

where $c_1, c_2 \leq k^{O(i)}$ and the last step follows from noting that there are at most $\text{poly}(i)$ intersection patterns. □

1.8.1 Proving Claim 1.8.2

The main technical work comes in proving Claim 1.8.2, which relies heavily on Garland's method and our new localization strategy for decorrelating variables (Corollary 1.6.4).

We split the proof into two parts. First, we will show that any pattern which has a unique element (i.e. some x_i which appears only in one of the four sets) may be disregarded.

Proposition 1.8.3. *If σ is a pattern in which any variable is unique (appears in only one I_j), then:*

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell} \left[g_i(I_1^\sigma) g_i(I_2^\sigma) g_i(I_3^\sigma) g_i(I_4^\sigma) \right] \leq 2^{O(i)} \gamma \epsilon^2 \|D_i^k f\|_2^2.$$

Proof. To simplify notations in the proof, let $I = \{x_1, \ldots, x_\ell\}$ and $I_j = I_j^\sigma$. Assume without loss of generality that I_4 has a unique variable x_ℓ, and set $J = I \setminus \{x_\ell\}$ and

$J_4 = I_4 \setminus \{x_\ell\}$. We can re-write our expectation as:

$$(*) = \mathop{\mathbb{E}}_{J}\left[g_i(I_1)g_i(I_2)g_i(I_3)\mathop{\mathbb{E}}_{x_\ell \in X_J(1)}\left[g_i|_{J_4}(x_\ell)\right]\right].$$

By Corollary 1.6.4, the inner expectation can be replaced with $D_i g_i(J_4)$ up to γ error in the following sense. Consider any fixing of the variables J (namely, fixing $x_1, \ldots, x_{\ell-1}$), we have:

$$\begin{aligned}\mathop{\mathbb{E}}_{x_\ell \in X_J(1)}\left[g_i|_{J_4}(x_\ell)\right] &= \mathop{\mathbb{E}}_{x_\ell \in X_{J_4}(1)}\left[g_i|_{J_4}(x_\ell)\right] + \Gamma g_i|_{J_4}(J \setminus J_4) \\ &= D_i g_i(J_4) + \Gamma g_i|_{J_4}(I \setminus I_4)\end{aligned}$$

where $\|\Gamma\| \leq O(i\gamma)$ by Corollary 1.6.4. Plugging this back into our original expectation gives:

$$(*) = \mathop{\mathbb{E}}_{J}\left[g_i(I_1)g_i(I_2)g_i(I_3)D_i g_i(J_4)\right] + \mathop{\mathbb{E}}_{J}\left[g_i(I_1)g_i(I_2)g_i(I_3)\Gamma g_i|_{J_4}(I \setminus I_4)\right]$$

The idea is now to split each term into two parts: the first three terms $g_i(I_1)g_i(I_2)g_i(I_3)$ and the last term. Let's first split these portions by Cauchy-Schwarz to get:

$$\begin{aligned}(*) \leq\ & \mathop{\mathbb{E}}_{J}\left[g_i(I_1)^2 g_i(I_2)^2 g_i(I_3)^2\right]^{1/2} \\ & \cdot \left(\mathop{\mathbb{E}}_{J}\left[D_i g_i(J_4)^2\right]^{1/2} + \mathop{\mathbb{E}}_{J_4}\left[\mathop{\mathbb{E}}_{I \setminus I_4 \in X_{J_4}}\left[\Gamma g_i|_{J_4}(I \setminus I_4)^2\right]\right]^{1/2}\right)\end{aligned}$$

where we have re-arranged variable for convenience in the last term. We now bound each term separately.

The first term can be bounded by the observation that $\|g_i\|_\infty \leq 2^{O(i)}\epsilon\|f\|_\infty$, and hence:

$$\mathop{\mathbb{E}}_{J}\left[g_i(I_1)^2 g_i(I_2)^2 g_i(I_3)^2\right]^{1/2} \leq 2^{O(i)}\epsilon^2\|g_i\|_2$$

where we simply bounded two of the three g_i^2 terms by their infinity norm and applied Garland's lemma for localizations (Lemma 1.6.2) to remove the extra variables.

We next analyze the second term. The first summand is exactly $\|D_i g_i\|$, which by Lemma 1.7.5 is at most $O(\gamma \|D_i^k f\|)$. The second summand is more involved, but can be analyzed through a combination of standard spectral bounds and Garland's lemma for restrictions (Lemma 1.6.1). In particular, re-writing the inner expectation as an inner-product we get:

$$\begin{aligned}
\mathop{\mathbb{E}}_{J_4}\left[\mathop{\mathbb{E}}_{I \setminus I_4 \in X_{J_4}}\left[\Gamma g_i|_{J_4}(I \setminus I_4)^2\right]\right]^{1/2} &= \mathop{\mathbb{E}}_{J_4}\left[\langle \Gamma g_i|_{J_4}, \Gamma g_i|_{J_4}\rangle\right]^{1/2} \\
&\leq c\gamma \mathop{\mathbb{E}}_{J_4}\left[\langle g_i|_{J_4}, g_i|_{J_4}\rangle\right]^{1/2} \\
&= c\gamma \|g_i\|_2,
\end{aligned}$$

where $c \leq O(i)$ and we have applied the fact that $\|\Gamma\| \leq O(i\gamma)$ and Garland's lemma for restrictions (Lemma 1.6.1). Recalling from Claim 1.7.7 that $\|g_i\|_2 \leq 2^i \|D_i^k f\|_2$ completes the result. □

We may now restrict our analysis to patterns in which every variable appears at least twice. Note that this implies $\ell \leq 2i$, which is important because we expect our expectation to scale at best with k^{-2i}, so any terms with $\ell > 2i$ would cause difficulty. As in [252], we break this analysis into two steps. Let $I_1, \ldots, I_4$ satisfy intersection pattern σ as above (we drop the σ superscript for convenience), and let H_i for $i \in \{2, 3, 4\}$ denote the set of variables that appear i times.

We'll start by handling H_2 through a combination of Cauchy-Schwarz, Garland's method, and our localization technique for higher moments. Unlike the case of the complete complex studied in [252], these latter components are necessary due to the fact that local-spectral expanders are generally far from product spaces (a crucial property of the complete complex exploited in [252]). The proof is fairly technical, so we'll start by laying out

some convenient notation. For any $0 \le m \le \ell$, let $T^m = \{x_1, \dots, x_m\}$. Let $j = |H_3 \cup H_4|$ where $0 \le j \le \ell$. Noting that re-ordering the variables $x_1, \dots, x_\ell$ has no effect on the distribution, we may assume without loss of generality that $H_3 \cup H_4 = \{x_1, \dots, x_j\}$ (where if $j = 0$ then $H_3 \cup H_4$ is empty). Finally, we introduce two useful notations: for $m \le \ell$ let $I_r^m = I_r \cap \{x_1, \dots, x_m\}$ and $s_r^m = i - |I_r^m|$.

Proposition 1.8.4.

$$\mathop{\mathbb{E}}_{x_1,\dots,x_\ell}[g_i(I_1)g_i(I_2)g_i(I_3)g_i(I_4)] \le \mathop{\mathbb{E}}_{x_1,\dots,x_j}\left[\sqrt{\prod_{r=1}^{4}\mathop{\mathbb{E}}_{\tau_r \sim X_{T^j}(s_r^j)}[g_i^2|_{I_r^j}(\tau_r)]}\right] + 2^{O(i)}\gamma^{1/2}\epsilon^2\|g_i\|_2^2$$

Proof. The proof follows from an inductive argument where we pull one variable in $x \in H_2$ inside the sum in each step by de-correlating the two copies of g_i which do not take x as an input, and then applying Cauchy-Schwarz. In particular, we will show by induction that for all $\ell \ge m \ge j$:

$$\mathop{\mathbb{E}}_{x_1,\dots,x_\ell}[g_i(I_1)g_i(I_2)g_i(I_3)g_i(I_4)] \le \mathop{\mathbb{E}}_{x_1,\dots,x_m}\left[\sqrt{\prod_{r=1}^{4}\mathop{\mathbb{E}}_{\tau_r \sim X_{T^m}(s_r^m)}[g_i^2|_{I_r^m}(\tau_r)]}\right] + 2^{O(i)}\gamma^{1/2}\epsilon^2\|g_i\|_2^2$$

The base case ($m = \ell$) is trivial (as I_r^m then contains all relevant variables and the inner expectations are trivial). Since we also done if $m = j$, we may assume that $x_m \in H_2$ and therefore lies in exactly two of I_1, I_2, I_3, I_4 by definition. Assume without loss of generality that $x_m \in I_3, I_4$. We'd like to pull x_m inside the expectation. The issue is that despite the fact that x_m does not participate in I_1 or I_2, these terms actually depend on x_m regardless since τ_1 and τ_2 are drawn from a link that includes x_m. To fix this, we can use Corollary 1.6.4 to de-correlate these terms from x_m:

$$\sqrt{\prod_{r=1}^{2}\mathop{\mathbb{E}}_{\tau_r \sim X_{T^m}(s_r^m)}[g_i^2|_{I_r^m}(\tau_r)]} = \sqrt{\prod_{r=1}^{2}\left(\mathop{\mathbb{E}}_{\tau_r \sim X_{T^{m-1}}(s_r^{m-1})}\left[g_i^2|_{I_r^{m-1}}(\tau_r)\right] + \Gamma g_i^2|_{I_r^{m-1}}(x_m)\right)}$$

$$
\begin{aligned}
&\leq \sqrt{\mathop{\mathbb{E}}_{\tau_1 \sim X_{T^{m-1}}(s_1^{m-1})} \left[g_i^2|_{I_1^{m-1}}(\tau_1) \right]} \sqrt{\mathop{\mathbb{E}}_{\tau_2 \sim X_{T^{m-1}}(s_2^{m-1})} \left[g_i^2|_{I_2^{m-1}}(\tau_1) \right]} \\
&+ \sqrt{\Gamma g_i^2|_{I_2^{m-1}}(x_m)} \sqrt{\mathop{\mathbb{E}}_{\tau_1 \sim X_{T^{m-1}}(s_1^{m-1})} \left[g_i^2|_{I_1^{m-1}}(\tau_1) \right]} \\
&+ \sqrt{\Gamma g_i^2|_{I_1^{m-1}}(x_m)} \sqrt{\mathop{\mathbb{E}}_{\tau_2 \sim X_{T^{m-1}}(s_2^{m-1})} \left[g_i^2|_{I_2^{m-1}}(\tau_2) \right]} \\
&+ \sqrt{\Gamma g_i^2|_{I_1^{m-1}}(x_m)} \sqrt{\Gamma g_i^2|_{I_2^{m-1}}(x_m)}
\end{aligned}
$$

where $\|\Gamma\| \leq O(i\gamma)$ and we have used the fact that by assumption $I_r^m = I_r^{m-1}$ for $r = 1, 2$. For the moment, denote the last 3 terms by $err(g)$. Then by the inductive hypothesis we have:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell} [g_i(I_1) g_i(I_2) g_i(I_3) g_i(I_4)] \\
\leq &\mathop{\mathbb{E}}_{x_1,\ldots,x_m} \left[\sqrt{\prod_{r=1}^{4} \mathop{\mathbb{E}}_{\tau_r \sim X_{T^m}(s_r)} [g_i^2|_{I_r^m}(\tau_r)]} \right] + 2^{O(i)} \gamma^{1/2} \epsilon^2 \|g_i\|_2^2 \\
\leq &\mathop{\mathbb{E}}_{x_1,\ldots,x_{m-1}} \left[\sqrt{\prod_{r=1}^{2} \mathop{\mathbb{E}}_{\tau_r \sim X_{T^{m-1}}(s_r)} [g_i^2|_{I_r^{m-1}}(\tau_r)]} \right. \\
&\qquad \left. \cdot \mathop{\mathbb{E}}_{x_m} \left[\sqrt{\mathop{\mathbb{E}}_{\tau_3 \sim X_{T^m}(s_3)} [g_i^2|_{I_3^m}(\tau_3)]} \sqrt{\mathop{\mathbb{E}}_{\tau_4 \sim X_{T^m}(s_4)} [g_i^2|_{I_4^m}(\tau_4)]} \right] \right] \\
&\quad + \mathop{\mathbb{E}}_{x_1,\ldots,x_m} \left[err(g) \sqrt{\mathop{\mathbb{E}}_{\tau_3 \sim X_{T^m}(s_3^m)} [g_i^2|_{I_3^m}(\tau_3)]} \sqrt{\mathop{\mathbb{E}}_{\tau_4 \sim X_{T^m}(s_4^m)} [g_i^2|_{I_4^m}(\tau_4)]} \right] + 2^{O(i)} \gamma^{1/2} \epsilon^2 \|g_i\|_2^2
\end{aligned}
$$

By Cauchy-Schwarz, the first term can be bounded by:

$$
\mathop{\mathbb{E}}_{x_1,\ldots,x_{m-1}} \left[\sqrt{\prod_{r=1}^{4} \mathop{\mathbb{E}}_{\tau_r \sim X_{T^{m-1}}(s_r^{m-1})} [g_i^2|_{I_r^{m-1}}(\tau_r)]} \right],
$$

so it is enough to show that the latter error term is small. We'll analyze each term in

$err(g)$ independently using Cauchy-Schwarz, Garland's method, and our bound on $\|g\|_\infty$. Starting with the first term, an application of Cauchy-Schwarz gives:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x_1,\ldots,x_m}\left[\sqrt{\Gamma g_i^2|_{I_2^{m-1}}(x_m)}\sqrt{\mathop{\mathbb{E}}_{\tau_1\sim X_{T^{m-1}}(s_1^{m-1})}\left[g_i^2|_{I_1^{m-1}}(\tau_1)\right]}\right.\\
&\qquad\left.\cdot\sqrt{\mathop{\mathbb{E}}_{\tau_3\sim X_{T^m}(s_3^m)}[g_i^2|_{I_3^m}(\tau_3)]}\sqrt{\mathop{\mathbb{E}}_{\tau_4\sim X_{T^m}(s_4^m)}[g_i^2|_{I_4^m}(\tau_4)]}\right]\\
\leq&\mathop{\mathbb{E}}_{x_1,\ldots,x_m}\left[\Gamma g_i^2|_{I_2^{m-1}}(x_m)\mathop{\mathbb{E}}_{\tau_1\sim X_{T^{m-1}}(s_1^{m-1})}\left[g_i^2|_{I_1^{m-1}}(\tau_1)\right]\right]^{1/2}\\
&\cdot\mathop{\mathbb{E}}_{x_1,\ldots,x_m}\left[\mathop{\mathbb{E}}_{\tau_3\sim X_{T^m}(s_3^m)}[g_i^2|_{I_3^m}(\tau_3)]\mathop{\mathbb{E}}_{\tau_4\sim X_{T^m}(s_4^m)}[g_i^2|_{I_4^m}(\tau_4)]\right]^{1/2}.
\end{aligned}$$

The righthand expectation is easy to analyze using the fact that $\|g_i\|_\infty\leq 2^{O(i)}\epsilon$:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x_1,\ldots,x_m}\left[\mathop{\mathbb{E}}_{\tau_3\sim X_{T^m}(s_3^m)}[g_i^2|_{I_3^m}(\tau_3)]\mathop{\mathbb{E}}_{\tau_4\sim X_{T^m}(s_4^m)}[g_i^2|_{I_4^m}(\tau_4)]\right]^{1/2}\\
\leq&2^{O(i)}\epsilon\mathop{\mathbb{E}}_{x_1,\ldots,x_m}\left[\mathop{\mathbb{E}}_{\tau_4\sim X_{T^m}(s_4^m)}[g_i^2|_{I_4^m}(\tau_4)]\right]^{1/2}\\
=&2^{O(i)}\epsilon\mathop{\mathbb{E}}_{\tau\sim X(|I_4^m|)}[\langle g_i|_\tau,g_i|_\tau\rangle]^{1/2}\\
=&2^{O(i)}\epsilon\|g_i\|_2
\end{aligned}$$

where the last two equalities follow from Garland's method. Turning our attention to the lefthand expectation, we can apply Cauchy-Schwarz to get:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x_1,\ldots,x_m}\left[\Gamma g_i^2|_{I_2^{m-1}}(x_m)\mathop{\mathbb{E}}_{\tau_1\sim X_{T^{m-1}}(s_1^{m-1})}\left[g_i^2|_{I_1^{m-1}}(\tau_1)\right]\right]^{1/2}\\
\leq&\mathop{\mathbb{E}}_{x_1,\ldots,x_{m-1}}\left[\langle\Gamma g_i^2|_{I_2^{m-1}},\Gamma g_i^2|_{I_2^{m-1}}\rangle\right]^{1/4}\mathop{\mathbb{E}}_{x_1,\ldots,x_{m-1}}\left[\mathop{\mathbb{E}}_{\tau_1\sim X_{T^{m-1}}(s_1^{m-1})}\left[g_i^2|_{I_1^{m-1}}(\tau_1)\right]^2\right]^{1/4}\\
\leq&2^{O(i)}\gamma^{1/2}\mathop{\mathbb{E}}_{x_1,\ldots,x_{m-1}}\left[\langle g_i^2|_{I_2^{m-1}},g_i^2|_{I_2^{m-1}}\rangle\right]^{1/4}\mathop{\mathbb{E}}_{x_1,\ldots,x_{m-1}}\left[\langle g_i^2|_{I_1^{m-1}},g_i^2|_{I_1^{m-1}}\rangle\right]^{1/4}
\end{aligned}$$

where in the last step we have applied the fact that $\|\Gamma\|\leq O(i\gamma)$. Analysis of the remaining

expectations follows exactly as before. In particular, re-arranging variables by symmetry and applying Garland's method, we can continue the above inequality as follows:

$$\begin{aligned}
&= \gamma^{1/2} 2^{O(i)} \mathop{\mathbb{E}}_{\tau \sim X(|I_2^{m-1}|)} \left[\langle g_i^2|_\tau, g_i^2|_\tau \rangle\right]^{1/4} \mathop{\mathbb{E}}_{\tau \sim X(|I_1^{m-1}|)} \left[\langle g_i^2|_\tau, g_i^2|_\tau \rangle\right]^{1/4} \\
&= \gamma^{1/2} 2^{O(i)} \langle g_i^2, g_i^2 \rangle^{1/2} \\
&\leq \gamma^{1/2} 2^{O(i)} \epsilon \|g_i\|_2
\end{aligned}$$

where in the final step we have again applied our bound on $\|g_i\|_\infty$. Putting the analysis of these two terms together, we get the desired bound on the first summand of $err(g)$:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x_1,\ldots,x_m} \Bigg[\sqrt{\Gamma g_i^2|_{I_2^{m-1}}(x_m)} \sqrt{\mathop{\mathbb{E}}_{\tau_1 \sim X_{T^{m-1}}(s_1^{m-1})} \left[g_i^2|_{I_1^{m-1}}(\tau_1) \right]} \\
&\qquad \cdot \sqrt{\mathop{\mathbb{E}}_{\tau_3 \sim X_{T^m}(s_3^m)} [g_i^2|_{I_3^m}(\tau_3)]} \sqrt{\mathop{\mathbb{E}}_{\tau_4 \sim X_{T^m}(s_4^m)} [g_i^2|_{I_4^m}(\tau_4)]} \Bigg] \\
\leq &\gamma^{1/2} 2^{O(i)} \epsilon^2 \|g_i\|_2^2.
\end{aligned}$$

The analysis of second summand in $err(g)$ is exactly the same, and the third term differs only in that the lefthand expectation in the previous analysis becomes:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_m} \left[\Gamma g_i^2|_{I_1^{m-1}}(x_m) \Gamma g_i^2|_{I_2^{m-1}}(x_m) \right]^{1/2} \leq \gamma 2^{O(i)} \epsilon \|g_i\|_2$$

by the same arguments. Combining these together, we get that our error term is bounded by $\gamma^{1/2} 2^{O(i)} \epsilon^2 \|g_i\|_2^2$, which completes the proof. □

It is left to analyze H_3 and H_4. Recalling that we've assumed $\{x_1, \ldots, x_j\} = H_3 \cup H_4$, Proposition 1.8.4 can be restated as:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell} [g_i(I_1) g_i(I_2) g_i(I_3) g_i(I_4)] \leq \mathop{\mathbb{E}}_{H_3 \cup H_4} \left[\sqrt{\prod_{r=1}^{4} \mathop{\mathbb{E}}_{\tau_r \sim X_{H_3 \cup H_4}(s_r^j)} [g_i^2|_{I_r^j}(\tau_r)]} \right] + 2^{O(i)} \gamma^{1/2} \epsilon^2 \|g_i\|_2^2.$$

The key is now to apply Proposition 1.7.11, which says that the maximum of the inner restricted expectations are small, where the factor is better the fewer variables we restrict. In order to minimize the number of restrictions, we use Cauchy-Schwarz to separate out I_1 and I_2 from I_3 and I_4:

$$\mathop{\mathbb{E}}_{H_3 \cup H_4} \left[\sqrt{\prod_{r=1}^{4} \mathop{\mathbb{E}}_{\tau_r \sim X_{H_3 \cup H_4}(s_r^j)} [g_i^2|_{I_r^j}(\tau_r)]} \right] \leq \mathop{\mathbb{E}}_{H_3 \cup H_4} \left[\prod_{r=1}^{2} \mathop{\mathbb{E}}_{\tau_r \sim X_{H_3 \cup H_4}(s_r^j)} [g_i^2|_{I_r^j}(\tau_r)] \right]^{1/2} \cdot \mathop{\mathbb{E}}_{H_3 \cup H_4} \left[\prod_{r=3}^{4} \mathop{\mathbb{E}}_{\tau_r \sim X_{H_3 \cup H_4}(s_r^j)} [g_i^2|_{I_r^j}(\tau_r)] \right]^{1/2}.$$

Analysis of these two terms is the same, so we focus on the former. The idea is to bound one of the two inner expectations (say I_1) by its maximum, and note that the other term then simply returns $\|g_i\|$. Unfortunately, there is a slight issue with this strategy naively: H_3 may contain variables that are not in I_1, so we cannot directly apply Proposition 1.7.11. We can fix this by de-correlating I_1 from the extraneous variables in H_3 using Corollary 1.6.4 (similar to our strategy in Proposition 1.8.4). More formally, let $B_{12} = (H_3 \cap I_1 \cap I_2) \cup H_4$ for simplicity of notation. By exactly the same inductive argument used in Proposition 1.8.4 we have:

$$\mathop{\mathbb{E}}_{H_3 \cup H_4} \left[\prod_{r=1}^{2} \mathop{\mathbb{E}}_{\tau_r \sim X_{H_3 \cup H_4}(s_r^j)} [g_i^2|_{I_r^j}(\tau_r)] \right]^{1/2} \leq \mathop{\mathbb{E}}_{B_{12} \sim X} \left[\mathop{\mathbb{E}}_{\tau_1 \sim X_{B_{12}}} [g_i^2|_{I_1^{B_{12}}}(\tau_1)] \mathop{\mathbb{E}}_{\tau_2 \sim X_{B_{12}}} [g_i^2|_{I_2^{B_{12}}}(\tau_2)] \right]^{1/2} + 2^{O(i)} \varepsilon^2 \gamma^{1/2} \|g_i\|_2$$

where for the moment we have omitted the sizes of B_{12}, τ_1, and τ_2 for simplicity (these will be computed soon). Pulling out the maximal I_1 term and applying Proposition 1.7.11

with $j = |B_{12}|$, we then get:

$$\begin{aligned}\mathop{\mathbb{E}}_{H_3 \cup H_4}\left[\prod_{r=1}^{2} \mathop{\mathbb{E}}_{\tau_r \sim X_{H_3 \cup H_4}(s_r^j)} [g_i^2|_{I_r^j}(\tau_r)]\right]^{1/2} &\leq \max_{B_{12}} \left(\mathop{\mathbb{E}}_{\tau_1 \sim X_{B_{12}}} [g_i^2|_{I_1^{B_{12}}}(\tau_1)]\right)^{1/2} \|g_i\|_2 \\ &\quad + 2^{O(i)}\varepsilon^2\gamma^{1/2}\|g_i\|_2 \\ &\leq \frac{2^{O(i)}\epsilon^{1/2}}{\binom{k-|B_{12}|}{i-|B_{12}|}^{1/2}}\|g_i\|_2 + 2^{O(i)}\varepsilon^2\gamma^{1/2}\|g_i\|_2\end{aligned}$$

The same argument holds for the latter product over I_3 and I_4. Letting $B_{12} = (H_3 \cap I_3 \cap I_4) \cup H_4$, and putting everything together, we finally get the bound:

$$\begin{aligned}\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}[g_i(I_1)g_i(I_2)g_i(I_3)g_i(I_4)] &\leq \frac{2^{O(i)}\epsilon}{\binom{k-|B_{12}|}{i-|B_{12}|}^{1/2}\binom{k-|B_{34}|}{i-|B_{34}|}^{1/2}}\|g_i\|_2^2 + 2^{O(i)}\varepsilon^2\gamma^{1/2}\|g_i\|_2^2 \\ &\leq \frac{2^{O(i)}\epsilon}{\binom{k-|B_{12}|}{i-|B_{12}|}^{1/2}\binom{k-|B_{34}|}{i-|B_{34}|}^{1/2}}\frac{1}{\binom{k}{i}}\|f_i\|_2^2 + 2^{O(i)}\varepsilon^2\gamma^{1/2}\|D_i^k f\|_2^2 \\ &\leq 2^{O(i)}\epsilon\left(\frac{k}{i}\right)^{\frac{|B_{12}|+|B_{34}|}{2}-2i}\|f_i\|_2^2 + 2^{O(i)}\varepsilon^2\gamma^{1/2}\|D_i^k f\|_2^2\end{aligned}$$

where we have applied the basic binomial bound $\binom{n}{p} \geq \left(\frac{n}{p}\right)^p$. To complete the result, it suffices show that $\frac{|B_{12}|+|B_{34}|}{2} = 2i - \ell$. This follows similarly to the analogous argument in [252], but we'll give a simplification of their proof for completeness. Recall that B_{12} consists of variables in H_3 and H_4 that appear in both I_1 and I_2, and B_{34} similarly consists of variables in H_3 and H_4 that appear in both I_3 and I_4. Since every variable in H_3 occurs in exactly one of $(I_1 \cap I_2)$ and $(I_3 \cap I_4)$ by definition, we get that

$$\frac{|B_{12}| + |B_{34}|}{2} = |H_4| + \frac{|H_3|}{2}.$$

To compute the righthand side, note that by definition we have the following two relations:

1. Since each term has $4i$ total variables (with repetition):

$$4|H_4| + 3|H_3| + 2|H_2| = 4i$$

2. Since there are ℓ unique variables:

$$|H_4| + |H_3| + |H_2| = \ell$$

Combining these equations gives the desired equality:

$$|H_4| + \frac{|H_3|}{2} = \frac{(4|H_4| + 3|H_3| + 2|H_2|) - 2(|H_4| + |H_3| + |H_2|)}{2} = 2i - \ell.$$

Putting everything together, we finally get

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}[g_i(I_1)g_i(I_2)g_i(I_3)g_i(I_4)] \leq 2^{O(i)}\epsilon\left(\frac{i}{k}\right)^{\ell}\|f_i\|^2 + c_{k,i}\varepsilon\gamma^{1/2}\|D_i^k f\|^2,$$

as desired.

1.9 Characterizing Expansion in HD-walks

One traditional application of hypercontractivity on the discrete hypercube lies in showing that the noisy hypercube graph (given by randomizing each bit of a binary string x with some probability $1-\rho$) is a small-set expander. This result is also often thought of as stating "sparse functions on the hypercube are noise-sensitive," an interpretation we'll discuss in the next section. Unlike the noisy hypercube, it is well known that HD-walks are far from being small set expanders [38]. Before we quantify this further, let's recall the definition of (edge) expansion in the general weighted setting.

Definition 1.9.1 (Weighted Edge Expansion)**.** Let (X, Π) be a weighted simplicial complex,

M a k-dimensional HD-Walk over (X, Π), and $S \subset X(k)$ a subset. The weighted edge expansion of S is the average probability of leaving S after one step of M:

$$\Phi(S) = \mathop{\mathbb{E}}_{v \sim \pi_k|_S} \left[M(v, X(k) \setminus S) \right],$$

where $\pi_k|_S$ is the (normalized) restriction of π_k to S,

$$M(v, X(k) \setminus S) = \sum_{y \in X(k) \setminus S} M(v, y),$$

and $M(v, y)$ is the transition probability from v to y.

A small-set expander is simply a graph where all small sets expand. To understand why HD-walks fail this condition, let's consider the Johnson graph. The Johnson graph $J(n, k, \ell)$ is the graph on $\binom{[n]}{k}$ whose edges are given by sets with intersection size ℓ. Well-studied object in their own right, the Johnson graphs are a fundamental example higher order random walks on the complete complex [9]. In our context, we usually think of n as being much larger than k, and ℓ as being (at least) ck for some constant $0 < c < 1$. In this case, one can show by direct computation that the expansion of any i-link X_τ is bounded away from 1:

$$\Phi(X_\tau) \approx 1 - c^{-i},$$

despite the fact that its density is vanishingly small: $\mathbb{E}[1_{X_\tau}] \approx (k/n)^i$.

Recently, BHKL proved a general variant of this result for all HD-walks (see [38, Theorem 9.2]). They show that spectrum of any k-dimensional walk M on a sufficiently strong local-spectral expander is divided up into $k+1$ strips of width $O_{k,M}(\gamma)$[18] centered around some set of approximate eigenvalues $\{\lambda_i(M)\}_{i=0}^k$, and that the expansion of any link at level i is almost exactly $1 - \lambda_i(M)$. They also prove a weak converse to this

[18]BHKL actually only prove the width is $O_{k,M}(\sqrt{\gamma})$, the improvement to $O_{k,M}(\gamma)$ was given soon after by Zhang [363].

result: *any* non-expanding set must be concentrated in a link. It is convenient to state the contrapositive. For any $\delta > 0$, let $R_\delta(M) = r$ denote the number of approximate eigenvalues of M that are greater than δ (a quantity BHKL call the **ST-Rank** of M). BHKL [38, Theorem 9.5] prove that the expansion of any set $S \subset X(k)$ is at least:

$$\Phi(S) \gtrsim 1 - \delta - c_1 \binom{k}{r} \varepsilon - c_2 \gamma \tag{1.10}$$

where S is (ε, r)-pseudorandom.[19] This is great when $\varepsilon \ll \binom{k}{r}$, but for many applications of interest (e.g. in hardness of approximation), we think of ε as fixed and of k as going to infinity. In this regime, the above characterization is useless, as the bound reduces to the trivial fact $\Phi(S) \geq 0$. Using hypercontractivity, we can completely resolve this issue by offering a variant of Equation (1.10) with no dependence on k. Before we give the statement, however, we note that both BHKL and our result require the approximate eigenvalues of the HD-walk $\{\lambda_i(M)\}_{i=0}^k$ to decrease monotonically. BHKL proved that this property holds for a broad class of walks they call **complete walks**, which includes all HD-walks of interest studied in the literature.

Definition 1.9.2 (Complete HD-Walk ([38] Definition 7.10))**.** Let (X, Π) be a weighted, pure simplicial complex and $M = \sum_{Y \in \mathcal{Y}} \alpha_Y Y$ an HD-walk on (X, Π). M is called *complete* if for all $n \in \mathbb{N}$ there exist $n_0 > n$ and d such that $\sum_{Y \in \mathcal{Y}} \alpha_Y Y$ is also an HD-walk when taken to be over the d-dimensional complete complex on n_0 vertices.

All walks we have seen so far (canonical walks, swap walks, pure walks, affine combinations thereof, etc.) are complete, so restricting to this class does not lose much generality. With this in mind, we can finally state our dimension independent bound on the expansion of pseudorandom sets.

[19] Note that we have simplified BHKL's result here somewhat for simplicity of presentation, but it is an accurate representation of their result in most cases of interest.

Theorem 1.9.3 (Pseudorandom Sets Expand)**.** *Let (X, Π) be a two-sided γ-local-spectral expander, M a complete k-dimensional HD-walk, and $S \subseteq X(k)$ of density α. Then for any $\delta > 0$ and $r = R_\delta(M) - 1$, the expansion of S is at least:*

$$\Phi(S) \geq 1 - \delta - (1-\delta)2^{O(r)}\epsilon^{1/3} - c\gamma$$

where $c \leq 2^{O(k)} w(M) h(M)^2$ and S is (ε, r)-pseudorandom.

The proof of Theorem 1.9.3 goes through a level-i inequality for pseudorandom functions of independent interest.

Theorem 1.9.4 (Level-i Inequality)**.** *Let (X, Π) be a γ-local-spectral expander with $\gamma < 2^{-\Omega(k)}$ and $f \in C_k$ a boolean, (ϵ, i)-pseudorandom function. Then:*

$$\langle f, f_{\uparrow i} \rangle \leq 2^{O(i)}\epsilon^{1/3}\mathbb{E}[f].$$

Let's first prove Theorem 1.9.3 given Theorem 1.9.4.

Proof of Theorem 1.9.3. The argument is standard and follows from the identity $\Phi(S) = 1 - \frac{1}{\alpha}\langle f, Mf \rangle$ (where $\alpha = \mathbb{E}[f]$), and expanding $f = \sum_{i=0}^{k} f_{\downarrow i}$ (the HD-Level-Set Decomposition). Namely, since $f_{\downarrow i}$ is an approximate eigenvector, we can write:

$$\begin{aligned}\Phi(S) &= 1 - \frac{1}{\alpha}\sum_{i=0}^{k}\langle f, Mf_{\downarrow i}\rangle \\ &\geq 1 - \frac{1}{\alpha}\sum_{i=0}^{k}\lambda_i(M)\langle f, f_{\downarrow i}\rangle - c_1\gamma\end{aligned}$$

where $c_1 \leq w(M)h(M)^2 2^{O(k)}$. We can now apply Theorem 1.7.8 to switch between

decompositions to get:

$$\Phi(S) \geq 1 - \frac{1}{\alpha}\sum_{i=0}^{k} \lambda_i(M)\langle f, f_{\uparrow i}\rangle - c_2\gamma$$

where $c_2 = c_1 + 2^{O(k)}$. Since M is a complete walk, its eigenvalues decay monotonically [38], we can therefore simplify the above to:

$$\begin{aligned}\Phi(S) &\geq 1 - \frac{1}{\alpha}\sum_{i=0}^{r} \lambda_i(M)\langle f, f_{\uparrow i}\rangle - \frac{\lambda_{r+1}(M)}{\alpha}\sum_{i=r+1}^{k}\langle f, f_{\uparrow i}\rangle - c_2\gamma \\ &= 1 - \frac{1}{\alpha}\sum_{i=0}^{r} \lambda_i(M)\langle f, f_{\uparrow i}\rangle - \frac{\lambda_{r+1}(M)}{\alpha}\left(\alpha - \sum_{i=0}^{r}\langle f, f_{\uparrow i}\rangle\right) - c_2\gamma.\end{aligned}$$

Recall that by definition $\lambda_{r+1}(M) \leq \delta$, and hence

$$\begin{aligned}\Phi(S) &\geq 1 - \delta - \frac{1-\delta}{\alpha}\sum_{i=0}^{r}\langle f, f_{\uparrow i}\rangle - c_2\gamma \\ &\geq 1 - \delta - (1-\delta)2^{O(r)}\varepsilon^{1/3} - c_2\gamma\end{aligned}$$

where in the last step we have applied Theorem 1.9.4. □

It is left to prove Theorem 1.9.4, which also follows from fairly standard arguments given Theorem 1.8.1.

Proof of Theorem 1.9.4. To simplify notations, we write f_i instead of $f_{\uparrow i}$. Notice that by Hölder's inequality for $p = 4/3, q = 4$ we have:

$$\langle f, f_i\rangle \leq \|f\|_{4/3}\|f_i\|_4 = \alpha^{3/4}\,\mathbb{E}[f_i^4]^{1/4}$$

Combining this with Theorem 1.8.1 gives the following relation:

$$\frac{\langle f, f_i\rangle^4}{\alpha^3} \leq \mathbb{E}[f_i^4] \leq 2^{O(i)}\epsilon\,\mathbb{E}[f_i^2] + c_1\gamma^{1/2}\alpha \leq 2^{O(i)}\epsilon\langle f, f_i\rangle + c_2\gamma^{1/2}\alpha \tag{1.11}$$

where $c_1, c_2 \leq 2^{O(k)}$ by approximate orthogonality (Lemma 1.7.6). We can simplify the above via two observations. First, note that we can assume without loss of generality that $\gamma^{1/4} \leq c_1^{-1}\epsilon$. This follows from observing that:

$$\langle f, f_i \rangle = \binom{k}{i} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} \langle D_j^k f, D_j^k f \rangle.$$

Appealing to arguments from [38, Lemma 8.8], we have that $\langle D_j^k f, D_j^k f \rangle \leq \varepsilon\alpha$, which gives the naive bound:

$$\langle f, f_i \rangle \leq \binom{k}{i} 2^{O(i)} \varepsilon\alpha.$$

If $\varepsilon \leq \binom{k}{i}^{-3/2}$ then $\binom{k}{i}\epsilon \leq \epsilon^{1/3}$ and our desired bound follows. Otherwise, we may assume from now on that $\varepsilon \geq \binom{k}{i}^{-3/2} \geq 2^{-(3/2)k}$. Since $\gamma \leq 2^{-\Omega(k)}$, we are therefore free to assume $\gamma^{1/4} \leq c_1^{-1}\epsilon$ as well. Second, we can also assume $\langle f, f_i \rangle \geq \gamma^{1/4}\alpha$, since otherwise we are done by our previous assumptions on γ and ε. Combining these with Equation (1.11) then gives:

$$\begin{aligned} \frac{\langle f, f_i \rangle^4}{\alpha^3} &\leq 2^{O(i)}\epsilon\langle f, f_i \rangle + c_1\gamma^{1/2}\alpha \\ &\leq 2^{O(i)}\epsilon\langle f, f_i \rangle + \varepsilon\gamma^{1/4}\alpha \\ &\leq 2^{O(i)}\varepsilon\langle f, f_i \rangle \end{aligned}$$

as desired. □

1.10 Fourier Analysis on HDX

In this section we further develop the theory of Fourier analysis on simplicial complexes, and show how hypercontractivity for pseudorandom functions (Theorem 1.8.1) recovers tight analogues of the KKL Theorem and noise-sensitivity of sparse functions. This requires introducing a number of new analog definitions of classic Fourier analytic

quantities on simplicial complexes. To get an idea for what these should look like, it will be useful to start by considering a natural embedding of the hypercube itself into a simplicial complex.

Definition 1.10.1 (Hypercube Complex)**.** The hypercube complex $X = X_{\{0,1\}^n}$ is the complete n-partite complex on $X(1) = [n] \times \{0, 1\}$, where the first coordinate denotes the color of the vertex. That is, the top level faces are $X(n) = \{\{(1, x_1), \ldots, (n, x_n)\} : x \in \{0, 1\}^n\}$.

We make a few notes on this definition. First, it is clear from definition that $X(n)$ can equivalently be thought of as the hypercube $\{0, 1\}^n$, where each color in $[n]$ corresponds to a coordinate in $\{0, 1\}^n$. Further, classic graphs on $\{0, 1\}^n$ such as the hypercube or noisy hypercube can be expressed as simple higher order random walks. The hypercube graph, for instance, is simply the non-lazy lower walk $UD^+ = 2U_{n-1}D_n - I$. This embedding will serve as our guiding principle for developing analog Fourier-analytic definitions on simplicial complexes—whenever possible, our definitions will reduce to the standard notion when applied to the hypercube complex. We note that the same embedding can be used for any product distribution and all of our definitions will generalize appropriately. We focus on the simple case of the hypercube for ease of exposition.

1.10.1 Total Influence and the KKL Theorem

We'll start with a fundamental concept in classical Fourier analysis, *influence*. Let's first recall the definition of (total) influence on the discrete hypercube. Influence can be formalized in a number of equivalent ways. It is often thought of, for instance, as a measure of average sensitivity. In our context, it will be most convenient to view influence as a statement about the expansion of a function with respect to the hypercube graph. More formally, let Q_n denote the normalized adjacency matrix of the hypercube graph, and $Q_n^{\text{lazy}} = \frac{I+Q_n}{2}$ its lazy variant. We will write total influence in terms of the (un-normalized)

Laplacian operator $L = n(I - Q_n^{\text{lazy}})$.

Definition 1.10.2 (Total Influence (hypercube))**.** Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. The total influence of f, denoted $I[f]$, is:

$$I[f] = \langle f, Lf \rangle.$$

Expressed in this sense, there is a natural generalization to simplicial complexes. It is not hard to see that on the hypercube complex, Q_n^{lazy} is exactly the lower walk UD. As a result, we'll define influence using the Laplacian of the lower walk.

Definition 1.10.3 (Total Influence)**.** Let (X, Π) be a pure, weighed simplicial complex and $f \in C_k$. The influence of f, denoted $I[f]$ is:

$$I_{(X,\Pi)}[f] = \langle f, L_{UD} f \rangle$$

where $L_{UD} = k(I - U_{k-1}D_k)$. When clear from context, we will simply write $I[f]$.

When (X, Π) is sufficiently expanding, Definition 1.10.3 acts much like standard influence on the cube. For instance, recalling standard bounds on the spectral expansion of L_{UD} [239], it is not hard to see the total influence of any function on a γ-local-spectral expander lies between $(1+O_k(\gamma))\text{Var}(f) \leq I_{(X,\Pi)}[f] \leq k\text{Var}(f)$, which returns the standard bounds as γ goes to 0. Similarly, it is obvious that the total influence of any function on the hypercube complex is equivalent to its total influence on the hypercube, as the lower walk $U_{n-1}D_n$ is exactly Q_n^{lazy}.

Observation 1.10.4. Let $f : \{0,1\}^n \to \mathbb{R}$ be any function and $f_X : X_{\{0,1\}^n}(n) \to \mathbb{R}$ its equivalent on the hypercube complex, that is:

$$f(x_1, \ldots, x_n) = f_X((1, x_1), \ldots, (n, x_n))$$

Then:

$$I_{X_{\{0,1\}^n}}[f] = I[f].$$

One of the most well-studied problems in the analysis of boolean functions is understanding the structure of functions with low influence. The seminal result in this area is called the "KKL Theorem" [226]. Informally, the KKL Theorem states that if a function has low total influence, there must exist an *influential coordinate* (in the sense that on average over $\{0,1\}^n$, the coordinate has a large affect on the value of f). Morally, this can also be thought of as strong notion of the following statement: "functions with low influence are not pseudorandom." While the KKL Theorem itself does not extend beyond the hypercube (at least in unbalanced settings), this latter interpretation does. In particular, Bourgain [158] proved a similar statement over any product space: functions with low influence must have some influential *set* of coordinates, and are therefore not pseudorandom. We prove a variant of Bourgain's result for local-spectral expanders.

Theorem 1.10.5 (Bourgain's Theorem for HDX)**.** *Let (X, Π) be a two-sided γ-local-spectral expander with $\gamma \leq 2^{-\Omega(k)}$ and $f \in C_k$ a boolean function. Then for any $0 \leq K \leq k$, if $I[f] \leq K\,Var(f)$, there exists an $(i \leq K)$-link τ with large density:*

$$\mathop{\mathbb{E}}_{X_\tau}[f] \geq 2^{-O(K)}.$$

Proof. This follows without too much difficulty from the expansion of pseudorandom sets (Theorem 1.9.3). In particular, notice that our assumption on the influence implies the following bound on the expansion of f with respect to the lower walk $U_{k-1}D_k$:

$$\Phi(f) = \frac{\langle f, L_{UD} f\rangle}{k\,\mathbb{E}[f]} = \frac{I[f](1-\mathbb{E}[f])}{k\mathrm{Var}(f)} \leq \frac{K}{k}.$$

Recall that Theorem 1.9.3 states that for any $\delta > 0$ and $r = R_\delta(UD) - 1$, the expansion

of an (ε, r)-pseudorandom boolean function g with respect to the lower walk is at least:

$$\Phi(g) \geq 1 - \delta - (1-\delta)2^{O(r)}\varepsilon^{1/3} - c\gamma.$$

Using this fact, we'll show that f cannot be (ε, K)-pseudorandom for $\varepsilon \leq 2^{-\Omega(K)}$, which gives the result.

To this end, assume f is (ε, K)-pseudorandom for some $\varepsilon = 2^{-\Omega(K)}$ to be determined soon (else we are done), and let δ be $1 - \frac{K(1+\epsilon^{1/6})}{k}$ such that $1 - \delta < (K+1)/k$. Since the eigenvalues of UD are concentrated around $1, 1-1/k, 1-2/k, \cdots, 1-K/k$ for small enough γ [38], the ST-Rank $R_\delta(UD) = K+1$, and $r = K$. Theorem 1.9.3 then implies:

$$\Phi(f) \geq \frac{K}{k} \cdot (1+\epsilon^{1/6})(1 - 2^{O(r)}\epsilon^{1/3}),$$

where we have again used our assumption on the size of γ. Re-arranging the above using the upper bound on expansion then gives a lower bound on ε of:

$$\epsilon^{1/3} \geq \frac{\epsilon^{1/6}}{1+\epsilon^{1/6}} \cdot \frac{1}{2^{O(r)}} \geq \frac{1}{2^{O(r)}},$$

which implies the result.

□

Before moving on, we'll prove that this result is tight.

Proposition 1.10.6. *Let $c \geq 1$ be any constant. Then for all integers $K, k > 1$ satisfying $k \geq \Omega_c(K)$ and any n sufficiently larger than k, there exists a Boolean function $f \in C_k$ on the k-dimensional complete complex on n vertices satisfying:*

1. *The influence of f is small:*

$$I[f] \leq K\, Var(f)$$

2. For every $i \leq cK$, all i-links are sparse:

$$\forall i \leq cK, \tau \in X(i) : \mathop{\mathbb{E}}_{X_\tau}[f] \leq 2^{-\Omega(K)}.$$

Proof. Our construction is based on a careful analysis of the anti-tribes function (a.k.a the AND of ORs function) similar to [245, Example 5.8]. Concretely, let $T_1, \ldots T_m$ (called "tribes") be $m = 2cK$ disjoint sets of $c_1 \frac{n}{k}$ vertices for some $c_1 \geq \log(\Omega(c))$. We define our candidate function $f \in C_k$ to be 1 on a k-face S exactly when S contains some vertex from each T_i:

$$f(S) = \begin{cases} 1 & \text{if } \forall 1 \leq i \leq m\text{: } |S \cap T_i| > 0 \\ 0 & \text{else.} \end{cases}$$

For simplicity, it will actually be more convenient to analyze f as a function over $[n]^k$ rather than $X(k) = \binom{n}{[k]}$. Since the probability of repeated vertices in the former is $o_{n,k}(1)$, this has no effect on our final result when n is sufficiently larger than k.

Let's start by proving the density of every cK-link is at most $2^{-\Omega(k)}$.[20] Note that this implies the same for every i-link for $i \leq cK$. It is not hard to see that the largest density link comes from fixing an element in each of cK tribes. For simplicity, fix such a cK-link T with a vertex in each T_i for $cK + 1 \leq i \leq m$ (all such links are symmetric, so it suffices to analyze this case). For a uniformly drawn element $S \in [n]^k$, let E_i denote the event that S contains a vertex in T_i. We'd like to bound:

$$\mathop{\mathbb{E}}_{X_T}[f] = \mathop{\mathbb{P}}_{S \sim [n]^k}\left[\bigcap_{i=1}^{cK} E_i \;\middle|\; S \supset T\right] = \mathop{\mathbb{P}}_{S \sim [n]^{k-cK}}\left[\bigcap_{i=1}^{cK} E_i\right],$$

where we have used the fact that $S \setminus T$ is independent of T since we are working over $[n]^k$.

[20] Formally we note this should really be shown for $\lfloor cK \rfloor$-links, but this makes no significant difference in the analysis so we ignore it for simplicity.

Since the E_i are negatively correlated, we can bound this probability by:

$$\begin{aligned}\mathop{\mathbb{P}}_{S\sim[n]^{k-cK}}\left[\bigcap_{i=1}^{cK} E_i\right] &\leq \prod_{i=1}^{cK} \mathop{\mathbb{P}}_{S\sim[n]^{k-cK}}[E_i] \\ &\leq (1-(1-c_1/k)^k)^{cK} \\ &\leq \left(1-\frac{1}{O(c)}\right)^{cK} \\ &\leq 2^{-\Omega(K)}\end{aligned}$$

where we've used the fact that $e^{-x} \geq 1-x \geq e^{-x/(1-x)}$ for $x < 1$ and our assumptions on the size of k.

We now move on to analyzing the influence of f, which will follow from similar computations. To start, notice that it is instead sufficient to bound the expansion of f with respect to the lower walk by:

$$\Phi(f) \leq \frac{K}{k}\frac{\mathrm{Var}(f)}{\mathbb{E}[f]} = \frac{K(1-\mathbb{E}[f])}{k},$$

as then:

$$I[f] = \langle f, L_{UD} f\rangle = k\Phi(f)\mathbb{E}[f] \leq K\mathrm{Var}(f),$$

where we recall L_{UD} is the un-normalized Laplacian of UD.

To this end, recall that the expansion of f can also be defined as the average probability of leaving $\mathrm{supp}(f)$ after applying the walk, that is:

$$\Phi(f) = \mathop{\mathbb{E}}_{S\sim\mathrm{supp}(f)}[\phi(S)],$$

where $\phi(S)$ denotes the probability of leaving S in a single step of the lower walk. To compute this value, recall that in the down step of the walk, a uniformly random vertex is

removed from S. In order to leave the support of f in the up step, the removed element must have been selected from a tribe T_i such that $|S \cap T_i| = 1$. The idea is then to show that for most samples, only a small fraction of tribes have exactly one element. With this in mind, let B_i be the event $|S \cap T_i| = 1$ over the randomness of $S \sim \text{supp}(f)$. Formalizing the above argument, we can bound $\phi(S)$ by the sum over B_i:

$$\phi(S) \leq \sum_{i=1}^{m} \frac{B_i(S)}{k},$$

and therefore the expansion $\Phi(f)$ by:

$$\Phi(f) \leq \frac{1}{k} \mathop{\mathbb{E}}_{S \sim \text{supp}(f)} [B_i(S)].$$

By a similar argument to our density calculations, the probability that any fixed tribe T_i has exactly one element from $S \sim \text{supp}(f)$ is at most:

$$\begin{aligned} \mathbb{E}[B_i] &= \left(1 - \frac{c_1}{k}\right)^{k-m} \\ &\leq e^{-c_1 \frac{k-m}{k}} \\ &\leq \frac{1}{\Omega(c)} \end{aligned}$$

since we have by assumption that k is much larger than m. Plugging this into our expression for expansion then gives:

$$\Phi(f) \leq \frac{m}{k} \cdot \frac{1}{\Omega(c)} \leq c_2 \frac{K}{k}$$

for some $c_2 < 1$. Noting that $\mathbb{E}[f] = 2^{-\Omega(K)}$ then implies the result for the appropriate setting of constants. $\square$

1.10.2 Stability and the Noise Operator

Another fundamental notion in boolean Fourier analysis is the *noise operator* T_ρ. It is convenient to express the definition in terms of the following process on an element $x \in \{0,1\}^n$:

1. Remove each bit with probability $1-\rho$.
2. Replace each removed bit uniformly[21] at random.

We write the distribution over y given by this process as $N_\rho(x)$. The noise operator T_ρ is simply the averaging operator over ρ-correlated strings.

Definition 1.10.7 (Noise Operator (Hypercube))**.** Let $f : \{0,1\} \to \mathbb{R}$ be any function. The noise operator T_ρ averages f over N_ρ:

$$T_\rho f(x) = \mathop{\mathbb{E}}_{y \sim N_\rho(x)}[f(y)].$$

To extend the noise operator to simplicial complexes, consider the following reformulation of the distribution $N_\rho(x)$: instead of removing each coordinate independently with probability $1-\rho$, we remove a uniformly random *set of i coordinates* with probability $\binom{n}{i}\rho^{n-i}(1-\rho)^i$, and replace them uniformly at random. This equivalent process does have a natural analog on simplicial complexes: simply replace "uniformly random set of i coordinates" with "uniformly random i-face." We can formalize this through the averaging operators.

Definition 1.10.8 (Noise Operator (Simplicial Complex))**.** Let (X, Π) be a pure, weighted simplicial complex. The noise operator $T^k_\rho(X, \Pi)$ at level k of the complex is:

$$T^k_\rho(X, \Pi) = \sum_{i=0}^{k} \binom{k}{i} (1-\rho)^i \rho^{k-i} U^k_{k-i} D^k_{k-i}.$$

[21] In more general settings like the p-biased cube, this is replaced with respect to the underlying distribution.

We write T_ρ when the level and complex are clear from context.

Let's take a moment to check that, as with influence, when applied to the hypercube complex this definition recovers T_ρ.

Observation 1.10.9. Let $f : \{0,1\}^n \to \mathbb{R}$ be any function and $f_X : X_{\{0,1\}^n}(n) \to \mathbb{R}$ its equivalent on the hypercube complex, then:

$$T_\rho^n(X_{\{0,1\}^n}) f_X = T_\rho f.$$

Proof. $T_\rho^n(X_{\{0,1\}^n})$ is also an averaging operator, so it is enough to confirm it averages over the ρ-noisy distribution N_ρ. We claim this is clear from definition. In particular, notice that $U_{n-i}^n D_{n-i}^n$ on $X_{\{0,1\}^n}$ is exactly the process of removing i-coordinates uniformly at random, and replacing them with uniformly random bits. As we mentioned above, this is an equivalent way to define $N_\rho(x)$, is applying this process with probability $\binom{n}{i}(1-\rho)^i \rho^{n-i}$, which exactly matches the definition of $T_\rho(X_{\{0,1\}^n})$. □

The noise operator has a wide variety of applications across boolean Fourier analysis. One classical application is to analyze the *noise-sensitivity* of a boolean function, that is the likelihood that the function flips on a noisy input. It is convenient to define the opposite concept first, *stability*.

Definition 1.10.10 (Stability (Hypercube))**.** Let $f : \{0,1\} \to \mathbb{R}$ be any function. The stability of f with respect to ρ, denoted $\mathrm{Stab}_\rho(f)$, is:

$$\mathrm{Stab}_\rho(f) = \langle f, T_\rho f \rangle.$$

Since we already defined T_ρ on simplicial complexes, stability has an obvious analog.

Definition 1.10.11 (Stability (Simplicial Complex))**.** Let (X, Π) be a weighted, pure simplicial complex and $f \in C_k$. The noise stability of f with respect to ρ, denoted $\mathrm{Stab}_\rho(f)$, is:

$$\mathrm{Stab}_\rho^{(X,\Pi)}(f) = \langle f, T_\rho^k(X, \Pi) f \rangle.$$

We drop (X, Π) from the notation when clear from context.

Similarly, it is clear that our definition of stability for complexes returns the original definition when applied to the hypercube complex.

Observation 1.10.12. Let $f : \{0,1\}^n \to \mathbb{R}$ be any function and $f_X : X_{\{0,1\}^n}(n) \to \mathbb{R}$ its equivalent on the hypercube complex, then:

$$\mathrm{Stab}_\rho(f) = \mathrm{Stab}_\rho^{X_{\{0,1\}^n}}(f_X)$$

A function is called *noise-sensitive* if is has poor stability. One classical result in boolean Fourier analysis is that sparse functions on the hypercube are noise-sensitive, which is equivalent to saying that the noisy hypercube graph is a small-set expander. Since the noise operator is just a specific instance of a (complete) higher order random walk, Theorem 1.9.3 implies an analogous statement for functions on HDX: pseudorandom functions are noise-sensitive.

Corollary 1.10.13 (Pseudorandom functions are noise sensitive)**.** *Let (X, Π) be a two-sided γ-local-spectral expander, $f \in C_k$ an (r, δ)-pseudorandom boolean function for $\delta \leq 2^{-\Omega(r)}\epsilon^3$ and $r = \log(2/\epsilon)/\log(1/\rho) + 2$. Then f is noise sensitive:*

$$Stab_\rho(f) \leq (\epsilon + c\gamma)\,\mathbb{E}[f]$$

for $c \leq 2^{O(k)}$.

Proof. One can directly compute from [38, Corollary 7.6] that the approximate eigenvalues

of T_ρ are exactly $\lambda_i = \rho^i$. As a result, for small enough γ, the $(\varepsilon/2)$-ST-Rank of T_ρ is at most:

$$R_{\varepsilon/2}(T_\rho) \leq \log(\varepsilon/2)/\log(1/\rho) + 2.$$

Since T_ρ is a higher order random walk, Theorem 1.9.3 states that the non-expansion of any (δ, r)-pseudorandom function f of density α is at most:

$$\begin{aligned}\frac{1}{\alpha}\langle f, T_\rho f\rangle &\leq \alpha + (1-\alpha)\varepsilon/2 + 2^{O(r)}\delta + c\gamma \\ &\leq \alpha + (1-\alpha)\varepsilon/2 + \varepsilon/4 + c\gamma \\ &\leq \varepsilon + c\gamma,\end{aligned}$$

where we've used the fact that $\alpha \leq \delta \leq \varepsilon/4$. □

The noise operator is actually also commonly used to define hypercontractivity. In this form, the standard hypercontractive inequality generally states:

$$\|T_\rho f\|_4 \leq \|f\|_2$$

for some $\rho = \Theta(1)$. It is well known that on the hypercube this statement is in fact equivalent to Bonami's lemma. We can show a similar equivalence between our variant of Bonami's lemma (Theorem 1.8.1) and a noise operator based form of hypercontractivity for pseudorandom functions. To state the strongest form of the result, it will be useful to extend the classic notion of *degree* to simplicial complexes.

Definition 1.10.14 (Function Degree)**.** Let (X, Π) be a pure, weighted simplicial complex, and $f \in C_k$ any function. The degree of f, denoted $\deg(f)$, is the largest i such that $f_{\uparrow i}$ is non-zero.

We now show how to translate Theorem 1.8.1 into noise operator form for degree i, (ε, i)-pseudorandom functions.

Proposition 1.10.15. *Let (X, Π) be a two-sided γ-local-spectral expander satisfying $\gamma \leq 2^{-\Omega(k)}$ and $f \in C_k$ a degree i,[22] (ε, i)-pseudorandom function. Then for some constant $\rho = \Theta(1)$, we have:*

$$\|T_\rho f\|_4^4 \leq \epsilon \|f\|_2^2 \|f\|_\infty^2.$$

Proof. The overall proof follows a fairly standard reduction from hypercontractivity to Bonami's lemma (see e.g. [308, Exercise 9.6]), but requires some extra work due to the fact that the Bottom-Up decomposition is only approximately an eigenbasis for T_ρ (in an ℓ_2-sense). Namely, by [38, Proposition 7.5] and Lemma 1.7.5, we can write:

$$T_\rho f_{\uparrow j} = \rho^k f_{\uparrow j} + \mathrm{err}_j$$

where $\|\mathrm{err}_j\|_2 \leq k^{O(j)}\gamma\|f\|_2$, and $\|\mathrm{err}_j\|_\infty \leq k^{O(j)}\varepsilon\|f\|_\infty$. The last of these facts is slightly less standard, and follows from noting that err_j is really a linear combination of at most $k^{O(j)}$ averaging operators applied to g_j (see [38, Proposition 7.5]), and that $\|g_j\|_\infty \leq 2^{O(j)}\varepsilon\|f\|_\infty$. With this in mind, we can expand out $\|T_\rho f\|$ by the Bottom-Up Decomposition and apply Theorem 1.8.1 to get:

$$\begin{aligned}
\|T_\rho f\|_4 &\leq \sum_{j=0}^{i} \|T_\rho f_{\uparrow j}\|_4 \\
&\leq \sum_{j=0}^{i} \rho^i \|f_{\uparrow j}\|_4 + \|\mathrm{err}_j\|_4 \\
&\leq \frac{1}{2}\varepsilon^{1/4}\|f\|_2^{1/2}\|f\|_\infty^{1/2} + \sum_{j=0}^{i} \|\mathrm{err}_j\|_4 \\
&\leq \frac{1}{2}\varepsilon^{1/4}\|f\|_2^{1/2}\|f\|_\infty^{1/2} + \sum_{j=0}^{i} \|\mathrm{err}_j\|_2^{1/2}\|\mathrm{err}_j\|_\infty^{1/2} \\
&\leq \varepsilon^{1/4}\|f\|_2^{1/2}\|f\|_\infty^{1/2}
\end{aligned}$$

[22] When $i \ll k$, we can replace the condition on γ with $\gamma \leq k^{-\Omega(i)}$

where we have assumed that ρ is a sufficiently small constant. Taking the fourth power of both sides completes the proof. $\square$

This chapter, in full, is based on the material as it appears in Symposium on Theory of Computing 2022. Bafna, Mitali; Hopkins, Max; Kaufman, Tali; Lovett, Shachar. "Hypercontractivity on High Dimensional Expanders". The dissertation author was a primary investigator and author of this material.

Chapter 2

Hypercontractivity on HDX II: Symmetrization and q-Norms

2.1 Introduction

Recent years have seen the development of a powerful theory of boolean function analysis *beyond the cube*, with breakthrough applications in extremal and probabilistic combinatorics [158, 139], quantum communication [33], approximate sampling [24, 96, 22], and hardness of approximation [128, 126, 255, 297]. Among the most powerful tools in boolean analysis, and the core of all above applications, is the theory of *hypercontractivity* [71], an analytic inequality bounding the high norm behavior of low degree functions on the cube. Popularized in the 80s by Kahn, Kalai, and Linial [226], variants of hypercontractivity have since been established for many objects, including products [78, 349, 160, 158], the slice [252], the shortcode [46], the Grassmann [255, 138], and various groups [147, 139].

Many potential applications of hypercontractivity (e.g. to low-soundness PCPs [128], sparsest cut [227], the unique games conjecture [47]) are still hindered by the fact that most known hypercontractive inequalities only hold for *dense* domains. Motivated in this context, Gur, Lifshitz, and Liu [187] and Bafna, Hopkins, Kaufman, and Lovett [39] recently proved the first hypercontractive inequalities for sparse systems, in particular for the class of hypergraphs known as *high dimensional expanders* (HDX), an extension of expander graphs to high dimensions that have garnered significant interest in boolean

analysis and hardness of approximation [124, 111, 38, 119, 162, 208, 40, 113]. Their works gave the first example of hypercontractivity for sparse domains, but fell short of known inequalities on products: suffering from sub-optimal dependence on degree (a critical parameter in application),[1] and requiring restrictive assumptions on the function in question.

In this work, inspired by O'Donnell and Zhao's [364] proof of hypercontractivity for products, we take a new approach for high dimensional expanders that completely avoids these issues: *symmetrization*. Symmetrization is a classical technique of Kahane [225] and Bourgain [77] which analyzes high moment behavior of functions on product spaces $f : \Omega_i^{\otimes d} \to \mathbb{R}$ by convolving them with a random boolean string $r \in \{\pm 1\}^d$. Up to the application of a small amount of noise, Bourgain proved that the resulting function $\tilde{f} : \{\pm 1\}^d \times \Omega_i^{\otimes n} \to \mathbb{R}$ has the same behavior of f, allowing one to analyze $\tilde{f}$ instead using standard analysis on its boolean component. We prove Bourgain's theorem holds for sufficiently strong high dimensional expanders up to a $(1 \pm o(1))$ factor. As applications, we prove an optimal hypercontractive inequality for high dimensional expanders, and a booster theorem for general low influence functions generalizing Bourgain's classical result on products [158]. This resolves two of the main open questions of [187, 39]

Our proof of the symmetrization theorem is based on two new ideas in the theory of high dimensional expansion. First, we introduce the notion of *q-norm HDX*, which bounds the local divergence of a hypergraph from a product space in q-norm, and show any strong enough spectral HDX is a q-norm HDX. This allows us to manipulate higher norms directly, avoiding expensive error terms appearing in [39, 187]. Second, we introduce a basic variant of coordinate-wise analysis for the noise operator T_r on q-norm HDX, showing it can be approximately broken into coordinate-wise components $T^i_{r_i}$ that correctly localize to the standard 1-dimensional operators T_{r_i} when restricted to the ith coordinate. This allows

[1] Technically [39] do achieve the correct dependence, but only for the much stricter class of two-sided HDX.

for the replacement method as used in Bourgain's original proof of the symmetrization theorem, greatly simplifying prior analysis.

2.1.1 Background

High Dimensional Expanders:.

We study (weighted) *partite hypergraphs* $X \subset [n]^d$, which we view as d-dimensional distributions $X = (X_1, \ldots, X_d)$. X is called a γ-*product*[2] if for every $i \neq j$

1. The bipartite graph corresponding to the marginal (X_i, X_j) is a γ-spectral expander
2. This holds even conditioning on any (feasible) values of X_S for any $i, j \notin S \subseteq [d]$.

We write the space of functions on X as $C_d := \{f : X \to \mathbb{R}\}$. The weights of X, denoted $\Pi(x)$, induce a natural expectation operator and L_p-space over C_d defined by $\mathbb{E}[f] = \sum_{x \in X} \Pi(x) f(x)$ and $\|f\|_p = \mathbb{E}[|f|^p]^{1/p}$. Given a linear operator $T : C_d \to C_d$, we define its p-norm in the standard way as $\max_{f \in C_d} \frac{\|Tf\|_p}{\|f\|_p}$.

Noise Operator and Total Influence:.

The *noise operator* $T_\rho : C_d \to C_d$ is a classical smoothing procedure in boolean analysis which given a string $(x_1, \ldots, x_d) \in X$, averages over y generated by re-sampling each coordinate with probability $1 - \rho$. The *total influence* of a $\{\pm 1\}$-valued function $f \in C_d$ measures its average sensitivity to re-sampling an *individual* coordinate

$$\mathbf{I}[f] = 2 \sum_{i=1}^{d} \mathbb{E}_x[\mathbb{P}[f(x) \neq f(x^{(i)})]],$$

where $x^{(i)}$ is generated by re-sampling the ith coordinate of x.

[2] γ-products, introduced formally in [187], are essentially equivalent to the standard notion of spectral HDX [109].

Fourier Analysis, Symmetrization, and the Generalized Noise Operator:.

The classical theory of Fourier analysis on products is given by the *Efron-Stein Decomposition*, whose components $\{f^{=S}\}_{S\subseteq[d]}$ are given by

$$f^{=S}(y) = \sum_{T\subseteq S}(-1)^{|S\setminus T|}\mathbb{E}[f(x) \mid x_T = y_T]$$

It is easy to check that $f = \sum_{S\subseteq[d]} f^{=S}$. We write $f^{\leq i} = \sum_{|S|\leq i} f^{=S}$ as the degree-i component of f. On product spaces, $\{f^{=S}\}$ is an orthogonal eigenbasis for T_ρ:

$$T_\rho f = \sum_{S\subseteq[d]} \rho^{|S|} f^{=S}$$

This leads to a standard generalization of the noise operator to vectors $r \in \mathbb{R}^d$

$$T_r f \coloneqq \sum_{S\subseteq[d]} r_S f^{=S},$$

where $r_S = \prod_{i\in S} r_i$. Finally, the *symmetrization* of f, denoted $\tilde{f} : \{\pm 1\}^d \times X \to \mathbb{R}$ is the function

$$\widetilde{f}(r, x) = T_r f(x) = \sum_{S\subseteq[d]} r_S f^{=S}(x).$$

Note $\tilde{f}$ is the unique function whose X-restrictions $\tilde{f}|_x(r) = \tilde{f}(x, r)$ have Fourier coefficients $f^{=S}(x)$.

2.1.2 Results

Our main contribution is the following extension of Bourgain's symmetrization theorem to HDX:

Theorem 2.1.1 (The Symmetrization Theorem (Informal Corollary 2.5.2))**.** *Let $q > 1$*

and X be a d-partite γ-product satisfying $\gamma \leq 2^{-\Omega_q(d)}$. Then for any $f : X \to \mathbb{R}$

$$(1 - o_\gamma(1))\|\widetilde{T_{c_q} f}\|_q \leq \|f\|_q \leq (1 + o_\gamma(1))\|\widetilde{T_2 f}\|_q$$

The symmetrization theorem is a powerful tool for extending classical structure theorems on the cube to product spaces. In this work, we focus mainly on its application to the *hypercontractive inequalities*. On the boolean cube, the simplest form of hypercontractivity (called the Bonami Lemma [71]) states that the degree-i part of any function $f : \{0,1\}^d \to \mathbb{R}$ is 'smooth' in the sense that

$$\|f^{\leq i}\|_4 \leq 2^{O(i)}\|f^{\leq i}\|_2.$$

This simple statement has far reaching consequences in analysis and theoretical computer science, including the famous KKL Theorem [226] and its many applications.

Unfortunately, the Bonami lemma is false on unbalanced product spaces. The obvious counter-examples are dictators, or more generally 'local' functions that are dense in some restriction. Consider, for instance, the ith dictator on the p-biased cube. While $\mathbf{1}_i$ is a degree-1 function, it is easy to compute:

$$\mathbb{E}[\mathbf{1}_i^4] = p \gg \mathbb{E}[\mathbf{1}_i^2]^2 = p^2.$$

Global hypercontractivity aims to show that these local functions are the *only* counter-examples. In other words, as long as f isn't too dense in any restriction, the Bonami lemma should still hold. Variants of such a bound have been known since the 90s [158], but a tight bound was only shown recently by O'Donnell and Zhao [364] and Keevash, Lifshitz, Long, and Minzer [248]. Using symmetrization, we prove strong enough partite high dimensional expanders enjoy essentially the same bound.

Theorem 2.1.2 (Global Hypercontractivity on HDX (Bonami Form, Informal Theorem 2.6.1)). *Let X be a d-partite γ-product satisfying $\gamma \leq 2^{-\Omega(d)}$. Then for any $f : X \to \mathbb{R}$ and $i \leq d$:*

$$\|f^{\leq i}\|_4^4 \leq 2^{O(i)}\|f\|_2^2 \max_{|S|\leq i, x_S}\{\|f|_{x_S}\|_2^2\}$$

In other words, so long as f is not dense upon restricting the values of any i coordinates, i.e. $\|f|_{x_S}\|_2^2 \leq 2^{O(i)}\|f\|_2^2$, we recover the standard Bonami lemma. Note that on a balanced product space such as the cube, such a bound indeed holds generically for all functions.[3] Theorem 2.1.2 improves on the best known bounds for γ-products [187] which scale with $i^{O(i)}$ and have an additional error term scaling with $\gamma\|f\|_\infty$, limiting their regime of application. We discuss further in Section 2.1.4.

As hinted above, the $2^{O(i)}$ dependence in Theorem 2.1.2 is tight. We emphasize this dependence is often critical in application because it matches the eigenvalues of the noise operator, allowing one to transfer to a variant of the classical operator-form of hypercontractivity.

Corollary 2.1.3 (Global Hypercontractivity (Operator Form, Informal Corollary 2.6.2)). *Let X be a d-partite γ-product satisfying $\gamma \leq 2^{-\Omega(d)}$. There exists a constant $\rho \in (0, 1)$ such that for any mean-0 function $f : X \to \mathbb{R}$ and $i \leq d$*

$$\|T_\rho f^{\leq i}\|_4^4 \leq \|f\|_2^2 \max_{|S|\leq i, x_S}\{\|f|_{x_S}\|_2^2\}$$

Combined with the Fourier analytic machinery of [187], these results immediately give a tight characterization of low influence functions on partite HDX.

Corollary 2.1.4 (Low Influence Functions are Local (Informal Corollary 2.6.3)). *Let X be a d-partite γ-product and $f : X \to \mathbb{F}_2$ a constant variance function satisfying $\mathbf{I}[f] \leq K$,*

[3] In fact even on unbalanced products the same argument recovers so-called 'generalized hypercontractivity', which scales with the worst marginal probability of the product, see [308].

then $\exists S \subset [d]$ *with* $|S| \leq O(K)$ *and* $x_S \in X[S]$ *such that*

$$\mathbb{E}[f|_{x_S}] \geq 2^{-O(K)}$$

Corollary 2.1.4 is only meaningful when f is sparse (namely $\mathbb{E}[f] \leq 2^{-\Omega(K)}$), otherwise it is trivially the case there exist dense restrictions. One of the main open questions raised in [39] was whether an analogous result could be proved for any density function. The classical such result on product spaces is Bourgain's booster theorem, which states that any low influence function has many restrictions on which it *deviates* substantially from its expectation. We extend this result to HDX.

Theorem 2.1.5 (A Booster Theorem for HDX (Informal Theorem 2.7.1))**.** *Let* X *be a* d*-partite* γ*-product with* $\gamma \leq 2^{-\Omega(d)}$*, and* $f : X \to \{\pm 1\}$ *a constant variance function satisfying* $\mathbf{I}[f] \leq K$*. Then*

$$\mathbb{P}_{x \sim X}\left[\exists T \subset [d] : \ |T| \leq O(K), \ |\mathbb{E}[f|_{x_T}] - \mathbb{E}[f]| > 2^{-O(K^2)}\right] \geq 2^{-O(K^2)}$$

Classically, Bourgain's booster theorem lead to the famous theory of sharp thresholds for graph properties [158]. It is an interesting open problem whether Theorem 2.1.2 or Theorem 2.1.5 could be used in this context.

2.1.3 Technical Overview

Our proof of the symmetrization theorem relies on two elementary new tools in the theory of high dimensional expansion. We first introduce our notion of 'q-norm HDX' which bounds the local divergence of X from a product in q-norm. Second, we introduce our coordinate-wise treatment of the noise operator, and sketch how the technique is used to prove the symmetrization theorem. Finally, we sketch how symmetrization is used to prove optimal global hypercontractivity, combining the elegant argument of O'Donnell

and Zhao [364] for products with distinct elements of [187]'s method and [39]'s proof of hypercontractivity for two-sided HDX. We omit any technical overview of the booster theorem which follows from applying similar tricks to Bourgain's original proof for product spaces.

(q, γ)-Products:.

A core problem identified in [187] is that standard high dimensional expanders and γ-products only bound the *spectral* behavior of the complex, which a priori seems insufficient to control higher norm behavior. [187] handle this issue in part by losing factors in infinity norm, e.g. by relating q and 2 norms via the elementary bound $\|f\|_q^q \leq \|f\|_2^2 \|f\|_\infty^{q-2}$. Unfortunately, this method leads to an error term which dominates the main expression for strongly global f, and is insufficient for our purposes.

We handle this and analogous issues by introducing q-norm HDX. Given a partite complex X, let $A_{i,j}$ denote the (normalized) bi-partite adjacency matrix of (X_i, X_j), and $\Pi_{i,j}$ be $A_{i,j}$'s stationary operator.[4]

Definition 2.1.6 ((q, γ)-Products)**.** A d-partite complex X is a (q, γ)-product if for every distinct $i, j \in [d]$:

1. The marginal (X_i, X_j) is 'q-norm-expander'

$$\|A_{i,j} - \Pi_{i,j}\|_q \leq \gamma$$

2. This holds under all feasible conditionings $X_S = z_S$.

Note that setting $q = 2$ exactly recovers the notion of a γ-product. While q-norms beyond the spectral setting are typically much harder to analyze, an elementary application of Riesz-Thorin interpolation actually implies any sufficiently strong γ-product is also a (q, γ_q)-product.

[4]In this case, this is simply the matrix where every row is the marginal distribution over X_i. In the feasible conditioning, this is replaced with the conditioned marginal over X_i (similarly for $A_{i,j}$).

Lemma 2.1.7 (Informal Lemma 2.3.4)**.** *Any γ-product is a (q, γ_q)-product for $\gamma_q \leq \gamma^{2/q} 2^{1-2/q}$.*

This simple observation allows us to handle q-norms directly, avoiding dependence on $\|f\|_\infty$.

Coordinate-Wise Analysis on HDX:.

The second critical tool in our analysis is a simplified method of coordinate-wise analysis on HDX, inspired by Bourgain's proof of the symmetrization theorem. We focus in particular on the noise operator, though the same method can be applied to other random walks on HDX.[5]

On a product space, the noise operator T_ρ can naturally be expressed as the product of coordinate-wise operators T^i_ρ. When $\rho \in [0, 1]$, this simply corresponds to re-sampling the ith coordinate with probability $1 - \rho$, and keep all other coordinates fixed. For general $r \in \mathbb{R}^d$, the corresponding coordinate noise operator is often defined as

$$T^i_{r_i} f \coloneqq \sum_{S \ni i} r_i f^{=S} + \sum_{S \not\ni i} f^{=S}.$$

Unfortunately, while this generalization makes manipulation of coordinate-wise noise operators simpler, it does not interact well with γ-products where the Efron-Stein decomposition is not closed (i.e. the Efron-Stein decomposition of $f^{=S}$ itself is not exactly $f^{=S}$). Instead, it turns out when dealing with general partite complexes, one should instead extend T_r by generalizing its form in terms of the *projection operators*

$$E_S f(y) = \mathbb{E}_x[f(x) \mid x_S = y_S],$$

which re-sample all coordinates outside of S. The extended noise operator may then be

[5]Indeed 'coordinate-wise' or 'tensorization' methods have appeared in the sampling literature analyzing entropic inequalities for the so-called down-up walk on related objects, though not in the sense we consider. See Section 2.1.4 for further discussion.

defined as a 'binomial' combination of projection operators:

$$T_r = \sum_{S \subseteq [d]} r_S \prod_{i \notin S} (1 - r_i) E_S,$$

where $r_S = \prod_{i \in S} r_i$. The coordinate-wise noise operators are then naturally defined as $T^i_{r_i} \coloneqq r_i I + (1 - r_i) E_{[d]\setminus i}$. On a product space, it is easy to check that for either definition $T_r = T^1_{r_1} \dots T^d_{r_d}$. Our first key lemma shows this continues to hold approximately on (q, γ)-products.

Lemma 2.1.8 (Decorrelation (Informal Lemma 2.4.2))**.** *Let X be a d-partite (q, γ)-product and $r \in \mathbb{R}^d$. Then*

$$\|T_r f - T^1_{r_1} \dots T^d_{r_d} f\|_q \leq O_\gamma(\|f\|_q)$$

Note this holds for any ordering of $T^i_{r_i}$, and does not strongly rely on which definition we take. The core reason to use the operator-extension is that this is the notion which *localizes* correctly to the marginals of X.

Lemma 2.1.9 (Localization (Informal Lemma 2.4.3))**.** *Let X be a d-partite complex, and $r \in \mathbb{R}^d$. Then for any $S \subseteq [d]$:*

$$T^i_r f(x) = T_{r_i} f|_{x_{-i}}(x_i),$$

where $f|_{x_{-i}} : X[i] \to \mathbb{R}$ is the localization of f to the marginal $X_{[d]\setminus\{i\}}$ defined by $f|_{x_{-i}}(x_i) = f(x)$.[6]

Decorrelation and localization allows us to reduce analysis of the 'high dimensional' quantity $T_\rho f(x)$ to the *1-dimensional quantities* $T^i_\rho f|_{x_{-i}}(x_i)$. One can then apply standard 1-dimensional inequalities, and 'lift' the result back to the original complex X by reversing the lemmas. This is a powerful (and very standard) trick on the hypercube and product

[6]Formally, this function lives on the *link* of x_{-i}, see Section 2.2.1.

spaces in boolean analysis—many classical results are proved via this type of reduction to the 1-D case (often properties proved this way, like hypercontractivity, are said to 'tensorize').

The Symmetrization Theorem:.

We now sketch the proof of the symmetrization theorem based on these three components: a (standard) 1-D version of the inequality, decorrelation into coordinate-wise operators, and localizing to the 1-D case. We will show only the upper bound; the lower bound follows via similar reasoning. The proof very closely follows the original proof of symmetrization (as presented in [308]).

Slightly re-phrasing the theorem statement, our goal is to show

$$\|T_{\frac{1}{2}} f\|_q \leq (1 + o_\gamma(1))\|T_r f\|_q.$$

By Lemma 2.1.8, we can break these terms into their coordinate-wise components and instead argue

$$\|T^1_{\frac{1}{2}} \dots T^d_{\frac{1}{2}} f\|_q \leq (1 + o_\gamma(1))\|T^1_{r_1} \dots T^d_{r_d} f\|_q$$

Written in this form, there is really only one natural strategy: the *replacement method*. In particular, we'll argue that each of the coordinate-wise operators $T^i_{1/2}$ can be sequentially replaced by $T^i_{r_i}$, incurring only $O(\gamma)\|f\|_q$ error in each step. Toward this end, define the 'partially symmetrized' operator

$$T^{(j)} \coloneqq T^1_{r_1}, \dots T^j_{r_j} T^{j+1}_{\frac{1}{2}} \dots T^d_{\frac{1}{2}}.$$

Since $T^{(0)} = T^1_{1/2} \dots T^d_{1/2} \approx T_\rho$, and $T^{(d)} = T^r_{r_1} \dots T^d_{r_d} \approx T_r$ by telsecoping it is enough to show for all j

$$\|T^{(j)} f\|_q \leq (1 + o_\gamma(1))\|T^{(j+1)} f\|_q$$

Un-wrapping the lefthand side, by Lemma 2.1.8 we can permute the order of the T_j's to move the $(j+1)$st operator to the front, and localize to the $(j+1)$st coordinate via Lemma 2.1.9

$$\begin{aligned}\|T^{(j)}f\|_q &= \|T^1_{r_1},\dots T^j_{r_j}T^{j+1}_{\frac{1}{2}}\dots T^d_{\frac{1}{2}}f\|_q \\ &\approx \|T^{j+1}_{\frac{1}{2}}(T^1_{r_1},\dots T^j_{r_j}T^{j+2}_{\frac{1}{2}}\dots T^d_{\frac{1}{2}}f)\|_q \\ &= \|\|T_{\frac{1}{2}}(T^1_{r_1},\dots T^j_{r_j}T^{j+2}_{\frac{1}{2}}\dots T^d_{\frac{1}{2}}f)|_{x_{-j}}\|_{q,x_j}\|_{q,x_{-j}}\end{aligned}$$

The inner norm is now over a 1-dimensional domain, so we can apply the standard 1-D symmetrization theorem to replace $T_{1/2}g$ with $T_{r_{j+1}}g$, then 'un'-localize and 'un'-permute by the reverse directions of Lemma 2.1.9 and Lemma 2.1.8:

$$\begin{aligned}&\|\|T_{\frac{1}{2}}(T^1_{r_1},\dots T^j_{r_j}T^{j+2}_{\frac{1}{2}}\dots T^d_{\frac{1}{2}}f)|_{x_{-j}}\|_{q,x_j}\|_{q,x_{-j}} \\ \leq&\|\|T_{r_{j+1}}(T^1_{r_1},\dots T^j_{r_j}T^{j+2}_{\frac{1}{2}}\dots T^d_{\frac{1}{2}}f)|_{x_{-j}}\|_{q,x_j}\|_{q,x_{-j}} \\ =&\|T^{j+1}_{r_{j+1}}T^1_{r_1},\dots T^j_{r_j}T^{j+2}_{\frac{1}{2}}\dots T^d_{\frac{1}{2}}\|_q \\ \approx&\|T^1_{r_1},\dots T^j_{r_{j+1}}T^{j+2}_{\frac{1}{2}}\dots T^d_{\frac{1}{2}}\|_q \\ =&\|T^{(j+1)}f\|_q,\end{aligned}$$

as desired.

Global Hypercontractivity:.

Our proof of global hypercontractivity combines the elegant product proof of O'Donnell and Zhao [308] with q-norm variants of techniques from both [187]'s bound on γ-products and [39]'s proof for the two-sided case. On the analytical side, the key fact is that the Efron-Stein decomposition is an approximate eigenbasis of the noise operator on (q,γ)-products in the following sense.

Lemma 2.1.10 (Approximate Eigenbasis)**.** *Let X be a $(4,\gamma)$-product. For any f and $r \in \mathbb{R}^d$*

$$\|T_r f^{=S} - r_S f^{=S}\|_4 \leq O_{d,r}(\gamma)\|f\|_4.$$

In fact, this is not quite strong enough for our result, as Theorem 2.1.2 requires avoiding any error term scaling in $\|f\|_4$. We handle this by instead proving a finer-grained bound for global functions:

$$\|T_r f^{=S} - r_S f^{=S}\|_4 \leq O_{d,r}(\gamma)\|f\|_2^{1/2} \max_{T \subseteq S, x_T} \{\|f|_{x_T}\|_2^{1/2}\}.$$

With this in mind, the core idea of O'Donnell and Zhao is simply to bound the 4-norm by symmetrizing, then applying standard hypercontractivity. In particular by Theorem 2.1.1 and Lemma 2.1.10:

$$\begin{aligned}
\|f^{\leq i}\|_4^4 &\lesssim \|T_r T_2 f^{\leq i}\|_4^4 && \text{(Symmetrization)} \\
&\lesssim \| \left(\sum_{|S| \leq i} r_S 2^{|S|} f^{=S} \right) \|_4^4 && \text{(4-Approximate Eigenbasis)} \\
&= \mathbb{E}_x \left[\mathbb{E}_r \left[\left(\sum_{|S| \leq i} r_S 2^{|S|} f^{=S}(x) \right)^4 \right] \right].
\end{aligned}$$

Notice that the inner expectation, now over the boolean hypercube, is a degree-i function with Fourier coefficients $2^{|S|} f^{=S}(x)$. Thus we may apply the standard Bonami Lemma and Parseval's Theorem to move from the 4-norm to the 2-norm:

$$\begin{aligned}
&\leq 2^{O(i)} \mathbb{E}_x \left[\mathbb{E}_r \left[\left(\sum_{|S| \leq i} r_S f^{=S}(x) \right)^2 \right]^2 \right] && \text{(Bonami Lemma)} \\
&= \mathbb{E}_x \left[\left(\sum_{|S| \leq i} f^{=S}(x)^2 \right) \left(\sum_{|T| \leq i} f^{=T}(x)^2 \right) \right]. && \text{(Parseval)}
\end{aligned}$$

Notice that if the two inner terms were independent, we'd be done, as each term individually is roughly the 2-norm $\|f^{\leq i}\|_2^2$ by Parseval. Unfortunately the terms are correlated by shared variables within x. To handle this, [308] re-index by pulling out the shared variables as the intersections $I = S \cap T$:

$$\mathbb{E}_x \left[\left(\sum_{|S| \leq i} f^{=S}(x)^2 \right) \left(\sum_{|T| \leq i} f^{=T}(x)^2 \right) \right] \leq \sum_{I \leq i} \mathbb{E}_x \left[\sum_{S \supset I} f^{=S}(x)^2 \sum_{T : T \cap S = I} f^{=T}(x)^2 \right]. \quad (2.1)$$

Since $f^{=S}$ depends only on x_S, and $f^{=T}$ depends only on x_T. On a product space, we can then factor out these variables and bound the above by

$$\sum_{I \leq i} \mathbb{E}_{x_I} \left[\left(\sum_{S \supset I} \mathbb{E}_{x_{S \setminus I}} [f^{=S}(x_S)^2] \right) \left(\sum_{T \supset I} \mathbb{E}_{x_{T \setminus I}} [f^{=T}(x_T)^2] \right) \right]$$

One then factors out the righthand term by its maximum, which can be bounded by $2^{O(i)} \max_{|T| \leq i, x_T} \|f|_{x_T}\|_2^2$, and observes that the remaining term is at most $2^{O(i)} \|f^{\leq i}\|_2^2$, completing the proof.

Unfortunately, on a γ-product, the variables $x_{S \setminus I}$ and $x_{T \setminus I}$ are not independent, so this strategy fails naively. To break their correlation, we'll use a now standard tool in the HDX literature called the *swap walks*:

$$A_{S,T} f(y) = \mathbb{E}[f(x_S) \mid x_T = y_T],$$

which correspond to the (normalized) bipartite adacency matrix of (X_S, X_T). It is by now a widely used fact that the swap walks on HDX expand (see e.g. [109, 9, 187, 12, 114]).

Formally, expanding out Equation (2.1) one encounters many terms of the form

$$\mathop{\mathbb{E}}_{x_{S \setminus I}} \left[f^{=S}(x_S)^2 \mathop{\mathbb{E}}_{x_{T \setminus I} \sim X_{x_S}} \left[f^{=T}(x_T)^2 \right] \right]$$

Similar to the strategy of [39] in the two-sided case, the trick is to observe that we can

re-write the inner expectation as an application of the swap-walk, and de-correlate the terms using spectral expansion:

$$\mathop{\mathbb{E}}_{x_{S\setminus I}}\left[f^{=S}(x_S)^2 \mathop{\mathbb{E}}_{x_{T\setminus I}\sim X_{x_S}}\left[f^{=T}(x_T)^2\right]\right] = \left\langle (f^{=S}|_{x_I})^2, A^{x_I}_{T\setminus I, S\setminus I}(f^{=T}|_{x_I})^2\right\rangle$$
$$\lesssim \mathop{\mathbb{E}}_{x_{S\setminus I}\sim X_{x_I}}[(f^{=S}|_{x_I})^2]\mathop{\mathbb{E}}_{x_{T\setminus I}\sim X_{x_I}}[(f^{=T}|_{x_I})^2].$$

A careful analysis of the error terms reveals the resulting approximate inequality holds up to error
$\gamma\|f\|_2^2 \max_{|S|\le i, x_S}\{\|f|_{x_S}\|_2^2\}$. This is dominated by the main term, so we may proceed as in the product case to complete the proof.

2.1.4 Related Work

Boolean Analysis Beyond the Cube:.

Our work fits into a long line of boolean analysis (and in particular hypercontractive inequalities) beyond the cube (see e.g. [78, 349, 160, 157, 194, 252, 255, 111, 147, 187, 39, 139] among many others). Our results are most closely related to the line of work on product spaces and high dimensional expanders [245, 364, 39, 187]. In particular, we rely heavily on generalizing ideas of Gur, Lifshitz, and Liu [187], who introduced the approximate Efron-Stein Decomposition for γ-products, and proved a similar hypercontractive inequality roughly of the form

$$\|f^{\le i}\|_4^4 \le i^{O(i)}\|f\|_2^2 \max_{|S|\le i, x_S}\{\|f|_{x_S}\|_2^2\} + 2^{O(d)}\gamma\|f\|_2^2\|f\|_\infty^2 \tag{2.2}$$

As discussed in the previous section, Equation (2.2) is restrictive in two senses. First, due to the $\|f\|_\infty$ scaling in the error term, the bound is only meaningful when $\max_{|S|\le i, x_S}\{\|f|_{x_S}\|_2^2\}$ is at least $2^{O(d)}\gamma\|f\|_\infty^2$. This means that on any *fixed* complex, the bound is not meaningful for all ranges of global functions (in particular the error term dominates for strongly global

functions). By leveraging connections with $(4, \gamma)$-products, Theorem 2.1.2 removes this limitation completely.

Second, and perhaps more importantly, the leading $i^{O(i)}$-dependence in Equation (2.2) does not scale appropriately with the eigenvalues of corresponding noise operators. This scaling is often critical in application, where the dependence must cancel the corresponding eigenvalues of relevant operators. Similar improvements for the Grassmann and Bilinear Scheme [138, 143], for instance, recently led to breakthrough progress of PCP theory [297].

On two-sided local-spectral expanders, Bafna, Hopkins, Kaufman, and Lovett [39] prove a version of Equation (2.2) with the correct $2^{O(i)}$ scaling on the main term, but similarly suffering from dependence in $\|f\|_\infty$ in the error term. Because two-sided HDX can be embed into a partite one-sided HDX with a corresponding Fourier decomposition, our result also improves the state of the art for the two-sided case.

Finally, it is worth highlighting that due to its use of symmetrization (and more generally the methods of [308]), the proof of Theorem 2.1.2 is substantially simpler than prior proofs of hypercontractivity for HDX [187, 39]. The proof of hypercontractivity on other extended spaces such as the Grassmann [255, 138, 143] or Lie groups [139], despite recent simplifications to the former, still remains prohibitively complicated to extend to related objects such as higher rank tensors. Our result is the first to extend the simple symmetrization technique beyond products—could such a method also simplify and extend modern 'group-based' hypercontractive inequalities?

Coordinate-Wise Analysis on (near)-Product Spaces:.

Coordinate-wise methods are among the most classical techniques for proving high dimensional analytic inequalities (see [308]). Some of these techniques, such as tensorization of variance and entropy, have already found great success in the study of approximate products and high dimensional expanders. This was first made explicit by Chen, Liu, and

Vigoda [94], though similar ideas were explored earlier by Kaufman and Mass [236]. Chen and Eldan [93] observed that many methods in the approximate sampling literature, such as spectral and entropic independence (variants of high dimensional expansion), can also be viewed from the of standpoint coordinate-wise localization schemes.

These works differ from our method in two main facets. First, all coordinate-wise methods in the literature beyond the study of variance strongly relied on *density* of the underlying complex. Indeed many of these works (see e.g. [96, 217] among others) actually prove full hypercontractive inequalities, which cannot hold in the sparse setting. Second, unlike tensorization methods, we localize more explicitly at the level of the *operator*, splitting T_ρ itself into coordinate operators still acting over the full space X, localizing the operators one-by-one. A similar approach was recently taken in concurrent and independent work[7] of Alev and Parzanchevski [12], who consider coordinate-wise variants of the down-up walk.

2.2 Preliminaries

2.2.1 Simplicial Complexes

A *pure simplicial complex* is a collection of disjoint sets

$$X = X(0) \cup \ldots \cup X(d)$$

where $X(d) \subseteq \binom{[n]}{d}$ is an arbitrary d-uniform hypergraph and $X(i) \subseteq \binom{[n]}{i}$ is given by downward closure, that is the family of i-sets that sit inside any element of $X(d)$. A *weighted* simplicial complex is a pure simplicial complex equipped with a measure Π_d over $X(d)$. This induces a natural measure Π_i over $X(i)$ given by drawing a d-set s, then an i-set $t \subset s$ uniformly at random. For simplicity, we will typically drop the weight function

[7] An informal version of this work more contemporaneous with Alev and Parzanchevski can be found in [199].

from notation and simply write X to mean a pure weighted simplicial complex.

We denote the space of functions over i-faces of X as $C_i \coloneqq \{f : X(i) \to \mathbb{R}\}$. The measures Π_i induce a natural inner product over this space defined by

$$\langle f, g \rangle = \mathop{\mathbb{E}}_{t \sim \Pi_i}[f(t)g(t)]$$

We will usually just write $t \sim X(i)$, where it is understood that Π_i is the underlying distribution.

Links:.

It will frequently be useful to look at the local structure of a given simplicial complex X. Given a face $t \in X$, the *link* of t is the simplicial complex

$$X_t \coloneqq \{s : t \cup s \in X\}$$

whose underlying distribution Π^t is the natural induced weight function generated by sampling $s' \in X(d)$ conditioned on s' containing t, that is $\Pi^t(s) = \mathbb{P}_{s' \sim \Pi_d}[s' = s \cup t \mid t \subset s']$. Equivalently, one may think of the link X_t as being the distribution over X conditioned on t appearing in the face, marginalized to the remaining variables.

Partite Complexes:.

A simplicial complex X is called *partite* if it is possible to partition its vertices $X(0) = \Omega_1 \cup \ldots \cup \Omega_d$ such that every face $s \in X(d)$ has exactly one vertex from each component. We can always think of a partite simplicial complex as a (possibly very sparse) distribution over $\bigotimes_{i=1}^{d} \Omega_i$. With this in mind, we will typically denote elements in $X(d) = \bigotimes_{i=1}^{d} \Omega_i$ as d-dimensional tuples $x = (x_1, \ldots, x_d)$. This allows us to align with more typical notation in the analysis literature over product spaces.

It will frequently be useful to work on the *projection* of $X(d)$ to a certain subset of

colors (coordinates). In particular, given $S \subset [d]$, a face $x_S \in X[S]$ is simply generated by drawing $x \in X(d)$ and projecting onto the S-coordinates of x.

2.2.2 High Order Random Walks

Hypergraphs come equipped with a sequence of natural random walks generalizing the standard random walk on graphs. In the context of high dimensional expanders, such walks were first studied by Kaufman and Mass [234], and have since become an integral part of almost all work on high dimensional expanders.

In this work, we focus mostly on partite complexes. In this setting, there is a natural family of generalizations of the bi-partite graph operator. Given subsets $S, T \subset [d]$, the random walk operator $A_{S,T}$ maps functions on $X[S]$ to functions on $X[T]$ by averaging. In particular, given a function $f : X[S] \to \mathbb{R}$ define

$$A_{S,T}f(y_T) = \mathbb{E}[f(x_S) \mid x_T = y_T].$$

Equivalently, this is the expected value of f when sampled from the link of y_T. Note that in the graph case, this is simply the underlying random walk on a bi-partite graph. We write $\Pi_{S,T}$ to denote the stationary distribution of the operator $A_{S,T}$. Note this is simply the distribution given by drawing $x \in X(d)$ and projecting onto S.

We will frequently make use of the special case where $S = [d]$, which averages f outside the specified coordinate subset T. In particular, given a subset $T \subset [d]$ and function $f : X(k) \to \mathbb{R}$, we define

$$E_T f(y_T) \coloneqq A_{[d],T}f(y_T) = \mathbb{E}[f(x) \mid x_T = y_T].$$

Similarly, $E_T f(y_T)$ is simply expectation of f over the link of y_T.

The Noise Operator:.

The noise operator is one of the best studied random walks in the analysis of boolean functions. The standard operator is typically defined via the following probabilistic interpretation describing its transition matrix as a Markov chain: given a face $x \in X(d)$, fix each coordinate in x with probability ρ, and re-sample all remaining coordinates. Formally, we can write the noise operator as a convex combination of the $\{E_S\}_{S\subseteq[d]}$ operators defined above.

Definition 2.2.1 (Noise Operator)**.** Let X be a d-partite complex, $f \in C_d$, and $\rho \in [0,1]$. The noise operator T_ρ acts on f by re-randomizing over each coordinate with probability $1-\rho$:

$$T_\rho f = \sum_{S\subset[d]} \rho^{|S|}(1-\rho)^{d-|S|} E_S f$$

Finally, we note that when working on a link X_τ, we will write any corresponding walk operator as M^τ to denote the specification to the link. E.g. $A^\tau_{i,j}$ is the bipartite graph between color i and color j within the link of τ, and $\Pi^\tau_{i,j}$ to denote its stationary distribution.

Total Influence:.

Total influence is a critical notion in boolean analysis measuring the total 'sensitivity' of a function to flipping individual coordinates. On product spaces, total influence is most naturally defined via the *Laplacian operators*

$$L_i = I - A_{[d]\setminus\{i\}}.$$

The total influence of a function $f \in C_d$ is

$$\mathbf{I}[f] = \sum_{i\in[d]} \langle f, L_i f\rangle$$

When f is boolean, it is easy to check that the inner product $\langle f, L_i f \rangle$ is proportional to the expected probability over x that re-sampling ith coordinate flips the value of f. The study of the structure of low influence functions is core to boolean analysis, and a major motivation behind this work.

2.2.3 High Dimensional Expanders and γ-Products

We focus in this work on an elegant reformulation of spectral high dimensional expansion of Dikstein and Dinur [109], and Gur, Lifshitz, and Liu [187] called a *γ-product*, which promises every bi-partite operator $A_{i,j}$ (in X and its links) is expanding.

Definition 2.2.2 (γ-Products)**.** A d-partite complex X is called a γ-product if for every link X_τ of co-dimension at least 2 and colors $i, j \notin \tau$, the spectral norm of the walk A^τ_{ij} satisfies:

$$\lambda_2(A^\tau_{ij}) \leq \gamma.$$

A similar notion was considered in the non-partite case by Bafna, Hopkins, Kaufman, and Lovett [39]. γ-products are closely related to the more standard notion of *local-spectral expansion* of Dinur and Kaufman [124] and Oppenheim [309].

Definition 2.2.3 (Local-Spectral Expander)**.** A weighted complex X is a (one-sided) γ-local-spectral expander if the graph underlying every non-trivial link is a (one-sided) γ-spectral expander.

Dikstein and Dinur [109] prove that any γ-one-sided partite local-spectral expander is an $O_d(\gamma)$-product. On the other hand, seminal work of Oppenheim [309] implies any γ-product is roughly an $O(\gamma)$-one-sided local-spectral expander for small enough γ (see [12] for a more detailed conversion between the two).

Gur, Lifshitz, and Liu [187] also observe that any two-sided local-spectral expander can be *embedded* into a partite complex as a γ-product. This is done by embedding of a d-dimensional complex X into $X(1)^d$ by adding every permutation of each top level face.

Observation 2.2.4. If X is a d-dimensional two-sided γ-local-spectral expander, then the partite complex[8]

$$X^d \coloneqq \{\pi(x) : \pi \in S_d, x \in X(d)\}$$

with distribution

$$\mu^d(x) = \Pi_d(x)/d!$$

is a γ-product.

Proof. It is enough to observe that the color walk A_{ij} is exactly the non-lazy upper walk on the graph underlying the original complex X. □

As a result, most results proved for γ-products carry over to two-sided local-spectral expanders. In fact, it turns out the approximate Fourier decomposition proposed by [39], when passed through the above embedding, exactly corresponds to the Efron-Stein components. As a result, all Fourier analytic results on γ-products indeed transfer to two-sided HDX as well.

2.3 Fourier Analysis on (q, γ)-Products

In this section, we introduce a natural generalization of local-spectral expanders to higher moments we call (q, γ)-Products, and extend the useful Fourier analytic machinery of [187, 39] to this setting.

Definition 2.3.1 ((q, γ)-Products)**.** For any $q \in (1, \infty]$, $\gamma > 0$, and $d \in \mathbb{N}$, a d-partite complex X is a (q, γ)-Product if for every link X_τ of co-dimension at least 2 and every $i \neq j$

$$\|A^\tau_{i,j} - \Pi^\tau_{i,j}\|_q \leq \gamma.$$

For a set $Q \subset (1, \infty]$, we say X is a (Q, γ)-product if it is a q-gamma product for all $q \in Q$.

[8]Technically, one should first fix an ordering on $X(1)$ for this to be well-defined, though the resulting complex is independent of choice of ordering.

Since $\|A^{\tau}_{i,j} - \Pi_{i,j}\|_2$ is exactly the second largest eigenvalue of $A^{\tau}_{i,j}$, setting $p = 2$ recovers the notion of a γ-product. One can define the notion of (q, γ)-HDX on non-partite complexes analogously.

Definition 2.3.2 ((q, γ)-HDX)**.** A simplicial complex X is a (q, γ)-HDX if every link X_τ of co-dimension at least 2 satisfies:

$$\|A_\tau - \Pi^{\tau}_1\|_q \leq \gamma$$

Similarly, setting $q = 2$ recovers the notion of a two-sided γ-local-spectral expander. It is easy to see that embedding a (q, γ)-HDX into a partite complex by including every ordering of the faces results in a (q, γ)-product by the same argument as above. All results we cover in the partite case therefore translate to the former, and we focus only on the partite case in what follows.

To relate (q, γ)-products to standard HDX, we will rely on the following special case of the classical Riesz-Thorin Interpolation theorem

Theorem 2.3.3 (Riesz-Thorin Interpolation [330, 353])**.** *Let X be a d-partite complex and $T : C_d \to C_d$ a linear operator. For any $0 < p_0 < p_1 \leq \infty$, $\theta \in (0, 1)$, and $\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$*

$$\|T\|_{p_\theta} \leq \|T\|^{1-\theta}_{p_0} \|T\|^{\theta}_{p_1}.$$

Because the infinity norm of the operators $A - \Pi$ is universally bounded, Theorem 2.3.3 implies any γ-product is a (q, γ_q)-product for γ_q an appropriate function of γ and q:

Lemma 2.3.4. *For any $q > 1$ and $\gamma > 0$, any γ-product is a (q, γ_q)-product for*

$$\gamma_q = \begin{cases} \gamma^{2/q} 2^{1-2/q} & q \geq 2 \\ \gamma^{\frac{2(q-1)}{q}} 2^{1-\frac{2(q-1)}{q}} & 1 < q < 2 \end{cases}$$

Proof. It is enough to prove the $q \geq 2$ case. In particular, since the adjoint of $A_{i,j}$ is $A_{j,i}$, the $1 < q < 2$ case follows from Hölder duality. Namely the fact that for any operator M:

$$\|M\|_q = \|M^*\|_{q'}$$

where $q' = \frac{q}{q-1}$ is q's Hölder conjugate. Assuming now that $q \geq 2$, for all $A^\tau_{i,j}$ we have:

$$\|A^\tau_{i,j} - \Pi^\tau_{i,j}\|_2 \leq \gamma \quad \text{and} \quad \|A^\tau_{i,j} - \Pi^\tau_{i,j}\|_\infty \leq 2,$$

where the former is from definition and the latter is due to stochasticity of $A^\tau_{i,j}$ and $\Pi^\tau_{i,j}$. Riesz-Thorin Interpolation then gives

$$\|A^\tau_{i,j} - \Pi^\tau\|_q \leq \|A^\tau_{i,j} - \Pi^\tau\|_2^{2/q} \|A^\tau_{i,j} - \Pi^\tau\|_\infty^{1-2/q} \leq \gamma^{2/q} 2^{1-2/q}$$

as desired. □

Lemma 2.3.4 allows us to easily control the higher norm behavior of classical constructions of high dimensional expanders such as the Ramanujan complexes [290] and coset complexes [238] over sufficiently large fields.

2.3.1 q-Norm Expansion of Swap Walks

A critical property of classical high dimensional expanders, and the core of [187]'s Fourier analysis on HDX, is that the operators $A_{S,T}$ are strongly expanding when $S \cap T = \emptyset$ (these specific operators are often called the 'swap-walks') [9, 109, 187, 12]. Prior arguments for expansion of the swap walks take specific advantage of being in the spectral ($q = 2$) setting. We give an elementary argument showing swap walks expand in q-norm on (q, γ)-products.

Lemma 2.3.5. *Let X be a (q,γ)-product and $q' = \frac{q}{q-1}$. For all coordinate subsets $S \cap T = \emptyset$:*

$$\|A_{S,T} - \Pi_{S,T}\|_q \leq |S||T|\gamma \quad \textit{and} \quad \|A_{S,T} - \Pi_{S,T}\|_{q'} \leq |S||T|\gamma$$

Proof. We induct on $|S| + |T|$. For the base case $|S| = |T| = 1$, observe we have $\|A_{S,T} - \Pi_{S,T}\|_q \leq \gamma$ by definition. To bound the q'-norm, recall Hölder duality promises that for any operator M

$$\|M\|_q = \|M^*\|_{q'},$$

where M^* is the adjoint. Thus we also have

$$\|A_{S,T} - \Pi_{S,T}\|_{q'} = \|A_{T,S} - \Pi_{T,S}\|_q \leq \gamma$$

as desired. Note by the same argument, this also holds within every link of X.

Assume now by induction that, for some fixed $|S| + |T| > 2$, all S', T' such that $|S'| + |T'| < |S| + |T|$ and $\tau \in X$ satisfy

$$\|A^{\tau}_{S',T'} - \Pi^{\tau}_{S',T'}\|_q \leq |S'||T'|\lambda.$$

We first argue we may assume without loss of generality that $|T| > 1$. This is again by Hölder duality since when $|S| > |T|$, we have

$$\|A_{S,T} - \Pi_{S,T}\|_q = \|A_{T,S} - \Pi_{T,S}\|_{q'} \leq \gamma.$$

Thus we are done if we show the result for all $q \in [1,\infty]$ and $|T| \geq |S|$ (in particular when $|T| > 1$).

Assume without loss of generality that $1 \in T$. Fix any $f : X[S] \to \mathbb{R}$. As is typically the case, the idea is to sample $t \in X[T]$ by first sampling $t' \in X[T \setminus \{1\}]$, then

$v \in X_t[1]$:

$$\|A_{S,T}f - \Pi_{S,T}f\|_q = \left\| \|(A_{S,T}f)|_{t'} - (\Pi_{S,T}f)|_{t'}\|_{q,v\in X_t[1]} \right\|_{q,t'\in X[T\setminus\{1\}]}$$

where for any $g : X[T] \to \mathbb{R}$, $g|_{t'}(v) = g(t \cup v)$ denotes the localization of g to the link of t'. The trick is now to observe that as a function of $X_{t'}[1]$, $A_{S,T}f|_{t'}$ is exactly $A^{t'}_{S,1}f^{t'}$, where $f^{t'} : X_{t'}[S] \to \mathbb{R}$ is the restriction $f^s(\tau) = f(\tau)$. Then by adding and subtracting the corresponding local stationary operator:

$$\begin{aligned}
\|S_{S,T}f - \Pi_{S,T}f\|_q &= \left\| \|A^{t'}_{S,1}f^{t'} - \Pi^{t'}_{S,1}f^{t'} + \Pi^{t'}_{S,1}f^{t'} - (\Pi_{S,T}f)|_{t'}\|_{q,v\in X_{t'}(1)} \right\|_{q,t'\in X[T\setminus\{1\}]} \\
&\leq \left\| \|A^{t'}_{S,1}f^{t'} - \Pi^{t'}_{S,1}f^{t'}\|_{q,v\in X_{t'}[1]} \right\|_{q,s\in X[T\setminus\{1\}]} \\
&+ \left\| \|\Pi^{t'}_{S,1}f^{t'} - (\Pi_{S,T}f)|_{t'}\|_{q,v\in X_{t'}[1]} \right\|_{q,t'\in X[T\setminus\{1\}]}
\end{aligned}$$

by the triangle inequality. The first term is now bounded by the inductive hypothesis applied in the link of t':

$$\begin{aligned}
\left\| \|A^{t'}_{S,1}f^{t'} - \Pi^{t'}_{S,1}f^{t'}\|_{q,v\in X_{t'}[1]} \right\|_{q,t'\in X[T\setminus\{1\}]} &\leq i\gamma \left\| \|f^{t'}\|_{q,X_{t'}[S]} \right\|_{q,t'\in X[T\setminus\{1\}]} \\
&= \mathop{\mathbb{E}}_{t'\in X[T\setminus 1]} \left[\mathop{\mathbb{E}}_{s\sim X_{t'}[S]} [|f(s)|^q] \right]^{1/q} \\
&= |S|\gamma\|f\|_q
\end{aligned}$$

where in the final step we have used the fact that drawing $t' \in X[T \setminus 1]$, then s from $X_{t'}[S]$ is equidistributed with simply drawing $s \in X[S]$ directly.

Toward the second term, observe that, as a function of t':

$$\Pi^{t'}_{S,1}f^{t'}(v) = A_{S,T'}f(t') \quad \text{and} \quad (\Pi_{S,T}f)|_{t'}(v) = \Pi_{S,T'}f(t')$$

so we finally have

$$\left\| \|\Pi_{S,1}^{t'} f^{t'} - (\Pi_{S,T} f)|_{t'}\|_{q,v\in X_{t'}[1]} \right\|_{q,t'\in X[T'\setminus\{1\}])} = \|S_{i-1,j} f - \Pi_{i-1,j} f\|_{q,t'\in X[T\setminus\{1\}]}$$
$$\leq (|T|-1)|S|\gamma\|f\|_q$$

by the inductive hypothesis. Altogether this gives

$$\|S_{S,T} f - \Pi_{S,T} f\|_q \leq \gamma|S| + \gamma(|T|-1)|S| = |S||T|\gamma$$

as desired. □

2.3.2 The Efron-Stein Decomposition

Most of our analysis of (q,γ)-products is based on the Efron-Stein decomposition, introduced in the context of HDX independently in [187] (for γ-products) and [39] (for two-sided HDX).

Definition 2.3.6 (Efron-Stein Decomposition)**.** Let X be a d-partite complex, $0 \leq k \leq d$, and $f \in C_k$. The Efron-Stein Decomposition of f is given by the collection of functions $\{f^{=S}\}_{S\subseteq[d]}$ where:

$$f^{=S} \coloneqq \sum_{T\subset S} (-1)^{|S\setminus T|} E_T f$$

We write $f^{\leq i} = \sum_{|S|\leq i} f^{=S}$ to denote the degree at most i components of f, and $f^{=i}$ degree exactly i.

Note that $f^{=S}$ is only a function of the S-projection of x. With this in mind, we sometimes write $f^{=S}(x_S)$ instead of $f^{=S}(x)$.

In the remainder of the section, we argue that essentially any useful property of the Efron-Stein decomposition continues to hold approximately on (q,γ)-products. We

start with the core result of [187], who show the Efron-Stein decomposition on γ-products is an approximate Fourier basis.

Theorem 2.3.7 ([187])**.** *Let X be a d-partite γ-product and $f : X \to \mathbb{R}$. The Efron-Stein Decomposition satisfies the following properties:*

1. *Decomposition:* $f = \sum_{S \subset [d]} f^{=S}$
2. *Approximate Orthogonality:* $\langle f^{=S}, f^{=T} \rangle \leq 2^{O(|S|+|T|)} \gamma \|f\|_2^2$
3. *Approximate Parseval:* $\left| \langle f^{\leq i}, f^{\leq i} \rangle - \sum_{|S| \leq i} \langle f^{=S}, f^{=S} \rangle \right| \leq 2^{O(d)} \gamma \|f\|_2^2$

We will require variants of a few other standard properties of Efron-Stein extended to high norms. First, we observe the q-norm of $f^{=S}$ is not much larger than the q-norm of f itself.

Lemma 2.3.8. *Let X be a d-partite complex. Then for any $f \in C_d$, $S \subseteq [d]$, and $q \geq 1$*

$$\|f^{=S}\|_q \leq 2^{|S|} \|f\|_q$$

When $q = 4$, we can further bound

$$\|f^{=S}\|_4 \leq \|f\|_2^{1/2} \max_{T \subseteq S, x_T} \left\{ \|f|_{x_T}\|_2^{1/2} \right\}$$

Proof. E_T is an averaging operator and therefore contracts q-norms. As such we have:

$$\begin{aligned} \|f^{=S}\|_q &= \|\sum_{T \subset S} (-1)^{|S|-|T|} E_T f\|_q \\ &\leq 2^{|S|} \|f\|_q \end{aligned}$$

by the triangle inequality. For $q = 4$, we can instead bound

$$\begin{aligned}
\|\sum_{T\subset S}(-1)^{|S|-|T|}E_T f\|_q &= \mathbb{E}\left[\left(\sum_{T\subset S}(-1)^{|S|-|T|}E_T f\right)^4\right]^{1/4} \\
&\leq 2^{O(|S|)}\mathbb{E}\left[\left(\sum_{T\subset S}(-1)^{|S|-|T|}(E_T f)^2\right)^2\right]^{1/4} \\
&\leq 2^{O(|S|)}\max_{T\subseteq S, x_T}\left\{(E_T f)^2\right\}\mathbb{E}\left[\sum_{T\subset S}(-1)^{|S|-|T|}(E_T f)^{1/2}\right]^{1/4} \\
&\leq 2^{O(|S|)}\max_{T\subseteq S, x_T}\left\{\|f|_{x_T}\|_2^{1/2}\right\}\|f\|_2^{1/2}
\end{aligned}$$

where we've used the fact that $\mathbb{E}[f|_{x'_T}]^2 \leq \|f|_{x_{T'}}\|_2^2$ and E_T contracts all p-norms by Jensen's inequality. □

Second, we will crucially rely on the fact that $E_T f^{=S}$ is small so long as T does not contain S.

Lemma 2.3.9. *Let X be a d-partite (q,γ)-product and $f \in C_d$. Then for any $T, S \subset [d]$ such that $T \not\supset S$:*

$$\|E_T f^{=S}\|_q \leq |T|2^{O(|S|)}\gamma\|f\|_q$$

For $q = 4$, we can further bound

$$\|E_T f^{=S}\|_q \leq |T|2^{O(|S|)}\gamma\|f\|_2^{1/2}\max_{T'\subseteq S, x_{T'}}\left\{\|f|_{x'_T}\|_2^{1/2}\right\}$$

Proof. The proof is similar to the $q = 2$ case given in [187], but with somewhat more care required for the latter bound. The key is to prove the following claim

Claim 2.3.10. For any $f : X(d) \to \mathbb{R}$ and $q \geq 2$:

$$\|E_T E_{T'} - E_{T\cap T'}f\|_q \leq \gamma|T||T'|\min\left\{\|f\|_q, \|f\|_{q-2}^{1-2/q}\max_{T''\subset T}\left\{\|f|_{x''_T}\|_2^{2/q}\right\}\right\}$$

Proof. It is enough to consider the setting where $T \cap T' = \emptyset$. In particular, assume both results hold in this case and let $I = T \cap T'$. We localize within I and apply the claim:

$$\begin{aligned}
&\|E_T E_{T'} f - E_{T\cap T'} f\| \\
=&\|\|(E^{x_I}_{T\setminus I} E^{x_I}_{T'\setminus I} - E^{x_I}_{\emptyset}) f|_{x_I}\|_{q,x_{\bar{I}}}\|_{x_I} \\
\leq&\gamma|T||T'|\|\min\left\{\|f|_{x_I}\|_{q,x_{\bar{I}}}, \|f|_{x_I}\|_{q-2}^{1-2/q} \max_{T''\subset T'\setminus I}\left\{\|f|_{x_{T''\cup I}}\|_2^{2/q}\right\}\right\}\|_{q,x_I} \\
=&\gamma|T||T'|\min\left\{\|f\|_q, \|f\|_{q-2}^{1-2/q} \max_{T''\subset T}\left\{\|f|_{x''_T}\|_2^{2/q}\right\}\right\}.
\end{aligned}$$

We therefore turn our attention to the case $T \cap T' = \emptyset$ and show

$$\|(E_T E_{T'} - E_\emptyset) f\|_q \leq |T||T'|\gamma \min\left\{\|f\|_q, \|f\|_{q-2}^{1-2/q} \max_{T''\subset T}\left\{\|f|_{x''_T}\|_2^{2/q}\right\}\right\}.$$

We appeal to the argument of [187], who observe that one may write $E_T E_{T'} = A_{T',T} E_{T'}$ and $\mathbb{E}[E_{T'} f] = E_\emptyset[f]$, so

$$\begin{aligned}
\|E_T E_{T'} f - E_\emptyset f\|_q &= \|(E_{T',T} - \Pi_{T',T}) E_{T'} f]\|_q \\
&\leq |T'||T|\gamma\|E_{T'} f\|_q
\end{aligned}$$

Since averaging contracts q-norms by Jensen's inequality, this is at most $|T'||T|\gamma\|E_{T'} f\|_q$. Alternatively, similar to the prior lemma we can pull out a factor of the restricted 2-norm before contracting

$$\begin{aligned}
\|E_{T'} f\|_q &= \mathbb{E}_x\left[\mathop{\mathbb{E}}_{y_{\bar{T'}}\sim x_{T'}}[f|_{x_{T'}}]^q\right]^{1/q} \\
&\leq \max_{x_{T'}}\left\{\mathbb{E}[f|_{x_{T'}}]_2^{2/q}\right\} \mathbb{E}_x[\mathbb{E}_{y_{\bar{T'}}\sim x_{T'}}[f|_{x_{T'}}]^{q-2}]^{1/q} \\
&\leq \max_{x_{T'}}\left\{\|f|_{x_{T'}}\|_2^{2/q}\right\} \|f\|_{q-2}^{1-2/q}
\end{aligned}$$

since $\mathbb{E}[f|_{x'_T}]^2 \leq \|f|_{x_{T'}}\|_2^2$ and $E_{T'}$ contracts all p-norms by Jensen's inequality. $\square$

The result now follows simply by expanding $f^{=S}$:

$$\begin{aligned}
\|E_T f^{=S}\|_q &= \|E_T \sum_{T' \subset S} (-1)^{|S|-|T'|} E_{T'} f\|_q \\
&= \|\sum_{T' \subset S} (-1)^{|S|-|T'|} E_T E_{T'} f\|_q \\
&= \|\sum_{T' \subset S} (-1)^{|S|-|T'|} E_{T \cap T'} f + \sum_{T' \subset S} (-1)^{|S|-|T'|} (E_T E_{T'} - E_{T \cap T'}) f\|_q \\
&\leq 2^{O(|S|)} \|(E_T E_{T'} - E_{T \cap T'}) f\|_q \\
&\leq 2^{O(|S|)} |T| \gamma \min \left\{ \|f\|_q, \|f\|_{q-2}^{1-2/q} \max_{T' \subset S} \left\{ \|f|_{x'_T}\|_2^{2/q} \right\} \right\}
\end{aligned}$$

where we have used the standard fact that $\sum_{T' \subset S} (-1)^{|S|-|T'|} E_{T \cap T'}$ is identically zero when $T \not\supset S$. $\square$

A critical property of the standard Efron-Stein decomposition on product spaces is that it is an eigenbasis for standard random walks such as the noise operator. Using the above lemma, it is elementary to show this extends approximately to (q, γ)-products.

Lemma 2.3.11. *Let X be a d-partite (q, γ)-product and $M = \sum_{T \subset [d]} \alpha_T E_T$ for some $\alpha_T \in \mathbb{R}$. Then for $f \in C_d$ and $S \subset [d]$:*

$$\|M f^{=S} - \lambda_S f^{=S}\|_q \leq \|\alpha\|_1 2^{O(d)} \gamma \|f\|_q$$

where $\lambda_S = \sum_{T \supseteq S} \alpha_T$. For $q = 4$ we further have

$$\|M f^{=S} - \lambda_S f^{=S}\|_q \leq \|\alpha\|_1 2^{O(d)} \gamma \|f\|_2^{1/2} \max_{T \subseteq S, x_T} \left\{ \|f|_{x_T}\|_2^{1/2} \right\}$$

Proof. expanding $Mf^{=S}$ we have

$$\begin{aligned} Mf^{=S} &= \sum_{T\subset[d]} \alpha_T E_T f^{=S} \\ &= \sum_{T\supseteq S} \alpha_T f^{=S} + \sum_{T\not\supseteq S} \alpha_T E_T f^{=S} \end{aligned}$$

where the second equality holds since $f^{=S}$ deqends only on coordinates in S, and E_T fixes such coordinates when $T \supseteq S$. Applying Lemma 2.3.9 and the triangle inequality then comqletes the proof, as the remaining terms are all small in q-norm. □

A similar application also shows that the Efron-Stein basis is 'approximately closed'.

Lemma 2.3.12. *Let X be a d-partite (q,γ)-product. Then for any $f \in C_d$ and $T, S \subseteq [d]$*

- *For $S \neq T$:*

$$\|(f^{=S})^{=T}\|_q \leq 2^{O(d)}\gamma\|f\|_q$$

- *For $S = T$:*

$$\|(f^{=S})^{=S} - f^{=S}\|_q \leq 2^{O(d)}\gamma\|f\|_q$$

When $q = 4$, we may replace $\|f\|_q$ by $\|f\|_2^{1/2} \max_{T\subseteq S, x_T} \left\{\|f|_{x_T}\|_2^{1/2}\right\}$.

Proof. Using the definition of Efron-Stein, we have

$$(f^{=S})^{=T} = \sum_{T'\subseteq T} (-1)^{T\setminus T'} E_{T'} f^{=S}$$

If $S \neq T$, then $T' \not\supset S$ for every term in the sum, and we are done by the triangle inequality and Lemma 2.3.9. If $S = T$, the only 'surviving' term is $E_S f^{=S} = f^{=S}$, and we are similarly done. □

Finally, we will also need the following behavior of the Efron-Stein decomposition under restriction.

Lemma 2.3.13. *Let X be a d-partite complex and $f \in C_d$. Let $I, B \subset [d]$ be any two disjoint sets. Then*

$$f^{=I\cup B}(y_I, x_B, z) = \sum_{J\subset I}(-1)^{|I|-|J|}(f|_{y_J})^{=B}(x_B, z).$$

Proof. This is a standard fact on product spaces (see, e.g., [364]). We give a proof that generalizes to any partite complex. Expanding the righthand side:

$$\begin{aligned}
\sum_{J\subset I}(-1)^{|I|-|J|}(f|_{y_J}^{=B})(x_B, z) &= \sum_{J\subset I}(-1)^{|I|-|J|}\sum_{A\subset B}(-1)^{|B|-|A|}E_A^{y_J}f|_{y_J}(x_B, z)\\
&= \sum_{J\subset I}(-1)^{|I|-|J|}\sum_{A\subset B}(-1)^{|B|-|A|}\mathop{\mathbb{E}}_{z'\sim X_{y_J\cup x_A}}[f(y_J, x_A, z')]\\
&= \sum_{J\subset I}(-1)^{|I|-|J|}\sum_{A\subset B}(-1)^{|B|-|A|}E_{J\cup A}f(y_I, x_B, z)\\
&= \sum_{T\subset I\cup B}(-1)^{|I\cup B|-|T|}E_T f(y_I, x_B, z)\\
&= f^{=I\cup B}(y_I, x_B, z)
\end{aligned}$$

as desired. □

2.3.3 Efron-Stein, Total Influence, and the Noise Operator

While the individual Efron-Stein components are only approximate eigenvectors, it turns out the full decomposition behaves exactly as in the product case with respect to the standard operators [187]. We will make use this fact for the generalized noise operator that both allows varying parameters by coordinate, and may take 'noise' values outside of $[0, 1]$. This no longer has a natural interpretation as a noisy random walk, but remains very useful in application.

Definition 2.3.14 (Vector-Valued Noise Operator)**.** Let X be a d-partite complex, $f \in C_d$, and $r \in \mathbb{R}^d$. The noise operator T_r is defined as:

$$T_r f = \sum_{S \subset [d]} \prod_{i \in S} r_i \prod_{i \notin S} (1 - r_i) E_S f$$

The following claim is reasonably standard, but we include it for completeness.

Claim 2.3.15. Let X be a d-partite complex, $f \in C_d$, and $r \in \mathbb{R}^n$. Then

$$T_r f = \sum_{S \subset [d]} r_S f^{=S},$$

where $r_S = \prod_{i \in S} r_i$

Proof. By inclusion-exclusion, we have

$$E_S f = \sum_{T \subseteq S} f^{=T}.$$

With this in hand, expanding $T_r f$ we have:

$$\begin{aligned}
T_r f &= \sum_{S \subset [d]} \prod_{i \in S} r_i \prod_{i \notin S} (1 - r_i) E_S f \\
&= \sum_{S \subset [d]} \prod_{i \in S} r_i \prod_{i \notin S} (1 - r_i) \sum_{T \subseteq S} f^{=T} \\
&= \sum_{T \subset [d]} f^{=T} \left(\sum_{S \supset T} \prod_{i \in S} r_i \prod_{i \notin S} (1 - r_i) \right) && \text{(Re-indexing)} \\
&= \sum_{T \subset [d]} r_T f^{=T} \left(\sum_{S \supset T} \prod_{i \in S \setminus T} r_i \prod_{i \notin S} (1 - r_i) \right) \\
&= \sum_{T \subset [d]} r_T f^{=T}
\end{aligned}$$

where in the final step we have observed that

$$\sum_{S \supset T} \prod_{i \in S \setminus T} r_i \prod_{i \notin S} (1 - r_i) = \prod_{i \in [d] \setminus T} (1 - r_i + r_i) = 1.$$

□

One can also view this as an equivalent method of defining the operator T_r. We will move freely between these two equivalent notions in what follows. We can similarly express the Laplacians in terms of Efron-Stein. The proofs are standard (see e.g. [187, Lemma 6.2]) and similar to the above so we omit them.

Lemma 2.3.16. *Let X be a d-partite complex and $f \in C_d$. Then:*

1. $L_i f = \sum_{S \ni i} f^{=S}$
2. $\sum_{i \in [d]} L_i f = \sum_{S \subseteq [d]} |S| f^{=S}$

This implies the following useful relation between total influence and the Efron-Stein decomposition.

Corollary 2.3.17. *Let X be a d-partite complex and $f \in C_d$. Then*

$$\mathbf{I}[f] = \sum_{i=0}^{d} i \langle f, f^{=i} \rangle$$

2.4 Coordinate-Wise Analysis on HDX

The cornerstone of our analysis is a new, but completely elementary method of coordinate-wise analysis on high dimensional expanders inspired by the proof of Bourgain's symmetrization theorem. We break the argument into two main parts, a *de-correlation* step where we break the noise operator into coordinate-wise components, and a *localization* step where we show the coordinate-wise noise operators may be viewed as the standard

noise operator on the projection/localization to the relevant coordinates. We will see in the next section that these lemmas facilitate classic analytic tools such as the replacement method.

We first define a coordinate-wise version of the noise operator that operates only over a specific subset.

Definition 2.4.1 (Coordinate-Wise Noise Operator)**.** Let X be a d-partite complex, $S \subseteq [d]$, and $r \in \mathbb{R}^S$. The coordinate-wise noise operator T_r^S acts on f by:

$$T_r^S f \coloneqq \sum_{T \subseteq S} r_{S \setminus T} \prod_{i \in T} (1 - r_i) E_{[d] \setminus T} f$$

Note that we have inverted the sum in the sense that index T corresponds to $E_{[d]\setminus T}$. When $r \in [0,1]^d$, the above should still be thought of as re-sampling each coordinate within S with probability $1 - r_i$.

Our first key lemma shows that T_r can be approximately decomposed into its constituent coordinate operators (under any ordering) up to some error in q-norm. For simplicity of notation, given a permutation $\pi \in S_d$ and vector $r \in \mathbb{R}^d$ we write

$$T_r^\pi \coloneqq T_{r_{\pi(1)}}^{\pi(1)} \dots T_{r_{\pi(d)}}^{\pi(d)}.$$

Lemma 2.4.2 (Decorrelation)**.** *Let X be a d-partite (q,γ)-product, $r \in \mathbb{R}^d$, and $\pi \in S_d$ any permutation, then:*

$$\|T_r f - T_r^\pi f\|_q \leq c_{d,r} \gamma \|f\|_q$$

where $c_{d,r} = d^3 \sum_S \left| r_S \prod_{i \notin S} (1 - r_i) \right|$

Proof. The proof is essentially an immediate application of Claim 2.3.10. Assume $\pi = I$ without loss of generality (the analysis is invariant under ordering). Expanding out the

coordinate-wise product we have

$$\begin{aligned} T^1_{r_1} \dots T^d_{r_d} &= \prod_{i \in [d]} (r_i I + (1 - r_i) E_{[d]\setminus i}) \\ &= \sum_S \left(r_S \prod_{i \notin S} (1 - r_i) \right) \prod_{i \notin S} E_{[d]\setminus i} \end{aligned}$$

It is therefore enough to bound the q-norm

$$\| E_S - \prod_{i \notin S} E_{[d]\setminus i} \|_q \leq d^3 \gamma.$$

This follows from iterated application of Claim 2.3.10. More formally, assume without loss of generality that $S = \{1, \dots, j\}$. Then we can write the following telescopic sum:

$$\begin{aligned} E_S - \prod_{i \notin S} E_{[d]\setminus i} &= \sum_{\ell=1}^{d-j} \left(\left(\prod_{i=j+1}^{d-\ell} E_{[d]\setminus i} \right) E_{[d-\ell]} - \left(\prod_{i=j+1}^{d-\ell+1} E_{[d]\setminus i} \right) E_{[d-\ell+1]} \right) \\ &= \sum_{\ell=1}^{d-j} \left(\left(\prod_{i=j+1}^{d-\ell} E_{[d]\setminus i} \right) E_{[d-\ell]} - \left(\prod_{i=j+1}^{d-\ell} E_{[d]\setminus i} \right) E_{[d]\setminus\{d-\ell+1\}} E_{[d-\ell+1]} \right) \\ &= \sum_{\ell=1}^{d-j} \left(\prod_{i=j+1}^{d-\ell} E_{[d]\setminus i} \right) \left(E_{[d-\ell]} - E_{[d]\setminus\{d-\ell+1\}} E_{[d-\ell+1]} \right) \end{aligned}$$

By Claim 2.3.10 and the triangle inequality, we finally have:

$$\begin{aligned} \| E_S - \prod_{i \notin S} E_{[d]\setminus i} \|_q &\leq \sum_{\ell=1}^{d-j} \| \left(\prod_{i=j+1}^{d-\ell} E_{[d]\setminus i} \right) \left(E_{[d-\ell]} - E_{[d]\setminus\{d-\ell+1\}} E_{[d-\ell+1]} \right) \|_q \\ &\leq \sum_{\ell=1}^{d-j} \| E_{[d-\ell]} - E_{[d]\setminus\{d-\ell+1\}} E_{[d-\ell+1]} \|_q \\ &\leq d^3 \gamma \end{aligned}$$

where we have additionally taken advantage of the fact that averaging operators contract q-norm. □

The second critical lemma is the 'localization' process, where we argue that the decomposed operators can equivalently be viewed as local noise operators on the links of the complex.

Lemma 2.4.3 (Localization)**.** *Let X be a d-partite simplicial complex, and $r \in \mathbb{R}^d$. Then for any $S \subseteq [d]$:*

$$T_r^S f(x) = T_r^{x_{\bar{S}}} f|_{x_{\bar{S}}}(x_S)$$

Proof. The proof is essentially immediate from the fact that the partite averaging operators 'respect' restriction, namely for any $T \subseteq S$ we have by definition:

$$E_{[d]\setminus T} f(x) = E_{S\setminus T}^{x_{\bar{S}}} f|_{x_{\bar{S}}}(x_S).$$

Then expanding out $T_r^S f$ gives:

$$\begin{aligned} T_r^S f(x) &= \sum_{T\subseteq S} r_{T\setminus S} \prod_{i\in T}(1-r_i) E_{[d]\setminus T} f(x) \\ &= \sum_{T\subseteq S} r_{T\setminus S} \prod_{i\in T}(1-r_i) E_{S\setminus T}^{x_{\bar{S}}} f|_{x_{\bar{S}}}(x_S) \\ &= T_r^{x_{\bar{S}}} f|_{x_{\bar{S}}}(x_S) \end{aligned}$$

as desired. □

2.5 The Symmetrization Theorem

We now turn our attention to our core technical result, Bourgain's symmetrization theorem for HDX.

Theorem 2.5.1 (Symmetrization on HDX)**.** *Let $q > 1$ and X be a d-partite (q,γ)-product for $\gamma \leq 2^{-\Omega(d)}$. Then for any $f \in C_d$:*

$$(1 - 2^{O(d)}\gamma)\|\widetilde{T_{c_q} f}\|_q \leq \|f\|_q \leq (1 + 2^{O(d)}\gamma)\|\widetilde{T_2 f}\|_q$$

for some constant $0 \leq c_q \leq 1$ *dependent only on* q.

Combined with Lemma 2.3.4, we immediately get symmetrization for standard γ-products and HDX.

Corollary 2.5.2. *Let* $q \geq 1$ *and* X *be a* γ*-product with* $\gamma \leq 2^{-\Omega(\max\{q,q'\}d)}$. *For any* $f \in C_d$*:*

$$\left(1 - 2^{O(d)}\gamma^{\frac{2}{\max\{q,q'\}}}\right) \|\widetilde{T_{c_q} f}\|_q \leq \|f\|_q \leq \left(1 + 2^{O(d)}\gamma^{\frac{2}{\max\{q,q'\}}}\right) \|\widetilde{T_2 f}\|_q$$

for q' *the Hölder conjugate of* q. *For* $q \in \{4, 4/3\}$, *one may take* $c_q = 2/5$.

The proof closely follows the ideas of Bourgain, as presented by O'Donnell [308]. Based on the machinery developed in the previous section, the idea is to decompose T_ρ into coordinate-wise operators and handle each coordinate as a single-variate problem, replacing each copy of T^i_ρ with T^i_r for $r \in \{-1, 1\}$. To this end, we first need the single-variate symmetrization theorem.

Lemma 2.5.3 (Symmetrization for random variables (10.14,10.15 in [308]))**.** *Let* X *be a* 0*-mean, real-valued random variable satisfying* $\|X\|_q \leq \infty$. *Then for any* $a \in \mathbb{R}$, *we have:*

$$\|a + \frac{1}{2}X\|_q \leq \|a + rX\|_q$$

where $r \sim \{-1, 1\}$ *is a uniformly distributed random bit.*

With this in hand, we can prove a single-coordinate variant of the result by restricting our function to the relevant variable and applying the above.

Lemma 2.5.4. *Let* X *be a* d*-partite complex and* $f \in C_d$. *For any* $i \in [d]$*:*

$$\|T^i_{1/2} f\|_q \leq \|T^i_r f\|_q$$

where $r \sim \{-1,1\}$ is a uniformly distributed random bit.

Proof.

$$\begin{aligned}
\|T^i_{1/2} f\|_q &= \|\|T^{x_{-i}}_{1/2} f|_{x_{-i}}\|_{q,x_i}\|_{q,x_{-i}} && \text{(Lemma 2.4.3)} \\
&= \|\|\left(f|_{x_{-i}}\right)^{\{\emptyset\}} + \frac{1}{2}\left(f|_{x_{-i}}\right)^{=\{i\}}\|_{q,x_i,r}\|_{q,x_{-i}} \\
&\leq \|\|\left(f|_{x_{-i}}\right)^{\{\emptyset\}} + r\left(f|_{x_{-i}}\right)^{=\{i\}}\|_{q,x_i,r}\|_{q,x_{-i}} && \text{(Lemma 2.5.3)} \\
&= \|\|T_r f|_{x_{-i}}\|_{q,x_i}\|_{q,x_{-i},r} \\
&= \|T^i_r f\|_q && \text{(Lemma 2.4.3)}
\end{aligned}$$

□

We are now ready to prove the upper bound of Theorem 2.5.1 via the replacement method.

Proof of Theorem 2.5.1 (Upper Bound). We first argue it is sufficient to show

$$\|T_{1/2} f\|_q \leq \|T_r f\|_q + 2^{O(d)}\gamma\|f\|_q \tag{2.3}$$

In particular, for any arbitrary function g let $f = T_2 g$. Then applying the above and Claim 2.3.15 we have

$$\begin{aligned}
\|g\|_q &\leq \|T_r T_2 g\|_q + 2^{O(d)}\gamma\|T_2 g\|_q \\
&\leq \|T_r T_2 g\|_q + 2^{O(d)}\gamma\|\sum_{S\subseteq[d]} 2^{|S|} g^{=S}\|_q \\
&\leq \|T_r T_2 g\|_q + 2^{O(d)}\gamma\|g\|_q && \text{(Lemma 2.3.8)}
\end{aligned}$$

Re-arranging and using the fact that $\frac{1}{1-2^{O(d)}\gamma} \leq 1 + 2^{O(d)}\gamma$ for small enough γ gives the desired bound.

We prove Equation (2.3) by the replacement method. In particular, we first start with de-correlating the noise operator into its coordinate-wise components by Lemma 2.4.2:

1. $\|T_{1/2}f - T^1_{1/2} \dots T^d_{1/2} f\|_q \leq 2^{O(d)}\gamma\|f\|_q$.
2. $\|T_r f - T^1_{r_1} \dots T^d_{r_d} f\|_q \leq 2^{O(d)}\gamma\|f\|_q$.

Thus it is sufficient to prove

$$\|T^1_{1/2} \dots T^d_{1/2} f\|_q \leq \|T^1_{r_1} \dots T^d_{r_d} f\|_q + 2^{O(d)}\gamma\|f\|_q.$$

We now iterate through coordinates one by one replacing $T_{1/2}$ with T_{r_i}. Toward this end, define T_j to be the partially replaced operator at step j:

$$T_j \coloneqq T^1_{r_1} \dots T^j_{r_j} T^{j+1}_{1/2} \dots T^d_{1/2}.$$

We'll prove for all $j \in \{0, \dots, d\}$:

$$\|T_j f\|_q \leq \|T_{j+1} f\|_q + 2^{O(d)}\gamma\|f\|_q. \tag{2.4}$$

Then we clearly have

$$\begin{aligned}
\|T^1_{1/2} \dots T^d_{1/2} f\|_q &= \|T_0 f\|_q \\
&\leq \|T_d f\|_q + 2^{O(d)}\gamma\|f\|_q \\
&= \|T^1_{r_1} \dots T^d_{r_d} f\|_q + 2^{O(d)}\gamma\|f\|_q
\end{aligned}$$

as desired.

It remains to prove Equation (2.4), which now follows exactly as in the standard proof up to the accumulation of error. In particular, we may simply permute the $(j+1)$st operator to the front using Lemma 2.4.2, apply our single-coordinate symmetrization theo-

rem to the remaining function, and permute back. Formally, observe that by Lemma 2.4.2 any two orderings π and σ and $r \in [-1,1]^d$ satisfy:

$$\|T_r^\pi f - T_r^\sigma f\|_q = \|(T_r^\pi f - T_r f) + (T_r f - T_r^\sigma f)\|_q \leq 2^{O(d)}\gamma\|f\|_q \tag{2.5}$$

by the triangle inequality. Thus we can write:

$$\begin{aligned}
\|T_j f\| &= \|T_{r_1}^1 \dots T_{r_j}^j T_{1/2}^{j+1} \dots T_{1/2}^d f\|_q \\
&\leq \|T_{1/2}^{j+1}(T_{r_1}^1 \dots T_{r_j}^j T_{1/2}^{j+2} \dots T_{1/2}^d f)\|_q + 2^{O(d)}\gamma\|f\|_q && \text{(Equation (2.5))} \\
&\leq \|T_{r_{j+1}}^{j+1}(T_{r_1}^1 \dots T_{r_j}^j T_{1/2}^{j+2} \dots T_{1/2}^d f)\|_q + 2^{O(d)}\gamma\|f\|_q && \text{(Lemma 2.5.4)} \\
&\leq \|(T_{r_1}^1 \dots T_{r_{j+1}}^{j+1} T_{1/2}^{j+2} \dots T_{1/2}^d f)\|_q + 2^{O(d)}\gamma\|f\|_q && \text{(Equation (2.5))} \\
&= \|T_{j+1} f\|_q + 2^{O(d)}\gamma\|f\|_q,
\end{aligned}$$

completing the proof. □

The proof of the lower bound is very similar. We start again with a closely related single-variate lemma.

Lemma 2.5.5 ([308, Lemma 10.43])**.** *For any $q \geq 2$, there is an absolute constant $c_q \in (0,1)$ such that for any mean-0, real-valued random variable X satisfying $\|X\|_q \leq \infty$ and $a \in \mathbb{R}$:*

$$\|a - c_q X\|_q \leq \|a + X\|_q.$$

For $q \in \{4, 4/3\}$, we may take $c_q = 2/5$.

We can now prove the lower bound in Theorem 2.5.1 by similar arguments to the above.

Proof of Theorem 2.5.1 (Lower Bound). Similar to the upper bound, it suffices to show

$$\|T^1_{r_1 c_q} \dots T^d_{r_d c_q} f\|_q \leq \|f\|_q + 2^{O(d)} \gamma \|f\|_q.$$

Observe that we may re-write the lefthand side of the above inequality by breaking the norm into its boolean and product components:

$$\|T^1_{r_1 c_q} \dots T^d_{r_d c_q} f\|_q = \|\|T^1_{r_1 c_q} \dots T^d_{r_d c_q} f\|_{q,X}\|_{q,r}$$

It is therefore sufficient to prove that the operators $T^i_{r_i c_q}$ contract the inner q-norm for fixed $r_i \in \{\pm 1\}$. Since $0 \leq c_q \leq 1$, $T^i_{c_q}$ is an averaging operator and contracts q-norms by Jensen's inequality. Thus it is enough to show that for any g with finite q-norm:

$$\|T^i_{-c_q} g\|_q \leq \|g\|_q$$

By our localization lemma, Lemma 2.4.3, it is therefore sufficient to show the 'single-variate' version of this contraction for T_{-c_q} on any co-dimension 1 link, as then:

$$\|T^i_{-c_q} g\|_q = \|\|T^{x_{-i}}_{-c_q} g|_{x^{-i}}\|_{q,x_i}\|_{q,x_{-i}} \leq \|\|g|_{x^{-i}}\|_{q,x_i}\|_{q,x_{-i}} \leq \|g\|_q.$$

Since $T^{x_{-i}}_{-c_q}$ is self-adjoint, it is also enough to prove the inequality just for $q \geq 2$ case by Hölder conjugation (c.f. Lemma 2.3.5). We can now appeal to Lemma 2.5.5. In particular for any single variate function h

$$\begin{aligned} \|T_{-c_q} h\|_q &= \|h^{=\emptyset} - c_q h^{=1}\|_q \\ &\leq \|h^{=\emptyset} + h^{=1}\| && \text{(Lemma 2.5.5)} \\ &= \|h\|_q \end{aligned}$$

since $h^{=1}$ is mean-0 and bounded. This holds in particular on the co-dim 1 links of X, completing the proof. □

2.6 Optimal Global Hypercontractivity

We now give our first application of the symmetrization theorem, a simple proof of global hypercontractivity on HDX with (asymptotically) optimal parameters. We follow the high level strategy of [364], adapting the argument where necessary due to lack of independence.

Theorem 2.6.1 (Hypercontractivity on γ-products)**.** *Let X be a d-partite $(\{2,4\},\gamma)$-product with $\gamma \le 2^{-\Omega(d)}$. For any $i \le d$ and (ε, i)-global function $f \in C_d$:*

$$\|f^{\le i}\|_4^4 \le 2^{O(i)}\|f^{\le i}\|_2^2 \max_{|S|\le i, x_S}\{\|f|_{x_S}\|_2^2\} + 2^{O(d)}\gamma\|f\|_2^2 \max_{|S|\le i, x_S}\{\|f|_{x_S}\|_2^2\}.$$

Note that this is a finer-grained result than claimed in Theorem 2.1.2, which follows from Lemma 2.3.4 and by replacing $\|f^{\le i}\|_2^2$ with $(1+2^{O(d)}\gamma)\|f\|_2^2$ (the error term is then everywhere dominated by the main term for γ sufficiently small).

Before proving the result, we give the two mentioned corollaries regarding the operator-form and implications for low-influence functions.

Corollary 2.6.2. *Let X be a d-partite γ-product satisfying $\gamma \le 2^{-\Omega(d)}$ and $0 \le i \le d$. There exists a constant $\rho \in (0,1)$ such that for any mean-0 function $f : X(d) \to \mathbb{R}$ and $i \le d$*

$$\|T_\rho f^{\le i}\|_4^4 \le \|f\|_2^2 \max_{|S|\le i, x_S}\{\|f|_{x_S}\|_2^2\}.$$

Proof. By Lemma 2.3.11 and repeated application of Cauchy-Schwarz, we have

$$\|T_\rho f\|_4^4$$

$$
\begin{aligned}
&\leq 8\|\sum_{S\leq i}\rho^i f^{=S}\|_4^4 + 2^{O(d)}\gamma\|f\|_2^2 \max_{|S|\leq i, x_S}\{\|f|_{x_S}\|_2^2\} \\
&\leq 8\sum_{j\leq i}(8\rho)^j\|f^{=j}\|_4^4 + 2^{O(d)}\gamma\|f\|_2^2 \max_{|S|\leq i, x_S}\{\|f|_{x_S}\|_2^2\} \\
&\leq \sum_{j\leq i}2^{O(j)}(8\rho)^j\|f^{=j}\|_2^2 \max_{|S|\leq j, x_S}\{\|f|_{x_S}\|_2^2\} + 2^{O(d)}\gamma\|f\|_2^2 \max_{|S|\leq i, x_S}\{\|f|_{x_S}\|_2^2\} \quad \text{(Theorem 2.6.1)} \\
&\leq \frac{1}{2}\max_{|S|\leq i, x_S}\{\|f|_{x_S}\|_2^2\}\sum_{j\leq i}\|f^{=j}\|_2^2 + 2^{O(d)}\gamma\|f\|_2^2 \max_{|S|\leq i, x_S}\{\|f|_{x_S}\|_2^2\} \\
&\leq \max_{|S|\leq i, x_S}\{\|f|_{x_S}\|_2^2\}\|f\|_2^2 \quad \text{(Theorem 2.3.7)}
\end{aligned}
$$

for small enough γ, as desired. Note we have repeatedly used the fact that $f^{=\emptyset} = \mathbb{E}[f]\mathbf{1} = 0$. □

Corollary 2.6.3. *Let X be a d-partite γ-product and $f : X \to \mathbb{F}_2$ any function with influence $\mathbf{I}[f] \leq K\,Var(f)$. There exists $S \subset [d]$ with $|S| \leq O(K)$ and $x_S \in X[S]$ such that*

$$\mathbb{E}[f|_{x_S}] \geq 2^{-O(K)}.$$

Proof. The proof follows from a standard 'level-i inequality' implied by global hypercontractivity. Namely for any boolean function f, we claim the degree at most i Fourier mass is bounded by

$$\langle f, f^{\leq i}\rangle \leq 2^{O(i)}\mathbb{E}[f]\max_{|S|\leq i, x_S}\{\mathbb{E}[f|_{x_S}]^{1/4}\}$$

This follows from a basic application of Hölder's inequality:

$$
\begin{aligned}
\langle f, f^{\leq i}\rangle &\leq \|f\|_{4/3}\|f^{\leq i}\|_4 && \text{(Hölder)} \\
&\leq 2^{O(i)}\|f\|_{4/3}\|f\|_2^{1/2}\max_{|S|\leq i, x_S}\{\|f|_{x_S}\|_2^{1/2}\} && \text{(Theorem 2.6.1)} \\
&= 2^{O(i)}\mathbb{E}[f]\max_{|S|\leq i, x_S}\{\mathbb{E}[f|_{x_S}]^{1/4}\} && \text{(Booleanity)}
\end{aligned}
$$

On the other hand, recall the influence of f can be written as $\mathbf{I}[f] = \sum_{i=0}^{d} i\langle f, f^{=i}\rangle$, so in particular

$$K\mathrm{Var}(f) \geq \mathbf{I}[f] \geq (K+1)\left(\mathrm{Var}(f) - \sum_{i=1}^{K}\langle f, f^{=i}\rangle\right) - 2^{O(d)}\gamma\|f\|_2^2.$$

Re-arranging and using Booleanity, for small enough γ we then have

$$\langle f, f^{\leq K}\rangle \geq \frac{1}{2}\mathbb{E}[f],$$

which combined with the level-i inequality implies the claimed bound. □

We now turn to the proof of Theorem 2.6.1, which relies on the following corollary of the approximate Fourier properties of Efron-Stein from Section 2.3.

Claim 2.6.4.

$$\|T_r T_2 f^{\leq i} - \sum_{|S|\leq i} 2^{|S|} f^{=S} r_S\|_4 \leq 2^{O(d)}\gamma\|f\|_2^{1/2} \max_{|S|\leq i, x_S}\{\|f|_{x_S}\|_2^{1/2}\}$$

The proof is essentially immediate from Lemma 2.3.11 and Lemma 2.3.12, and is deferred to the end of the section. We can now prove global hypercontractivity.

Proof of Theorem 2.6.1. By symmetrization, Claim 2.6.4, and Cauchy-Schwarz:

$$\begin{aligned}
\|f^{\leq i}\|_4^4 &\leq 2\|T_r T_2 f^{\leq i}\|_4^4 \\
&= 2\mathbb{E}\left[\left(\sum_{|S|\leq i} 2^{|S|} r_S f^{=S} + \left(T_r T_2 f^{\leq i} - \sum_{|S|\leq i} 2^{|S|} r_S f^{=S}\right)\right)^4\right] \\
&\leq 16\mathbb{E}\left[\left(\sum_{|S|\leq i} 2^{|S|} r_S f^{=S}\right)^4\right] + 16\mathbb{E}\left[\left(T_r T_2 f^{\leq i} - \sum_{|S|\leq i} 2^{|S|} r_S f^{=S}\right)^4\right]
\end{aligned}$$

$$\leq 16\mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{r}\left[\left(\sum_{|S|\leq i}2^{|S|}f^{=S}(x)r_S\right)^4\right]\right]+2^{O(d)}\gamma\|f\|_2^2\max_{|S|\leq i,x_S}\{\|f|_{x_S}\|_2^2\}.$$

Notice that the inner expectation is now over a degree-i boolean function with Fourier coefficients $2^{|S|}f^{=S}(x)$. Thus applying the standard Bonami Lemma and Parseval gives:

$$\begin{aligned}
&16\mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{r}\left[\left(\sum_{|S|\leq i}2^{|S|}f^{=S}(x)r_S\right)^4\right]\right]\\
\leq&2^{O(i)}\mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{r}\left[\left(\sum_{|S|\leq i}2^{|S|}f^{=S}(x)r_S\right)^2\right]^2\right] && \text{(Bonami Lemma)}\\
=&2^{O(i)}\mathop{\mathbb{E}}_{x}\left[\left(\sum_{|S|\leq i}2^{2|S|}f^{=S}(x)^2\right)^2\right] && \text{(Parseval)}\\
\leq&2^{O(i)}\mathop{\mathbb{E}}_{x}\left[\left(\sum_{|S|\leq i}f^{=S}(x)^2\right)^2\right].
\end{aligned}$$

We now turn our attention to the term:

$$\mathop{\mathbb{E}}_{x}\left[\left(\sum_{|S|\leq i}f^{=S}(x)^2\right)^2\right]=\mathop{\mathbb{E}}_{x}\left[\left(\sum_{|S|\leq i}f^{=S}(x)^2\right)\left(\sum_{|T|\leq i}f^{=T}(x)^2\right)\right]$$

Our goal is now to isolate and pull out the shared variables between the two terms, so we can handle each term 'independently'. To this end, following [364] we re-index the sum over intersections $I=S\cap T$. In particular, we can re-write the above as

$$\sum_{|I|\leq i}\mathop{\mathbb{E}}_{x_I}\left[\sum_{S\supset I:|S|\leq i}\mathop{\mathbb{E}}_{x_{S\setminus I}\sim X_{x_I}}\left[f^{=S}(x_S)^2\left(\sum_{T:|T|\leq i,S\cap T=I}\mathop{\mathbb{E}}_{x_{T\setminus I}\sim X_{x_S}}\left[f^{=T}(x_T)^2\right]\right)\right]\right]. \tag{2.6}$$

On a product space, $x_{T\setminus I}$ and $x_{S\setminus I}$ are independent, which allows the above to upper

bounded by the much simpler expression

$$\sum_{|I|\le i} \mathop{\mathbb{E}}_{x} \left[\left(\sum_{S \supset I : |S| \le i} f^{=S}(x)^2 \right) \left(\sum_{T \supset I : |T| \le i} f^{=T}(x)^2 \right) \right].$$

Unfortunately, this strategy does not work as is on a γ-product. In particular, $x_{S \setminus I}$ and $x_{T \setminus I}$ are *not independent*, since the latter is drawn from the link of the former. A similar issue appeared in the analysis of hypercontractivity on two-sided HDX of [39]. The fix is to observe that since the right-hand function does not actually depend on the $x_{S \setminus I}$ variables in its input, we can use the swap walks to 'de-correlate' $x_{S \setminus I}$ and $x_{T \setminus I}$ up to an appropriate error term in γ.

Claim 2.6.5.

$$(2.6) \le \sum_{|I| \le i} \mathop{\mathbb{E}}_{x_I} \left[\left(\sum_{S \supset I : |S| \le i} \mathop{\mathbb{E}}_{x_{S \setminus I}} \left[f^{=S}(x_S)^2 \right] \right)^2 \right] + 2^{O(d)} \gamma \|f\|_2^2 \max_{|S| \le i, x_S} \left\{ \|f|_{x_S}\|_2^2 \right\}.$$

We defer the proof of the claim and complete the argument. Pulling out the maximum we can bound the main term by

$$\max_{|I| \le i, y_I} \left(\sum_{T \supset I : |T| \le i} \mathop{\mathbb{E}}_{x_{T \setminus I} \sim X_{y_I}} \left[f^{=T}(y_I, x_{T \setminus I})^2 \right] \right) \sum_{|I| \le i} \mathop{\mathbb{E}}_{x_I} \left[\left(\sum_{S \supset I : |S| \le i} \mathop{\mathbb{E}}_{x_{S \setminus I} \sim X_{x_I}} \left[f^{=S}(x_S)^2 \right] \right) \right]$$

We bound the two terms separately. The latter is the easier of the two, and can be bounded by approximate Parseval (Theorem 2.3.7):

$$\begin{aligned} \sum_{|I| \le i} \mathop{\mathbb{E}}_{x_I} \left[\left(\sum_{S \supset I : |S| \le i} \mathop{\mathbb{E}}_{x_{S \setminus I} \sim X_{x_I}} \left[f^{=S}(x_S)^2 \right] \right) \right] &= \left(\sum_{|I| \le i} \sum_{S \supset I : |S| \le i} \mathop{\mathbb{E}}_{x_S} \left[f^{=S}(x_S)^2 \right] \right) \\ &\le 2^i \sum_{|S| \le i} \langle f^{=S}, f^{=S} \rangle \\ &\le 2^i \|f^{\le i}\|_2^2 + 2^{O(d)} \gamma \|f\|_2^2. \end{aligned}$$

To bound the maximum, recall from Lemma 2.3.13 that

$$f^{=I\cup B}(y_I, x_B) = \sum_{J\subset I}(-1)^{|I|-|J|}(f|_{y_J})^{=B}(x_B).$$

Thus for any possible $y_I \in X[I]$, a simple application of Cauchy-Schwarz gives:

$$\begin{aligned}
\sum_{T\supset I:|T|\le i} \mathop{\mathbb{E}}_{x_{T\setminus I}\sim X_{y_I}} \left[f^{=T}(y_I, x_{T\setminus I})^2\right] &\le \sum_{T\supset I} \mathop{\mathbb{E}}_{x_{T\setminus I}\sim X_{y_I}} \left[f^{=T}(y_I, x_{T\setminus I})^2\right] \\
&\le 2^i \sum_{J\subset I}\sum_{S\subset \bar{I}} \mathop{\mathbb{E}}_{x_S}[(f_{y_J})^{=S}(x_S)^2] \\
&\le 2^i \sum_{J\subset I}\sum_{S\subset \bar{I}} \mathop{\mathbb{E}}_{x_S}[(f_{y_J})^{=S}(x_S)^2] \\
&\le 2^{O(i)} \max_{|J|\le|I|,x_J}\{\|f|_{x_J}\|_2^2\}.
\end{aligned}$$

Putting everything together, we get the final bound

$$\|f^{\le i}\|_4^4 \le 2^{O(i)}\|f^{\le i}\|_2^2 \max_{|S|\le i,x_S}\left\{\|f|_{x_S}\|_2^2\right\} + 2^{O(d)}\gamma\|f\|_2^2 \max_{|S|\le i,x_S}\left\{\|f|_{x_S}\|_2^2\right\}$$

as desired. □

2.6.1 Auxiliary Proofs

We now give the proofs of Claim 2.6.4 and Claim 2.6.5.

Proof of Claim 2.6.4. The proof is immediate from linearity of the noise operator and the triangle inequality if we can show for any $|S| \le i$:

$$\|T_r T_2 f^{=S} - 2^{|S|} r_S f^{=S}\|_4 \le 2^{O(d)}\gamma\|f\|_2^{1/2} \max_{|S|\le i,x_S}\{\|f|_{x_S}\|_2^{1/2}\}.$$

By Claim 2.3.15 we can write

$$T_r T_2 f^{=S} = \sum_{S' \subseteq [d]} 2^{|S|} T_r (f^{=S})^{=S'}.$$

Since the projection operators $\{E_T\}$ contract p-norms and T_ρ is a (bounded) linear combination of projections, Lemma 2.3.12 implies the righthand side satisfies

$$\| \sum_{S' \subseteq [d]} 2^{|S|} T_r (f^{=S})^{=S'} - 2^{|S|} T_r f^{=S} \|_4 \leq 2^{O(d)} \gamma \|f\|_2^{1/2} \max_{|S| \leq i, x_S} \{ \|f|_{x_S}\|_2^{1/2} \}$$

Expanding out T_r, the main term is then

$$2^{|S|} \sum_{T \subseteq [d]} r_T \prod_{i \notin T} (1 - r_i) E_T f^{=S}.$$

Now by Lemma 2.3.9, $\|E_T f^{=S}\|_4 \leq 2^{O(d)} \gamma \|f\|_2^{1/2} \max_{|S| \leq i, x_S} \{\|f|_{x_S}\|_2^{1/2}\}$ unless $S \subset T$, in which case $E_T f^{=S} = f^{=S}$. Thus the above is close in 4-norm to

$$2^{|S|} \sum_{T \supset S} r_T \prod_{i \notin T} (1 - r_i) f^{=S} = 2^{|S|} r_S f^{=S}$$

as in Claim 2.3.15. □

Proof of Claim 2.6.5. Recall our goal is to de-correlate $x_{S \setminus I}$ and $x_{T \setminus I}$ to approximately upper bound

$$\sum_{|I| \leq i} \mathop{\mathbb{E}}_{x_I} \left[\sum_{S \supset I : |S| \leq i} \mathop{\mathbb{E}}_{x_{S \setminus I} \sim X_{x_I}} [f^{=S}(x_S)^2] \left(\sum_{T : |T| \leq i, S \cap T = I} \mathop{\mathbb{E}}_{x_{T \setminus I} \sim X_{x_S}} [f^{=T}(x_T)^2] \right) \right]$$

by the 'independent' variant

$$\sum_{|I|\leq i} \mathop{\mathbb{E}}_{x_I} \left[\left(\sum_{S \supset I : |S| \leq i} \mathop{\mathbb{E}}_{x_{S\setminus I} \sim X_{x_I}} \left[f^{=S}(x_S)^2 \right] \right)^2 \right]. \tag{2.7}$$

To do this, first note we may pull out the inner sum in the former expectation and write

$$\sum_{|I|\leq i} \mathop{\mathbb{E}}_{x_I} \left[\sum_{S \supset I : |S| \leq i} \sum_{T : |T| \leq i, S \cap T = I} \mathop{\mathbb{E}}_{x_{S\setminus I} \sim X_{x_I}} \left[f^{=S}(x_S)^2 \mathop{\mathbb{E}}_{x_{T\setminus I} \sim X_{x_S}} \left[f^{=T}(x_T)^2 \right] \right] \right]$$

where we first draw x_I, then $x_{S\setminus I}$ conditional on x_I, and finally $x_{T\setminus I}$ conditional on x_S.

The trick is now to re-write each inner expectation as an application of the swap walk on the link of x_I:

$$\begin{aligned} \mathop{\mathbb{E}}_{x_{S\setminus I} \sim X_{x_I}} \left[f^{=S}(x_S)^2 \mathop{\mathbb{E}}_{x_{T\setminus I} \sim X_{x_S}} \left[f^{=T}(x_T)^2 \right] \right] &= \left\langle (f^{=S}|_{x_I})^2, A^{x_I}_{T\setminus I, S\setminus I} (f^{=T}|_{x_I})^2 \right\rangle_{X_{x_I}} \\ &\leq \mathop{\mathbb{E}}_{x_{S\setminus I} \sim X_{x_I}} [(f^{=S}|_{x_I})^2] \mathop{\mathbb{E}}_{x_{T\setminus I} \sim X_{x_I}} [(f^{=T}|_{x_I})^2] \\ &\quad + 2^{O(d)} \gamma \|f^{=S}|_{x_I}\|_4^2 \|f^{=T}|_{x_I}\|_4^2 \end{aligned}$$

where the final step is by Cauchy-Schwarz and expansion of the swap walk. Since $x_{S\setminus I}$ and $x_{T\setminus I}$ are now independent (conditioned on x_I), we can plug them back into the above and upper bound by Equation (2.7) as in the product setting.

It is left to handle the error term, which can similarly be upper bounded as

$$\begin{aligned} 2^{O(d)} \gamma \sum_{|I|\leq i} \mathop{\mathbb{E}}_{x_I} \left[\left(\sum_{S \supset I : |S| \leq i} \mathbb{E}[f^{=S}(x)^4]^{1/2} \right)^2 \right] &\leq 2^{O(d)} \max_{|S|\leq i} \left\{ \|f^{=S}\|_4^4 \right\} \\ &\leq 2^{O(d)} \gamma \|f\|_2^{1/2} \max_{|S|\leq i, x_S} \{ \|f|_{x_S}\|_2^{1/2} \} \end{aligned}$$

where the first inequality is by Cauchy-Schwarz and the latter Lemma 2.3.8. □

2.7 Bourgain's Booster Theorem

In this section we prove γ-products satisfy a booster theorem, resolving a main open question of [39].

Theorem 2.7.1. *Let X be a $(\{2,4\},\gamma)$-product for $\gamma \le 2^{-\Omega(d)}$ and $f : X(d) \to \{\pm 1\}$ a function with $\mathbf{I}[f] \le K$. If $Var(f) \ge .01$, there is some $|\tau| \ge 2^{-O(K^2)}$ such that*

$$\mathbb{P}_{x\sim X(d)}[\exists T \subset [d] : \ |T| \le O(K) \text{ and } x_T \text{ is a } \tau\text{-booster}] \ge |\tau|$$

Our proof closely follows O'Donnell's treatment of Bourgain's Theorem for product spaces in [308], adjusting where necessary to handle lack of independence. The key technical component is to show that for any low influence function, it is possible to identify a small *input-dependent* set of coordinates which account for most of the Fourier mass. This is similar to the proof of Friedgut's Junta theorem, which does this in an input-independent fashion. Unfortunately, the latter is not possible on products or HDX.

Proposition 2.7.2. *Let X be a $(\{2,4\},\gamma)$-product with $\gamma \le 2^{-\Omega(d)}$, $\varepsilon \in (2^{O(d)}\gamma, 1/2)$, and $f : X(d) \to \{\pm 1\}$ any function satisfying $Var(f) \ge .01$. Let $\ell = \mathbf{I}[f]/\varepsilon$. There exists a family of 'notable coordinates' $\{J_x\}_{x\in X(d)}$ such that*

1. *J_x is small:*

$$\forall x, |J_x| \le 2^{O(\ell)}$$

2. *Small subsets of J_x contain of most of the Fourier mass:*

$$\mathbb{E}\left[\sum_{S\notin \mathcal{F}_x} f^{=S}(x)^2\right] \le 3\varepsilon$$

where $\mathcal{F}_x \coloneqq \{S : S \subseteq J_x, |S| \le \ell\}$.

We first prove the booster theorem assuming this fact. The proof is essentially exactly as in the product case (see e.g. [308, Page 311]). We include the argument for completeness.

Proof of Theorem 2.7.1. Set ε of Proposition 2.7.2 to .001 (note this is admissible when $\gamma \leq 2^{-\Theta(d)}$ is small enough), and let $|\mathcal{F}_x| \leq 2^{O(K^2)}$ be the resulting family of subsets. Since we have assumed $\mathrm{Var}(f) \geq .01$, we have $f^{=\emptyset}(x)^2 \leq .99$, so

$$\mathbb{E}_x \left[\sum_{S \in \mathcal{F}_x \setminus \{\emptyset\}} f^{=S}(x)^2 \right] \geq 1 - 3\varepsilon - .99 - 2^{O(d)}\gamma \geq .005$$

for small enough γ. As a result, we must have that the maximum is at least $2^{-O(K^2)}$:

$$\mathbb{E}_x \left[\max_{S \in \mathcal{F}_x \setminus \{\emptyset\}} f^{=S}(x)^2 \right].$$

In particular, this means for every x we can define a set $0 < |S_x| \leq O(K)$ such that

$$\mathbb{E}_x \left[f^{=S_x}(x)^2 \right] \geq 2^{-O(K^2)}$$

Since $f^{=S_x}(x) = \sum_{T \subseteq S_x} (-1)^{|S_x|-|T|} E_T f(x) \leq 2^{|S_x|}$ for $\{\pm 1\}$-valued f, so we have $f^{=S_x}(x)^2 \leq 2^{O(K)}$ for all terms above and therefore

$$\mathbb{P}_x[f^{=S_x}(x)^2 \geq 2^{-O(K^2)}] \geq 2^{-O(K^2)}.$$

Finally, we argue that whenever $f^{=S_x}(x)^2 \geq 2^{-O(K^2)}$, there exists $T \subseteq S_x$ which is a $2^{-O(K^2)}$-booster, completing the proof.

Toward this end, let $g = f - \mathbb{E}[f]$, and note for all $S \neq \emptyset$, $g^{=S} = f^{=S}$. Since $S_x \neq \emptyset$,

we therefore have $|g^{=S_x}(x)| \geq 2^{-O(K^2)}$. Finally, recall

$$g^{=S_x}(x) = \sum_{\emptyset \neq T \subseteq S_x} (-1)^{|S_x|-|T|} E_T g(x),$$

so there must exist some $0 < T \leq O(K)$ such that $|E_T g| \geq 2^{-O(K^2)}$. However since $g = f - \mathbb{E}[f]$, we then have

$$|E_T g(x)| = |\mathbb{E}[f|_{x_T}] - \mathbb{E}[f]| \geq 2^{-O(K^2)}$$

as desired. □

We now prove the core Proposition.

Proof of Proposition 2.7.2. First observe that the expected Fourier mass on components beyond level ℓ is small

$$\mathbb{E}_x \left[\sum_{|S|>\ell} f^{=S}(x)^2 \right] \leq 2\varepsilon.$$

This follows by viewing the components $\sum_{|S|=i} \mathbb{E}_x[f^{=S}(x)^2]$ as an approximate distribution over $[d]$. Namely using the fact that

$$1 = \mathbb{E}_x[f^2] \in (1 \pm 2^{O(d)}\gamma) \sum_{S \subseteq [d]} \mathbb{E}_x[f^{=S}(x)^2],$$

for small enough γ, there is some normalizing factor $c \in [.9, 1.1]$ such that $c\mathbb{E}_x[f^{=S}(x)^2]$ is a distribution. Now consider the random variable Z that takes value $|S|$ with probability $c\mathbb{E}_x[f^{=S}(x)^2]$. Markov's inequality implies

$$\mathbb{P}[Z > \ell] \leq \frac{\mathbb{E}[Z]}{\ell} = \frac{c^{-1}\mathbf{I}[f] + c_d\gamma}{\ell} = c^{-1}\varepsilon + \frac{\varepsilon c_d \gamma}{I[f]} \leq 2\varepsilon$$

for small enough γ, where we've used Corollary 2.3.17 and the standard fact that $\mathbf{I}[f] \geq$

$d\mathrm{Var}[f]$.

We may now restrict our attention to the components of degree at most ℓ. In particular, we need to find small sets J_x such that

$$\mathbb{E}_x \left[\sum_{|S| \le \ell, S \not\subseteq J_x} f^{=S}(x)^2 \right] \le \varepsilon$$

The natural strategy to define such a set is simply to take any coordinate with large influence, that is

$$J'_x := \left\{ j \in [d] : \sum_{S \ni j} f^{=S}(x)^2 \ge \tau \right\}$$

where $\tau = 2^{-\Theta(\ell)}$ is a sufficiently small constant. In fact we will show for this definition

$$\mathbb{E}_x \left[\sum_{|S| \le \ell, S \not\subseteq J'_x} f^{=S}(x)^2 \right] \le \varepsilon/2 \tag{2.8}$$

The issue is that $|J'_x|$ may not be bounded. We will argue that truncating J'_x as

$$J_x = \begin{cases} J'_x & \text{if } |J'_x| \le C^\ell \\ \emptyset & \text{otherwise} \end{cases}$$

for some large enough constant $C > 0$, gives the desired set family. In particular, it is enough to additionally show

$$\mathbb{E} \left[\mathbf{1}[|J'_x| > C^\ell] \sum_{0 < |S| \le \ell} f^{=S}(x)^2 \right] \le \varepsilon/2. \tag{2.9}$$

We first show Equation (2.8). Toward this end, we'll require the following Lemma (generalizing [308, Lemma 10.48] to γ-products) which follows from standard hypercontractivity and approximate Parseval. We defer the proof.

Lemma 2.7.3. *Fix $x \in X(d)$ and $i \notin J'_x$. If $\gamma \leq 2^{-\Omega(d)}$ is sufficiently small, for $g = T_{2/5}L_i f$ we have:*

$$\|T_{\frac{1}{\sqrt{3}}}\tilde{g}|_x\|_2^2 \leq 2\tau^{1/3}\|\tilde{g}|_x\|_{4/3}^{4/3}$$

We now prove Equation (2.8):

$$\begin{aligned}
\mathbb{E}_x\left[\sum_{|S|\leq \ell, S\not\subseteq J'_x} f^{=S}(x)^2\right] &\leq 2^{O(\ell)}\mathbb{E}_x\left[\sum_{S\not\subseteq J'_x}(T_{1/\sqrt{3}}T_{2/5}f^{=S})(x)^2\right] + 2^{O(d)}\gamma \\
&\leq 2^{O(\ell)}\mathbb{E}_x\left[\sum_{i\notin J'_x}\sum_{S\ni i}(T_{1/\sqrt{3}}T_{2/5}f^{=S})(x)^2\right] + 2^{O(d)}\gamma \\
&\leq 2^{O(\ell)}\mathbb{E}_x\left[\sum_{i\notin J'_x}(T_{1/\sqrt{3}}T_{2/5}L_i f)(x)^2\right] + 2^{O(d)}\gamma
\end{aligned}$$

where the last step follows from observing

$$\begin{aligned}
&\mathbb{E}_x\left[\sum_{S\ni i}(T_{1/\sqrt{3}}T_{2/5}f^{=S})(x)^2\right] \\
=&\mathbb{E}_x\left[(T_{1/\sqrt{3}}T_{2/5}L_i f)(x)^2\right] - \sum_{S\neq S'\ni i}\langle T_{1/\sqrt{3}}T_{2/5}f^{=S}, T_{1/\sqrt{3}}T_{2/5}f^{=S'}\rangle \\
\leq& 2^{O(d)}\gamma
\end{aligned}$$

by approximate orthogonality and the fact that

$$\|T_{1/\sqrt{3}}T_{2/5}f^{=T} - (\frac{2}{\sqrt{35}})^{|T|}f^{=T}\|_2 \leq 2^{O(d)}\gamma\|f\|_2$$

for any T by the same argument as in Claim 2.6.4.

Define $g^i \coloneqq T_{2/5}L_i f$. We'd like to pass to the symmetrization of g^i so we can apply Lemma 2.7.3. To this end, we observe that

$$\mathbb{E}_x\left[T_{1/\sqrt{3}}g^i(x)^2\right] \leq \mathbb{E}_x\left[\|T_{1/\sqrt{3}}\widetilde{g^i}|_x\|_{2,r}^2\right] + 2^{O(d)}\gamma$$

This follows since by definition $\widetilde{g^i}|_x$ is the boolean function whose Fourier coefficients are $(g^i)^{=S}(x)$, so by Parseval

$$\begin{aligned}\mathbb{E}_x\left[\|T_{1/\sqrt{3}}\widetilde{g^i}|_x\|_{2,r}^2\right] &= \mathbb{E}_x\left[\sum_{S\subseteq[d]}\left(\frac{1}{\sqrt{3}}\right)^{|S|}(g^i)^{=S}(x)^2\right] \\ &\geq \mathbb{E}_x\left[T_{1/\sqrt{3}}g^i(x)^2\right] - 2^{O(d)}\gamma\end{aligned}$$

where the last step is again by the fact that Efron-Stein is approximately orthogonal and approximately an eigenbasis.

This allows us to apply Lemma 2.7.3 and 'un-symmetrize'. Namely we have:

$$\begin{aligned}\mathbb{E}_x\left[\sum_{|S|\leq\ell,S\not\subseteq J'_x} f^{=S}(x)^2\right] &\leq 2^{O(\ell)}\sum_{i\notin J'_x}\mathbb{E}_x\left[\|T_{1/\sqrt{3}}\widetilde{g^i}|_x\|_{2,r}^2\right] + 2^{O(d)}\gamma \\ &\leq 2^{O(\ell)}\tau^{1/3}\sum_{i\notin J'_x}\mathbb{E}_x\left[\|\widetilde{g^i}|_x\|_{4/3,r}^{4/3}\right] + 2^{O(d)}\gamma \\ &\leq 2^{O(\ell)}\tau^{1/3}\sum_{i\in[d]}\|L_i f\|_{4/3}^{4/3} + 2^{O(d)}\gamma \\ &\leq 2^{O(\ell)}\tau^{1/3}\mathbf{I}[f] + 2^{O(d)}\gamma \\ &\leq \varepsilon/2\end{aligned}$$

for small enough τ, γ. Note here we've used the standard inequality $\|L_i f\|_{4/3}^{4/3} \leq O(\langle f, L_i f\rangle)$ (see e.g. [308, Exercise 8.10]) which holds for $\{\pm 1\}$-valued functions via relating the coordinate-wise influence to the probability $f(x) \neq f(x^{(i)})$ up re-randomizing the ith coordinate.

We now need to show that $|J'_x|$ is typically small (Equation (2.9)). We first separate

the indicator and Fourier sum by Cauchy-Schwarz:

$$\mathbb{E}\left[\mathbf{1}[|J'_x| > C^\ell] \sum_{0<|S|\le\ell} f^{=S}(x)^2\right] \le \mathop{\mathbb{P}}_x[|J'_x| > C^\ell]^{1/2} \cdot \mathbb{E}\left[\left(\sum_{0<|S|\le\ell} f^{=S}(x)^2\right)^2\right]^{1/2}$$

We bound the two terms separately. The first can be bounded by a simple application of Markov and

$$\begin{aligned}
\mathop{\mathbb{P}}_x[|J'_x| > C^\ell] &\le C^{-\ell}\mathbb{E}_x[|J'_x|] \\
&\le C^{-\ell}\mathbb{E}_x[\frac{1}{\tau}\sum_{i\in[d]}\sum_{S\ni i} f^{=S}(x)^2] \\
&\le \frac{C^{-\ell}}{\tau}(I[f] + 2^{O(d)}\gamma)
\end{aligned}$$

for large enough C and small enough γ.

Bounding the latter term requires slightly more care. Let $h = T_{2/5}(f - f^{=\emptyset})$. We have

$$\begin{aligned}
\mathbb{E}_x\left[\left(\sum_{0<|S|\le\ell} f^{=S}(x)^2\right)^2\right] &\le 2^{O(\ell)}\mathbb{E}\left[\left(\sum_{S\neq\emptyset} T_{2/5} f^{=S}(x)^2\right)^2\right] + 2^{O(d)}\gamma \\
&\le 2^{O(\ell)}\mathbb{E}_x\left[\|\tilde{h}|_x\|_{2,r}^4\right] + 2^{O(d)}\gamma \\
&\le 2^{O(\ell)}\mathbb{E}_x\left[\|\tilde{h}|_x\|_{4,r}^4\right] + 2^{O(d)}\gamma \\
&\le 2^{O(\ell)}\|f - f^{\emptyset}\|_4^4 + 2^{O(d)}\gamma \\
&\le 2^{O(\ell)}\mathrm{Var}(f) + 2^{O(d)}\gamma \\
&\le 2^{O(\ell)}\mathbf{I}[f] + 2^{O(d)}\gamma
\end{aligned}$$

where we have used the standard fact that $\mathrm{Var}(f) \le \mathbf{I}[f]$, and in the second inequality have applied the same properties of approximate Efron-Stein as in the prior arguments.

Combining these we have

$$\mathbb{E}\left[\mathbf{1}[|J'_x| > C^\ell] \sum_{0<|S|\le\ell} f^{=S}(x)^2\right] \le (2^{O(\ell)}\mathbf{I}[f] + 2^{O(d)}\gamma)^{1/2}\left(\frac{C^{-\ell}}{\tau}(I[f] + 2^{O(d)}\gamma)\right)^{1/2} \le \varepsilon/2$$

for large enough constant $C > \tau^{-1}$ and small enough γ. □

It is left to prove Lemma 2.7.3.

Proof of Lemma 2.7.3. We first appeal to $(4/3, 2)$-hypercontractivity which states that $\|T_{1/\sqrt{3}}f\|_2 \le \|f\|_{4/3}$ for any $f : \{\pm 1\}^d \to \mathbb{R}$. Then we have:

$$\|T_{1/\sqrt{3}}\tilde{g}|_x\|_2^2 \le \left(\|\tilde{g}|_x\|_{4/3}^2\right)^{1/3} \cdot \|\tilde{g}|_x\|_{4/3}^{4/3} \le \left(\|\tilde{g}|_x\|_2^2\right)^{1/3} \cdot \|\tilde{g}|_x\|_{4/3}^{4/3}$$

Our goal is now to bound $\|\tilde{g}|_x\|_2^2 \le \tau + 2^{O(d)}\gamma \le 2\tau$. By standard Parseval

$$\begin{aligned}\|\tilde{g}|_x\|_2^2 &= \sum_{S\subseteq[d]} (T_{2/5}L_i f)^{=S}(x)^2 \\ &\le \sum_{S\ni i} (2/5)^{2|S|} f^{=S}(x)^2 + 2^{O(d)}\gamma \\ &\le \tau + 2^{O(d)}\gamma\end{aligned}$$

since $i \notin J'_x$. The second inequality comes expanding $L_i f$ as $\sum_{T\ni i} f^{=T}$ and applying Lemma 2.3.11 and Lemma 2.3.12. □

Acknowledgements

We thank Vedat Alev, Nima Anari, Venkat Guruswami, Tali Kaufman, and Shachar Lovett for helpful discussions on hypercontractivity and q-norm HDX. We thank the Weizmann Institute and the Simons Institute for the Theory of Computing for hosting the author for part of the completion of this work.

Chapter 3

On Low Influence Functions on Weak HDX

3.1 Introduction

In 1988, Kahn, Kalai, and Linial [226] proved the 'KKL Theorem': a characterization of low influence (non-expanding) functions[1] on the hypercube that revolutionized the field of boolean function analysis. In recent years, variants of the KKL Theorem *beyond the cube*, including on products [78, 349, 160, 157, 194, 245, 280], Lie groups [139], and the Grassmann have become a powerful tool in hardness of approximation, leading to the theory of sharp thresholds [158], the 2-2 Games Conjecture [253, 126, 125, 47, 252, 255], and PCPs with near-optimal alphabet-soundness tradeoff [297]. Despite this, our understanding of KKL-type theorems on general spaces remains poor. As a result, many approaches in this direction (e.g. towards the unique games conjecture) have hit substantial barriers.

Toward this end, a series of works [239, 111, 38, 39, 187, 162] proposed a new framework for unifying our understanding of KKL-type Theorems beyond the cube: *high dimensional expanders* (HDX). HDX are a generalization of expanders to hypergraphs and ranked posets that have seen a recent explosion of application within theoretical computer science [239, 111, 38, 39, 187, 162, 124, 109, 236, 9, 119, 208, 52, 25, 11, 24, 95, 96, 94, 146,

[1] Informally, a low-influence function on the hypercube is one which has low expected probability of changing value under a random coordinate flip.

218, 286, 66, 117, 315, 127, 114, 113, 40]. Recently, Bafna, Hopkins, Kaufman, and Lovett [38, 39] and Gur, Lifshitz, and Liu [187] proved a KKL-Type Theorem for high dimensional expanders stating that any low influence function must be *local*, highly concentrated in some restriction of the hypergraph. Unfortunately, BHKL and GLL's techniques only work in the regime of *near-perfect* expansion, a very strong requirement only known for three algebraic families of sparse constructions [290, 238, 108], and as such remain difficult to handle in application where simpler constructions might be more amenable (e.g. in classical PCP reductions).

In this work, we study KKL type theorems on weak high dimensional expanders. It is typical in this setting tow ork from the contrapositive view, which states that any function f which is '(ε, i)-global', i.e. f does not increase by more than ε upon restricting i vertices, expands. Our main result is an expansion theorem which scales with the underlying local-spectral parameters of the hypergraph X, and a corresponding family of weak HDX exhibiting a similar lower bound.

Theorem 3.1.1 (Structure of Low Influence Functions (Informal))**.** *For any hypergraph X and (ε, i)-global function $f : X(k) \to \mathbb{F}_2$ on k-faces of X:*

$$\Phi(f) \geq (1 - \mathbb{E}[f]) \frac{1}{k-i} \prod_{j=i}^{k-2} (1 - \gamma_j) + c_{k,\gamma}\varepsilon$$

Moreover, for every $i \leq k$ there exists an infinite family of hypergraphs $\{X_n\}$ and $(0, i)$-global functions $\{f_n : X_n(k) \to \mathbb{F}_2\}$ satisfying

$$\Phi(f) \leq (1 - \mathbb{E}[f]) \frac{i+1}{k} \prod_{j=i}^{k-2} (1 - \gamma_j)$$

Here $\Phi(f)$ is the total influence (expansion) of f and $\{\gamma_j\}$-correspond to the largest non-trivial *positive* eigenvalues on j-links of X (see Section 3.1.1). Under weak assumptions, we emphasize $c_{k,\gamma} \leq O(1)$ is an *absolute constant*. Here the result should be compared to

[39, 187] who show

$$\Phi(f) \geq (1 - \mathbb{E}[f])\frac{i+1}{k} + O_i(\varepsilon) + \exp(-k),$$

essentially matching the lower bound, but only when $\gamma_j, |\gamma_j^{(-)}| \leq \exp(-k)$. It remains an important question whether weak HDX match this behavior. See Section 3.1.4 for further discussion.

Leveraging Theorem 3.1.1, we give new characterizations of low-influence functions on natural families of combinatorial high dimensional expanders, including *clique complexes* and *product-based HDX*. As an immediate corollary, we also derive a new 'Kruskal-Katona' theorem for global functions on such objects. Finally, using similar ideas, we prove a 'local-to-global' *small-set expansion theorem* (a KKL variant corresponding to the noise operator), and as an application give the first such result for *Ramanujan complexes* [290], a seminal construction of high dimensional expanders failing the requirements of [39, 187].

At a technical level, our work builds on a growing body of Fourier analytic tools for high dimensional expanders [239, 111, 38, 39, 187, 162, 109, 9, 176] and relies in particular on a recent decomposition theorem of Gotlib and Kaufman (GK) [176]. Our work makes two main technical contributions beyond these prior methods. First, we show that as long as the complex X has *negative* local spectra bounded away from -1, global functions have bounded projection onto low levels of the GK-decomposition. Second, inspired by GK's proof technique, we consider an *inductive* approach for bounding expansion through a new variant of Garland's Lemma, breaking the expansion of f as a function-dependent expectation across local parts of the complex. This latter component is critical to ensure the density dependence in the resulting characterization is independent of dimension. In comparison, directly applying GK's decomposition gives an expansion bound for arbitrary one-sided HDX (that is with no assumption on the negative spectra), but deteriorates with the *number of vertices* of X [175]. Extending our results to the true one-sided case

remains an interesting open problem.

3.1.1 Background

High Dimensional Expanders:.

We focus on weighted *pure simplicial complexes* (X, Π) where

$$X = X(0) \cup \ldots \cup X(d)$$

for $X(d) \subset \binom{n}{d}$ an arbitrary d-uniform hypergraph and $X(i) \subset \binom{n}{i}$ given by downward closure, and

$$\Pi = (\pi_0, \ldots, \pi_d)$$

for π_d an arbitrary distribution over $X(d)$ and π_i given by selecting a uniformly random i-set from $\tau \sim \pi_d$.

Our bounds will depend on a recently popular local notion of high dimensional expansion which examines the structure of local components called *links*. For every face $\tau \in X(i)$, the link of τ is given by:

$$X_\tau \coloneqq \{\sigma \in X : \sigma \cup \tau \in X\}.$$

Given a function $f : X(k) \to \mathbb{R}$, its localization $f|_\tau : X_\tau(k - |\tau|) \to \mathbb{R}$ is $f_\tau(\sigma) = f(\tau \cup \sigma)$.

Denote the normalized adjacency matrix of the graph underlying X_τ by A_τ, and for $0 \leq j \leq d-2$ denote the worst-case spectral parameters of j-links by

$$\gamma_j \coloneqq \max_{\tau \in X(j)} \{\lambda_2(A_\tau)\}, \qquad \gamma_j^{(-)} \coloneqq \min_{\tau \in X(j)} \{\lambda_{\min}(A_\tau)\}.$$

A complex is called *strongly connected* if every $\gamma_j < 1$, a (one-sided) *local-spectral expander* if every γ_j is bounded away from 1 [309], and a (two-sided) local-spectral expander if every $\gamma_j^{(-)}$ is additionally bounded away from -1 [124].

Influence and the Down-Up Walk:.

As in the seminal work of KKL, we are interested in characterizing the structure of boolean functions $C_k(X, \mathbb{F}_2) \coloneqq \{f : X(k) \to \mathbb{F}_2\}$ with low total influence. In our setting it will be convenient to use an equivalent definition (up to normalization) concerning the expansion of a celebrated random walk on complexes called the *lower* or *down-up* walk. The lower walk at level k, denoted N_k^1, moves between k-faces of the complex via a shared $(k-1)$-face, removing a vertex uniformly at random and re-sampling from the appropriate conditional distribution on the remaining $(k-1)$-face. The *edge expansion* of a set $S \subset X(k)$ with respect to the lower walk is the expected probability of leaving S in a single step, or equivalently:

$$\Phi(S) = 1 - \frac{\langle N_k^1 1_S, 1_S \rangle}{\langle 1_S, 1_S \rangle}.$$

It is not hard to see that expansion, which can also be written as $\frac{\langle 1_S, L 1_S \rangle}{\langle 1_S, 1_S \rangle}$ for the 'Laplacian Operator' L, is equivalent to standard notions of influence up to normalization (see Section 3.2 for details), so we will focus on this notion throughout instead.

Small-Set Expansion and the Noise Operator:.

The small-set expansion theorem is a classical and closely related result to the KKL theorem [3, 226]. In its standard form, it states that small sets on the noisy cube expand near perfectly. To prove an analog result beyond the cube, we need to define a corresponding notion of the noise operator T_ρ on simplicial complexes. Informally, T_ρ is the random walk on k-faces of X which fixes each vertex with probability ρ, and re-samples the remaining vertices. Formally, the operator can be expressed as a convex combination of 'longer' down-up walks

$$T_\rho \coloneqq \sum_{i=0}^{k} \binom{k}{i} (1-\rho)^{k-i} \rho^i N_k^{k-i},$$

where N_k^j moves between k-faces via a shared $(k-j)$-face. Note this recovers the standard operator when X is a product space (see [39]). We will always write expansion with respect to the noise operator at Φ_{T_ρ}, and write just Φ for the down-up walk. Finally, it will sometimes be useful to reference the opposite variant of the lower walks, the 'upper' walks $\widehat{N}_k^i$, which moves between k-faces via a shared $(k+i)$-face.

3.1.2 Results

We are now ready to more formally cover our results. We split the subsection into two parts: structure theorems for low-influence functions on arbitrary HDX, and a local-to-global small-set expansion theorem.

A Structure Theorem for Low Influence Functions

The canonical examples of low influence sets on simplicial complexes are *links*.[2] Our goal is to show these are the *only* such examples. To this end, we call a function f (ε, i)-global if it is uncorrelated with i-links:

$$\forall \tau \in X(i) : \mathop{\mathbb{E}}_{X_\tau}[f_\tau] \leq \mathbb{E}[f] + \varepsilon.$$

We prove that global functions expand, where the quantitative parameters depend on the underlying local-spectral expansion of the complex.

Theorem 3.1.2 (Expansion of Global Functions (Informal Theorem 3.3.1))**.** *Let (X, Π) be a k-dimensional simplicial complex and $f \in C_k$ any (ε, i)-global boolean function. Then*

$$\Phi(f) \geq \frac{1 - \mathbb{E}[f]}{k-i} \prod_{j=i}^{k-2}(1-\gamma_j) - c_{k,i,\gamma}\varepsilon.$$

[2]Note here we really mean the set of k-face including some fixed i-face (rather than removing the common element as in the formal definition).

where $c_{k,i,\gamma} \leq (k-i+1)\left(\frac{1}{k-i}\prod_{j=i}^{k-2}(1-\gamma_j) - \frac{1}{k-i+1}\prod_{j=i-1}^{k-2}(1-\gamma_j)\right)\left(1+(k-i)\gamma_{i-1}^{(-)}\right)^{-1}$

For many regimes of interest, e.g. when either γ_{k-2} or $|\gamma_{k-2}^{(-)}|$ is less than roughly $\frac{1}{k}$, $c_{\gamma,k,i} \leq O(1)$ is an *absolute constant*, with no dependence on dimension. In such cases the contrapositive of Theorem 3.1.2 states any non-expanding function has constant density in some restriction.

Perhaps the most important question regarding Theorem 3.1.2 is to what extent the bound is tight. Using a combinatorial HDX construction of Golowich [170, 172] and Liu, Mohanty, and Yang [287], we construct families of global functions on weak HDX matching the above up to replacing the dependence on $\frac{1}{k-i}$ with $\frac{i+1}{k}$.

Theorem 3.1.3 ((Non)-Expansion of Global Functions (Informal Theorem 3.3.2))**.** *For every* $0 < i < k$*, there exists a* k*-dimensional family of complexes* $\{X_n\}$ *and* $(0,i)$*-global functions* $\{f_n\}$ *such that*

$$\Phi(f_n) \leq (1-\mathbb{E}[f])\frac{i+1}{k}\prod_{j=i}^{k-2}(1-\gamma_j).$$

We discuss the two dependencies in these bounds and their relation to bounds in the literature further in Section 3.1.4. The former, $\frac{1}{k-i}$, is the natural consequence of an inductive approach, while $\frac{i+1}{k}$ corresponds to the approximate eigenvalues of HDX with $\exp(-k)$-local-spectral expansion [111, 38]. It remains an interesting question which dependence is correct for weak HDX.

As applications of Theorem 3.1.2, we give the first structure theorems for simpler combinatorial constructions of high dimensional expanders. We'll start by looking at one of the most classical constructions of simplicial complexes: *clique-complexes*. Given a graph $G = (V, E)$, the k-dimensional clique-complex $K_{G,k}$ is the complex induced by taking the uniform distribution over k-cliques of G. Using Theorem 3.1.2, we give the first non-trivial characterization of non-expanding sets on (dense) clique-complexes.

Corollary 3.1.4 (Expansion in Clique-Complexes (Informal Theorem 3.5.2))**.** *Fix $k \in \mathbb{N}$ and let $G = (V, E)$ be any graph with minimum degree at least $\Delta_{min} \geq \frac{2k-2}{2k-1}|V|$. Then the expansion of any (ε, i)-global boolean function f on the k-dimensional clique-complex $K_{G,k}$ is at least:*

$$\Phi(f) \geq (1 - \mathbb{E}[f])\frac{(i+1)}{k(k-i)} - O(\varepsilon)$$

While Corollary 3.1.4 is non-trivial in any dimension, it is strongest in low dimensions. For instance, an interesting implication for 3-clique (triangle) complexes is that any function of triangles with expansion worse than 1/3 must have a dense vertex.

Corollary 3.1.5 (Low Influence Functions on Triangle Complexes (Informal Corollary 3.5.3))**.** *Let $G = (V, E)$ be any graph of minimum degree at least $\Delta_{min} \geq \frac{5}{6}|V|$ and $S \subset K_{G,3}(3)$ any set with expansion*

$$\Phi(S) \leq \frac{1}{3} - \delta.$$

Then there exists a vertex $v \in G$ such that S contains at least a $\frac{\delta}{2}$-fraction of the triangles touching v.

Of course, clique-complexes arising from dense graphs are themselves dense objects. A major selling points of prior work was the ability to characterize low-influence functions on several families of *sparse* complexes. While our work cannot match the quantitative strength of these results, it does lead to new insight on the structure of functions on simpler constructions of bounded-degree HDX with quantitatively weaker parameters. In particular, we consider (cutoffs of) the elementary combinatorial HDX of Golowich, which tensors a bounded-degree expander G with the complete complex to give a weak bounded-degree two-sided HDX.

Corollary 3.1.6 (Expansion in Combinatorial HDX (Informal Corollary 3.5.5))**.** *Let X be a $2k$-dimensional 'product-HDX' of any graph G and $\Delta_{4k}(2k)$, and $f \in C_k$ any (ε, i)-global*

boolean function. Then

$$\Phi(f) \geq \frac{(1-\mathbb{E}[f])i}{(k-1)(k-i)} - O(\varepsilon)$$

As for the case of clique-complexes, this is especially powerful in low dimensions. For concreteness, we again look at the first non-trivial setting of $k = 3$.

Corollary 3.1.7 (Low Influence Functions on Combinatorial HDX (Informal Corollary 3.5.6))**.** *Let X be a 3-dimensional 'product-HDX' of any graph G and $\Delta_n(6)$, and $S \subset X(3)$ be any subset of triangles with expansion at most*

$$\Phi(S) \leq \frac{1}{4} - \delta.$$

Then there exists a vertex $v \in X(1)$ such that S contains a $\Omega(\delta)$ fraction of triangles including the vertex v.

Finally, we note that as an immediate corollary of Theorem 3.1.2 we also get a 'Kruskal-Katona' type theorem in all the above settings. Kruskal-Katona is a seminal and broadly used result in extremal combinatorics which given a set $S \subset X(k)$, lower bounds the number of $(k-1)$-faces that sit inside S, called its 'lower shadow' and denoted ∂S.

Corollary 3.1.8 (Kruskal-Katona for Weak HDX)**.** *Let (X, Π) be a k-dimensional simplicial complex and $S \subset X(k)$ any (ε, i)-global set. Then*

$$\mathbb{E}[\partial S] \geq \mathbb{E}[S]\left(1 + \frac{1-\mathbb{E}[S]}{k-i}\prod_{j=i}^{k-2}(1-\gamma_j) - c_{k,i,\gamma}\varepsilon.\right)$$

where $c_{k,i,\gamma} \leq (k-i+1)\left(\frac{1}{k-i}\prod_{j=i}^{k-2}(1-\gamma_j) - \frac{1}{k-i+1}\prod_{j=i-1}^{k-2}(1-\gamma_j)\right)\left(1+(k-i)\gamma_{i-1}^{(-)}\right)^{-1}$

Small-Set Expansion

Similar to the setting of the lower walk, links also provide a canonical family of non-expanding sets for the noise operator on simplicial complexes. In this setting, it

is typical [245, 187, 39] to use a slightly stronger notion of global functions we'll call '(ε, i)-strongly global', which simply assumes f is sparse on every link:

$$\forall \tau \in X(i) : \mathop{\mathbb{E}}_{X_\tau}[f_\tau] \leq \varepsilon.$$

BHKL [39] and GLL [187] prove small-set expansion for strongly global functions on near-perfect two-sided and one-sided partite high dimensional expanders. While these properties are satisfied by some constructions (e.g. [239]), the seminal construction of good HDX, the so-called *Ramanujan complexes* of Lubotzky, Samuels, and Vishne [290], actually fail these conditions. These complexes (and their variants, e.g. [113, 40]) satisfy a variety of useful topological properties [119, 208]

One of the advantages of the Ramanujan complexes is their simple *local* structure. In particular, their links are near-perfect partite HDX, and therefore satisfy a small-set expanison theorem. This raises a natural question: is any complex X whose links satisfy global small-set expansion itself a global small-set expander? We answer this question in the affirmative.

Theorem 3.1.9 (Local-to-Global SSE (Informal Corollary 3.4.6))**.** *Let X be a complex such that $\forall v \in X(1)$:*

1. *Any $(\varepsilon, i-1)$-strongly global function f on X_v has expansion*

$$\Phi_{T_\rho}(f) \geq \phi(\varepsilon, k, i)$$

2. *X_v is a γ-two-sided or γ-one-sided partite HDX*

Then any (ε, i)-global boolean function f has expansion at least

$$\Phi_{T_\rho}(f) \geq \phi(\varepsilon, k, i) - 2^{O(k)}\gamma - 2^{-\Omega(k)}.$$

We note the requirement that links of X are HDX can be removed up to worsened dependence in the dimensional error term (see Corollary 3.4.6). The inverse exponential error above is essentially negligible, since any function with respect to the noise operator has non-expansion at least inverse exponential in k.

As an immediate corollary, we prove the Ramanujan complexes are global small-set expanders.

Corollary 3.1.10 (Ramanujan Complexes are Global SSEs (Informal Corollary 3.5.9))**.** *Let $f \in C_k$ be an (ε, i)-strongly global boolean function on a Ramanujan Complex with $2^{-\Omega(k)}$ one-sided local-spectral expansion. The expansion of f is at least*

$$\Phi_{T_\rho}(f) \geq 1 - \rho^{i+1} - O_i(\varepsilon) - 2^{-\Omega(k)}.$$

3.1.3 Techniques

We sketch the proofs of Theorem 3.1.2 and Theorem 3.1.9, starting with the former.

Upper Bound:.

Our general approach to prove global sets expand is broken into three core components. The first is a new function-dependent variant of 'Garland's Lemma', a classical method of breaking global functions into local components over the complex. In particular, we show the expansion of a function f with respect to the lower walks can be written as an expectation of its expansion over links of X.

Lemma 3.1.11 (Garland's Lemma for Expansion (Informal Lemma 3.3.5)**.** *Let (X, Π) be a weighted, pure simplicial complex and $f \in C_k$ any function on k-faces. Then for all $i \leq k$ and $j \leq k - i$:*

$$\Phi(f) = \mathop{\mathbb{E}}_{\tau \sim \pi_j^f} [\Phi_{X_\tau}(f|_\tau)], \tag{3.1}$$

where $f|_\tau(\sigma) = \tau \cup \sigma$ is the localization of f to the link of τ, and π_j^f is some f-dependent distribution over i-faces of X.

The second core component is a variant of the GK's decomposition theorem [176] for the lower walk.

Theorem 3.1.12 (GK-Decomposition (Informal Theorem 3.7.1))**.** *Let (X, Π) be a k-dimensional simplicial complex. Then for any $f \in C_k$ there is an orthogonal decomposition $f = \sum_{i=0}^{k} f_i$ such that:*

$$\langle N_1^k f, f\rangle \leq \mathbb{E}[f]^2 + \sum_{i=1}^{k-1} \left(1 - \frac{1}{k-i+1} \prod_{j=i-1}^{k-2} (1-\gamma_j)\right) \langle f, f_i\rangle.$$

Finally, the third main component is a 'level-1 inequality' showing any $(1, \varepsilon)$-global function has low projection onto the first level of the GK-decomposition.

Proposition 3.1.13 (Level-1 Inequality (Informal Corollary 3.3.9))**.** *Let (X, Π) be a k-dimensional complex and f any $(\varepsilon, 1)$-global boolean function $f \in C_k$. Then*

$$\langle f, f_1\rangle \leq \frac{k}{1 + \gamma_0^{(-)}(k-1)}\varepsilon.$$

We remark that this is a corollary of a more general level i inequality that scales with the minimum non-zero eigenvalue of the upper walks (see Theorem 3.3.8). One could instead apply this directly with the GK decomposition, but the resulting dependence on ε is substantially worse than can be achieved by combining these with Garland's method.

Given these three components, Theorem 3.1.2 follows from a fairly simple induction. We first prove the statement directly for $(\varepsilon, 1)$-global functions using the GK-Decomposition and Level-1 inequality. For general i, we use Garland's lemma to reduce to $i = 1$ inside links of X, writing the expansion as $\Phi(f) = \mathbb{E}_{\tau \sim \pi_{i-1}^f}[\Phi_\tau(f|_\tau)]$, and observe that the localization $f|_\tau$ of a (ε, i) global function is itself $(1, \varepsilon - \delta_\tau)$ global for some $|\delta_\tau| \leq \varepsilon$. Applying the statement for 1-global functions inside the expectation (carefully handling dependence on $\mathbb{E}_{\tau \sim \pi_{i-1}^f}[\delta_\tau]$) then gives the claimed bound.

Lower Bound:.

Our lower bound relies on the 'product-HDX' of Golowich [170, 172] and Liu, Mohanty, and Yang [287]. Roughly speaking, these complexes are constructed by tensoring a 'low-dimensional' complex X (say of dimension j) and a 'high-dimensional' complete complex $Y = \Delta_k(n)$ in the following fashion:

$$F_{X,Y}(k) = \{s = (x_1, y_1), \ldots, (x_k, y_k) : \Pi_X(s) \in X(j) \wedge \Pi_Y(s) \in Y(k)\}$$

where Π_X and Π_Y are the projections onto the first and second coordinates respectively. Equipped with the correct weight function, $F_{X,Y}$ has top level local-spectral expansion roughly $\gamma_{k-2}, |\gamma_{k-2}^{(-)}| \approx \frac{1}{k-1}$ [172].

For every $i \leq k$, we'd like to construct an i-global function $f \in C_k$ with expansion as close as possible to Theorem 3.1.2. To do this, we will use a product complex where X has dimension $i+1$ and Y has dimension k. The idea of the construction is to 'lift' an i-global boolean function f on X to the product. In particular, to construct f, for every i-face $\tau \in X$, we set $f = 1$ on exactly a $\mu = \mathbb{E}[f]$ fraction of the $(i+1)$-faces containing τ. We may then take the final function f' on the product to simply be the value of f on the face's X-projection. It is easy to check that if X and Y are sufficiently regular, this results in a perfectly balanced function at level i of the product F.

Recall that the down-up walk operates on a face $s = \{(x_1, y_1), \ldots, (x_k, y_k)\}$ by removing a uniformly random vertex (x_i, y_i), and re-sampling (x_i', y_i') such that $s \setminus (x_i, y_i) \cup (x_i', y_i') \in X(k)$. To analyze (non)-expansion of this walk, we observe the only way to escape f' is for the lower walk to escape the X projection of s, $\pi_X(s) = \{x_1, \ldots, x_k\}$. This only occurs when down-step removes a vertex (x_i, y_i) where the x_i value is unique within the face. The weighting of product complexes ensures that every multiset pattern of X-values appears with equal probability. As a result, a direct multinomial computation shows a unique x is selected in the down step exactly with probability $\frac{i+1}{k} \prod_{j=i}^{k-2}(1 - \gamma_j)$. Finally, if a

unique vertex is selected, the conditional probability of then escaping f' in the up-step is exactly $(1-\mu)$ (the fraction of possible resulting new X-projections that land in f) as desired.

SSE and the Noise Operator:.

The first main component of our low influence characterization, Lemma 3.1.11, already shows that expansion of the lower walk exhibits local-to-global behavior. The proof of Theorem 3.1.9 would follow immediately if the same form of localization were to hold for the noise operator. Unfortunately, while the lemma may be extended to lower walks of arbitrary length, such a statement does not hold directly for the noise operator.

To see why, recall the noise operator is a convex combination of lower walks $T_\rho = \sum_{i=0}^{k} B_k^\rho(i) N_k^{k-i}$ where $B_k^\rho(i) = \binom{k}{i}(1-\rho)^{k-i}\rho^i$. Since expansion is linear, for any function f we can write

$$\Phi_{T_\rho}(f) = \sum_{i=0}^{k} B_k^\rho(i) \Phi_{N_k^{k-i}}(f).$$

Localizing T_ρ naively hits two main issues. First, a walk N_k^i can only be appropriately localized to j-links when $j \leq k-i$ (otherwise the walk goes below the level of the link, and cannot be captured at this level of locality). Second, the *coefficients* of T_ρ, which are binomially distributed, depend on dimension and therefore do not localize correctly. In particular, after localizing 'permissible' the lower walks in the sum, we'd get:

$$\Phi_{T_\rho}(f) = \sum_{i=0}^{j-1} B_k^\rho(i) \Phi_{N_k^i}(f) + \sum_{i=j}^{k} B_k^\rho(i) \Phi_{N_{k-j}^{k-i-j}}(f|_\tau).$$

To correspond to the local noise operator on the right-hand side, we'd instead need the corresponding coefficients to be $B_{k-j}^\rho(i)$, the binomial distribution on $k-j$ trials, not k trials as in the original instance. A first approach to solving these issues is to define a 'shifted' noise operator over the original space that simply throws out the lefthand sum, which is negligible, and replaces the remaining coefficients with $B_{k-j}^\rho(i)$ over the

original space. One can then argue that the expansion with respect to the shifted operator 'localizes correctly' (allowing us to apply local SSE), and is moreover close to the expansion of the original noise operator. Unfortunately, this procedure results in error scaling with the TV-distance between Binomial distributions on k and $k-j$ trials. This costs $\text{poly}(k^{-1})$ additive error, and therefore does not give the desired global SSE theorem.

To fix this issue, we instead make the stronger assumption that links of X are good two-sided or partite one-sided high dimensional expanders, a property satisfied by our main motivating example (the Ramanujan complexes). This allows us to appeal to the theory of Fourier analysis developed in [111, 38, 39, 187], which shows one may decompose expansion with respect to the lower walks as

$$\Phi_{N_k^i}(f) \approx \frac{1}{\langle f, f\rangle} \sum_{i=0}^{k} \lambda(N_k^i)\langle f_\ell, f_\ell\rangle$$

for some decomposition $f = \sum_{i=0}^{k} f_i$ and approximate eigenvalues $\lambda(N_k^i)$. With this in mind, instead of shifting the noise operator, we can decompose the localized expansion in the Fourier basis as

$$\Phi_{T_\rho}(f) \lesssim \sum_{i=0}^{j-1} B_k^\rho(i) + \mathop{\mathbb{E}}_{\tau\sim\pi_j^f}\left[\frac{1}{\langle f|_\tau, f|_\tau\rangle}\sum_{i=j}^{k} B_k^\rho(i) \sum_{\ell=0}^{k-j} \lambda_\ell(N_{k-j}^{k-i-j}|_\tau)\langle (f|_\tau)_\ell, (f|_\tau)_\ell\rangle\right]$$

Finally for each fixed ℓ, we show its corresponding coefficient in the above sum is at most ρ^ℓ, the (approximate) eigenvalue of the local noise operator itself [39, 187]. This allows us to write the desired localized inequality

$$\Phi_{T_\rho}(f) \lesssim \sum_{i=0}^{j-1} B_k^\rho(i) + \mathop{\mathbb{E}}_{\tau\sim\pi_j^f}\left[\Phi_{T_\rho}(f|_\tau)\right]$$

which may finally be bounded by the local SSE theorem using the fact that any j-restriction $f|_\tau$ of an (ε, i)-strongly global function is $(2\varepsilon, i-j)$-strongly global. The error in this

analysis scales only with the cost of cutting off small i (the first term) and the approximation error of the Fourier decomposition. For constant j, the former is at most $\exp(-k)$, while the latter is $2^{O(k)}\gamma$ when j-links are γ-HDX. Thus for complexes with sufficiently expanding links, we get an SSE theorem with negligible additive error as desired.

3.1.4 Discussion and Related Work

We close the section with a more in-depth comparison with prior work and discuss a natural open problem such comparisons raise. At a high level, our work fits into a long line of research examining the structure of low-influence functions beyond the hypercube, starting not long after the KKL Theorem itself with the study of product spaces and the p-biased cube [78, 349, 160, 157, 194, 245] and more recently in extended settings such as the slice [252], multi-slice [148, 335], symmetric group [147], Grassmannian [255, 137], high dimensional expanders [38, 39, 187, 162, 176], and Lie groups [139]. We rely heavily on the Fourier analytic machinery initiated by Dikstein, Dinur, Filmus, and Harsha [111], and Kaufman and Oppenheim [239], and further extended by Bafna, Hopkins, Kaufman, and Lovett [38, 39], Gur Lifshitz, and Liu [187], Hopkins [200], and Gotlib and Kaufman [176].

Concretely, several bounds are known for the expansion of global sets with respect to the lower walk. In the regime of near-perfect expansion, BHKL [38] prove any (ε, i)-global function has expansion at least

$$\Phi(f) \gtrsim (1 - \mathbb{E}[f])\frac{i+1}{k} - \binom{k}{i}\varepsilon.$$

This is tight in the regime that $\varepsilon \ll \frac{1}{\binom{k}{i}}$, and mimics the structure one would expect to see on general k-fold product spaces. For strongly global functions, BHKL [39], GLL [187],

and Hopkins [200] improve the ε-dependence to be *independent of dimension*:

$$\Phi(f) \gtrsim (1 - \mathbb{E}[f])\frac{i+1}{k} - 2^{O(i)}\varepsilon.$$

In the regime of weak high dimensional expansion, the only known bound for global functions is due to Gotlib and Kaufman [175], who show that on any *one-sided* HDX:

$$\Phi(f) \geq (1 - \mathbb{E}[f])\frac{1}{k-i}\prod_{j=i}^{k-2}(1-\gamma_j) + |X(k)|\varepsilon.$$

Unlike the prior bounds, however, this is only really meaningful when $\varepsilon \approx 0$, a non-trivial but extremely strong notion of globalness that does not lead to meaningful structural characterizations.

In some sense our bound, Theorem 3.1.2, sits somewhere between the latter two. it has the beneficial properties of the latter in that it holds for global functions and under arbitrary (two-sided) local-spectral expansion, and of the former in that the dependence on ε is typically constant. On the other hand, due to being obtained through an inductive approach, it also inherits the main term of Gotlib-Kaufman [176], scaling with $\frac{1}{k-i}$ instead of $\frac{i+1}{k}$. When k is large, the former dependence is substantially worse.

At a technical level, this difference in main term stems from the fact that our 'advantage' for i-global functions, like in Gotlib-Kaufman [176], comes from the second eigenvalue of the lower walk inside i-links of X, which scales as $1 - \frac{1}{k-i+1}\prod_{j=i-1}^{k-2}(1-\gamma_j)$. On the other hand, on very strong HDX, the advantage of i-global functions comes from the ith approximate eigenvalue of the global lower walk on X itself, which scales with $1 - \frac{i}{k}$, combined with a strong level-i inequality. It is natural to conjecture the 'correct' dependence sits between these two bounds, as $1 - \frac{i}{k}\prod_{j=i-1}^{k-2}(1-\gamma_j)$. This is exactly the form appearing in our lower bound for global expansion of product-HDX. Unfortunately, both the techniques of this paper and Fourier analytic techniques from prior work (which

typically incur exponential error) are very far from proving such a bound.

3.2 Background

We now give a more formal introduction to the language of high dimensional expansion, including simplicial complexes, local-spectral expansion, high order random walks, and influence/expansion of sets.

3.2.1 Simplicial Complexes and Local-Spectral Expansion

We study the structure of low influence functions on weighted *pure simplicial complexes*.

Definition 3.2.1 (Simplicial Complex)**.** A weighted, d-dimensional[3] pure simplicial complex (X, Π) on n vertices consists of a complex

$$X = X(0) \cup \ldots \cup X(d)$$

where $X(d) \subseteq \binom{[n]}{d}$ is a d-uniform hypergraph and each $X(i) \subseteq \binom{[n]}{i}$ is given by downward closure:

$$X(i) \coloneqq \left\{ \sigma \in \binom{[n]}{i} : \exists \tau \in X(d), \sigma \subset \tau \right\},$$

and a joint distribution $\Pi = (\pi_1, \ldots, \pi_d)$ over $X(0) \times \ldots \times X(d)$ where π_i is induced from π_d by sampling a uniformly random size-i subset:

$$\pi_i(\sigma) = \frac{1}{\binom{k}{i}} \sum_{\tau \supset \sigma : |\tau| = d} \pi_d(\tau).$$

Since all complexes considered in this work will be pure, weighted, and d-dimensional, we drop these monikers throughout and usually refer to (X, Π) as just a simplicial complex.

[3]We note that our notion of dimension is off by 1 from much of the historical HDX literature, but generally leads to simpler expressions in our setting.

We write Δ_n to denote the complete simplicial complex on n vertices.

We will sometimes work in the special setting of *partite* complexes, which often play a special role in the theory of high dimensional expansion.

Definition 3.2.2 (Partite Complex)**.** A simplicial complex (X, Π) is called partite if its vertices can be partitioned into d 'colors':

$$X(1) = X^1 \amalg \ldots \amalg X^d$$

such that every d-face has exactly one vertex from each color.

There are many notions of high dimensional expansion on simplicial complexes. In this work we take the recently popular *local-spectral* approach that analyzes the spectrum of local components called *links*.

Definition 3.2.3 (Link)**.** Given a complex (X, Π) and a face $\tau \in X(i)$, the link of τ is the $(d-i)$-dimensional sub-complex (X_τ, Π_τ) where

$$X_\tau \coloneqq \{\sigma \in X : \sigma \cup \tau \in X\}$$

and $\Pi_\tau = (\pi_{\tau,0}, \ldots, \pi_{\tau,d-i})$ is the distribution induced by restricting Π to X_τ in the natural manner

$$\pi_{\tau,d-i}(\sigma) = \frac{\pi_d(\sigma \cup \tau)}{\sum_{\sigma' \in X_\tau} \pi_d(\sigma' \cup \tau)}.$$

Typically, a complex is said to be a *local-spectral expander* if the graph underlying each (co-dimension ≥ 2) link is a spectral expander [309, 239].

Definition 3.2.4 (Underlying Graph)**.** Given a simplicial complex (X, Π) and a link (X_τ, Π_τ), the (edge)-weighted graph underlying X_τ is given by $G_{X_\tau} = (X_\tau(1), X_\tau(2), \pi_2)$,

and has weighted adjacency (random walk) matrix

$$A_\tau(v, w) = \pi_{\tau \cup v, 1}(w).$$

In our work, it will be convenient to have finer grain control over the local spectral structure of these operators than is typically considered in the literature.

Definition 3.2.5 (Local Spectra)**.** Let (X, Π) be a simplicial complex. We define the local-spectral parameters $\{\gamma_i\}_{i=0}^{d-2}$ and $\{\gamma_i^{(-)}\}_{i=0}^{d-2}$ to be the worst non-trivial positive and negative eigenvalues respectively across links of each dimension:

$$\gamma_i \coloneqq \max_{\tau \in X(j)} \{\lambda_2(A_\tau)\}, \qquad \gamma_i^{(-)} \coloneqq \min_{\tau \in X(j)} \{\lambda_{\min}(A_\tau)\}.$$

In more standard language, an infinite family of complexes are called *one-sided* local-spectral expanders [309] if every γ_i is bounded away from 1, and *two-sided* local-spectral expanders [124] if additionally each $\gamma_i^{(-)}$ is also bounded away from -1.

3.2.2 Averaging Operators and the KO-Decomposition

Given a complex (X, Π), let $C_k = C_k(X, \mathbb{R}) \coloneqq \{f : X(k) \to \mathbb{R}\}$ denote the space of real-valued functions over k-sets of X. The function spaces associated with any complex come with a set of standard analysis tools, including a natural weighted inner product:

$$\langle f, g \rangle \coloneqq \mathop{\mathbb{E}}_{\pi_i}[fg],$$

and *averaging operators* (sometimes called 'unsigned boundary operators') that map between the C_i. In particular, for every $0 \leq i < d$, the *Up Operator* $U_i : C_i \to C_{i+1}$ lifts

functions between levels:

$$\forall \sigma \in X(i+1): \; U_i f(\sigma) = \frac{1}{i+1} \sum_{\tau \subset \sigma: \; |\tau|=i} f(\tau),$$

and for every $0 < i \leq d$ the *Down Operator* $D_i : C_i \to C_{i-1}$ conversely lowers functions:

$$\forall \sigma \in X(i-1): \; D_i f(\sigma) = \sum_{\tau \supset \sigma: \; |\tau|=i} \pi_{\sigma,1}(\tau \setminus \sigma) f(\tau).$$

It will often be convenient to lift or lower a function multiple levels, which can be done through the following compositions

$$U_i^k \coloneqq U_{k-1} \circ \ldots \circ U_i, \qquad D_i^k \coloneqq D_{i+1} \circ \ldots \circ D_k.$$

A crucial and well-known fact is that the averaging operators D and U are adjoint.

Lemma 3.2.6. *Let (X, Π) be a simplicial complex and $0 \leq k < d$. Then for all $f \in C_k$ and $g \in C_{k+1}$ we have*

$$\langle U_i f, g \rangle = \langle f, D_{i+1} g \rangle.$$

The averaging operators also give a natural formalization of what it means for a function to 'come from below,' namely being in $\mathrm{Im}(U_i^k)$. This can be used in a number of ways to build natural function decompositions on simplicial complexes that break up a function into contributions from each level. In this work, we rely on the basis developed by Kaufman and Oppenheim [239] which can be stated in terms of the averaging operators as follows.

Definition 3.2.7 (KO-Decomposition [239])**.** Let (X, Π) be a simplicial complex, $0 \leq k \leq d$, and $f \in C_k$. There exists an orthogonal decomposition $f = \sum_{i=0}^{k} f_i$ such that each $f_i \in Im(U_i^k) \cap Ker(D_{i-1}^k)$.

3.2.3 High Order Random Walks

Just as expander graphs are inextricably tied to their underlying random walk, high dimensional expanders also bear close connection to high order analogs on their underlying complex. We will focus in particular in this work on two classical settings, the *down-up walks*, and the *noise operator*.

Definition 3.2.8 (Down-up Walk [234, 124])**.** Given a complex (X, Π), for any $0 < k \leq d$ the down-up walk on k-faces, denoted N_k^1, walks between k-faces of X via a shared $(k-1)$-face, and can be written formally as

$$N_k^1 \coloneqq U_{k-1} D_k.$$

It will also be useful for the sake of analysis to have access to longer versions of the down-up walk, as well as their corresponding up-down walks. With this in mind, define

$$N_k^i \coloneqq U_{k-i}^k D_{k-i}^k, \qquad \widehat{N_k^i} \coloneqq D_{k-i}^k U_{k-i}^k$$

to be the walks that move between k-faces via a shared $(k-i)$-face and $(k+i)$-face respectively. Finally, we'll also need the following 'non-lazy' variant of the up-down walk

$$M_k^+ \coloneqq \frac{k}{k+1} D_{k+1} U_k - \frac{1}{k+1} I.$$

Note that by adjointness of D and U, all these walks are self-adjoint and therefore have spectral decompositions.

The lower walk can be phrased as the random process that removes a uniformly random element $v \in \sigma$, and re-samples a new element conditioned on the remaining face $\sigma \setminus \{v\}$. We will also study a variant of this process called the *noise operator*, which removes *each* element in σ with some fixed probability, then jointly re-samples these

elements conditioned on the rest.

Definition 3.2.9 (The Noise Operator)**.** Given a complex (X, Π), for any $0 < k \leq d$ and $0 \leq \rho \leq 1$, the ρ-noise operator on k-faces, denoted T^k_ρ, re-samples each vertex with probability $1 - \rho$:

$$T^k_\rho \coloneqq \sum_{i=0}^{k} \binom{k}{i} (1-\rho)^i \rho^{k-i} N^i_k,$$

3.2.4 Boolean Function Analysis and Expansion

The down-up walk and its variants capture a broad variety of structures studied throughout theoretical computer science. On the complete complex it gives the Johnson graphs, on spin-systems it gives the celebrated Glauber Dynamics [24], and on the natural complex embedding of $\mathbb{F}_2^d$ it simply results in the standard hypercube graph [39] (up to laziness). It is this final connection which allows us to generalize the classical notion of *total influence* of a boolean function, traditionally defined the hypercube as[4]

$$I(f) = \sum_{i \in d} \mathbb{E}_{x \sim \mathbb{F}_2^d} \left[\left(\frac{f(x) - f(x^{\oplus i})}{2} \right)^2 \right] = d\langle f, Lf \rangle$$

where $L = I - U_{k-1} D_k$ is the standard Laplacian operator. The latter equality leads to a natural interpretation on general simplicial complexes.

Definition 3.2.10 (Total Influence on Complexes [111, 39, 187])**.** Given a complex (X, Π) and $0 < k \leq d$, the total influence of a boolean function $f \in C_k$ on k-sets is

$$I(f) \coloneqq k\langle f, (I - U_{k-1} D_k) f \rangle$$

The total influence of a boolean function is also classically related to its *combinatorial expansion.*

[4]Here $x^{\oplus i}$ denotes the string x with the ith bit flipped.

Definition 3.2.11 (Combinatorial (Edge)-Expansion)**.** Let M be random walk on universe Ω with stationary distribution π. The edge expansion of a subset $S \subset \Omega$ is the expected probability of leaving S in a single step of M:

$$\Phi_M(f) \coloneqq \mathbb{E}_{\pi|_S}\left[M(v, X \setminus S)\right]$$

where $\pi|_S(x)$ is the natural distribution induced from restricting π to S and $M(v, X \setminus S)$ is the total outgoing weight from v. It will also sometimes be useful to refer to **non-expansion**, denoted $\bar{\Phi}_M \coloneqq 1 - \Phi_M$.

It is a simple exercise to show that the expansion of S can also be written as the Rayleigh quotient of 1_S with respect to $I - M$:

$$\Phi_M(1_S) = \frac{\langle 1_S, (I - M)1_S\rangle_\pi}{\langle 1_S, 1_S\rangle_\pi}$$

where $\langle f, g\rangle_\pi = \mathbb{E}_\pi[fg]$. With this in mind, it is an easy observation that since the stationary distribution of N_k^1 is π_k, the total influence of a function is (up to normalization) exactly its expansion with respect to the down-up walk.

Expansion of the noise operator is also a classical object of study in boolean function analysis, where it is sometimes referred to as a function's *noise-sensitivity*. It is a classical result that sparse functions on the hypercube are noise-sensitive

Theorem 3.2.12. *Let $X = \mathbb{F}_2^k$ be the hypercube complex. Then for any $0 \leq \rho \leq 1$ and $S \subset X(k)$:*

$$\Phi_{T_\rho}(S) \geq 1 - \mathbb{E}[1_S]^{\frac{1-\rho}{1+\rho}}$$

This result is also called the *small-set expansion theorem*, as it states that small sets on the noisy hypercube expand near perfectly.

3.2.5 Global Functions

The small-set expansion theorem fails inherently over unbalanced domains, even in settings as simple as the p-biased cube. In particular, 'local' functions like indicators or juntas which would be large sets on the cube become small, but maintain their otherwise poor expansion. A long and influential line of work has studied the extension of small-set expansion theorems to this and other settings byeond the cube by showing these are the *only* bad examples. To this end, we call a function *global* if it does not become much denser upon any restriction.

Definition 3.2.13 (Global Function)**.** A function $f : X(k) \to \mathbb{F}_2$ is called (ε, i)-weakly global if

$$\forall \tau \in X(i) : \mathop{\mathbb{E}}_{X_\tau}[f_\tau] \leq \mathbb{E}[f] + \varepsilon.$$

KKL-type theorems for global functions are sometimes called 'booster' theorems, since they state any function with low total influence must see a density 'bump' or 'boost' inside some small restriction. This is the content, for instnace, of Bourgain's influential Booster Theorem used in analysis of sharp thresholds [158].

For the case of small-set expansion, a more typical definition in the literature beyond the cube is to require the stronger guarantee that the set is not strongly correlated with any restriction. We call such functions *strongly* global.

Definition 3.2.14 (Strongly Global Function)**.** A function $f : X(k) \to \mathbb{F}_2$ is called (ε, i)-strongly global if

$$\forall \tau \in X(i) : \mathop{\mathbb{E}}_{X_\tau}[f_\tau] \leq \varepsilon.$$

Small-set expansion for strongly global functions roughly translates to statements of the form:

$$\Phi_{T_\rho}(S) \geq 1 - \rho^i - c_i \varepsilon^{O_\rho(1)}.$$

In other words, any function with expansion substantially worse than $1-\rho^i$ must be dense inside some i-link.

3.2.6 Localization, Garland's Method, and the Trickling-Down Theorem

One of the most powerful tools in the study of high dimensional expansion is the idea of localization: breaking analysis of global functions or properties into localized parts. Given $f \in C_k$ and $\tau \in X(i)$, we define the *localization* of f to τ, $f|_\tau : X_\tau(k-i) \to \mathbb{R}$, as

$$\forall \sigma \in X_\tau(k-|\tau|) : f|_\tau(\sigma) = f(\sigma \cup \tau).$$

We will similarly need to consider localizations of the averaging operators. For any $\tau \in X(i)$, we write $U|_\tau$ and $D|_\tau$ to denote the up and down operators respectively on the link (X_τ, Π_τ), and similarly denote the localized upper and lower walks as $N^i_{k-|\tau|}|_\tau$ and $\widehat{N}^i_{k-|\tau|}|_\tau$.

The key observation, often referred to as Garland's method,[5] is that the natural inner product on X 'respects localization.' We will use the following three formalizations of this notion which can be found across a variety of works [309, 239, 111].[6]

Lemma 3.2.15 (Garland's Method [309, 239, 111])**.** *On any simplicial complex (X, Π):*

$$\forall 0 \le i \le k \le d \text{ and } f, g \in C_k : \langle f, g\rangle = \mathop{\mathbb{E}}_{\tau \sim \pi_i}\left[\langle f|_\tau, g|_\tau\rangle\right] \tag{3.2}$$

$$\forall 0 < k \le d \text{ and } f \in C_k : \langle N^1_k f, f\rangle = \mathop{\mathbb{E}}_{v \sim \pi_1}\left[\langle N^1_{k-1}|_v f|_v, f|_v\rangle\right] \tag{3.3}$$

$$\forall 0 \le k < d \text{ and } f \in C_k : \langle M^+_k f, f\rangle = \mathop{\mathbb{E}}_{\tau \sim \pi_{k-1}}\left[\langle A_\tau f|_\tau, f|_\tau\rangle\right] \tag{3.4}$$

The first modern application of Garland's method (albeit applied to a different type

[5] In reference to Garland's original work [163] using similar ideas far before the invent of high dimensional expansion.

[6] We note that some of these equalities hold more generally, but the above is sufficient for our purposes.

of localization) is Oppenheim's Trickling-Down Theorem [309], which shows the spectral behavior of co-dimension 2 links is inherited throughout the rest of the complex.

Theorem 3.2.16 (Trickling-Down [309])**.** *Let (X, Π) be a strongly connected complex. Then for all $0 \leq i \leq d-2$:*

1. $\gamma_i \leq \frac{\gamma_{d-2}}{1-(d-2-i)\gamma_{d-2}}$
2. $\gamma_i^{(-)} \geq \frac{\gamma_{d-2}^{(-)}}{1-(d-2-i)\gamma_{d-2}^{(-)}}$.

Moreover the latter holds without the assumption of strong connectivity.

We will make liberal use of Garland's method and the Trickling-Down Theorem throughout.

3.3 A Booster Theorem for Arbitrary HDX

Our first application of the basic inductive method is a weak KKL-type theorem for *arbitrary* (two-sided) HDX. In particular, we will show that any function with low total influence on a complex whose local-spectral expansion is bounded away from 1 must be *local* (in particular, must have a booster). We first state the result in the contrapositive, that global sets expand.

Theorem 3.3.1 (Global Sets Expand)**.** *Let (X, Π) be a simplicial complex, $0 \leq k \leq d$, and $f \in C_k$ any (ε, i)-global boolean function. Then the expansion of f with respect to the lower walk is at least*

$$\Phi(f) \geq \frac{1-\mathbb{E}[f]}{k-i}\prod_{j=i}^{k-2}(1-\gamma_j) - c_{k,i,\gamma}\varepsilon$$

where $c_{k,i,\gamma} \leq (k-i+1)\left(\frac{1}{k-i}\prod_{j=i}^{k-2}(1-\gamma_j) - \frac{1}{k-i+1}\prod_{j=i-1}^{k-2}(1-\gamma_j)\right)\left(1+(k-i)\gamma_{i-1}^{(-)}\right)^{-1}$

While the constant $c_{k,i,\gamma}$ in Theorem 3.3.1 is somewhat hard to interpret a priori, we emphasize that for many reasonable settings $c_{k,i,\gamma} \leq O(1)$ is an *absolute constant*, with

no dependence on the dimension. This is the case, for instance, if $|\gamma_{i-1}^{(-)}| \ll \frac{1}{k-i}$,[7] or when $\gamma_{k-2} \leq \frac{1}{k-1}$ and $\gamma_{d-2}^{(-)}$ is bounded away from one.

We prove a lower bound matching within a factor of $\frac{(k-i)(i+1)}{k}$ (at most i, and a constant when $i = k - O(1)$).

Theorem 3.3.2. *For every $i \leq k$ and rational $\mu \in [0,1]$, there exists an infinite family of k-dimensional complexes $\{Z_n\}$ and corresponding functions $\{f_n : Z_n(k) \to \mathbb{F}_2\}$ of expectation μ satisfying*

1. *f_n is $(0,i)$-global and (μ,i)-strongly global*
2. $\Phi(f_n) \leq (1-\mathbb{E}[f])\frac{i+1}{k}\prod_{j=i}^{k-2}(1-\gamma_j)$

Theorem 3.3.1 is strongest in low dimensions due to the leading coefficient, but still has the following interesting interpretation in the high dimensional regime: any low-influence function must have a *constant* booster inside some link of reasonably large co-dimension.[8]

Corollary 3.3.3 (Booster Theorem for general HDX)**.** *There is a universal constant $c > 0$ such that for any $i \in \mathbb{N}$, $\frac{1}{i} > \delta > 0$, and k-dimensional complex (X,Π), any subset $S \subset X(k)$ with expansion at most:*

$$\Phi(S) \leq \frac{1-\mathbb{E}[1_S]}{i}\prod_{j=k-i}^{k-2}(1-\gamma_j) - \delta$$

has constant correlation with a link of co-dimension i:

$$\exists \tau \in X(k-i) : \mathop{\mathbb{E}}_{X_\tau}[1_S] \geq \mathbb{E}[1_S] + c'\delta$$

where $c' \leq O(c_{k,k-i,\gamma}^{-1})$.

[7] Note $|\gamma_{i-1}^{(-)}|$ is always at most $\frac{1}{k-i}$ by Trickling-Down, so this is not such a strong assumption.

[8] We thank Yotam Dikstein for pointing out this interpretation.

We remark this is actually not implied by prior expansion theorems for the lower walk, even on stronger HDX, which would instead roughly imply the existence of an $O_\delta(k)$-link with $2^{-O_\delta(k)}$ density.

Finally, GLL observe that a bound on expansion of the lower walk of any simplicial complex X immediately implies a Kruskal-Katona Theorem (see [187, Section 8.3]). Recall that give a set $S \subset X(k)$, the lower shadow of S, denoted ∂S, is the set of all $(k-1)$-faces τ contained in some $s \in S$. We get the following corollary for general simplicial complexes.

Corollary 3.3.4. *Let (X, Π) be a k-dimensional simplicial complex and $S \subset X(k)$ any (ε, i)-global set. Then*

$$\mathbb{E}[\partial S] \geq \mathbb{E}[S]\left(1 + \frac{1-\mathbb{E}[S]}{k-i}\prod_{j=i}^{k-2}(1-\gamma_j) - c_{k,i,\gamma}\varepsilon.\right)$$

where $c_{k,i,\gamma} \leq (k-i+1)\left(\frac{1}{k-i}\prod_{j=i}^{k-2}(1-\gamma_j) - \frac{1}{k-i+1}\prod_{j=i-1}^{k-2}(1-\gamma_j)\right)\left(1+(k-i)\gamma_{i-1}^{(-)}\right)^{-1}$

3.3.1 Garland's Lemma for Expansion

The first key component of Theorem 3.3.1 is a new variant of Garland's Lemma. While the tool is already broadly used in the study of expansion of walks on HDX (see e.g. [111, 38, 39]), it is typically applied *indirectly* throughout the analysis. We rely on the key (albeit basic) insight that Garland's Method also applies directly to expansion itself.

Lemma 3.3.5 (Garland's Lemma for Expansion)**.** *Let (X, Π) be a weighted, pure simplicial complex and $f \in C_k$ any function on k-faces. Then the expansion of f with respect to the lower walk N_k^i can be written as an expectation of the expansion of f restricted to links. That is for all $i \leq k$ and $j \leq k-i$:*

$$\bar{\Phi}_{N_k^i}(f) = \mathop{\mathbb{E}}_{\tau\sim\pi_j^f}[\bar{\Phi}_{N_{k-j}^i|_\tau}(f|_\tau)], \tag{3.5}$$

where π_j^f is the distribution given by:

$$\pi_j^f(\tau) = \pi_j(\tau)\frac{\langle f|_\tau, f|_\tau\rangle}{\langle f, f\rangle}.$$

Proof. Localizing via Equation (3.3), we have:

$$\begin{aligned}
\bar{\Phi}(f) &= \frac{1}{\langle f, f\rangle}\sum_{\tau\in X(j)} \pi_j(\tau)\langle f|_\tau, N_{k-j}^i|_\tau f|_\tau\rangle \\
&= \frac{1}{\langle f, f\rangle}\sum_{\tau\in X(j)} \pi_j(\tau)\langle f|_\tau, f|_\tau\rangle\bar{\Phi}_{N_{k-j}^i|_\tau}(f|_\tau) \\
&= \sum_{\tau\in X(j)} \pi_j^f(\tau)\bar{\Phi}_{N_{k-j}^i|_\tau}(f|_\tau).
\end{aligned}$$

The fact that π_j^f is a distribution follow from noting that $\pi_j^f > 0$ by construction, and by a simple application of Equation (3.2):

$$\begin{aligned}
\sum_{\tau\in X(j)} \pi_j^f(\tau) &= \sum_{\tau\in X(j)} \pi_j(\tau)\frac{\langle f|_\tau, f|_\tau\rangle}{\langle f, f\rangle} \\
&= \frac{1}{\langle f, f\rangle}\mathop{\mathbb{E}}_{\tau\sim\pi_j}[\langle f|_\tau, f|_\tau\rangle] \\
&= 1
\end{aligned}$$

□

We emphasize that this variant of Garland's lemma differs from the typical form in that the distribution over links is *function dependent*. Thus slightly more care must be taken in its application, as typical terms which average out over the standard link distribution may not do so in the function dependent one.

3.3.2 The KO/GK Decomposition

The proof of Theorem 3.3.1 also relies on recent decomposition of Gotlib and Kaufman [176], itself based on the following decomposition of Kaufman and Oppenheim [239].

Definition 3.3.6 (KO-Decomposition [239])**.** Let (X, Π) be a simplicial complex, $0 \leq k \leq d$, and $f \in C_k$. There exists an orthogonal decomposition of f into levels

$$f = \sum_{i=0}^{k} f_i$$

such that $f_i \in Im(U_i^k) \cap Ker(D_{i-1}^k)$.

The KO-basis is not known to be an eigenbasis (even approximately) on weak HDX, but Gotlib and Kaufman [176] recently showed it is still possible to gain some advantage by understanding how a function projects onto its components. We will use a variant of their result for the lower walk.

Theorem 3.3.7. *For any simplicial complex (X, Π), $0 < k \leq d$, and $f \in C_k$, we have:*

$$\langle N_k^1 f, f \rangle \leq \mathbb{E}[f]^2 + \sum_{i=1}^{k-1} \left(1 - \frac{1}{k-i+1} \prod_{j=i-1}^{k-2} (1-\gamma_j)\right) \langle f, f_i \rangle.$$

The proof is similar to [176, Theorem 7.9], which focuses on the non-lazy upper walk instead, and is given in Section 3.7.

3.3.3 A Level-i Inequality and the Upper Bound

With this in mind, a natural goal is now to show that global functions project mostly onto large components, and therefore expand well. In fact, it turns out a variant of this statement is true for *all* simplicial complexes, where the strength of the bound scales with the smallest non-zero eigenvalue of the upper walks.

Theorem 3.3.8 (Level-i Inequality)**.** *Let (X, Π) be a simplicial complex. Then for any $0 < k \leq d$ and (ε, i)-global boolean function $f \in C_k$, the following level-i inequality holds:*

$$\sum_{j=1}^{i} \langle f, f_j \rangle \leq \frac{1}{\lambda_{min}^{\emptyset}(\widehat{N}_i^{k-i})} \varepsilon \mathbb{E}[f].$$

Unfortunately, while Theorem 3.3.8 is a non-trivial statement for any fixed complex, it may not be bounded away from 0 *asymptotically* (i.e. the right-hand side may depend on the number of vertices in the complex). The crucial observation is that this issue can be fixed whenever the complex is locally far from bipartite.

We focus in particular on the special case of $i = 1$, which is the only setting we need for our approach.

Corollary 3.3.9 (Level-1 Inequality)**.** *Let (X, Π) be a simplicial complex, $k \leq d$, and $f \in C_k$ any function on k-faces. Then*

$$\langle f, f_1 \rangle \leq \frac{k}{1 + \gamma_0^{(-)}(k-1)} \varepsilon \mathbb{E}[f] \leq \frac{k}{1 + \gamma_{d-2}^{(-)} \cdot \frac{k-1}{1-(d-2)\gamma_{d-2}^{(-)}}} \varepsilon \mathbb{E}[f]$$

Proof. The second inequality follows from the Trickling-Down Theorem. Then by Theorem 3.3.8, it is enough to lower bound $\lambda_{\min}(D_1^k U_1^k) \leq \lambda_{\min}^{\emptyset}(D_1^k U_1^k)$, i.e. to show that for any $f \in C_k$:

$$\frac{\langle f, D_1^k U_1^k f \rangle}{\langle f, f \rangle} \geq \left(\frac{1}{k} + \gamma_0^{(-)} \frac{k-1}{k} \right).$$

We appeal to the fact that $D_1^k U_1^k = \frac{k-1}{k} M_1^+ + \frac{1}{k} I$ (see e.g. [9, 176]) and observe that for any $f \in C_k$:

$$\begin{aligned} \langle f, D_1^k U_1^k f \rangle &= \frac{1}{k} \langle f, f \rangle + \frac{k-1}{k} \langle f, M_1^+ f \rangle \\ &\geq \left(\frac{1}{k} + \gamma_0^{(-)} \frac{k-1}{k} \right) \langle f, f \rangle \end{aligned}$$

as desired. □

We now prove Theorem 3.3.8 itself. Similar to the approach of [38], we first show a level-i inequality in terms of variance before reducing to the standard setting.

Lemma 3.3.10 (Level-i Inequality (Variance))**.** *Let (X, Π) be a simplicial complex, $k \leq d$, and $f : C_k \to \mathbb{R}$ any function on k-faces. Then for all $i > 0$:*

$$Var(D_i^k f) \geq \lambda_{min}^{\emptyset}(\widehat{N}_i^{k-i}) \sum_{j=1}^{i} \langle f, f_j \rangle$$

Proof. First, observe that by adjointness of the averaging operators we can write:

$$\begin{aligned} \langle D_i^k f, D_i^k f \rangle &= \langle f, U_i^k D_i^k f \rangle \\ &= \sum_{j=0}^{k} \langle f, U_i^k D_i^k f_j \rangle \\ &= \sum_{j=0}^{i} \langle f, U_i^k D_i^k f_j \rangle \end{aligned}$$

since $f_j \in \mathrm{Ker}(D_{j-1}^k)$, and therefore that

$$\mathrm{Var}(D_i^k f) = \sum_{j=1}^{i} \langle f, U_i^k D_i^k f_j \rangle$$

since $\langle f, f_0 \rangle = \mathbb{E}[f]^2 = \mathbb{E}[D_i^k f]^2$.

The trick is now to observe that each f_j is completely orthogonal to the kernel of $U_i^k D_i^k$. If this is the case, we then have

$$\begin{aligned} \mathrm{Var}(D_i^k f) &\geq \lambda_{\min}^{\emptyset}(U_i^k D_i^k) \sum_{j=1}^{i} \langle f, f_j \rangle \\ &= \lambda_{\min}^{\emptyset}(\widehat{N}_i^{k-i}) \sum_{j=1}^{i} \langle f, f_j \rangle \end{aligned}$$

since $U_i^k D_i^k$ and $D_i^k U_i^k$ are PSD and have the same non-zero spectrum.

To see that f_j has no projection onto $\mathrm{Ker}(U_i^k D_i^k)$, observe that $\mathrm{Ker}(U_i^k D_i^k) = \mathrm{Ker}(D_i^k)$ by adjointness of D and U. Namely if $U_i^k D_i^k f = 0$, then $\langle f, U_i^k D_i^k f\rangle = \|D_i^k f\|^2 = 0$ so $f \in Ker(D_i^k)$ (the inclusion in the other direction is trivial). Finally since

$$f_j \in \mathrm{Im}(U_j^k) \supset \mathrm{Im}(U_i^k) = \mathrm{Ker}(D_i^k)^\perp$$

by construction, we are done. □

The proof of Theorem 3.3.8 now follows immediately from the ℓ_∞-to-ℓ_2 reduction introduced in [38].

Proof of Theorem 3.3.8. BHKL observed that any (ε, i)-global boolean function satisfies $\mathrm{Var}(D_i^k f) \leq \varepsilon \mathbb{E}[f]$. Combined with Lemma 3.3.10 this completes the proof. □

We note that this result can be generalized to real-valued functions under some light regularity conditions (and arbitrary non-negative functions) as in [38], but we focus on the boolean case here for simplicity.

We are finally equipped to prove Theorem 3.3.1.

Proof of Theorem 3.3.1. For notational convenience, for $1 \leq \ell < k \leq d$ define

$$\lambda_{k,\ell} \coloneqq 1 - \frac{1}{k-\ell+1} \prod_{j=\ell-1}^{k-2} (1-\gamma_j)$$

and set $\lambda_{k,0} = 1$ and $\lambda_{k,k} = 0$.

We first prove the result for $i = 1$, then show how to reduce to this setting in general via localization. When $i = 1$, we actually show a slightly stronger bound

$$(1 - \mathbb{E}[f])(1 - \lambda_{k,2}) - \frac{(\lambda_{k,1} - \lambda_{k,2})}{\frac{1}{k} + \frac{k-1}{k}\gamma_0^{(-)}}\varepsilon. \tag{3.6}$$

This follows fairly immediately from Theorem 3.3.7 and properties of the $\{\lambda_{k,i}\}$ and $\{f_i\}$ decomposition:

$$\begin{aligned}
\Phi(f) &\geq 1 - \frac{1}{\mathbb{E}[f]} \sum_{\ell=0}^{k} \lambda_{k,\ell} \langle f, f_\ell \rangle \\
&\geq 1 - \frac{1}{\mathbb{E}[f]} \sum_{\ell=0}^{1} \lambda_{k,\ell} \langle f, f_\ell \rangle - \frac{\lambda_{k,2}}{\mathbb{E}[f]} \sum_{\ell=2}^{k} \langle f, f_\ell \rangle && (\lambda_{k,i} \text{ are non-increasing}) \\
&\geq 1 - \frac{1}{\mathbb{E}[f]} \sum_{\ell=0}^{1} \lambda_{k,\ell} \langle f, f_\ell \rangle - \frac{\lambda_{k,2}}{\mathbb{E}[f]} \left(\mathbb{E}[f] - \sum_{\ell=0}^{1} \langle f, f_\ell \rangle \right) && (\text{Booleanity}) \\
&= 1 - \frac{1}{\mathbb{E}[f]} \sum_{\ell=0}^{1} (\lambda_{k,\ell} - \lambda_{k,2}) \langle f, f_\ell \rangle - \lambda_{k,2} \\
&= (1 - \mathbb{E}[f])(1 - \lambda_{k,2}) - \frac{1}{\mathbb{E}[f]} (\lambda_{k,1} - \lambda_{k,2}) \langle f, f_1 \rangle && (\langle f, f_0 \rangle = \mathbb{E}[f]^2) \\
&\geq (1 - \mathbb{E}[f])(1 - \lambda_{k,2}) - \frac{(\lambda_{k,1} - \lambda_{k,2})}{\frac{1}{k} + \frac{k-1}{k} \gamma_0^{(-)}} \varepsilon && (\text{Corollary 3.3.9})
\end{aligned}$$

We observe by the same proof, the above also holds even when $\lambda_{k,1}$ and $\lambda_{k,2}$ are replaced with any values a,b such that $a \geq \lambda_{k,1}$ and $b \geq \lambda_{k,2}$.

The inductive step relies on the observation that global functions are approximately closed under localization. This requires some care to achieve the stated bound. In particular, observe that if f is (ε, i)-global, for any $\tau \in X(i-1)$ we can write

$$\mathbb{E}[f|_\tau] = \mathbb{E}[f] + \delta_\tau$$

where $|\delta_\tau| \leq \varepsilon$, and further that $f|_\tau$ is then $(\varepsilon - \delta_\tau, 1)$-global. With this in mind, applying Garland's method for expansion (Lemma 3.3.5), we have

$$\Phi(f) = \mathop{\mathbb{E}}_{\tau \sim \pi_{i-1}^f} [\Phi_\tau(f|_\tau)].$$

Observe that for all $\tau \in X(i-1)$, the 'local' $\lambda_{i,j}$ values in the link X_τ (denoted $\lambda_{i,j}^\tau$ and defined analogously with respect to X_τ) are upper bounded by their corresponding values

in the original complex, that is $\lambda^{\tau}_{k-i+1,1} \leq \lambda_{k,i}$ and $\lambda^{\tau}_{k-i+1,2} \leq \lambda_{k,i+1}$. Then applying the base case to the above localization, we get

$$\begin{aligned}
\Phi(f) &\geq \mathop{\mathbb{E}}_{\tau\sim\pi^f_{i-1}}\left[(1-\mathbb{E}[f]-\delta_\tau)(1-\lambda_{k,i+1}) - \frac{(\lambda_{k,i}-\lambda_{k,i+1})}{\frac{1}{k-i+1}+\frac{k-i}{k-i+1}\gamma^{(-)}_{i-1}}(\varepsilon-\delta_\tau)\right] \\
&= (1-\mathbb{E}[f])(1-\lambda_{k,i+1}) - \frac{(\lambda_{k,i}-\lambda_{k,i+1})}{\frac{1}{k-i+1}+\frac{k-i}{k-i+1}\gamma^{(-)}_{i-1}}\varepsilon \\
&\quad + \left(\frac{(\lambda_{k,i}-\lambda_{k,i+1})}{\frac{1}{k-i+1}+\frac{k-i}{k-i+1}\gamma^{(-)}_{i-1}} - (1-\lambda_{k,i+1})\right)\mathop{\mathbb{E}}_{\tau\sim\pi^f_{i-1}}[\delta_\tau] \\
&\geq (1-\mathbb{E}[f])(1-\lambda_{k,i+1}) - \frac{(\lambda_{k,i}-\lambda_{k,i+1})}{\frac{1}{k-i+1}+\frac{k-i}{k-i+1}\gamma^{(-)}_{i-1}}\varepsilon
\end{aligned}$$

where in the final inequality we have used the fact that

$$\left(\frac{(\lambda_{k,i}-\lambda_{k,i+1})}{\frac{1}{k-i+1}+\frac{k-i}{k-i+1}\gamma^{(-)}_{i-1}} - (1-\lambda_{k,i+1})\right) \geq 0.$$

This in turn follows from noting that $\gamma^{(-)}_{i-1} \leq 0$, and

$$\begin{aligned}
(k-i+1)(\lambda_{k,i}-\lambda_{k,i+1}) &= (k-i+1)\left(\frac{1}{k-i} - \frac{1}{k-i+1}(1-\gamma_{i-1})\right)\prod_{j=i}^{k-2}(1-\gamma_j) \\
&= \left(\frac{1}{k-i}+\gamma_{i-1}\right)\prod_{j=i}^{k-2}(1-\gamma_j) \\
&\geq \frac{1}{k-i}\prod_{j=i}^{k-2}(1-\gamma_j) \\
&= 1-\lambda_{k,i+1}.
\end{aligned}$$

□

3.3.4 A Lower Bound for Global Expansion

We now move to the proof of our corresponding lower bound. The construction is based on Golowich's [170, 172] 'product-based' HDX, which product a (typically sparse)

low-dimensional complex X with a (typically dense) high dimensional complex Y to get a resulting HDX with roughly $\gamma_j = \frac{1}{j+1}$ local-spectral expansion. We follow the exposition of [172] closely.

Definition 3.3.11 (Product-Complexes [172, Definition 13])**.** Let X and Y be simplicial complexes with measures π_X and π_Y respectively. Let $H_X = \dim(X)$, $H_Y = \dim(Y)$, and assume that $1 \leq H_X \leq H_Y$. Define the high-dimensional product $Z = X \circledh Y$ to be the simplicial complex of dimension $H_Z = H_Y$ with vertex set $Z(0) = X(0) \times Y(0)$ such that

$$Z(H_Z) \coloneqq \{\sigma \subset \sigma_X \times \sigma_Y : \sigma_X \in X(H_X),\ \sigma_Y \in Y(H_Y),\ \Pi_X(\sigma) = \sigma_X,\ \Pi_Y(\sigma) = \sigma_Y\}.$$

Define a weight function π_Z on $Z = X \circledh Y$ so that for any $\sigma \in Z(H_Z)$ with $\Pi_X(\sigma) = \{x_1, \ldots, x_{H_X}\}$ such that for $1 \leq i \leq H_X$, the value x_i occurs with multiplicity $M_i = |(\{x_i\} \times Y) \cap \sigma|$ in σ, then

$$\pi_Z(\sigma) = \frac{\pi_X(\Pi_X(\sigma))\pi_Y(\Pi_Y(\sigma))}{\binom{H_Y - H_X}{M_1 - 1, \ldots, M_{H_X} - 1}},$$

where $\binom{H_Y - H_X}{M_1 - 1, \ldots, M_{H_X} - 1} = \frac{(H_Y - H_X)!}{(M_1 - 1)! \cdots (M_{H_X} - 1)!}$ denotes the multinomial coefficient.

It is important for our purposes that the complexes we build have expansion *exactly* $\gamma_i = \frac{1}{i+1}$. To ensure this, we need to introduce a non-degeneracy condition on the first complex in the product. In particular, we call a k-dimensional complex X non-degenerate if there exists $\sigma \in X(k-1)$ such that $|X_\sigma(1)| > 1$. Golowich [172] proves the following local-spectral bound for non-degenerate products.

Theorem 3.3.12. *Let X and Y be simplicial complexes of dimensions H_X and H_Y respectively, such that X is non-degenerate, $2 \leq H_X < H_Y$ and $\gamma_{(H_Y - 2)}(Y) \leq \frac{1}{H_Z - 1}$. Furthermore, if $\dim(X) \geq 2$, then assume that $\max\{\gamma_{H_X - 2}(X), \gamma^{(-)}_{H_X - 2}\} \leq \frac{1}{H_Z - 1}$. Let $Z = X \circledh Y$ and $H_Z = \dim(Z) = H_Y$. Then for all $H_X - 1 \leq i \leq H_Z - 2$, $\gamma_i(Z) = \frac{1}{i+1}$.*

We are now ready to prove Theorem 3.3.2. Roughly speaking, the idea is to construct an i-global function by producting an $(i+1)$-dimensional complex X with a higher dimensional complete complex. For each i-face τ of X, we'll define our function on the tensor to be 1 when its projection onto X lies on some specified (say) half the $(i+1)$-faces containing τ. This forces the function to be perfectly global at level i, but on the other hand results in 'poor' expansion since the high dimension face must 'escape' from its current X-component to leave the set. We formalize the idea below.

Proof of Theorem 3.3.2. For notational simplicity, we will work with $(i-1)$-global functions instead of i-global functions. Let $\mu = a/b$ for $a, b \in \mathbb{N}$ and take the complex family $\{Z_n\}$ to be the products $X \textcircled{h} Y$ where $X = \Delta_{nb/\mu}(i)$ and $Y = \Delta_{2k}(k)$. Let X_f be any sub-complex of X such that for every $\tau \in X(i-1)$, exactly a μ fraction of faces $\sigma \supset \tau$ are in X_f. We define $f_n : Z_n(k) \to \mathbb{F}_2$ to be the indicator of X_f on the projection of Z to its X component:

$$f_n(\sigma) = \begin{cases} 1 & \text{if } \Pi_X(\sigma) \in X_f \\ 0 & \text{else.} \end{cases}$$

We first argue that f_n is $(0, i-1)$-global and $(\mu, i-1)$-strongly global. Since a the X-projection of a random k-face $\sigma \in Z_n(k)$ is uniformly distributed on $X(i)$, the expectation of f_n is μ. Let $\tau \in Z(i-1)$. If the projection of $\Pi_X(\tau) \in X(i-1)$ (that is every X-component is unique), then the projection of a random $\sigma \in Z_\tau$ is a uniformly i-face containing $\Pi_X(\tau)$. By definition, a μ fraction of these are in X_f. A similar argument holds for general τ, though the distribution is now uniform over faces whose projections which contain $\Pi_X(\tau)$. One can check this still has expectation μ.

It is left to bound the expansion of f_n. Consider the random walk starting from a face $\sigma \in f_n$. The only way for the walk to *leave* f_n is for it to switch projections onto X. This can only happen if the down-step samples a 'singleton', that is a vertex $(a, i) \in \sigma$ such that no other vertex has a as its X-coordinate. In particular, if σ has j singletons, the

probability of escaping the set during the down-up walk is exactly $\frac{j}{k}(1-\mu)$, the probability of selecting a singleton in the down-step, and an i-face from X_f in the up-step. Thus the total expansion is

$$\Phi(f_n) \leq (1-\mu)\sum_{j=0}^{i-1}\frac{j}{k}\mathop{\mathbb{P}}_{\sigma\in Z}[\sigma \text{ has } j \text{ singletons}]$$

The key is now to observe that each possible multiset of X projections occurs in Z with *equal probability*, due to the weighting scheme imposed on Z. This means we can compute the probability of j singletons simply by counting what fraction of multisets have j singletons. This boils down to a 'stars-and-bars' type computation and gives the bound

$$\mathbb{P}[\sigma \text{ has } j \text{ singletons}] = \frac{\binom{i}{j}\binom{k-i-1}{i-j-1}}{\binom{k-1}{i-1}}.$$

Combining these, a second combinatorial calculation gives the simplified bound

$$\Phi(f_n) \leq (1-\mu)\sum_{j=0}^{i-1}\frac{j}{k}\frac{\binom{i}{j}\binom{k-i-1}{i-j-1}}{\binom{k-1}{i-1}} \leq (1-\mu)\frac{(i-1)i}{(k-1)k}.$$

Finally, by Theorem 3.3.12, the local spectral expansion of Z implies

$$\prod_{j=i-1}^{k-2}(1-\gamma_j) = \prod_{j=i-1}^{k-2}\left(1-\frac{1}{j+1}\right) = \frac{i-1}{k-1},$$

so altogether we have

$$\Phi(f_n) \leq \frac{i}{k}\prod_{j=i-1}^{k-2}(1-\gamma_j)$$

as claimed. □

We note that while we have picked a specific instantiation of the product complexes, this lower bound will work with any sufficiently regular product complex whose lower-dimensional component has an appropriate degree sub-complex. This holds (approximately)

for any regular complex of large enough degree, simply by taking a random sub-complex.

3.4 A Local-to-Global Small-Set Expansion Theorem

In the previous section, we analyzed the structure of non-expanding functions with respect to the down-up walk. A second key operator of interest in application is the noise operator, where the analogous bounds on e.g. the cube lead to classical 'small-set expansion' theorems. In this section, we show a 'local-to-global' theorem for (global) small-set expansion: any complex whose links satisfy such a notion must itself be a global small-set expander. Later, we'll see this gives as an application the first small-set expansion theorem for (non-partite) Ramanujan complexes [290], a regime in which prior Fourier analytic techniques fail.

3.4.1 Localizing the Noise Operator

The first step in our proof is to generalize Lemma 3.3.5 to the noise operator. To this end, let $\mathcal{B}_n^p$ denote the standard mean p binomial distribution on n trials, and let d_{TV} denote the *Total Variation Distance*, that is the distance measure between distributions D and D' over universe Ω defined by:

$$d_{TV}(D, D') \coloneqq \max_{E \subseteq \Omega} |D(E) - D'(E)| = \frac{1}{2} \sum_{x \in \Omega} |D(x) - D'(x)|$$

We first argue the noise operator can be localized up to an error term that scales with the TV-distance between adjacent binomial distributions.

Proposition 3.4.1 (Garland's Lemma for Noise Operators)**.** *Let (X, Π) be a weighted simplicial complex. Let $\rho \in [0,1]$ and $j, k \in \mathbb{N}$. The expansion of T_ρ can be approximately localized over j-links:*

$$\bar{\Phi}_{T_\rho^k}(f) = \mathbb{E}_{\tau \sim \pi_j^f}[\bar{\Phi}_{T_\rho^{k-j}|_\tau}(f|_\tau)] \pm 2 d_{TV}(\mathcal{B}_k^\rho, \mathcal{B}_{k-j}^\rho).$$

Proof. Recall that the noise operator is given by a binomially-distributed convex combination of lower walks:

$$T_\rho^k \coloneqq \sum_{i=0}^{k} \mathcal{B}_k^\rho(i) N_k^{k-i},$$

Naively, one might wish to apply Lemma 3.3.5 directly to move to the localized noise operators, but this has two issues. First, we can only apply Lemma 3.3.5 whenever $i \leq k-j$. Second even accounting for this fact, the binomial coefficients need to be shifted to dimension $k-j$ as well. To account for this, we introduce the *shifted* noise operator:

$$T_\rho^{k,j\downarrow} \coloneqq \sum_{i=0}^{k-j} \mathcal{B}_{k-j}^\rho(i) N_k^{k-i}.$$

By Lemma 3.3.5, the shifted noise operator localizes to the desired form:

$$\begin{aligned}
\bar{\Phi}_{T_\rho^{k,j\downarrow}}(f) &= \sum_{i=0}^{k-j} \mathcal{B}_{k-j}^\rho(i) \bar{\Phi}_{N_k^{k-i}}(f) \\
&= \sum_{i=0}^{k-j} \mathcal{B}_{k-j}^\rho(i) \mathop{\mathbb{E}}_{\tau \sim \pi_j} [\bar{\Phi}_{N_k^{k-i}|_\tau}(f|_\tau)] \\
&= \mathop{\mathbb{E}}_{\tau \sim \pi_j} [\bar{\Phi}_{T_\rho^{k-j}|_\tau}(f|_\tau)]
\end{aligned}$$

where the final step is by linearity of inner products. Thus it is enough to show that

$$|\bar{\Phi}_{T_\rho^k}(f) - \bar{\Phi}_{T_\rho^{k,j\downarrow}}(f)| \leq 2 d_{TV}(\mathcal{B}_k^\rho, \mathcal{B}_{k-j}^\rho).$$

This is immediate from the fact that $\bar{\Phi}_{N_k^i}(f) \in [0,1]$, since by the triangle inequality:

$$\begin{aligned}
|\bar{\Phi}_{T_\rho^k}(f) - \bar{\Phi}_{T_\rho^{k,j\downarrow}}(f)| &\leq \sum_{i=0}^{k} |B_k^\rho(i) - B_{k-j}^\rho(i)| \bar{\Phi}_{N_k^{k-i}}(f) \\
&\leq \sum_{i=0}^{k} |B_k^\rho(i) - B_{k-j}^\rho(i)|
\end{aligned}$$

$$= 2d_{TV}(\mathcal{B}_k^\rho, \mathcal{B}_{k-j}^\rho)$$

as desired. □

For fixed j, the additive error term in Proposition 3.4.1 goes to 0 at a rate of about $\frac{1}{\sqrt{k}}$. While this is sufficient for most use cases, it only recovers small-set expansion for sets of size at least $\text{poly}(k^{-1})$, while one might hope to show such a bound for sets of size up to $\exp(-k)$. We show this is possible if the links are additionally good high dimensional expanders.

Proposition 3.4.2 (Refined Garland's Lemma for Noise Operators on HDX)**.** *Let (X, Π) be a k-dimensional weighted simplicial complex such that every j-link is a γ-two-sided or partite one-sided HDX. Then for any $\rho \in [0, 1 - 1/e]$ and $j \leq d$ The expansion of T_ρ can be approximately localized over j-links:*

$$\bar{\Phi}_{T_\rho^k}(f) \leq \mathbb{E}_{\tau \sim \pi_j^f}[\bar{\Phi}_{T_\rho^{k-j}|_\tau}(f|_\tau)] + F_{\mathcal{B}_k^\rho}(j-1) + c_k\gamma$$

where $c_k \leq 2^{O(k)}$ and $F_{\mathcal{B}_k^\rho}(\cdot)$ denotes the CDF of $\mathcal{B}_k^\rho$.

For constant j and γ sufficiently small, both error terms are inverse exponential in k. Such error is negligible since the noise operator *always* has non-expansion at least inverse exponential in k anyway.

The proof of Proposition 3.4.2 combines the basic localization of Lemma 3.3.5 with the Fourier analytic machinery of [111, 39, 187]. To this end, we first introduce a basic version of the latter.

Theorem 3.4.3 (Fourier Analysis on HDX [111, 39, 187])**.** *Let (X, Π) be a k-dimensional γ-two-sided or partite one-sided HDX. Then there exists a decomposition $f = \sum_{i=0}^{k} f_i$ such that:*

$$\bar{\Phi}_{N_k^{k-i}}(f) = \sum_{\ell=0}^{i} \frac{\binom{k-\ell}{i-\ell}}{\binom{k}{i}} \frac{\langle f_i, f_i \rangle}{\langle f, f \rangle} \pm c_k\gamma$$

where $c_k \leq 2^{O(k)}$. Similarly:

$$\bar{\Phi}_{T_\rho^k}(f) = \sum_{i=0}^{k} \rho^i \frac{\langle f_i, f_i \rangle}{\langle f, f \rangle} \pm c_k \gamma.$$

We note that Theorem 3.4.3 does not appear anywhere in the literature, but follows without too much difficulty from techniques developed in [111, 39, 187]. We give the proof for completeness in Section 3.6. We can now prove Proposition 3.4.2.

Proof of Proposition 3.4.2. At a high level, the proof is similar to Proposition 3.4.1, but we will separate the 'shifting' of T_ρ into two parts. The first, which will introduce our additive error, is to truncate the terms for $i \leq j$ which cannot be localized via Lemma 3.3.5:

$$\begin{aligned}\bar{\Phi}_{T_\rho^k}(f) &= \sum_{i=j}^{k} \mathcal{B}_k^\rho(i) \bar{\Phi}_{N_k^{k-i}}(f) + \sum_{i=0}^{j-1} \mathcal{B}_k^\rho(i) \bar{\Phi}_{N_k^{k-i}}(f) \\ &\leq \sum_{i=j}^{k} \mathcal{B}_k^\rho(i) \bar{\Phi}_{N_k^{k-i}}(f) + F_{\mathcal{B}_k^\rho}(j-1)\end{aligned}$$

once again using the fact that $\bar{\Phi}_{N_k^j} \in [0,1]$. We can apply now Lemma 3.3.5 to the left-hand term directly:

$$\sum_{i=j}^{k} \mathcal{B}_k^\rho(i) \bar{\Phi}_{N_i^{k-i}}(f) = \mathop{\mathbb{E}}_{\tau \sim \pi_j^f} \left[\sum_{i=j}^{k} \mathcal{B}_k^\rho(i) \bar{\Phi}_{N_{k-j|_\tau}^{k-i}}(f|_\tau) \right].$$

The second step is to give a finer-grained analysis of the inner expectation for each $\tau \in X(j)$ (rather than simply shifting $\mathcal{B}_k^\rho$ and bounding by TV-distance). In particular, we'll argue that for every $\tau \in X(j)$:

$$\sum_{i=j}^{k} \mathcal{B}_k^\rho(i) \bar{\Phi}_{N_{k-j|_\tau}^{k-i}}(f|_\tau) \leq \bar{\Phi}_{T_\rho^{k-j}|_\tau}(f|_\tau) + c_k \gamma$$

which completes the proof. To see this, first expand the LHS in terms of its Fourier basis:

$$\begin{aligned}
&\sum_{i=j}^{k} \mathcal{B}_k^\rho(i) \bar{\Phi}_{N_{k-j}^{k-i}|_\tau}(f|_\tau) \\
\leq &\frac{1}{\langle f|_\tau, f|_\tau\rangle} \sum_{i=0}^{k-j} \mathcal{B}_k^\rho(i+j) \sum_{\ell=0}^{i} \frac{\binom{k-j-\ell}{i-\ell}}{\binom{k-j}{i}} \langle (f|_\tau)_\ell, (f|_\tau)_\ell\rangle \\
= &\frac{1}{\langle f|_\tau, f|_\tau\rangle} \sum_{\ell=0}^{k-j} \left(\sum_{i=\ell}^{k-j} \binom{k}{i+j} \frac{\binom{k-j-\ell}{i-\ell}}{\binom{k-j}{i}} \rho^{i+j}(1-\rho)^{k-j-i} \right) \langle (f|_\tau)_\ell, (f|_\tau)_\ell\rangle.
\end{aligned}$$

The key is now to observe that the inner sum can be bounded by the (approximate) eigenvalues of T_ρ, namely that the following binomial inequality holds.

Claim 3.4.4. For all $k \in \mathbb{N}$, $j \leq k$, and $\ell \leq k - j$:

$$\sum_{i=\ell}^{k-j} \binom{k}{i+j} \frac{\binom{k-j-\ell}{i-\ell}}{\binom{k-j}{i}} \rho^{i+j}(1-\rho)^{k-j-i} \leq \rho^\ell.$$

Then re-applying Theorem 3.4.3 we'd have:

$$\begin{aligned}
\sum_{i=j}^{k} \mathcal{B}_k^\rho(i) \bar{\Phi}_{N_{k-j}^{k-i}|_\tau}(f|_\tau) &\leq \frac{1}{\langle f|_\tau, f|_\tau\rangle} \sum_{\ell=0}^{k-j} \rho^\ell \langle (f|_\tau)_\ell, (f|_\tau)_\ell\rangle \\
&\leq \bar{\Phi}_{T_\rho^{k-j}|_\tau}(f|_\tau) + c_k\gamma
\end{aligned}$$

as desired.

We leave the proof of Claim 3.4.4, which follows from fairly elementary binomial manipulation, to Section 3.6. □

3.4.2 The Small-Set Expansion Theorem

We now give the general statement of our local-to-global small-set expansion theorem. We call a complex j-locally (global) SSE if its links satisfy an SSE Theorem for strongly global functions.

Definition 3.4.5 (Locally SSE Complexes)**.** Let (X, Π) be a simplicial complex. We say (X, Π) is j-locally ϕ-SSE if for every $i \leq d - j$, $\varepsilon > 0$, j-link X_τ, and $(\varepsilon, i - j)$-strongly global function $h \in C_{k-j}(X_\tau)$ the expansion of h is at least:

$$\Phi_{T_\rho^{d-j}|_\tau}(f|_\tau) \geq \phi(\varepsilon, i)$$

It is an almost immediate corollary of Propositions 3.4.1 and 3.4.2 that global sets on nice complexes expand.

Corollary 3.4.6. *Let (X, Π) be a j-locally ϕ-nice simplicial complex. Then the expansion of any (ε, i)-strongly global function $f \in C_k$ of density at most ε satisfies:*

$$\Phi_{T_\rho^k}(f) \geq \phi(\varepsilon, i) - 2d_{TV}(\mathcal{B}_k^\rho, \mathcal{B}_{k-j}^\rho).$$

Moreover, if the j-links are γ-two-sided or partite one-sided HDX then

$$\Phi_{T_\rho^k}(f) \geq \phi(\varepsilon, i) - F_{\mathcal{B}_k^\rho}(j-1) - c_k\gamma$$

where $c_k \leq 2^{O(k)}$.

Proof. This is essentially immediate from combining Propositions 3.4.1 and 3.4.2 with the observation that any (i, ε)-strongly global function is $(i - j, \varepsilon)$-global upon localization to any j-link. We therefore have:

$$\Phi_{T_\rho^k}(f) \geq \mathop{\mathbb{E}}_{\tau \sim \pi_j^f}[\Phi_{T_\rho^{k-j}|_\tau}(f|_\tau)] - 2d_{TV}(\mathcal{B}_k^\rho, \mathcal{B}_{k-j}^\rho) \geq \phi(\varepsilon, i) - 2d_{TV}(\mathcal{B}_k^\rho, \mathcal{B}_{k-j}^\rho)$$

in the general case and

$$\Phi_{T_\rho^k}(f) \geq \mathop{\mathbb{E}}_{\tau \sim \pi_j^f}[\Phi_{T_\rho^{k-j}|_\tau}(f|_\tau)] - F_{\mathcal{B}_k^\rho}(j-1) - c_k\gamma \geq \phi(\varepsilon, i) - F_{\mathcal{B}_k^\rho}(j-1) - c_k\gamma$$

when the j-links are γ-HDX as desired. $\square$

3.5 Applications

In this section we give applications of our framework to the behavior of non-expanding sets on several families of well-studied complexes. In particular, we give a new characterization of low influence functions on clique and product complexes, and a small-set expansion theorem for the seminal Ramanujan complexes.

3.5.1 Low Influence Functions on Weak HDX

Bafna, Hopkins, Kaufman, and Lovett [38, 39], and independently Gur, Lifshitz, and Liu [187] characterized the structure of non-expanding sets on near-perfect local-spectral expanders. Unfortunately, very few families of HDX are known to satisfy such strong requirements, and those that do are highly algebraic in nature, making them a difficult fit for some potential lines of application such as PCP-type reductions. Applying our framework for weak expansion, we show such characterizations also holds on more 'everyday' objects, albeit in a quantitatively weaker sense. To illustrate this fact, we'll first take a closer look at a classical combinatorial setting: clique-complexes.

Definition 3.5.1 (Clique-complex)**.** Given a graph $G = (V, E)$, the k-dimensional clique-complex $K_{G,k}$ is the simplicial complex induced by the uniform distribution over k-cliques of G.

We show global sets expand on clique-complexes of dense graphs.

Theorem 3.5.2 (Expansion in Clique-complexes)**.** *Fix $k \in \mathbb{N}$ and let $G = (V, E)$ be any graph with minimum degree at least $\Delta_{min} \geq \frac{k-2}{k-1}|V| + \frac{k||A_{\bar{G}}||-1}{k-1}$, where $A_{\bar{G}}$ denotes the un-normalized adjacency matrix of G's complement. Then the expansion of any boolean*

(ε, i)-global function f is at least:

$$\Phi(f) \geq (1 - \mathbb{E}[f]) \frac{i+1}{k(k-i)} - \frac{5}{4}\varepsilon.$$

Proof. It will be convenient to phrase the result in the equivalent setup of k-size independent sets, denoted $I_{G,k}$. In particular, it is enough to prove the result holds over $I_{G,k}$ so long as

$$\Delta_{\max} \leq \frac{|V| - k||A_G|| + 1}{k-1} \tag{3.7}$$

as we can then apply this to the independent set complex of the complement of G (which is exactly $K_{G,k}$).

In their celebrated work on the spectral gap of high order random walks, Alev and Lau [11] showed that $I_{G,k}$ satisfies $\gamma_{k-2} \leq \frac{1}{k}$ as long as[9]

$$\Delta_{\max} \leq \frac{|V| - k|\lambda_{min}(A_G)| + 1}{k-1}.$$

Note that this always holds for any graph satisfying Equation (3.7).

Thus to apply our characterization theorem, it is enough to bound the negative local spectra. We will do this by proving that under Equation (3.7), the co-dimension 2 links of $I_{G,k}$ also satisfy $\gamma_{k-2}^{(-)} \geq -\frac{1}{k}$. Then by Oppenheim's Trickling-Down Theorem we have

$$\gamma_i^{(-)} \geq -\frac{1}{2k-2-i}$$

and the result follows immediately from plugging this bound into Theorem 3.3.1.

Our bound on $\gamma_{k-2}^{(-)}$ follows largely the same strategy as Alev and Lau's original analysis. Let G_S denote the link of a co-dimension 2 face $S \in X(k-2)$, and let $N[S]$ denote the union of S and vertices of G with a neighbor in S. As in Alev-Lau, the idea is

[9]This is not exactly the statement given in [11, Lemma 4.3], but it is immediate from the proof solving for dependence on $\Delta_{\max}$ instead of k.

now to observe that since the top level faces are distributed uniformly, the random walk matrix underlying M_S can be written as

$$M_S = D_S^{-1}(J - I - A_{G[\overline{N[S]}]}),$$

where $A_{G[\overline{N[S]}]}$ is the un-normalized adjacency matrix of the induced graph on $G[\overline{N[S]}]$ and D_S is the standard diagonal degree matrix of G_S. We can then lower bound the smallest eigenvalue of M_S by upper bounding the largest eigenvalue of $-M_S$ as:

$$\begin{aligned}
&\lambda_1(-M_S) \\
&\leq \lambda_1(D_S^{-1/2}(A_{G[\overline{N[S]}]} + I)D_S^{-1/2}) - \lambda_1(D_S^{-1}J) \\
&= \lambda_1(D_S^{-1/2}(A_{G[\overline{N[S]}]} + I)D_S^{-1/2}) - \sum_{v \in G_S} \frac{1}{\deg_{G_S}(v)} \\
&\leq \|D_S^{-1}\|\lambda_1(A_{G[\overline{N[S]}]} + I) - \sum_{v \in G_S} \frac{1}{\deg_{G_S}(v)} && \text{(Variational Characterization)} \\
&= \|D_S^{-1}\|(\lambda_1(A_{G[\overline{N[S]}]}) + 1) - \sum_{v \in G_S} \frac{1}{\deg_{G_S}(v)} \\
&\leq \|D_S^{-1}\|(\lambda_1(A_G) + 1) - \sum_{v \in G_S} \frac{1}{\deg_{G_S}(v)} && \text{(Cauchy's Interlacing Theorem).} \\
&\leq \|D_S^{-1}\|(\lambda_1(A_G) - 1)
\end{aligned}$$

where the last step follows from observing that $\|D_S^{-1}\| = \frac{1}{\text{min-deg}(G_S)}$ and $\sum_{v \in G_S} \frac{1}{\deg_{G_S}(v)} \geq \frac{2}{\text{min-deg}(G_S)}$.

Finally as observed in [11, Lemma 4.3], note that we can bound the degree of any $v \in G_S$ by

$$\deg_{G_S}(v) = |V| - |N[S]| - (\deg_{G[\overline{N[S]}]}(v) + 1) \geq |V| - (\Delta_{\max} + 1)(k - 1).$$

Thus we have

$$\|D_S^{-1}\| \leq \frac{1}{|V| - (\Delta_{\max} + 1)(k-1)},$$

and altogether

$$|\lambda_{\min}(M_S)| \leq \frac{\lambda_1(A_G) - 1}{|V| - (\Delta_{\max} + 1)(k-1)}.$$

Combined with Alev and Lau's bound on $\lambda_2(M_S)$ and our assumption on degree, this gives

$$\max\{\lambda_2(M_S), |\lambda_{\min}(M_S)|\} \leq \frac{\|A_G\| - 1}{|V| - (\Delta_{\max} + 1)(k-1)} \leq \frac{1}{k}$$

as desired. □

Since $\|A_{\bar{G}}\| \leq \Delta_{\max}(\bar{G}) = |V| - \Delta_{\min}(G)$, the simplified version in the introduction follows as an immediate corollary. We now take a look at two corollaries of this bound: first to the structure of non-expanding sets on 3-dimensional clique complexes, then for general k.

Corollary 3.5.3. *Let $G = (V, E)$ be a graph with $\Delta_{min} \geq \frac{5}{6}|V|$. Then any subset $S \subset K_{G,3}(3)$ of the 3-clique complex with expansion at most*

$$\Phi(S) \leq \frac{1}{3} - \delta$$

must contain a constant fraction of the triangles touching some vertex $v \in G$:

$$\exists v \in V : \mathop{\mathbb{E}}_{X_v}[1_S] \geq c\delta.$$

Proof. Assume for the sake of contradiction that f is $(1, \varepsilon)$-global. We will show it must be the case that $\varepsilon \geq \frac{\delta}{2}$. By Theorem 3.5.2 we have that

$$\frac{1}{3} - \delta \geq \Phi(f) \geq \frac{1}{3} - \frac{1}{3}\mathbb{E}[f] - \frac{5}{4}\varepsilon$$

and therefore that

$$\varepsilon \geq \frac{4}{5}\delta - \frac{4}{15}\mathbb{E}[f].$$

We can assume without loss of generality that $\mathbb{E}[f] \leq \frac{1}{2}\delta$ (otherwise the result is trivial by averaging), which in turn gives $\varepsilon \geq \frac{\delta}{2}$ as desired. □

We are not aware of any other method for showing such a characterization on low-dimensional clique complexes. It is further worth noting that while requiring min-degree at least $\frac{5}{6}|V|$ seems very restrictive (and indeed does imply the resulting complex is dense), clique complexes of much lower degree complexes are not even guaranteed to be connected (consider, e.g. two disjoint copies of the complete graph $K_{n/2}$, which has min-degree roughly $\frac{5}{6}|V|$).

For large k, the same proof strategy can be used to show that any non-expanding set has constant density in some $k - O(1)$ dimension link.

Corollary 3.5.4. *There is a universal constant $c > 0$ such that for any $i, k \in \mathbb{N}$, $0 < \delta < 1$, and graph $G = (V, E)$ of minimum degree at least $\Delta_{min} \geq \frac{2k-2}{2k-1}|V|$, any subset $S \subset X_{G,k}(k)$ with expansion at most*

$$\Phi(S) \leq \frac{1}{i}(1 - \delta)$$

has constant correlation with a link of co-dimension i:

$$\exists \tau \in X(k-i) : \mathop{\mathbb{E}}_{X_\tau}[1_S] \geq c\delta.$$

While characterizing non-expanding functions on common dense complexes is interesting in its own right, much of the promise of high dimensional expanders comes from their ability to give sparse models for dense objects such as the complete complex or product spaces. Prior analysis of small set expansion on such objects was limited to *algebraic* constructions of HDX [290, 238, 108]. In part due to their more complicated nature, it has

been difficult to find natural embeddings into these constructions, and there are no known ways to use them for PCPs or reductions within hardness of approximation. Towards this direction, there is a great deal of interest in simpler, combinatorial constructions of high dimensional expanders.

The most recent work in this direction is the product construction of Golowich [170], discussed in Section 3.3.4. As an immediate application of our expansion framework, we prove the following global expansion theorem and characterization of non-expanding sets on product complexes. We state the result for cutoffs for simplicity.

Corollary 3.5.5. *Let $k \geq 3$, G be any graph on n vertices, $X = G \circledcirc \Delta_{4k}(2k)$, $0 < i < k$, and $f \in C_k$ be any (ε, i)-global boolean function. The expansion of f with respect to the lower walk is at least:*

$$\Phi(f) \geq \frac{(1 - \mathbb{E}[f])i}{(k-1)(k-i)} - 4\varepsilon$$

We omit the proof, which is immediate from the fact that the product satisfies $\max\{\gamma_{k-2}, |\gamma_{k-2}^{(-)}\} \leq \frac{1}{k}$ by Theorem 3.3.12 and Theorem 3.2.16. As for the case of clique-complexes, this leads to a novel characterization of non-expanding sets in low dimensions. For concreteness, we again look at the setting of $k = 3$.

Corollary 3.5.6. *Let G be any graph on n vertices, and $X = G \circledcirc \Delta_{12}(6)$. There exists a universal constant $c > 0$ such that any $S \subset X(3)$ with expansion at most*

$$\Phi(S) \leq \frac{1}{4} - \delta$$

must contain a constant fraction of the triangles touching some vertex $v \in X(1)$:

$$\exists v \in X(1) : \mathop{\mathbb{E}}_{X_v}[1_S] \geq c\delta.$$

3.5.2 Small-Set Expansion of the Ramanujan Complexes

The Ramanujan complexes are the seminal construction of high dimensional expanders [290], but are *one-sided* expanders and may be non-partite, failing the strong conditions required by prior work. Nevertheless, the Ramanujan complexes come with a variety of other substantial benefits not necessarily enjoyed by other constructions, leading to applications ranging from property and agreement testing [236, 233] to quantum codes [142, 244] and Sum-of-Squares lower bounds [119]. It is natural therefore to ask whether these complexes also satisfy an analogous small-set expansion theorem to other constructions with stronger spectral guarantees.

We refer the reader to [290] for the rather involved details of these constructions, and state here only the salient points for our applications.

Theorem 3.5.7 (The LSV-Complexes [290])**.** *For every $\gamma > 0$ and $d \in \mathbb{N}$, there exists an infinite family of complexes $\{(X, \Pi)_n\}$ satisfying:*

1. *$(X, \Pi)_n$ is a d-dimensional, **non-partite** γ-one-sided HDX on n vertices*
2. *For all $v \in X(1)$: $(X_v, \Pi_v)_n$ is a **partite** γ-one-sided HDX on $O_{d,\gamma}(1)$ vertices.*

Since the Ramanujan complexes are one-sided and non-partite, they do not admit any (known) theory of Fourier analysis, and as a result the tools and results developed in [111, 38, 39, 187] do not apply. However, since their *links* are partite HDX, we can appeal to the following result of [208] to show the complexes are 1-locally nice.

Theorem 3.5.8 (SSE for Partite HDX [187])**.** *Let (X, Π) be a d-dimensional partite γ-one-sided local-spectral expander. If $f \in C_d$ is (ε, i)-global, then the expansion of f with respect to T_ρ is at least:*

$$\Phi_{T_\rho}(f) \geq 1 - \rho^{i+1} - c_i \varepsilon^{1/2} - c_k \gamma$$

where each $c_i \leq i^{O(i)}$.

We give the proof in Section 3.6 for completeness, which is essentially immediate from [187] and earlier work of Dikstein and Dinur [109]. As an immediate corollary, we get the first KKL-type theorem for any non-partite one-sided HDX.

Corollary 3.5.9. *Let $(X, \Pi)_n$ be a Ramanujan complex and $f \in C_k$ a (ε, i)-global set of density at most ε. Then the expansion of f is at least:*

$$\Phi_{T_\rho}(f) \geq 1 - \rho^i - c_i \varepsilon^{1/2} - c_k \gamma - \rho^k$$

where each $c_i \leq i^{O(i)}$.

Proof. By Theorem 3.5.8, the Ramanujan complexes $(X, \Pi)_n$ are 1-locally ϕ-nice for

$$\phi(\varepsilon, i) = 1 - \rho^{i+1} - c_i \varepsilon^{1/2} - c_k \gamma$$

The result is now immediate from Corollary 3.4.6. □

Equivalently, we can view this result via its contrapositive: any non-expanding set is local.

Corollary 3.5.10. *For any $\rho \in [0, 1]$, $k > \Omega(\rho^{-1})$, LSV-complex $(X, \Pi)_n$ with $\gamma \leq k^{-\Omega(k)}$ and any $\delta > 0$, there exist constants $r = r(\delta, \rho)$ and $s = s(\delta, \rho)$ such that for any boolean $f \in C_k$ with expansion $\bar{\Phi}_{T_\rho}(f) > \delta$, f has density at least s in some r-link:*

$$\exists \tau \in X(r) : \mathop{\mathbb{E}}_{X_\tau}[f|_\tau] \geq s$$

Moreover, we can take $r \leq O(\frac{\log \delta^{-1}}{\log \rho^{-1}})$, and $s \geq r^{-O(r)}$.

Acknowledgements

We thank Yotam Dikstein, Louis Golowich, Tali Kaufman, Shachar Lovett, and Anthony Ostuni for discussion and comments on an earlier draft of this work.

3.6 Fourier Analysis on HDX

Most of the Fourier analytic results needed for Proposition 3.4.2 follow without too much difficulty from existing techniques. However since the results are not explicitly stated in any work (and often require pulling from several sources), we give a contained version here. Our main goal is to prove the following result:

Theorem 3.6.1 (Theorem 3.4.3 Restated)**.** *Let (X, Π) be a k-dimensional γ-two-sided or partite one-sided HDX. Then there exists a decomposition $f = \sum_{i=0}^{k} f_i$ such that:*

$$\bar{\Phi}_{N_k^{k-i}}(f) = \sum_{\ell=0}^{i} \frac{\binom{k-\ell}{i-\ell}}{\binom{k}{i}} \frac{\langle f_i, f_i \rangle}{\langle f, f \rangle} \pm c_k \gamma$$

where $c_k \leq k^{O(k)}$. Similarly:

$$\bar{\Phi}_{T_\rho^k}(f) = \sum_{i=0}^{k} \rho^i \frac{\langle f_i, f_i \rangle}{\langle f, f \rangle} \pm c_k \gamma$$

The two-sided variant is actually proven explicitly in [38, 39], so we focus on the partite one-sided case which is studied implicitly in [187]. More formally, the authors study a notion of expansion on partite complexes they call 'γ-products.' To define these, we need to introduce a new type of random walk between colors on a partite complex, originally introduced by Dikstein and Dinur [109] in the context of agreement testing.

Definition 3.6.2 (Colored Swap-Walks [109])**.** Given a partite simplicial complex (X, Π) and $i, j \in [d]$, the colored swap-walk $M^{i,j}$ walks from X^i to X^j and has bipartite adjacency operator:

$$M^{i,j}(v, w) = \frac{\pi_{w,1}(v)}{\sum_{z \in X^i} \pi_{w,1}(z)}.$$

On a product space, every $M^{i,j}$ is a complete bipartite graph (since colors are completely independent). Gur, Lifshitz, and Liu introduced the notion of a γ-product to

relax this constraint to simply requiring these operators expand.

Definition 3.6.3 (γ-Product [187])**.** A partite complex (X, Π) is a γ-product if every link X_τ of co-dimension at least 2, every colored swap-walk is a γ-one-sided spectral expander:

$$\forall i, j : \lambda_2(M_\tau^{i,j}) \leq \gamma,$$

It is not hard to show that γ-products are *equivalent* to one-sided partite HDX up to a factor in dimension. Indeed the 'hard' direction (HDX $\to$ γ-product) was already shown in prior work of Dikstein and Dinur [109], while the 'easy' direction follows from Oppenheim's Trickling-Down Theorem.

Theorem 3.6.4 (γ-Product $\iff$ Partite HDX)**.** *Let (X, Π) be a simplicial complex and $\gamma < \frac{1}{d-1}$. Then:*

1. *If (X, Π) is a γ-one-sided HDX, then it is a $\frac{\gamma}{1-(d-1)\gamma}$-product*
2. *If (X, Π) is a γ-product, it is a $\frac{\gamma}{1-(d-1)\gamma}$-one-sided HDX*

Proof. The first result is [109, Corollary 7.6], and follows from a partite variant of the Trickling-Down Theorem. To prove the second, observe that

1. For $\tau \in X(d-2)$: $A_\tau = M_\tau^{1,2}$,
2. For any $i \leq d-2$ and $\tau \in X(i)$: A_τ is connected.[10]

Since we are promised $\lambda_2(M_\tau^{1,2}) \leq \gamma < \frac{1}{d-1}$ and all links are connected, Trickling-Down (Theorem 3.2.16) implies the result. □

GLL show that γ-products, and therefore partite one-sided HDX, have an approximate Fourier basis. Their result is based off a classical basis for product spaces called

[10] if A_τ is disconnected, there exist disconnected vertices v_i and v_j of colors i and j. By construction, v_i and v_j are then also disconnected in $M_\tau^{i,j}$, which violates the expansion assumption.

the *Efron-Stein Decomposition*. To state this, it will first be useful to define a family of partite averaging operators

Definition 3.6.5 (Partite Averaging Operators)**.** Given a partite simplicial complex (X, Π), the partite averaging operators $\{E_T\}_{T \subseteq [d]} : C(d) \to C(d)$ average over elements in $[n] \setminus T$:

$$E_T f(x) = \mathop{\mathbb{E}}_{X_{x_T}} [f|_{x_T}],$$

i.e. $E_T f$ is the conditional expectation of f given that coordinates T are fixed to x_T.

The Efron-Stein decomposition divides f into components $f^{=S}$ that make up the contribution to f from each coordinate set, which essentially amounts to applying inclusion-exclusion to $E_S f$.

Definition 3.6.6. Given a partite simplicial complex (X, Π) and f, define

$$f^{=S} = \sum_{T \subseteq S} (-1)^{|S \setminus T|} E_T f$$

We denote the sum of components at a given level as $f_i \coloneqq \sum_{|S|=i} f^{=S}$

It is not hard to check that $\{f^{=S}\}$ decompose f, that is.

$$f = \sum_{S \subseteq [d]} f^{=S}$$

GLL proved that the Efron-Stein decomposition is an approximate Fourier basis for the partite averaging operators in the following sense.

Theorem 3.6.7 (Efron-Stein Decomposition on HDX)**.** *Let (X, Π) be a partite γ-one-sided HDX. Then the Efron-Stein Decomposition is approximately orthogonal:*

$$\forall S, T : \langle f^{=S}, f^{=T} \rangle \leq 2^{O(|S|+|T|)} \gamma \|f\|_2^2,$$

and approximately an eigenbasis for every E_T:

1. *if $S \subseteq T : E_T f^{=S} = f^{=S}$*
2. *If $S \nsubseteq T : \|E_T f^{=S}\|_2 \leq \sqrt{|S||T|}2^{|S|}\gamma\|f\|$*

high order random walks on partite complexes can typically be written as convex combinations of the partite averaging operators. With this in mind, the proof of Theorem 3.4.3 follows easily from expanding f into the Efron-Stein basis.

Proof of Theorem 3.4.3. Observe that the lower walks can be written as the following convex combination of the colored averaging operators:

$$N_k^{k-i} = \frac{1}{\binom{k}{i}} \sum_{|T|=i} E_T$$

It then follows from Theorem 3.6.7 that Efron-Stein is also an approximate eigenbasis for N_k^i:

$$\begin{aligned} N_k^{k-i} f^{=S} &= \frac{1}{\binom{k}{i}} \sum_{|T|=i} E_T f^{=S} \\ &= \frac{1}{\binom{k}{i}} \sum_{\substack{|T|=i \\ T \supseteq S}} E_T f^{=S} + \frac{1}{\binom{k}{i}} \sum_{\substack{|T|=i \\ T \not\supseteq S}} E_T f^{=S} \\ &= \frac{\binom{k-|S|}{i-|S|}}{\binom{k}{i}} f^{=S} + e\vec{r}r \end{aligned}$$

where $\|e\vec{r}r\|_2 \leq 2^{O(k)}\gamma$. Note that the lefthand term is really 0 for $|S| > i$ (which we've denoted as a negative binomial coefficient for convenience).

The result for N_k^{k-i} then follows from expanding out $\Phi(f)$ in this basis and applying approximate orthogonality:

$$\Phi(f) = 1 - \frac{1}{\langle f, f\rangle}\langle f, \widehat{N}_k^i f\rangle$$

$$
\begin{aligned}
&= 1 - \frac{1}{\langle f, f\rangle} \sum_{\ell=0}^{k} \langle f, \widehat{N}_k^i f_\ell \rangle \\
&= 1 - \frac{1}{\langle f, f\rangle} \sum_{\ell=0}^{i} \frac{\binom{k-|S|}{i-|S|}}{\binom{k}{i}} \langle f, f_\ell \rangle + c_1 \gamma \\
&= 1 - \frac{1}{\langle f, f\rangle} \sum_{\ell=0}^{i} \frac{\binom{k-|S|}{i-|S|}}{\binom{k}{i}} \langle f, f_\ell \rangle + c_2 \gamma
\end{aligned}
$$

where $c_1, c_2 \leq 2^{O(k)}$. The result for T_ρ follows similarly, in fact as GLL ([187, Claim 8.3]) note the expansion decomposition in this case is exact:

$$
\Phi_{T_\rho}(f) = \frac{1}{\langle f, f\rangle} \sum_{i=0}^{k} \rho^i \langle f, f_i, \rangle
$$

so approximate orthogonality again implies the desired result. □

3.7 Gotlib-Kaufman for the Lower Walk

Theorem 3.7.1. *For any simplicial complex (X, Π), $0 < k \leq d$, and $f \in C_k$, we have:*

$$
\langle N_k^1 f, f\rangle \leq \mathbb{E}[f]^2 + \sum_{i=1}^{k-1} \left(1 - \frac{1}{k-i+1} \prod_{j=i-1}^{k-2} (1-\gamma_j)\right) \langle f, f_i\rangle.
$$

Following GK [176], the proof revolves around reducing to level-1 by (inductively) localizing the decomposition. The crucial observation is that in each step, it is enough to analyze the contribution coming from the constant part of level 1, which is handled by the following lemma.

Lemma 3.7.2 ([176, Lemma 7.8, Theorem 7.9])**.** *Let (X, Π) be a simplicial complex and $g \in C_k$ any function orthogonal to $\vec{1}$. Then the parallel part of g under restriction is small in expectation:*

$$
\mathop{\mathbb{E}}_{v\sim\pi_1} \left[\|(N|_v)_{k-1}^{k-1} g|_v\|_2^2\right] \leq \left(1 - \frac{k-1}{k}(1-\gamma_0)\right) \|g\|_2^2
$$

Proof. First, observe that since $(D|_v)_0^{k-1} g|_v$ is constant, we have:

$$\|(N|_v)_{k-1}^{k-1} g|_v\|_2^2 = D_1^k g(v)^2$$

and therefore that

$$\mathop{\mathbb{E}}_{v \sim \pi_1} \left[\|(N|_v)_{k-1}^{k-1} g|_v\|_2^2 \right] = \langle D_1^k g, D_1^k g \rangle = \langle N_k^{k-1} g, g \rangle$$

by adjointness of D and U. Since g is orthogonal to $\vec{1}$, it is enough to upper bound $\lambda_2(N_k^{k-1})$. It will be convenient to instead work with the upper walk $\widehat{N}_1^{k-1}$, which has the same non-zero spectrum.

In particular, it is well known that the non-lazy component of $\widehat{N}_1^{k-1}$ is simply M_1^+ (see e.g. [9, 176]), so we can write:

$$\lambda_2(\widehat{N}_1^{k-1}) = \lambda_2 \left(\frac{1}{k} I + \frac{k-1}{k} M_1^+ \right) \leq \frac{1}{k} + \frac{k-1}{k} \gamma_0 = 1 - \frac{k-1}{k}(1 - \gamma_0)$$

as desired. □

The proof of the main theorem then follows from combining this fact with a basic induction on k, performed by successive restrictions.

Proof of Theorem 3.7.1. The proof follows largely the same strategy as [176, Theorems 5.8,7.9] where the only real difference lies in the base case. The proof is by induction on k. Note that since $f_0 = \mathbb{E}[f]\vec{1}$, and the remaining f_i are orthogonal, it is enough to consider the case where $f_0 = 0$.

Because the KO-decomposition does not behave well under restriction, as in [176] we will instead prove the more general statement for any sum $\sum_{i=1}^{k} f_i$ satisfying the weaker

constraint $f_i \in Ker(D^k_{i-1})$:

$$\left\langle N^1_k \sum_{i=1}^{k} f_i, \sum_{i=1}^{k} f_i \right\rangle \leq \sum_{i=1}^{k-1} \left(1 - \frac{1}{k-i+1} \prod_{j=i-1}^{k-2} (1-\gamma_j)\right) \langle f_i, f_i \rangle + \sum_{i \neq j} c_{i,j} \langle f_i, f_j \rangle \quad (3.8)$$

Since the KO-decomposition is orthogonal and satisfies $f_i \in Ker(D^k_{i-1})$, this is sufficient to prove the result. For notational convenience, we continue to write $f = \sum_{i=1}^{k} f_i$ throughout. With this in mind, we start with the base case of graphs. Notice that by adjointness of U and D we have:

$$\langle N^1_2 f, f \rangle = \langle U_1 D_2 (f_1 + f_2), f_1 + f_2 \rangle = \langle D_2 f_1, D_2 f_1 \rangle$$

since $f_2 \in \mathrm{Ker}(D_2)$ by definition. The result now follows from [176, Lemma 7.8], who shows that

$$\langle D_2 f_1, D_2 f_1 \rangle \leq \left(1 - \frac{1}{2}(1-\gamma_0)\right) \|f_1\|^2_2$$

as desired (this also follows from the proof of Lemma 3.7.2 above).

The inductive step follows essentially the same argument as Gotlib-Kaufman (replacing relevant parameters throughout), but we include the proof for completeness. Because restriction is linear and respected by the lower walk, we have

$$\begin{aligned} \langle N^1_k f, f \rangle &= \left\langle N^1_k \sum_{i=1}^{k} f_i, \sum_{i=1}^{k} f_i \right\rangle \\ &= \mathop{\mathbb{E}}_{v \sim \pi_1} \left[\left\langle (N|_v)^1_{k-1} \sum_{i=1}^{k} f_i|_v, \sum_{i=1}^{k} f_i|_v \right\rangle \right]. \end{aligned}$$

Now that we have reduced to dimension $k-1$ we'd hope to apply the inductive hypothesis, but this requires a sum of the form $\sum_{i=1}^{k-1} h_i$ for $h_i \in \mathrm{Ker}((D|_v)^{k-1}_{i-1})$. Gotlib and Kaufman observed that the only part of the previous expression that fails to achieve this form is the

constant part of $f|_v$. In particular, letting

$$\begin{aligned} h_i^v &\coloneqq (f_{i+1})|_v && (1 < i < k) \\ h_i^v &\coloneqq (f_2)|_v + \left(1 - (N|_v)_{k-1}^{k-1}\right)(f_1)|_v && (i = 1), \end{aligned}$$

we have $h_i^v \in \mathrm{Ker}((D|_v)_{i-1}^{k-1})$ and can write:

$$\begin{aligned} \langle N_k^1 f, f\rangle = &\mathop{\mathbb{E}}_{v\sim\pi_1}\left[\left\langle (N|_v)_{k-1}^1 \sum_{i=1}^{k-1} h_i^v, \sum_{i=1}^{k-1} h_i^v \right\rangle\right] \\ &+ \mathop{\mathbb{E}}_{v\sim\pi_1}\left[\left\langle (N|_v)_{k-1}^1 (N|_v)_{k-1}^{k-1}(f_1)|_v, (N|_v)_{k-1}^{k-1}(f_1)|_v\right\rangle\right]. \end{aligned}$$

The first term can now be analyzed by the inductive hypothesis:

$$\begin{aligned} \left\langle (N|_v)_{k-1}^1 \sum_{i=1}^{k-1} h_i^v, \sum_{i=1}^{k-1} h_i^v \right\rangle &\le \sum_{i=1}^{k-2}\left(1 - \frac{1}{k-i}\prod_{j=i}^{k-2}(1-\gamma_j)\right)\langle h_i^v, h_i^v\rangle + \sum_{i\neq j} c_{i,j}\langle h_i^v, h_j^v\rangle \\ &= \sum_{i=2}^{k-1}\left(1 - \frac{1}{k-i+1}\prod_{j=i-1}^{k-2}(1-\gamma_j)\right)\langle (f_i)|_v, (f_i)|_v\rangle \\ &+ \left(1 - \frac{1}{k-1}\prod_{j=1}^{k-2}(1-\gamma_j)\right)\left\|\left(1-(N|_v)_{k-1}^{k-1}\right)(f_1)|_v\right\|_2^2 \\ &+ \sum_{i\neq j} c'_{i,j}\langle (f_i)|_v, (f_j)_v\rangle \end{aligned}$$

where we have re-applied the definition of h_i^v and collected mixed terms as in [176, Lemma 5.10]. Recall that by Garland's method for any two functions g, g' we have

$$\mathop{\mathbb{E}}_{v\sim\pi_1}\left[\langle g|_v, g'|_v\rangle\right] = \langle g, g'\rangle,$$

and thus altogether that

$$\langle N_k^1 f, f\rangle \le \sum_{i=2}^{k-1}\left(1 - \frac{1}{k-i+1}\prod_{j=i-1}^{k-2}(1-\gamma_j)\right)\langle f_i, f_i\rangle$$

$$+ \mathop{\mathbb{E}}_{v\sim\pi_1}\left[\left(1-\frac{1}{k-1}\prod_{j=1}^{k-2}(1-\gamma_j)\right)\|\left(1-(N|_v)_{k-1}^{k-1}\right)(f_1)|_v\|_2^2+\|(N|_v)_{k-1}^{k-1}(f_1)|_v\|_2^2\right]$$
$$+\sum_{i\neq j}c'_{i,j}\langle f_i,f_j\rangle$$

where we have used the fact that $N_k^1(N|_v)_{k-1}^{k-1}(f_1)|_v = (N|_v)_{k-1}^{k-1}(f_1)|_v$ since the latter is constant. For notational simplicity let

$$\lambda_{i,k} \coloneqq \left(1-\frac{1}{k-i+1}\prod_{j=i-1}^{k-2}(1-\gamma_j)\right).$$

It is therefore left to prove the following claim:

Claim 3.7.3 ([176, Theorems 5.8,7.9]).

$$\mathop{\mathbb{E}}_{v\sim\pi_1}\left[\lambda_{k,2}\|\left(1-(N|_v)_{k-1}^{k-1}\right)(f_1)|_v\|_2^2+\|(N|_v)_{k-1}^{k-1}(f_1)|_v\|_2^2\right]\leq\lambda_{k,1}\|f_1\|_2^2 \tag{3.9}$$

This follows from combining several more general claims in Gotlib-Kaufman, but we'll repeat the direct version here for completeness. GK observe the lefthand side can be re-written as

$$\mathop{\mathbb{E}}_{v\sim\pi_1}\left[\lambda_{k,2}\|\left(1-(N|_v)_{k-1}^{k-1}\right)(f_1)|_v\|_2^2+\lambda_{k,2}\|(N|_v)_{k-1}^{k-1}(f_1)|_v\|_2^2+(1-\lambda_{k,2})\|(N|_v)_{k-1}^{k-1}(f_1)|_v\|_2^2\right]$$
$$=\lambda_{k,2}\mathop{\mathbb{E}}_{v\sim\pi_1}\left[\|(f_1)|_v\|_2^2\right]+(1-\lambda_{k,2})\mathop{\mathbb{E}}_{v\sim\pi_1}\left[\|(N|_v)_{k-1}^{k-1}(f_1)|_v\|_2^2\right]$$
$$=\lambda_{k,2}\|f_1\|_2^2+(1-\lambda_{k,2})\mathop{\mathbb{E}}_{v\sim\pi_1}\left[\|(N|_v)_{k-1}^{k-1}(f_1)|_v\|_2^2\right]$$

With this in mind, we are finally ready to apply Lemma 3.7.2

$$\mathop{\mathbb{E}}_{v\sim\pi_1}\left[\|(N|_v)_{k-1}^{k-1}(f_1)|_v\|_2^2\right]\leq\left(1-\frac{k-1}{k}(1-\gamma_0)\right)\|f_1\|_2^2$$

which completes the proof since

$$\begin{aligned}\lambda_{k,2} + (1-\lambda_{k,2})\left(1 - \frac{k}{k+1}(1-\gamma_0)\right) &= 1 - (1-\lambda_{k,2})\left(1 - \frac{k}{k+1}(1-\gamma_0)\right) \\ &= 1 - \frac{1}{k}\prod_{j=0}^{k-2}(1-\gamma_j) \\ &= \lambda_{k,1}\end{aligned}$$

as desired. □

3.8 Garland's Lemma

We prove the lower-walk variant of Garland's Lemma from Lemma 3.2.15.

Lemma 3.8.1 (Garland's Lemma for Lower Walks)**.** *Let (X, Π) be a weighted d-dimensional simplicial complex. Then for any $i < k < d$, $f \in C_k$, $j \leq k - i$ and $\tau \in X(j)$:*

$$\langle f, N_k^i f\rangle = \mathbb{E}_{\tau\sim\pi_j}\left[\langle f|_\tau, N_{k-j}^i|_\tau f|_\tau\rangle\right]$$

where $f|_\tau : X_\tau(k-i) \to \mathbb{R}$ is given by $f_\tau(\sigma) = f(\tau \cup \sigma)$, and $N_{k-j}^i|_\tau$ is the lower walk on the link X_τ.

Proof. Since D and U are adjoint, we can write:

$$\begin{aligned}\mathbb{E}_{\tau\sim\pi_j}\left[\langle f|_\tau, N_{k-j}^i|_\tau f|_\tau\rangle\right] &= \mathbb{E}_{\tau\sim\pi_j}\left[\langle D_{k-i-j}^{k-j}|_\tau f|_\tau, D_{k-i-j}^{k-j}|_\tau f|_\tau\rangle\right] \\ &= \sum_{\tau\in X(j)}\pi_j(\tau)\sum_{\sigma\in X_\tau(k-i-j)}\pi_{\tau,k-i-j}(\sigma)(D_{k-i-j}^{k-j}|_\tau f|_\tau(\sigma))^2 \\ &= \frac{1}{\binom{k-i}{j}}\sum_{\tau\in X(j)}\sum_{\sigma\in X_\tau(k-i-j)}\pi_{k-i}(\tau\cup\sigma)(D_{k-i-j}^{k-j}|_\tau f|_\tau(\sigma))^2.\end{aligned}$$

The key is now to observe that $D_{k-i-j}^{k-j}|_\tau f|_\tau(\sigma)$, the expected value of $f|_\tau$ over the link of σ in X_τ is equivalent to the expectation of $f|_{\tau\cup\sigma}$ over the link $X_{\tau\cup\sigma}$, which is just

$D^k_{k-i} f(\tau \cup \sigma)$. Therefore we have:

$$\begin{aligned}
&\frac{1}{\binom{k-i}{j}} \sum_{\tau \in X(j)} \sum_{\sigma \in X_\tau(k-i-j)} \pi_{k-i}(\tau \cup \sigma)(D^{k-j}_{k-i-j}|_\tau f|_\tau(\sigma))^2 \\
=&\frac{1}{\binom{k-i}{j}} \sum_{\tau \in X(j)} \sum_{\sigma \in X_\tau(k-i-j)} \pi_{k-i}(\tau \cup \sigma)(D^{k}_{k-i} f(\tau \cup \sigma))^2 \\
=& \sum_{T \in X(k-i)} \pi_{k-i}(T)(D^k_{k-i} f(T))^2 \\
=&\langle f, N^i_k f \rangle,
\end{aligned}$$

where the second to last step follows from noting that the inner term depends only on the union $\tau \cup \sigma \in X(k-i)$, each of which is hit $\binom{k-i}{j}$ times in the sum. □

3.9 Proof of Claim 3.4.4

Claim 3.9.1 (Claim 3.4.4 Restated). For all $k \in \mathbb{N}$, $j \leq k$, and $\ell \leq k - j$:

$$\sum_{i=\ell}^{k-j} \binom{k}{i+j} \frac{\binom{k-j-\ell}{i-\ell}}{\binom{k-j}{i}} \rho^{i+j}(1-\rho)^{k-j-i} \leq \rho^\ell$$

Proof. We first re-write the sum to be 0-indexed for simplicity:

$$\sum_{i=\ell}^{k-j} \binom{k}{i+j} \frac{\binom{k-j-\ell}{i-\ell}}{\binom{k-j}{i}} \rho^{i+j}(1-\rho)^{k-j-i} = \sum_{i=0}^{k-j-\ell} \frac{\binom{k}{i+j+\ell}\binom{k-j-\ell}{i}}{\binom{k-j}{i+\ell}} \rho^{i+j+\ell}(1-\rho)^{k-j-\ell-i}$$

The key is to observe that by standard counting arguments:

$$\frac{\binom{k}{i+j+\ell}\binom{k-j-\ell}{i}}{\binom{k-j}{i+\ell}} \leq \binom{k-\ell}{i+j}, \tag{3.10}$$

since by the Binomial Theorem we then have

$$\sum_{i=0}^{k-j-\ell} \binom{k}{i+j+\ell} \frac{\binom{k-j-\ell}{i}}{\binom{k-j}{i+\ell}} \rho^{i+j+\ell}(1-\rho)^{k-j-\ell-i} \leq \rho^\ell \sum_{i=0}^{k-j-\ell} \binom{k-\ell}{i+j} \rho^{i+j}(1-\rho)^{k-j-\ell-i}$$

$$\leq \rho^{\ell} \sum_{z=0}^{k-\ell} \binom{k-\ell}{z} \rho^{z} (1-\rho)^{k-\ell-z}$$
$$= \rho^{\ell}$$

as desired. Proving Equation (3.10) is tedious but elementary. Standard binomial manipulations give

$$\frac{\binom{k}{i+j+\ell}\binom{k-j-\ell}{i}}{\binom{k-j}{i+\ell}\binom{k-\ell}{i+j}} = \frac{\binom{k}{j}}{\binom{k-\ell}{j}} \cdot \frac{\binom{i+j}{j}}{\binom{i+j+\ell}{j}},$$

and the result follows from observing the RHS is increasing in i, and is exactly 1 when $i = k - j - \ell$. □

Chapter 4

Eigenstripping, Spectral Decay, and Edge-Expansion on Posets

4.1 Introduction

Random walks on high dimensional expanders (HDX) have been the object of intense study in theoretical computer science in recent years. Starting with their original formulation by Kaufman and Mass [234], a series of works on the spectral structure of these walks [239, 111, 11] led to significant breakthroughs in approximate sampling [25, 11, 24, 95, 96, 94, 146, 218, 286, 66], CSP-approximation [9, 38], error-correcting codes [220, 221], agreement testing [124, 109, 236], and more. Most of these works focus on the structure of expansion in *hypergraphs* (typically called *simplicial complexes* in the HDX literature). On the other hand, it has become increasingly clear that hypergraphs are not always the right tool for the job—recent breakthroughs in locally testable [117] and quantum LDPC codes [315, 281, 277], for instance, all rely crucially on *cubical* structure not seen in hypergraphs, while many agreement testing results like the proof of the 2-2 Games Conjecture [255] rely crucially on *linear algebraic* rather than simplicial structure.

In this work, we study a generalized notion of high dimensional expansion on *partially ordered sets* (posets) introduced by Dikstein, Dinur, Filmus, and Harsha (DDFH) [111] called *expanding posets* (eposets). Random walks on eposets capture a broad range of important structures beyond their hypergraph analogs, including natural sparsifications

of the Grassmann graphs that recently proved crucial to the resolution of the 2-2 Games Conjecture [255, 253, 126, 125, 47, 252]. DDFH's notion of eposets is a *global* definition of high dimensional expansion based on a relaxation of Stanley's [346] sequentially differential posets, a definition originally capturing both the Grassmanian and complete simplicial complex. More recently, Kaufman and Tessler (KT) [243] have extended the study of eposets in two important aspects. First, in contrast to DDFH's original global definition, KT introduced the local-to-global study of high dimensional expansion in eposets. Second, they identified *regularity* as a key parameter controlling expansion. In particular, the authors showed strengthened local-to-global theorems for strongly regular posets like the Grassmann, giving the first general formulation for characterizing expansion based on an eposet's underlying architecture.

While analysis of the second eigenvalue is certainly an important consideration (e.g. for mixing applications), a deeper understanding of the spectral structure of eposets is required for applications like the proof of the 2-2 Games Conjecture. As such, our main focus in this work lies in characterizing the spectral and combinatorial behavior of random walks on eposets *beyond the second eigenvalue*. Strengthening DDFH and recent work of Bafna, Hopkins, Kaufman, and Lovett (BHKL) [38], we prove that at a coarse level (walks on) eposets indeed exhibit the same spectral and combinatorial characteristics as expanding hypergraphs (e.g. spectral stripping, expansion of pseudorandom sets). On the other hand, as in KT, we show that the finer-grained properties of these objects are actually controlled by the underlying poset's regularity, including the *rate of decay* of the spectrum and combinatorial expansion of associated random walks. This gives a stronger separation between structures like hypergraphs with weak (linear) eigenvalue decay, and Grassmann-based eposets with strong (exponential) eigenvalue decay (a crucial property in the proof of the 2-2 Games Conjecture [255]).

In slightly more detail, we show that all eposets exhibit a behaviour called "eigen-stripping" [239, 111, 38]: the spectrum of any associated random walk concentrates around

a few unique approximate eigenvalues. Moreover, the approximate eigenvalues of walks on eposets are tightly controlled by the poset architecture's regularity[1] $R(j, i)$, which denotes the total number of rank-i elements[2] less than any fixed rank-j element (see Section 4.1.1 for standard definitions). For simplicity, we specialize our result below to the popular "lower" or "down-up" walk (this simply corresponds to taking a random step down and back up the poset, again see Section 4.1.1); a more involved version holds for higher order random walks in full generality.

Theorem 4.1.1 (Eigenstripping and Regularity (informal Corollary 4.4.5 and Theorem 4.4.7))**.** *The spectrum of the lower walk UD on a k-dimensional γ-eposet is concentrated in $(k+1)$ strips:*

$$Spec(UD) \in \{1\} \cup \bigcup_{i=1}^{k} [\lambda_i(UD) + O_k(\gamma), \lambda_i(UD) - O_k(\gamma)],$$

where the approximate eigenvalues $\lambda_i(UD)$ are determined by the poset's regularity:

$$\lambda_i(UD) = \frac{R(k-1, i)}{R(k, i)}.$$

Theorem 4.1.1 generalizes and tightens recent work on expanding hypergraphs of BHKL [38, Theorem 2.2] (which itself extended a number of earlier works on the topic [239, 111, 9]). Additionally, our result on the connection between regularity and approximate eigenvalues generalizes the work of KT [243], who show an analogous result for λ_2. Theorem 4.1.1 reveals a stark contrast between the spectral behavior of eposets with different regularity parameters. As a prototypical example, consider the case of hypergraphs versus subsets of the Grassmann (k-dimensional vector spaces over $\mathbb{F}_q^n$). In the former, each k-set contains $\binom{k}{i}$ i-sets, leading to approximate eigenvalues that decay

[1] We will additionally assume a slightly stronger condition introduced in KT [243] called *middle regularity* throughout. See Section 4.2.1 for details.

[2] We consider *regular graded posets*, where each element has a corresponding rank. In a hypergraph, for instance, rank is given by the size of a set, while in the Grassmann poset it is given by subspace dimension.

linearly ($\lambda_i \approx (k-i)/k$). On the other hand, each k-dimensional vector space contains $\binom{k}{i}_q$ i-dimensional subspaces, which leads to eigenvalues that decay *exponentially* ($\lambda_i \approx q^{-i}$). The latter property, which we call *strong decay* is often crucial in applications (e.g. for hardness of approximation [255] or fast algorithms [38]), and while it is possible to recover strong decay on weaker posets by increasing the length of the walk [38], this is often untenable in application due to the additional degrees of freedom it affords.[3]

The spectral structure of walks on eposets is closely related to their *edge-expansion*, an important combinatorial property that has recently played a crucial role both in algorithms for [37, 38] and hardness of unique games [255]. The key insight in both cases lay in understanding *the structure of non-expanding sets*. We give a tight understanding of this phenomenon across all eposets in the so-called ℓ_2-regime [38], where we show that expansion is tightly controlled by the behavior of local restrictions called *links* (see Definition 4.1.7).

Theorem 4.1.2 (Expansion in the ℓ_2-Regime (informal Theorem 4.6.7))**.** *The expansion of any i-link is almost exactly $1 - \lambda_i(M)$. Conversely, any set with expansion less than $1 - \lambda_{i+1}(M)$ has* ***high variance*** *across i-links.*

In [38], it was shown this characterization allows for the application of a local-to-global algorithmic framework for unique games on such walks. This remains true on eposets, and it is an interesting open question whether there are significant applications beyond those given in BHKL's original work.[4]

Finally, as an application of our structure theorems, we give an in-depth analysis of the ℓ_2-structure of walks on expanding subsets of the Grassmann poset called q-eposets (first studied in [111]). We focus in particular on the natural q-analog of an important

[3]For instance such a walk might take exponential time to implement, or correspond to a more complicated agreement test than desired.

[4]While one can apply the framework to playing unique games on the Grassmann poset (or sparsifications thereof), the spectral parameters are such that this does not give a substantial improvement over standard algorithms [29].

set of walks called *partial-swap walks* introduced by Alev, Jeronimo, and Tulsiani [9] that generalize the Johnson graphs when applied to expanding hypergraphs. We show that applied to q-eposets, these objects give a natural set of walks generalizing the Grassmann graphs and further prove that our generic analysis for eposets gives a tight characterization of non-expansion in this setting. We note that this does not recover the result used for the proof of the 2-2 Games Conjecture which lies in the ℓ_∞*-regime* (replacing variance above with maximum) and requires a *dimension-independent* bound. This issue was recently (and independently) resolved for simplicial complexes in [39] and [187], and we view our work as an important step towards a more general understanding for families like the Grassmann beyond hypergraphs.

4.1.1 Background

Before jumping into our results in any further formality, we'll briefly overview the theory of expanding posets and higher order random walks. All definitions are covered in full formality in Section 4.2. A d-dimensional graded poset is a set X equipped with a partial order "$<$" and a ranking function $r : X \to [d]$ that respects the partial order and partitions X into levels $X(0) \cup \ldots \cup X(d)$. When $x < y$ and $r(y) = r(x) + 1$, we write $x \lessdot y$ or equivalently $y \gtrdot x$.[5] Finally, we will assume throughout this work that our posets are *downward regular*: there exists a regularity function $R(k, i)$ such that every k-dimensional element is greater than exactly $R(k, i)$ i-dimensional elements.[6]

Graded posets come equipped with a natural set of averaging operators called the *up* and *down* operators. Namely, for any function $f : X(i) \to \mathbb{R}$, these operators average f up or down one level of the poset respectively:

$$U_i f(x) = \mathop{\mathbb{E}}_{y \lessdot x} [f(y)],$$

[5]This is traditionally called a 'covering relation.'

[6]For notational convenience, we also define $R(i, i) = 1$ and $R(j, i) = 0$ whenever $j < i$.

$$D_i f(y) = \mathop{\mathbb{E}}_{x \succcurlyeq y}[f(x)].$$

Composing the averaging operators leads to a natural notion of random walks on the underlying poset called *higher order random walks* (HD-walks). The simplest example of such a walk is the *upper walk* $D_{i+1}U_i$ which moves between elements $x, x' \in X(i)$ via a common element $y \in X(i+1)$ with $y > x, x'$. Similarly, the *lower walk* $U_{i-1}D_i$ walks between $x, x' \in X(i)$ via a common $y \in X(i-1)$ with $y < x, x'$. It will also be useful at points to consider longer variants of the upper and lower walks called *canonical walks* $\widehat{N}_k^i = D_{k+1} \circ \ldots \circ D_{k+i} \circ U_{k+i-1} \circ \ldots \circ U_k$ and $\check{N}_k^i = U_k \circ \ldots \circ U_{k-i} \circ D_{k-i+1} \circ \ldots \circ D_k$ which similarly walk between k-dimensional elements in $X(k)$ via a shared element in $X(k+i)$ or $X(k-i)$ respectively.

Following DDFH [111], we call a poset a (δ, γ)-expander for $\delta \in [0,1]^{d-1}$ and $\gamma \in \mathbb{R}_+$ if the upper and lower walks are spectrally similar up to a laziness factor:

$$\|D_{i+1}U_i - (1-\delta_i)I - \delta_i U_{i-1}D_i\| \leq \gamma.$$

This generalizes standard spectral expansion which can be equivalently defined as looking at the spectral norm of $A_G - U_0 D_1$, where A_G (the adjacency matrix) is exactly the non-lazy upper walk. We note that under reasonable regularity conditions (see [243, 111]), this definition is equivalent to *local-spectral expansion* [124], which requires every local restriction of the poset to be an expander graph. While most of our results hold more generally, it will also be useful to assume a weak non-laziness condition on our underlying posets throughout that holds in most cases of interest (see Definition 4.2.8).

4.1.2 Results

With these definitions in mind, we can now cover our results in somewhat more formality. We split this section into three parts for readability: spectral stripping, charac-

terizing edge expansion, and applications to the Grassmann.

Eigenstripping

We start with our generalized spectral stripping theorem for walks on expanding posets.

Theorem 4.1.3 (Spectrum of HD-Walks (informal Corollary 4.4.5))**.** *Let M be an HD-walk on the kth level of a (δ, γ)-eposet. Then the spectrum of M is highly concentrated in $k+1$ strips:*

$$Spec(M) \in \{1\} \cup \bigcup_{i=1}^{k} [\lambda_i(M) - e, \lambda_i(M) + e]$$

where $e \leq O_{k,\delta}(\gamma)$. Moreover, the span of eigenvectors in the ith strip approximately correspond to functions lifted from $X(i)$ to $X(k)$.

This generalizes and improves an analogous result of BHKL [38] on expanding hypergraphs, which had sub-optimal error dependence of $O_k(\gamma^{1/2})$. The main improvement stems from an optimal spectral stripping result for arbitrary inner product spaces of independent interest.

Theorem 4.1.4 (Eigenstripping (informal Theorem 4.3.2))**.** *Let M be a self-adjoint operator over an inner product space V, and $V = V^1 \oplus \ldots \oplus V^k$ be an "approximate eigendecomposition" in the sense that there exist $\{\lambda_i\}_{i=1}^k$ and sufficiently small error factors $\{c_i\}_{i=1}^k$ such that for all $f_i \in V^i$:*

$$\|Mf_i - \lambda_i f_i\|_2 \leq c_i \|f_i\|.$$

Then the spectrum of M is concentrated around each λ_i:

$$Spec(M) \subseteq \bigcup_{i=1}^{k} [\lambda_i - c_i, \lambda_i + c_i].$$

Note that this result is tight—when there is c_i "error" in our basis we cannot expect to have better than c_i error in the resulting spectral strips. Theorem 4.1.4 improves over a preliminary result to this effect in [38] which had substantially worse dependence on c_i and required much stronger assumptions.[7] Theorem 4.1.3 then follows by work of DDFH ([111, Theorem 8.6]), who introduced a natural approximate eigendecomposition on eposets we call the HD-Level-Set Decomposition.

In full generality, the approximate eigenvalues in Theorem 4.1.3 depend on the eposet parameters δ, and can be fairly difficult to interpret. However, we show that under weak assumptions (see Section 4.2) the eigenvalues can be associated with the regularity of the underlying poset. We focus on the lower walks for simplicity, though the result can be similarly extended to general walks on eposets.

Theorem 4.1.5 (Regularity Controls Spectral Decay (informal Theorem 4.4.7))**.** *The approximate eigenvalues of the lower walk $\check{N}_k^{k-i}$ on a (δ, γ)-eposet are controlled by the poset's regularity function:*

$$\lambda_j(\check{N}_k^{k-i}) \in \frac{R(i,j)}{R(k,j)} \pm O_{k,\delta}(\gamma).$$

As discussed in Section 4.1, this generalizes work of Kaufman and Tessler [243] for the second eigenvalue of the upper/lower walks, and reveals a major distinction among poset architectures: posets with higher regularity enjoy faster decay of eigenvalues. We note that Theorem 4.1.1 can also be obtained by combining Theorem 4.1.4 with recent independent work of Dikstein, Dinur, Filmus, and Harsha on connections between eposets and regularity (namely in the recent update of their seminal eposet paper, see [111, Section 8.4.1]).

On a more concrete note, Theorem 4.1.5 gives a new method of identifying potential

[7] It is also worth noting that the proof in this work is substantially simplified from [38], requiring no linear algebraic manipulations at all.

poset architectures exhibiting strong spectral decay in the sense that for any $\delta > 0$, the lower walk only contains $O_\delta(1)$ approximate eigenvalues larger than δ (rather than a number growing with dimension). This property, referred to as *constant ST-Rank* in the context of hypergraphs in [38], is an important factor not only for the run-time of approximation algorithms on HDX [38], but also for the soundness of the Grassmann-based agreement test in the proof of the 2-2 Games Conjecture [255].

Characterizing Edge Expansion

Much of our motivation for studying the spectrum of HD-walks comes from the desire to understand a fundamental combinatorial quantity of graphs called *edge expansion*.

Definition 4.1.6 (Edge Expansion)**.** Let X be a graded poset and M an HD-Walk on $X(k)$. The edge expansion of a subset $S \subset X(k)$ with respect to M is

$$\Phi(S) = \mathop{\mathbb{E}}_{v \sim S} \left[M(v, X(k) \setminus S) \right],$$

where

$$M(v, X(k) \setminus S) = \sum_{y \in X(k) \setminus S} M(v, y)$$

and $M(v, y)$ denotes the transition probability from v to y.

As mentioned in the introduction, characterizing the edge-expansion of sets in HD-walks has recently proven crucial to understanding both algorithms for [37, 38] and hardness of unique games [255]. On expanding hypergraphs, it has long been known that *links* give the canonical example of small non-expanding sets.

Definition 4.1.7 (Link)**.** Let X be a d-dimensional graded poset. The k-dimensional link

of an element $\sigma \in X$ is the set of rank k elements greater than σ:[8]

$$X_\sigma^k = \{y \in X(k) : y > \sigma\}.$$

We call the link of a rank-i element an "i-link." When the level k is clear from context, we write X_σ for X_σ^k for simplicity.

In greater detail, BHKL [38] proved that on hypergraphs, the expansion of links is exactly controlled by their corresponding spectral strip. While their proof of this fact relied crucially on simplicial structure, we show via a more general analysis that the result can be recovered for eposets.

Theorem 4.1.8 (Expansion of Links (informal Theorem 4.6.3))**.** *Let X be a (δ, γ)-eposet and M an HD-walk on $X(k)$. Then for all $0 \le i \le k$ and $\tau \in X(i)$:*

$$\Phi(X_\tau) = 1 - \lambda_i(M) \pm O_{M,k,\delta}(\gamma).$$

As an immediate consequence, we get that when M is not a small-set expander, links are examples of small non-expanding sets. One might reasonably wonder whether the converse is true as well: *are all non-expanding sets explained by links*? This requires a bit of formalization. Following BHKL's exposition [38], given a set S consider the function $L_{S,i} : X(i) \to \mathbb{R}$ that encodes the behavior of $S \subset X(k)$ on links:

$$\forall \tau \in X(i) : L_{S,i}(\tau) = \mathop{\mathbb{E}}_{X_\tau}[\mathbb{1}_S] - \mathbb{E}[\mathbb{1}_S].$$

The statement "Non-expansion is explained by links" can then be interpreted as saying that a non-expanding set S should be detectable by some simple measure of $L_{S,i}$. There are two standard formalizations of this idea studied in the literature: the ℓ_2-regime, and

[8] We note that in the literature a link is usually defined to be all such elements, not just those of rank k. We adopt this notation since we are mostly interested in working at a fixed level of the complex.

the ℓ_∞-regime. These are captured by the following notion of *pseudorandomness* based on $L_{S,i}$.

Definition 4.1.9 (Pseudorandom Sets [38] (informal Definitions 4.5.2, 4.5.5))**.** We say a set S is (ε, ℓ)-ℓ_2-pseudorandom if[9]

$$\forall i \leq \ell : \|L_{S,i}\|_2^2 \leq \varepsilon \mathbb{E}[\mathbb{1}_S].$$

A set is (ε, ℓ)-ℓ_∞-pseudorandom if:

$$\forall i \leq \ell : \|L_{S,i}\|_\infty \leq \varepsilon.$$

In cases that ℓ_2 and ℓ_∞-pseudorandomness can be used interchangeably, we will simply write (ε, ℓ)-pseudorandom.

We prove that pseudorandom sets expand near-optimally.

Theorem 4.1.10 (Pseudorandom Sets Expand (informal Theorem 4.6.7))**.** *Let X be a (δ, γ)-eposet and M a walk on $X(k)$. Then the expansion of any (ε, i)-pseudorandom set S is at least:*

$$\Phi(S) \geq 1 - \lambda_{i+1} - O_\delta(R(k,i)\varepsilon) - O_{k,\delta,M}(\gamma).$$

In other words, any set with expansion less than $1 - \lambda_{i+1}$ must have appreciable variance across links at level i. We note that the formal version of this result is essentially tight in the ℓ_2-regime, but can be improved in many important cases in the ℓ_∞-regime. We'll discuss this further in the next section, especially in the context of the Grassmann poset.

Before this, however, it is worth separately mentioning the main technical component behind Theorem 4.1.10, a result traditionally called a "level-i" inequality.

[9]Throughout, $\|\cdot\|_2$ will always refer to the *expectation* norm $\|f\|_2 = \mathbb{E}[f^2]^{1/2}$.

Theorem 4.1.11 (Level-i inequality (informal Theorem 4.5.7))**.** *Let X be a (δ, γ)-eposet and $S \subset X(k)$ a (ε, ℓ)-pseudorandom set. Then for all $1 \leq i \leq \ell$:*

$$|\langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle| \leq (R(k,i)\varepsilon + O_{k,\delta}(\gamma)) \langle \mathbb{1}_S, \mathbb{1}_S \rangle$$

where $\mathbb{1}_{S,i}$ is the projection of $\mathbb{1}_S$ onto the ith eigenstrip.[10]

In other words, pseudorandomness controls the projection of S onto eigenstrips. Theorems 4.1.10 and 4.1.11 recover the analogous optimal bounds for simplicial high dimensional expanders in [38], where the regularity parameter $R(k,i) = \binom{k}{i}$, and are tight in a number of other important settings such as the Grassmann (discussed below). Theorem 4.1.10 and Theorem 4.1.11 can also be viewed as another separation between eposet architectures, this time in terms of *combinatorial* rather than *spectral* properties.

Application: q-eposets and the Grassmann Graphs

Finally, we'll discuss the application of our results to a particularly important class of eposets called "q-eposets." Just like standard high dimensional expanders arise from expanding subsets of the complete complex (hypergraph), q-eposets arise from expanding subsets of the Grassmann Poset.

Definition 4.1.12 (Grassmann Poset)**.** The Grassmann Poset is a graded poset $(X, <)$ where X is the set of all subspaces of $\mathbb{F}_q^n$ of dimension at most d, the partial ordering "$<$" is given by inclusion, and the rank function is given by dimension.

We call a (downward-closed) subset of the Grassmann poset a q-simplicial complex, and an expanding q-simplicial complex a q-eposet (see Section 4.2.5 for exact details). Using our machinery for general eposets, we prove a tight level-i inequality for pseudorandom sets.

[10]Note that since walks on eposets are simultaneously diagonalizable, the decomposition of X into eigenstrips is independent of the choice of walk.

Corollary 4.1.13 (Grassmann level-i inequality (informal Theorem 4.7.7))**.** *Let X be a γ-q-eposet and $S \subseteq X(k)$. If S is (ε, ℓ)-pseudorandom, then for all $1 \leq i \leq \ell$:*

$$|\langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle| \leq \left(\binom{k}{i}_q \varepsilon + O_{q,k}(\gamma) \right) \langle \mathbb{1}_S, \mathbb{1}_S \rangle$$

where $\binom{k}{i}_q = \frac{(1-q^k)\cdots(1-q^{k-i+1})}{(1-q^i)\cdots(1-q)}$ is the Gaussian binomial coefficient.

Corollary 4.1.13 is tight in a few senses. First, we prove the bound cannot be improved by any constant factor, even in the ℓ_∞-regime. In other words, for every $c < 1$, it is always possible to find an (ε, i)-pseudorandom function satisfying:

$$|\langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle| > c \left(\binom{k}{i}_q \varepsilon + O_{q,k}(\gamma) \right) \langle \mathbb{1}_S, \mathbb{1}_S \rangle.$$

Furthermore, it is well known the dependence on k in this result is necessary [253], even if one is willing to suffer a worse dependence on the pseudorandomness ε. This is different from the case of standard simplicial complexes, where the dependence can be removed in the ℓ_∞-regime [252, 39, 187]. However, there is a crucial subtlety here. It is likely that the k-dependence in this result can be removed by *changing the definition of pseudorandomness*. On the Grassmann poset itself, for instance, it is known that this can be done by replacing links with a closely related but finer-grained local structure known as "zoom-in zoom-outs" [255]. Indeed, more generally it is an interesting open problem whether there always exists a notion of locality based on the underlying poset structure that gives rise to k-independent bounds in the ℓ_∞-regime.

We close out the section by looking at an application of this level-i inequality to studying edge-expansion in an important class of walks that give rise to the well-studied *Grassmann graphs*.

Definition 4.1.14 (Grassmann Graphs)**.** The Grassmann Graph $J_q(n, k, t)$ is the graph on k-dimensional subspaces of $\mathbb{F}_q^n$ where $(V, W) \in E$ exactly when $\dim(V \cap W) = t$.

It is easy to see that the non-lazy upper walk on the Grassmann poset is exactly the Grassmann graph $J_q(n,k,k-1)$. In fact, it is possible to express any $J_q(n,k,t)$ as a sum of standard higher order random walks.

Proposition 4.1.15 (Grassmann Graphs are HD-Walks (informal Proposition 4.7.5))**.** *The Grassmann graphs are a hypergeometric sum of canonical walks:*

$$J_q(n,k,t) = \frac{1}{q^{(k-t)^2}\binom{k}{t}_q}\sum_{i=0}^{k-t}(-1)^{k-t-i}q^{\binom{k-t-i}{2}}\binom{k-t}{i}_q\binom{k+i}{i}_q N_k^i.$$

In Section 4.7 we prove a more general version of this result for any q-simplicial complex. This leads to a set of natural sparsifications of the Grassmann graphs that may be of independent interest for agreement testing, PCPs, and hardness of approximation. For simplicity, on a given q-simplicial complex X, we'll refer to these "sparsified" Grassmann graphs as $J_{X,q}(n,k,t)$ for the moment (more formally they are the "partial-swap walks," see Section 4.2.5). With this in mind, let's take a look at what our level-i inequality implies for the edge-expansion of these graphs.

Corollary 4.1.16 (q-eposets Edge-Expansion (informal Corollary 4.7.10))**.** *Let X be a d-dimensional γ-q-eposet and $S \subset X(k)$ a (ε,ℓ)-pseudorandom set. Then the expansion of S with respect to the sparsified Grassmann graph $J_{X,q}(n,k,t)$ is at least:*

$$\Phi(S) \geq 1 - \mathbb{E}[\mathbb{1}_S] - \varepsilon\sum_{i=1}^{\ell}\binom{t}{i}_q - q^{-(\ell+1)j} - O_{q,k}(\gamma).$$

In practice, t is generally thought of as being $\Omega(k)$ (or even $k-O(1)$), which results in a k-dependent bound. It remains an open problem whether a k-independent version can be proved for any q-eposet beyond the Grassmann poset itself. We conjecture such a result should indeed hold (albeit under a different notion of pseudorandomness), and may follow from q-analog analysis of recent work proving k-independent bounds for standard expanding hypergraphs [39, 187].

4.1.3 Related Work

Higher Order Random Walks..

Higher order random walks were first introduced in 2016 by Kaufman and Mass [234]. Their spectral structure was later elucidated in a series of works by Kaufman and Oppenheim [239], DDFH [111], Alev, Jeronimo, and Tulsiani [9], Alev and Lau [11], and finally BHKL [38]. With the exception of DDFH (who only worked with approximate eigenvectors without analyzing the true spectrum), all of these works focused on hypergraphs rather than general posets. Our spectral stripping theorem for eposets essentially follows from combining eposet machinery developed by DDFH with our improved variant of BHKL's general spectral stripping theorem.

Higher order random walks have also seen an impressive number of applications in recent years, frequently closely tied to analysis of their spectral structure. This has included breakthrough works on approximate sampling [25, 11, 24, 95, 96, 94, 146, 218, 286, 66], CSP-approximation [9, 38], error-correcting codes [220, 221], and agreement testing [124, 109, 236]. In this vein, our work is most closely related to that of Bafna, Barak, Kothari, Schramm, and Steurer [37], and BHKL [38], who used the spectral and combinatorial structure of HD-walks to build new algorithms for unique games. As previously discussed, the generalized analysis in this paper also lends itself to the algorithmic techniques developed in those works, but we do not know of any interesting examples beyond those covered in BHKL.

High Dimensional Expansion Beyond Hypergraphs..

Most works listed above (and indeed in the high dimensional expansion literature in general) focus only on the setting of hypergraphs. However, recent years have also seen the nascent development and application of expansion beyond this setting [117, 315, 281, 277, 208], including the seminal work of DDFH [111] on expanding posets as well as more recent breakthroughs on locally testable and quantum codes [117, 315]. While DDFH

largely viewed eposets as having similar structure (with the exception of the Grassmann), we strengthen the case that different underlying poset architectures exhibit different properties. This complements the recent result of Kaufman and Tessler [243], who showed that expanding posets with strong regularity conditions such as the Grassmann exhibit more favorable properties with respect to the second eigenvalue. Our results provide a statement of the same flavor looking at the entire spectrum, along with additional separations in more combinatorial settings. We note that a related connection between poset regularity and the approximate spectrum of walks on eposets was independently developed by DDFH in a recent update of their seminal work [111].

Expansion and Unique Games..

One of the major motivations behind this work is towards building a more general framework for understanding the structure underlying the Unique Games Conjecture [250], a major open problem in complexity theory that implies optimal hardness of approximation results for a large swath of combinatorial optimization problems (see e.g. Khot's survey [259]). In 2018, Khot, Minzer, and Safra [255] made a major breakthrough towards the UGC in proving a weaker variant known as the 2-2 Games Conjecture, completing a long line of work in this direction [253, 126, 125, 47, 252, 255]. The key to the proof lay in a result known as the "Grassmann expansion hypothesis," which stated that any non-expanding set in the Grassmann graph $J_q(d, k, k-1)$ had to be non-trivially concentrated inside a local-structure called "zoom-in zoom-outs." As noted in the previous section, this result differs from our analysis in two key ways: it lies in the ℓ_∞-regime, and must be totally independent of dimension.

Unfortunately, very little progress has been made towards the UGC since this result. This is in part because KMS' proof of the Grassmann expansion hypothesis, while a tour de force, is complicated and highly tailored to the exact structure of the Grassmann. To our knowledge, the same proof cannot be used, for instance, to resolve the related

"shortcode expansion hypotheses" beyond degree-2, similar conjectures offered by Barak, Kothari, and Steurer [47] in an effort to push beyond hardness of 2-2 Games. Just as the ℓ_2-regime analysis of DDFH and BHKL recently lead to a dimension independent bound in the ℓ_∞-regime for standard HDX [39, 187], we expect the groundwork laid in this paper will be important for proving generalized dimension independent expansion hypotheses in the ℓ_∞-regime beyond the special case of the Grassmann graphs.

4.2 Preliminaries

Before jumping into the details in full formality, we give a more careful review of background definitions regarding expanding posets, higher order random walks, and the Grassmann.

4.2.1 Graded Posets

We start with eposets' underlying structure, graded posets. A partially ordered set (poset) $P = (X, <)$ is a set of elements X endowed with a partial order "$<$". A graded poset comes equipped additionally with a rank function $r : X \to \mathbb{N}$ satisfying two properties:

1. r preserves "$<$": if $y < x$, then $r(y) < r(x)$.
2. r preserves cover relations: if x is the smallest element greater than y, then $r(x) = r(y) + 1$.

In other words, the function r partitions X into subsets by rank:

$$X(0) \cup \ldots \cup X(d),$$

where $\max_X(r) = d$, and $X(i) = r^{-1}(i)$. We refer to a poset with maximum rank d as "d-dimensional", and elements in $X(i)$ as "i-faces". Throughout this work, we will consider only d-dimensional graded posets with two additional restrictions:

1. They have a unique minimal element, i.e. $|X(0)| = 1$.

2. They are "pure": all maximal elements have rank d.

Finally, many graded posets of interest satisfy certain regularity conditions which will be crucial to our analysis. The first condition of interest is a natural notion called *downward* regularity.

Definition 4.2.1 (Downward Regularity)**.** We call a d-dimensional graded poset downward regular if for all $i \leq d$ there exists some constant $R(i)$ such that every element $x \in X(i)$ covers exactly $R(i)$ elements $y \in X(i-1)$.

Second, we will also need a useful notion called *middle regularity* that ensures uniformity across multiple levels of the poset.

Definition 4.2.2 (Middle Regularity)**.** We call a d-dimensional graded poset middle-regular if for all $0 \leq i \leq k \leq d$, there exists a constant $m(k,i)$ such that for any $x_k \in X(k)$ and $x_i \in X(i)$ satisfying $x_k > x_i$, there are exactly $m(k,i)$ chains[11] of elements $x_k > x_{k-1} > \ldots > x_{i+1} > x_i$ where each $x_j \in X(j)$.

We call a poset regular if it is both downward and middle regular. We note that regular posets also have the nice property that for any dimensions $i < k$, there exists a higher order regularity function $R(k,i)$ such that any $x \in X(k)$ is greater than exactly $R(k,i)$ elements in $X(i)$ (see Section 4.8). We will use this notation throughout. For notational convenience, we also define $R(i,i) = 1$ and $R(j,i) = 0$ whenever $j < i$.

Important examples of regular posets include pure simplicial complexes and the Grassmann poset (subspaces of $\mathbb{F}_q^n$ ordered by inclusion). We will assume all posets we discuss in this work are regular from this point forward.

[11] Such objects are sometimes called flags, e.g. in the case of the Grassmann poset.

4.2.2 Measured Posets and The Random Walk Operators

Higher order random walks may be defined over posets in a very similar fashion to simplicial complexes. The main difference is simply that "inclusion" is replaced with the poset order relation. Just as we might want these walks on HDX to have non-uniform weights, the same is true for posets, which can be analogously endowed with a distribution over levels. In slightly more detail, a *measured poset* is a graded poset X endowed with a distribution $\Pi = (\pi_0, \ldots, \pi_d)$, where each marginal π_i is a distribution over $X(i)$. While measured posets may be defined in further generality (cf. [111, Definition 8.1]), we will focus on the case in which the distribution Π is induced entirely from π_d, analogous to weighted simplicial complexes. More formally, we have that for every $0 \leq i < d$:

$$\pi_i(x) = \frac{1}{R(i+1,i)} \sum_{y \succ x} \pi_{i+1}(y).$$

In other words, each lower dimensional distribution π_i may be induced through the following process: an element $y \in X(i+1)$ is selected with respect to π_{i+1}, and an element $x \in X(i)$ such that $x < y$ is then chosen uniformly at random.

The averaging operators U and D are defined analogously to their notions on simplicial complexes, with the main change being the use of the general regularity function $R(i+1,i)$:

$$U_i f(y) = \frac{1}{R(i+1,i)} \sum_{x \preccurlyeq y} f(x),$$
$$D_{i+1} f(x) = \frac{1}{\pi_{i+1}(X_x)} \sum_{y \succ x} \pi_{i+1}(y) f(y),$$

where for $i < k$ and $x \in X(i)$,

$$\pi_k(X_x) = \sum_{y \in X(k) : y > x} \pi_k(y) = R(k,i)\pi_i(x)$$

is the appropriate normalization factor (we will use this notation throughout). On regular posets, it is useful to note that the up operators compose nicely, and in particular that:

$$U_i^k f(y) \coloneqq U_{k-1} \circ \ldots \circ U_i f(y) = \frac{1}{R(k,i)} \sum_{x \in X(i): x < y} f(x)$$

(see Section 4.8). Furthermore, just like on simplicial complexes, the down and up operators are adjoint with respect to the standard inner product on measured posets:

$$\langle f, g \rangle_{X(k)} = \sum_{\tau \in X(k)} \pi_k(\tau) f(\tau) g(\tau),$$

that is to say for any $f : X(k) \to \mathbb{R}$ and $g : X(k-1) \to \mathbb{R}$:

$$\langle f, U_{k-1} g \rangle_{X(k)} = \langle D_k f, g \rangle_{X(k-1)}.$$

Note that we'll generally drop the $X(k)$ from the notation when clear from context. This useful fact allows us to define basic self-adjoint notions of higher order random walks just like on simplicial complexes.

4.2.3 Higher Order Random Walks

Let C_k denote the space of functions $f : X(k) \to \mathbb{R}$. We define a natural set of random walk operators via the averaging operators.

Definition 4.2.3 (k-Dimensional Pure Walk [234, 111, 9])**.** Given a measured poset (X, Π), a k-dimensional pure walk $Y : C_k \to C_k$ on (X, Π) (of height $h(Y)$) is a composition:

$$Y = Z_{2h(Y)} \circ \cdots \circ Z_1,$$

where each Z_i is a copy of D or U, and there are $h(Y)$ of each type.

Following AJT and BHKL, we define general higher order random walks to be affine combinations[12] of pure walks.

Definition 4.2.4 (HD-walk)**.** Let X be a graded poset. Let $\mathcal{Y}$ be a family of pure walks $Y : C_k \to C_k$ on (X, Π). We call an affine combination

$$M = \sum_{Y \in \mathcal{Y}} \alpha_Y Y$$

a k-dimensional HD-walk on (X, Π) if it is stochastic and self-adjoint. The height of M, denoted $h(M)$, is the maximum height of any pure $Y \in \mathcal{Y}$ with a non-zero coefficient. The weight of M, denoted $w(M)$, is $|\alpha|_1$.

While most of our results will hold for general HD-walks (or at least some large subclass), we pay special attention to a basic class of pure walks that have seen the most study in the literature: *canonical walks*.

Definition 4.2.5 (Canonical Walk)**.** Given a d-dimensional measured poset (X, Π) and parameters $k + j \leq d$, the upper canonical walk $\widehat{N}_k^j$ is:

$$\widehat{N}_k^j = D_k^{k+j} U_k^{k+j},$$

and for $j \leq k$ the lower canonical walk $\check{N}_k^j$ is:

$$\check{N}_k^j = U_{k-j}^k D_{k-j}^k,$$

where $U_\ell^k = U_{k-1} \dots U_\ell$, and $D_\ell^k = D_{\ell+1} \dots D_k$.

Since the non-zero spectrum of $\widehat{N}_k^j$ and $\check{N}_{k+j}^j$ are equivalent (c.f. [11]), we focus in this work mostly on the upper walks which we write simply as N_k^j.

[12] An affine combination is a linear combination whose coefficients sum to 1.

For certain specially structured posets, we will also study an important class of HD-walks known as (partial) *swap walks*. We will introduce these well-studied walks in more detail in Section 4.2.5, and for now simply note that they give a direct generalization of the Johnson and Grassmann graphs when applied to the complete complex and Grassmann poset respectively.

4.2.4 Expanding Posets and the HD-Level-Set Decomposition

Dikstein, Dinur, Filmus, and Harsha [111] observed that one can use the averaging operators to define a natural extension of spectral expansion to graded posets. Their definition is inspired by the fact that γ-spectral expansion on a standard graph G can be restated as a bound on the spectral norm of the adjacency matrix minus its stationary operator:

$$\|A_G - UD\| \leq \gamma.$$

Informally, DDFH's definition can be thought of as stating that this relation holds for every level of a higher dimensional poset.

Definition 4.2.6 (eposet [111])**.** Let (X, Π) be a measured poset, $\delta \in [0,1]^{d-1}$, and $\gamma < 1$. X is an (δ, γ)-eposet if for all $1 \leq i \leq d-1$:

$$\|D_{i+1}U_i - (1-\delta_i)I - \delta_i U_{i-1}D_i\| \leq \gamma$$

.

We note that for a broad range of posets, this definition is actually equivalent (up to constants) to *local-spectral expansion*, a popular notion of high dimensional expansion introduced by Dinur and Kaufman [124]. This was originally proved for simplicial complexes by DDFH [111], and later extended to a more general class of posets by Kaufman and Tessler [243]. It is also worth noting that when $\gamma = 0$, posets satisfying the guarantee in

Definition 4.2.6 are known as *sequentially differential*, and were actually introduced much earlier by Stanley [346] in the late 80s.

Much of our analysis in this work will be based off of an elegant approximate Fourier decomposition for eposets introduced by DDFH [111].

Theorem 4.2.7 (HD-Level-Set Decomposition, Theorem 8.2 [111])**.** *Let (X, Π) be a d-dimensional (δ, γ)-eposet with γ sufficiently small. For all $0 \leq k \leq d$, let*

$$H^0 = C_0, H^i = Ker(D_i), V_k^i = U_i^k H^i.$$

Then:

$$C_k = V_k^0 \oplus \ldots \oplus V_k^k.$$

In other words, every $f \in C_k$ has a unique decomposition $f = f_0 + \ldots + f_k$ such that $f_i = U_i^k g_i$ for $g_i \in Ker(D_i)$.

It is well known that the HD-Level-Set Decomposition is approximately an eigenbasis for HD-walks on simplicial complex [111, 9, 38]. We show this statement extends to all eposets in Section 4.4 (extending DDFH's similar analysis of the upper walk N_k^1).

Finally, before moving on, we will assume for simplicity throughout this work an additional property of eposets we called (approximate) non-laziness.

Definition 4.2.8 (β-non-Lazy Eposets)**.** Let (X, Π) be a d-dimensional measured poset. We call (X, Π) β-non-lazy if for all $1 \leq i \leq d$, the laziness of the lower walk satisfies:

$$\max_{\sigma \in X(i)} \{\mathbb{1}_\sigma^T U_{i-1} D_i \mathbb{1}_\sigma\} \leq \beta.$$

Another way to think about this condition is that no element in the poset carries too much weight, even upon conditioning. All of our results hold for general eposets,[13] but

[13]The one exception is the lower bound of Theorem 4.1.8.

their form is significantly more interpretable when the poset is additionally non-lazy. In fact, most γ-eposets of interest are $O(\gamma)$-non-lazy. It is easy to see for instance that any "γ-local-spectral" expander satisfies this condition, an equivalent notion of expansion to γ-eposets under suitable regularity conditions [243]. We discuss this further in Section 4.8.

4.2.5 The Grassmann Poset and q-eposets

At the moment, there are only two known families of expanding posets of significant interest in the literature: those based on pure simplicial complexes (the downward closure of a k-uniform hypergraph), and pure q-simplicial complexes (the analogous notion over subspaces). The ℓ_2-structure of the former set of objects is studied in detail in [38]. In this work, we will focus on the latter which has seen less attention in the literature, but is responsible for a number of important results including the resolution of the 2-to-2 Games Conjecture [255].

Definition 4.2.9 (q-simplicial complex)**.** Let $G_q(n,d)$ denote the d-dimensional subspaces of $\mathbb{F}_q^n$. A weighted, pure q-simplicial complex (X,Π) is given by a family of subspaces $X \subseteq G_q(n,d)$ and a distribution Π over X. We will usually consider the downward closure of X in the following sense:

$$X = X(0) \cup \ldots \cup X(d),$$

where $X(i) \subseteq G_q(n,i)$ consists of all i-dimensional subspaces contained in some element in $X = X(d)$. Further, on each level $X(i)$, Π induces a natural distribution π_i:

$$\forall V \in X(i) : \pi_i(V) = \frac{1}{\binom{d}{i}_q} \sum_{W \in X(d): W \supset V} \pi_d(W),$$

where $\pi_d = \Pi$ and $\binom{d}{i}_q = \frac{(1-q^d)\cdots(1-q^{d-i+1})}{(1-q^i)\cdots(1-q)}$ is the Gaussian binomial coefficient.

The most basic example of a q-simplicial complex is the Grassmann poset, which corresponds to taking $X = G_q(n,d)$. This is the q-analog of the complete simplicial

complex. The Grassmann poset is well known to be a expander in this sense (see e.g. [346])—in fact it is a sequentially differential poset with parameters

$$\delta_i = \frac{(q^i - 1)(q^{n-i+1} - 1)}{(q^{i+1} - 1)(q^{n-i} - 1)},$$

the q-analog of the eposet parameters for the complete complex [111]. With this in mind, let's define a special class of eposets based on q-simplicial complexes.

Definition 4.2.10 (γ-q-eposet [111])**.** A pure, d-dimensional weighted q-simplicial complex (X, Π) is a γ-q-eposet if it is a (δ, γ)-eposet satisfying $\delta_i = q\frac{q^i-1}{q^{i+1}-1}$ for all $1 \leq i \leq d-1$.

Constructing bounded-degree q-eposets (a problem proposed by DDFH [111]) remains an interesting open problem. Kaufman and Tessler [243] recently made some progress in this direction, but the expansion parameter of their construction is fairly poor (around 1/2).

Finally, in our applications to the Grassmann we'll focus our attention on a particularly important class of walks called *partial-swap walks*. These should essentially be thought of as non-lazy variants of the upper canonical walks.

Definition 4.2.11 (Partial-Swap Walk)**.** Let (X, Π) be a weighted, d-dimensional q-simplicial complex. The partial-swap walk S_k^j is the restriction of the canonical walk N_k^j to faces whose intersection has dimension $k - j$. In other words, if $|V \cap W| > k - j$ then $S_k^j(V, W) = 0$, and otherwise $S_k^j(V, W) \propto N_k^j(V, W)$.

When applied to the Grassmann poset itself, it is clear by symmetry that the partial-swap walk S_k^j returns exactly the Grassmann graph $J_q(d, k, k-j)$. On the other hand, it is not immediately obvious these objects are even HD-walks when applied to a generic q-simplicial complex. We prove this is the case in Section 4.7.

4.3 Eigendecompositions and Eigenstripping

With preliminaries out of the way, we can move on to understanding HD-walks' spectral structure. It turns out that on expanding posets, these walks exhibit almost exactly the same properties as on the special case of simplicial complexes studied in [239, 111, 9, 38]: a walk's spectrum lies concentrated in strips corresponding to levels of the HD-Level-Set Decomposition. The key to proving this lies in a more general theorem characterizing the spectral structure of any inner product space admitting a "approximate eigendecomposition."

Definition 4.3.1 (Approximate Eigendecomposition [38])**.** Let M be an operator over an inner product space V. A decomposition $V = V^1 \oplus \ldots \oplus V^k$ is called a $(\{\lambda_i\}_{i=1}^k, \{c_i\}_{i=1}^k)$-approximate eigendecomposition if for all i and $v_i \in V^i$, Mv_i is close to $\lambda_i v_i$:

$$\|Mv_i - \lambda_i v_i\| \leq c_i \|v_i\|.$$

We will always assume for simplicity (and without loss of generality) that the λ_i are sorted: $\lambda_1 \geq \ldots \geq \lambda_k$.

BHKL [38] proved that as long as the c_i are sufficiently small, each V^i (loosely) corresponds to an "eigenstrip," the span of eigenvectors with eigenvalue closely concentrated around λ_i, and that these strips account of the entire spectrum of M. While sufficient for their purposes, their proof of this result was complicated and resulted in a variety of sub-optimal parameters. We give a tight variant of this result and significantly simplify the proof.

Theorem 4.3.2 (Eigenstripping)**.** *Let M be a self-adjoint operator over an inner product space V, and $V = V^1 \oplus \ldots \oplus V^k$ a $(\{\lambda_i\}_{i=1}^k, \{c_i\}_{i=1}^k)$-approximate eigendecomposition. Then*

as long as $c_i + c_{i+1} < \lambda_i - \lambda_{i+1}$, the spectrum of M is concentrated around each λ_i:

$$Spec(M) \subseteq \bigcup_{i=1}^{k} [\lambda_i - c_i, \lambda_i + c_i]$$

Proof. The idea is to examine for each i the operator $M_i^2 = (M - \lambda_i I)^2$. In particular, we claim it is enough to show the following:

Claim 4.3.3. For all $1 \leq i \leq k$, $\text{Spec}(M_i^2)$ contains $dim(V^i)$ eigenvalues less than c_i^2.

Let's see why this implies the desired result. Notice that the eigenvalues of M_i^2 are exactly $(\mu - \lambda_i)^2$ for each μ in $Spec(M)$ (with matching multiplicities), and therefore that any eigenvalue $\mu_i \in Spec(M_i^2)$ less than c_i^2 implies the existence of a corresponding eigenvalue of M in $[\lambda_i \pm c_i]$. If each M_i^2 has $\dim(V^i)$ eigenvalues less than c_i^2, then M has at least $\dim(V^i)$ eigenvalues in each interval $[\lambda_i \pm c_i]$. Moreover, since these intervals are disjoint by assumption and $\sum \dim(V^i) = \dim(V)$, this must account for all eigenvalues of M.

It remains to prove the claim, which is essentially an immediate application of Courant-Fischer theorem [149].

Proof of Claim 4.3.3. The Courant-Fischer theorem states that the kth smallest eigenvalue of a self-adjoint operator A is:

$$\lambda_{n-k+1} = \min_{U} \left\{ \max_{f \in U} \left\{ \frac{\langle f, Af \rangle}{\langle f, f \rangle} \right\} \;\middle|\; \dim(U) = k \right\}.$$

Setting $U = V^i$, $A = M_i^2$ and $k = \dim(V^i)$ gives the claim:

$$\lambda_{n-k+1}(M_i^2) \leq \max_{f \in V^i} \left\{ \frac{\langle f, M_i^2 f \rangle}{\langle f, f \rangle} \right\} = \max_{f \in V^i} \left\{ \frac{\|(M - \lambda_i I) f\|_2^2}{\langle f, f \rangle} \right\} \leq c_i^2$$

since $(M - \lambda_i I)$ is self-adjoint and $\bigoplus_{i \in [k]} V^i$ is a $(\{\lambda_i\}_{i=1}^k, \{c_i\}_{i=1}^k)$-approximate eigendecomposition. □

□

Note that this result is also trivially tight, as any true eigendecomposition is also a $(\{\lambda_i \pm c_i\}, \{c_i\})$-approximate eigendecomposition. We also note that similar strategies have been used in the numerical analysis literature (see e.g. [210]).

4.4 Spectra of HD-walks

Given Theorem 4.3.2, it is enough to prove that the HD-Level-Set Decomposition is an approximate eigenbasis for any HD-walk. This follows by the same inductive argument as for local-spectral expanders in [38], where the only difference is that somewhat more care is required to deal with general eposet parameters. To start, it will be useful to lay out some notation along with a simple observation from repeated application of Definition 4.2.6.

Lemma 4.4.1 ([111, Claim 8.8])**.** *Let (X, Π) be a d-dimensional (δ, γ)-eposet. Then*

$$\|D_{k+1}U_{k-j}^{k+1} - (1-\delta_j^k)U_{k-j}^k - \delta_j^k U_{k-j-1}^k D_{k-j}\| \leq \gamma_j^k,$$

where

$$\delta_{-1}^k = 1,\ \delta_j^k = \prod_{i=k-j}^{k} \delta_i,\ \gamma_j^k = \gamma \sum_{i=-1}^{j-1} \delta_i^k.$$

Applying this fact inductively implies that functions in the HD-Level-Set Decomposition are close to being eigenvectors.

Proposition 4.4.2. *Let (X, Π) be a (δ, γ)-eposet, and Y the pure balanced walk of height j, with down operators at positions $(i_1, \ldots, i_j)$. For $1 \leq \ell \leq k$, let $f_\ell = U_\ell^k g_\ell$ for some $g_\ell \in H^\ell$, and let*

$$\delta_j^k = \prod_{i=k-j}^{k} \delta_i,\ \gamma_j^k = \gamma \sum_{i=-1}^{j-1} \delta_i^k,$$

where $\delta_i^k = 1$ *for any* $i < 0$ *for notational convenience. Then* f_ℓ *is an approximate eigenvector of* Y*:*

$$\|Y f_\ell - \prod_{s=1}^{j} \left(1 - \delta_{k-2s+i_s-\ell}^{k-2s+i_s}\right) f_\ell\| \leq \|g_\ell\| \sum_{s=1}^{j} \gamma_{k-2s+i_s-\ell}^{k-2s+i_s} \prod_{t=1}^{s-1} \left(1 - \delta_{k-2t+i_t-\ell}^{k-2t+i_t}\right) \leq (j+k) j \gamma \|g_\ell\|.$$

Proof. We prove a slightly stronger statement to simplify the induction. For $b > 0$, let $Y_j^b : C_\ell \to C_{\ell+b}$ denote an unbalanced walk with j down operators, and $j + b$ up operators. If Y_j^b has down operators in positions $(i_1, \ldots, i_j)$ and $g_\ell \in H^\ell$, we claim:

$$\|Y_j^b g_\ell - \prod_{s=1}^{j} \left(1 - \delta_{i_s-2s}^{i_s+\ell-2s}\right) Y_0^b g_\ell\| \leq \|g_\ell\| \sum_{s=1}^{j} \gamma_{i_s-2s}^{i_s+\ell-2s} \prod_{t=1}^{s-1} \left(1 - \delta_{i_t-2t}^{i_t+\ell-2t}\right),$$

which implies the result (notice that the indices i_s shift by $b = k - \ell$). The base case $j = 0$ is trivial. Assume the inductive hypothesis holds for all $Y_i^b, i < j$. By Lemma 4.4.1 and recalling $g_\ell \in \ker(D_\ell)$, we have:

$$Y_j^b g_\ell = \left(1 - \delta_{i_1-2}^{i_1+\ell-2}\right) Y_{j-1}^b g_\ell + \Gamma g_\ell,$$

where Γ has spectral norm

$$\|\Gamma\| \leq \gamma_{i_1-2}^{i_1+\ell-2}.$$

Notice that Y_{j-1}^b has down operator indices $\{i_2 - 2, \ldots, i_j - 2\}$. The inductive hypothesis then implies:

$$\begin{aligned} Y_j^b g_\ell &= \left(1 - \delta_{i_1-2}^{i_1+\ell-2}\right) \prod_{s=2}^{j} \left(1 - \delta_{i_s-2s}^{i_s+\ell-2s}\right) Y_0^b g_\ell + \left(1 - \delta_{i_1-2}^{i_1+\ell-2}\right) \Gamma' g_\ell + \Gamma g_\ell \\ &= \prod_{s=1}^{j} \left(1 - \delta_{i_s-2s}^{i_s+\ell-2s}\right) g_\ell + \left(1 - \delta_{i_1-2}^{i_1+\ell-2}\right) \Gamma' g_\ell + \Gamma g_\ell, \end{aligned}$$

where $\Gamma' g_\ell$ has norm

$$\|\Gamma' g_\ell\| \leq \|g_\ell\| \sum_{s=2}^{j} \gamma_{i_s-2s}^{i_s+\ell-2s} \prod_{t=2}^{s-1} \left(1 - \delta_{i_t-2t}^{i_t+\ell-2t}\right).$$

Thus we may bound the norm of the righthand error term by:

$$\begin{aligned}\|\left(1 - \delta_{i_1-2}^{i_1+\ell-2}\right) \Gamma' g_\ell + \Gamma g_\ell\| &\leq \left(1 - \delta_{i_1-2}^{i_1+\ell-2}\right) \|\Gamma'\|\|g_\ell\| + \|\Gamma\|\|g_\ell\| \\ &\leq \sum_{s=1}^{j} \gamma_{i_s-2s}^{i_s+\ell-2s} \prod_{t=1}^{s-1} \left(1 - \delta_{i_t-2t}^{i_t+\ell-2t}\right) \|g_\ell\|,\end{aligned}$$

as desired. Recalling the shift in i_s by $k - \ell$, we can then bound the resulting error by $(j + k)j\gamma\|g_\ell\|$ since $\delta \in [0, 1]^{d-1}$. □

It is worth noting that when $\gamma = 0$, this implies that the HD-Level-Set decomposition is a true eigendecomposition. Since balanced walks are simply affine combinations of pure walks, this immediately implies a similar result for the more general case. To align with our definition of approximate eigendecompositions and Theorem 4.3.2, we'll also need the following general relation between $\|g_\ell\|$ and $\|f_\ell\|$ for eposets proved in [111] (albeit without the exact parameter dependence).

Lemma 4.4.3 ([111, Lemma 8.11])**.** *Let (X, Π) be a d-dimensional (δ, γ)-eposet, $0 \leq \ell \leq k < d$, and let*

$$\rho_\ell^k = \prod_{i=1}^{k-\ell} \left(1 - \delta_{k-\ell-i}^{k-i}\right), \quad \rho_{min} = \min_{0 \leq \ell \leq k} \{\rho_\ell^k\}.$$

Then for any $f_\ell = U_\ell^k g_\ell$ for $g_\ell \in Ker(D_\ell)$ we have:

$$\langle f_\ell, f_\ell \rangle \in (\rho_\ell^k \pm k^2\gamma)\langle g_\ell, g_\ell \rangle,$$

and for all $i \neq \ell$:

$$\langle f_\ell, f_i \rangle \leq O\left(\frac{k^2}{\rho_{min}} \gamma \|f_\ell\|\|f_i\|\right).$$

As an aside, we remark that the parameter ρ_ℓ^k turns out to be a crucial throughout much of our work, and while it is difficult to interpret on general eposets, we prove it has a very natural form as long as non-laziness holds.

Claim 4.4.4 (ρ_ℓ^k for regular eposets). Let (X, Π) be a regular, γ-non-lazy[14] d-dimensional (δ, γ)-eposet. Then for any $i \leq k < d$, we have:

$$\rho_i^k \in \frac{1}{R(k,i)} \pm err,$$

where $err \leq O\left(\frac{i^3 k^2 R_{\max}}{\delta_i(1-\delta_{i-1})}\gamma\right)$. Likewise as long as $\gamma \leq O\left(\frac{\max_i\{\delta_i(1-\delta_{i-1})\}}{i^3 k^2 R_{\max}^2}\right)$ we have

$$\rho_{\min}^{-1} \leq O(R_{\max}),$$

where $R_{\max} \coloneqq \max_{0\leq i\leq k}\{R(k,i)\}$.

This gives a nice generalization of the interpretation of ρ_i^k on hypergraphs, which is well known to be $\frac{1}{\binom{k}{i}}$ [111]. We prove this claim in Section 4.8. For simplicity, we will assume throughout the rest of this work that our eposets are γ-non-lazy, which is true for most cases of interest (see Section 4.8). All results holds in the more general case using ρ_i^k unless otherwise noted.

Combining Proposition 4.4.2 and Lemma 4.4.3 immediately implies that the HD-Level-Set Decomposition is an approximate eigendecomposition in the sense of Definition 4.3.1.

Corollary 4.4.5. *Let (X, Π) be a (δ, γ)-eposet and let $M = \sum_{Y\in\mathcal{Y}} \alpha_Y Y$ be an HD-walk. For $1 \leq \ell \leq k$, if $f_\ell = U_\ell^k g_\ell$ for some $g_\ell \in H^\ell$, then for $\gamma \leq O\left(\frac{\max_i\{\delta_i(1-\delta_{i-1})\}}{k^5 R_{max}^2}\right)$:*

$$\|Mf_\ell - \left(\sum_{Y\in\mathcal{Y}} \alpha_Y \lambda_{Y,\delta,\ell}\right) f_\ell\| \leq c\gamma\|f_\ell\|,$$

[14] One can prove this claim more generally for any β-non-laziness, but most γ-eposets of interest are additionally γ-non-lazy, so this simplified version is generally sufficient.

where $\lambda_{Y,\delta,\ell}$ is the corresponding eigenvalues of the pure balanced walk Y on a $(\delta,0)$-eposet (see Proposition 4.4.2), and $c \leq O\left((h(M)+k)h(M)R(k,\ell)w(M)\right)$.

Thus as long as the walk in question is self-adjoint (e.g. canonical or swap walk), Theorem 4.3.2 immediately implies that the true spectrum is concentrated around these approximate eigenvalues.

Before moving on it is instructive (and as we will soon see quite useful) to give an example application of Corollary 4.4.5 to a basic higher order random walk.

Corollary 4.4.6 (Spectrum of Lower Canonical Walks)**.** *Let (X,Π) be a (δ,γ)-eposet. The approximate eigenvalues of the canonical lower walk $\breve{N}_k^{k-\ell}$ are:*

$$\lambda_j(\breve{N}_k^{k-\ell}) = \prod_{s=1}^{k-\ell}(1-\delta_{k-s-j}^{k-s}).$$

Proof. The lower canonical walk $\breve{N}_k^{k-\ell} = U_\ell^k D_\ell^k$ is of height $k-\ell$, and has down operator at positions $\{1,\ldots,k-\ell\}$. In the language of Proposition 4.4.2 we therefore have $i_s = s$, which therefore gives:

$$\lambda_j(\breve{N}_k^{k-\ell}) = \prod_{s=1}^{k-\ell}(1-\delta_{k-s-j}^{k-s}).$$

Note this is 0 when $j > \ell$. □

Similar to the case of ρ_i^k, while this is difficult to interpret in the general setting, the eigenvalues have a very natural form on non-lazy eposets given by the regularity parameters.

Theorem 4.4.7. *Let (X,Π) be a γ-non-lazy (δ,γ)-eposet. The approximate eigenvalues of the canonical lower walk $\breve{N}_k^{k-i}$ are:*

$$\lambda_j(\breve{N}_k^{k-i}) \in \frac{R(i,j)}{R(k,j)} \pm c\gamma,$$

where $c \leq O\left(\frac{i^4k^2R_{max}}{\delta_i(1-\delta_{i-1})}\gamma\right)$.

The proof requires machinery developed in Section 4.6 and Section 4.8, and is given in Section 4.8.

4.5 Pseudorandomness and the HD-Level-Set Decomposition

Now that we know the spectral structure of HD-walks, we shift to studying their combinatorial structure. In particular, we will focus on how natural notions of pseudorandomness control the projection of functions onto the HD-Level-Set Decomposition.

Before proceeding, we state a simple corollary of Lemma 4.4.3 that will prove useful going forward:

Corollary 4.5.1. *Let (X, Π) be a (δ, γ)-eposet and suppose $f \in C_k$ has HD-Level-Set Decomposition $f = f_0 + \ldots + f_k$. If $\gamma \leq \frac{c'\rho_{\min}}{k^3}$ for a sufficiently small constant $c' > 0$, then*

$$\sum_{j=0}^{k} \|f_j\| \leq O(\sqrt{k}\|f\|). \tag{4.1}$$

Moreover, for any subset of indices I, it holds that

$$-\sum_{j\in I}\langle f, f_j\rangle \leq O\left(\frac{k^3\gamma\|f\|^2}{\rho_{min}}\right).$$

In particular, if $I = \{j : \langle f, f_j\rangle \leq 0\}$, then

$$\sum_{j\in I}|\langle f, f_j\rangle| \leq O\left(\frac{k^3\gamma\|f\|^2}{\rho_{min}}\right).$$

Proof. For the first claim, recall that by the approximate orthogonality of the HD-Level-Set Decomposition (Lemma 4.4.3), we have for all $i \neq j$:

$$|\langle f_i, f_j\rangle| \leq O\left(\frac{k^2}{\rho_{\min}}\gamma\|f_i\|\|f_j\|\right).$$

Then, applying Cauchy-Schwarz gives:

$$\begin{aligned}\left(\sum_{j=1}^{k}\|f_j\|\right)^2 &\leq k\sum_{j=1}^{k}\|f_j\|^2 \\ &\leq k\langle f,f\rangle - k\sum_{i\neq j\neq 0}\langle f_i,f_j\rangle \\ &\leq k\langle f,f\rangle + c\gamma\sum_{i\neq j\neq 0}\|f_i\|\|f_j\| \\ &\leq k\langle f,f\rangle + c\gamma\left(\sum_{j=1}^{k}\|f_j\|\right)^2\end{aligned}$$

where $c \leq O\left(\frac{k^3}{\rho_{\min}}\right)$. By our assumption on γ, we have $c\gamma \leq \frac{1}{2}$, and therefore rearranging yields

$$\sum_{i=1}^{k}\|f_j\| \leq O(\sqrt{k}\|f\|).$$

We now show how the second claim is a consequence of the first. For any subset I, we have

$$\begin{aligned}-\sum_{j\in I}\langle f,f_j\rangle &\leq -\sum_{j\in I}\sum_{i\neq j}\langle f,f_j\rangle \\ &\leq \frac{Ck^2\gamma}{\rho_{\min}}\sum_{i,j}\|f_i\|\|f_j\| \\ &= \frac{O(k^2\gamma)}{\rho_{\min}}\left(\sum_{i=0}^{k}\|f_i\|\right)^2 \\ &\leq O\left(\frac{k^3\gamma\|f\|}{\rho_{\min}}\right).\end{aligned}$$

□

4.5.1 ℓ_2-pseudorandomness

We start with pseudorandomness in the ℓ_2-regime, which measures the variance of a set across links.

Definition 4.5.2 (ℓ_2-Pseudorandom functions [38])**.** A function $f \in C_k$ is $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_2$-pseudorandom if its variance across i-links is small for all $1 \leq i \leq \ell$:

$$\mathrm{Var}(D_i^k f) \leq \varepsilon_i |\mathbb{E}[f]|.$$

In their work on simplicial complexes, BHKL [38] observed a close connection between ℓ_2-pseudorandomness, the HD-Level-Set Decomposition, and the spectra of the lower canonical walks. We'll show the same connection holds in general for eposets.

Theorem 4.5.3. *Let (X, Π) be a (δ, γ)-eposet with $\gamma \leq O\left(\frac{\max_i\{\delta_i(1-\delta_{i-1})\}}{k^5 R_{max}^2}\right)$. If $f \in C_k$ has HD-Level-Set Decomposition $f = f_0 + \ldots + f_k$, then for any $\ell \leq k$, $\mathrm{Var}(D_\ell^k f)$ is controlled by its projection onto $V_k^0 \oplus \ldots \oplus V_k^\ell$ in the following sense:*

$$\mathrm{Var}(D_\ell^k f) \in \sum_{j=1}^{\ell} \lambda_j(\check{N}_k^{k-\ell}) \langle f, f_j \rangle \pm c_k \gamma \|f\|^2,$$

where $c_k \leq O(k^{5/2} R_{max})$ and $\lambda_j(\check{N}_k^{k-\ell}) = \prod_{s=1}^{k-\ell}(1 - \delta_{k-s-j}^{k-s})$.

Proof. To start, notice that since $\langle D_\ell^k f, D_\ell^k f \rangle = \langle \check{N}_k^{k-\ell} f, f \rangle$ it is enough to analyze the application of $\check{N}_k^{k-\ell}$ to f. By Corollary 4.4.6, we know that each f_j is an approximate eigenvector satisfying:

$$\|\check{N}_k^{k-\ell} f_j - \lambda_j(\check{N}_k^{k-\ell}) f_j\| \leq O(k^2 R(k, \ell)\gamma) \|f_j\|,$$

where $\lambda_j(\check{N}_k^{k-\ell}) = 0$ for $j > \ell$. Combining these observations gives:

$$\begin{aligned}
\mathrm{Var}(D_\ell^k f) &= \left\langle D_\ell^k f, D_\ell^k f \right\rangle - \mathbb{E}[D_\ell^k f]^2 \\
&= \left\langle f, U_\ell^k D_\ell^k f \right\rangle - \langle f, f_0 \rangle \\
&= \sum_{j=1}^{k} \langle f, U_\ell^k D_\ell^k f_j \rangle
\end{aligned}$$

$$\in \sum_{j=1}^{\ell} \lambda_j(\check{N}_k^{k-\ell})\langle f, f_j\rangle \pm O\left(\frac{k^2}{\rho_{\min}}\gamma\|f\|\sum_{j=1}^{k}\|f_j\|\right).$$

where we have additionally used the fact that $\langle f, f_0\rangle = \mathbb{E}[f]^2 = \mathbb{E}[D_\ell^k f]^2$ and $\lambda_0(\check{N}_k^{k-\ell}) = 1$. Applying Equation (4.1) from Corollary 4.5.1 to bound the sum in the error term and replacing ρ with the relevant regularity parameters by Claim 4.4.4 then gives the result. □

As an immediate corollary, we get a level-i inequality for pseudorandom functions.

Corollary 4.5.4. *Let (X, Π) be a (δ, γ)-eposet with $\gamma \leq O\left(\frac{\max_i\{\delta_i(1-\delta_{i-1})\}}{k^5 R_{max}^2}\right)$ and let $f \in C_k$ be an $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_2$-pseudorandom function. Then for any $1 \leq i \leq \ell$:*

$$|\langle f, f_i\rangle| \leq R(k,i)\varepsilon_i|\mathbb{E}[f]| + c\gamma\|f\|^2,$$

where $c \leq O\left(\frac{k^5 R_{max}^2}{\max_i\{\delta_i(1-\delta_{i-1})\}}\right)$.

Proof. By Corollary 4.5.1, for any given $1 \leq i \leq k$, it holds that $-\sum_{j\neq i}\langle f, f_j\rangle \leq O\left(\frac{k^3}{\rho_{\min}}\gamma\|f\|^2\right)$. It follows from Theorem 4.5.3 that for all $0 \leq i \leq k$, the variance of $D_i^k f$ is lower bounded by its projection onto f_i:

$$\mathrm{Var}(D_i^k f) \geq \lambda_i(\check{N}_k^{k-i})\langle f, f_i\rangle - c\gamma\langle f, f\rangle,$$

where $c \leq O(\frac{k^3}{\rho_{\min}})$. Noting that $\lambda_i(\check{N}_k^{k-i}) = \rho_i^k$, if $i \leq \ell$, re-arranging the above and applying the pseudorandomness assumption gives:

$$\begin{aligned}\langle f, f_i\rangle &\leq \frac{1}{\rho_i^k}\mathrm{Var}(D_i^k f) + c_2\gamma\langle f, f\rangle \\ &\leq \frac{1}{\rho_i^k}\varepsilon_i|\mathbb{E}[f]| + c_2\gamma\langle f, f\rangle,\end{aligned}$$

where $c_2 \leq O(\frac{k^3}{\rho_{\min}^2})$. The lower bound on $\langle f, f_i\rangle$ is immediate from Corollary 4.5.1 with the set $I = \{i\}$. Applying Claim 4.4.4 then gives the result. □

As mentioned previously, this also recovers the tight inequality for simplicial complexes given in [38] where $R(k,i) = \binom{k}{i}$, as well as providing the natural q-analog for q-simplicial complexes where $R(k,i) = \binom{k}{i}_q$.

4.5.2 ℓ_∞-pseudorandomness

While ℓ_2-pseudorandomness is useful in its own right (e.g. for local-to-global algorithms for unique games [37, 38]), there is also significant interest in a stronger ℓ_∞-variant in the hardness of approximation literature [252, 255].

Definition 4.5.5 (ℓ_∞-Pseudorandom functions)**.** A function $f \in C_k$ is $(\varepsilon_1, \dots, \varepsilon_\ell)$-$\ell_\infty$-pseudorandom if for all $1 \le i \le \ell$ its local expectation is close to its global expectation:

$$\left\| D_i^k f - \mathbb{E}[f] \right\|_\infty \le \varepsilon_i.$$

In their recent work on ℓ_2-structure of expanding simplicial complexes, BHKL prove a basic reduction from ℓ_∞ to ℓ_2-pseudorandomness that allows for an analogous level-i inequality for this notion as well. Here, we'll show the same result holds for general eposets. As in their work, we'll take advantage of a weak local-consistency property called locally-constant sign.

Definition 4.5.6 (locally-constant sign [38])**.** Let (X, Π) be a graded poset. We say a function $f \in C_k$ has ℓ-local constant sign if:

1. $\mathbb{E}[f] \neq 0$,
2. $\forall s \in X(\ell)$ s.t. $\underset{X_s}{\mathbb{E}}[f] \neq 0 : \operatorname{sign}\left(\underset{X_s}{\mathbb{E}}[f]\right) = \operatorname{sign}\left(\mathbb{E}[f]\right)$.

With this in mind, we now state ℓ_∞-variant of Corollary 4.5.4:

Theorem 4.5.7. *Let (X, Π) be a (δ, γ)-eposet with $\gamma \le O\left(\frac{\max_i\{\delta_i(1-\delta_{i-1})\}}{k^5 R_{max}^2}\right)$ and let $f \in C_k$ have HD-Level-Set Decomposition $f = f_0 + \ldots + f_k$. If f is $(\varepsilon_1, \dots, \varepsilon_\ell)$-$\ell_\infty$-pseudorandom,*

then for all $1 \leq i \leq \ell$*:*

$$|\langle f, f_i \rangle| \leq (R(k,i) + c\gamma)\, \varepsilon_i^2 + c\gamma \|f\|^2,$$

and if f *has* i*-local constant sign:*

$$|\langle f, f_i \rangle| \leq (R(k,i) + c\gamma)\, \varepsilon_i |\mathbb{E}[f]| + c\gamma \|f\|^2$$

where in both cases $c \leq O\left(\frac{k^5 R_{max}^2}{\max_i \{\delta_i(1-\delta_{i-1})\}}\right)$.

We note that when f is boolean, this bound simplifies to

$$|\langle f, f_i \rangle| \leq (R(k,i)\varepsilon_i + c\gamma)\, \mathbb{E}[f],$$

which we'll see in the next section is a particularly useful form for analyzing edge expansion. The proof of Theorem 4.5.7 relies mainly on a reduction to the ℓ_2-variant for functions with locally-constant sign. This reduction is almost exactly the same as in [38], but we include it for completeness.

Lemma 4.5.8. *Let* (X, Π) *be a graded poset and* $f \in C_k$ *a* $(\varepsilon_1, \ldots, \varepsilon_\ell)$*-*$\ell_\infty$*-pseudorandom function with* i*-local constant sign for any* $i \leq \ell$*. Then* f *is* $(\varepsilon_1, \ldots, \varepsilon_\ell)$*-*$\ell_2$*-pseudorandom.*

Proof. As in [38], the idea is to notice that locally constant sign allows us to rewrite $\|D_i^k f\|_2^2$ as an expectation over some related distribution P_i:

$$\begin{aligned}
\frac{1}{\mathbb{E}[f]} \langle D_i^k f, D_i^k f \rangle &= \sum_{s \in X(i)} \pi_i(s) \left(\frac{1}{\mathbb{E}[f]} \sum_{t \in X_s} \frac{\pi_k(t) f(t)}{\pi_k(X_s)} \right) D_i^k f(s) \\
&= \sum_{s \in X(i)} \left(\frac{1}{R(k,i)} \frac{\sum_{t \in X_s} \pi_k(t) f(t)}{\mathbb{E}[f]} \right) D_i^k f(s) \\
&= \mathop{\mathbb{E}}_{P_i}[D_i^k f],
\end{aligned}$$

where P_i being a probability distribution follows from the locally-constant sign of f, and the second step follows from the fact that $\pi_k(X_s) = \sum_{t \in X_s} \pi_k(t) = R(k,i)\pi_i(s)$. The result then follows easily from averaging:

$$\left| \frac{1}{\mathbb{E}[f]} \mathrm{Var}(D_i^k f) \right| = \left| \mathop{\mathbb{E}}_{P_i}[D_i^k f] - \mathbb{E}[f] \right| \leq \|D_i^k f - \mathbb{E}[f]\|_\infty.$$

When $\mathbb{E}[f] > 0$, the ℓ_∞-norm here may be replaced with maximum. □

The proof of Theorem 4.5.7 now follows from reducing to the case of locally-constant sign. The argument is exactly as in the proof of [38, Theorem 8.7], but we include it for completeness.

Proof of Theorem 4.5.7. We focus on the general bound, since the result for functions with locally constant sign is immediate from Lemma 4.5.8 and Corollary 4.5.4. The argument for general functions f follows simply from noting that we can always shift f to have locally constant sign. With this in mind, assume $\mathbb{E}[f] \geq 0$ for simplicity (the negative case is similar). Let $f' = f + (\varepsilon_i - \mathbb{E}[f])\mathbb{1}$ be the aforementioned shift. As long as $\varepsilon_i > 0$, it is easy to see that f' has i-local constant sign and further that

$$f' = f'_0 + f_i + \ldots + f_k,$$

where $f'_0 = f_0 + (\varepsilon_i - \mathbb{E}[f])\mathbb{1}$. Since shifts have no effect on ℓ_∞-pseudorandomness, f' is $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_\infty$-pseudorandom by assumption, and therefore $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_2$-pseudorandom by Lemma 4.5.8. We can now apply Corollary 4.5.4 to get:

$$\begin{aligned} \langle f + (\varepsilon_i - \mathbb{E}[f])\mathbb{1}, f_i \rangle &\leq \frac{1}{\rho_i^k} \varepsilon_i \mathbb{E}[f + (\varepsilon_i - \mathbb{E}[f])\mathbb{1}] + c\gamma \langle f + (\varepsilon_i - \mathbb{E}[f])\mathbb{1}, f + (\varepsilon_i - \mathbb{E}[f])\mathbb{1} \rangle \\ &\leq \left(\frac{1}{\rho_i^k} + c\gamma \right) \varepsilon_i^2 + c\gamma \langle f, f \rangle, \end{aligned}$$

since $\langle f_i, \mathbb{1} \rangle = 0$ for all $i > 0$. Finally, as this holds for all $\varepsilon_i > 0$, a limiting argument

implies the result for $\varepsilon_i = 0$. Applying Claim 4.4.4 completes the proof. □

4.6 Expansion of HD-walks

It is well known that higher order random walks on simplicial complexes (e.g. the Johnson graphs) are not small-set expanders. BHKL gave an exact characterization of this phenomenon for local-spectral expanders: they showed that the expansion of any i-link with respect to an HD-walk M is almost exactly $1 - \lambda_i(M)$. Moreover, using the level-i inequality from the previous section, BHKL proved a tight converse to this result in an ℓ_2-sense: *any* non-expanding set must have high variance across links. This gave a complete ℓ_2-characterization of non-expanding sets on local-spectral expanders, and lay the structural groundwork for new algorithms for unique games over HD-walks.

In this section, we'll show that these results extend to general expanding posets. To start, let's recall the definition of edge expansion.

Definition 4.6.1 (Weighted Edge Expansion)**.** Let (X, Π) be a graded poset and M a k-dimensional HD-Walk. The weighted edge expansion of a subset $S \subset X(k)$ with respect to M is

$$\Phi(S) = \mathop{\mathbb{E}}_{v \sim \pi_k|_S} \left[M(v, X(k) \setminus S) \right],$$

where

$$M(v, X(k) \setminus S) = \sum_{y \in X(k) \setminus S} M(v, y)$$

and $M(v, y)$ denotes the transition probability from v to y.

Before we prove the strong connections between links and expansion, we need to introduce an important property of HD-walks, monotonic eigenvalue decay.

Definition 4.6.2 (Monotonic HD-walk)**.** Let (X, Π) be a (δ, γ)-eposet. We call an HD-walk M monotonic if its approximate eigenvalues $\lambda_i(M)$ (given in Corollary 4.4.5) are non-increasing.

Most HD-walks of interest (e.g. pure walks, partial-swap walks on simplicial or q-simplicial complexes, etc.) are monotonic. This property will be crucial to understanding expansion. To start, let's see how it allows us to upper bound the expansion of links.

Theorem 4.6.3 (Local Expansion vs Global Spectra)**.** *Let (X, Π) be a (δ, γ)-eposet and M be a k-dimensional monotonic HD-walk. Then for all $0 \le i \le k$ and $\tau \in X(i)$:*

$$\Phi(X_\tau) \in 1 - \lambda_i(M) \pm c\gamma,$$

where $c \le O\left(\frac{k^5 R_{max}^2 (h(M)+k)h(M)w(M)}{\delta_{k-i}^k (1-\delta_{i-1})}\right)$.

The key to proving Theorem 4.6.3 is to show that the weight of an i-link lies almost entirely on level i of the HD-Level-Set Decomposition. To show this, we'll rely another connection between regularity and eposet parameters for non-lazy posets.

Claim 4.6.4. Let (X, Π) be a d-dimensional (δ, γ)-eposet. Then for every $1 \le k \le d$ and $0 \le i \le k$, the following approximate relation between the eposet and regularity parameters holds:

$$\lambda_i(N_k^1) \in \frac{R(k,i)}{R(k+1,i)} \pm \left(\gamma_{k-i}^k + R(k,i)\delta_{k-i}^k \gamma\right)$$

where we recall $\lambda_i(N_k^1) = 1 - \prod_{j=i}^{k} \delta_j$.

We prove this relation in Section 4.8. With this in hand, we can show links project mostly onto their corresponding level.

Lemma 4.6.5. *Let (X, Π) be a d-dimensional (δ, γ)-eposet with $\gamma \le O\left(\frac{\max_i\{\delta_i(1-\delta_{i-1})\}}{k^5 R_{max}^2}\right)$. Then for all $0 \le i \le k < d$ and $\tau \in X(i)$, $\mathbb{1}_{X_\tau}$ lies almost entirely in V_k^i. That is for all $j \ne i$:*

$$\left|\frac{\langle \mathbb{1}_{X_\tau,i}, \mathbb{1}_{X_\tau,j}\rangle}{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle}\right| \le O\left(\frac{k^3 R_{max}}{\delta_{k-i}^k (1-\delta_{i-1})}\gamma\right).$$

Proof. We'll show that the expansion of $\mathbb{1}_{X_\tau}$ with respect to the upper walk N_k^1 is almost exactly $1 - \lambda_i(N_k^1)$, which implies most of the weight must lie on V_i^k. We'll start by analyzing the expansion of $\mathbb{1}_{X_\tau}$ through a simple combinatorial argument. First, since D and U are adjoint we have:

$$\begin{aligned}\bar{\Phi}(\mathbb{1}_{X_\tau}) &= \frac{\langle \mathbb{1}_{X_\tau}, D_{k+1}U_k\mathbb{1}_{X_\tau}\rangle}{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle} \\ &= \frac{\langle U_k\mathbb{1}_{X_\tau}, U_k\mathbb{1}_{X_\tau}\rangle}{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle}.\end{aligned}$$

The trick is now to notice $\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle = R(k,i)\pi_i(\tau)$, and $\langle U_k\mathbb{1}_{X_\tau}, U_k\mathbb{1}_{X_\tau}\rangle = \frac{R(k,i)^2}{R(k+1,i)}\pi_i(\tau)$. As a result, applying Claim 4.6.4 gives:

$$\bar{\Phi}(\mathbb{1}_{X_\tau}) \in \lambda_i(N_k^1) \pm (c\gamma + R(k,i)\gamma),$$

for $c \leq k\gamma$. To see why this implies that most of the weight lies on V_i^k, note that we can also unfold the expansion of $\mathbb{1}_{X_\tau}$ in terms of the HD-Level-Set decomposition:

$$\begin{aligned}\bar{\Phi}(\mathbb{1}_{X_\tau}) &= \frac{1}{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle}\sum_{j=0}^{i}\langle \mathbb{1}_{X_\tau}, N_k^1\mathbb{1}_{X_\tau,j}\rangle \\ &\in \frac{1}{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle}\sum_{j=0}^{i}\lambda_i(N_k^1)\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,j}\rangle \pm c_2\gamma\end{aligned}$$

where $c_2 \leq \frac{k\sqrt{k}}{\rho_{\min}}$. Recall from Corollary 4.5.1 that for the set I of indices with negative inner product, it holds that $-\sum_{j\in I}\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,j}\rangle \leq O\left(\frac{k^3}{\rho_{\min}}\gamma\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle\right)$. Moreover, the positive inner products (i.e. the indices not in I) must sum to at least $\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle$. Then if there exists some $j \neq i$ such that $\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,j}\rangle > c_3\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle$ for large enough $c_3 \leq O\left(\frac{1}{\delta_{k-i}^k(1-\delta_{i-1})}\cdot\left(\frac{k^3}{\rho_{\min}}\gamma + R(k,i)\gamma\right)\right)$, the non-expansion would be strictly larger than $\lambda_i(N_k^1) + c\gamma + R(k,i)\gamma$ giving the desired contradiction (note that $(1-\delta_{i-1})\delta_{k-1}^k$ is the gap between the $i-1$st and ith approximate eigenvalue). The form in the theorem statement

then follows from applying Claim 4.4.4. □

We note that the above is the only result in our work that truly relies on non-laziness (it is used only to replace ρ with regularity in all other results). It is possible to recover the upper bound in Theorem 4.6.3 for general eposets via arguments used in [38], but the lower bound remains open for concentrated posets. With that in mind, we now prove Theorem 4.6.3.

Proof of Theorem 4.6.3. By the previous lemma, we have

$$\left|\frac{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,j}\rangle}{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle}\right| \leq O\left(\frac{1}{\delta_{k-i}^k(1-\delta_{i-1})} \cdot \left(\frac{k^3}{\rho_{\min}}\gamma + R(k,i)\gamma\right)\right).$$

Expanding out $\bar{\Phi}(\mathbb{1}_{X_\tau})$ then gives:

$$\begin{aligned}
\bar{\Phi}(\mathbb{1}_{X_\tau}) &= \frac{1}{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle}\sum_{j=0}^{i}\langle \mathbb{1}_{X_\tau}, M\mathbb{1}_{X_\tau,j}\rangle \\
&\leq \frac{1}{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle}\sum_{j=0}^{i}\lambda_i(M)\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,j}\rangle + c_2\gamma \\
&\leq \lambda_i(M)\frac{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,i}\rangle}{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau}\rangle} + err_1 \\
&\leq \lambda_i(M) + err_2.
\end{aligned}$$

where $c_2, err_1, err_2 \leq O\left(\frac{k}{\delta_{k-i}^k(1-\delta_{i-1})}\left(\frac{k^2(h(M)+k)h(M)w(M)}{\rho_{\min}}\gamma + R(k,i)\gamma\right)\right)$ and the last step follows from the approximate orthogonality. As usual, the form in the theorem statement then follows from applying Claim 4.4.4. □

Altogether, we've seen that for sufficiently nice expanding posets, the expansion of any i-link with respect to an HD-walk is almost exactly $1-\lambda_i(M)$. Since HD-walks are generally poor expanders (have large $\lambda_1(M)$), Theorem 4.6.3 implies that links are examples of small, non-expanding sets. Following BHKL, we'll now prove a converse to this result: any non-expanding set must be explained by some structure inside links. To

help give a precise statement, we first recall BHKL's notion of Stripped Threshold Rank (specialized to eposets for convenience).

Definition 4.6.6 (Stripped Threshold Rank [38])**.** Let (X, Π) be a (δ, γ)-eposet and M a k-dimensional HD-walk with γ small enough that Theorem 4.3.2 implies the HD-Level-Set Decomposition has a corresponding decomposition of disjoint eigenstrips $C_k = \bigoplus W_k^i$. The ST-Rank of M with respect to η is the number of strips containing an eigenvector with eigenvalue at least η:

$$R_\eta(M) = |\{W_k^i : \exists f \in V^i, Mf = \lambda f, \lambda > \eta\}|.$$

We often write just R_η when M is clear from context.

With this in mind, we'll show a converse to Theorem 4.6.3 in both ℓ_2 and ℓ_∞ senses (respectively that any non-expanding set must have high variance over links, and must be more concentrated than expected in some particular link). It is convenient to express these results through their contrapositives: that pseudorandom sets expand. The proof is the same as in [38] for simplicial complexes, but we include it for completeness.

Theorem 4.6.7. *Let (X, Π) (δ, γ)-eposet, M a k-dimensional, monotonic HD-walk, and γ small enough that the eigenstrip intervals of Theorem 4.3.2 are disjoint. For any $\eta > 0$, let $r = R_\eta(M) - 1$. Then the expansion of a set $S \subset X(k)$ of density α is at least:*

$$\Phi(S) \geq 1 - \alpha - (1-\alpha)\eta - \sum_{i=1}^{r} (\lambda_i(M) - \eta) R(k,i) \varepsilon_i - c\gamma$$

where S is $(\varepsilon_1, \ldots, \varepsilon_r)$-pseudorandom and $c \leq O\left(\frac{k^5 R_{max}^2 (h(M)+k) h(M) w(M)}{\max_i \{\delta_i (1-\delta_{i-1})\}}\right)$.

Proof. Let $\mathbb{1}_S = \mathbb{1}_{S,0} + \ldots + \mathbb{1}_{S,k}$ be the HD-Level-Set Decomposition of the indicator of

S. By linearity of the inner product, we may then write:

$$\begin{aligned}
\Phi(S) &= 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]}\langle \mathbb{1}_{X_\tau}, M\mathbb{1}_{X_\tau}\rangle \\
&= 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{j=0}^{k}\langle \mathbb{1}_S, M\mathbb{1}_{S,j}\rangle \\
&= 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{j=0}^{k}\lambda_j(M)\langle \mathbb{1}_S, \mathbb{1}_{S,j}\rangle - \frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{j=0}^{k}\langle \mathbb{1}_S, \Gamma_j\mathbb{1}_{S,j}\rangle
\end{aligned}$$

where $\|\Gamma_j\| \leq O\left((h(M)+k)h(M)\frac{w(M)}{\rho_{\min}}\right)$. The trick is now to notice we can bound the righthand error term using Cauchy-Schwarz:

$$\begin{aligned}
\left|\frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{j=0}^{k}\langle \mathbb{1}_S, \Gamma_j\mathbb{1}_{S,j}\rangle\right| &\leq \frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{j=0}^{k}|\langle \mathbb{1}_S, \Gamma_j\mathbb{1}_{S,j}\rangle| \\
&\leq \frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{j=0}^{k}\|\Gamma_j\|\|\mathbb{1}_S\|\|\mathbb{1}_{S,j}\| \\
&\leq c\gamma\frac{\|\mathbb{1}_S\|}{\mathbb{E}[\mathbb{1}_S]}\sum_{j=0}^{k}\|\mathbb{1}_{S,j}\| \\
&\leq c_1\gamma,
\end{aligned}$$

where $c \leq O\left((h(M)+k)h(M)\frac{w(M)}{\rho_{\min}}\right)$ and $c_1 \leq O(\sqrt{k}c)$ by Equation (4.1). Since M is a monotonic walk, we can further write:

$$\begin{aligned}
\Phi(S) &\geq 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{i=0}^{r}\lambda_i(M)\langle \mathbb{1}_S, \mathbb{1}_{S,i}\rangle - \frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{i=r+1}^{k}\lambda_i(M)\langle \mathbb{1}_S, \mathbb{1}_{S,i}\rangle - c_1\gamma \\
&\geq 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{i=0}^{r}\lambda_i(M)\langle \mathbb{1}_S, \mathbb{1}_{S,i}\rangle - \frac{\eta}{\mathbb{E}[\mathbb{1}_S]}\sum_{i=r+1}^{k}\langle \mathbb{1}_S, \mathbb{1}_{S,i}\rangle - c_2\gamma \\
&= 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{i=0}^{r}\lambda_i(M)\langle \mathbb{1}_S, \mathbb{1}_{S,i}\rangle - \eta\left(1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{i=0}^{r}\langle \mathbb{1}_S, \mathbb{1}_{S,i}\rangle\right) - c_2\gamma \\
&= 1 - \eta - \frac{1}{\mathbb{E}[\mathbb{1}_S]}\sum_{i=0}^{r}(\lambda_i(M)-\eta)\langle \mathbb{1}_S, \mathbb{1}_{S,i}\rangle - c_2\gamma
\end{aligned}$$

$$= 1 - \eta - (1-\eta)\alpha - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=1}^{r} (\lambda_i(M) - \eta) \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle - c_2\gamma,$$

where $c_2 \leq O\left(k^2(h(M)+k)h(M)\frac{w(M)}{\rho_{\min}}\right)$. To justify the second inequality, observe that for any $r < i \leq k$ such that $\langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle \geq 0$, replacing $\lambda_i(M)$ with η is valid. For the set I of $r < i \leq k$ with negative inner product, Corollary 4.5.1 implies that the sum over I is $O(k^3\gamma/\rho_{\min})$, so the inequality remains valid by absorbing the small error into c_2. Applying Corollary 4.5.4 to bound $\langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle$ then gives the ℓ_2-variant result, Theorem 4.5.7 gives the ℓ_∞-variant, and Claim 4.4.4 gives the form given in the theorem statement. □

We note that Theorem 4.6.7 recovers the analogous result for simplicial complex in [38] by plugging in the appropriate value $R(k,i) = \binom{k}{i}$. BHKL also prove this special case is tight in two senses. First, they show that if one wants to retain linear dependence on the pseudorandomness parameter ε, Theorem 4.6.7 is tight in both the ℓ_2 and ℓ_∞-regimes. Second, they show that the dependence on k is necessary in the ℓ_2-regime, even if we allow sub-optimal dependence on ε. In the next section, we'll generalize this result to q-simplicial complexes as well. In both cases the proofs are highly structural and depend on the underlying structure of the poset—it remains an interesting open problem whether this bound is tight for all poset structures.

4.7 The Grassmann and q-eposets

In this section, we examine the specification of our results on eposets to expanding subsets of the Grassmann poset. We show that our analysis is tight in this regime via a classic example of a small non-expanding set in the Grassmann graphs called co-links.

4.7.1 Spectra

We'll start by examining the spectrum of HD-walks on the Grassmann and q-eposets. We'll focus our attention in this section on the most widely used walks in the literature,

the canonical and partial-swap walks. To start, recall that the Grassmann poset itself is sequentially differential with parameters

$$\delta_i = \frac{(q^i - 1)(q^{n-i+1} - 1)}{(q^{i+1} - 1)(q^{n-i} - 1)}. \tag{4.2}$$

Plugging this into Proposition 4.4.2 gives a nice exact form for the spectra of canonical walks.

Corollary 4.7.1 (Grassmann Poset N_k^j Spectra)**.** *Let $X = G_q(n, d)$ be the Grassmann Poset, $k + j \leq d$, and $f_\ell = U_\ell^k g_\ell$ for some $g_\ell \in H^\ell$. Then:*

$$N_k^j f_\ell = \lambda_\ell f_\ell,$$

where,

$$\lambda_\ell = q^{\ell j} \frac{\binom{k+j-\ell}{j}_q}{\binom{k+j}{j}_q} \frac{\binom{n-k-\ell}{j}_q}{\binom{n-k}{j}_q} \approx q^{-\ell j}.$$

Proof. By Proposition 4.4.2 we have that

$$\begin{aligned}
\lambda_\ell(N_k^j) &= \prod_{s=1}^{j} \left(1 - \prod_{i=\ell}^{k-s+j} \delta_i\right) \\
&= \prod_{s=1}^{j} \left(1 - \prod_{i=\ell}^{k-s+j} \frac{(q^i - 1)(q^{n-i+1} - 1)}{(q^{i+1} - 1)(q^{n-i} - 1)}\right).
\end{aligned}$$

The result then follows from telescoping the interior product and simplifying:

$$\begin{aligned}
&= \prod_{s=1}^{j} \left(1 - \frac{(q^\ell - 1)(q^{n-\ell+1} - 1)}{(q^{k-s+j+1} - 1)(q^{n+s-k-j} - 1)}\right) \\
&= q^{\ell j} \left(\prod_{s=1}^{j} \frac{\left(q^{k+j-s-\ell+1} - 1\right)}{(q^{k+j-s+1} - 1)}\right) \left(\prod_{s=1}^{j} \frac{\left(q^{n+s-k-j-\ell} - 1\right)}{(q^{n+s-k-j} - 1)}\right) \\
&= q^{\ell j} \frac{\binom{k+j-\ell}{j}_q}{\binom{k+j}{j}_q} \frac{\binom{n-k-\ell}{j}_q}{\binom{n-k}{j}_q}
\end{aligned}$$

as desired. □

This recovers a very simple proof of classical results to this effect (see e.g. [106]). An analogous computation gives an approximate bound on the spectrum of N_k^j on q-eposets as well.

Corollary 4.7.2 (q-eposets N_k^j Spectra)**.** *Let (X, Π) be a d-dimensional γ-q-eposet with $\gamma \leq q^{-\Omega(k^2)}$, $k + j \leq d$, and $f_\ell = U_\ell^{k-1} g_\ell$ for some $g_\ell \in H^\ell$. Then:*

$$\|N_k^j f_\ell - \frac{\binom{k+j-\ell}{j}_q}{\binom{k+j}{j}_q} f_\ell\| \leq O\left(j(j+k)\binom{k}{\ell}_q\right)\gamma\|f_\ell\|$$

Note that for small enough γ, Theorem 4.3.2 implies that the true spectra is then concentrated around these values as well. It is also worth noting that these eigenvalues are, as one would expect, the natural q-analog of the corresponding eigenvalues on simplicial complexes.

It turns out that this fact will carry over to the important class of partial-swap walks as well. Partial-swap walks on simplicial complexes were originally analyzed by AJT [9], who showed they can be written as a hypergeometric combination of canonical walks. Their proof is specific to the structure of simplicial complexes, and some work is required to generalize their ideas to the Grassmann case. Following the overall proof strategy of AJT, it will be helpful to first show that the canonical walks themselves can be written as an expectation of swap walks over a q-hypergeometric distribution, and then use the q-binomial inversion theorem to derive the desired result.

Lemma 4.7.3 (q-analog of [9, Lemma 4.11])**.** *Let (X, Π) be a pure, measured q-simplicial complex. Then:*

$$N_k^j = \sum_{i=0}^{j} q^{i^2} \frac{\binom{j}{i}_q \binom{k}{k-i}_q}{\binom{k+j}{k}_q} S_k^i$$

Proof. We follow the structure and notation of [9, Lemma 4.11]. Assume that the canonical

walk starts at a subspace $V \in X(k)$, and walks up to $W \in X(k+j)$. We wish to analyze the probability that upon walking back down to level k, a subspace V' with intersection $k-i$ is chosen, that is:

$$dim(V \cap V') = k - i.$$

Let such an event be denoted $\mathcal{E}_i(W)$. It follows from elementary q-combinatorics (see e.g. [82, Lemma 9.3.2]) that

$$\mathop{\mathbb{P}}_{V' \subset W}[\mathcal{E}_i(W) \mid W] = q^{i^2} \frac{\binom{j}{i}_q \binom{k}{k-i}_q}{\binom{k+j}{k}_q},$$

where $V' \in X(k)$ is drawn uniformly from the k-dimensional subspaces of W. In essence, we wish to relate this process to the swap walk S_k^i. To do so, note that while the swap walk (as defined) only walks up to $X(k+i)$, walking up to $X(k+j)$ and conditioning on intersection i, a process called the *i-swapping j-walk* by [9], is exactly the same due to symmetry (via the regularity condition, see [9][Proposition 4.9] for a more detailed explanation). Thus consider the i-swapping j-walk, and let T_i' denote the variable standing for the subspace chosen by the walk. Conditioned on picking the same W as the canonical walk in its ascent, we may relate T_i' to the canonical walk:

$$\mathbb{P}[T_i' = T \mid W] = \mathbb{P}[V' = T \mid W \text{ and } \mathcal{E}_i(W)]$$

We may now decompose the canonical walk by intersection size:

$$\begin{aligned} N_k^j(V,T) &= \sum_{i=0}^{j} \sum_{W \in X(k+j)} \mathbb{P}[W]\, \mathbb{P}[\mathcal{E}_i(W) \mid W]\, \mathbb{P}[V' = T \mid W \text{ and } \mathcal{E}_i(W)] \\ &= \sum_{i=0}^{j} \sum_{W \in X(k+j)} q^{i^2} \frac{\binom{j}{i}_q \binom{k}{k-i}_q}{\binom{k+j}{k}_q} \mathop{\mathbb{E}}_{W \supset V}[\mathbb{P}[V' = T \mid W \text{ and } \mathcal{E}_i(W)]] \end{aligned}$$

$$= \sum_{i=0}^{j} \sum_{W \in X(k+j)} q^{i^2} \frac{\binom{j}{i}_q \binom{k}{k-i}_q}{\binom{k+j}{k}_q} \mathop{\mathbb{E}}_{W \supset V} [\mathbb{P}[T'_i = T \mid W \,]]$$
$$= \sum_{i=0}^{j} \sum_{W \in X(k+j)} q^{i^2} \frac{\binom{j}{i}_q \binom{k}{k-i}_q}{\binom{k+j}{k}_q} \mathbb{P}[T'_i = T]$$
$$= \sum_{i=0}^{j} \sum_{W \in X(k+j)} q^{i^2} \frac{\binom{j}{i}_q \binom{k}{k-i}_q}{\binom{k+j}{k}_q} S_k^i(V, T)$$

□

This results in the q-analog of the analogous result on simplicial complexes [9, Lemma 4.11]. To recover the analogous statement writing partial-swap walks in terms of canonical walks, we can now apply a q-Binomial inversion theorem.

Lemma 4.7.4 (q-Binomial Inversion (Theorem 2.1 [365]))**.** *Suppose* $\{a_i\}_{i \geq 1}$, $\{b_i\}_{i \geq 1}$ *are two sequences. If:*

$$a_j = \sum_{i=1}^{j} (-1)^i \binom{j}{i}_q b_i,$$

then

$$b_j = \sum_{i=1}^{j} (-1)^i q^{\binom{j-i}{2}} \binom{j}{i}_q a_i$$

We note that [365, Theorem 2.1] is stated in slightly more generality in the original work, but the above lemma is an immediate application. With this in hand, we can finally prove that swap walks on the Grassmann poset are indeed HD-walks:

Proposition 4.7.5. *Let* (X, Π) *be a weighted pure* q*-simplicial complex. Then for* $k+j \leq d$*:*

$$S_k^j = \frac{1}{q^{j^2} \binom{k}{k-j}_q} \sum_{i=0}^{j} (-1)^{j-i} q^{\binom{j-i}{2}} \binom{j}{i}_q \binom{k+i}{i}_q N_k^i,$$

and similarly,

$$J_q(n,k,t) = S_k^{k-t} = \frac{1}{q^{(k-t)^2} \binom{k}{t}_q} \sum_{i=0}^{k-t} (-1)^{k-t-i} q^{\binom{k-t-i}{2}} \binom{k-t}{i}_q \binom{k+i}{i}_q N_k^i$$

Proof. The proof is an easy application of Lemma 4.7.4 and the q-Binomial theorem. In particular, for any $V, V' \in X(k)$, let

$$a_i = (-1)^i q^{i^2} \binom{k}{k-i}_q S_k^i(V, V').$$

Noting that $N_0^j = S_0^j = I$, Lemma 4.7.3 gives the following equality:

$$\binom{k+j}{k}_q \left(N_k^j(V, V') - \frac{1}{\binom{k+j}{k}_q} I(V, V') \right) = \sum_{i=1}^{j} (-1)^i \binom{j}{i}_q a_i.$$

Setting the second sequence $\{b_i\}_{i \geq 1}$ to

$$b_i = \binom{k+i}{k}_q \left(N_k^i(V, V') - \frac{1}{\binom{k+i}{k}_q} I(V, V') \right),$$

Lemma 4.7.4 then implies:

$$\begin{aligned} q^{j^2} \binom{k}{k-j}_q S_k^j(V, V') &= \sum_{i=1}^{j} (-1)^{j-i} q^{\binom{j-i}{2}} \binom{k+i}{k} \left(N_k^i(V, V') - \frac{1}{\binom{k+i}{k}_q} I(V, V') \right) \\ &= \sum_{i=1}^{j} (-1)^{j-i} q^{\binom{j-i}{2}} \binom{k+i}{k} N_k^i(V, V') - \sum_{i=1}^{j} (-1)^{j-i} q^{\binom{j-i}{2}} I(V, V') \\ &= \sum_{i=0}^{j} (-1)^{j-i} q^{\binom{j-i}{2}} \binom{k+i}{k} N_k^i(V, V') \end{aligned}$$

where the last step follows from the q-Binomial theorem. □

Once again, we note that this is unsurprisingly the q-analog of the analogous statement on simplicial complexes (see [9, Corollary 4.13]). Finally, we'll use this to show that the eigenvalues of partial-swap walks on q-simplicial complexes are given by the natural q-analog of the simplicial complex case.

Corollary 4.7.6. *Let (X, Π) be a d-dimensional γ-q-eposet with γ sufficiently small,*

$k + j \leq d$, and $f_\ell = U_\ell^k g_\ell$ for some $g_\ell \in H^\ell$. Then:

$$\|S_k^j f_\ell - \frac{\binom{k-j}{\ell}_q}{\binom{k}{\ell}_q} f_\ell\| \leq O\left(\left(\frac{q}{q-1}\right)^{\min(j,k-j)+2} k^2 \binom{k}{\ell}_q\right) \gamma \|f_\ell\|$$

Proof. This follows from combining Corollary 4.4.5, Corollary 4.7.1, and Proposition 4.7.5. Let $t = k - j$. In particular, it is sufficient to note that (in the notation of Corollary 4.4.5):

$$\begin{aligned}\sum_{Y \in \mathcal{Y}} \alpha_Y \lambda_{Y,\delta,\ell} &= \frac{1}{q^{(k-t)^2}\binom{k}{t}_q} \sum_{i=0}^{k-t} (-1)^{k-t-i} q^{\binom{k-t-i}{2}} \binom{k-t}{i}_q \binom{k+i-\ell}{i}_q \\ &= \frac{\binom{k-j}{\ell}_q}{\binom{k}{\ell}_q}.\end{aligned}$$

and further that:

$$\begin{aligned}w(S_k^j) &= \frac{1}{q^{j^2}\binom{k}{k-j}_q} \sum_{i=0}^{j} q^{\binom{j-i}{2}} \binom{j}{i}_q \binom{k+i}{i}_q \\ &\leq \frac{q^{jk}}{q^{j^2}\binom{k}{k-j}_q} \sum_{i=0}^{j} q^{-i} \\ &\leq \left(\frac{q}{q-1}\right)^{\min(j,k-j)+1}\end{aligned}$$

□

Again, since the swap walks are self-adjoint Theorem 4.3.2 implies that for small enough γ the true spectra is closely concentrated around these values as well. It is worth noting that if the above analysis is repeated using the exact eposet parameters for the Grassmann (see Equation (4.2)), this recovers the standard eigenvalues of the Grassmann graphs (see e.g. [106]).

4.7.2 Pseudorandom Functions and Small Set Expansion

With an understanding of the spectra of HD-walks on q-simplicial complexes, we move to studying its combinatorial structure. By direct computation, it is not hard to show that on q-eposets, $\rho_\ell^k = \frac{1}{\binom{k}{\ell}_q}$ (Claim 4.4.4 would only imply this is approximately true). As a result, we get a level-i inequality for q-simplicial complexes that is the natural q-analog of BHKL's inequality for basic simplicial complexes.

Theorem 4.7.7. *Let (X, Π) be a γ-q-eposet with $\gamma \leq q^{-\Omega(k^2)}$, and let $f : C_k \to \mathbb{R}$ be any function on k-faces with HD-Level-Set Decomposition $f = f_0 + \ldots + f_k$. If f is $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_\infty$-pseudorandom, then for all $1 \leq i \leq \ell$:*

$$|\langle f, f_i \rangle| \leq \left(\binom{k}{i}_q + c\gamma \right) \varepsilon_i^2 + c\gamma \|f\|^2.$$

If f additionally has i-local constant sign or is $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_2$-pseudorandom, then

$$|\langle f, f_i \rangle| \leq \binom{k}{i}_q \varepsilon_i |\mathbb{E}[f]| + c\gamma \|f\|^2$$

where in both cases $c \leq q^{O(k^2)}$

For large enough q, γ^{-1}, this result is exactly tight. The key to showing this fact is to examine a local structure unique to the Grassmann called *co-links*. The co-link of an element $W \in X(k')$, is all of the subspaces *contained in* W:

$$\bar{X}_W = \{V \in X(k) : V \subseteq W\}.$$

Just like links, co-links of dimension i (that is $k' = d - i$) also come through levels 0 through i of the complex, although this is somewhat trickier to see.

Lemma 4.7.8 (HD-level-set decomposition of co-links)**.** *Let $X = G_q(d, k)$ and $\mathcal{S} = \overline{X}_W$*

be a co-link of dimension i for $W \in X(d-i)$. Then, we have

$$\mathbb{1}_{\mathcal{S}} \in V_k^0 \oplus \cdots \oplus V_k^i.$$

Proof. Since we know that $V_k^0 \oplus \cdots \oplus V_k^i = \mathrm{Im}(U_i^k C_i)$ (see e.g. [111]), all we need to do is to show that there exists an $f \in C_i$ such that $\mathbb{1}_{\mathcal{S}} = U_i^k f$. More specifically, we can write $f = \sum_{U \in X(i)} \alpha_U \mathbb{1}_U$. Then, we have

$$(U_i^k f)(V) = \sum_{U \in X(i)} \alpha_U (U_i^k \mathbb{1}_U)(V) = \frac{1}{R(k,i)} \sum_{U \in X(i), U \subset V} \alpha_U.$$

Suppose $\alpha_U = g(\dim(U \cap W))$ for some function $g : \{0, \ldots, i\} \to \mathbb{R}$. We will prove that there exists a unique g that satisfies the desired equations.

Consider the dimension of $V \cap W$. If $V \subset W$, i.e., $\dim(V \cap W) = k$, then for all $U \in X(i)$ s.t. $U \subset V$, $\dim(U \cap W) = i$. Then, for all $V \subset W$ we must have:

$$U_i^k \mathbb{1}_V = \frac{1}{R(k,i)} \sum_{U \in X(i), U \subset V} g(i) = g(i) = 1.$$

On the other hand, consider $V \not\subset W$. In this case we must have $\dim(V \cap W) = k - j$ for some $i \geq j > 0$ and further that $\dim(U \cap W) \in \{i - j, \ldots, i\}$ for all $U \in X(i)$ s.t. $U \subset V$. This gives the following set of linear equations:

$$U_i^k \mathbb{1}_V = \sum_{\ell = i-j}^{i-1} c_{j,\ell} g(\ell) + c_{j,i} = 0 \quad \forall 1 \leq j \leq i,$$

where $c_{j,\ell} := R(k,i)^{-1} \cdot |\{U \in X(i) : U \subset V, \dim(U \cap W) = \ell, \dim(V \cap W) = k - j\}|$ is a constant for all $\ell \in \{i-j, \ldots, i\}$. Since this system can be written as a triangular form with positive diagonal, it is invertible and there exists a unique solution for $g(0), \ldots, g(i-1)$ as desired. By definition, such a solution must satisfy $f = \sum_{U \in X(i)} g(\dim(U \cap W)) \mathbb{1}_U$, so we

have constructed $f \in C_i$ such that $U_i^k f = \mathbb{1}_{\mathcal{S}}$, which completes the proof of the claim. $\square$

Using this fact, we can show that our level-i inequality is exactly tight.

Proposition 4.7.9. *Let $X = G_q(d,k)$ be the Grassmann poset. For any $i \leq k \in \mathbb{N}$ and $c < 1$, there exist large enough q, d and a set $S \subset X(k)$ such that*

$$\langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle > c \binom{k}{i}_q \varepsilon_i \langle \mathbb{1}_S, \mathbb{1}_S \rangle$$

where S is (i, ε_i)-pseudorandom.

Proof. The proof goes through examining a "co-link" of dimension i, that is for $W \in X(d-i)$:

$$\bar{X}_W = \{V \in X(k) : V \subset W\}.$$

For simplicity, let $S \coloneqq \bar{X}_W$. The density of the co-link S in any j-link X_V is:

$$\alpha_j = \frac{(q^{d-i-j}-1)\dots(q^{d-k+1-i}-1)}{(q^{d-j}-1)\dots(q^{d-k+1}-1)} = q^{-i(k-j)} + o_{q,d}(1).$$

The idea is now to examine the (non)-expansion of the co-link with respect to the lower walk $U_{k-1}D_k$. By direct computation, the probability of returning to $\bar{X}_W$ after moving to a $(k-1)$-dimensional subspace is exactly:

$$\bar{\Phi}(\bar{X}_W) = \frac{q^{d-i} - q^{k-1}}{q^d - q^{k-1}} = q^{-i} \pm q^{-\Omega(d)} \tag{4.3}$$

On the other hand, by Proposition 4.4.2 the approximate eigenvalues of the lower walk are given by

$$\lambda_j = \frac{q^{k-j}-1}{q^k - 1} = q^{-j} - O(q^{-k})$$

Since a dimension-i co-link has no projection onto levels $i+1$ through k, we can also write

the non-expansion as:

$$\bar{\Phi}(\bar{X}_W) = \frac{1}{\langle \mathbb{1}_S, \mathbb{1}_S \rangle} \sum_{j=0}^{i} q^{-j} \langle \mathbb{1}_S, \mathbb{1}_{S,j} \rangle - O(q^{-k})$$

for large enough q, d. Combining this with our previous formula for the non-expansion in Equation (4.3), we get that there exists a universal constant c' such that for large enough q and d, $\mathbb{1}_{\bar{X}_W}$ cannot have more than a $\frac{c'}{q}$ fraction of its mass on levels 1 through $i-1$. Finally, noticing that:

$$\binom{k}{i}_q \alpha_i = 1 + o_q(1)$$

we have

$$\frac{\langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle}{\langle \mathbb{1}_S, \mathbb{1}_S \rangle} \geq \frac{q - c'}{q} \geq c \binom{k}{i}_q \alpha_i$$

since the latter is strictly bounded away from 1 for large enough q. This completes the result since $\bar{X}_W$ is (α_i, i)-pseudorandom. □

We'll close the section by giving an immediate application of Theorem 4.7.7 to the expansion of pseudorandom sets, and briefly discuss connections with the proof of the 2-2 Games Conjecture and algorithms for unique games. Namely, as corollary of Theorem 4.7.7, we show that for both the canonical and partial-swap walks, sufficiently pseudorandom functions expand near perfectly.

Corollary 4.7.10 (q-eposets Edge-Expansion)**.** *Let (X, Π) be a d-dimensional γ-q-eposet, $S \subset X(k)$ a subset whose indicator function $\mathbb{1}_S$ is $(\varepsilon_1, \ldots, \varepsilon_\ell)$-pseudorandom. Then the edge expansion of S with respect to the canonical walk N_k^j is bounded by:*

$$\Phi_{\pi_k}(N_k^j, S) \geq 1 - \mathbb{E}[\mathbb{1}_S] - \sum_{i=1}^{\ell} \frac{\binom{k+j-i}{j}_q}{\binom{k+j}{j}_q} \binom{k}{i}_q \varepsilon_i - q^{-(\ell+1)j} - q^{O(k^2)}\gamma$$

Further, the edge expansion of S with respect to the partial-swap walk S_k^j is bounded by:

$$\Phi_{\pi_k}(S_k^j, S) \geq 1 - \mathbb{E}[\mathbb{1}_S] - \sum_{i=1}^{\ell} \binom{k-j}{i}_q \varepsilon_i - q^{-(\ell+1)j} - q^{O(k^2)}\gamma$$

Note that S_k^j on q-eposets is a generalization of the Grassmann Graphs (and are equivalent when X is the Grassmann Poset). While our definition of pseudorandomness is weaker than that of [255] and therefore necessarily depends on the dimension k, we take the above as evidence that the framework of expanding posets may be important for making further progress on the Unique Games Conjecture. In particular, combined with recent works removing this k-dependence on simplicial complexes [39, 187], it seems plausible that the framework of expanding posets may lead to a more general understanding of the structure underlying the unique games conjecture.

4.8 Eposet Parameters and Regularity

In this section we will discuss connections between notions of regularity, the averaging operators, and eposet parameters. To start, we'll show that downward and middle regularity (which are defined only on adjacent levels of the poset) imply extended regularity between any two levels.

Proposition 4.8.1. *Let (X, Π) be a d-dimensional regular measured poset. Then for any $i < k \leq d$, there exist regularity constant $R(k, i)$ such that for any $x_k \in X(k)$, there are exactly $R(k, i)$ elements $x_i \in X(i)$ such that $x_k > x_i$.*

Proof. Given any element $x_k \in X(k)$, downward regularity promises there are exactly $\prod_{j=i+1}^{k} R(j)$ unique chains $x_k < x_{k-1} < \ldots < x_{i+1} < x_i$. By middle regularity, any fixed $x_i \in X(i)$ which appears in this fashion appears in exactly $m(k, i)$ chains. Noting that $x_i < x_k$ if and only if x_i appears in such a chain, the total number of $x_i < x_k$ must be

exactly:

$$R(k,i) = \frac{\prod_{j=i+1}^{k} R(j)}{m(k,i)}.$$

□

A similar argument shows that regularity allows the up operators to compose in the natural way.

Proposition 4.8.2. *Let (X, Π) be a d-dimensional regular measured poset. Then for any $i < k \leq d$ we have:*

$$U_i^k f(x_k) = \frac{1}{R(k,i)} \sum_{x_i < x_k} f(x_i)$$

Proof. Expanding out $U_i^k f(y)$ gives:

$$U_i^k f(x_k) = \frac{1}{\prod_{j=i+1}^{k} R(j)} \sum_{x_{k-1} < x_k} \cdots \sum_{x_i < x_{i+1}} f(x_i)$$

The number of times each x_i appears in this sum is exactly the number of chains starting at x_k and ending at x_i, so by middle regularity:

$$\begin{aligned} \frac{1}{\prod_{j=i+1}^{k} R(j)} \sum_{x_{k-1} < x_k} \cdots \sum_{x_{i+1} < x_i} f(x_i) &= \frac{m(k,i)}{\prod_{j=i+1}^{k} R(j)} \sum_{x_i < x_k} f(x_i) \\ &= \frac{1}{R(k,i)} \sum_{x_i < x_k} f(x_i). \end{aligned}$$

as desired. □

We'll now take a look at the connection between eposet parameters and regularity. It is convenient to first start with a lemma stating that non-laziness is equivalent to bounding the maximum transition probability of the lower walk.

Lemma 4.8.3. *Let (X, Π) be a d-dimensional measured poset. Then for any $0 < i \leq d$, the maximum laziness of the lower walk is also the maximum transition probability:*

$$\max_{\sigma \in X(i)} \left\{ \mathbb{1}_\sigma^T U_{i-1} D_i \mathbb{1}_\sigma \right\} = \max_{\sigma, \tau \in X(i)} \left\{ \mathbb{1}_\sigma^T U_{i-1} D_i \mathbb{1}_\tau \right\}.$$

Proof. Assume that $\tau \neq \sigma$. Then the transition probability from τ to σ is exactly

$$\begin{aligned} \mathbb{1}_\sigma^T U_{i-1} D_i \mathbb{1}_\tau &= \frac{\pi_\tau(\sigma \setminus \tau)}{R(i, i-1)} \\ &\leq \frac{1}{R(i, i-1)} \sum_{\tau \lessdot \sigma} \pi_\tau(\sigma \setminus \tau) \\ &= \mathbb{1}_\sigma^\tau U_{i-1} D_i \mathbb{1}_\sigma, \end{aligned}$$

which implies the result. □

We now prove our two claims relating the eposet parameters to regularity.

Claim 4.8.4. Let (X, Π) be a d-dimensional (δ, γ)-eposet. Then for every $1 \leq k \leq d$ and $0 \leq i \leq k$, the following approximate relation between the eposet and regularity parameters holds:

$$\lambda_i(N_k^1) \in \frac{R(k, i)}{R(k+1, i)} \pm \left(\gamma_{k-i}^k + R(k, i) \delta_{k-i}^k \gamma \right)$$

where we recall $\lambda_i(N_k^1) = 1 - \prod_{j=i}^{k} \delta_j$.

Proof. One of our main analytical tools so far has been the relation between the upper and lower walks given in Lemma 4.4.1:

$$\| D_{k+1} U_i^{k+1} - (1 - \delta_{k-i}^k) U_i^k - \delta_{k-i}^k U_{i-1}^k D_i \| \leq \gamma_{k-i}^k.$$

For this result, we'll actually need a refinement of this result given in [38, Lemma A.1]:[15]

$$D_{k+1}U_i^{k+1} = (1-\delta_{k-i}^k)U_i^k + \delta_{k-i}^k U_{i-1}^k D_i + \sum_{j=-1}^{k-i-1} U_{k-j-1}^k \Gamma_j U_i^{k-j-1} \tag{4.4}$$

where $\sum \|\Gamma_j\| \leq \gamma_{k-i}^k$. The idea is now to examine the "laziness" of the two sides of this equality. In other words, given a starting k-face τ, what is the probability that the resulting i-face σ satisfies $\sigma < \tau$?

To start, we'll argue that the laziness of the lefthand side is exactly $\frac{R(k,i)}{R(k+1,i)}$. This follows from noting that there are $R(k,i)$ i-faces σ satisfying $\sigma < \tau$, and $R(k+1,i)$ options after taking the initial up-step of the walk to $\tau' > \tau$. After the down-steps, the resulting i-face is uniformly distributed over these $R(k+1,i)$ options $\sigma < \tau'$, and since every $\sigma < \tau < \tau'$, all original $R(k,i)$ lazy options are still viable after the up-step to τ'.

Analyzing the right-hand side is a bit trickier. The initial term $(1-\delta_{k-i}^k)U_i^k$ is completely lazy, so it contributes exactly $(1-\delta_{k-i}^k) = \lambda_i(N_k^1)$. We'll break the second term into two steps: walking from $X(k)$ to $X(i)$ via U_i^k, then from $X(i)$ to $X(i)$ via the lower walk $U_{i-1}D_i$. Starting at a k-face τ, notice that after applying the down step U_i^k we are uniformly spread over $\sigma < \tau$. Computing the laziness then amounts to asking what the probability of staying in this set is after the application of UD, which one can naively bound by the maximum transition probability times the set size $R(k,i)$. By non-laziness, the maximum transition probability is at most γ (see Lemma 4.8.3).

The third term can be handled similarly. The first down step U_{k-j-1}^k spreads τ evenly across $\sigma < \tau$ in $X(k-j-1)$. The resulting i-face σ' after applying $\Gamma_j U_i^{k-j-1}$ is less than τ if and only if the intermediary $(k-j-1)$-face after applying Γ_j is less than τ, which is bounded by the spectral norm $\|\Gamma_j\|$.[16]

Putting everything together, since both sides of Equation (4.4) must have equivalent

[15]Formally the result is only stated for simplicial complexes in [38], but the same proof holds for eposets.

[16]We note that Γ_j is not stochastic, but it is self-adjoint and an easy exercise to see that the analogous reasoning still holds.

laziness, we get that $\lambda_i(N_k^1)$ must be within $\sum\|\Gamma_j\| + \delta_{k-i}^k R(k,i)\gamma$ as desired. $\square$

Claim 4.4.4 and Theorem 4.4.7 can both be proving an analogous theorem for the upper walk.

Claim 4.8.5 (Regularity and Upper Walk Spectrum). Let (X, Π) be a d-dimensional (δ, γ)-eposet. Then for any $j \leq i \leq k < d$, we have:

$$\lambda_j(N_i^{k-i}) \in \frac{R(i,j)}{R(k,i)} \pm err,$$

where $err \leq O\left(\frac{i^4 k^2 R_{\max}}{\delta_i(1-\delta_{i-1})}\gamma\right)$.

Proof. This follows almost immediately from the fact that i-links lie almost entirely on the ith eigenstrip (Lemma 4.6.5). In particular, it is enough to examine the expansion of i-links with respect to the upper canonical walk N_i^{k-i}. On the one hand, for any $j \leq i$ and $\tau \in X(j)$ we have:

$$\begin{aligned}
\bar{\Phi}(X_\tau^i) &= \frac{\langle \mathbb{1}_{X_\tau^i}, N_i^{k-i}\mathbb{1}_{X_\tau^i}\rangle}{\langle \mathbb{1}_{X_\tau^i}, \mathbb{1}_{X_\tau^i}\rangle} \\
&= \frac{\langle U_j^k \mathbb{1}_\tau, U_j^k \mathbb{1}_\tau\rangle}{\langle U_j^i \mathbb{1}_\tau, U_j^i \mathbb{1}_\tau\rangle} \\
&= \frac{R(i,j)^2}{R(k,i)^2}\frac{\langle \mathbb{1}_{X_\tau^k}, \mathbb{1}_{X_\tau^k}\rangle}{\langle \mathbb{1}_{X_\tau^i}, \mathbb{1}_{X_\tau^i}\rangle} \\
&= \frac{R(i,j)}{R(k,i)}\frac{\langle \mathbb{1}_\tau, \mathbb{1}_\tau\rangle}{\langle \mathbb{1}_\tau, \mathbb{1}_\tau\rangle} \\
&= \frac{R(i,j)}{R(k,i)}.
\end{aligned}$$

where we have applied the fact that $\langle X_\tau^\ell, X_\tau^\ell\rangle = R(\ell, j)\langle \mathbb{1}_\tau, \mathbb{1}_\tau\rangle$. On the other hand, by Lemma 4.6.5 we also have that:

$$\bar{\Phi}(\mathbb{1}_{X_\tau^i}) = \frac{1}{\langle \mathbb{1}_{X_\tau^i}, \mathbb{1}_{X_\tau^i}\rangle}\sum_{\ell=0}^{i}\langle \mathbb{1}_{X_\tau^i}, N_i^{k-i}\mathbb{1}_{X_\tau^i,\ell}\rangle$$

$$
\begin{aligned}
&\in \frac{1}{\langle \mathbb{1}_\tau, \mathbb{1}_\tau \rangle} \sum_{\ell=0}^{i} \lambda_j(N_i^{k-i}) \langle \mathbb{1}_{X_\tau^i}, \mathbb{1}_{X_\tau^i,\ell} \rangle + c\gamma \\
&\in \lambda_j(N_i^{k-i}) \frac{\langle \mathbb{1}_\tau, \mathbb{1}_{\tau,j} \rangle}{\langle \mathbb{1}_\tau, \mathbb{1}_\tau \rangle} + \sum_{j=0}^{i} err_1 \\
&\in \lambda_j(N_i^{k-i}) + err_2
\end{aligned}
$$

where as in the proof of Theorem 4.6.3, $c, err_1, err_2 \leq O\left(\frac{i^4 k^2 R_{\max}}{\delta_{i-j}^i (1-\delta_{j-1})}\gamma\right)$. □

Claim 4.4.4 follows immediately from observing that $\rho_i^k = \lambda_i(N_i^{k-i})$ (by Proposition 4.4.2). Theorem 4.4.7 follows from observing that $\widehat{N}_i^{k-i}$ and $\check{N}_k^{k-i}$ have the same approximate eigenvalues (similarly by Proposition 4.4.2).

Finally we close out the section by discussing the connection between non-laziness and a variant of eposets called local-spectral expanders [243]. To start, let's recall this latter definition.

Definition 4.8.6 (Local-Spectral Expander [124, 243])**.** A d-dimensional measured poset (X, Π) is a γ-local-spectral expander if the graph underlying every link[17] of dimension at most $d-2$ is a γ-spectral expander.[18]

Under suitable regularity conditions (see [243]), local-spectral expansion is equivalent to the notion of expanding posets used in this work. A simple argument shows that γ-local-spectral expanders are γ-non-lazy.

Lemma 4.8.7. *Let (X, Π) be a d-dimensional γ-local-spectral expander, and $0 < i < d$. The laziness of the lower walk on level i is at most:*

$$
\max_{\sigma \in X(i)} \left\{ \frac{\langle \mathbb{1}_\sigma, U_{i-1} D_i \mathbb{1}_\sigma \rangle}{\langle \mathbb{1}_\sigma, \mathbb{1}_\sigma \rangle} \right\} \leq \gamma.
$$

[17]Here the link of τ is not just its top level faces, but the complex given by taking this set, removing τ from each face, and downward closing.

[18]A graph is a γ-spectral expander if its weighted adjacency matrix has no non-trivial eigenvalues greater than γ in absolute value.

Proof. Through direct computation, the laziness probability of the lower walk at $\sigma \in X(i)$ is exactly

$$\frac{\langle \mathbb{1}_\sigma, U_{i-1}D_i\mathbb{1}_\sigma\rangle}{\langle \mathbb{1}_\sigma, \mathbb{1}_\sigma\rangle} = \frac{1}{R(i,i-1)}\sum_{\tau\lessdot\sigma}\pi_\tau(\sigma\setminus\tau)$$

It is therefore enough to argue that $\pi_\tau(\sigma\setminus\tau) \leq \gamma$. This follows from the fact that the graph underlying the link X_τ is a γ-spectral expander. In particular, recall that an equivalent formulation of this definition states that:

$$\|A_\tau - UD_\tau\| \leq \gamma,$$

where A_τ is the standard (non-lazy upper) walk and UD_τ is the lower walk on the graph underlying X_τ. This implies that the weight of any vertex v in the graph is at most γ, as:

$$\frac{\langle \mathbb{1}_v, UD_\tau\mathbb{1}_v\rangle}{\langle \mathbb{1}_v, \mathbb{1}_v\rangle} = \frac{\langle \mathbb{1}_v, (UD_\tau - A_\tau)\mathbb{1}_v\rangle}{\langle \mathbb{1}_v, \mathbb{1}_v\rangle} \leq \|A_\tau - UD_\tau\| \leq \gamma$$

where we have used the fact that A_τ is non-lazy by definition. Since $\pi_\tau(\sigma\setminus\tau)$ is exactly the weight of the vertex $\sigma\setminus\tau$ in X_τ, this completes the proof. □

This chapter, in full, is based on the material as it appears in Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques 2022. Gaitonde, Jason; Hopkins, Max; Kaufman, Tali; Lovett, Shachar; Zhang; Ruizhe. “Eigenstripping, Spectral Decay, and Edge-Expansion on Posets". The dissertation author was a primary investigator and author of this material.

Chapter 5

Chernoff Bounds and Reverse Hypercontractivity on HDX

5.1 Introduction

Recent years have seen the emergence of *high dimensional expanders* (HDX) as a core tool in theoretical computer science, with breakthrough applications in approximate sampling [25, 24], coding theory [117, 315], and quantum complexity [27], and strong promise toward longstanding problems in hardness of approximation and probabilistically checkable proofs [124, 109, 236, 177, 208, 40, 113]. One central force behind the success of HDX in application is their *concentration of measure*. Consider (as a warmup) the following fundamental question: given a k-uniform hypergraph X and a function $f : X(1) \to [0, 1]$ on its vertices, how concentrated is f to its mean across a random hyperedge?

$$\mathop{\mathbb{P}}_{\{v_1,\ldots,v_k\}\in X(k)}\left[\left|\frac{1}{k}\sum_{i=1}^{k} f(v_i) - \mu\right| > \varepsilon\right] \overset{?}{\leq} \beta(\varepsilon, k) \tag{5.1}$$

When X is the complete hypergraph, this was classically resolved by Chernoff and Hoeffding [97, 197] who showed $\beta(\varepsilon, k) \leq \exp(-\varepsilon^2 k)$. Over the years, near-matching bounds have been shown for more general hypergraph families (see e.g. [317, 273, 232]), and even for certain bounded degree systems like walks on expanders [6, 165]. These objects, called *sampler graphs* or *extractors*, are by now a core tool in the field with applications in

de-randomization, complexity, and cryptography [167, 362].

High dimensional expanders are known to satisfy a 'Chebyshev-type' bound $\beta(\varepsilon, k) \leq \frac{1}{k\varepsilon^2}$ for (5.1) [124], but this is of course not what makes them powerful. Rather, many modern applications in complexity require concentration of X against *broader classes of functions*. Of particular interest is the extension of (5.1) to functions $f : X(i) \to [0, 1]$ sitting on *i-sets* of X. Complexes satisfying such a bound are called *inclusion samplers*; they arise naturally in the study of agreement and PCPs (forming consistency checks), and play a critical role in low soundness constructions toward the Sliding Scale Conjecture [215, 129, 299, 128]. Unfortunately, inclusion samplers are notoriously difficult to construct, and no bounded degree families were known for many years.

This changed with the advent of spectral HDX [163, 291, 141, 234, 124, 309]. A key motivation behind the modern incarnation of these objects [234, 124, 309] was the fact that their 'inclusion graphs' (the bipartite adjacency matrix between k-sets and i-sets) are spectral expanders, and therefore satisfy 'Chebyshev-type' concentration $\beta(\varepsilon, i, k) \leq \frac{i}{\varepsilon^2 k}$. This observation played a central role in recent progress in agreement testing [124, 121, 109, 236, 177, 110, 41], but left an exponential gap from the Chernoff-type concentration needed for other applications.

In this work we take a major step toward closing this gap, resolving the problem completely in many regimes of interest. In particular, drawing on ideas from the concentration of measure [76], PCP [215], and HDX [109, 9] literature, we prove HDX are optimal samplers between every two levels.

Theorem 5.1.1 (HDX are Optimal Inclusion Samplers (Informal))**.** *Let X be a d-dimensional HDX. Then for any $i \leq k \leq d$ and any $f : X(i) \to [0, 1]$ of expectation μ:*

$$\mathop{\mathbb{P}}_{s \in X(k)} \left[\left| \mathop{\mathbb{E}}_{t \subset s} [f(t)] - \mu \right| \geq \varepsilon \right] \leq \exp\left(-\varepsilon^2 \frac{k}{i} \right).$$

Moreover, this is essentially tight—no inclusion sampler achieves $\exp\left(-\omega\left(\varepsilon^2 \frac{k}{i}\right)\right)$ *error.*

Theorem 5.1.1 is powerful when $i = o(k)$, but our corresponding lower bounds show non-trivial inclusion sampling is essentially impossible in the critical regime $i = \Theta(k)$. In fact this exact issue stood as a barrier for many years on the complete complex, where Impagliazzo, Kabanets, and Wigderson [215] proved Theorem 5.1.1 and used it to construct sub-optimal testers and PCPs by taking $i = \Theta(\sqrt{k})$. In 2017, Dinur and Livni-Navon [128] resolved this problem by appealing to a somewhat surprising player: an underutilized tool from boolean function analysis called *reverse hypercontractivity* (RHC).

Reverse hypercontractivity is a classic functional inequality due to Borell [73] that lower bounds the correlation between sets on the noisy hypercube. Roughly speaking, it implies that for any two subsets $A, B \subseteq \{0,1\}^d$, the probability a ρ-correlated edge crosses A and B is at least some power of their measure:

$$\mathop{\mathbb{P}}_{(s,s')\sim T_\rho}[s \in A, s' \in B] \geq \mathbb{P}[A]^{O_\rho(1)}\,\mathbb{P}[B]^{O_\rho(1)}.$$

Reverse hypercontractivity is an even rarer phenomenon than inclusion sampling—while there are some moderately de-randomized inclusion graphs (namely the Grassmann [215]) satisfying Chebyshev-type variants of Theorem 5.1.1, none are known to admit reverse hypercontractivity. At outset this may even seem necessary. The inequality is closely tied to tensorization and (modified) Log-Sobolev Inequalities [304], which inherently fail on sparse hypergraphs.

In this work, we circumvent this issue by a new combinatorial argument showing any hypergraph satisfying Theorem 5.1.1 'locally' is also reverse hypercontractive. This leads to the following corollary for HDX:

Theorem 5.1.2 (Reverse Hypercontractivity for HDX (Informal))**.** *let X be a d-dimensional HDX. Then for any $\rho \in (0,1)$, $k \leq d$, and $A, B \subset X(k)$:*

$$\mathop{\mathbb{P}}_{(s,s')\sim T_\rho(X)}[s \in A, s' \in B] \geq \mathbb{P}[A]^{O_\rho(1)}\,\mathbb{P}[B]^{O_\rho(1)}.$$

Here $T_\rho(X)$ is the natural generalization of the noise operator to hypergraphs which, roughly speaking, generates s' by re-sampling each vertex in s with probability $1-\rho$. We also prove a more standard analytical version of reverse hypercontractivity for all functions by reduction to the above, as well as analog statements for the heavily studied 'down-up' walks which play a critical role in almost all HDX applications.

Applications.

The remainder of our work is devoted to proving a variety of applications of Theorem 5.1.1 and Theorem 5.1.2 including new families of agreement tests, bounded degree complexes with optimal geometric overlap, double samplers with near-optimal overhead, new degree lower bounds for HDX, large distance list-decodable and locally-testable codes, and several extensions of classic combinatorial and analytic results to HDX. Here we choose to highlight just two of these applications, agreement testing and geometric overlap. We refer the reader to Section 5.2.5 and Section 5.2.6 for a more in-depth discussion of our applications.

Agreement testing is a general property testing paradigm that generalizes classical subroutines in PCP theory like the plane vs. plane test (e.g. [31, 168, 116]). Roughly speaking, an agreement test consists of a universe U and a family of overlapping subsets $S \subset \mathbb{P}(U)$ (think of U as the vertices of X, and S as its hyperedges). Given a family of *local* assignments $\{f_s : s \to \{0,1\}\}_{s\in S}$, we'd like to test whether f_s actually agrees with a *global* function $g : U \to \{0,1\}$ in the sense that $f_s = g|_s$ for many s. These systems occur in PCP reductions to ensure the provers (who answer on subsets) can't 'cheat' by using answers not corresponding to a real solution of the initial problem (corresponding to global functions on U).

Agreement testing has two major regimes, the '99%-regime' (where we'd like to infer global structure only if the test passes with high probability) and the '1%-regime' (where we'd like to infer global structure even when the test only passes with *non-trivial*

probability). Leveraging reverse hypercontractivity, we give new tests in both regimes. In the 99%-regime, we give a new 2-query test where the universe U corresponds to *i-sets* of a high dimensional expander, extending a similar result of [118] on the complete complex. Our test, along with Theorem 5.1.1, has since appeared as an important sub-routine in the construction of the first bounded-degree 1% agreement testers in [110, 113].

In the 1%-regime, we give a new 3-query tester for a variant of ℓ_∞-expanders (a strengthened notion of HDX [232]) with *optimal soundness*, meaning that if the test passes with probability asymptotically better than a random function ($2^{-\Theta(k)}$ for a k-dimensional ℓ_∞-expander), we are able to infer global structure. In the context of PCPs, such a soundness guarantee is critical to ensure the alphabet size stays polynomial in the soundness. Our result gives rise to the first families of optimal testers beyond the complete complex and products [128], including (dense) random complexes, skeletons of many well-studied spin systems, the full linear matroid, and more. We remark that while all such examples are dense, we prove an optimal 'local' agreement theorem only under the assumption of local spectral expansion. We lift this guarantee to the general result using ℓ_∞-expansion, but it is possible this requirement could be relaxed using recent techniques based on coboundary expansion [177, 41, 110, 40, 113] to give sparse optimal testers.

The *Geometric Overlap Property* is one of the earliest notions of high dimensional expansion [182, 288]. A d-dimensional complex X has c-geometric overlap if for any embedding of X into $\mathbb{R}^d$, there is a point $p \in \mathbb{R}^d$ that is covered by a c-fraction of X's hyperedges. Geometric overlap was first proven for the complete complex in 2-dimensions by Boros and Füredi [74], and later extended to all dimensions by Bárány [50]. Gromov [182] asked whether there are bounded degree complexes satisfying geometric overlap. This was resolved in [153], who gave both an optimal random construction and several explicit constructions with bounded but sub-optimal overlap (including one from high dimensional expanders). Several later works [316, 140, 309] continued to build the connection between

high dimensional expanders and geometric overlap, but it remained open whether a construction achieving the best of both worlds (explicit and optimal) could be achieved. Leveraging a variant of Theorem 5.1.1, we resolve this problem, showing under mild assumptions that *any* sufficiently strong high dimensional expander has near-optimal geometric overlap.

The Trickling-Down Threshold.

High dimensional expanders exhibit a *phase transition* at a certain expansion parameter called the 'Trickling-Down (TD)-Threshold' [309]. Any complex breaking this barrier immediately has 'local-to-global' structure, meaning properties such as global expansion and fast-mixing can be inferred just from *local* structure of the complex [163, 234, 309, 11]. This barrier shows up as a major point in constructions of HDX as well. It is relatively easy to build complexes *at* the threshold by tensoring an expander with the complete complex [287, 170], but *breaking* the barrier requires complicated algebraic machinery and is considered the 'gold standard' for HDX—indeed only three such constructions are known [291, 238, 108].

Theorem 5.1.1 and Theorem 5.1.2 hold for $\sqrt{d}$-dimensional skeletons of any complex breaking the TD-Threshold. In fact, we prove that the top level of such complexes satisfy exponential concentration for all Lipschitz functions, corresponding to $\beta(\varepsilon, i, k) \leq \exp(-\varepsilon\frac{\sqrt{k}}{i})$ in the inclusion sampling setting. On the other hand, we give examples of complexes *at* the TD-threshold that exhibit arbitrarily poor concentration. We view our results in this sense as a significant strengthening of HDX's 'local-to-global' phase transition, and a partial explanation to the great difficulty of constructing hypergraphs beyond the TD-Threshold. Whereas graph expansion and mixing are 'common' phenomenon, Theorem 5.1.1 and Theorem 5.1.2 imply as soon as one breaks the TD-Threshold X not only satisfies exponential concentration of measure, its skeletons satisfy optimal concentration and reverse hypercontractivity—rare properties not known for any other sparse systems.

5.2 Main Results and Proof Overview

5.2.1 Background

Before stating our results in somewhat more formality, we cover some basic background on high dimensional expanders, random walks, and sampling. We refer the reader to Section 5.4 for formal details and discussion.

Simplicial Complexes.

A d-**uniform simplicial complex** X consists of a d-uniform hypergraph $X(d)$ together with its downward closure

$$X = X(1) \cup \ldots \cup X(d),$$

where $X(i) \subseteq \binom{[n]}{i}$, called the '$i$-faces', are all i-size subsets that sit in some hyperedge in $X(d)$. Given a face $t \in X(i)$, the **link** of t is the sub-complex induced by localizing to faces that include t, that is $X_t \coloneqq \{s : s \cup t \in X\}$. We say X is **connected** if the base graph $(X_t(1), X_t(2))$ of every link is connected. A d-uniform complex is **partite** if its vertices can be partitioned into d 'parts' such that each top-level face has one part from each component. Finally, the **inclusion graph** $(X(k), X(i))$ is the bipartite graph between k-sets and i-sets of X where edges are given by inclusion.

We emphasize that we have changed convention here from dimension, where $X(d)$ refers to sets of size $d+1$, to uniformity where $X(d)$ consists of sets of size d. We will use the latter notation throughout the rest of the paper to simplify both the statements and proofs of our results.

High dimensional expanders.

A simplicial complex X is a λ**-one-sided HDX** (resp. two-sided) if for every $s \in X$ of co-dimension[1] at least 2 $(X_t(1), X_t(2))$ is a λ-one-sided spectral expander (resp. two

[1]Here by co-dimension j we mean sets of size $d - j$.

sided).

The weakest form of high dimensional expanders are λ-**Trickling-Down (TD) Complexes**. We call X λ-TD if it is connected and every co-dimension 2 link is a $\frac{\lambda}{d-1}$-one-sided expander.

Our results hold for a broad class of HDX we call **Nice Complexes**. We defer the formal definition to Section 5.4.8, and here give three basic examples of nice complexes for which our results hold:

- X is the $\sqrt{d}$-skeleton of a λ-TD complex
- X is the $\sqrt{d}$-skeleton of a partite 2^{-d}-one-sided HDX[2]
- X is a 2^{-d}-two-sided HDX

High Order Random Walks.

Simplicial complexes admit a variety of 'high order random walks' generalizing the standard random walk on a graph. There are three natural walks critically important to our results. The first are the classical **down-up walks** of [234], denoted $U_{k,d}D_{d,k}$, which walk between $s, s' \in X(d)$ through a shared k-face $t \subset s, s'$. The second are the **swap walks** of [109, 9], denoted $S_{i,j}$, which walk between *disjoint* faces $s \in X(i)$ and $s' \in X(j)$ via $s \cup s' \in X(i+j)$. The final walk is the **noise operator** T_ρ, which walks between $s, s' \in X(d)$ via a random i-subface of s for $i \sim Bin(1-\rho, d)$.[3]

Sampler Graphs.

A bipartite graph $G = (L, R, E)$ is an (ε, β)-function additive sampler if $\forall f : R \to [0,1]$

$$\mathop{\mathbb{P}}_{v \in L}\left[\left|\mathop{\mathbb{E}}_{u \in R, u \sim v}[f(u)] - \mathop{\mathbb{E}}_{u \in R}[f(u)]\right| \geq \varepsilon\right] < \beta.$$

[2]The reader may notice this is subsumed by the above case. In reality, we prove stronger (but somewhat more involved) bounds for partite complexes, and show under plausible assumptions that optimal sampling even holds at the *top* level of such objects.

[3]We remark that while definition this may seem odd at outset, it is the natural extension of the noise operator to hypergraphs and recovers the standard notion on products. See [39] or Section 5.4 for further details.

G is called an (α, β, δ)-multiplicative sampler if for all $f : R \to [0, 1]$ of density at least $\mathbb{E}_{u \in R}[f(u)] \geq \alpha$:

$$\mathop{\mathbb{P}}_{v \in L}\left[\left|\mathop{\mathbb{E}}_{u \in R, u \sim v}[f(u)] - \mathop{\mathbb{E}}_{u \in R}[f(u)]\right| \geq \delta \mathop{\mathbb{E}}_{u \in R}[f(u)]\right] < \beta.$$

It is a well-known fact (c.f. [215]) that one can move between additive and multiplicative samplers without much parameter loss, and moreover that one can 'flip' L and R in the sense that if $G = (L, R, E)$ is an (α, β, δ)-sampler, then $G^{op} = (R, L, E)$ is roughly a (β, α, δ)-sampler up to slight decay in parameters. We refer the reader to Section 5.4 for details.

5.2.2 Sampling and Concentration

We can now state and overview the proof of optimal sampling on HDX in somewhat more formality.

Theorem 5.2.1 (Sampling on HDX (Informal Theorem 5.5.1))**.** *Let X be a nice, k-uniform HDX and $i \leq k$. The containment graph $(X(k), X(i))$ is a (ε, β)-function sampler for*

$$\beta \leq \exp\left(-\Omega\left(\varepsilon^2 \frac{k}{i}\right)\right).$$

In other words, for any $f : X(i) \to [0, 1]$ of expectation μ:

$$\mathop{\mathbb{P}}_{s \in X(k)}\left[\left|\mathop{\mathbb{E}}_{t \subseteq s}[f(t)] - \mu\right| \geq \varepsilon\right] < \exp\left(-\Omega\left(\varepsilon^2 \frac{k}{i}\right)\right).$$

Since $\sqrt{d}$-skeletons of λ-TD complexes are nice, we get the following immediate corollary.

Corollary 5.2.2 (Sampling for TD-Complexes (Informal Claim 5.4.28))**.** *Fix $\lambda < 1$ and let X be a d-uniform λ-TD complex. Then for any $i < k \leq \sqrt{d}$, the containment graph of*

$(X(k), X(i))$ *is a* (ε, β)*-sampler for*

$$\beta \leq \exp\left(-(1-\lambda)\varepsilon^2 \frac{k}{i}\right).$$

We remark that this (and all following) results also hold in some variation under the popular notion of *spectral independence*, which for our purposes act similar to TD complexes. We discuss in the main body.

En route to Theorem 5.2.1, we also show the *top level* of any λ-TD complex satisfies exponential concentration for Lipschitz functions, which may be of independent interest.

Theorem 5.2.3 (Exponential Concentration of λ-TD Complexes (Informal Corollary 5.7.7))**.** *For any* $\lambda < 1$*,* $\nu > 0$*, and* $k \in \mathbb{N}$*, let* X *be a* k*-uniform* λ*-TD complex. For any* ν*-Lipschitz function* $f : X(k) \to \mathbb{R}$*:*

$$\mathop{\mathbb{P}}_{s \in X(d)}[|f(s) - \mathbb{E}[f]| \geq t] \leq \exp\left(-\frac{t}{\sqrt{c_\lambda \nu}}\right)$$

for $c_\lambda \leq 1 + e^{\frac{\lambda}{1-\lambda}}$*.*

We refer the reader to Section 5.4.6 for the definition of ν-Lipschitz, and here just note that specialized to inclusion sampling this implies $(X(k), X(i))$ is roughly a $(\varepsilon, \exp(-\varepsilon\frac{\sqrt{k}}{i}))$-sampler. A similar statement holds under spectral independence, albeit for a more restricted class than Lipschitz functions (see Corollary 5.7.8).

Proof Overview.

The proof of Theorem 5.2.1 is broken into two main components: the 'Chernoff' ($i = 1$) setting, and a bootstrapping argument lifting Chernoff to general i. Here we focus just on the simplest version of the proof, the case of *two-sided HDX*, and refer the reader to Section 5.7 for the general case.

Part I: Chernoff-Hoeffding.

Toward Theorem 5.2.1, we first prove a fairly general "k vs 1" concentration theorem for any HDX X whose strength depends on X's underlying quantitative expansion. Under strong assumptions (2^{-k}-HDX), we recover true Chernoff. Perhaps more surprisingly, even under very weak assumptions ($\frac{1}{k}$-HDX) we still recover strong concentration of the form $\exp(-\varepsilon^2\sqrt{k})$. Taking a $\sqrt{k}$-skeleton of the latter recovers Chernoff (see Section 5.4.6). We give a simplified statement here:

Theorem 5.2.4 (Chernoff-Hoeffding for HDX (Informal Theorem 5.5.2))**.** *Let X be a k-uniform $\frac{1}{k}$-two-sided HDX. Then for any $f : X(k) \to [0,1]$:*

$$\mathbb{P}_{\{v_1,\ldots,v_k\}\in X(k)}\left[\left|\frac{1}{k}\sum_{i=1}^{k} f(v_i) - \mu\right| > \varepsilon\right] \leq \exp(-\Omega(\varepsilon^2\sqrt{k}))$$

Moreover, if X is a 2^{-k}-two-sided HDX stronger concentration holds:

$$\mathbb{P}_{\{v_1,\ldots,v_k\}\in X(k)}\left[\left|\frac{1}{k}\sum_{i=1}^{k} f(v_i) - \mu\right| > \varepsilon\right] \leq \exp(-\Omega(\varepsilon^2 k))$$

For simplicity let $f = 1_A$ be the indicator of some $A \subset X(1)$. The proof of Theorem 5.2.4 follows the standard Chernoff-style method of bounding the moment generating function (MGF). Namely for any $r > 0$:

$$\begin{aligned}\mathbb{P}_{\{v_1,\ldots,v_k\}\in X(k)}\left[\sum_{i=1}^{k} 1_A(v_i) \geq k(\mu+\varepsilon)\right] &= \mathbb{P}\left[\exp\left(\sum_{i=1}^{k} r1_A(v_i)\right) \geq \exp(rk\,(\mu+\varepsilon))\right] \\ &\leq \frac{\mathbb{E}\left[\prod_{i=1}^{k} e^{r1_A(v_i)}\right]}{e^{rk(\mu+\varepsilon)}}.\end{aligned}$$

In the classical proof of Chernoff, the variables v_i are independent, so one can bound the

MGF by

$$\mathop{\mathbb{E}}_{\{v_1,\ldots,v_k\}\in X(k)}\left[\prod_{i=1}^{k} e^{r1_A(v_i)}\right] = \mathop{\mathbb{E}}_{v\in X(1)}\left[e^{r1_A(v)}\right]^k \lesssim e^{rk(\mu+r)} \tag{5.2}$$

and set $r = \Theta(\varepsilon)$ to get the desired bound.

When X is an HDX, the variables $\{v_1, \ldots, v_k\}$ may be extremely correlated, so the above approach breaks down naively. Instead, we take inspiration from Healy's proof [195] of the Expander-Chernoff Theorem and recursively bound the MGF by 'splitting' it into components on lower levels of X using the swap walks. Toward this end, for any $\ell \le k$, define the partial MGF $z_\ell : X(\ell) \to \mathbb{R}$ on a face $t = \{v_1, v_2, \ldots, v_\ell\}$ as

$$z_\ell(v_1, v_2, \ldots, v_\ell) = \prod_{i=1}^{\ell} e^{r1_A(v_i)} = e^{r|t\cap A|}.$$

Our goal is to bound $\mathbb{E}[z_k] \approx \mathbb{E}[z_1]^k$. To set up a recursion, we appeal to the elementary observation that a k-set $s \in X(k)$ can be sampled by first drawing a $\frac{k}{2}$-set $t_1 \in X(k/2)$, then $t_2 \in X_{t_1}(k/2)$ conditionally from its link. We set $s = t_1 \cup t_2$. This allows us to 'split' z_k into two correlated copies of $z_{k/2}$:

$$\begin{aligned}
\mathop{\mathbb{E}}_{s\in X(k)}[z_k(s)] &= \mathop{\mathbb{E}}_{s\in X(k)}[e^{r|s\cap A|}] \\
&= \mathop{\mathbb{E}}_{t_1\cup t_2}[e^{r|t_1\cap A|}e^{r|t_2\cap A|}] \\
&= \mathop{\mathbb{E}}_{t_1\cup t_2}[z_{k/2}(t_1)z_{k/2}(t_2)].
\end{aligned}$$

The trick is now to observe that (t_1, t_2) is distributed exactly as an edge in the swap walk $S_{\frac{k}{2},\frac{k}{2}}$. Since swap walks on HDX have excellent spectral expansion [9, 109, 187, 12], we can 'de-correlate' the above and write the relation

$$\mathbb{E}[z_k] \le (1-\gamma)\,\mathbb{E}[z_{k/2}]^2 + \gamma\mathbb{E}[z_{k/2}^2] \tag{5.3}$$

where γ is the expansion of the swap walk.

We could now recurse if not for the error term $\gamma\mathbb{E}[z_{k/2}^2]$, which could very well be the dominating factor. The key is show $\mathbb{E}[z_{k/2}^2]$ actually can't be too much bigger than $\mathbb{E}[z_k]$ itself:

$$\mathbb{E}[z_{k/2}^2] \lesssim e^{r^2k}\,\mathbb{E}[z_k]. \tag{5.4}$$

Setting r such that $\gamma e^{r^2k} \ll 1$, (5.3) becomes

$$\mathbb{E}[z_k] \lesssim \frac{1-\gamma}{1-e^{r^2k}\gamma}\,\mathbb{E}[z_{k/2}]^2 \approx \mathbb{E}[z_{k/2}]^2,$$

allowing us to recurse to show an upper bound similar to (5.2). For 'weak' HDX, this strategy requires setting $r = \Theta(\varepsilon k^{-1/2})$, while for strong HDX we may set $r = \Theta(\varepsilon)$. This results in the gap in the stated bounds.

To prove (5.4) we in some sense 'reverse' the sampling process above, and draw a $\frac{k}{2}$-set t by first sampling a k-set $s \in X(k)$, then take $t \subset s$ uniformly at random. This means we can write the error term as:

$$\mathop{\mathbb{E}}_{t\in X(k/2)}[z_{k/2}(t)^2] = \mathop{\mathbb{E}}_{s\in X(k)}\left[\mathop{\mathbb{E}}_{t\subseteq s}\left[z_{k/2}(t)^2\right]\right] = \mathop{\mathbb{E}}_{s\in X(k)}\left[\mathop{\mathbb{E}}_{t\subseteq s}\left[e^{2r|t\cap A|}\right]\right].$$

Crucially, the inner expectation is now over the $\frac{k}{2}$-uniform *complete complex* (whose vertices are the k-set s), so standard concentration implies the exponent, $2r|t\cap A|$, is close to its expectation $r|s\cap A|$ up to a factor of roughly r^2k with high probability. Thus:

$$\mathop{\mathbb{E}}_{s\in X(k)}\left[\mathop{\mathbb{E}}_{t\subseteq s}\left[e^{2r|t\cap A|}\right]\right] \lesssim \mathop{\mathbb{E}}_{s\in X(k)}\left[e^{r|s\cap A|+r^2k}\right] = e^{r^2k}\mathop{\mathbb{E}}_{s\in X(k)}[z_k].$$

Part II: Bootstrapping.

We now argue one can lift Chernoff for HDX to optimal sampling for their inclusion graphs $(X(k), X(i))$. The idea, based on the method in [215] for the complete complex, is

to try to reduce to a $\frac{k}{i}$-uniform complex Y whose vertices are i-sets in $X(i)$, and whose $\frac{k}{i}$-faces correspond in some way to the k-sets in X. We could then hope to apply Chernoff to this system to bound the k vs. i sampling behavior in the original complex.

Somewhat more formally, we'll consider the $\frac{k}{i}$-uniform complex generated by the following process: draw a random k-face $\{v_1, \ldots, v_k\} \in X(k)$, and partition the face randomly into $\frac{k}{i}$ subsets of size i denoted $\{I_1, \ldots, I_{\frac{k}{i}}\}$. The resulting complex, called the *faces complex* of X, inherits local-spectral expansion from X's swap walks so we can apply Chernoff for HDX to get concentration of the form $\exp(-\varepsilon^2 \frac{k}{i})$.

Unfortunately, this does not prove concentration for quite the right object. In particular, given a function $f : X(i) \to [0,1]$, the above really states that a random partitioning of a k-face s into $\frac{k}{i}$ sub i-faces satisfies

$$\mathop{\mathbb{P}}_{s \in X(k),\ \uplus_j I_j = s} \left[\left| \frac{i}{k} \sum_{j=1}^{k/i} f(I_j) - \mu \right| > \varepsilon \right] \leq \exp\left(-\varepsilon^2 \frac{k}{i} \right),$$

whereas we'd like to bound $\mathop{\mathbb{P}}_{s \in X(k)} \left[\left| \mathop{\mathbb{E}}_{t \subset s} [f(t)] - \mu \right| > \varepsilon \right]$. Here we are saved by the fact that each I_j is *marginally* distributed as a random i-set $t \subset s$. We prove in such cases it is possible to inherit sampling from the correlated bound up to a small loss in parameters, completing the proof.

5.2.3 Reverse Hypercontractivity

Leveraging Theorem 5.2.1 *locally* in X, we prove reverse hypercontractivity for high dimensional expanders.

Theorem 5.2.5 (Reverse Hypercontractivity (Informal Theorem 5.6.1))**.** *Fix $\rho \in (0,1)$ and let X be a k-uniform nice complex for k sufficiently large. Then there exist constants*

C, q (depending only on ρ) such that

$$\langle f, T_\rho g \rangle \geq C \|f\|_q \|g\|_q.$$

In other words T_ρ is "(C, q, q')-reverse hypercontractive.[4]*"*

The core of Theorem 5.2.5 is really a version of the result for the down-up walk which simply states any two subsets $A, B \subset X(k)$ of non-trivial size must have correspondingly many edges between them.

Lemma 5.2.6 (Reverse Hypercontractivity of the Down-Up Walk (Informal Theorem 5.6.4))**.** *Fix $\gamma \in (0,1)$ and let X be a k-uniform nice complex for k sufficiently large. There exist constants c, q (depending only on γ) such that for any $A, B \subset X(k)$ of measure at most* $\exp(-ck)$*:*

$$\mathop{\mathbb{P}}_{s,s' \sim U_{\gamma k,k} D_{k,\gamma k}} [s \in A, s' \in B] \geq \mathbb{P}[A]^q \, \mathbb{P}[B]^q.$$

Note that the assumption on the set size in Lemma 5.2.6 is essentially tight. Even if one walks from s to s' while going down to a single vertex it is possible to find sets A and B of size roughly $\exp(-k)$ which are totally disconnected, even on the complete complex.

Proof Overview.

The proof of Lemma 5.2.6 relies on the following simple observation: since the down-up walk samples an edge by first sampling $t \in X(\gamma k)$ and then *independently* sampling $s, s' \in X(k)$ containing t, we can express the γ-correlated mass of A and B as:

$$\mathop{\mathbb{P}}_{s,s' \sim U_{\gamma k,k} D_{k,\gamma k}} [s \in A, s' \in B] = \mathop{\mathbb{E}}_{t \in X(\gamma k)} \left[\mathop{\mathbb{P}}_{s} [A \mid \ s \supseteq t] \mathop{\mathbb{P}}_{s} [B \mid \ s \supseteq t] \right].$$

[4]Here q' is the Hölder Conjugate of q. See Section 5.4 for a more detailed background on reverse hypercontractivity.

Now the connection with inclusion sampling becomes clear—if even a non-negligible fraction of $t \in X(\gamma k)$ see a reasonable portion of A and B simultaneously, the righthand side should be large!

Naturally our first thought might be to appeal to Theorem 5.2.1 which bounds this type of behavior, but when $i = \gamma k$ the resulting bound is far too weak to be useful. In particular, after swapping from additive to multiplicative sampling and 'flipping' the graph, Theorem 5.2.1 roughly states that $(X(i), X(k))$ is an (α, β, δ)-sampler for $\alpha \approx \exp(-\delta^2 \beta \frac{k}{i})$. When $i = \Theta(k)$, this only gives a guarantee for sets of constant size.

The key is to realize that because X expands *locally*, we don't have to sample $t \in X(\gamma k)$ all at once. Instead, we'll sample t in 'steps' as $t = t_1 \cup \ldots \cup t_m$ where each t_j is sampled conditionally (i.e. from the link of) its predecessors, and apply Theorem 5.2.1 to each t_j individually. Toward this end, assume $\mathbb{P}[A] \leq \mathbb{P}[B]$ and fix the 'step size' $\ell = |t_j|$ roughly such that $\exp(-c' \frac{k}{\ell}) = \mathbb{P}[A]$ for some sufficiently small $c' > 0$.

Drawing our first component t_1, observe for small enough c' we have set our parameters such that Theorem 5.2.1 promises $(X(\ell), X(k))$ is an $(\alpha, \frac{1}{4}, \frac{1}{2})$-sampler for $\alpha \leq \min\{\mathbb{P}[A], \mathbb{P}[B]\}$, so with probability at least $\frac{1}{2}$ the conditional measures of A and B in t_1 are at least $\frac{1}{2}\mathbb{P}[A]$ and $\frac{1}{2}\mathbb{P}[B]$ respectively. Moreover, moving into the link of t_1 we can apply exactly the same process to t_2 (and so forth), so by the i-th step one can inductively show that with probability at least 2^{-i}, the conditional mass of A and B are at least $2^{-i}\mathbb{P}[A]$ and $2^{-i}\mathbb{P}[B]$ respectively. On the other hand, we only take $m = \frac{\gamma k}{\ell} = O(\log(\mathbb{P}[A]^{-1}))$ total steps by our choice of ℓ, so at the end of the procedure we have that with probability at least $2^{-m} = \mathbb{P}[A]^{O(1)}$, the density of A and B in t is at least $\mathbb{P}[A]^{O(1)}$ as well. Putting everything together, we have

$$\begin{aligned}\mathop{\mathbb{P}}_{s,s' \sim U_{\gamma k,k} D_{k,\gamma k}}[s \in A, s' \in B] &= \mathop{\mathbb{E}}_{t \in X(\gamma k)}\left[\mathop{\mathbb{P}}_{s}[A \mid \ s \supseteq t] \mathop{\mathbb{P}}_{s}[B \mid \ s \supseteq t]\right] \\ &\geq 2^{-m} \cdot (2^{-m}\mathbb{P}[A]) \cdot (2^{-m}\mathbb{P}[B])\end{aligned}$$

$$\geq \mathbb{P}[A]^{O(1)}\,\mathbb{P}[B]$$

as desired.

Once one has Lemma 5.2.6, Theorem 5.2.5 follows by passing to discrete approximations of f and g that carefully balance the quality of approximation (namely regarding the mass of f and g and pointwise closeness) and coarseness (number of discrete values in the approximation). Done correctly, one can then divide the functions into boolean level-sets and apply Lemma 5.2.6 to achieve full reverse hypercontractivity at the cost of a slight decrease in the corresponding norm and an additional constant factor.

5.2.4 Optimality and the Trickling-Down Threshold

Oppenheim's Trickling Down Theorem is among the most fundamental results in the theory of spectral high dimensional expansion [309]. It states that any connected complex whose top links have expansion *strictly* better than $\frac{1}{d-1}$ exhibits 'local-to-global' behavior: one can immediately give quantitative bounds on the spectral expansion of *all* links of X (including its 'global' 1-skeleton), as well as on the down-up walks [11]. Our results are in some sense a strengthening of the local-to global behavior of Oppenheim: essentially as soon as one passes this threshold, not only can one infer 'Chernoff'-type concentration on skeletons of X, X itself actually satisfies exponential concentration for any Lipschitz function. This is in stark contrast to complexes *at* the TD-Barrier, which we observe may have no concentration properties whatsoever:

Theorem 5.2.7 (Lower Bounds at the TD-Barrier (Proposition 5.11.7))**.** *For every $\beta < 1$ and $k \in \mathbb{N}$, there exists a family of k-uniform 1-TD complexes $\{X_n\}$ such that $(X_n(k), X_n(1))$ is not a $(\frac{1}{2}, \beta)$-additive sampler.*

Theorem 5.2.1 also gives essentially the best possible parameters for inclusion sampling on any complex.

Theorem 5.2.8 (Inclusion Sampling Lower Bounds (Informal Theorem 5.11.1))**.** *Let $i < k \in \mathbb{N}$, $\varepsilon \in (0, 0.1)$, and let X be a k-uniform complex, then one of the following holds:*

1. *$(X(k), X(i))$ is not an (ε, β)-sampler for $\beta \leq \exp\left(-O\left(\varepsilon^2 \frac{k}{i}\right)\right)$.*
2. *$(X(k), X(i))$ is not an $(\frac{1}{10i}, \frac{\varepsilon}{6i})$-sampler.*

Classical lower bounds on samplers (e.g. [87]) are based on *degree* and are far from tight in our setting where inclusion structure is the main barrier. Nevertheless, we are able to argue our bounds are optimal by reducing to $i = 1$ where the degree bound is tight. The idea is to show any set A which is a counter-example to sampling of $(X(k), X(1))$ can be 'lifted' to a counter example to $(X(k), X(i))$ by taking A' to be the family of i-sets that hit A. This works when the measure of A' is roughly $i\mu(A)$, and therefore when 1) A is small, and 2) $(X(k), X(i))$ is a sampler. We show the original technique of [87] can be used to analyze sets of any density, thereby giving the desired counter-example.

5.2.5 Agreement Testing

Leveraging Theorem 5.2.5, we prove several new agreement theorems in both the 99% and 1%-regimes. In the former we consider a general setup of [118]. Given a complex X and an ensemble of functions $\mathcal{F} = \{f_s : \binom{s}{i} \to \mathbb{F}_2\}_{s \in X(k)}$, we'd like to test whether $\mathcal{F}$ 'comes from' a global function $g : X(i) \to \mathbb{F}_2$. To be concrete, consider $i = 1$. Here we are given $\mathcal{F} = \{f_s : s \to \mathbb{F}_2\}$ and can interpret each f_s as the indicator of some subset $r_s \subseteq s$. We want to test whether there is a 'global' subset $R \subseteq X(1)$ that approximately determines most local subsets in the sense that $r_s \approx R \cap s$. For $i = 2$, a similar interpretation holds but for *subgraphs* of X. Namely, think of $f_s : \binom{s}{2} \to \mathbb{F}_2$ now as a subgraph of s, i.e. corresponding to the indicator of an edge-set $e_s \subseteq \binom{s}{2}$. Now 'coming from' a global function means there is a global subgraph $E \subseteq X(2)$ such that for most local graphs, $e_s \approx E \cap s$. For $i > 2$ one interprets this similarly as a question about local vs. global

sub-complexes of X. Such 'subcomplex' tests have proved useful both in boolean analysis [118] and as a tool for analyzing the more challenging 1%-regime [110, 113].

We study the classical V-test in this context, which draws a pair of k-faces $(s, s') \sim U_{k/2,k}D_{k,k/2}$ intersecting on $k/2$ vertices and checks if s and s' agree on most i-faces in the intersection. Let $Agree^V_\eta(\mathcal{F})$ denote the probability (s, s') agree on at least a $1-\eta$ fraction of their shared i-faces. Similarly, for any global function $g : X(i) \to \mathbb{F}_2$, write $f_s \overset{\eta}{\approx} g|_s$ if f_s agrees with g on a $1-\eta$ fraction of its i-faces. Then:

Theorem 5.2.9 (The Subcomplex V-Test, 99%-Regime (Informal Theorem 5.8.1))**.** *Let X be a k-uniform nice complex. For any $\eta, \varepsilon > 0$ if $Agree^V_\eta(\mathcal{F}) \geq 1-\varepsilon$:*

$$\exists g : X(1) \to \mathbb{F}_2, \quad \mathbb{P}_{s \in X(k)}[f_s \overset{\eta'}{\approx} g|_s] \geq 1-\varepsilon'.$$

where $\eta' \leq O(\eta + \varepsilon)$ and $\varepsilon' \leq \varepsilon^{O_\eta(1)} + e^{-\Omega_\eta(k)}$.

In the '1%'-regime, our goal is to infer global structure of $\mathcal{F}$ even when the test passes with small (but non-negligible) probability. We focus on the $i = 1$ setting and aim to construct a test $\mathcal{T}$ that infers global structure whenever $Agree^{\mathcal{T}}(\mathcal{F}) \geq \exp(-\Omega(k))$, the best possible bound since a random function passes any such test with this probability. To this end, we consider a variant of [215]'s Z-test on high dimensional expanders which samples a triple $(\sigma, \sigma', \sigma'') \in X(k)$ roughly such that (σ, σ') is distributed as the V-test, and σ'' is drawn from the link of $\sigma' \setminus \sigma$. The Z-test passes if (σ, σ') and (σ', σ'') are consistent on their intersections. We prove that the Z-test is sound under the stronger assumption that the complex is "λ-global", meaning the swap walk $S_{\frac{k}{2},\frac{k}{2}}$ is λ-close to its stationary distribution in ℓ_∞-norm (see Definition 5.8.6).

Theorem 5.2.10 (The Z-Test, 1%-Regime (Informal Theorem 5.8.7))**.** *$\forall \lambda, \eta > 0$ and large enough k, let X be a λ-global k-uniform nice complex. Then for any $\delta \in (8\lambda + e^{-\Omega(\eta k)}, \frac{1}{8})$*

if $Agree_0^Z(\mathcal{F}) \geq \delta$*:*

$$\exists g : X(1) \to \mathbb{F}_2, \quad \mathop{\mathbb{P}}_{s \in X(k)}[f_s \overset{\eta}{\approx} g|_s] \geq \delta/8.$$

We will not give a true proof overview of our testing results (though it should be mentioned that the high level strategy closely follows [215, 128]), but wish to highlight that the core of Theorem 5.8.7 is really a 'weak' *local* agreement theorem for the V-test on nice HDX of independent interest. The statement is quite technical, so we give a very informal version here that captures it in spirit.

Theorem 5.2.11 (The Local Agreement Theorem (Informal Theorem 5.8.12))**.** *Let* X *be a nice complex. Then for any* $\delta \geq \exp(-\Omega(k))$ *if* $Agree_0^V \geq \delta$*, for an* $\Omega(\delta)$*-fraction of* $t \in X(k/2)$ *there is a 'smoothing' of* $\mathcal{F}$ *which maintains its structure and has* $(1-\delta^2)$*-agreement on the link* X_t*.*

In other words, any non-trivial agreement of the V-test *must* come from the fact that $\mathcal{F}$ is locally consistent with a global function. It is possible Theorem 5.2.11 could be propogated to a true Z-test under much weaker conditions than λ-globality, e.g. under the recent topological notions of [177, 110, 41]. We leave this as an open question for the 1%-regime.

5.2.6 Further Applications

Finally, we discuss our applications of Theorem 5.2.1 and Theorem 5.2.5 beyond agreement testing.

Geometric Overlap.

A complex X has (d, c)-geometric overlap if for every embedding $X(1)$ into $\mathbb{R}^{d-1}$, there is a point $q \in \mathbb{R}^{d-1}$ that lies in the convex hulls of at least a c-fraction of X's embedded d-faces. [153] proved every d-uniform bounded degree complex has geometric overlap at best c_d, where c_d is the overlap of the complete complex. We prove sufficiently strong high dimensional expanders match this bound.

Theorem 5.2.12 (Geometric Overlap (Informal Theorem 5.9.13))**.** *For every $d \in \mathbb{N}$ and $\varepsilon > 0$, there exists $\lambda > 0$ such that any λ-two-sided HDX with uniform vertex weights has $(d, c_d - \varepsilon)$-geometric overlap.*

We remark that while geometric overlap is a classical problem in mathematics (dating back to [74]), we are not aware of any applications in computer science. Nevertheless other instances of overlap theorems have proven powerful (see e.g. [92]), so it seems plausible such results may be of future use.

Double Samplers.

Double samplers are a powerful variant of sampler graphs used in [121] to construct good list-decodable codes and in [132] for the heavy hitters problem. Roughly speaking, double samplers are 'three-wise' inclusion structures $(X(k), X(j), X(1))$ such that

1. $(X(k), X(j))$ is a $(1/2, \beta)$-additive sampler
2. For all $s \in X(k)$, the restriction of $(X(j), X(1))$ to vertices in s is a $(1/2, \beta)$-additive sampler

The best known prior construction of double samplers [121] had overhead $\frac{|X(k)|}{|X(1)|} = \exp(\text{poly}(\beta^{-1}))$, leading to poor rate of their resulting list-decodable codes. One of the main open questions asked in [121] was to determine the optimal overhead of a double sampler.

We nearly resolve this problem for 'typical' complexes X which are non-contracting (i.e. $|X(j)| > |X(i)|$ whenever $j > i$), and satisfy a weak 'hitting-set' type guarantee. In particular, Theorem 5.2.1 gives double samplers with only *quasi-polynomial overhead*, while for such typical complexes Theorem 5.2.8 implies a corresponding near-matching lower bound.

Theorem 5.2.13 (Near-Optimal Double Samplers (Informal Theorem 5.9.3))**.** *For every $\beta > 0$, there exists an infinite family of double samplers with $\exp(O(\log^8(\beta^{-1})))$ overhead.*

Moreover, any 'typical' double sampler has overhead at least $\exp(\Omega(\log^2(\beta^{-1})))$.

We refer the reader to Section 5.9.1 for the exact restrictions on X for which the lower bound holds.

Locally Testable and List-Decodable Codes.

While it is likely applying our double sampler machinery to the arguments of [121] would give list-decodable codes with substantially improved rate, this was already achieved subsequent to [121]'s work using walks on expanders [10]. Instead, using a variant of our machinery we give a closely related application to distance amplification for locally testable codes with near-optimal distance-alphabet trade-off. The resulting codes are list-decodable and preserve testability up to a log factor in the alphabet.

Theorem 5.2.14 (Large Distance List-Decodable LTCs (Informal Theorem 5.10.21))**.** *For all large enough* $k \in \mathbb{N}$ *and* $\varepsilon > 0$*, there exists an explicit family of* $\mathbb{F}_2$*-linear codes with*

1. *Distance:* $1 - 2^{-k} - \varepsilon$
2. *Alphabet Size:* 2^k
3. *Rate:* $\varepsilon^{O(k)}$

Moreover, the codes are efficiently list decodable up to distance $1 - 2^{-\Omega(k)}$ *and locally testable in* $O(1)$ *queries with soundness* $\Omega(\frac{1}{k \log(1/\varepsilon)})$.

Prior distance amplification techniques for LTCs [268] have exponentially worse distance-alphabet trade-off, but improved rate. To our knowledge the above are the first large distance list-decodable LTCs. We also give a candidate technique for amplifying distance while maintaining *constant* soundness (removing the k factor in the denominator above) using HDX, and prove the method works on the complete complex. We refer the reader to Section 5.10 for further details on these notions.

Degree Lower Bounds for HDX.

One of the formational results in the study of expansion is the Alon-Boppana theorem [307], roughly stating any family of bounded-degree λ-expanders have degree at least $\frac{2}{\lambda^2}$. Despite degree being a critical parameter in application (degree controls the 'blow-up' induced by using an HDX as a gadget), in high dimensions we understand very little about optimal degree. The best known constructions of λ-TD complexes have degree $\lambda^{\tilde{O}(d^2)}$. Is this optimal? Could $\lambda^{o(d^2)}$ be achieved?

We take the first step toward answering this question. Combining our sampling results with degree lower bounds of [87], we show super-exponential lower bounds for certain special classes of HDX, including polynomial skeletons of any hyper-regular λ-TD complex.

Theorem 5.2.15 (Degree Lower Bounds (Informal Theorem 5.9.8))**.** *Fix $\lambda < 1$, $d \in \mathbb{N}$, and $k \leq \sqrt{d}$. Let X be the k-skeleton of a d-uniform hyper-regular λ-TD complex. Then*

$$deg(X) \geq 2^{\Omega_\lambda(k^2)}$$

Theorem 5.2.15 exhibits yet another threshold phenomenon at the TD-barrier, since there exist 1-TD complexes [170] which have exponential degree at any level. While the latter are not hyper-regular, they are still reasonably balanced and our techniques extend to this regime.[5]

If one could prove a true Chernoff bound for λ-TD complexes (rather than our concentration of $\exp(-\sqrt{k})$), Theorem 5.2.15 could be improved to showing degree $2^{\Omega(d^2)}$ at the *top level* of X, albeit still under the assumption of hyper-regularity. While bounded-degree hyper-regular HDX do exist [159], their degree is substantially worse and removing

[5]Namely, we only really require that the underlying graph of every link does not have a (roughly) $\exp(-\sqrt{d})$ fraction of vertices making up 1/2 the measure. Unfortunately, constructions such as the Ramanujan complex have extremely unbalanced links, and actually do fail this property. See Section 5.9.2 for further discussion.

this constraint from our lower bound remains an important open problem.

Separating MLSI and Reverse Hypercontractivity.

The Modified Log-Sobolev Inequality (MLSI) is a powerful analytic inequality used to bound the mixing time of Markov chains and previous weakest sufficient condition for reverse hypercontractivity [304]. To the best of our knowledge, it was open until this work whether MLSI was *necessary*. We resolve this question for the setting allowing a leading constant $C > 1$.

Corollary 5.2.16 (Separating RHC and MLSI (Informal Corollary 5.9.17))**.** *There exist constants C, q, q' such that for infinitely many $N \in \mathbb{N}$, there exist (C, q, q')-reverse hypercontractive operators on N vertices with vanishing MLSI constant:*

$$\rho_{MLSI} \leq \tilde{O}\left(\frac{1}{\log(N)}\right)$$

This separation is essentially the strongest possible for operators with constant expansion, since any such operator has LSI (and therefore MLSI) at least $\Omega(\frac{1}{\log(N)})$ [107].

A Frankl–Rödl Theorem.

The Frankl–Rödl Theorem [154] is a broadly applied result in extremal combinatorics bounding the independence number of the graph on $\{0,1\}^n$ connecting strings of fixed intersection size. Lemma 5.2.6 implies an analogous result for the γk-step down-up walk on HDX.

Corollary 5.2.17 (Frankl–Rödl for HDX (Informal Claim 5.9.19))**.** *Fix $\gamma > 0$ and let X be a k-uniform nice complex for k sufficiently large. The γk-step down-up walk has independence number*

$$\alpha \leq \exp(-\Omega_\gamma(k)) \cdot |X(k)|$$

In comparison to Frankl–Rödl, the above bounds the size of sets avoiding inter-

sections of size *at most* instead of *exactly* $(1-\gamma)k$. A similar issue appears when using reverse hypercontractivity toward this end in the cube and is handled by a separate Fourier analytic argument in [61]. It would be interesting to extend our results to the exact case to give a true analog of Frankl–Rödl.

'It Ain't Over Till It's Over'.

Reverse hypercontractivity was used to resolve Friedgut and Kalai's 'It Ain't Over Till It's Over' Conjecture [302], which bounds the tail of random restrictions in the hypercube. The core of their proof was an analog for the closely related noise operator previously used in analysis of non-interactive correlation distillation protocols [303]. We generalize the latter theorem to HDX:

Theorem 5.2.18 (Noise Operator Tail Bounds (Informal Theorem 5.9.18))**.** *Fix $\rho \in (0,1)$ and $k \in \mathbb{N}$ sufficiently large. For any k-uniform nice complex X and $f : X(k) \to [0,1]$ of density μ:*

$$\mathbb{P}\left[T_\rho f \notin [\delta, 1-\delta]\right] \leq \delta^{O_{\mu,\rho}(1)}$$

5.3 Related Work

High Dimensional Expanders.

Spectral high dimensional expansion was developed over a series of works [141, 234, 309, 124] building on prior notions of high dimensional expansion in topology [163, 291, 283, 182, 233, 230]. Basic higher order random walks were introduced by Kaufman and Mass [234], with the swap operators later introduced independently in [109, 9]. A great deal of work has been done at the intersection of high dimensional expanders and analysis, including determining the optimal spectral gap of the down-up walks [234, 239, 111, 11, 243], developing a general theory of Fourier analysis [236, 111, 162, 39, 187, 176], and applications thereof to agreement testing [124, 109, 236, 177], mixing of Markov Chains [25, 24, 96, 94, 286, 66, 146, 23] (among many others), and algorithms [132, 38, 52].

Two works [242, 232] in the HDX literature also study Chernoff-style bounds, albeit in very different regimes from our results. Kaufman and Sharakanski [242] study concentration for global functions over repeated random walks on k-faces, which is incomparable to our setting. Kaufman, Kyng, and Song [232] prove scalar and matrix Chernoff bounds under ℓ_∞-*independence*, a strengthening of spectral HDX that only holds for dense complexes. While in this work we are primarily interested in proving concentration for *bounded-degree* complexes, it is interesting to ask whether their stronger result (Chernoff for matrix-valued functions) extends to the sparse regime.

In [39, 187], the authors prove a notion called "global" hypercontractivity for high dimensional expanders. While it is known that reverse hypercontractivity follows from the standard hypercontractive inequality [304], we are not aware of a reduction from the weaker global variant. Our work also differs substantially from [39, 187] in its tools and regime of application.

Samplers and Chernoff-Hoeffding.

Samplers are among the most classical tools in theoretical computer science, see e.g. [167, 343, 362] for surveys of their many constructions and applications. De-randomized Chernoff bounds are an important sub-family of sampler graphs that exhibit optimal tails in the degree of the graph. Sparse examples were constructed in [6, 165] using walks on expanders, and there is a great deal of literature toward understanding what general families of hypergraphs admit such concentration, e.g. under limited independence [337, 338, 339, 318], negative correlation [133], for edge colorings [313], under Dobrushin uniqueness [317], for strongly Raleigh distributions [273], and most recently under ℓ_∞-independence [232]. Such bounds have many interesting applications beyond those discussed in this work.

Inclusion samplers were first introduced in Impagliazzo, Kabanets, and Wigderson's [215] work on agreement testers and PCPs with strong soundness. Prior to Theorem 5.2.1,

there were only two known systems of inclusion samplers with optimal tails: the complete complex [215], and curves [299]. Both examples were used toward the construction of PCPs with strong soundness, but failed to achieve optimal parameters due to their blow-up in size. Several sparse variants of inclusion samplers were known to admit weaker 'Chebyshev' type bounds, including high dimensional expanders [124] and the Grassmann [215]. The latter also lead to PCPs with strong soundness [129], but failed to achieve optimal parameters due to its polynomial tail. Theorem 5.2.1 is the first to achieve the best of both worlds, though there remain substantial barriers toward its use for PCPs.

Reverse Hypercontractivity.

Reverse hypercontractivity was first shown by Borell [73] and first utilized in theoretical computer science by [303] who observed the inequality implies fine-control of mixing between sets (akin to Lemma 5.2.6). The result has since found great use in hardness of approximation [302, 145, 344, 61, 229], social choice [305, 249, 301], and extremal combinatorics [61, 229]. In 2013, [304] extended known bounds on reverse hypercontractivity to general product spaces and settings with bounded modified Log-Sobolev Inequalities, a surprising result given the failure of standard hypercontractivity in these settings. Dinur and Livni-Navon [128] were the first to apply these results to agreement testing, where they resolved the conjecture of [215] regarding exponential soundness for the Z-test and built the first 'combinatorial' PCPs with optimal soundness.

We note that MLSI bounds (and therefore RHC) are known for many 'dense' settings of HDX studied in the sampling literature (see e.g. [96]), with the most general prior condition being the notions of entropic independence and fractional log-concavity [22] which are significant strengthenings of spectral HDX. These methods all necessarily rely on density and cannot capture our theory.

Agreement Tests.

Agreement testing, also known as direct product testing, is a powerful tool in the construction of PCPs and locally testable codes [333, 31, 328, 168, 130, 116, 112, 122]. Agreement tests in the 99%-regime were studied by [168] and [115] in the complete complex, and extended to high dimensional expanders in [124, 109, 236]. [118] gave the first 'subcomplex' agreement theorem in the 99%-regime (the setting of Theorem 5.2.9) in the complete complex. Agreement testing in the 1%-regime was studied in the complete complex in [120, 215, 128], with the lattermost work giving optimal bounds. [215] also gave a de-randomized test over the Grassmann complex, which is polynomial size, but suffers from exponentially worse soundness. A different 1%-regime test on the Grassmann was also used to resolve the 2-2 Games Conjecture [254, 125, 47].

Finally, [177, 41, 110, 40, 113] studied the 1% (or closely related *list* agreement) regime on high dimensional expanders. These works identified key *topological* properties any complex must exhibit to be a good tester, and subsequently constructed complexes satisfying these notions, giving the first bounded degree 1%-agreement testers. The methods of [110, 113] in particular rely on Theorem 5.2.1 and Theorem 5.2.9. The soundness achieved by the above works, however, is inverse *logarithmic* in the dimension instead of inverse exponential (or even polynomial). Achieving inverse exponential soundness as in Theorem 5.2.10 for families of bounded degree complexes remains an important open question.

Open questions

1. While we are able to show optimal concentration for a fairly broad class of high dimensional expanders, in the weakest settings (namely under spectral independence and at the TD-Threshold), we are only able to prove exponential concentration at the top level. It is unclear whether this is a fundamental or purely technical barrier:

do such complexes satisfy a true Chernoff bound? As discussed above, a resolution of this question in the positive leads to better degree lower bounds for HDX, and in particular a $2^{\Omega(d^2)}$ lower bound for the top level of hyper-regular λ-TD complexes.

2. The best known constructions of high dimensional expanders are not hyper-regular. Is it possible to remove this constraint from our degree lower bound? While our technique holds even for 'reasonably balanced' complexes, it cannot handle objects like the Ramanujan complexes have extremely unbalanced links. It seems likely this is a technical rather than inherent barrier, and we conjecture some finer notion of concentration or application thereof may be able to remove this constraint.

3. Our sampling bounds for partite HDX depend on the concentration of the swap complex. Using negative correlation, we are only able to show Chernoff on $C_{\ell,k,n}$ in the regime where $n \geq \Omega(k\ell)$, which roughly corresponds to $(\varepsilon, e^{-\varepsilon^2 k/\ell^2})$-sampling at the top level of partite HDX. If one could prove Chernoff for $C_{\ell,k,n}$ when $n = \Theta(k)$, it would imply optimal bounds for the top level of partite HDX. Using the local-to-global entropy contraction framework of [96, 186], the former follows from showing optimal entropy contraction of the Kneser graphs. Does such a bound hold? Can one prove optimal inclusion sampling even in the more restrictive setting of product spaces?

4. We show that 'standard' simplicial complexes used in the construction of double samplers cannot achieve better than quasipolynomial overhead. Are there non-standard or non-simplicial constructions better served for this purpose?

5. Our 1%-regime test only holds on dense complexes due to the assumption of ℓ_∞-expansion. However, the main argument only requires reverse hypercontractivity and spectral gap of the down-up walk. Can the argument be completed without the assumption of ℓ_∞-expansion to give new sparse agreement testers in the low

acceptance regime?

Roadmap

In Section 5.4 we give more detailed and formal preliminaries on high dimensional expanders, random walks, samplers, and reverse hypercontractivity. In Section 5.5 we give the formal statement of our optimal sampling theorem and prove the two-sided case, deferring the remaining cases to Section 5.7. In Section 5.6 we prove reverse hypercontractivity. In Section 5.8 and Section 5.9 we give applications of reverse hypercontractivity to agreement testing, and analysis and combinatorics respectively. In Section 5.10 we give basic sampling lemmas for splitting trees and construct high distance list-decodable LTCs. Finally in Section 5.11 we show optimality of our sampling results in terms of requirements on X and strength of sampling.

5.4 Preliminaries and Notation

5.4.1 Graphs and Spectral Expansion

A weighted un-directed graph $G = (V, E, \mathbb{P})$ consists of a finite vertex set V, a set of edges $E \subseteq \binom{V}{2}$, and a distribution on the edges $\mathbb{P} : E \to (0, 1]$. For a vertex v we denote by $\mathbb{P}[v] = \frac{1}{2}\sum_{uv \in E} \mathbb{P}[uv]$. In this paper all graphs are assumed to be weighted (the weight is implicit in notation). We also assume that there are no isolated vertices.

Let $\ell_2(V) = \{f : V \to \mathbb{R}\}$. The graph G induces an inner product $\langle f, g\rangle = \mathbb{E}_v[f(v)g(v)]$ on $\ell_2(V)$, as well as a normalized adjacency operator $A : \ell_2(V) \to \ell_2(V)$ defined as

$$Af(v) := \mathop{\mathbb{E}}_{u\sim v}[f(u)] = \sum_{uv\in E} \frac{\mathbb{P}[uv]}{\sum_{w\sim v}\mathbb{P}[wv]} f(u).$$

It is well known that A is diagonalizable and has eigenvalues

$$1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{|V|} \geq -1,$$

where the first 'trivial' eigenvalue corresponds to the space of constant functions. A graph is called a spectral expander if all non-trivial eigenvalues are small.

Definition 5.4.1 (Spectral expander)**.** A graph $G = (V, E)$ is called a λ-*one-sided spectral expander* if $\lambda_2 \leq \lambda$. We say that G is a λ-*two-sided spectral expander* if $\lambda_2 \leq \lambda$ and $\lambda_{|V|} \geq -\lambda$.

We record the following inequality for λ-two-sided spectral expanders, a variant of the classical *expander-mixing lemma*. For every $f, g \in \ell_2(V)$,

$$|\langle f, Ag\rangle - \mathbb{E}[f]\,\mathbb{E}[g]| \leq \lambda \|f\|_2 \|g\|_2. \tag{5.5}$$

Bipartite Graphs and Bipartite Expanders

A bipartite graph is a graph where the vertex set can be partitioned to two independent sets $V = L \cup R$, called sides. We sometimes denote such graphs by $G = (L, R, E)$.

The Bipartite Adjacency Operator.

In a bipartite graph, we view each side as a separate probability space, where for any $v \in L$ (resp. R), $\mathbb{P}[v] = \sum_{w \sim v} \mathbb{P}[wv]$. We can define the *bipartite adjacency operator* as the operator $B : \ell_2(L) \to \ell_2(R)$ by

$$\forall f \in \ell_2(L), v \in R,\ Bf(v) = \mathop{\mathbb{E}}_{w \sim v}[f(u)]$$

where the expectation is taken with respect to the probability space L, conditioned on being adjacent to v. There is a similar operator $B^* : \ell_2(R) \to \ell_2(L)$ as the bipartite operator for the opposite side. As the notation suggests, B^* is adjoint to B with respect to the inner products of $\ell_2(L), \ell_2(R)$.

We denote by $\lambda(B)$ the spectral norm of B when restricted to $\ell_2^0(L) = \{\mathbf{1}\}^\perp$, the orthogonal complement of the constant functions (according to the inner product the

measure induces on L). Namely

$$\lambda(B) = \sup\{\langle Bf, g\rangle \mid \|g\|, \|f\| = 1, f \perp \mathbf{1}_L\}.$$

Definition 5.4.2 (Bipartite Expander)**.** Let G be a bipartite graph, let $\lambda < 1$. We say G is a λ-*bipartite expander*, if $\lambda(B) \leq \lambda$.

It is easy to show that a bipartite graph is a λ-bipartite expander if and only if it is a λ-one-sided spectral expander. So we use these terms interchangeably on bipartite graphs.

We record the following inequality for λ-bipartite expanders similar to (5.5). For every $f \in \ell_2(L), g \in \ell_2(R)$,

$$|\langle f, Ag\rangle - \mathbb{E}[f]\,\mathbb{E}[g]| \leq \lambda \|f\|_2 \|g\|_2. \tag{5.6}$$

5.4.2 Reverse Hypercontractivity

Hypercontractivity and reverse hypercontractivity are powerful analytic inequalities from boolean function analysis that bound the contraction behavior of operators between normed spaces. Recall that for $p \neq 0$ and a function f on a probability space, we denote by $\|f\|_p = \mathbb{E}[|f|^p]^{1/p}$ (for $p \leq 0$ this is only defined for f that are non-zero almost everywhere). We note that $p \mapsto \|f\|_p$ is monotone increasing.

Let V be a finite probability space and let $\ell_2(V) = \{f : V \to \mathbb{R}\}$. A monotone operator is an operator such that for any non-negative function $f : V \to [0, \infty)$, Af is also non-negative (i.e. for any $x \in X$, $Af(x) \geq 0$). For example, adjacency operators of graphs are always monotone.

Let A be a monotone operator. Let $1 \leq p < q$ and let $C > 0$. A is typically called

(p, q, C)-*hypercontractive* if for every $f : V \to \mathbb{R}$,

$$\|Af\|_q \leq C\|f\|_p.$$

It is well know that this is equivalent to 'two-function hypercontractivity', that for every $f, g : V \to \mathbb{R}$

$$\langle Af, g\rangle \leq C\|f\|_p\|g\|_{q'}$$

where $q' = \frac{q}{q-1}$ is q's Hölder conjugate [308]. *Reverse hypercontractivity* 'flips' this inequality in multiple ways: both in the direction of the inequality, and in the relation of p and q.

Definition 5.4.3 (Reverse hypercontractivity)**.** Let $q < p < 1$ such that $p, q \neq 0$. Let $C > 0$. Let A be a monotone operator. We say that A is (p, q, C)-reverse hypercontractive if for every $f : V \to \mathbb{R}^{\geq 0}$,

$$\|Af\|_q \geq C\|f\|_p$$

or equivalently if for every $f, g : V \to \mathbb{R}^{\geq 0}$ it holds that

$$|\langle Af, g\rangle| \geq C\|f\|_p\|g\|_{q'}$$

where $q' = \frac{q}{q-1}$ is q's Hölder conjugate (the equivalence is similar to standard hypercontractivity).

It is sometimes convenient to substitute $f' := f^p, g' = g^{q'}$ and write the two-function reverse hypercontractivity inequality as

$$\left|\langle Af'^{1/p}, g'^{1/q'}\rangle\right| \geq \mathbb{E}[f']^{1/p}\,\mathbb{E}[g']^{1/q'}.$$

Probably the simplest (and most useful) interpretation of this inequality is when f and g are indicators of sets $M, N \subseteq V$ respectively, and that A is an adjacency operator of

a graph (whose vertices are the points in V). Then this inequality has a combinatorial interpretation as a type of 'mixing' lemma. Namely that the probability a random edge hits M and N:

$$\mathop{\mathbb{P}}_{uv\in E}[u \in M, v \in N] \geq \mathbb{P}[M]^{1/p}\,\mathbb{P}[N]^{1/q'}.$$

This becomes a powerful tool in the regime where the sets are smaller than can be controlled by the expander-mixing lemma.

5.4.3 Simplicial Complexes

Simplicial complexes.

A pure d-uniform simplicial complex X is a finite set system (hypergraph) consisting of an arbitrary collection of sets of size d together with all their subsets (individually called 'faces'). The sets of size i in X are denoted by $X(i)$, and in particular, the vertices of X are denoted by $X(1)$ (we do not distinguish between vertices and the singletons of vertices). We write $X^{\leq j}$ to denote faces of X up to size j.

We note that this notation departs somewhat from the typical convention in the literature which denotes faces of size $i+1$ by $X(i)$, and what we have called a d-uniform simplicial complex is more typically referred to as a $(d-1)$-*dimensional* simplicial complex. We have chosen to adopt the former notation since most theorems in this work require a substantial amount of arithmetic on set sizes, and the latter convention quickly becomes overly cumbersome.

Probability over simplicial complexes.

Let X be a simplicial complex and $\mathbb{P}_d : X(d) \to (0,1]$ a density function on $X(d)$. This density function induces densities on lower level faces $\mathbb{P}_k : X(k) \to (0,1]$ by $\mathbb{P}_k(t) = \frac{1}{\binom{d}{k}} \sum_{s\in X(d), s\supset t} \mathbb{P}_d(s)$. Equivalently $\mathbb{P}_k$ is the density induced by drawing a d-face from $s \sim \mathbb{P}_d$, and a uniformly random k-face $t \subset s$.

When clear from the context, we omit the level of faces and just write $\mathbb{P}[T]$ or

$\mathbb{P}_{t\in X(k)}[T]$ for an event $T \subseteq X(k)$.

Links of faces.

Let X be a d-uniform simplicial complex. Let $k < d$ and $s \in X(k)$. The link of s (called a k-link) is a $(d-k)$-uniform simplicial complex defined by $X_s = \{t \setminus s \mid \ t \in X, t \supseteq s\}$. Note the link of the emptyset is X itself $X_\emptyset = X$.

Let $s \in X(k)$ for some $k \leq d$. The density function $\mathbb{P}_d$ on X induces a density function $\mathbb{P}^s_{d-k} : X(d-k) \to (0,1]$ on the link where $\mathbb{P}^s_{d-k}[t] = \frac{\mathbb{P}[t\cup s]}{\mathbb{P}[s]\binom{d}{k}}$. We usually omit s in the probability, and for $T \subseteq X_s(k)$ write $\mathbb{P}_{t\in X_s(k)}[T]$.

Connected Complexes.

We call a complex connected if the graph underlying every link is connected.

Partite complexes.

A d-partite simplicial complex is a d-uniform complex whose vertices can be partitioned into d disjoint sets (typically called "parts" or "colors")

$$X(0) = X[1] \cup X[2] \cup \ldots \cup X[d]$$

such that every $s \in X(d)$ has exactly one vertex from each part.

Let X be a d-partite simplicial complex. The *color* of a face $t \in X(k)$ is $col(t) = \{i \in [d] \mid \ t \cap X[i] \neq \emptyset\}$. Let $F \subseteq [d]$. We denote by $X[F] = \{s \in X \mid \ col(s) = F\}$ the faces of X with color F. The projection of the *complex* X onto F is the sub-simplicial complex $X^F = \{s \in X \mid \ col(s) \subseteq F\}$. When X is endowed with a density $\mathbb{P}_d$, X^F has a naturally induced density $\mathbb{P}^F$ by sampling $\sigma \sim \mathbb{P}_d$, and outputting the projection σ_F. Quantitatively, the induced distribution can be written as

$$\mathbb{P}^F(s) = \sum_{t\in X(d), s\subseteq t} \mathbb{P}[t].$$

Finally, when $F = \{i\}$ is a singleton we just write $X[i]$, X^i, and x_i for brevity.

Given a generic complex X, it will often be useful to consider X's 'partitification' $P = P(X)$, which simply includes every possible ordering of the faces of X as tuples. Formally, endow the faces of X with an arbitrary order and define

$$P(k) \coloneqq \left\{(s,\pi) \coloneqq \left\{(s_{\pi(1)},1),(s_{\pi(2)},2),\ldots,(s_{\pi(k+1)},k)\right\} \;\middle|\; s \in X(k), \pi \in S_k\right\},$$

where S_k is the group of permutations on k-letters and the measure of a face (s,π) is inherited naturally as $\frac{1}{k!}\mathbb{P}_X[s]$. Note that $P(X)$ does not depend on the choice of ordering, which is simply a notational convenience.

Degree of Simplicial Complexes.

Degree is a critical parameter of simplicial complexes capturing the (local) 'blow-up' incurred by moving to higher uniform faces. We define the (max) degree of a complex with respect to i-faces as:

$$\deg^i(X) = \max_{v \in X(1)} |\{s \in X(i) : v \in s\}|$$

We write just $\deg(X) \coloneqq \deg^d(X)$ to denote degree with respect to top level faces of X, and $\deg^{(i)}(v)$ for the degree of a specific vertex. An infinite family of d-uniform simplicial complexes $\{X_n\}$ is called *bounded-degree* if there exists a (dimension-dependent) constant C such that $\deg^i(X) \le C$ for all $i \le d$.

Hitting Set.

It will occasionally be useful for us to use a common variant of sampling on simplicial complexes called *hitting set*. We call a complex (γ, i)-hitting if for any $A \subset X(1)$:

$$\mathop{\mathbb{P}}_{\sigma \in X(i)}[\sigma \subset A] \le \mathbb{P}[A]^i + \gamma$$

In other words, if the probability $\sigma \in X(i)$ *hits* $X(1) \setminus A$ is at least $1 - \mu(A)^i - \gamma$. If a complex is (γ, i)-hitting for all sizes i, we just call it γ-hitting.

5.4.4 Higher Order Random Walks

Simplicial complexes come equipped with several natural families of random walks generalizing the standard random walk on a graph. Toward this end, let X be a d-uniform simplicial complex and $\ell < k \leq d$ we define the standard 'averaging' or 'random walk' operators that move up and down the complex:

Down and up operators.

The *down operator* $D_{k,\ell} : \ell_2(X(k)) \to \ell_2(X(\ell))$ is the bipartite operator of the containment graph of $(X(k), X(\ell))$, that is:

$$D_{k,\ell} f(s) = \mathop{\mathbb{E}}_{t \supseteq s} [f(t)] .$$

The adjoint of $D_{k,\ell}$ is the *up operator*, $U_{\ell,k} : \ell_2(X(\ell)) \to \ell_2(X(k))$, given by:

$$U_{\ell,k} g(t) = \mathop{\mathbb{E}}_{s \subseteq t} [g(s)] .$$

We also write $U_k \coloneqq U_{k-1,k}$ and $D_k \coloneqq D_{k,k-1}$ as shorthand throughout.

The composition of the up and down operators $U_{\ell,k} D_{k,\ell}$, called the "down-up" walk, is the normalized adjacency matrix of the graph whose vertices are $X(k)$ and whose edge distribution is defined by sampling $t \sim \mathbb{P}_\ell$, then $s, s' \sim \mathbb{P}_k$ conditioned on containing t. We denote the corresponding (bipartite) graph by $(X(k), X(\ell))$. For brevity, we sometimes just write $UD_{k,\ell}$ instead of $U_{\ell,k} D_{k,\ell}$.

Noise operator.

The noise operator is one the most classical objects of study in Boolean function analysis. Given $\rho \in [0, 1]$, the standard noise operator $T_{k,\rho}$ operates on the hypercube $\mathbb{F}_2^k$

by re-sampling each bit uniformly with probability $1-\rho$.

The noise operator has a natural extension to simplicial complexes [39, 187]. It is convenient to define $T_{k,\rho} : \ell_2(X(k)) \to \ell_2(X(k))$ by its action on a face $\sigma \in X(k)$ by the following process:

1. Sample $\ell \sim Bin(k, \rho)$
2. Sub-sample $\tau \subset \sigma$ of size ℓ
3. sample $\sigma' \in X(k)$ conditioned on containing τ.

It is easy to check that when X is a product space,[6] this exactly recovers the standard noise operator. We note it is also possible to write the noise operator as a convex combination of down-up walks:

$$T_{k,\rho} = \sum_{\ell=0}^{k} \binom{k}{\ell} \rho^{\ell}(1-\rho)^{k-\ell} UD_{k,\ell}.$$

Swap walks.

Let X be a d-uniform simplicial complex. Let i, j be so that $i + j \leq d$. The swap walk $S_{i,j} = S_{i,j}(X)$ is the bipartite adjacency operator of the graph $(X(i), X(j), E)$. An edge $\{s_i, s_j\}$ is chosen in this graph by first selecting a face $t \in X(i+j)$, and then partitioning it to $t = s_i \cup s_j$ where $s_i \in X(i)$ and $s_j \in X(j)$ uniformly at random. This walk was defined and studied independently by [9] and by [109].

Colored swap walks.

The standard swap walks are not well behaved on partite complexes, but there is a useful analog for this setting defined in [109]. Let X be a d-uniform partite simplicial complex, and $F_1, F_2 \subseteq [d]$ be two disjoint subsets. The colored swap walk $S_{F_1,F_2} = S_{F_1,F_2}(X)$ is the bipartite adjacency operator of the graph $(X[F_1], X[F_2], E)$. An edge $\{s_1, s_2\}$ is

[6] One can always express a product space $\otimes_i \Omega_i$ as a partite simplicial complex by defining each coordinate as a part, see [39] for further details.

chosen in this graph by first selecting a face $t \in X[F_1 \cup F_2]$, and then partitioning it to $t = s_1 \cup s_2$ according to its colors.

A useful observation is that the spectral expansion of swap walks (and color swap walks) is monotone in the following sense.

Observation 5.4.4. Let X be a d-uniform simplicial complex. Then for every i and $j' < j$, $\lambda(S_{i,j'}) \leq \lambda(S_{i,j})$. The same holds for partite complexes. For every disjoint F_1, F_2 and $F_2' \subseteq F_2$, $\lambda(S_{F_1,F_2'}) \leq \lambda(S_{F_1,F_2})$.

Proof. The proof follows from the fact that one can factor

$$S_{i,j'}(X) = D_{j,j'} S_{i,j}(X)$$

and the observation that $D_{j,j'}$ contracts 2-norms. A similar argument is true for colored swap walks in the partite case, replacing $D_{j,j'}$ with the corresponding bipartite graph between the F_2-colored faces and the F_2'-colored faces they contain. □

5.4.5 Sampler Graphs

Sampler graphs are bipartite graphs $G = (L, R, E)$ where a random vertex $v \in L$ "sees" any large enough set in R with approximately the correct probability. Here we discuss a few classical variants of samplers and their relations, and refer the reader to [167] for further discussion.

Definition 5.4.5 (Multiplicative sampler)**.** Let $G = (L, R, E)$ be a bipartite graph and $\alpha, \beta, \delta > 0$. G is an (α, β, δ)-*multiplicative sampler* if for every set $A \subseteq R$ of size $\mathbb{P}[A] \geq \alpha$ it holds that

$$\mathop{\mathbb{P}}_{v \in L}\left[\left|\mathop{\mathbb{P}}_{u \sim v}[u \in A] - \mathbb{P}[A]\right| > \delta \mathbb{P}[A]\right] \leq \beta.$$

Note that the definition of a sampler is not a priori symmetric, L and R have different roles. We will also study a related *additive* notion of samplers.

Definition 5.4.6 (Additive sampler)**.** Let $G = (L, R, E)$ be a bipartite graph and $\beta, \varepsilon > 0$. G is an (ε, β)-*additive sampler* if for every set $A \subseteq R$ it holds that

$$\mathop{\mathbb{P}}_{v \in L}\left[\left|\mathop{\mathbb{P}}_{u \sim v}[u \in A] - \mathbb{P}[A]\right| > \varepsilon\right] \leq \beta.$$

Finally, sometimes it is also useful to sample *functions* instead of sets. toward this we introduce the following definition in additive notation.

Definition 5.4.7. Let $G = (L, R, E)$ be a bipartite graph and $\varepsilon, \beta > 0$. G is a (ε, β)-*function additive sampler* if for every $f : R \to [0, 1]$ with expectation $\mathbb{E}_{v \in R}[f(v)] = \mu$, it holds that

$$\mathop{\mathbb{P}}_{v \in L}\left[\left|\mathop{\mathbb{E}}_{u \sim v}[f(v)] - \mu\right| > \varepsilon\right] < \beta,$$

and similarly for a (α, β, δ)-*function multiplicative sampler*.

Basic Sampler Properties

While the definition of samplers are not symmetric with respect to L and R, a near-tight correspondence (L, R, E) and (R, L, E) is given in [215, Lemma 2.5]. We repeat their proof in Section 5.14 in a more general setup.

Claim 5.4.8. Let $\beta, \delta > 0$, let $\delta' > \delta$ and $\alpha < \frac{\min\{\delta, 0.5\}}{1+\delta}$. Then for every (α, β, δ)-sampler $G = (L, R, E)$, it holds that $G_{op} \coloneqq (R, L, E)$ is a $(\frac{1-\alpha(1+\delta)}{\alpha(\delta'-\delta)}\beta, 2\alpha, \delta')$-sampler.

We also note that additive samplers and multiplicative are equivalent, at least in the weak sense as below. We prove this claim in Section 5.14.

Claim 5.4.9. Let $G = (L, R, E)$ be a bipartite graph.

1. If G is a (β, δ)-additive sampler then G is a $(C\delta, \beta, \frac{1}{C})$-multiplicative sampler for any $C > 1$.

2. If G is a (α, β, δ)-multiplicative sampler for $\alpha \leq \frac{1}{2}$. Then G is a (β, δ)-additive sampler, where $\delta = \max\{\delta, (1 + \delta)(\alpha + p)\}$ and $p = \max_{v \in R} \mathbb{P}[v]$.

Under a slightly stronger assumption one can also remove the dependence on δ in the second item.

Claim 5.4.10. Let $\beta, \alpha_0 > 0$. If for every $\alpha > \alpha_0$ it holds that G is an $(\alpha, \beta, \frac{\alpha_0}{\sqrt{\alpha}})$-multiplicative sampler, then G is a$(\beta, 2(\alpha + p))$-additive sampler where $p = \max_{v \in R} \mathbb{P}[v]$.

Below we show that every additive sampler is also a function additive sampler (albeit with worse parameters). We did not try to optimize parameters and it could be the case that a better reduction exists.

Claim 5.4.11. Let $G = (L, R, E)$ be an (ε, β)-additive sampler such that every $r \in R$ has degree at least k and every $v \in L$ has probability at most $\frac{1}{k}$. Assume that $\exp(-0.01\varepsilon^2 k)) < \frac{1}{4}$. Then G is also a $(4\beta, 2\varepsilon)$-function additive sampler.

We prove the claim in Section 5.14. We note that [167] also presents a reduction from an additive sampler to a function additive sampler is presented. The reduction there changes the underlying graph, which is why we prove a reduction more suitable to our needs.

5.4.6 Concentration of Measure

Sampler graphs are a special case of the powerful concept of concentration of measure, a viewpoint we will also take throughout this work. We follow the standard notation of [76] adapted in the natural way from product measures to simplicial complexes. Given a simplicial complex X and a function $f : X(k) \to \mathbb{R}$, let $Z = f(x_1, \ldots, x_k)$ denote the random variable distributed as $f(x)$ where $x \in X(k)$. We are interested in understanding the concentration of Z around its mean. Z is said to satisfy *subexponential* concentration if there exists a constant $c > 0$ such that

$$\mathbb{P}[|Z - \mathbb{E}[Z]| > t] \leq \exp(-ct),$$

and *subgaussian* concentration if

$$\mathbb{P}[|Z - \mathbb{E}[Z]| > t] \leq \exp(-ct^2).$$

Classical concentration bounds (e.g. on the cube, products) typically hold for functions f satisfying certain Lipschitz-type conditions. Toward this end, define the Z-dependent variable $Z'_{(i)} = f(x_1, \dots z_i, \dots, x_k)$ where z_i is sampled conditional on x_{-i}.[7] Abusing notation slightly, we will call a function ν-Lipschitz if its squared difference with respect to the down-up walk is bounded:

Definition 5.4.12 (ν-Lipschitz function)**.** Let X be a simplicial complex and $\nu > 0$. We call a function $f : X(d) \to \mathbb{R}$ ν-Lipschitz if with probability 1:

$$\sum_{i=1}^{d} (Z - Z'_{(i)})_+^2 \leq \nu,$$

where $(z)_+ = \max\{z, 0\}$.

We will occasionally rely on a stricter variant that requires the difference of f on any two neighboring k-faces be bounded.

Definition 5.4.13 (ν-bounded difference)**.** We say a function $f : X(d) \to \mathbb{R}$ has ν-bounded difference if for every $s \in X(d)$ and neighboring s' of the down-up walk:

$$(f(s) - f(s'))^2 \leq \frac{\nu}{d}.$$

Most of our work focuses on a special case of functions with ν-bounded difference we call *lifted* functions.

[7] Recall that even in the non-partite setting, we assign each k-face an arbitrary order, making $Z'_{(i)}$ well defined.

Definition 5.4.14 (Lifted functions)**.** Let X be a d-uniform simplicial complex and $f : X(k) \to \mathbb{R}$ any function. For any $k < k' \leq d$, the k'-lift of f is the function:

$$U_{k,k'} f(s) = \mathbb{E}_{s' \subset s}[f(s')]$$

We remark that lifted functions (sometimes called 'degree' or 'level'-k functions) are fundamental objects in boolean analysis and heavily studied in the HDX literature [111, 38, 162]. As discussed in the introduction, inclusion sampling is in one-to-one correspondence with concentration bounds for lifted functions. More generally, bipartite sampling guarantees on $G = (L, R, E)$ correspond to concentration for functions of the form $A_G f$, where $f : R \to [0, 1]$ and $A_G : \mathbb{R}^R \to \mathbb{R}^L$ is the normalized bipartite adjacency matrix of G.

On simplicial complexes, we will typically be interested in understanding concentration at various levels of the complex. One frequently useful fact is that once concentration has been established for k-faces, it can typically be 'raised' or 'lowered' to functions on other levels via simple reductions to concentration of the complete complex. In cases, this can even be used to *improve* bounds on lower levels by bootstrapping from stronger bounds from larger set sizes. We record the following basic results to this effect here and use them throughout. Their proofs are given in Section 5.14.

Given integers $k \leq d$, a d-uniform simplicial complex X, and a function $f : X(k) \to \mathbb{R}$, denote

$$\pi_{up}^{d,k,f}(t) = \max_{s \in X(d)} \left\{ \mathop{\mathbb{P}}_{r \subset s} \left[f(r) - \mathop{\mathbb{E}}_{r \subset s}[f] > t \right] \right\}$$

and similarly

$$\pi_{low}^{d,k,f}(t) = \max_{s \in X(d)} \left\{ \mathop{\mathbb{P}}_{r \subset s} \left[f(r) - \mathop{\mathbb{E}}_{r \subset s}[f] < -t \right] \right\}$$

to be the worst-case concentration of f restricted to the induced complete complex on d-faces of X.

Lemma 5.4.15 (Raising Concentration)**.** *Let X be a d-uniform simplicial complex, $k \leq d$, and $f : X(k) \to [0,1]$ a function satisfying*

1. ***Upper Tail:*** $\mathbb{P}_{X(k)}[f - \mathbb{E}[f] > t] \leq up(t)$
2. ***Lower Tail:*** $\mathbb{P}_{X(k)}[f - \mathbb{E}[f] < -t] \leq low(t)$.

for some functions $up, low : \mathbb{R}_+ \to [0,1]$. Then the d-lift $U_{k,d}f : X(d) \to \mathbb{R}$ satisfies:

1. ***Upper Tail:*** $\mathbb{P}_{X(d)}[U_{k,d}f - \mathbb{E}[f] > t] \leq up(\frac{t}{2})(1 - \pi_{low}^{d,k,f}(\frac{t}{2}))^{-1}$
2. ***Lower Tail:*** $\mathbb{P}_{X(d)}[U_{k,d}f - \mathbb{E}[f] < -t] \leq low(\frac{t}{2})(1 - \pi_{up}^{d,k,f}(\frac{t}{2}))^{-1}$.

We remark that since concentration in the complete complex is typically quite good, for most function classes of interest the latter term is close to 1 and very little is lost lifting the concentration to level d. We will also use a variant of this result for 'lowering' concentration bounds specified to Lipschitz functions.

Lemma 5.4.16 (Lowering Concentration)**.** *Let X be a d-uniform simplicial complex and $k \leq d$. Assume there exist functions $up(t,\nu)$ and $low(t,\nu)$ such that any ν-Lipschitz $f : X(d) \to \mathbb{R}$ satisfies:*

1. ***Upper Tail:*** $\mathbb{P}[f - \mathbb{E}[f] > t] \leq up(t,\nu)$
2. ***Lower Tail:*** $\mathbb{P}[f - \mathbb{E}[f] < -t] \leq low(t,\nu)$.

Then any function $f' : X(k) \to \mathbb{R}$ with ν-bounded difference satisfies:

1. ***Upper Tail:*** $\mathbb{P}[f' - \mathbb{E}[f'] > t] \leq up(\frac{t}{2}, \frac{k}{d}\nu) + e^{-\frac{t^2}{4\nu}}$
2. ***Lower Tail:*** $\mathbb{P}[f' - \mathbb{E}[f'] < -t] \leq low(\frac{t}{2}, \frac{k}{d}\nu) + e^{-\frac{t^2}{4\nu}}$.

5.4.7 High Dimensional Expanders

In this section we define the notions of high dimensional expansion used throughout this work and dicsuss their relation with high order random walks.

Local Spectral Expanders.

The most standard notion of spectral high dimensional expansion is 'local-spectral expansion', due to [124, 309].

Definition 5.4.17 (Local-spectral expander)**.** Let X be a d-uniform simplicial complex. We say that X is a λ-one-sided (two sided) high dimensional expander if for every $i \leq d-2$ and $s \in X(i)$, the graph underlying X_s is a λ-one-sided (two sided) spectral expander.

A key property of local-spectral expanders is that they imply the expansion of associated high order random walks [234, 124, 239, 111, 11]. We first state such a bound for the single and multi-step down-up walks (or more accurately their corresponding bipartite operator).

Theorem 5.4.18 ([11])**.** *Let X be a λ-one-sided d-uniform high dimensional expander. Then for every $\ell < k \leq d$ it holds that $\lambda(D_{k,\ell}) \leq \sqrt{\frac{\ell}{k}} \cdot (1+\lambda)^{\frac{k-\ell}{2}}$.*

In the special case of the single step operator, we will occasionally use the following refined bound.

Theorem 5.4.19 ([11])**.** *Let X be a λ_i-one-sided d-uniform high dimensional expander. Then for every $\ell < k \leq d$ it holds that $\lambda(D_k) \leq \sqrt{\frac{1}{k}\prod_{i=0}^{k-2}(1-\lambda_i)}$*

A critical observation of [9, 109] is that by removing the laziness inherent in the down-up walk, one can substantially improve expansion. We state an improvement of their bounds for these swap walks by Gur, Lifshitz, and Liu [187]:[8]

Theorem 5.4.20 ([187])**.** *Let X be a λ-two-sided high dimensional expander. Then*

$$\lambda(S_{\ell_1,\ell_2}) \leq \sqrt{\ell_1 \ell_2}\lambda$$

[8]Formally this result is stated in [187] for λ-products (defined below), but follows for λ-two-sided HDX since their partifications are λ-products.

Recently, [12] proved tighter quantitative bounds that are non-trivial for any underlying local-spectral expansion. We will not need such fine-grain control in this work.

In the partite case, the colored swap walks also have known expansion bounds assuming partite local-spectral expansion [109, 187].

Theorem 5.4.21 ([109]). *Let X be a partite λ-one-sided high dimensional expander. Then*

$$\lambda(S_{F_1,F_2}) \leq |F_1|\,|F_2|\,\lambda.$$

We will also need a bound on the spectrum of the averaging operators based on local-spectral expansion. This has been the subject of intense study [234, 124, 111, 239, 11], with the best known bounds given by Alev and Lau [11].

λ-Products.

Another useful notion of partite high dimensional expansion is to look directly at the expansion of the bipartite graphs between different components. We say that X is a λ-product (a term coined in [187]), if for every face $s \in X$ and every two sides $X_s[i], X_s[j]$, the bipartite graph between the two sides is a λ-one-sided spectral expander. Dikstein and Dinur showed any partite local-spectral expander is a λ-product.

Claim 5.4.22 ([109]). Let X be a k-partite $\frac{\lambda}{1+\lambda}$-one-sided local-spectral expander. Then X is also a λ-product.

λ-Trickling-Down Complexes.

Oppenheim's trickling-down theorem [309] shows that in any connected complex, expansion in the top links 'trickles down' to expansion at lower levels.

Theorem 5.4.23 (The Trickling-Down Theorem [309]). *Let X be a connected d-uniform simplicial complex. Assume that for every $s \in X(d-2)$, it holds that all non-trivial eigenvalues of X_s are in $[-\tau, \lambda]$ for some $\tau, \lambda \geq 0$. Then all non-trivial eigenvalues of X*

are in

$$\left[-\frac{\tau}{1+(d-2)\tau}, \frac{\lambda}{1-(d-2)\lambda}\right].$$

As discussed in the introduction, the Trickling-Down Theorem introduces a 'phase transition' in the local-to-global behavior of HDX when co-dimension 2 links have expansion $\frac{1}{d-1}$. Past this point, expansion of the 1-skeleton of X can be inferred from expansion of the links. Before it, no guarantee is possible. This threshold-type behavior naturally suggests studying following definition of complexes 'approaching' the TD-barrier.

Definition 5.4.24 (λ-Trickling-Down Complex)**.** We call a d-uniform simplicial complex λ-Trickling-Down (λ-TD) if it is connected, and all co-dimension 2 links have (one-sided) expansion $\frac{\lambda}{d-1}$.

Spectral Independence.

Spectral independence is a closely related notion to local-spectral expansion introduced in the sampling literature [24]. Here we give an equivalent version of the definition in terms of link expansion, and refer the reader to [24] for the original definition and their equivalence.

Definition 5.4.25 (Spectral independence)**.** For $\eta > 0$, a d-uniform complex X is called η-spectrally independent (η-SI) if the graph underlying every co-dimension j link is a (one-sided) $\frac{\eta}{j}$-expander.

5.4.8 Nice Complexes

In the following sections of the paper we prove results for various standard notions of high dimensional expansion including two sided high dimensional expanders, and skeletons of sufficiently low-dimensional skeltons of partite and one-sided HDX. To compactify our statements, we bundle these notions into one definition henceforth dubbed 'nice complexes'. We first give a simplified version of the definition.

Definition 5.4.26 (Nice complex (Simplified))**.** Let $c > 0$. A k-uniform complex X is *c-nice* if it meets one of the following conditions:

1. X is a 2^{-cd}-two-sided HDX.
2. X is the k-skeleton of a 2^{-cd}-one-sided d-partite HDX for $d \geq 2k^2$.
3. X is the k-skeleton of a d-uniform complex Y with $\lambda(U_{d-1}D_d) = 1 - \frac{c}{d}$ and $d \geq 2k^2$.

We call X *c-locally nice* if every link of the complex is also c-nice.

In the above, the reader may notice that the second case is subsumed by the third by Theorem 5.4.18. In fact, we prove more general results for both two-sided and partite HDX in which this is not the case. The partite setting requires an object we call the *swap complex*, $C_{\ell,k,d}$, which is the $\frac{k}{\ell}$-uniform complex whose vertices are all ℓ-sets $\binom{[d]}{\ell}$ and whose $\frac{k}{\ell}$-faces are all possible families of fully disjoint ℓ-sets. We define the function $\mathbf{swap(k)} \in [k, \infty]$ to be the smallest $d \geq k$ such that for all ℓ dividing $k/2$, the swap complex $C_{\ell,k/2,d}$ satisfies a Chernoff bound, i.e. such that $(C_{\ell,k/2,d}(\frac{k}{2\ell}), C_{\ell,k/2,d}(1))$ is an $(\varepsilon, \exp(-\Omega(\varepsilon^2 \frac{k}{\ell}))$-additive function sampler for any $\varepsilon > 0$.

Definition 5.4.27 (Nice complex)**.** Let $c > 0$. A k-uniform complex X is *c-nice* if it meets one of the following conditions:

1. X is a 2^{-cd}-two-sided HDX.
2. X is the k-skeleton of a 2^{-cd}-one-sided d-partite HDX for $d \geq \mathrm{swap}(k)$.
3. X is the k-skeleton of a $\frac{1}{d}2^{-cd^\alpha}$-two-sided d-uniform HDX for $d \geq k^{\frac{2}{1+\alpha}}$ for $\alpha \in [0, 1]$

We call X *c-locally nice* if every link of the complex is also c-nice.

It is immediate from negative correlation that $\mathrm{swap}(k) \leq O(k^2)$, but we conjecture in fact $\mathrm{swap}(k) = k$. If this holds, all our results apply to the top level of sufficiently strong partite HDX (including, e.g., to product spaces).

Note that skeletons of nice complexes are nice (see [276, Theorem 3.5] for the third case), and that if X is c-nice for one of the first two items, it is also c-locally nice. It is simple computation to verify that skeletons of λ-TD and η-SI complexes are locally nice.

Claim 5.4.28. Let $\lambda < 1$ and $\eta > 0$. For any d-uniform complex Y:

1. If Y is λ-TD, then its $\sqrt{(1-\lambda)d}$-skeleton is e^{-1}-locally nice.
2. If Y is η-SI, then its $\sqrt{\frac{d}{\max\{2,\eta\}}}$-skeleton is e^{-2}-locally nice.

Proof. We first prove niceness. The first item follows from observing by Theorem 5.4.23 and Theorem 5.4.19 that the down-up walk on the $k = (1-\lambda)d$ skeleton of a λ-TD complex has $\lambda(U_{k-1}D_k) = 1 - \frac{e^{-1}}{d}$.

For the second item, let $k' = \frac{d}{\max\{2,\eta\}}$. By definition one can check the k'-skeleton of Y is a $\min\{\frac{3}{d}, \frac{\eta}{(1-1/\eta)d}\}$-one-sided local-spectral expander and Theorem 5.4.19 implies $\lambda(U_{k'-1}D_{k'}) \leq 1 - \frac{e^{-2}}{k'}$.

For local niceness, note that η-SI and λ-TD complexes have η-SI and λ-TD links respectively. Further, one can check directly that if X is a k-skeleton of d-uniform Y, then X_s is a $(k-|s|)$-skeleton of Y_s, which is $(d-|s|)$-uniform. The rate $\frac{d-|s|}{k-|s|} \geq \frac{d}{k}$ thus we can apply the same proof of niceness to every link as well. □

5.5 Concentration on (Two Sided) HDX

We first give the formal statement of our optimal inclusion sampling theorem in the general case.

Theorem 5.5.1 (Sampling on HDX)**.** *Fix $c > 0$. There are constants $c_1, c_2 > 0$ such that for any $k \in \mathbb{N}$, $i < k$, and k-uniform c-nice complex X, the containment graph $(X(k), X(i))$ is an (ε, β)-function sampler for*

$$\beta = \frac{c_1}{\varepsilon} \exp\left(-c_2 \varepsilon^2 \frac{k}{i}\right).$$

Moreover, if X is a k-skeleton of a d-uniform complex Y, then for every $k < k' \leq d$, the graph $(Y(k'), Y(i))$ is also an $(\varepsilon, O(\beta))$-function sampler.

We refer the reader to Section 5.7.3 for the exact λ dependence in the λ-TD case. In this section we will only prove the two-sided case of Theorem 5.5.1. We start by analyzing the more typical 'Chernoff-Hoeffding'-type setting, then show how to bootstrap this bound to prove the two-sided case of Theorem 5.5.1.

5.5.1 Chernoff-Hoeffding

In this section we introduce a new technique to prove 'k vs. 1' concentration on simplicial complexes X whose underlying strength depends on the expansion of the *swap walks* of X. The general form is fairly cumbersome (we refer the reader to Proposition 5.5.4, which bounds the moment generating function in terms of swap walk expansion), and for readability only state the result here for $\frac{1}{k}2^{-k^\alpha}$-two-sided HDX, interpolating between $\frac{1}{2k}$-HDX at $\alpha = 0$, and $2^{-\Omega(k)}$-HDX at $\alpha = 1$.

Theorem 5.5.2. *Let $\alpha \in [0,1]$ and X be a k-uniform $\frac{1}{k}2^{-k^\alpha}$-two-sided local-spectral expander. Then for any $f : X(1) \to [0,1]$ and $\varepsilon > 0$:*

1. *Upper tail:*

$$\mathop{\mathbb{P}}_{\{v_1,\ldots,v_k\}\in X(k)}\left[\frac{1}{k}\sum_{i=1}^{k} f(v_i) - \mu > \varepsilon\right] \leq 2\min\{k, c_\alpha\}\exp\left(-\frac{1}{12}\varepsilon^2 k^{\frac{1+\alpha}{2}}\right)$$

2. *Lower tail:*

$$\mathop{\mathbb{P}}_{\{v_1,\ldots,v_k\}\in X(k)}\left[\frac{1}{k}\sum_{i=0}^{k} f(v_i) - \mu < -\varepsilon\right] \leq 2\min\{k, c_\alpha\}\exp\left(-\frac{1}{12}\varepsilon^2 k^{\frac{1+\alpha}{2}}\right)$$

where $c_\alpha \leq e^{\frac{(\frac{3}{\alpha})^{\frac{1}{\alpha}}}{e^\alpha}}$.

We remark it is possible to use this approach to prove similar concentration for any λ-TD complex, but in Section 5.7.2 we will see a stronger bound for this specific regime based on a variant of the Herbst argument. As such we allow ourselves the additional constant factor here in expansion for convenience. Second, note that by setting α appropriately, Theorem 5.5.2 shows for any $c > 0$ that the containment graph $(X(k), X(1))$ of any 2^{-ck}-two-sided local spectral expanders is a (ε, β)-sampler for $\beta = \exp(-\Omega_c(-\varepsilon^2 k))$.

The core of Theorem 5.5.2 centers around the notion of *splittability* [9], which will allow us to de-correlate the variables in our complex to bound the moment generating function of $U_{1,k}f$. We introduce a slight variant of the standard definition convenient for our purposes.

Definition 5.5.3 (Balanced Splittability)**.** We call a d-uniform complex X $\{\gamma_i\}_{i\in[\lfloor \log(d)\rfloor -1]}$-balanced-splittable if for every i the 2^i-th swap walk expands:

$$\lambda(S_{2^i,2^i}) \leq \gamma_i$$

All applications of splittability in the literature lose *additive* factors in the γ_i. We introduce a new approach combining splittability with an 'internal' reduction to the complete complex that allows finer control over this error and instead incurs more manageable *multiplicative* loss.

Proposition 5.5.4. *For any $j \in \mathbb{N}$ and $r \in [0,1]$, let $k = 2^j$, and X be a k-uniform $\{\gamma_i\}$-balanced-splittable complex with $e^{\frac{5}{4}2^{-i}r^2k}\gamma_{j-1-i} < 1$. Then for any $f : X(1) \to [0,1]$:*

$$\mathop{\mathbb{E}}_{\{v_1,\ldots,v_k\}\in X(k)}\left[\prod_{i=0}^{k} e^{rf(v_i)}\right] \leq c_{r,\gamma} \mathop{\mathbb{E}}_{X(1)}[e^{rf}]^k$$

where $c_{r,\gamma} \leq \left(\prod_{i=0}^{j-1}\left(\frac{1-\gamma_{j-1-i}}{1-e^{\frac{5}{4}2^{-i}r^2k}\gamma_{j-1-i}}\right)^{2^i}\right)$.

While the definition of $c_{r,\gamma}$ may be somewhat hard to parse, note at least that for

any fixed r, $c_{r,\gamma}$ goes to 1 as γ tends to 0.

Proof. For any even ℓ, denote by $z_\ell : X(\ell) \to \mathbb{R}$ the 'partial' moment generating function:

$$z_\ell(\{v_1, \ldots, v_\ell\}) = \prod_{i=1}^{\ell} e^{rf(v_i)}.$$

Our goal is to bound $\mathbb{E}[z_k] = \mathbb{E}[\exp(r\sum_{i=1}^{k} f(v_i))]$ by induction on z_ℓ. Denote by $z_\ell^0 = \mathbb{E}[z_\ell]\mathbf{1}$ the projection of z_ℓ to the space of constant functions and by $z_\ell^\perp = z_\ell - z_\ell^0$ its orthogonal part. The first key observation is that we may split z_ℓ into two 'copies' of $z_{\ell/2}$ using the swap operator:

$$\mathop{\mathbb{E}}_{X(\ell)}[z_\ell] = \mathop{\mathbb{E}}_{X(\ell/2)}\left[z_{\ell/2} S_{\ell/2} z_{\ell/2}\right].$$

Now splitting $z_{\ell/2}$ itself into parallel and perpendicular components, we can bound this as

$$\begin{aligned}
\mathop{\mathbb{E}}_{X(\ell/2)}\left[z_{\ell/2} S_{\ell/2} z_{\ell/2}\right] &= \mathbb{E}[z_{\ell/2} S_{\ell/2}(z_{\ell/2}^0 + z_{\ell/2}^\perp)] \\
&= \mathbb{E}[z_{\ell/2}]^2 + \langle z_{\ell/2}^\perp, S_{\ell/2} z_{\ell/2}^\perp \rangle \\
&\leq \mathbb{E}[z_{\ell/2}]^2 + \gamma_{\log(\ell)-1} \|z_{\ell/2}^\perp\|_2^2 \\
&= (1 - \gamma_{\log(\ell)-1})\,\mathbb{E}[z_{\ell/2}]^2 + \gamma_{\log(\ell)-1} \|z_{\ell/2}\|_2^2. \qquad (5.7)
\end{aligned}$$

The first of these two terms can clearly be bounded by induction. The difficulty lies in the error term, as the 2-norm of the partial MGF may be much larger than its expectation. To handle this, we argue via reduction to the complete complex the second moment of $z_{\ell/2}$ is actually quite close to the original expectation of z_ℓ.

Claim 5.5.5. For any even ℓ and $r \in [0, 1]$:

$$\|z_{\ell/2}\|_2^2 \leq e^{\frac{5}{4} r^2 \ell} \mathbb{E}[z_\ell]$$

Let's first complete the proof given this fact. Plugging the claim back into Equa-

tion (5.7) and applying induction we get the following bound:

$$\begin{aligned}\mathbb{E}[z_k] &\leq \frac{1-\gamma_{j-1}}{1-e^{\frac{5}{4}r^2\ell}\gamma_{j-1}}\mathbb{E}[z_{k/2}]^2 \\ &\leq \frac{1-\gamma_{j-1}}{1-e^{\frac{5}{4}r^2k}\gamma_{j-1}}\left(\prod_{i=0}^{j-2}\left(\frac{1-\gamma_{j-2-i}}{1-e^{2^{-i}\frac{5}{4}r^2\frac{k}{2}}\gamma_{j-2-i}}\right)^{2^{i+1}}\right)\mathbb{E}[z_1]^k \\ &= \left(\prod_{i=0}^{j-1}\left(\frac{1-\gamma_{j-1-i}}{1-e^{2^{-i}\frac{5}{4}r^2(k+1)}\gamma_{j-1-i}}\right)^{2^i}\right)\mathbb{E}[e^{rf}]^k\end{aligned}$$

as desired. □

It is left to prove the claim relating $\|z_{\ell/2}\|_2^2$ to $\mathbb{E}[z_\ell]$.

Proof of Claim 5.5.5. Recall our goal is to show

$$\mathbb{E}[z_{\ell/2}^2] = \mathop{\mathbb{E}}_{t\in X(\ell/2)}\left[\prod_{v\in t}e^{2rf(v)}\right] \leq e^{\frac{5}{4}r^2\ell}\mathbb{E}[z_\ell]$$

The key is to observe that we can draw $t \in X(\ell/2)$ by first drawing $s \in X(\ell)$, and then drawing $t \subset s$ uniformly at random:

$$\mathop{\mathbb{E}}_{t\in X(\ell/2)}\left[\prod_{v\in t}e^{2rf(v)}\right] = \mathop{\mathbb{E}}_{s\in X(\ell)}\left[\mathop{\mathbb{E}}_{t\subset s}\left[\prod_{v\in t}e^{2rf(v)}\right]\right]$$

The inner expectation is now within the complete complex which is negatively correlated, so may therefore 'pull out' the product (e.g. by [197, Theorem 4]) and write:

$$\mathop{\mathbb{E}}_{t\subset s}\left[\prod_{v\in t}e^{2rf(v)}\right] \leq \mathop{\mathbb{E}}_{v\in s}[e^{2rf(v)}]^{\ell/2}.$$

Let $\mu_s = \mathbb{E}_{v\in s}[f(v)]$ and observe $\ell\mu_s = \sum_{v\in s} f(v)$. Combining the above, we have

$$\mathbb{E}[z_{\ell/2}^2] \leq \mathop{\mathbb{E}}_{s\in X(\ell)}\left[\mathop{\mathbb{E}}_{v\in s}[e^{2rf(v)}]^{\ell/2}\right]$$

$$
\begin{aligned}
&\leq \mathop{\mathbb{E}}_{s\in X(\ell)}\left[\mathop{\mathbb{E}}_{v\in s}[1+2rf(v)+\frac{5}{2}r^2f(v)]^{\ell/2}\right]\\
&= \mathop{\mathbb{E}}_{s\in X(\ell)}\left[(1+2r\mu_s+\frac{5}{2}r^2\mu_s)^{\ell/2}\right]\\
&\leq \mathop{\mathbb{E}}_{s\in X(\ell)}\left[e^{(r+\frac{5}{4}r^2)\ell\mu_s}\right]\\
&= \mathop{\mathbb{E}}_{s\in X(\ell)}\left[\prod_{v\in s}e^{(r+\frac{5}{4}r^2)f(v)}\right]\\
&\leq e^{\frac{5}{4}r^2\ell}\mathbb{E}\left[z_\ell\right]
\end{aligned}
$$

where in the second inequality we've used that $e^x < 1+x+\frac{5}{4}x^2$ for $x\in[0,2]$ and that $f^2(v)\leq f(v)$. □

The proof of Theorem 5.5.2 is now essentially immediate from Chernoff's method.

Proof of Theorem 5.5.2. We prove the upper tail. The lower tail follows from applying the upper tail to $1-f$. Let k' be the largest power of 2 such that $k'\leq k$. Fix $r=\frac{1}{2}\varepsilon k'^{\frac{\alpha-1}{2}}$, and recall by Theorem 5.4.20 that $\gamma_i\leq\frac{2^i}{k'}2^{-(k')^\alpha}$. By Proposition 5.5.4, we therefore have:

$$
\begin{aligned}
&\mathop{\mathbb{P}}_{\{v_1,\ldots,v_{k'}\}\in X(k')}\left[\sum_{i=1}^{k'}f(v_i)\geq\mu k'+\varepsilon k'\right]\\
&= \mathop{\mathbb{P}}_{\{v_1,\ldots,v_{k'}\}\in X(k')}\left[\exp\left(r\sum_{i=1}^{k}f(v_i)\right)\geq\exp\left(rk'\mu+rk'\varepsilon\right)\right]\\
&\overset{Markov}{\leq} \frac{\mathbb{E}\left[\prod_{i=0}^{k'}e^{rf(v_i)}\right]}{\exp(rk'\mu+rk'\varepsilon)}\\
&\leq \left(\prod_{i=0}^{j-1}\left(\frac{1}{1-2^{-i-1}e^{-\frac{1}{3}(k')^\alpha}}\right)^{2^i}\right)\exp\left(-\frac{\varepsilon^2}{3}(k')^{\frac{1+\alpha}{2}}\right)
\end{aligned}
$$

where in the final inequality we have applied Proposition 5.5.4 and bounded $\mathbb{E}[e^{rf}]^{k'}$ by standard manipulations. To bound the leading coefficient first note that for any $\alpha\geq 0$ we have

$$
\prod_{i=0}^{j-1}\left(\frac{1}{1-2^{-i-1}e^{-\frac{1}{3}(k')^\alpha}}\right)^{2^i}=\prod_{i=0}^{j-1}\left(\frac{1}{1-2^{-i-1}}\right)^{2^i}\leq k'
$$

since each individual term in the product is at most $\frac{1}{2}$ and $j = \log_2(k')$. For $\alpha > 0$ we may also write

$$\begin{aligned}\prod_{i=0}^{j-1}\left(\frac{1}{1-2^{-i-1}e^{-\frac{1}{3}(k')^\alpha}}\right)^{2^i} &\leq \prod_{i=0}^{j-1}\left(\frac{1}{1-\frac{e^{-\frac{1}{3}(k')^\alpha}}{2}}\right)^{2^i}\\ &\leq \left(1+e^{-\frac{1}{3}(k'+1)^\alpha}\right)^{k'}\\ &\leq e^{\frac{k'}{e^{\frac{1}{3}(k')^\alpha}}}\\ &\leq e^{\frac{(\frac{3}{\alpha})^{\frac{1}{\alpha}}}{e^\alpha}}\end{aligned}$$

where the last inequality follows from observing $e^{\frac{k'}{e^{\frac{1}{3}(k')^\alpha}}}$ is maximized at $k' = (\frac{3}{\alpha})^{-\alpha}$.

Finally we apply Lemma 5.4.15 to lift this guarantee to $X(k)$:

$$\begin{aligned}\mathop{\mathbb{P}}_{X(k)}[U_{1,k}f - \mu > \varepsilon] &\leq \min\{k', c_\alpha\}\exp\left(-\frac{\varepsilon^2}{12}(k'+1)^{\frac{1+\alpha}{2}}\right)\left(1-\pi_{low}^{k,k',U_0^{k'}f}(\varepsilon/2)\right)^{-1}\\ &\leq \min\{k, c_\alpha\}\exp\left(-\frac{\varepsilon^2}{12}(k+1)^{\frac{1+\alpha}{2}}\right)\left(1-e^{-\frac{\varepsilon^2(k+1)}{6}}\right)^{-1}\\ &\leq 2\min\{k, c_\alpha\}\exp\left(-\frac{\varepsilon^2}{12}k^{\frac{1+\alpha}{2}}\right)\end{aligned}$$

where in the final inequality we have assumed $\varepsilon^2 \geq \frac{6}{k}$ (else the stated bound is trivial). □

5.5.2 From Chernoff to Inclusion Sampling

We now show how to bootstrap Chernoff-type concentration to optimal sampling for the general inclusion graph $(X(k), X(i))$. We focus on the case of two-sided HDX, and give the substantially more involved argument for the full theorem in Section 5.7. We restate the theorem in the two-sided case here for convenience.[9]

Theorem 5.5.6 (Sampling on two-sided HDX)**.** *For any $c > 0$ there exists $c' > 0$ such that for any $k \in \mathbb{N}$, $i < k$, and k-uniform 2^{-ck}-two-sided local spectral expander X, $(X(k), X(i))$*

[9]We prove only the $\alpha = 1$ setting, the general version follows the same argument.

is an (ε, β) *sampler for*

$$\beta = \frac{c'}{\varepsilon} \exp\left(-\Omega_c\left(\varepsilon^2 \frac{k}{i}\right)\right).$$

Moreover, if X *is a* k*-skeleton of a* d*-uniform complex* Y*, then for every* $k < k' \leq d$*, the graph* $(Y(k'), Y(i))$ *is also an* $(\varepsilon, O(\beta))$*-function sampler.*

The proof of Theorem 5.5.6 is based on a refinement of an idea of [215], who observed on the complete complex one can prove an analogous bound by partitioning k-sets into i disjoint $\frac{k}{i}$-sets, and treating these as independent variables. In the world of general simplicial complexes, this type of subdivision is called the *faces complex* of X [110].

Definition 5.5.7 (Faces Complex [110])**.** Let X be a d-uniform simplicial complex. For any $i \leq d$, the faces complex $F_X^{(i)}$ is the $\lfloor \frac{d}{i} \rfloor$-uniform complex with vertices $X(i)$ and top level faces:

$$F_X^{(i)}\left(\left\lfloor \frac{d}{i} \right\rfloor\right) = \left\{ \{s_1, \ldots, s_{\lfloor \frac{d}{i} \rfloor}\} \;\middle|\; \bigcup_{j=1}^{\lfloor \frac{d}{i} \rfloor} s_j \in X,\ \bigcup_{j \neq \ell} s_j \cap s_\ell = \emptyset \right\}$$

The weight of a face is proportional to the weight of its union in X. We drop X from the notation when clear from context.

In our setting, subdivided faces have extreme dependencies so [215]'s analysis for the complete complex fails. Nevertheless, it is in fact possible to generalize this approach to any simplicial complex through somewhat more careful analysis.

Proposition 5.5.8. *Let* X *be a* d*-uniform complex and* $i \leq k \leq d$ *be such that* $\left(F^{(i)}\left(\lfloor \frac{k}{i} \rfloor\right), F^{(i)}(1)\right)$ *is an* (ε, β)*-function additive sampler. Then* $(X(k), X(i))$ *is a* $(2\varepsilon, \frac{2}{\varepsilon}\beta)$*-function additive sampler.*

The proof of Proposition 5.5.8 itself centers around a lemma for general bipartite samplers stating that sampling over correlated sub-divisions is sufficient to imply sampling

for the general graph as long as the components are appropriately marginally distributed.

Lemma 5.5.9. *Given a graph $G = (L, R, E)$, parameters $\varepsilon, \beta > 0$ and $k \in \mathbb{N}$, and $f : L \to [0, 1]$, suppose there is a distribution $(v_1, v_2, \ldots, v_k, u) \sim D$ over $L^k \times R$ satisfying*

1. *For every $i \in [k]$, the marginal (v_i, u) is distributed according to the edges of G.*
2. $\mathbb{P}_{(v_1,v_2,\ldots,v_k,u)\sim D}\left[\frac{1}{k}\sum_{i=1}^{k} f(v_i) > \mathbb{E}[f] + \varepsilon\right] \leq \beta.$

Then

$$\mathbb{P}_{u\sim R}\left[\mathop{\mathbb{E}}_{v\in L, v\sim u}[f(v)] > \mathbb{E}[f] + 2\varepsilon\right] \leq \frac{\beta}{\varepsilon}. \tag{5.8}$$

The same statement applies replacing $>$ with $<$ and ε with $-\varepsilon$ inside the probabilities.

We defer the proof to the end of the subsection and first prove Proposition 5.5.8.

Proof of Proposition 5.5.8. Let $m = \lfloor \frac{k}{i} \rfloor$ and fix a function $f : X(i) \to [0, 1]$ with expectation $\mu = \mathbb{E}_{s\in X(i)}[f(v)]$. Consider the distribution D which samples $(s_1, s_2, \ldots, s_m, t) \subset X(i)^m \times X(k)$ by drawing $\{s_i\}_{i=1}^m$ from the faces complex, and draws $t \supset \bigcup s_i$ from $X(k)$. We claim D satisfies the requirements of Lemma 5.5.9, namely:

1. $\forall i \in [m]$, (s_i, t) is marginally distributed according to the inclusion graph edges $(X(k), X(i))$
2. $\mathbb{P}_{\{s_1,s_2,\ldots,s_m,t\}\sim D}\left[\left|\frac{1}{k}\sum_{i=1}^{k} f(s_i) - \mu\right| > \varepsilon\right] \leq \beta.$

The first fact follows from the definition of the faces complex, namely that marginally each s_i is distributed as $X(i)$, and t is drawn from $X(k)$ conditional on containing s_i. The second fact is immediate from our assumption that the faces complex is an (ε, β)-sampler.

□

We are now ready to prove the two-sided case of Theorem 5.5.1.

Proof of Theorem 5.5.6 (Theorem 5.5.1 two sided case). By Theorem 5.5.2 if the faces complex of X is a 2^{-ck}-two-sided HDX, then it is an (ε, β)-sampler for $\beta \leq c_1 e^{\Omega_c\left(\varepsilon^2 \lfloor \frac{k}{i} \rfloor\right)}$ and $c_1 > 0$ some constant dependent only on c. Then by Proposition 5.5.8, X would be a $(2\varepsilon, \frac{\beta}{\varepsilon})$-sampler. The proof is completed by observing $2^{-\Omega(k)}$-local-spectral expansion of the faces complex is immediate from Theorem 5.4.20, since the graph underlying any link of the faces complex is a swap walk within (a link of) X. The 'moreover' statement follows directly from Lemma 5.4.15. □

It remains to prove Lemma 5.5.9.

Proof of Lemma 5.5.9. Let $Z : R \to [0, 1]$ record the probability $u_0 \in R$ ε-mis-samples f under D:

$$Z(u_0) \coloneqq \mathop{\mathbb{P}}_{(v_1,v_2,\ldots,v_k,u)\sim D}\left[\frac{1}{k}\sum_{i=1}^{k} f(v_i) > \mu + \varepsilon \;\middle|\; u = u_0\right].$$

By assumption, the expected mis-sampling probability is small

$$\mathop{\mathbb{E}}_{u_0}\left[Z(u_0)\right] = \mathop{\mathbb{P}}_{(v_1,v_2,\ldots,v_k,u)\sim D}\left[\frac{1}{k}\sum_{i=1}^{k} f(v_i) > \mu + \varepsilon\right] \leq \beta,$$

so Markov's inequality implies the following tail bound on $Z(u_0)$:

$$\mathop{\mathbb{P}}_{u_0\in R}\left[Z(u_0) > \varepsilon\right] \leq \frac{\beta}{\varepsilon}.$$

We argue if u_0 is such that $\mathbb{E}_{v\in L, v\sim u_0}\left[f(v_i)\right] > \mu + 2\varepsilon$ then $Z(u_0) > \varepsilon$, thus concluding that (5.8) holds.

The key is the simple observation that the (conditional) expectation A under D and G are equivalent:

$$\begin{aligned}\mathop{\mathbb{E}}_{(v_1,v_2,\ldots,v_k,u)\sim D}\left[f(v_i) \mid \; u = u_0\right] &= \frac{1}{k}\sum_{i=1}^{k}\mathop{\mathbb{E}}_{(v_i,u)}\left[f(v_i)|u = u_0\right] \\ &= \mathop{\mathbb{E}}_{v\in L}\left[f(v) \mid \; v \sim u_0\right] \qquad (5.9)\end{aligned}$$

where the first equality is by linearity of expectation, and the second is by our assumption that (v_i, u) is distributed as G. The result now follows roughly from arguing the random variable $\frac{1}{k}\sum_{i=1}^{k} f(v_i)$ restricted to u_0, is non-trivially concentrated around its expectation $\mathbb{E}_{v\in L}\left[f(v) \mid \ v \sim u_0\right]$, which is at least 2ε-far from μ.

In particular, fix u_0 such that $\mathbb{E}_{(v_1,v_2,\ldots,v_k,u)\sim D}\left[f(v_i) \mid \ u = u_0\right] = \mu_{u_0} > \mu + 2\varepsilon$. Then, we bound $Z(u_0)$ by its upper tail:

$$Z(u_0) \geq \mathop{\mathbb{P}}_{(v_1,v_2,\ldots,v_k,u)\sim D}\left[\frac{1}{k}\sum_{i=1}^{k} f(v_i) > \mu_{u_0} - \varepsilon \;\middle|\; u = u_0\right].$$

Denote the right-hand side by p, and as $f(v) \in [0,1]$ we have

$$\mu_{u_0} \leq p + (1-p)(\mu_{u_0} - \varepsilon).$$

Re-arranging gives

$$Z(u_0) \geq p \geq \frac{\varepsilon}{1 - \mu_{u_0} - \varepsilon} \geq \varepsilon.$$

The last inequality follows from the the denominator being in $(0,1)$ ($0 < \mu_{u_0} - \varepsilon < 1$ because $\mu_{u_0} - \varepsilon \geq \mu + \varepsilon > 0$ and $\mu_{u_0} - \varepsilon < \mu_{u_0} \leq 1$).

Finally, we note the usual trick (changing f in the proof to $1 - f$) gives the bounds for

$$\mathop{\mathbb{P}}_{u\in R}\left[\mathop{\mathbb{E}}_{v\sim u, v\in L}\left[f(v)\right] < \mu - 2\varepsilon\right].$$

□

Multiplicative Sampling

In the following sections, it will sometimes be more convenient to work with *multiplicative* sampling instead of the additive bounds of Theorem 5.5.1, and with the 'flipped' inclusion graph $(X(i), X(k))$. By the generic translations in Claim 5.4.9 and Claim 5.4.8 we get the following corollary for this setting.

Corollary 5.5.10. *Let X be a k-maximal c-*nice *complex. Then for every $\delta, \beta \in (0,1)$, $(X(i), X(k))$ is a (α, β, δ)-sampler for*

$$\alpha \leq \frac{c'}{\beta\delta} e^{-\Omega_c(\beta^2\delta^2 \frac{k}{i})}.$$

We remark that we have not attempted to optimize dependence on β and δ above. In the following sections we will only be interested in the setting where both are constant.

5.6 Reverse Hypercontractivity

In this section we recursively leverage optimal inclusion sampling to prove reverse hypercontractivity for high dimensional expanders.

Theorem 5.6.1 (Reverse Hypercontractivity)**.** *Fix $c > 0$, $\rho \in (0,1)$, and let X be a k-uniform c-locally nice complex for k sufficiently large. Then there exist constants C, q (dependent only on c, ρ) such that*

$$\langle f, T_\rho g \rangle \geq C \|f\|_q \|g\|_q.$$

At its core, the proof of Theorem 5.6.1 really only relies on X and its links satisfying optimal i vs k sampling. With this in mind, we will instead work with the following notion of a *Link up-sampler* that captures this core property:

Definition 5.6.2 (Link up-sampler (LUS))**.** Let $\tau \in (0,1)$ and let X be a k-uniform simplicial complex. We say that X is a τ-LUS if for every $i \leq k-2$, every $j \leq k-i$, and every $s \in X(i)$, it holds that $(X_s(j), X_s(k-i))$ is a $(\frac{1}{\tau}\exp(-\tau\frac{k-i}{j}), 0.1, 0.2)$-multiplicative sampler.

We note the constants $0.1, 0.2$ are essentially arbitrary and fixed for convenience. The notion can also be further relaxed in a number of ways such as asking for sampling

for j only up to a linear number of levels or only within links up to a certain level and the main results of this section would still hold.

We prove link-up samplers are reverse hypercontractive. Theorem 5.6.1 is an immediate corollary.

Theorem 5.6.3. *For every $\rho \in (0,1)$ and $\tau > 0$, there is some $\ell > 1$ and $C > 0$ such that the following holds for all sufficiently large $k \in \mathbb{N}$. Let X be k-uniform and τ-LUS. Then T_ρ is $(\frac{1}{\ell}, \frac{\ell}{1-\ell}, C)$-reverse hypercontractive.*

We split the proof of Theorem 5.6.3 below into two parts, roughly corresponding to reverse hypercontractivity for boolean functions, and a reduction from the general to boolean case.

5.6.1 The Boolean Case

The main workhorse behind Theorem 5.6.3 is to show that any linear-step *down-up walk* on an LUS is reverse hypercontractive for sets of size up to $\exp(-\Omega(k))$.

Theorem 5.6.4. *Let $\tau > 0$ and $\gamma \in (0,1)$. There exist constants $c, q > 0$ such that the following holds for any sufficiently large k. Let X be a k-uniform τ-LUS. Then for any sets $A, B \subseteq X(k)$ of relative size at least $\exp(-ck)$ and $\gamma' \leq \gamma$ it holds that*

$$\mathop{\mathbb{P}}_{t,t' \sim UD_{k,\lfloor \gamma' k \rfloor}} [t \in A, t' \in B] \geq \mathbb{P}[A]^q \mathbb{P}[B]^q .$$

It will be convenient for us to assume at least one of the sets A and B is not too large. To this end, we first handle the case that both are large separately.

Claim 5.6.5. For every $p > 0.5$ there exists $q \geq 1$ so that the following holds. Let X be a simplicial complex and $A, B \subseteq X(k)$ subsets such that $\mathbb{P}[A], \mathbb{P}[B] \geq p$. Then for any $\gamma' \leq \gamma$:

$$\mathop{\mathbb{P}}_{t,t' \sim UD_{k,\lfloor \gamma' k \rfloor}} [t \in A, t' \in B] \geq \mathbb{P}[A]^q \mathbb{P}[B]^q .$$

The proof is largely un-enlightening computation, so we postpone it to the end of the subsection. Henceforth, assume without loss of generality that $\mathbb{P}[A] \le \mathbb{P}[B]$ and has $\mathbb{P}[A] \le 0.8$.

With this out of the way, we overview the main proof strategy. Assume for the moment that $\gamma' k$ is integral. As discussed in the introduction, Theorem 5.6.4 relies on the following elementary observation: since the down-up walk samples an edge by first sampling $s \in X(\gamma k)$ and then *independently* sampling $t, t' \in X(k)$ containing s, we can express the γ'-correlated mass of A and B as:

$$\mathop{\mathbb{P}}_{t,t' \sim UD_{k,\gamma' k}}[t \in A, t' \in B] = \mathop{\mathbb{E}}_{s \in X(\gamma' k)} \left[\mathop{\mathbb{P}}_{t}[A \mid \ t \supseteq s] \cdot \mathop{\mathbb{P}}_{t}[B \mid \ t \supseteq s] \right].$$

This motivates us to consider the set of $s \in X(\gamma k)$ that *simultaneously* see a large enough fraction of A, B. Toward this end, for any integer $\ell_0 \le k$ and real number $\delta > 0$, let

$$G^X(\ell_0, \delta) = \left\{ s \in X(\ell_0) \,\middle|\, \mathop{\mathbb{P}}_{t}[A \mid \ t \supseteq s] \ge \delta \, \mathbb{P}[A] \text{ and } \mathop{\mathbb{P}}_{t}[B \mid \ t \supseteq s] \ge \delta \, \mathbb{P}[B] \right\}$$

denote the set of elements in $X(\ell_0)$ that see a noticeable fraction of both A and B.

Recalling $\mathbb{P}[B] \ge \mathbb{P}[A]$, to prove Theorem 5.6.4 it suffices to show for any $\gamma' \le \gamma$

$$\mathbb{P}\left[G^X(\gamma' k, \mathbb{P}[A]^{O(1)}) \right] \ge \mathbb{P}[A]^{O(1)}, \tag{5.10}$$

since (writing $G \coloneqq G^X(\gamma' k, \mathbb{P}[A]^{O(1)})$) we then have

$$\begin{aligned} \mathop{\mathbb{P}}_{t,t' \sim UD_{k,\gamma' k}}[t \in A, t' \in B] &\ge \mathop{\mathbb{P}}_{s \in X(\gamma' k)}[G] \cdot \mathop{\mathbb{P}}_{s \in X(\gamma' k)} \left[\mathop{\mathbb{P}}_{t}[t \in A | s \in G] \mathop{\mathbb{P}}_{t}[t \in B | s \in G] \right] \\ &\ge \mathbb{P}[A]^{O(1)} \mathbb{P}[B] \end{aligned} \tag{5.11}$$

As such, our main goal is to prove (5.10). A key observation toward this end is that

G satisfies certain useful 'composition' behavior: we can bound $G(\gamma' k, \cdot)$ by iteratively bounding $G(i, \cdot)$ for smaller i within repeatedly deeper links of the complex. In particular, for sets A and B and any $r \in X$, define the restrictions of A and B to the link of r as

$$A_r \coloneqq \{t \in X_r : r \cup t \in A\} \qquad \text{and} \qquad B_r \coloneqq \{t \in X_r : r \cup t \in B\}$$

and define the *local* subset $G^{X_r}(\ell_0, \delta)$ as the set of $s \in X_r(\ell_0)$ that see a sufficiently fraction of A_r and B_r:

$$G^{X_r}(\ell_0, \delta) = \left\{ s \in X_r(\ell_0) \middle| \begin{array}{r} \mathop{\mathbb{P}}_{t \in X_r(k-r)} [A_r \mid \ t \supseteq s] \geq \delta \, \mathbb{P}[A_r] \\ \text{and } \mathop{\mathbb{P}}_{t \in X_r(k-r)} [B_r \mid \ t \supseteq s] \geq \delta \, \mathbb{P}[B_r] \end{array} \right\}.$$

G and G^{X_r} compose in the following fashion:

Observation 5.6.6 (Composition of G^X). Let $A, B \subset X(k)$, $\ell_1, \ell_2 \in \mathbb{N}$, and let $\delta_1, \delta_2, \eta_1, \eta_2 > 0$. If

1. $\mathbb{P}\left[G^X(\ell_1, \delta_1)\right] \geq \eta_1$.
2. $\forall r \in G^X(\ell_1, \delta_1)$: $\mathbb{P}_{X_r(\ell_2)}\left[G^{X_r}(\ell_2, \delta_2)\right] \geq \eta_2$

Then

$$\mathbb{P}\left[G^X(\ell_1 + \ell_2, \delta_1 \delta_2)\right] \geq \eta_1 \eta_2.$$

Proof. Sample $s \in X(\ell_1 + \ell_2)$ by sampling $r_1 \in X(\ell_1)$ and then $r_2 \in X_r(\ell_0)$ setting $s = r_1 \cup r_2$. Fix $r_1 \in G^X(\ell_1, \delta_1)$. Observe that if $r_2 \in G^{X_{r_1}}(\ell_2, \delta_2)$ then

$$\begin{aligned} \mathop{\mathbb{P}}_{t}[A \mid \ r_1 \cup r_2 \subseteq t] &\geq \delta_2 \mathop{\mathbb{P}}_{t}[A \mid \ r_1 \subseteq t] \\ &\geq \delta_1 \delta_2 \, \mathbb{P}[A], \end{aligned}$$

and similarly for B. Therefore in this case $s \in G^X(\ell_1 + \ell_2, \delta_1\delta_2)$. Hence

$$\begin{aligned}\mathop{\mathbb{P}}_{s}\left[G^X(\ell_1+\ell_2,\delta_1\delta_2)\right] &\geq \mathop{\mathbb{P}}_{r_1}\left[G^X(\ell_1,\delta_1)\right]\cdot \mathop{\mathbb{P}}_{s=r_1\cup r_2}\left[r_2 \in G^{X_{r_1}}(\ell_2,\delta_2)\;\middle|\; r_1\in G^X(\ell_1,\delta_1)\right]\\ &\geq \eta_1\eta_2.\end{aligned}$$

□

The proof strategy (modulo details), is then simple. To bound $G(\gamma' k, \mathbb{P}[A]^{q_0})$, we divide $t \in X(k)$ into pieces as $r_1 \cup r_2 \cup r_3 \ldots$, where each r_i is viewed as being drawn from the link of $r_1 \cup \ldots \cup r_{i-1}$. The local G-set for each r_i can be lower bounded by the sampling properties of $X_{\cup_{j<i} r_j}$, and by carefully choosing the size of the r_i, we can ensure that composition gives the desired bound on G.

In practice, this approach hits some complications due to the fact that the localized restrictions of A and B shrink throughout the process. To handle this, we'll instead use a 'two-phase' version of the above. First, we show that there exists some small absolute constant γ_0, potentially much smaller than the desired γ, for which the above approach does work. We call complexes satisfying such a bound lovely:

Definition 5.6.7 (Lovely complex)**.** We say that X is $(c, 1-\gamma_0, k, q)$-*lovely*, if for all $A, B \subset X(k)$ with $\mathbb{E}[B] \geq \mathbb{E}[A] \geq \exp(-ck)$

$$\mathbb{P}\left[G^X(\gamma_0 k, \mathbb{P}\left[A\right]^q)\right] \geq \mathbb{P}\left[A\right]^q.$$

We first argue any LUS complex is γ'-lovely for some γ' depending only on τ and γ.

Lemma 5.6.8. *For every $\tau > 0$ and $\gamma < 1$ there exist c, q and $\gamma_0 > 0$ so that following holds for large enough k. Let X be τ-LUS, then for every $t \in X^{\leq \gamma k}$, X_t is $(c, 1-\gamma', k-|t|, q)$-lovely for any $\gamma' \leq \gamma_0$.*

Second, we give a similar argument showing that loveliness itself can be bootstrapped to higher levels.

Claim 5.6.9. Let X be a k-uniform simplicial complex. Let $q', q'' \geq 1$, and $c', c'', \gamma_1, \gamma_2 \geq 0$. Assume that:

1. X is $(c', 1-\gamma_1, k, q')$-lovely.
2. For every $r \in X(\gamma_1 k)$, X_r is $(c'', 1-\gamma_2, (1-\gamma_1)k, q'')$-lovely.

Then X is $(c, (1-\gamma_1)(1-\gamma_2), k, q)$-lovely for $q = q''q' + q'' + q'$ and $c = \min\{\frac{c''}{(q'+1)(1-\gamma_1)}, c'\}$.

Combined with some definition chasing, Lemma 5.6.8 and Claim 5.6.9 immediately imply Theorem 5.6.4.

Proof of Theorem 5.6.4. Recall that by (5.11) it suffices to show X is $(c, 1-\gamma, k, q)$-lovely for some constants c, q depending only on γ and τ. We first prove a slightly different statement by induction:

$$\forall \ell \in \left[\left\lceil \frac{\log(1-\gamma)}{\log(1-\gamma_0)} \right\rceil\right] : \quad X \text{ is } (c_\ell, (1-\gamma_0)^\ell, k, q_\ell)\text{-lovely.}$$

for constants c_ℓ, q_ℓ dependent only on τ, γ. Note this implicitly assumes $(1-\gamma_0)^\ell$ is integer. This is unecessary and we discuss removing this assumption at the end of the subsection. Finally, observe that for $\ell = \left\lceil \frac{\log(1-\gamma)}{\log(1-\gamma_0)} \right\rceil$ we have $(1-\gamma_0)^\ell \leq 1-\gamma$. We will argue a slight modification gives exactly $1-\gamma''$ for any $\gamma'' \leq \gamma$ after showing the above.

The base case $\ell = 1$ is immediate from Lemma 5.6.8. Assume the statement is true for some $\ell < \left\lceil \frac{\log(1-\gamma)}{\log(1-\gamma_0)} \right\rceil$, then X is $(c(\ell), (1-\gamma_0)^\ell, k, q(\ell))$-lovely. Moreover, for $j = 1-(1-\gamma_0)^\ell \leq \gamma k$, Lemma 5.6.8 implies every j-link X_t is $(c'', 1-\gamma', k-j, q'')$-lovely for any $\gamma' \leq \gamma_0$. Thus by Claim 5.6.9, we have that X is $(c_{\ell+1}, 1-(1-\gamma_0)^{\ell+1}, k-|t|, q_{\ell+1})$-lovely where $c_{\ell+1}$ and $q_{\ell+1}$ are functions of γ_0, $c(\ell)$, $q(\ell)$, c'', and q'' as described in Claim 5.6.9. Since γ_0, c'', and q'' are themselves functions only of τ and γ, this gives the desired bound.

In order to prove X is $(c, 1-\gamma, k, q)$-lovely, simply observe that in the final step of this induction we may choose $\gamma_0' \leq \gamma_0$ such that $(1-\gamma_0)^\ell(1-\gamma_0') = 1-\gamma$. Exactly the same argument as above then gives X is at least $(c_\ell, 1-\gamma, k, q_\ell)$-lovely for $\ell = \left\lceil \frac{\log(1-\gamma)}{\log(1-\gamma_0)} \right\rceil$. The same strategy can be used to achieve the same (or better) constants for any $\gamma' < \gamma$ as well. □

It remains to prove Lemma 5.6.8 and Claim 5.6.9. While Lemma 5.6.8 is the central component, it is instructive to first prove the simpler Claim 5.6.9 which is in some sense a 'one-step' variant of the former.

Proof of Claim 5.6.9. Let A, B be sets of size at least $\exp(-ck)$. Our goal can be rephrased as showing

$$G^X((\gamma_1 + \gamma_2(1-\gamma_1))k, \mathbb{P}[A]^{q''q'+q''+q'}) \geq \mathbb{P}[A]^{q''q'+q''+q'}$$

since $1-(\gamma_1+\gamma_2(1-\gamma_1)) = (1-\gamma_1)(1-\gamma_2)$. We prove this via Observation 5.6.6. In particular since $c' < c$, $(c', 1-\gamma_1, k, q')$-loveliness of X implies

$$G^X(\gamma_1 k, \mathbb{P}[A]^{q'}) \geq \mathbb{P}[A]^{q'}.$$

Moreover, for every $r \in G^X(\gamma_1 k, \mathbb{P}[A]^{q'})$ we have that $\mathbb{P}_{X_r}[A], \mathbb{P}_{X_r}[B] \geq \mathbb{P}[A]^{q'+1} \geq \exp(-c''(1-\gamma_1)k)$. Therefore by $(c'', 1-\gamma_2, (1-\gamma_1)k, q'')$-loveliness of X_r it holds that

$$G^{X_{t_1}}(\gamma_2(1-\gamma_1)k, \mathop{\mathbb{P}}_{X_r}[A]^{q''}) \geq \mathop{\mathbb{P}}_{X_r}[A]^{q''} \geq \mathbb{P}[A]^{(q'+1)q''}.$$

Combining the two bounds and applying Observation 5.6.6 we then have

$$G^X(\gamma_1 k + \gamma_2(1-\gamma_1)k, \mathbb{P}[A]^{q''q'+q''+q'}) \geq \mathbb{P}[A]^{q''q'+q''+q'}$$

as desired. □

Finally, we end with the proof of Lemma 5.6.8. Here, instead of decomposing $s = r \cup r'$, we will need to more carefully decompose $s = r_1 \cup r_2 \cup r_3 \ldots$ and control in every step how much A and B shrink using the link-up-sampling property.

Proof of Lemma 5.6.8. We start with a few simplifying assumptions. First, recall that by Claim 5.6.5 we may assume that $\mathbb{P}[A] \leq 0.8$. Second, note it is sufficient to prove the theorem just for X. The result then follows for all γk-links X_t since the link of a τ-LUS complex is itself τ-LUS, and the ambient dimension of each link is at least $(1-\gamma)k$ which we take to be still sufficiently large. Finally we introduce one notational simplification. Recall in the definition of a τ-LUS that for any i-face s the graph $(X_s(j), X_s(k-i))$ is a $(\frac{1}{\tau}\exp(-\tau\frac{k-i}{j}), 0.1, 0.2)$-multiplicative sampler. In the proof of this lemma we will only consider links of faces s with size $i \leq \gamma_0 k$ for some sufficiently small γ_0, so these are always $(\frac{1}{\tau}\exp(-\tau(1-\gamma_0)\frac{k}{j}), 0.1, 0.2)$-multiplicative samplers. Moreover, we will also always ensure $j \leq \gamma_0 k$, allowing us to subsume the constant $\frac{1}{\tau}$ into the exponent. In particular, as long as $\gamma_0 \leq \frac{\tau}{3\ln\frac{1}{\tau}}$, we may assume all (examined) links in the proof are $(\exp(-\tau'\frac{k}{j}), 0.1, 0.2)$-multiplicative samplers where $\tau' = \frac{\tau}{3}$.

Fix $\gamma_0 = \min\{\frac{\tau}{3\ln\frac{1}{\tau}}, \gamma\}$, let $c = \gamma_0\frac{\tau'}{3}$, and assume $\mathbb{P}[B] \geq \mathbb{P}[A] \geq \exp(-ck)$. Fix $m = \frac{-\log\mathbb{P}[A]}{c}$, and define the 'step size' $\ell = \frac{k}{m}$. Observe that $1 \leq \ell \leq \gamma_0 k$. Assume for the moment that ℓ is also integer (see Remark 5.6.11). We record the following immediate observations:

Observation 5.6.10.

1. $\mathbb{P}[A] = \exp(-cm)$.
2. $0.8^m = \mathbb{P}[A]^{q_0}$ where $q_0 = -c^{-1}\log 0.8$ is a constant.
3. For every $i \leq \gamma_0 m$:

$$0.8^i\,\mathbb{P}[A] \geq \exp(-\tau' m) = \exp(-\tau'\frac{k}{\ell}).$$

The first two items are by definition. The third is satisfied for any γ_0, c such that $c(1+\gamma_0 q_0) \le \tau'$. One can check that taking $\gamma_0 \le \min\{\gamma, \frac{\tau}{3\ln\frac{1}{\tau}}\}$ suffices. Finally, assume for simplicity k is such that $\gamma_0 k \in \mathbb{N}$.

We now move to the main argument. Let $j \in \{1, 2, \dots, j_{fin} = \lceil \gamma_0 m \rceil\}$ and for every $j \le \lfloor \gamma_0 m \rfloor$ define

$$G_j \coloneqq G^X(j\ell, (0.8)^j)$$

as above. If $j_{fin} > \gamma_0 m$, additionally define

$$G_{j_{fin}} \coloneqq G^X(\gamma_0 k, 0.8^{j_{fin}}) \subseteq G^X(\gamma_0 k, \mathbb{P}[A]^{\gamma_0 q_0 + 1}).$$

where in the last step we've used the assumption that $\mathbb{P}[A] \le 0.8$. We will show

$$\mathbb{P}\left[G_{j_{fin}}\right] \ge (0.8)^{j_{fin}} \ge \mathbb{P}[A]^{\gamma_0 q_0 + 1}. \tag{5.12}$$

Setting $q = \gamma_0 q_0 + 1$ proves the lemma. We prove (5.12) by inductively showing that for every $j = 1, 2, \dots, j_{fin}$,

$$\mathbb{P}[G_j] \ge (0.8)^j.$$

We start with the base case $j = 1$, which already contains the main idea. Given $S \subset X(k)$, define

$$T(S) \coloneqq \left\{ r \in X(\ell) : \mathbb{P}[S \mid r] < 0.8\,\mathbb{P}[S] \right\}$$

to be the set of 'terrible' ℓ-sets which see too little of S. Observe that by construction

$$G_1 = X(\ell) \setminus (T(A) \cup T(B)).$$

Since $\ell \le \gamma_0 k$, we have $(X(k), X(\ell))$ is a $(\exp(-\tau'\frac{k}{\ell}), 0.1, 0.2)$-sampler by our earlier discussion. On the other hand, by Observation 5.6.10, we can bound the sizes of A and B

as:

$$\mathbb{P}[B] \geq \mathbb{P}[A] = \exp(-cm) \geq \exp(-\tau' \frac{k}{\ell}).$$

Thus applying sampling we can bound $\mathbb{P}[G_1]$ by

$$\mathbb{P}[G_1] \geq 1 - \mathbb{P}[T(A)] - \mathbb{P}[T(B)] \geq 1 - 2 \cdot 0.1 = 0.8$$

as desired.

The inductive step is similar to the base case, only performed *locally* within the links of X together with Observation 5.6.6. By induction hypothesis $\mathbb{P}[G_i] \geq 0.8^i$. For every $s \in G_i$, by Observation 5.6.10

$$\underset{X_s}{\mathbb{P}}[B], \underset{X_s}{\mathbb{P}}[A] \geq (0.8)^i \mathbb{P}[A] \geq \exp(-\tau' m) \geq \exp(-\tau' \frac{k}{\ell}).$$

Thus by the same argument above for X_s and $A_s = \{t \setminus s \mid s \subseteq t \in A\}, B_s = \{t \setminus s \mid s \subseteq t \in B\}$, we have

$$\mathbb{P}\left[G^{X_s}(\ell, 0.8) \;\middle|\; s \in G_i \right] \geq 1 - \mathbb{P}[T(A)] - \mathbb{P}[T(B)] \geq 1 - 2 \cdot 0.1 = 0.8.$$

The inequality $\mathbb{P}[G_{i+1}] \geq 0.8^{i+1}$ then follows by Observation 5.6.6.

Finally, we note that for the last step, going between $j_{fin} - 1$ to j_{fin} (in the case that $j_{fin} > \gamma m$), we follow the same procedure except that we may need to sample less than ℓ new vertices. This only improves the sampling so the same analysis as above holds. □

Remark 5.6.11. Throughout the proof we have 'cheated' in the typical way assuming a priori that various parameters (ℓ, γk, $\gamma_0 k$, and $(1-\gamma_0)^i(1-\gamma')k$) are integer. Note that we did not assume above $\gamma_0 m$ was integer because of the careful interplay between m and $\mathbb{P}[A]$. In all other cases, however, these assumptions are benign and taking the appropriate

integer floor does not substantially change the proof. The only modification is that every 'step' in the process (e.g., drawing an $\lfloor \ell \rfloor$-size set) potentially moves one fewer level up the complex. In the worst case, this results in taking a constant multiplicative factor more steps than under integer assumptions, resulting in additional constant factors in c and q_0. We omit the details.

Proof of Claim 5.6.5. Denote by $E(X, Y)$ the set of directed edges (t, t') such that $t \in X, t' \in Y$. It is easy to see that $\mathbb{P}[E(A, B)] + \mathbb{P}[E(A, B^c)] = \mathbb{P}[A]$. Similarly we have that $\mathbb{P}[E(A, B^c)] + \mathbb{P}[E(A^c, B^c)] = \mathbb{P}[B^c]$ so $\mathbb{P}[B^c] \geq \mathbb{P}[E(A^c, B^c)]$. combining these together gives that $\mathbb{P}[E(A, B)] \geq \mathbb{P}[A] - \mathbb{P}[B^c]$. Assume that $\mathbb{P}[A] \leq \mathbb{P}[B]$, this implies that $\mathbb{P}[E(A, B)] \geq \mathbb{P}[A] - \mathbb{P}[A^c] = 2\mathbb{P}[A] - 1$. Hence it is sufficient to show that given A such that $\mathbb{P}[A] > p$, there exists a constant $q > 0$ such that $2\mathbb{P}[A] - 1 \geq \mathbb{P}[A]^q$ (which in turn is at least $\mathbb{P}[A]^q \mathbb{P}[B]^q$). Indeed let us consider $f(x) = 2x - 1 - x^q$ and show this function is non-negative in the range $[p, 1]$ for large enough q.

Assume without loss of generality that q is an integer. In this case $f(x) = (1 - x)(\sum_{j=1}^{q-1} x^j - 1)$. This function is greater than 0 in $[p, 1]$ if and only if $\sum_{j=1}^{q-1} x^j \geq 1$ in this range. In the range $[p, 0.9]$ the series $\sum_{j=1}^{q-1} x^j$ converges uniformly to $\frac{x}{1-x}$ which is always greater or equal $\frac{p}{1-p} > 1$. In the range $x \in [0.9, 1]$ we can take any $q \geq 3$ and one can verify that $\sum_{j=1}^{q-1} x^j > 0.9 + 0.9^2 > 1$. □

5.6.2 The General Case

We now prove Theorem 5.6.3 using our restricted reverse hypercontractive inequality for indicators. This is done via the following reduction from reverse hypercontractivity for sets to the general case.

Theorem 5.6.12. *For every $\ell \geq 1$ and $\varepsilon > 0$ there exists $\kappa' \geq \left(\frac{1}{5}\right)^{2\ell(1+\varepsilon)} \left(\frac{1}{18+\frac{12}{\varepsilon}}\right)^{2(\ell-1)}$ such that the following holds. Let V be a finite probability space and D a monotone linear*

operator such that for every $A, B \subseteq V$*,*

$$\langle 1_A, D1_B \rangle \geq \kappa \, \mathbb{P}[A]^{\ell} \, \mathbb{P}[B]^{\ell}. \tag{5.13}$$

Then for every two arbitrary *functions* $f_1, f_2 : V \to \mathbb{R}_{\geq 0}$ *it holds that*

$$\langle f_1, Df_2 \rangle \geq \kappa\kappa' \|f_1\|_{\frac{1}{\ell(1+\varepsilon)}} \|f_2\|_{\frac{1}{\ell(1+\varepsilon)}}. \tag{5.14}$$

That is, D *is* $(\frac{1}{\ell(1+\varepsilon)}, \frac{1}{1-\ell(1+\varepsilon)}, \kappa\kappa')$*-reverse hypercontractive.*

The proof involves careful discretization and thresholding of f_1 and f_2 into level sets in a way that largely maintains the functions' moments. We defer the details to Section 5.16 and prove the main Theorem assuming this result.

Proof of Theorem 5.6.3. By Theorem 5.6.12 it is enough that we show that there exist $\ell > 1, \kappa > 0$ such that for every $A, B \subseteq X(k)$

$$\langle 1_A, T_\rho 1_B \rangle \geq \kappa \, \mathbb{P}[A]^{\ell} \, \mathbb{P}[B]^{\ell}.$$

Since $\gamma > \rho$, observe we can always take k sufficiently large so that $\mathbb{P}_{Y \sim Bin(\rho,k)}[Y \leq \lfloor \gamma k \rfloor] \geq \frac{1}{2}$. Let q, c be the constants promised by Theorem 5.6.4 with respect to the $UD_{\lfloor \gamma k \rfloor, k}$. By monotonicity of q and c, these apply to all walks of intersection at most $\lfloor \gamma k \rfloor$). Thus if both $\mathbb{P}[A], \mathbb{P}[B] \geq \exp(-ck)$ then by Theorem 5.6.4:

$$\mathop{\mathbb{P}}_{t_1, t_2 \sim T_\rho}[t_1 \in A, t_2 \in B] \geq \mathop{\mathbb{P}}_{Y \sim Bin(\rho,k)}[Y \leq \gamma k] \, \mathbb{P}[A]^q \, \mathbb{P}[B]^q \geq \frac{1}{2} \, \mathbb{P}[A]^q \, \mathbb{P}[B]^q.$$

Otherwise, it holds that $\mathbb{P}[A] \, \mathbb{P}[B] \leq \exp(-ck)$. On the other hand, the probability of resampling *all vertices* in the noise operation is $(1-\rho)^k = \exp(-ck)^{\ell''}$ for some constant

$\ell'' > 1$ that depends on c and on ρ. In particular,

$$\mathop{\mathbb{P}}_{t_1,t_2\sim T_\rho}[t_1 \in A, t_2 \in B] \geq (1-\rho)^k \mathbb{P}[A]\mathbb{P}[B] \geq \mathbb{P}[A]^{\ell''+1}\mathbb{P}[B]^{\ell''+1}.$$

Taking $\ell = \max\{q, \ell'' + 1\}$ completes the proof. □

5.6.3 Proof of Theorem 5.6.1

Finally, we briefly give the formal proof for reverse hypercontractivity on high dimensional expanders. This follows from the fact that any c-locally nice complex is $\tau(c)$-LUS for some appropriate setting of constants.

Corollary 5.6.13. *For every $c < 1$ there exists a constant $\tau > 0$ such that any c-locally nice complex is τ-LUS.*

Proof. By definition every i-link X_s is a c-nice complex, so by Corollary 5.5.10 every $(X_s(j), X_s(k-i))$ is a $(\frac{1}{c_1}\exp(-c_2\frac{k-i}{j})), 0.1, 0.2)$-sampler for some constants c_1, c_2 dependent only on c. Taking $\tau = \min\{c_1, c_2\}$ then suffices. □

Proof of Theorem 5.6.1. The proof is immediate from Corollary 5.6.13 and Theorem 5.6.3. □

5.7 Concentration for All Nice HDX

In this section we study concentration of measure for partite complexes and complexes whose down-up walk have an $\Omega(\frac{1}{d})$-spectral gap, completing the proof of Theorem 5.5.1. Along the way, we'll show that complexes near the TD barrier satisfy exponential concentration for Lipschitz functions.

5.7.1 Chernoff-Hoeffding for Partite HDX

We'll start with the setting of partite HDX. As in the two-sided case, we first study the special case of $(X(k), X(1))$. In the partite case different components may behave differently, making the splittability argument for two-sided HDX fail. Instead, we will rely on a *localization* approach, leveraging the local sampling properties between bipartite components of the complex. The resulting quantitative bounds via this technique are largely incomparable with our prior two-sided analysis. While they are weaker for large λ, they approach the error of true independent sampling for small enough λ and remove the restrictive condition of two-sided expansion. The latter in particular actually leads to sparser complexes with subgaussian concentration due to better known degree bounds for partite constructions.

Before stating the result formally, we briefly introduce some relevant notation. Given a complex X, let X^{ind} denote the complete complex whose vertices are $X(1)$, that is:

$$X^{ind}(k) = \{\{v_1, v_2, \ldots, v_k\} : v_i \in X(1)\}$$

where each v_i is drawn independently from $X(1)$ *with repetition.*[10] Similarly for X partite, let $X^{ind,p}$ be the complete k-partite complex whose sides are $X[1], X[2], \ldots, X[k]$:

$$X^{ind,p}(k) = \{\{v_1, v_2, \ldots, v_k\} : v_1 \in X[1], v_2 \in X[2], \ldots, v_k \in X[k]\}$$

where each v_i is drawn independently from the marginal over $X[i]$.

Finally, let $\sigma(X, k, \varepsilon)$ denote the worst case 'k vs 1' ε-sampling error of X.

$$\sigma(X,k,\varepsilon) = \max\left\{ \mathop{\mathbb{P}}_{s\in X(k)}\left[\left| \mathop{\mathbb{P}}_{v\in X(1)}[A \mid v \in s] - \mathop{\mathbb{P}}_{v\in X(1)}[A]\right| > \varepsilon\right] \;\middle|\; A \subseteq X(1)\right\}$$

[10] We remark we have abused notation somewhat as this is not formally a simplicial complex and the faces are *multi*-sets, but this is irrelevant for our purposes.

We prove that sampling in two-sided and one-sided partite HDX match the bounds in their respective independent complex up to $\text{poly}(k)\lambda$-error.

Theorem 5.7.1. *Let $\eta > 0, \varepsilon \in (0, \frac{1}{2})$, and $\lambda \leq \frac{\eta^{2.5}}{16k^4}$. Then*

1. *If X is k-uniform λ-two-sided local-spectral expander:*

$$\sigma(X, k, \varepsilon) \leq \sigma(X^{ind}, k, \varepsilon) + \eta$$

2. *If X is a k-partite λ-one-sided local-spectral expander:*

$$\sigma(X, k, \varepsilon) \leq \sigma(X^{ind,p}, k, \varepsilon) + \eta$$

In particular, in both cases X is then a (ε, β)-sampler for $\beta = \eta + \exp(-\Omega(\varepsilon^2 k))$ by Chernoff-Hoeffding.

In the two-sided case, note that Theorem 5.5.2 is tighter when $\lambda > \exp(-\Omega(k))$, while for small enough λ Theorem 5.5.2 wins out achieving parameters arbitrarily close to independent sampling. We note that here we only give a bound on sampling sets, but this can be converted to concentration for lifts of bounded functions with essentially no loss by Claim 5.4.11.

The core of Theorem 5.7.1 is a lower bound on the k-wise correlation across coordinates of a λ-product.

Proposition 5.7.2. *Let X be a k-uniform λ-product. For every $\eta < 1$ and functions $f_i : X[i] \to [0, 1]$:*

$$\mathop{\mathbb{E}}_{\{v_1, v_2, \ldots, v_k\} \in X(k)} \left[\prod_{i=1}^{k} f_i(v_i) \right] \geq (1 - \eta)^k \prod_{i=1}^{k} \mathbb{E}[f_i]$$

whenever $\lambda \leq \frac{\eta^{1.5} \min_i \{\mathbb{E}[f_i]^2\}}{\sqrt{8k^3}}$.

Proof. The proof relies on the following adaptation of an elementary connection between bipartite expansion and sampling observed in [124]:

Claim 5.7.3. Let $G = (L, R, E)$ be a λ-bipartite expander and let A be its bipartite adjacency operator. Let $f : L \to [0, 1]$ be a function with $\mathbb{E}[f] \geq \mu$ and let $0 < \varepsilon < \mu$. Then

$$\mathbb{P}_{u \in R}[Af(u) < \mu - \varepsilon] < \frac{\lambda^2 \mu}{\varepsilon^2}.$$

We prove this in Section 5.14 for completeness. Using this fact, we prove the following more general claim. Write $\mu_i = \mathbb{E}[f_i]$ and fix $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k > 0$ such that $\mu_i - (i-1)\varepsilon_i > 0$. For $\varepsilon = \min_i\{\varepsilon_i\}$ we show

$$\mathop{\mathbb{E}}_{\{v_1, v_2, \ldots, v_k\} \in X(k)} \left[\prod_{i=1}^{k} f_i(v_i)\right] \geq \prod_{i=1}^{k} \left(\mu_i - (i-1)\varepsilon_i - \lambda^2 \frac{k-i}{\varepsilon^2 \cdot \min\limits_{j>i}(\mu_j - (i-1)\varepsilon_j)}\right). \quad (5.15)$$

Setting $\varepsilon_i = \frac{\mu_i \eta}{2i}$ and applying our assumed bound on λ then gives the desired result.

We prove (5.15) by induction on k, the uniformity of the complex. For $k = 1$ the bound is trivial since $\mathbb{E}[f_1] = \mu_1$. Assume the statement holds for any $(k-1)$-uniform λ-product, and for each $i = 2, \ldots, k$ let T_i denote the vertices in $X[1]$ that under-sample f_i:

$$T_i = \left\{ v_1 \in X[1] \;\middle|\; \mathop{\mathbb{E}}_{v_i \in X_{v_1}[i]}[f(v_i)] < \mu_i - \varepsilon_i \right\}.$$

Since each $(X[1], X[i])$ is a λ-bipartite expander, Claim 5.7.3 implies

$$\mathbb{P}\left[\bigcup_{i=2}^{k} T_i\right] \leq \sum_{i=2}^{k} \frac{\lambda^2}{\mu_i \varepsilon_i^2} \leq \lambda^2 \frac{k-1}{\varepsilon^2 \cdot \min_{j>1}\{\mu_j\}}.$$

We de-correlate our expected product along the first coordinate by conditioning on

$G = X[1] \setminus \{\bigcup_{i=2}^{k} T_i\}$:

$$\mathbb{E}\left[\prod_{i=1}^{k} f_i(v_i)\right] \geq \mathbb{E}\left[\mathbf{1}_G(v_0)\prod_{i=1}^{k} f_i(v_i)\right]$$
$$\geq \mathbb{E}[\mathbf{1}_G(v_0)f_0(v_0)] \cdot \min_{v_1 \in G}\left(\mathop{\mathbb{E}}_{\{v_2,v_3,\ldots,v_k\}\in X_{v_1}}\left[\prod_{i=2}^{k} f_i(v_i)\right]\right).$$

We bound the two terms separately. For the first:

$$\mathbb{E}[\mathbf{1}_G(v_0)f_0(v_0)] \geq \mathbb{E}[f_0(v_0)] - \mathbb{P}\left[\bigcup_{i=2}^{k} T_i\right] \tag{5.16}$$
$$\geq \mu_0 - \lambda^2 \frac{k-1}{\varepsilon^2 \cdot \min_{j>1}\{\mu_j\}}. \tag{5.17}$$

To bound the minimum term, we apply the inductive hypothesis on every link X_{v_1}. Toward this end, for notational simplicity re-index the sides of X_{v_1} to be from 1 to $k-1$ and set:

1. $f'_i = f_{i+1}$,
2. $\varepsilon'_i = \varepsilon_{i+1}$,
3. $\mu'_i = \mu_{i+1} - \varepsilon_{i+1}$ and $\mu' = \min_{j>1} \mu'_i$.

Since $\mathbb{E}_{X_{v_1}[i]}[f'_i] \geq \mu'_i$ by assumption on the links, applying the inductive hypothesis to X_{v_1} and $\{f'\}$ bounds the minimum by

$$\prod_{i=1}^{k-1}\left(\mu'_i - (i-1)\varepsilon'_i - \lambda^2 \frac{k-i}{\varepsilon^2 \cdot \min_{j>i}(\mu'_j - (i-1)\varepsilon'_j)}\right)$$

Combining this with the prior term (corresponding to $i = 0$) gives the desired bound

$$\prod_{i=1}^{k}\left(\mu_i - (i-1)\varepsilon_i - \lambda^2 \frac{k-i}{\varepsilon^2 \cdot \min_{j>i}(\mu_j - (i-1)\varepsilon_j)}\right).$$

□

We note a more involved variant of the above applying different bounds in each level of induction implies the following finer-grained bound:

Claim 5.7.4. Let X be a k-partite λ-product. Let $f_i : X[i] \to [0,1]$ be such that $\mathbb{E}_{v_i \in X[i]}[f(v_i)] \geq \mu_i$. Let $\{\varepsilon_{i,\ell} > 0\}_{1 \leq i < \ell \leq k}$ and let $s^{(m)}_{i,\ell} = \sum_{j=\ell}^{m} \varepsilon_{j,i}$ (and for brevity $s^{(m)}_{m+1,\ell} = 0$). Then

$$\mathop{\mathbb{E}}_{\{v_0,v_1,\ldots,v_k\} \in X(k)} \left[\prod_{i=1}^{k} f_i(v_i) \right] \geq \prod_{i=1}^{k} \left(\mu_i - s^{(k-1)}_{i,i} - \lambda^2 \sum_{j=0}^{i-1} \frac{\mu_j - s^{(k-1)}_{i,j}}{\varepsilon^2_{i,j}} \right). \tag{5.18}$$

We omit the proof which is an unenlightening technical extension of the above.

Using our lower bound on negative correlation, we show that for any function tuple $f = (f_1, f_2, \ldots, f_k)$, the random variables $f(X)$ and $f(X^{ind})$ are actually *distributionally* close. This will allow us to immediately recover any concentration bound on X^{ind} up to the distributional error. More formally, given a k-partite simplicial complex X, let D_{ind} denote the distribution supported by of $X^{ind,p}$, i.e. $\mathbb{P}_{D_{ind}}[\{v_1, v_2, \ldots, v_k\}] = \prod_{i=1}^{k} \mathbb{P}_{X[i]}[v_i]$. Moreover, for any distribution D over $\Omega = X[1] \times X[2] \times \cdots \times X[k]$, and a function $\bar{f} : \Omega \to \mathbb{R}^k$ define the *push forward* distribution $\bar{f}_* D : \mathbb{R}^k \to [0,1]$ as

$$\bar{f}_* D(\bar{x}) = \mathop{\mathbb{P}}_{\omega \sim D} \left[\bar{f}(\omega) = \bar{x} \right].$$

Let $D_X = \pi_k$ denote the distribution over k-sets of X. We show the pushforward distributions $\bar{f}_* D_X$ and $\bar{f}_* D_{ind}$ are close in TV-distance.

Lemma 5.7.5. *Let $\varepsilon, \eta > 0$ and let X be a k-partite λ-product for $\lambda \leq \frac{\eta^{2.5}}{8k^4}$. Then for any $f_i : X[i] \to \{0,1\}$ and $\bar{f} = (f_1, f_2, \ldots, f_k)$:*

$$d_{TV}(\bar{f}_* D_X, \bar{f}_* D_{ind}) \leq \eta.$$

Proof. Note that both distributions are supported in $\{0,1\}^k$ by our assumption on the domain of $\bar{f}$. Thus we need to show for every $B \subseteq \{0,1\}^k$:

$$\left| \mathop{\mathbb{P}}_{\bar{f}_* D_X}[B] - \mathop{\mathbb{P}}_{\bar{f}_* D_{ind}}[B] \right| \leq \eta.$$

Indeed since $\mathbb{P}_{\bar{f}_* D_X}[B] - \mathbb{P}_{\bar{f}_* D_{ind}}[B] = \mathbb{P}_{\bar{f}_* D_{ind}}\left[\bar{B}\right] - \mathbb{P}_{\bar{f}_* D_X}\left[\bar{B}\right]$ for $\bar{B} = \{0,1\}^k \setminus B$, it is enough to just show the one-sided bound

$$\mathop{\mathbb{P}}_{\bar{f}_* D_X}[B] \geq \mathop{\mathbb{P}}_{\bar{f}_* D_{ind}}[B] - \eta.$$

Fix any event $B \subseteq \{0,1\}^k$. We'll decompose B into two parts. First, we'll look at the subset B_1 on which f takes many 'unlikely' values, and argue the difference is small just because D_X and D_{ind} have the same marginals over each coordinate of $\{0,1\}^k$. In particular, let $I = \left\{ i \in [k] \,\middle|\, \mathbb{E}[f_i] \leq \frac{\eta}{2k} \right\}$ and $J = \left\{ i \in [k] \,\middle|\, \mathbb{E}[f_i] \geq 1 - \frac{\eta}{2k} \right\}$ be the sets of coordinates with extreme expectations and let

$$B_1 = B \cap \left\{ x \,\middle|\, \left(\bigvee_{i \in I} \{f_i(x) = 1\} \right) \vee \left(\bigvee_{j \in J} \{f_j(x) = 0\} \right) \right\}.$$

Set $B_2 = B \setminus B_1$.

We first show that $\mathbb{P}_{\bar{f}_* D_X}[B_1] - \mathbb{P}_{\bar{f}_* D_{ind}}[B_1] \leq \frac{\eta}{2}$. Since both distributions have the same marginals, we can write for either D_X or D_{ind} by a union bound:

$$\mathbb{P}[B_1] \leq \sum_{i \in I} \mathbb{P}[f_i = 1] + \sum_{j \in J} \mathbb{P}[f_j = 0] \leq (|I| + |J|)\frac{\eta}{2k} \leq \frac{\eta}{2}.$$

In particular the difference is no more than $\frac{\eta}{2}$.

Moving on to B_2, we decompose the probability as a sum over k-wise products:

$$\mathop{\mathbb{P}}_{\bar{f}_* D_X}[B_2] = \sum_{\bar{b} \in B_2} \mathop{\mathbb{P}}_{D_X}\left[\bigwedge_{i=1}^{k} f_i = b_i\right] = \sum_{\bar{b} \in B_2} \mathop{\mathbb{E}}_{\{v_1, v_2, \ldots, v_k\} \sim D_X}\left[\prod_{i=1}^{k} |b_i - f_i(v_i)|\right].$$

We note that the functions $|b_i - f_i(v_i)|$ have values in $[0,1]$ and by assumption have expectation $\geq \frac{\eta}{2}$ (since otherwise $\bar{b} \in B_1$). Then by Proposition 5.7.2 setting $\varepsilon = \frac{\eta}{2}$ we have

$$\begin{aligned}
\mathop{\mathbb{P}}_{D_X}[B_2] &\geq (1 - \frac{\eta}{2k})^k \sum_{\bar{b} \in B_2} \prod_{i=1}^{k} \mathop{\mathbb{E}}_{v_i \in X[i]}[|b_i - f_i(v_i)|] \\
&= (1 - \frac{\eta}{2k})^k \sum_{\bar{b} \in B_2} \prod_{i=1}^{k} \mathop{\mathbb{P}}_{D_{ind}}[\mathbf{1}_{f_i = b_i}] \\
&= (1 - \frac{\eta}{2k})^k \sum_{\bar{b} \in B_2} \prod_{i=1}^{k} \mathop{\mathbb{P}}_{D_{ind}}[\mathbf{1}_{f_i = b_i}] \\
&\geq \mathop{\mathbb{P}}_{D_{ind}}[B_2] - \frac{\eta}{2}
\end{aligned}$$

where we've used the marginal equivalence of D_X and D_{ind} and independence of D_{ind} in the second and third steps respectively. Combining our bounds on B_1 and B_2 gives the result. □

We can now easily prove Theorem 5.7.1.

Proof of Theorem 5.7.1. Let $f : X(1) \to [0,1]$. For the partite case, set $f_i = f|_{X[i]}$. Consider the set of 'bad' k-tuples that mis-sample f:

$$B = \left\{ \left| \sum_{j=1}^{k} f(v_j) - \mu \right| > \varepsilon \right\}.$$

By Lemma 5.7.5, the TV-distance between the distribution over values of $\bar{f}(x) =$

$f(x_1, x_2, \ldots, x_k)$ over X and $X^{ind,p}$ is at most η so we have:

$$\begin{aligned}\sigma(X,k,\varepsilon) &= \max_{\bar{f}} \mathop{\mathbb{P}}_{\bar{f}_* D_X}[B] \\ &\leq \max_{\bar{f}} \mathop{\mathbb{P}}_{\bar{f}_* D_{ind}}[B] + \eta \\ &\leq S(X^{ind,p}, k, \varepsilon) + \eta\end{aligned}$$

as desired. The two-sided case follows the same argument after taking the partitification $P(X)$ of X and setting $f_i(v,i) = f(v)$. □

5.7.2 Lipschitz Concentration for 'Weak' HDX

In this section we study concentration of measure under weaker quantitative assumptions on the underlying complex, requiring only that the down-up walk of X has a good spectral gap. Our results in this special case are incomparable to our prior strategies. While we cannot show Chernoff (subgaussian) strength bounds directly in this setting, we are able to give exponential concentration for the substantially broader class of *Lipschitz* functions. To achieve the subgaussian bounds needed for Theorem 5.5.1, we pass to lower level skeletons. Before stating the formal results we briefly recall some notation from Section 5.4.6.

Given a function $f : X(k) \to \mathbb{R}$, let $Z = f(x_1, \ldots, x_k)$ and $Z'_{(i)} = f(x_1, \ldots z_i, \ldots, x_k)$ where z_i is sampled conditional on x_{-i}. We call a f ν-Lipschitz if with probability 1 over Z, Z'

$$\sum_{i=1}^{k} (Z - Z'_{(i)})_+^2 \leq \nu$$

and say it has ν-bounded difference if for all adjacent (s, s') in the down-up walk:

$$(f(s) - f(s'))^2 \leq \frac{\nu}{k}$$

We prove that ν-Lipschitz functions on HDX satisfy exponential concentration of measure.

Theorem 5.7.6. *Let X be a d-uniform simplicial complex such that $U_{d-1}D_d$ has spectral gap at least $\frac{1}{Cd}$ for some $C > 0$. Then for any $0 \le k \le d$ and ν-Lipschitz function $f : X(k) \to \mathbb{R}$:*

1. *Upper Tail:* $\mathbb{P}[f - \mathbb{E}[f] \ge t] \le 2e^{-t/\sqrt{(C+1)\nu}}$
2. *Lower Tail:* $\mathbb{P}[f - \mathbb{E}[f] \le -t] \le 2e^{-t/\sqrt{(C+1)\nu}}$

As (almost) immediate corollaries, we get exponential concentration for TD and SI complexes.

Corollary 5.7.7 (Concentration up to the TD Barrier)**.** *Let X be a d-uniform λ-TD complex for $\lambda < 1$. Then for any $0 \le k \le d$ and ν-Lipschitz function $f : X(k) \to \mathbb{R}$*

1. ***Upper tail:*** $\mathbb{P}[f - \mathbb{E}[f] \ge t] \le 2e^{-\frac{t}{\sqrt{c_\lambda \nu}}}$
2. ***Lower tail:*** $\mathbb{P}[f - \mathbb{E}[f] \le -t] \le 2e^{-\frac{t}{\sqrt{c_\lambda \nu}}}$

where $c_\lambda \le 1 + e^{\frac{\lambda}{(1-\lambda)}}$.

Corollary 5.7.8 (Concentration under Spectral Independence)**.** *Fix $\eta > 0$ and let X be a d-uniform η-spectrally independent complex. Then for any $0 \le k \le d$, $g : X(1) \to [0, 1]$, and $f = U_{1,k}g$:*

1. ***Upper tail:*** $\mathbb{P}[f - \mathbb{E}[f] \ge \varepsilon] \le 4e^{-c_\eta \varepsilon \sqrt{k}}$
2. ***Lower tail:*** $\mathbb{P}[f - \mathbb{E}[f] \le -\varepsilon] \le 4e^{c_\eta \varepsilon \sqrt{k}}$

where $c_\eta \le \frac{1}{12 \max\{1, \sqrt{\eta}\}}$.

We note it is possible to extend the latter to lifted Lipschitz functions from an appropriate skeleton.

The proof of Theorem 5.7.6 is via a variant of the *Herbst Argument*, a classical strategy for proving concentration of random variables based on applying functional inequalities like MLSI and LSI to the moment generating function. While sparse simplicial complexes cannot have bounded (modified) log-sobolev constants, a somewhat lesser known variant of the Herbst argument using spectral gap was developed by Aida and Stroock [5], Bobkov and Ledoux [68], and Boucheron, Lugosi, and Massart [76]. We will give an elementary adaption of their method to HDX achieving the above bounds.

The key to Theorem 5.7.6 is really the following approximate variant of the Efron-Stein inequality, sometimes called 'approximate tensorization of variance'. We remark that similar inequalities for local-spectral expanders appear in the literature [236, 94].

Lemma 5.7.9 (Approximate Efron-Stein Inequality)**.** *Let X be a d-uniform simplicial complex such that $\lambda_2(U_{d-1}D_d) \leq 1 - \frac{1}{Cd}$ for some $C > 0$. Then for any $0 \leq k \leq d$, $f : X(k) \to \mathbb{R}$, and $Z = f(X)$ we have:*

$$Var(f) \leq (C+1) \sum_{i=1}^{k} \mathbb{E}\left[(Z - Z'_{(i)})_+^2\right]$$

Proof. It is sufficient to instead prove that any k-uniform complex whose down-up walk $U_{k-1}D_k$ has gap $\frac{1}{C'k}$ satisfies

$$Var(f) \leq C' \sum_{i=1}^{k} \mathbb{E}\left[(Z - Z'_{(i)})_+^2\right].$$

The result then follows from [276, Theorem 3.5], who shows that $\lambda_2(U_{d-1}D_d) \leq 1 - \frac{1}{Cd}$ implies $\lambda_2(U_{k-1}D_k) \leq 1 - \frac{1}{(C+1)k}$ for any $0 \leq k \leq d$.

With this in mind, let E_i be the ith averaging operator:

$$E_i f(s) = \mathop{\mathbb{E}}_{s_i \sim X_{s_{-i}}} [f(s_{-i} \cup s_i)],$$

where we recall s_{-i} is s without the ith vertex in its ordering. Define the ith Laplacian $L_i = I - E_i$, and the total laplacian operator as $L = I - UD = \frac{1}{k+1} \sum L_i$. By assumption on the spectral gap of UD, we have:

$$Var(f) \leq Ck\langle f, Lf \rangle$$

On the other hand expanding out the Laplacian gives:

$$\begin{aligned} Ck\langle f, Lf \rangle &= C \sum_{i=1}^{k} \langle f, L_i f \rangle \\ &= C \sum_{i=1}^{k} \langle L_i f, L_i f \rangle \\ &= C \sum_{i=1}^{k} \mathop{\mathbb{E}}_{s \sim X(k)} [Var_s^{(i)}(f)] \end{aligned}$$

where $Var_s^{(i)}(f)$ is the variance of f over the link $X_{s_{-i}}$. Finally examining this expected variance, we have:

$$\mathop{\mathbb{E}}_{s \sim X(k)} [Var_s^{(i)}(f)] = \frac{1}{2} \mathbb{E}[(Z - Z'_{(i)})^2] = \mathbb{E}[(Z - Z'_{(i)})_+^2]$$

where both equalities follow from the fact that Z and $Z'_{(i)}$ are i.i.d conditioned on s_{-i}. Combining with the above completes the proof. □

We can now prove Theorem 5.7.6 closely following the approach of [76].

Proof of Theorem 5.7.6. We prove only the upper tail. The lower tail follows similarly. It is sufficient to prove the following exponential integrability of the moment generating

function:

$$\mathbb{E}\left[e^{\frac{1}{\sqrt{C\nu}}(f-\mathbb{E}[f])}\right] \leq 2, \tag{5.19}$$

since we then have:

$$\begin{aligned}\mathbb{P}[f-\mathbb{E}[f]>t] &= \mathbb{P}\left[\exp\left(\frac{1}{\sqrt{C\nu}}(f-\mathbb{E}[f])\right) > \exp\left(\frac{1}{\sqrt{C\nu}}t\right)\right] \\ &\leq 2\exp\left(-\frac{1}{\sqrt{C\nu}}t\right)\end{aligned}$$

by Markov.

For notational simplicity, let $g = f - \mathbb{E}[f]$ and let $Z = g(X)$. Following [76], we use Efron-Stein to set up a recurrence relating $e^{\lambda Z}$ to $e^{\lambda Z/2}$ for small enough $\lambda > 0$. Applying Efron-Stein to the latter gives:

$$\begin{aligned}\mathbb{E}[e^{\lambda Z}] - \mathbb{E}[e^{\lambda Z/2}]^2 &\leq C\sum_{i=1}^{k}\mathbb{E}\left[(e^{\lambda Z/2} - e^{\lambda Z'_{(i)}/2})_+^2\right] \\ &\leq C\sum_{i=1}^{k}\mathbb{E}\left[e^{-\lambda Z}(1-e^{\lambda(Z'_{(i)}-Z)/2})_+^2\right] \\ &\leq \frac{C\lambda^2}{4}\mathbb{E}\left[e^{-\lambda Z}\sum_{i=1}^{k}(Z-Z'_{(i)})^2\right] \\ &\leq \frac{C\lambda^2\nu}{4}\mathbb{E}\left[e^{-\lambda Z}\right]\end{aligned}$$

since f (and therefore g) are ν-bounded. Re-arranging we have the recurrence:

$$\mathbb{E}[e^{\lambda Z}] \leq \frac{1}{1-\frac{\lambda^2 C\nu}{4}}\mathbb{E}[e^{\lambda Z/2}]^2.$$

It is then an elementary exercise to solve when $\lambda \leq \frac{1}{\sqrt{C\nu}}$ and derive Equation (5.19) (see [76, Section 3.6] for the full derivation). □

We now prove the corollaries for SI/TD-complexes

Proof of Corollary 5.7.7. By Theorem 5.4.23, X is a $\frac{\lambda}{(d-1)(1-\lambda)}$-one-sided local-spectral expander, so Theorem 5.4.19 gives $\lambda_2(U_{d-1}D_d) \le 1 - \frac{1}{d}e^{\frac{\lambda}{1-\lambda}}$. Plugging this into Theorem 5.7.6 gives the result. □

In the spectral independence regime, one must be slightly more careful since an η-spectrally independent system has $C \approx k^\eta$. We recover concentration by looking only at lifted functions and analyzing concentration on some lower skeleton.

Proof of Corollary 5.7.8. We prove the upper tail. The lower tail follows similarly. We break the proof into two cases. First, assume that $\eta \le 2$ and let $k' = \lfloor \frac{d}{3} \rfloor$. The k'-skeleton of X is a one-sided $\frac{1}{k'}$-local-spectral expander. Thus by Theorem 5.4.19 $\lambda(U_{k'-1}D_{k'}) \le 1 - \frac{1}{ek'}$ and Theorem 5.7.6 implies every $k \le k'$ has concentration:

$$\mathbb{P}[f - \mathbb{E}[f] > \varepsilon] \le 2e^{-\frac{1}{\sqrt{e+1}}\varepsilon\sqrt{k}}.$$

Further, by Lemma 5.4.15 every $k > k'$ has concentration:

$$\begin{aligned}\mathbb{P}[f - \mathbb{E}[f] > \varepsilon] &\le 2e^{-\frac{1}{2\sqrt{e+1}}\varepsilon\sqrt{k'}}(1 - e^{-\frac{\varepsilon^2}{12}k}) \\ &\le 4e^{-\frac{1}{2\sqrt{6(e+1)}}\varepsilon\sqrt{d}} \\ &\le 4e^{-\frac{1}{12}\varepsilon\sqrt{d}}\end{aligned}$$

assuming $k \ge 6$ and $\varepsilon^2 \ge \frac{12}{k}$ (else the bound is trivial).

Otherwise, assume $\eta > 2$ and let $k' = \lfloor \frac{d}{\eta} \rfloor$. The k'-skeleton of X is a $\frac{\eta}{(1-\frac{1}{\eta})d}$-one-sided local-spectral expander. Then by Theorem 5.4.19 $\lambda(U_{k'-1}D_{k'}) \le 1 - \frac{1}{e^2k'}$ and Theorem 5.7.6 implies for every $k \le k'$:

$$\mathbb{P}[f - \mathbb{E}[f] > \varepsilon] \le 2e^{-\frac{1}{\sqrt{e^2+1}}\varepsilon\sqrt{k}}.$$

Further, by Lemma 5.4.15 every $k > k'$ has concentration:

$$\begin{aligned}\mathbb{P}[f - \mathbb{E}[f] > \varepsilon] &\leq 2e^{-\frac{1}{2\sqrt{e^2+1}}\varepsilon\sqrt{k'}}(1 - e^{-\frac{\varepsilon^2}{12}k}) \\ &\leq 4e^{-\frac{1}{2\sqrt{4\eta(e^2+1)}}\varepsilon\sqrt{d}} \\ &\leq 4e^{-\frac{1}{12\sqrt{\eta}}\varepsilon\sqrt{d}}\end{aligned}$$

assuming $\eta \leq \frac{d}{2}$ and $\varepsilon^2 \geq \frac{12}{k+1}$ (again, otherwise the stated bound is trivial).

□

Subgaussian Concentration on Skeletons

We now show that it is possible to recover subgaussian concentration from the above analysis by passing to skeletons of size roughly $\Theta(\sqrt{d})$.

Corollary 5.7.10 (Subgaussian concentration for skeletons)**.** *Let X be a d-uniform simplicial complex with $\lambda_2(U_{d-1}D_d) \leq 1 - \frac{1}{Cd}$. Then for any $k \leq d$, $\nu > 0$, and function $f : X(k) \to \mathbb{R}$ with ν-bounded difference:*

1. ***Upper Tail**: $\mathbb{P}[f - \mathbb{E}[f] \geq t] \leq 2e^{-\frac{t}{\sqrt{4(C+1)\nu}}\cdot\sqrt{\frac{d}{k}}} + e^{-\frac{t^2}{4\nu}}$*
2. ***Lower Tail:** $\mathbb{P}[f - \mathbb{E}[f] \leq -t] \leq 2e^{-\frac{t}{\sqrt{4(C+1)\nu}}\cdot\sqrt{\frac{d}{k}}} + e^{-\frac{t^2}{4\nu}}$*

Proof. The proof is immediate from Theorem 5.7.6 and Lemma 5.4.16, setting $f_{up}(t,\nu) = f_{low}(t,\nu) = 2e^{\frac{-t}{\sqrt{(C+1)\nu}}}$. □

Corollary 5.7.10 does not imply full subgaussian concentration for all functions with bounded difference, but does for certain critical sub-classes such as bounded 1-lifts:

Corollary 5.7.11. *Let X be a d-uniform simplicial complex with $\lambda_2(U_{d-1}D_d) \leq 1 - \frac{1}{Cd}$. Then for any $k \leq d$ and $f : X(1) \to [0,1]$ of expectation μ:*

1. ***Upper Tail**: $\mathbb{P}_{s\sim X(k)}[U_{1,k}f(s) - \mu \geq \varepsilon] \leq 3e^{-\frac{1}{4\sqrt{C+1}}\varepsilon^2 k}$*

2. ***Lower Tail:*** $\mathbb{P}_{s\sim X(k)}[U_{1,k}f(s) - \mu \leq -\varepsilon] \leq 3e^{-\frac{1}{4\sqrt{C}+1}\varepsilon^2 k}$

Proof. The proof is immediate from Lemma 5.4.16 and the observation that $U_{1,k}f$ has $\frac{1}{k}$-bounded difference. □

Note that by similar arguments as above, this implies subgaussian concentration for the skeletons of λ-TD and η-SI complexes. This is already implicit in the statement of Theorem 5.5.1, since TD/SI-complexes are c-nice since an appropriate skeleton of these complexes are nice (Claim 5.4.28). Finally, we remark it is likely possible to generalize the above strategy from bounded difference to Lipschitz functions on sufficiently strong local-spectral expanders by more carefully analyzing the eigenspaces of the Laplacian and higher order random walks associated with the d-lift (e.g. as in [239, 111, 162]).

5.7.3 Inclusion Sampling and the Proof of Theorem 5.5.1

We are finally ready to prove Theorem 5.5.1 in full. We again restate the theorem in the final two cases for convenience.

Theorem 5.7.12 (Sampling on HDX)**.** *Let X be a k-uniform complex which is either*

1. *The skeleton of a d-uniform partite 2^{-cd}-one-sided HDX Y for $d \geq swap(k)$*
2. *The skeleton of a d-uniform complex Y with $\lambda(U_{d-1}D_d) \leq 1 - \frac{1}{cd}$ for $d \geq 2k^2$*

Then for any $i < k$ the containment graph $(X(k), X(i))$ is a (ε, β) sampler for

$$\beta = \frac{64}{\varepsilon}\exp(-\Omega_c(\varepsilon^2\frac{k}{i})).$$

Moreover, if X is a k-skeleton of a d-uniform complex Y, then for every $k < k' \leq d$, the graph $(Y(k'), Y(i))$ is also $(\varepsilon, O(\beta))$-function samplers.

Note that, combined with Claim 5.4.28, Theorem 5.2.3 is an immediate corollary.

Proof of Theorem 5.7.12 (partite and spectral gap case). In both cases the 'moreover' part follows directly from Lemma 5.4.15.

Spectral Gap.

Let Y be the promised d-uniform complex such that X is an (at most) $\sqrt{\frac{d}{2}}$-uniform skeleton of Y such that Y's down-up walk has spectral gap $\lambda(U_{d-1}D_d) \leq 1 - \frac{1}{cd}$. For any $g : X(i) \to [0,1]$ Theorem 5.7.6 implies $U_{i,d}g$ has the following exponential tail:

$$\mathbb{P}[U_{i,d}g(s) - \mathbb{E}[g] \geq \varepsilon] \leq 2e^{-\Omega(\varepsilon\frac{\sqrt{d}}{i})} \leq 2e^{-\Omega(\varepsilon\frac{k}{i})}$$

since $U_{i,d}g$ has $\frac{i^2}{d}$-bounded difference. By Lemma 5.4.16, $U_{i,k}g$ therefore has tail:

$$\mathbb{P}[U_{i,k}g(s) - \mathbb{E}[g] \geq \varepsilon] \leq 2e^{-\Omega(\varepsilon\frac{k}{i})} + \frac{2}{\varepsilon}e^{-\Omega(\varepsilon^2\frac{k}{i})},$$

where the latter term is from the following concentration bound on the complete complex:

Claim 5.7.13. For any $i \leq k \leq n$ and $g : \Delta_n(i) \to [0,1]$:

$$\mathbb{P}_{\Delta_n(k)}[U_{i,k}g - \mathbb{E}[g] > \varepsilon] \leq \frac{2}{\varepsilon}e^{-\Omega(\varepsilon^2\frac{k}{i})}$$

and likewise for the lower tail.

We prove this claim in Section 5.12 via a simple application of Lemma 5.5.9. The lower tail follows similarly.

Partite HDX.

The partite case of Theorem 5.5.1 introduces a new challenge. We cannot directly reduce to the complete complex since we are not on a sufficiently low dimensional skeleton, and the faces complex of X isn't a (sufficiently strong) HDX to blackbox apply Theorem 5.5.2. Nevertheless, lacking expansion does not necessarily mean the faces complex is a poor sampler. We will argue F_X satisfies Chernoff by splitting its analysing into two

components 1) sampling a good partition of *colors* 2) sampling a good subset conditioned on this partition. We show the first step may be reduced to concentration on the faces complex of the complete complex (the so-called *swap complex*), while the second reduces to concentration for partite HDX (Theorem 5.7.1).

Let $k' = k/2$. By Lemma 5.4.15, it is enough to show for any $f : X(\ell) \to [0,1]$:

$$\mathop{\mathbb{P}}_{X(k)}[|U_{\ell,k}f - \mathbb{E}[f]| > \varepsilon] \leq 8\exp\left(-c'\frac{k}{\ell}\right) \tag{5.20}$$

for some universal constants $c' > 0$. The stated result is then an immediate consequence of Lemma 5.4.15.

To prove this, let's recall the notion of the swap complex.

Definition 5.7.14 (The Swap Complex)**.** Let $n \geq k \geq \ell \in \mathbb{N}$. The (ℓ, k, n)-swap complex is the $\lfloor \frac{k}{\ell} \rfloor$-uniform simplicial complex $C = C_{\ell,k,n}$ whose vertices are all size-ℓ subsets $[n]$:

$$C(1) = \binom{[n]}{\ell},$$

and whose top-level faces are all possible pair-wise disjoint ℓ-sets:

$$C\left(\left\lfloor \frac{k}{\ell} \right\rfloor\right) = \left\{ \{s_1, s_2, \ldots, s_{\lfloor \frac{k}{\ell} \rfloor}\} : \bigcup_{i \neq j} s_i \cap s_j = \emptyset \right\}$$

endowed with the uniform distribution.

In other words, $C_{\ell,k,n}$ is exactly $F^{\ell}_{\Delta_n(k)}$, the faces complex of the k-uniform complete complex on n vertices. Recall that $\mathrm{swap}(k)$ is the smallest d such that $C_{\ell,k/2,d}$ satisfies a Chernoff bound. We first prove $\mathrm{swap}(k) \leq O(k^2)$.

Theorem 5.7.15. *Let $C = C_{\ell,k,n}$ for $n \geq (k+1)\ell$. For any $f : C(0) \to [0,1]$ and $\varepsilon \in (0,1)$*

1. ***Upper tail**: $\mathbb{P}[U_{1,\frac{k}{\ell}}f - \mathbb{E}[f] > \varepsilon] \leq \exp(-c_1\varepsilon^2\frac{k}{\ell})$*

2. ***Lower tail:*** $\mathbb{P}[U_{1,\frac{k}{\ell}} f - \mathbb{E}[f] < -\varepsilon] \leq \exp(-c_2 \varepsilon^2 \frac{k}{\ell})$

for some universal constants $c_1, c_2 > 0$.

The proof simply follows from the fact that $C_{\ell,k,n}(k/\ell)$ is negatively correlated under these parameters. We conjecture that the swap complex should exhibit optimal concentration even when $n = \Theta(k)$. We discuss this (and prove the above) in Section 5.13. With this in mind it is enough to prove the theorem assuming $d = \text{swap}(k)$ (namely this subsumes the partite case of the simplified niceness definition).

Let X be a k-uniform skeleton of a d-partite $2^{-\Omega(d)}$-HDX for $d \geq \text{swap}(k)$ and let $k' = k/2$. By Proposition 5.5.8 and Lemma 5.4.15, it is enough to prove the $F_X^{(\ell)}(k'/\ell)$ is an $(\varepsilon, 8e^{-\Omega(\varepsilon^2 \frac{k'}{\ell})})$-additive function sampler for any $\varepsilon > 0$.

Fix $f : X(\ell) \to [0,1]$ of expectation μ viewed as a function on the vertices of the faces complex, and let $m = \lfloor \frac{k'}{\ell} \rfloor$. Consider the random variable $\mathcal{J} \coloneqq (col(s_1), col(s_2), \ldots, col(s_m))$ corresponding to the partition generated by drawing $(s_1, \ldots, s_m)$ from the faces complex of X. For a fixed $\mathcal{J} = \{I^1, I^2, \ldots, I^m\}$, denote the conditional expectation of f under $\mathcal{J}$ as

$$\mu_{\mathcal{J}} \coloneqq \mathop{\mathbb{E}}_{s \sim X(\ell)}[f |\ col(s) \in \mathcal{J}].$$

Further let $C_{\mathcal{J}}$ denote the complex generated by restricting to top-level faces of C with partition $\mathcal{J}$. We define two 'bad' events based on $\mathcal{J}$ outside which our sampling is ε-accurate:

1. **(Bad Colors):** E_1 the event that $\mu_{\mathcal{J}}$ deviates significantly from μ

$$|\mu_{\mathcal{J}} - \mu| \geq \varepsilon/2.$$

2. **(Bad Sampling):** E_2 the event that $(s_1, \ldots, s_m) \sim C_{\mathcal{J}}$ samples $\mu_{\mathcal{J}}$ poorly

$$\left| \frac{1}{m} \sum_{i \in [m]} f(s_i) - \mu_{\mathcal{J}} \right| \geq \varepsilon/2.$$

It is then sufficient to prove the following three claims:

1. $\mathbb{P}_{(s_1,s_2,\ldots,s_m)\sim C} \left[\left| \frac{1}{m} \sum_{i \in [m]} f(s_i) - \mu \right| > \varepsilon \right] \leq \mathbb{P}[E_1] + \mathbb{P}[E_2 \wedge \neg E_1]$.
2. $\mathbb{P}[E_1] \leq 2\exp(-c'\varepsilon^2 m)$.
3. $\mathbb{P}[E_2 \wedge \neg E_1] \leq 4\exp(-c'\varepsilon^2 m)$.

The first claim is immediate from definition, as the event $|\sum f - \mu| > \varepsilon$ only occurs if $E_1 \vee E_2$ holds, the probability of which is bounded by $\mathbb{P}[E_1] + \mathbb{P}[E_2 \wedge \neg E_1]$ as desired.

Bounding $\mathbb{P}[E_1]$.

Observe $\mathcal{J}$ is equidistributed with the complex $C_{\ell,k/2,d}$ and define $g : C(1) \to [0,1]$ by $g(I^j) = \mathbb{E}_C[f \mid I^j \in col(s)]$. For $\mathcal{J} = \{I^1, \ldots, I^m\}$ we then have:

1. $\mathbb{E}[g] = \mathop{\mathbb{E}}_{I \sim C(1)} [\mathbb{E}[f \mid col(s) \in I]] = \mu$
2. $\frac{1}{m} \sum_{j=1}^{m} g(I^j) = \mu_{\mathcal{J}}$.

Since $d \geq \mathrm{swap}(k)$ by assumption, $C_{\ell,k',d}$ satisfies Chernoff and we can write

$$\begin{aligned}
\mathbb{P}[E_1] &= \mathop{\mathbb{P}}_{\mathcal{J}} \left[|\mu_{\mathcal{J}} - \mu| > \frac{\varepsilon}{2} \right] \\
&= \mathop{\mathbb{P}}_{\{I^1,\ldots,I^m\}\sim C(m)} \left[\left| \frac{1}{m} \sum_{j=1}^{m} g(I^j) - \mathbb{E}[g] \right| > \frac{\varepsilon}{2} \right] \\
&\leq 2\exp\left(-c'\varepsilon^2 m\right).
\end{aligned}$$

since $\frac{1}{m}\sum_i g(s_i)$ is $\frac{1}{m}$-bounded.

Bounding $\mathbb{P}[E_2 \wedge \neg E_1]$.

It is left to argue that for any fixed $\mathcal{J} = \{I^1, I^2, \ldots, I^m\}$, $C_{\mathcal{J}}$ is a partite $2^{-\Omega(k)}$-HDX. The result is then immediate from applying Theorem 5.7.1 with $\eta = \exp(-\frac{1}{4}\varepsilon^2 m)$:

$$\begin{aligned}
\mathbb{P}[E_2 \wedge \neg E_1] &\leq \mathop{\mathbb{P}}_{\{s_1,\ldots,s_m\}\sim D}[E_2 |\ \mathcal{J} \in \neg E_1] \\
&= \mathop{\mathbb{E}}_{\mathcal{J}\in\neg E_1}\left[\mathop{\mathbb{P}}_{s\sim C_{\mathcal{J}}}\left[\left|\frac{1}{m}\sum_{i\in[m]} f(s_i) - \mu_{\mathcal{J}}\right| \geq \frac{\varepsilon}{2}\right]\right] \\
&\leq 3\exp\left(-\frac{1}{4}\varepsilon^2 m\right).
\end{aligned}$$

By construction it is easily checked $C_{\mathcal{J}}$ is a partite complex whose links are isomorphic to swap walks S_{I_i,I_j} within a link of the original complex X. By Theorem 5.4.21, $C_{\mathcal{J}}$ is therefore a partite λ-one-sided HDX for $\lambda \leq 2^{-\Omega(d)}$ as well. □

5.8 Agreement Testing

In this section we use our reverse hypercontractive inequality to prove several new agreement testing theorems on HDX. In the first subsection, we give new agreement tests in the 99%-regime between any two levels of an HDX. In the second subsection, we explore the role of reverse hypercontractivity in the 1%-regime, and prove that HDX admit agreemment tests with optimal *local* soundness (a weaker notion of soundness only requiring globalness on links), and show under the stronger assumption of ℓ_∞-expansion that this guarantee can be propagated to a true tester with optimal soundness.

5.8.1 Background

Suppose U is some finite set of vertices and $S \subseteq \mathbb{P}(U)$ are subsets of U. An *agreement test* is a procedure that inputs a family of assignments to the subsets $\{f_s : s \to \{0,1\}\}_{s\in S}$, and aims to check whether the assignments are consistent with a *global* assignment to the ground set $G : U \to \{0,1\}$. This is a classical setup in hardness of approximation (in

particular in the construction of PCPs), where one repeats a problem in parallel and needs to check that answers across repetitions are consistent.

Let us give the simplest setup first, slightly generalizing it later on. Let (U, S, Σ, D) be such that

1. U is a finite set ("Universe").
2. $S \subseteq \mathbb{P}(U)$ ("Sets").
3. Σ is another finite set ("Alphabet").
4. A distribution $s_1, s_2 \sim D$ such that $s_1, s_2 \in S$ ("Distribution").

An *ensemble of functions* in this context is a family $\mathcal{F} = \{f_s : s \to \Sigma\}_{s \in S}$. An ensemble is called *global* if there exists a function $G : U \to \Sigma$ such that for every $s \in S$ it holds that $f_s = G|_s$.

We wish to relate the following two quantities. The first quantity is the "agreement" of an ensemble. For $\eta \in [0, 1)$ we denote by

$$Agree_\eta(\mathcal{F}) := \mathop{\mathbb{P}}_{s_1, s_2, t \sim D} \left[f_{s_1}|_{s_1 \cap s_2} \overset{\eta}{\approx} f_{s_2}|_{s_1 \cap s_2} \right], \tag{5.21}$$

where the notation inside the probability means that there is a set $t' \subseteq s_1 \cap s_2$ of size $|t'| \geq (1 - \eta)|s_1 \cap s_2|$ such that $f_{s_1}|_{t'} = f_{s_1}|_{t'}$. We will also use $\overset{\eta}{\not\approx}$ to denote the negation of this.

The second quantity is the distance from global functions. For $\eta, \varepsilon > 0$ we say that $\mathcal{F}$ is (η, ε)-close to a global function if the exists some $G : U \to \Sigma$ such that

$$\mathop{\mathbb{P}}_{s} \left[f_s \overset{\eta}{\approx} G|_s \right] \geq 1 - \varepsilon.$$

The probability distribution in which we take $s \in S$ above is the marginal distribution of

D.[11]

We denote by

$$\mathrm{dist}_\eta(\mathcal{F}, Glob) = \min\{\varepsilon \geq 0 \mid \text{ there exists } G : U \to \Sigma \text{ such that } \mathcal{F} \text{ is } (\eta, \varepsilon)\text{-close to } G\}.$$

A *high acceptance agreement theorem* (also known as a "99% agreement theorem") is a theorem that states that if a set of local functions pass the test with almost perfect probability ("99%"), then the functions are close to being global. In other words, this is a theorem that states for a specific (U, S, Σ, D) that for every ensemble $\mathcal{F}$,

$$Agree_\eta(\mathcal{F}) = 1 - \varepsilon \quad \Longrightarrow \quad \mathrm{dist}_{\eta'}(\mathcal{F}, Glob) \leq \varepsilon'.$$

where $\varepsilon, \varepsilon', \eta, \eta' > 0$. Ideally, one wants such a statement for every ε, η small enough, with $\varepsilon' = O(\varepsilon)$ and $\eta' = O(\eta)$. But other parameter regimes are interesting in applications as well.

Agreement tests originated in low-degree testing results such as [333] and were later abstracted by [168] to roughly the setup above. They are an important component of Dinur's proof of the PCP Theorem [116]. An optimal high acceptance agreement theorem was proven by [115] on the complete complex. This was extended by [118] to another setup related to the complete complex we discuss below. Agreement theorems on sparse complexes such as high dimensional expanders are also known, pioneered by [124].

A *low acceptance agreement theorem* (also known as a "1% agreement theorem") deals with the regime where it is only assumed that the functions pass the agreement test with small (sometimes even sub-constant) probability ("1%"). This is a theorem of the form

$$Agree_\eta(\mathcal{F}) = \delta \geq \varepsilon \quad \Longrightarrow \quad \mathrm{dist}_{\eta'}(\mathcal{F}, Glob) \leq 1 - \mathrm{poly}(\delta).$$

[11] Formally, draw $(s, s') \sim D$, then s or s' with probability $1/2$ independently.

for some $(\varepsilon, \eta, \eta')$ (and some explicit polynomial in δ). We note here that in this test even when $\eta = 0$, an $\eta' > 0$ is sometimes unavoidable (see discussions in [120, 128]).

Agreement theorems in the low acceptance regime are typically harder to prove than in the high acceptance regime, since the promised structure on the family of functions is mild. However, in applications such that a parallel repetition theorem [215] and low-degree testing [31], they are often crucial. Dinur and Goldenberg [120] proved a low acceptance agreement theorem on the complete complex. Their proof was simplified by [215], who also derived a more sophisticated Z-test (see below) and gave the first sparse structure supporting such agreement tests using subspaces. Work by [128] gave an improved analysis of [215], asymptotically matching the lower bound of a random set of local functions. Their work heavily relied on reverse hypercontractivity, a strategy we follow below. Recently, [110] and [41] gave characterizations of high dimensional expanders that support agreement theorems, and [113, 40] constructed such complexes.

There are also other kinds of theorems about agreement tests that don't strictly fall into any of these categories. One example such as [177] list agreement theorem, that inspired the former line [110, 41, 113, 40].

Varying the Test

We present two simple extensions to the agreement test above. The first comes from sampling more than two sets. For example, in the Z-test by [215], three sets s_1, s_2, s_3 are sampled, and the test passes if $f_{s_1} = f_{s_2}$ *and* $f_{s_2} = f_{s_3}$ (equality is with respect to the respective intersections). See the formal definition of this test below.

Another extension to the definition of a test, mainly for technical reasons, is a test that checks $f_{s_1}|_t = f_{s_2}|_t$ only on some $t \subset s_1 \cap s_2$ which is also sampled by the tester. This is sometimes more convenient to analyze than the test checking agreement on the whole intersection. To summarize this discussion more formally, the distribution D in our setup samples $(\{s_i\}_{i=1}^k, \{t_{i,j}\}_{1 \le i < j \le k})$ such that $\{s_i\}_{i=1}^k \subseteq S$ are sets, and for every $i < j$,

$t_{i,j} \subseteq s_i \cap s_j$. The agreement test passes if for all $i < j$, $f_{s_i}|_{t_{i,j}} \overset{\eta}{\approx} f_{s_j}|_{t_{i,j}}$.

5.8.2 Agreement Testing for Subsets

We now move to the first setting of interest, agreement tests on simplicial complexes in the 99% regime. In particular we study the following setup due to [118]. Let $j < k < d$ and let X be a d-uniform simplicial complex. Let us denote by $SS(X, \Sigma, j, k, d) \coloneqq (U, S, \Sigma, D)$ where:

1. $U = X(j)$.
2. $S = \left\{ \binom{s}{j} \;\middle|\; s \in X(d) \right\}$.
3. The distribution $D = D_{d,k}$ samples s_1, s_2 according to the down-up walk. The 'test' set $t \subseteq s_1 \cap s_2$ is a randomly chosen k-face.

For simplicity of notation we sometimes refer to the sets as s instead of $\binom{s}{j}$, and also index functions by f_s for $s \in X(d)$ (and not $f_{\binom{s}{j}}$). However, in this context when writing $v \in s$, we mean a j-set instead of a vertex.

In other words, we study the following variant of the classical 'V-test' in this setting:

The V-test on $X(k)$:

1. Draw $t \in X(k)$, and $s_1, s_2 \in X_t(d)$ independently
2. Accept if:

$$f_{t \cup s_1}|_t \overset{\eta}{\approx} f_{t \cup s_2}|_t$$

When $j = 1$, this reduces to the heavily studied setting of (de-randomized) 'direct product testing'.

We are now ready state the main result of the subsection, our generalized 99%-tester:

Theorem 5.8.1. *Let j, d be integers, let $\gamma, \eta, c, \nu < 1$. Denote by $k = \gamma d$. Suppose that $j < k$, $\eta\lfloor \frac{k}{j} \rfloor \geq 192 \log \frac{16}{\eta}$ and $\nu < 1$. Let X be a d-uniform c-locally nice complex such that for every link of a set $v \in X(j)$, $\lambda(UD_{d-j,k-j}(X_v)) \leq \nu$. Then the following holds for the set system $SS(X, \Sigma, j, k, d)$; for every alphabet Σ, family $\mathcal{F}$, and $\varepsilon > 0$*

$$Agree_\eta(\mathcal{F}) \geq 1 - \varepsilon \quad \Longrightarrow \quad \mathrm{dist}_{8\frac{\eta+\varepsilon}{1-\nu}}(\mathcal{F}, g_{maj}) \leq \varepsilon^{O_{\gamma,c}(1)} + 2^{-\Omega_{\gamma,c}(d)} \tag{5.22}$$

where $g_{maj} \coloneqq \underset{s \in X(d): s \supset v}{plurality}\{f_s(v)\}$ is the plurality decoding.[12]

We note that every c-nice complex has $\lambda(UD_{d-j,k-j}) \leq \frac{k-j}{d-j}(1 + \frac{1}{c'(d-j)})^{(1-\gamma)(d-j)}$, so this is not an extra requirement in most cases (see [276, Theorem 3.5]).

We remark that for the setup of $j = 1$ there exist agreement theorems that achieve (exact) closeness $1 - O(\varepsilon)$ for high dimensional expanders [124, 109, 236]. Theorem 5.8.1 is incomparable to these results. On the one hand, in Theorem 5.8.1 the resulting 'globalness' is weaker than [124, 109, 236] where one gets a guarantee of $\mathrm{dist}_0(\mathcal{F}, g_{maj}) \leq O(\varepsilon)$. On the other hand, the initial assumption on the sets of local functions in the aforementioned results is $Agree_0(\mathcal{F}) \geq 1 - \varepsilon$, wheras this result assumes $Agree_\eta(\mathcal{F}) \geq 1 - \varepsilon$ for $\eta > 0$. This relaxed assumption is crucial in the low-soundness regime theorem we give in the next section.

For $j > 1$, to our knowledge Theorem 5.8.1 is the first agreement theorem beyond the complete complex, where Dinur, Filmus, and Harsha [118] showed

$$Agree_0(\mathcal{F}) \geq 1 - \varepsilon \Rightarrow \mathrm{dist}_0(\mathcal{F}, Glob) \geq 1 - O_j(\varepsilon). \tag{5.23}$$

It is plausible that, with more effort, Equation (5.23) could be proven directly on HDX. The advantage of our method in this sense lies in its generality and (relative) simplicity.

[12] Note that ties may be broken arbitrarily and the weighting is according to the (normalized) complex weights π_d.

Proof of Theorem 5.8.1. Given a family $\mathcal{F} = \{f_s\}$ with high agreement, we show it can be (approximately) decoded to the majority function $G = g_{\mathrm{maj}}$. We argue this can be inferred directly from reverse hypercontractivity and the following claim bounding the *average* probability of disagreement with G for a random pair $v \subseteq s$ drawn from $(X(j), X(d))$:

Claim 5.8.2. $\mathbb{P}_{v \subset s}[f_s(v) \neq G(v)] \leq \frac{\varepsilon+\eta}{1-\nu}$.

We first show this implies the result. Define the set of 'bad' d-faces which noticeably disagree with G as

$$B \coloneqq \left\{ s \in S \;\middle|\; \mathop{\mathbb{P}}_{v \in s}[f_s(v) \neq G(v)] > \frac{8(\varepsilon+\eta)}{1-\nu} \right\}.$$

It is enough to show $\mathbb{P}[B] \leq \max\{\varepsilon^{O_{\gamma,c}(1)}, \exp(-\Omega_{\gamma,c}(d))\}$.

Toward bounding B, define an intermediary set A of 'good' faces:

$$A \coloneqq \left\{ s \in S \;\middle|\; \mathop{\mathbb{P}}_{v \in s}[f_s(v) \neq G(v)] \leq \frac{4(\varepsilon+\eta)}{1-\nu} \right\}$$

and note that $\mathbb{P}[A] \geq \frac{1}{2}$ by Claim 5.8.2 and Markov's Inequality. The idea is to argue that, on the one hand, reverse hypercontractivity implies the number of edges (s_1, s_2) between A and B is at least some power of $\mathbb{P}[B]$, while, on the other hand, agreement implies the number of such edges is at most $O(\varepsilon)$ because they (typically) contain many $v \in s_1 \cap s_2$ that disagree and therefore fail the test.

More formally, denote the event that $(s_1, s_2) \in (A \times B)$ as $E_{A,B}$. Assuming $\mathbb{P}[B] \geq \exp(-c'd)$ (for c' depending on c, γ as in Theorem 5.6.4), reverse hypercontractivity lower bounds the measure of this event by

$$\mathbb{P}[E_{A,B}] \geq \left(\frac{\mathbb{P}[B]}{2}\right)^q$$

for some constant q depending only on c, γ. On the other hand, since the test rejects with

probability ε, we can upper bound the measure of $E_{A,B}$ by $O(\mathbb{P}[\text{Test rejects}]) = O(\varepsilon)$ we will obtain a bound of the form $\mathbb{P}[B] \leq O(\varepsilon^{1/q})$. To this end, note

$$\mathbb{P}[E_{A,B}] \leq \frac{\varepsilon}{\mathbb{P}[\text{Test rejects}|\ E_{A,B}]}.$$

It is therefore enough to argue the conditional rejection probability is at least some constant (say $\frac{1}{2}$). Toward this end, for a face $s \in X(d)$, define

$$T_s \coloneqq \{v \in s : f_s(v) \neq G(v)\}$$

to be the j-faces on which f_s disagrees with majority, and recall that an edge in our test is sampled by first picking $s_1 \in X(d)$, then $t \subset s_1$ in $X(\gamma d)$, then $s_2 \supset t$ conditionally. Given t, denote by $t(j) = \binom{t}{j}$ the set of sub j-faces in t, for shortness. We define two bad events outside of which the test rejects and prove they occur with vanishing probability.

1. **Mis-sampling A:** E_1 the event that t sees 'too many' $v \subset s_1$ that disagree with majority:
$$E_1 \coloneqq \left\{ \frac{|t(j) \cap T_{s_1}|}{\binom{k}{j}} \geq 5\frac{\varepsilon + \eta}{1+\gamma} \right\}$$
2. **Mis-sampling B:** E_2 the event that t sees 'too few' $v \subset s_2$ that disagree with majority:
$$E_2 \coloneqq \left\{ \frac{|t(j) \cap T_{s_2}|}{\binom{k}{j}} \leq 7\frac{\varepsilon + \eta}{1+\gamma} \right\}$$

Conditioning on $\neg E_1$ and $\neg E_2$ the test rejects since s_1 and s_2 disagree on a $2\frac{\varepsilon+\eta}{1-\nu}$-fraction of t. Thus by a union bound:

$$\mathbb{P}[\text{Test Rejects}|\ E_{A,B}] \geq 1 - \mathbb{P}[E_1|\ E_{A,B}] - \mathbb{P}[E_2|\ E_{A,B}]$$

and it is enough to show that each $\mathbb{P}[E_i|\ E_{A,B}] \leq \frac{1}{4}$

Bounding $\mathbf{E_1}$ and $\mathbf{E_2}$.

Events E_1 and E_2 occur with vanishing probability due to the sampling properties of the complete complex Δ (Claim 5.12.1), namely that for any $\alpha, \delta > 0$ and $j \le k$, the inclusion graph $(\Delta(k), \Delta(j))$ is an $(\alpha, \frac{4}{\alpha\delta}\exp(\frac{\delta^2}{12}\alpha \left\lfloor \frac{k}{j} \right\rfloor), \delta)$-sampler.

With this in mind, recall $t \in X(\gamma d)$ is drawn uniformly at random from s_1, so we may equivalently view t as being drawn from a γd-uniform complete complex Δ on the d vertices of s_1. By assumption, $s_1 \in A$, so $T_{s_1} \subset \Delta(j)$ is of measure at most $4\frac{\varepsilon+\eta}{\gamma}$, and sampling then implies

$$\mathbb{P}\left[\frac{|t(j) \cap T_{s_1}|}{\binom{k}{j}} > 5\frac{\varepsilon+\eta}{1+\gamma}\right] \le \frac{4(1+\gamma)}{\varepsilon+\eta}\exp\left(-\frac{\eta}{96}\left\lfloor\frac{k}{j}\right\rfloor\right) \le \frac{1}{4}.$$

Since we can equivalently sample an edge (s_1, s_2) by first picking s_2, then $t \subset s_2$, then $s_1 \supset t$, E_2 can be bounded using the lower tail of Δ in exactly the same fashion to get

$$\mathbb{P}\left[\frac{|t(j) \cap T_{s_2}|}{\binom{k}{j}} < 7\frac{\varepsilon+\eta}{1+\gamma}\right] \le \frac{4(1+\gamma)}{\varepsilon+\eta}\exp\left(-\frac{\eta}{192}\left\lfloor\frac{k}{j}\right\rfloor\right) \le \frac{1}{4}.$$

□

Proof of Claim 5.8.2. Fix $v \in X(j)$ and let $D_{d,k}^{(v)}$ denote the conditional distribution over $(s_1, t, s_2) \sim D_{d,k}$ such that $v \subset t$. Note that by construction, $D_{d,k}^{(v)}$ (after removing v from each face) is distributed exactly as the down-up walk $D_{d-j,k-j}$ within the link of v. Partition the vertices of this graph into sets $S_\sigma^{(v)} \coloneqq \{s|\ f_s(v) = \sigma\}_{\sigma\in\Sigma}$ by the alphabet value they assign v, and observe that since the marginals of $D_{d,k}^{(v)}$ are distributed as $X_v(d-j)$ we have

$$1 - \mathop{\mathbb{P}}_{v\subset s}[f_s(v) \ne G(v)] = \mathop{\mathbb{P}}_{v\subset s}[f_s(v) = G(v)] = \mathop{\mathbb{E}}_{v\sim X(j)}\left[\max_{\sigma\in\Sigma}\left\{\mathbb{P}\left[S_\sigma^{(v)}\right]\right\}\right],$$

where the inner probability is over the v-conditioned marginal.

The idea, which is fairly standard (see e.g. [124, Claim 5.2]), is now to relate this maximum partition size to the *local* disagreement of $\mathcal{F}$ using expansion. In particular, for each $v \in X(j)$ define the v-local disagreement of $\mathcal{F}$ as:

$$\varepsilon_v \coloneqq \mathop{\mathbb{P}}_{(s_1,s_2)\sim D^{(v)}_{d,k}}[f_{s_1}(v) \neq f_{s_2}(v)].$$

It is easy to see that $\mathbb{E}_v[\varepsilon_v] \leq \varepsilon + \eta$. This follows from observing that sampling a random $v \in X(j)$, then $(s_1, t, s_2) \sim D^{(v)}_{d,k}$ is equivalent to *first* sampling $(s_1, t, s_2) \sim D_{d,k}$, then sampling v uniformly from t. The bound is clear from the latter interpretation since the tester rejects with probability at most ε, and otherwise agrees on all but an η fraction of the j-sets in t. It is therefore enough to prove

$$\max_{\sigma\in\Sigma}\left\{\mathbb{P}\left[S^{(v)}_\sigma\right]\right\} \geq 1 - \frac{\varepsilon_v}{1-\nu},$$

as taking expectation on both sides gives the result.

Toward this end, define $E = E^{(v)}$ to be the set of edges $(s_1, s_2) \sim D^{(v)}_{d,k}$ that cross the alphabet partition. Since the marginal is distributed as the down-up walk in v's link, this graph is a ν-spectral expander by assumption. Writing $S_\sigma = S^{(v)}_\sigma$ for simplicity of notation, we have

$$\begin{aligned}
\varepsilon_v = \mathbb{P}[E] &= \sum_{\sigma\in\Sigma} \mathop{\mathbb{P}}_{(s_1,s_2)}[s_1 \in S_\sigma, s_2 \notin S_\sigma] \\
&= \sum_{\sigma\in\Sigma} \langle \mathbf{1}_{S_\sigma}, D^{(v)}_{d,k}(\mathbf{1} - \mathbf{1}_{S_\sigma})\rangle \\
&= 1 - \sum_{\sigma\in\Sigma} \langle \mathbf{1}_{S_\sigma}, D^{(v)}_{d,k}\mathbf{1}_{S_\sigma}\rangle \qquad (5.24)
\end{aligned}$$

where $\mathbf{1}_{S_\sigma}$ is the indicator of S_σ and $\mathbf{1}$ is the all-ones function and the final equality is

from by stochasticity of $D^{(v)}_{d,k}$. Let $f_\sigma = \mathbf{1}_{S_\sigma} - \mathbb{P}[S_\sigma]\,\mathbf{1}$ be the projection of $\mathbf{1}_{S_\sigma}$ onto $\mathbf{1}^\perp$. By orthogonality it holds that $\|f_\sigma\|^2 = \|\mathbf{1}_{S_\sigma}\|^2 - \mathbb{P}[S_\sigma]^2\,\|\mathbf{1}\|^2 = \mathbb{P}[S_\sigma] - \mathbb{P}[S_\sigma]^2$. Therefore exploiting the expansion of $D^{(v)}_{d,k}$ we have

$$\begin{aligned}\langle \mathbf{1}_{S_\sigma}, D^{(v)}_{d,k}\mathbf{1}_{S_\sigma}\rangle &= \mathbb{P}[S_\sigma]^2\,\langle \mathbf{1},\mathbf{1}\rangle + \langle f_\sigma, D^{(v)}_{d,k} f_\sigma\rangle \\ &\le \mathbb{P}[S_\sigma]^2 + \nu\|f_\sigma\|^2 \\ &= \nu\,\mathbb{P}[S_\sigma] + (1-\nu)\,\mathbb{P}[S_\sigma]^2 .\end{aligned}$$

Inserting this back to (5.24) we have that

$$\begin{aligned}\varepsilon_v &\ge 1 - \sum_{\sigma\in\Sigma}\left(\nu\,\mathbb{P}[S_\sigma] + (1-\nu)\,\mathbb{P}[S_\sigma]^2\right) \\ &= (1-\nu)\left(1 - \sum_{\sigma\in\Sigma}\mathbb{P}[S_\sigma]^2\right) \\ &\ge (1-\nu)\left(1 - \max_{\sigma\in\Sigma}\{\mathbb{P}[S_\sigma]\}\right).\end{aligned}$$

Re-arranging gives the desired result. □

Testing all the intersection

We note that one could also consider the standard V-test where s and s' are still drawn from the down-up walk, but we test agreement over $t = s_1 \cap s_2$ (whereas in the definition of the test s_1 and s_2 may have some elements in their intersection not in t). The success in these tests are related by the following claim.

Claim 5.8.3. Let $\eta > 0$. Let D_1 be the test above. Let D_2 be the test with a similar distribution to D_1, with the distinction that $t = s_1 \cap s_2$. Then for any $\mathcal{F}$, $Agree_{\eta,D_2}(\mathcal{F}) \ge 1-\varepsilon$ implies that $Agree_{2\eta,D_1}(\mathcal{F}) \ge 1 - \varepsilon - 2^{-\Omega_\gamma(\eta^2 k/j)}$.

Proof of Claim 5.8.3. Let s_1, s_2 be a pair supported by D_2 such that $\mathcal{F}$ passes the D_2-test on the pair. Let A be the set of j-faces contained in $s_1 \cap s_2$ on which f_{s_1}, f_{s_2} disagree

on. By assumption $|A| \leq \eta|s_1 \cap s_2|$. By the sampling properties of the complete complex Claim 5.12.1 the fraction of $t \subseteq s_1 \cap s_2$ such that $|A \cap t| \geq 2\eta|t|$ is $2^{-\Omega_\gamma(\eta^2 k/j)}$. Thus the probability of *failing* the D_1 test is at most the probability of failing the D_2 test plus $2^{-\Omega_\gamma(\eta^2 k/j)}$. □

Remark 5.8.4. When $\eta \lesssim \varepsilon$, the 'additive' error $\exp(-\Omega(d))$ can be removed from Theorem 5.8.1 by interpolating the result with a standard averaging argument over Claim 5.8.2. This technique stops working as soon as η is much larger than ε. While the additional $\exp(-\Omega(d))$ error is not harmful in most applications, it is nevertheless interesting to ask whether the term can be removed for general η.

Agreement Testing for the non-Lazy Down-Up Walk

Using a similar technique, we can also prove an agreement test for the following distribution based on the swap walks. Define the test distribution $(s_1, s_2, t) \sim A_{d,k}$ by the following procedure:

1. Sample $t \in X(k)$.
2. Sample $r_1, r_2 \in X(d-k)$ in the swap-walk inside X_t.
3. Output $s_i = t \cup r_i$ and t.

The resulting edge distribution over (s_1, s_2) is sometimes called the 'partial' swap walk [9]. Note this walk has the property that the intersection is fixed. The following claim shows that when the local swap walks expand, an agreement theorem for this walk follows as well.

Claim 5.8.5. Let j, d be integers, let $\gamma, \eta, c, \lambda, \nu < 1$. Let $k = \gamma d$ and suppose $j < k$, $\eta\lfloor \frac{k}{j} \rfloor \geq 192 \log \frac{16}{\eta}$ and $\nu < 1$. Let X be a $(2d-k)$-uniform c-locally nice complex such that for every $v \in X(j)$, $\lambda(A_{d-j,k-j}(X_v)) \leq \nu$. Suppose additionally that for every $t \in X(k)$,

$S_{d-k,d-k}(X_t) \leq \lambda$. Then the following holds for the set system $SS(X, \Sigma, j, k, d)$; for every alphabet Σ, family $\mathcal{F}$, and $\varepsilon > 0$

$$Agree_\eta(\mathcal{F}) \geq 1 - \varepsilon \quad \Longrightarrow \quad \mathrm{dist}_{8\frac{\eta+\varepsilon}{1-\nu}}(\mathcal{F}, g_{\mathrm{maj}}) \leq \varepsilon^{O_{\gamma,C}(1)} + \lambda^{O_{\gamma,C}(1)}. \tag{5.25}$$

where $g_{\mathrm{maj}} \coloneqq \underset{s \in X(d): s \supset v}{\mathrm{plurality}}\{f_s(v)\}$ is the plurality decoding.

Proof of Claim 5.8.5. As in the proof of Theorem 5.8.1, we define

$$B \coloneqq \left\{ s \in S \;\middle|\; \underset{v \in s}{\mathbb{P}}\left[f_s(v) \neq G(v)\right] > \frac{8(\varepsilon+\eta)}{1-\nu} \right\}.$$

and

$$A \coloneqq \left\{ s \in S \;\middle|\; \underset{v \in s}{\mathbb{P}}\left[f_s(v) \neq G(v)\right] \leq \frac{4(\varepsilon+\eta)}{1-\nu} \right\}.$$

As in the proof of Theorem 5.8.1, $\mathbb{P}[A] \geq \frac{1}{2}$ by an argument similar to Claim 5.8.2 and Markov's inequality. More precisely, we need to repeat the proof of Claim 5.8.2, only instead of using the down-up walk conditioned on v, we use $A_{d,k}$ conditioned on $v \in t$. This is just $A_{d-j,k-j}$ inside X_v.

In addition, as in the proof of Theorem 5.8.1, it also holds that the relative fraction of edges crossing between A and B is small, i.e. $\mathbb{P}\left[E_{A,B}\right] \leq 2\varepsilon$.

Our goal is to show that $\mathbb{P}\left[B\right] \leq \max\{\varepsilon^{\Omega_{\gamma,c}(1)}, \exp(-\Omega_{\gamma,c}(d)), \lambda^{\Omega_{\gamma,c}(1)}\}$.

Assume otherwise, and that in particular reverse hypercontractivity for indicators applies for sets of size $\mathbb{P}\left[B\right]$. By Theorem 5.6.4

$$\underset{t \in X(k)}{\mathbb{P}}\left[\underset{s \supseteq t}{\mathbb{P}}\left[A\right] \cdot \underset{s \supseteq t}{\mathbb{P}}\left[B\right] \geq \left(\mathbb{P}\left[A\right]\mathbb{P}\left[B\right]\right)^{O_{\gamma,c}(1)}\right] \geq \mathbb{P}\left[A\right]\mathbb{P}\left[B\right]^{O_{\gamma,c}(1)}.$$

Let us denote by G the set of t satisfying the above condition. Suppose we sampled $t \in G$ in the first step of our agreement distribution. Recall that now we sample and edge by going in the swap walk in the link of t. The probability we sampled an edge between A

and B is therefore

$$\mathop{\mathbb{P}}_{r_1,r_2\sim S(X_t)}[t\cup r_1\in A, t\cup r_2\in B] \geq (\mathbb{P}[A]\,\mathbb{P}[B])^{O_{\gamma,c}(1)} - \lambda = p\cdot\mathbb{P}[B]^{O_{\gamma,c}(1)} - \lambda$$

by the expansion of the swap walk and the expander mixing lemma (for some constant p). Thus if $p\,\mathbb{P}[B]^{O_{\gamma,c}(1)} \geq 2\lambda$ then we have that

$$\Omega(\mathbb{P}[B]^{O_{c,\gamma}(1)}) \leq \mathbb{P}[G]\,\mathbb{P}[s_1\in A, s_2\in B \mid \ G] \leq \mathbb{P}[E_{A,B}] \leq 4\varepsilon$$

and the claim follows. □

5.8.3 Low Soundness and the Z-Test

In this section we will apply Theorem 5.8.1 and reverse hypercontractivity to prove soundness of [215]'s Z-test in the 1%-regime under certain stronger assumptions on the complex. Before moving on, we briefly comment on the notation in this subsection. Unlike before, we will use the convention that s,t are $\frac{k}{2}$-sets (and not k-sets as in the previous subsections). We will typically denote k-sets by a capital letter instead. We first recall their test:

The Z-test on $X(k)$:

1. Draw $t\in X\left(\frac{k}{2}\right)$, and $s,s'\in X_t\left(\frac{k}{2}\right)$ independently
2. Draw $s''\sim X_{s'}\left(\frac{k}{2}\right)$
3. Accept if:

$$f_{t\cup s} = f_{t\cup s'} \quad \text{and} \quad f_{s'\cup t} = f_{s'\cup s''}$$

We prove soundness for complexes which are 'λ-global', a slight strengthening of ℓ_∞-independence [232]:

Definition 5.8.6 (Global Complex)**.** A k-uniform complex X is called λ-global if $\forall t \in X\left(\frac{k}{2}\right)$:

$$\|S_t - \pi_{\frac{k}{2}}\|_{TV} \leq \lambda$$

where S_t is the distribution over neighbors of t in $S_{\frac{k}{2},\frac{k}{2}}$ and $\pi_{\frac{k}{2}}$ is the induced distribution on $X\left(\frac{k}{2}\right)$.

Globality of X essentially promises that every $\frac{k}{2}$-set in X 'sees' most other $\frac{k}{2}$-sets in X, and is equivalent to bounding the matrix infinity norm of the swap walk from its stationary operator $\frac{1}{2}\|S_{\frac{k}{2},\frac{k}{2}} - \Pi_{\frac{k}{2}}\|_\infty \leq \lambda$. In fact, we note we really only need the former weaker condition, which is an *average* instead of worst-case bound on the ℓ_1-norm of rows of $S_{\frac{k}{2},\frac{k}{2}} - \Pi_{\frac{k}{2}}$. However since bounding the infinity norm is somewhat more standard (and we do not have any examples of complexes satisfying an average but not worst-case bound), we stick with this stricter definition in what follows.

Theorem 5.8.7. *$\forall \lambda, \eta > 0$ and large enough k, let X be a k-uniform, λ-global, c-locally nice complex. Then for any $\delta \in (8\lambda + e^{-\Omega_c(\eta k)}, \frac{1}{8})$ if $Agree_0^Z(\mathcal{F}) \geq \delta$:*

$$\exists g : X(0) \to \Sigma, \quad \mathop{\mathbb{P}}_{s \in X(k)}[f_s \overset{\eta}{\approx} g(s)] \geq \delta/8.$$

A couple of remarks are in order. First, we mention that it is well known that no tester can do better than $\exp(-\Omega(k))$ soundness [120, 128], so our bound is essentially optimal for small λ. Second, we remark that many (dense) complexes of interest are indeed λ-global HDX, including basic examples such as the complete complex and random complexes, and more involved examples such as skeletons of various Ising models, independent sets, list-colorings, or more generally essentially any of the myriad complexes studied in the approximate sampling literature (see e.g. [24, 286, 96, 146, 66] among many others). Theorem 5.8.7 shows the Z-test is sound on these families in a blackbox fashion, even up to $\exp(-\eta k)$ for the appropriate parameters. We elaborate on these examples further in

the end of this section.

The proof of Theorem 5.8.7 largely follows the strategy of [215, 128]. The main difference lies in replacing certain ad-hoc arguments for the complete complex with more general methods based on reverse hypercontractivity and globality.

A Local Agreement Theorem

The core of Theorem 5.8.7 is a *local* structure theorem for the V-test on high dimensional expanders. Stating this requires a few definitions. We remark that these are standard notions in the literature, and we follow the notation of [128]. We first define *restrictions*.

Definition 5.8.8 (Restriction)**.** A *restriction* is a pair $\tau = (t, \sigma)$ such that $t \in X\left(\frac{k}{2}\right)$ and $\sigma : t \to \Sigma$ is an assignment to t.

Given a family of functions as above, there is a natural distribution over restrictions that comes from sampling a random function and a random restriction of it. More formally:

1. Sample $t \in X\left(\frac{k}{2}\right)$ and $s \in X_t\left(\frac{k}{2}\right)$.
2. Output $\tau = (t, f_{s\cup t}|_t)$.

Whenever we mention a distribution over restrictions τ we always mean $\tau = (t, f_{s\cup t}|_t)$ sampled as above. Given such a restriction τ, it will also be useful to have notation for the faces in X that are *consistent* with τ.

Definition 5.8.9 (Consistent Strings)**.** Given a restriction $\tau = (t, \sigma)$, denote by $\mathcal{V}_\tau$ the set of faces $s \in X_t\left(\frac{k}{2}\right)$ consistent with τ:

$$\mathcal{V}_\tau \coloneqq \left\{ s \in X_t\left(\frac{k}{2}\right) : f_{s\cup t}|_t = \sigma \right\}$$

A restriction is called *good* if it has many consistent faces.

Definition 5.8.10 (Good Restriction)**.** A restriction $\tau = (t, \sigma)$ is called δ-good if $\mathbb{P}_{s \in X_t}[\mathcal{V}_\tau] \geq \frac{\delta}{2}$.

When δ is clear from context, we just call such restrictions good. Finally, a restriction is called 'DP' if its consistent strings agree with a global function ('DP' is for 'direct product').

Definition 5.8.11 (DP-Restriction)**.** A restriction $\tau = (t, \sigma)$ is called (η, δ)-DP if there exists $g_\tau : X_t(1) \to \Sigma$ such that:

$$\mathop{\mathbb{P}}_{s \sim X_t}\left[f_{t \cup s} \overset{\eta/4}{\not\approx} g_\tau \;\middle|\; s \in \mathcal{V}_\tau \right] \leq \delta^2$$

As before, when η, δ are clear from context we just write 'DP'. We can now state the local agreement theorem for high dimensional expanders.

Theorem 5.8.12 (The Local Agreement Theorem)**.** *For any $\eta > 0$ and k sufficiently large, let X be a k-uniform c-locally nice complex. Then for all $\delta > e^{-\Omega(\eta k)}$, if $Agree_0^V(\mathcal{F}) \geq \delta$:*

1. *Many restrictions are good:*

$$\mathop{\mathbb{P}}_{\tau}[\tau \text{ is good}] \geq \frac{\delta}{2}$$

2. *Almost all 'good' restrictions are DP:*

$$\mathop{\mathbb{P}}_{\tau}[\tau \text{ is DP} \mid \tau \text{ is good}] \geq 1 - \delta^2$$

We remark that while Theorem 5.8.12 does not give a true agreement tester in its own right, it is a powerful tool independent of Theorem 5.8.7. It roughly states that good agreement implies a non-negligible fraction of good restrictions, *and* that if f_s agrees with a restriction of a face t it is almost certainly a restriction of a global assignment on the link X_t. This tool is a critical part of many of the previous works on agreement testing [120, 115, 215, 128].

We prove Theorem 5.8.12 in the next section. We first show it implies the main theorem under the additional assumption of λ-globality. The argument is a simple adaptation of [215].

Proof of Theorem 5.8.7. As in [215], we replace the standard Z-test in Section 5.8.3 with a related proxy test where we first sample (t, s) as before, but then draw (s', s'') *independent of* (t, s). If $s' \in X_t$ we continue as in the Z-test, and if $s' \notin X_t$ we accept. The probability of passing the new test is only greater or equal to the probability of passing the original test, but by λ-globalness, the swap walk starting at t is λ-close to the stationary distribution hence the probability that s' misses the link of t is at most λ.

With this in mind, let τ denote the restriction $(t, f_{s\cup t}|_t)$ and observe that if τ is not good, the new Z-test accepts with probability at most $\delta/2 + \lambda \leq \delta$. Thus it must be the case that conditioned on τ being good, the test still accepts with probability at least δ. Condition on this event, and observe that by Theorem 5.8.12, τ is also a DP-restriction except with probability at most $\delta^2 \leq \delta/8$ (recall that $\delta < \frac{1}{8}$).

Fix such a DP restriction τ, and denote by B the set of faces in $\mathcal{V}_\tau$ on which $\mathcal{F}$ is $\eta/4$-close to the DP function g_τ promised by Theorem 5.8.12 (if there is more than one, take any such function). Assume toward contradiction that $\mathcal{F}$ is not $(\eta, \delta/8)$-close to g_τ. Let $S = s' \cup s''$ and let H denote the event that $f_S \overset{\eta}{\approx} g_\tau$.[13] By assumption $\mathbb{P}[H] \leq \frac{\delta}{8}$, because S is drawn as in the distribution of $X(k)$. We define four events, outside which the test rejects:

1. E_1: The event that $s' \notin X_t$,
2. E_2: The event that $s' \in \mathcal{V}_\tau \setminus B$,
3. E_3: The event that $S \in H$,

[13]Formally here we extend g_τ to a function on $X(1)$ by arbitrarily assigning values to any vertex in $X(1) \setminus X_t(1)$.

4. E_4: The event that $S \notin H$, and $f_S|_{s'} \overset{\eta/2}{\approx} g_\tau|_{s'}$.

We first argue the test rejects if none of the E_i occurs. In particular, note by $\overline{E_1} \cap \overline{E_2}$, s' must either be in B, or in $X_t \setminus \mathcal{V}_\tau$. In the latter case, the first query of the proxy Z-test rejects since $s' \in X_t$ but $f_{t \cup s'} \neq f_{t \cup s}$. On the other hand, in the former case $s' \in B$ and therefore $f_{t \cup s'}|_{s'} \overset{\eta/4}{\approx} g_\tau|_{s'}$. However, by $\overline{E_3} \cap \overline{E_4}$, we have $S \notin H$ and, moreover, that the restriction to s' has disagreement at least $f_S|_{s'} \overset{\frac{\eta}{2}}{\not\approx} g_\tau|_{s'}$. Thus there must be a vertex on which $f_{t \cup s'}$ and f_S disagree and the second query rejects.

Finally by a union bound it is enough to prove $\sum \mathbb{P}[E_i] \leq \delta/2$. The first event occurs with probability at most $\delta/8$ by λ-globality where $\lambda < \frac{\delta}{8}$. The second event occurs with probability at most $\delta^2 \leq \delta/8$ by Theorem 5.8.12. The third event occurs with probability at most $\delta/8$ by assumption. The final event occurs with probability at most $e^{-\Omega(\eta k)} \leq \delta/8$ by Chernoff, since f_S η-disagrees with g_τ and s' is a random subset, which completes the proof. □

Proof of the Local Agreement Theorem

Let us begin by proving the first property in Theorem 5.8.12.

Claim 5.8.13. Under the assumptions of Theorem 5.8.12, $\mathbb{P}_\tau[\tau \text{ is good}] \geq \frac{\delta}{2}$.

Proof. Let $p = \mathbb{P}_\tau[\tau \text{ is good}]$. Since $Agree_0^V(\mathcal{F}) \geq \delta$ and we have

$$\begin{aligned}
\delta &\leq Agree_0^V(\mathcal{F}) \\
&= p \mathop{\mathbb{P}}_{s,t,s'}[f_{t \cup s} = f_{t \cup s'} \mid (t, f|_{s \cup t}) \text{ is good}] \\
&\quad + (1-p) \mathop{\mathbb{P}}_{s,t,s'}[f_{t \cup s} = f_{t \cup s'} \mid (t, f|_{s \cup t}) \text{ is not good}].
\end{aligned}$$

By definition $\mathbb{P}_{s,t,s'}[f_{t \cup s} = f_{t \cup s'} \mid (t, f|_{s \cup t}) \text{ is not good}] \leq \frac{\delta}{2}$ so $\delta \leq p + \frac{\delta}{2}$ which implies the claim. □

The rest of this section is dedicated to proving the second property in Theorem 5.8.12, which is more complicated. The proof has two main steps. First, we will need a variant of the notion of an *excellent restriction*, that appeared in [120, 215, 128]. Roughly, an excellent restriction is a restriction (t, σ) so that if $s_1, s_2 \in \mathcal{V}_t$, the probability that $f_{t\cup s_1}|_{s_1\cap s_2} \approx f_{t\cup s_2}|_{s_1\cap s_2}$ is $1 - \text{poly}(\delta)$. The probability in which we choose $s_1, s_2 \in X_t$ is stated below. We show most good restrictions are in fact excellent.

Second, we show that excellent restrictions are DP. Here we use a variant of the "smoothing" technique introduced by [128] in order to get an ensemble of local functions $\left\{ \tilde{f}_s : s \to \Sigma \;\middle|\; s \in X_t\left(\frac{k}{2}\right) \right\}$ that agree with high probability $1 - \text{poly}(\delta)$ and use our 99%-tester (Theorem 5.8.1) to show the restriction is DP.

Formally we require some additional notation to properly define these notions. First, denote the $\frac{k}{4}$-step down-up walk in the link of t as N_t and denote by $(s, t', s') \sim N_t$ the distribution where s, s' are chosen according to N_t and t' is the intermediate set chosen in the first down step (we note that $t' \subseteq s \cap s'$ but equality doesn't necessarily hold). Second, let N_t^3 denote three steps of N_t. Let $(s_1, t_1, s_2, t_2, s_3, t_3, s_4) \sim N_t^3$ be the full walk sampled. Finally, let $(s_1, t_2, t', s_4) \sim N_t^3$ be such that s_1, t_2, s_4 are the marginals as before, and $t' = t_1 \cap t_2 \cap t_3$ the vertices that are in all sets in the sampled walk.

Fix some sufficiently small constant $c_1 > 0$. We can now define excellence.

Definition 5.8.14 (Excellent Restriction)**.** A good restriction $\tau = (t, \sigma)$ is called excellent if:

1. $\mathbb{P}_{(s,t',s')\sim N_t}\left[(s, s' \in \mathcal{V}_\tau) \wedge \left(f_s|_{t'} \overset{\eta/1500}{\not\approx} f_{s'}|_{t'}\right)\right] \leq e^{-c_1\eta k}$ and,
2. $\mathbb{P}_{(s_1,t_2,t',s_4)\sim N_t^3}\left[(s, s' \in \mathcal{V}_\tau) \wedge \left(f_{s_1}|_{t'} \overset{\eta/1500}{\not\approx} f_{s_4}|_{t'}\right)\right] \leq e^{-c_1\eta k}$.

where we've abused notation and written f_s as shorthand for $f_{s\cup t}$.

We note that this definition is slightly cumbersome due to the fact that we don't measure the agreement of $f_s, f_{s'}$ (respectively f_{s_1}, f_{s_4}) on the intersection of $s \cap s'$ (re-

spectively $s_1 \cap s_4$), but instead on a fixed subset. This is for technical reasons and we encourage the readers to think of the case of N_t where $t' = s \cap s'$ (and analogously for N_t^3).

For notational convenience, let $\mu \coloneqq e^{-c_1 \eta k}$ and $\eta' = \eta/1500$. The following lemma (or variants thereof) is fairly standard [213, 215, 128]:

Lemma 5.8.15 (Good restrictions are excellent)**.** *There exists $c_2 > 0$ such that*

$$\mathop{\mathbb{P}}_{\tau=(t,f_{s\cup t}|_t)}[\tau \text{ is good but not excellent}] \leq e^{-c_2\eta k}.$$

Proof. We only prove that N_t^3 case, which is somewhat less standard. The N_t proof is analogous. It is sufficient to show that over a random restriction $\tau = (t, f_{s_0\cup t}|_t)$:

$$\mathop{\mathbb{P}}_{\tau=(t,f_{s_0\cup t}|_t)}\left[\mathop{\mathbb{P}}_{(s_1,t_2,t',s_4)\sim N_t^3}\left[s_1, s_4 \in \mathcal{V}_t \wedge f_{s_1}|_{t'} \overset{\eta'}{\not\approx} f_{s_4}|_{t'}\right] > e^{-c\eta k}\right] \leq e^{-\Omega(\eta k)} \tag{5.26}$$

where s_1, t_2, t', s_4 is chosen independently of s_0. In particular if this holds then, by claim Claim 5.8.13, conditioning on being good can increase the probability of this event to hold to at most $\frac{2e^{-\Omega(\eta k)}}{\delta} \leq e^{-c_2\eta k}$ for the right choice of constants.

The key to Equation (5.26) is to observe that this process can be sampled by the following equivalent method: sample $T \in X\left(\frac{k}{2} + \frac{k}{4}\right)$, and randomly partition it into $T = t \cup t_2$ so that $|t| = \frac{k}{2}$ and $|t_2| = \frac{k}{4}$. The faces s_1, t_2, t_3, and s_4 are now drawn independently from the "up-down-up" walk $U_{\frac{k}{4},\frac{k}{2}} D_{\frac{k}{2},\frac{k}{4}} U_{\frac{k}{4},\frac{k}{2}}$ within X_t starting from t_2.

Observe that $|t'| \geq \frac{k}{32}$ with probability $e^{-\Omega(k)}$ by a standard Chernoff bound. Conditioning on this event, observe that if $s_1, s_4 \in \mathcal{V}_\tau$ and $f_{s_1}|_{t'} \overset{\eta'}{\not\approx} f_{s_4}|_{t'}$, then it must be the case that

1. t_2 has at least $\frac{\eta' k}{32}$ elements on which $f_{s_1} \neq f_{s_4}$
2. t has no elements on which $f_{s_1}|_t \neq f_{s_4}|_t$.

However, t and t_2 are a uniformly random partition of T, so such a split occurs with

probability at most $e^{-\Omega(\eta k)}$ by Chernoff as desired. $\square$

To proving excellent restrictions are DP, we'll use a variant of Dinur and Livni-Navon's 'smoothing' operation that spreads the consistent strings $\mathcal{V}_\tau$ over the entire link. Instead of using the noise operator as in their work, we use the down-up walk which avoids a number of technical complications.

Definition 5.8.16 (Smoothed Assignment)**.** For every $\tau = (t, f_{t \cup s_0}|_t)$, define the smoothed assignment $\widetilde{\mathcal{F}}_\tau = \{\tilde{f}_s\}_{s \in X_t}$ as

$$\tilde{f}_s(v) = \underset{(t',s') \sim N_t(s): s' \in \mathcal{V}_\tau, t' \ni v}{\text{Plurality}} \{f_{s'}(v)\}.$$

We break ties arbitrarily. If the plurality is not well defined (i.e. there are no such $s' \in \mathcal{V}_\tau$ that contain v) we define it to be $\tilde{f}_s(v) = \perp$.

The main idea is to show that on excellent restrictions, $\widetilde{\mathcal{F}}_\tau$ is 1) highly consistent with the original family $\mathcal{F}$ on $\mathcal{V}_\tau$, and 2) has very high agreement on the *entire* link. This reduces the problem to an (approximate) 99%-regime test within the link of t, which we can solve via our tester from the previous section. More formally, the following lemmata suffice to prove Theorem 5.8.12.

Lemma 5.8.17. *Every excellent $\tau = (t, \sigma)$ satisfies:*

$$\mathbb{P}_{s \sim X_t} \left[f_s \overset{10\eta'}{\not\approx} \tilde{f}_s \;\middle|\; s \in \mathcal{V}_\tau \right] \leq \frac{\delta^2}{2},$$

that is, $\mathcal{F}$ and $\widetilde{\mathcal{F}}$ are close on $\mathcal{V}_\tau$ (note $f_s \overset{10\eta'}{\not\approx} \tilde{f}_s$ is over just the vertices of s, not $s \cup t$)

Lemma 5.8.18. *There exists $c_3 > 0$ such that every excellent $\tau = (t, \sigma)$ satisfies*

$$\mathbb{P}_{(s,t',s') \sim N_t} \left[\tilde{f}_s|_{t'} \overset{10\eta'}{\not\approx} \tilde{f}_{s'}|_{t'} \right] \leq e^{-c_3 \eta k},$$

that is, $\widetilde{\mathcal{F}}$ is highly agreeing on X_t.

Proof of Theorem 5.8.12. The first property is proved in Claim 5.8.13. Toward the second, we will prove that every excellent τ is a DP-restriction. Then by Lemma 5.8.15 we have that

$$\mathop{\mathbb{P}}_{\tau}[\tau \text{ is DP} \mid \ \tau \text{ is good}] \geq \mathop{\mathbb{P}}_{\tau}[\tau \text{ is excellent} \mid \ \tau \text{ is good}] \geq 1 - \delta^2$$

where we've used the assumption that $\delta \geq e^{-\Omega(\eta k)}$. Fix any excellent $\tau = (t, \sigma)$. One can check these satisfy the conditions of Theorem 5.8.1 with $\nu = 1/2$. Together with Lemma 5.8.18 and Claim 5.8.3 that translates between the agreement guarantee in Lemma 5.8.18 to that of Theorem 5.8.1, we have that there exists a global function $g : X_t(0) \to \Sigma$ such that

$$\mathop{\mathbb{P}}_{s \sim X_t}[\tilde{f}_s \overset{320\eta'}{\not\approx} g|_s] \leq e^{-c_4 \eta k}$$

where c_4 is some constant depending on c_3, η and the c-local niceness of X. We note that after conditioning over $s \in \mathcal{V}_\tau$, we still have that this is at most $\frac{2e^{-c_4\eta k}}{\delta} \leq \frac{\delta^2}{2}$, where this holds by assumption that δ is large enough.

Combining this with Lemma 5.8.17, we then have

$$\begin{aligned}\mathop{\mathbb{P}}_{s \sim X_t}\left[f_s \overset{\eta/4}{\not\approx} g(s) \,\middle|\, s \in \mathcal{V}_t\right] &\leq \mathop{\mathbb{P}}_{s \sim X_t}\left[f_s \overset{10\eta'}{\not\approx} \tilde{f}_s \,\middle|\, s \in \mathcal{V}_\tau\right] + \mathop{\mathbb{P}}_{s \sim X_t}\left[\tilde{f}_s \overset{320\eta'}{\not\approx} g(s) \,\middle|\, s \in \mathcal{V}_t\right] \\ &\leq \delta^2\end{aligned}$$

as desired. □

It is left to prove the key lemmata. This is the main place we use reverse hypercontractivity (other than the application of Theorem 5.8.1 in the proof of Theorem 5.8.12). The proofs for both properties follow the strategy of [128] adapted to our setting.

Proof of Lemma 5.8.17. We define two families of 'bad' sets outside of which f and $\tilde{f}$ approximately agree. First, we look at the set of 'lonely' strings that don't sufficiently see

$\mathcal{V}_\tau$.

$$L \coloneqq \left\{ s \in X_t : \mathop{\mathbb{P}}_{s' \sim N_t(s)}[s' \in \mathcal{V}_\tau] \leq \delta' \right\}$$

for $\delta' = e^{-c_5 \eta k}$ for some $c_5 > 0$ sufficiently small. We will assume below that $\delta' \leq \delta^4$ (which is possible due to the assumptions on δ). We record the following properties of L that we prove at the end of the subsection.

Claim 5.8.19. There exists a sufficiently small constant $c_5 > 0$ such that the following holds for $\delta' = e^{-c_5 \eta k}$.

1. $\mathbb{P}_{s \in X_t}[L] \leq \frac{\delta'^2}{2}$.
2. For any $s \notin L$, the fraction of vertices $v \in s$ s.t. $\mathbb{P}_{(t',s') \sim N_t(s)}[s' \in \mathcal{V}_\tau \mid t' \ni v] \leq \delta'/3$ is at most η'.

Second, we look at the family of sets within $\mathcal{V}_\tau$ with strong disagreement:

$$B \coloneqq \left\{ s \in \mathcal{V}_\tau : \mathop{\mathbb{P}}_{(t',s') \sim N_t(s)} \left[s' \in \mathcal{V}_\tau \wedge f_s|_{t'} \overset{\eta'}{\not\approx} f_{s'}|_{t'} \right] \geq \frac{\delta'}{30} \right\}.$$

We claim that both L and B have measure at most $\delta^2/2$ within $\mathcal{V}_\tau$. The former follows from the first item of Claim 5.8.19 since $\mathbb{P}[L|\mathcal{V}_\tau] \leq \frac{\delta'^2}{\delta} \leq \frac{\delta^2}{2}$. For the latter we have $\mathbb{P}[B] \leq \frac{30\mu}{\delta'} \leq \frac{\delta^2}{2}$ from Markov's inequality and excellence of τ (here we also need that the c_5 in the definition of δ' is sufficiently smaller than the c_1 in the definition of μ).

It is left to observe that $f_s \overset{10\eta'}{\approx} \tilde{f}_s$ for any face $s \in \mathcal{V}_\tau \setminus (L \cup B)$. Fix such an s, and define D_0 to be the set of vertices on which f and $\tilde{f}$ disagree:

$$D_0 \coloneqq \left\{ v \in s : f_s(v) \neq \tilde{f}_s(v) \right\}.$$

We need to argue $|D_0| \leq 10\eta'|s|$. Assume toward contradiction otherwise. Let D' denote the vertices in s such that $\mathbb{P}_{(t',s') \sim N_t(s)}[s' \in \mathcal{V}_\tau \mid t' \ni v] \leq \delta'/3$ and write $D = D_0 \setminus D'$. By Claim 5.8.19, we have $|D| \geq 9\eta'|s|$.

Consider now the following probabilistic experiment. Sample $v \in D$ uniformly, then some $t', s' \sim N_t(s)$ conditioned on $v \in t'$ and $s' \in \mathcal{V}_\tau$. We denote this distribution by $\mathcal{D}$. Note that for every fixed $v \in D$, this is exactly the distribution over which the plurality in the definition of $\tilde{f}_s(v)$ was defined.

With this in mind consider the event $P = \{f_s(v) \neq f_{s'}(v)\}$ and observe that $\mathbb{P}_{\mathcal{D}}[P] \geq \frac{1}{2}$ because for every fixed $v \in D$, by construction $f_s(v)$ disagrees with the plurality vote on v. On the other hand, we show toward contradiction that P is contained in events whose total probability is strictly less than $\frac{1}{2}$. Namely:

1. E_1: the event $\{f_s|_{t'} \overset{\eta'}{\not\approx} f_{s'}|_{t'}\}$
2. E_2: the event $|t' \cap D| < 6\eta'|t'|$
3. E_3: the event $P \setminus (E_1 \cup E_2)$.

Obviously $P \subseteq E_1 \cup E_2 \cup E_3$ so it is enough to bound the probability of the three events, starting with E_3. In this case we have that $f_s, f_{s'}$ disagree on at most an η'-fraction of t', but $|t' \cap D| \geq 6\eta'|t'|$. Conditioned on t', v is drawn uniformly from $|D \cap t'|$, so the probability that $f_s(v) \neq f_{s'}(v)$ is at most $\frac{1}{6}$.

Moving on to E_2, notice that if D is an $9\eta'$-fraction of s, a random $t' \subset s$ contains *less than a* $6\eta'$-fraction of D with probability $\exp(-\Omega(\eta'^2 k))$ by Chernoff. While the distribution $\mathcal{D}$ doesn't sample t' uniformly (since it conditions on $v \in t'$ and $s' \in \mathcal{V}_\tau$), we do know for any fixed $v \in D$, $v \in t'$ with probability at least $\frac{1}{2}$, and because $s \notin L$ and $v \notin D'$ the probability that $s' \in \mathcal{V}_\tau$ is at least $\frac{\delta'}{3}$ even after conditioning on v. Hence, even after conditioning on these events the probability that $|t' \cap D| < 6\eta'|t'|$-fraction of D is at most $2\delta^{-1}\exp(-\Omega(\eta'^2 k)) < \frac{1}{10}$ for large enough k.

Finally, we bound the probability of E_1 by similar reasoning.

$$\mathop{\mathbb{P}}_{\mathcal{D}}[E_1] = \mathop{\mathbb{P}}_{v \sim D, (t', s') \sim N_t(s)} \left[f_s|_{t'} \overset{\eta'}{\not\approx} f_{s'}|_{t'} \;\middle|\; s' \in \mathcal{V}_\tau, v \in t' \right]$$

$$\leq \frac{3}{\delta'} \mathop{\mathbb{P}}_{v\sim D, s'\sim N_t(s)} \left[s' \in \mathcal{V}_\tau \wedge f_s|_{t'} \overset{\eta'}{\not\approx} f_{s'}|_{t'} \;\middle|\; v \in t' \right]$$
$$\overset{s\notin B}{\leq} \frac{3}{\delta'} \cdot \frac{\delta'}{30} \cdot \mathbb{P}\left[v \in t'\right]^{-1} \leq \frac{1}{5}.$$

Thus $\frac{1}{2} \leq \mathbb{P}[P] \leq \mathbb{P}[E_1] + \mathbb{P}[E_2] + \mathbb{P}[E_3] < \frac{1}{2}$ and a contradiction is reached. □

Proof of Lemma 5.8.18. The proof is similar to the first property. Let L be as in Lemma 5.8.17. This time we define a set of bad *triples* as follows. For a set $(s, t', s') \sim N_t$ we consider the following quantity

$$F(s,t',s') \coloneqq \mathop{\mathbb{P}}_{(s_1,t_1,s_2,t_2,s_3,t_3,s_4)\sim N_t^3} \left[s_1, s_4 \in \mathcal{V}_\tau \wedge f_{s_1}|_{t''} \overset{\eta'}{\not\approx} f_{s_4}|_{t''} \;\middle|\; s_2 = s, t_2 = t', s_3 = s' \right]$$

where $t'' = t_2 \cap t_3 \cap t_4$. We define B to be the set of triples with large F-value:

$$B \coloneqq \left\{ (s,t',s') : F(s,t',s') \geq \frac{\delta'^2}{80} \right\},$$

and keep L as in the proof of Lemma 5.8.17. By excellence of τ and Markov's inequality $\mathbb{P}_{(s,t',s')\sim N_t}[B] \leq \frac{80\mu}{\delta'^2} \leq \exp(-\Omega(\eta k))$. Since s and s' are both marginally distributed as $X_t\left(\frac{k}{2}\right)$, we also have by Claim 5.8.19 and a union bound that either one of $s \in L$ or $s' \in L$ with probability at most $\delta'^2 = e^{-2c_5\eta k}$. It is therefore enough to show $\tilde{f}_s|_{t'} \overset{10\eta'}{\approx} \tilde{f}_{s'}|_{t'}$ under the assumption that $(s, t', s') \notin B$ and $s, s' \notin L$.

Toward this end, fix (s, t', s') and consider the set of disagreeing vertices on the smoothings at s and s':

$$D_0 \coloneqq \left\{ v \in t' : \tilde{f}_s(v) \neq \tilde{f}_{s'}(v) \right\}.$$

Assume for the sake of contradiction that $|D_0| \geq 10\eta'|t'|$. Similar to before, we actually consider $D = D_0 \setminus D'$ where D' is the set of vertices in t' such that $\mathbb{P}_{(t_3,s_4)\sim N_t(s')}[s_4 \in \mathcal{V}_\tau | t_3 \ni v]$ (likewise for $s_1, t_1 \sim N_t(s)$). By Claim 5.8.19, both s and s' have at most $\eta'|s|$ such vertices so $|D| \geq 6\eta'|t'|$.

We now consider the following probabilistic experiment. Sample $v \in D$ uniformly and $(s_1, t_1, s_2, t_2, s_3, t_3, s_4)$ conditioned on $s_2 = s, t_2 = t', s_3 = s'$. We then condition this experiment on:

1. $v \in t'' = t_1 \cap t_2 \cap t_3$.
2. $s_1, s_4 \in \mathcal{V}_\tau$.

Note that for a fixed $v \in D$, the marginals $(s_1, t_1, s_2 = s)$ and $(s_3 = s', t_3, s_4)$ are precisely the distributions that were used in the definition of $\tilde{f}_s(v), \tilde{f}_{s'}(v)$ respectively. As before, define the event $P = \{f_{s_1}(v) \neq f_{s_4}(v)\}$ and observe that $\mathbb{P}_{\mathcal{D}}[P] \geq \frac{1}{2}$ because for every $v \in D$ the plurality labeling for v over N_t under our conditioned process is different at s and s'.

We now define events whose union contains P which occur with total probability less than $\frac{1}{2}$. These are

1. E_1: the event $\{f_{s_1}|_{t''} \overset{\eta'}{\not\approx} f_{s_4}|_{t''}\}$
2. E_2: the event $|t'' \cap D| \leq 5\eta'|t''|$.
3. E_3: the event $P \setminus (E_1 \cup E_2)$.

By the same analysis as Lemma 5.8.17 we have $\mathbb{P}[E_3] < \frac{1}{5}$ and $\mathbb{P}[E_2] \leq \frac{1}{10}$ (we omit the repeated details). As for E_1 we have

$$\begin{aligned}
\mathop{\mathbb{P}}_{\mathcal{D}}[E_1] &= \mathop{\mathbb{P}}_{v \in D, s_1, t_1, t_3, s_4}\left[f_{s_1}|_{t''} \overset{\eta'}{\not\approx} f_{s_4}|_{t''} \;\middle|\; s_1, s_4 \in \mathcal{V}_\tau, v \in t''\right] \\
&\leq \frac{9}{\delta'^2} \mathop{\mathbb{P}}_{v \sim D, s' \sim N_t(s)}\left[s_1, s_4 \in \mathcal{V}_\tau \wedge f_{s_1}|_{t''} \overset{\eta'}{\not\approx} f_{s_4}|_{t''} \;\middle|\; v \in t''\right] \\
&\overset{(s,t',s') \notin B}{\leq} \frac{9}{\delta'^2} \cdot \frac{\delta'^2}{80} \cdot \mathbb{P}[v \in t']^{-1} \leq \frac{1}{5}.
\end{aligned}$$

We reach a similar contradiction as before. □

Proof of Claim 5.8.19. Let us begin with the first item. Recall under the assumptions of Theorem 5.8.7 N_t is reverse hypercontractive for indicators. Namely let c_6 be the constant promised by Theorem 5.6.4, and observe that by assumption $\mathbb{P}_{X_t}[\mathcal{V}_\tau] \geq \delta/2 \geq e^{-c_6\eta k}$. If $\mathbb{P}_{X_t}[L] \leq 2e^{-2c_5\eta k}$ we are done, so assume otherwise. Then by definition of L and Theorem 5.6.4 we have the chain of inequalities:

$$\mathbb{P}[L]\delta' \geq \mathop{\mathbb{P}}_{(s,s')\sim N_t}[s \in L, s' \in \mathcal{V}_\tau] \geq 2^{-q-1}\,\mathbb{P}[L]^q\delta^q.$$

where q is the constant given in Theorem 5.6.4. Re-arranging gives $\mathbb{P}[L] \leq (2^{q+1}\delta'\delta^{-q})^{\frac{1}{q-1}} \leq \exp(-c_7\eta k) \leq \frac{\delta^2}{2}$ for the appropriate choice of constant $c_7 > 0$. Thus if c_5 is sufficiently small we get the first item.

For the second item, let D' denote the set of $v \in s$ such that $\mathbb{P}_{(t',s')\sim N_t(s)}[s' \in \mathcal{V}_\tau \mid t' \ni v] \leq \frac{\delta'}{3}$. Define $g : \binom{s}{|t'|} \to [0,1]$ to be $g(t_0) = \mathbb{P}_{(t',s')\sim N_t(s)}[s' \in \mathcal{V}_\tau \mid\ t' = t_0]$. Observe that $\mathbb{E}[g] \geq \delta'$ since by assumption $s \notin L$. Moreover by construction we have $v \in D'$ if and only if $\mathbb{E}_{t_0 \ni v}[g(t_0)] \leq \frac{1}{3}\,\mathbb{E}[g(t_0)]$.

The problem is now reduced to sampling on the complete complex. In other words let $L = \binom{s}{|t'|}$, $R = s$ and $v \sim t$ if $v \in t$. By Chernoff for every α, this graph is an $(\alpha, \beta, \frac{1}{2})$-multiplicative sampler for $\beta = \exp(-\Omega(\alpha k))$. By Claim 5.4.8, this implies that the opposite graph is a $(\frac{6\beta}{\alpha}, 2\alpha, \frac{2}{3})$-multiplicative sampler. We take $\alpha = \eta'/2$, and the constant $c_5 > 0$ defined above small enough so that $\frac{6\beta}{\alpha} \leq \delta'$. As such we have for every function $g : L \to [0,1]$ of expectation at least δ' the fraction of vertices $v \in s$ such that $\mathbb{E}_{t' \subseteq s, v \in t'}[g] < \frac{1}{3}\,\mathbb{E}_{t' \subseteq s}[g(t')]$ is at most η' as desired. □

5.8.4 Examples of Global Complexes

We conclude the section with several examples of 'global HDX' to which Theorem 5.8.7 applies in a blackbox manner. We start with a few basic direct examples and

then argue a vast array of complexes studied in the approximate sampling literature give rise to global HDX. These are:

1. The complete complex.
2. Erdos-Renyi Hypergraphs.
3. Skeletons of the full linear matroid over $\mathbb{F}_q^d$.
4. Skeletons of ℓ_∞-independent complexes, including many classic spin systems.
5. Faces complexes of any of the above.

The first and simplest example of global complexes (other than the complete complex) come from the classical random model of Linial and Meshulam [283].

Definition 5.8.20 (Erdos-Renyi Hypergraphs)**.** The Erdos-Renyi hypergraph $X \sim G_k(n,p)$ is a random k-uniform simplicial complex whose $(k-1)$-skeleton is complete, and such that every k-face is sampled into X with probability p independently.

The second direct example we'll give is based on the full linear matroid over $\mathbb{F}_q^d$, whose faces consist of linearly independent vectors over $\mathbb{F}_q^d$:

$$X(i) = \{\{v_1, v_2, \ldots, v_i\} \mid \ v_1, v_2, \ldots, v_i \text{ are independent}\},$$

endowed with the uniform distribution.

Claim 5.8.21 (Global complexes)*.* The following families are global and HDX:

1. The k-uniform complete complex on n vertices for large enough n is:
   - (one-sided) 0-local-spectral HDX
   - $n^{-\Omega(k)}$-global
2. The Erdos-Renyi hypergraph $G_k(n,p)$ for large enough n is w.h.p:

- $o_n(1)$-local-spectral HDX
- $(1-p+o_n(1))$-global

3. The k-skeleton of the full linear matroid over $\mathbb{F}_q^d$ for large enough q, d is :
   - (one-sided) 0-local-spectral HDX
   - $q^{-\Omega(d)}$-global

Proof. We cover each case individually:

Complete Complex.

All links of the complete complex are complete and therefore 0-one-sided local spectral expanders. For globalness, observe that the only $\frac{k}{2}$-sets which do not appear in the link of some t are those that intersect it, which are an $1 - \frac{\binom{n-k}{k}}{\binom{n}{k}} \leq n^{-\Omega(k)}$ fraction.

Random Hypergraphs.

It is well known for fixed p that this complex is an $o_n(1)$-local-spectral expander with high probability so long as $p \gg \frac{k\log(n)}{n}$ (the top links are simply random $G(n,p)$ graphs). Since all other links are at least connected with high probability, Theorem 5.4.23 implies the local-spectral bound. Toward globality, fix any $\frac{k}{2}$-set $t \subset \binom{[n]}{k/2}$. By Chernoff the probability that more than $(1-p+o_n(1))$-fraction of $\frac{k}{2}$-sets disjoint from t fail to appear in X is much less than $n^{-\Omega(k)}$ for large enough n. Thus union bounding over t, the worst-case total variation of S_t from uniform over $\binom{n}{\frac{k}{2}}$ is at most $1-p+o_n(1)$ as desired.[14]

Linear Matroid.

All matroids (and therefore their skeletons) are 0-local spectral expanders [25] (for this matroid specifically, it is folklore). Toward globality, fix $\frac{k}{2}$ linearly independent vectors $B_1 = \{v_1, \ldots, v_{k/2}\}$. The set $B_2 = \{w_1, \ldots, w_{k/2}\}$ only fails to appear in the link of B_1 if B_1

[14] Formally, distance here should be measured from $\pi_{\frac{k}{2}}$, but this is within $o_n(1)$ of uniform in TV with high probability.

and B_2 are linearly dependent, which happens with probability roughly $\frac{q^{k+(k-1)d}}{q^{dk}} \leq q^{-\Omega(d)}$ (ignoring low-order terms). Since $\Pi_{k/2}$ is uniform over linearly independent sets this gives the result. □

The sampling literature is rife with examples of "ℓ_∞-independent" complexes (a term formally coined in [232]), which are closely related to global complexes. We'll consider the variant implicit in [96]. Let μ be a distribution over $[q]^d$, and observe μ induces a natural d-partite simplicial complex X_μ with

$$X_\mu(d) \coloneqq \{(1, a_1), \ldots, (d, a_d) : (a_1, \ldots, a_d) \in Supp(\mu)\}$$

and $\pi_X((1, a_1), \ldots, (d, a_d)) = \mu(a_1, \ldots, a_d)$. Let $S \subset [d]$ and $z \in [q]^S$ be such that $\mathbb{P}_{x\sim\mu}[x_S = z] > 0$ (we call such configurations 'feasible'). The (S, z)-*influence matrix* $\Psi_{z\to S}$ has entries

$$\Psi_{z\to S}((u, i), (v, j)) = \begin{cases} \mathbb{P}_{x\sim\mu}[x_i = u | x_j = v, x_S = z] - \mathbb{P}_{x\sim\mu}[x_i = u | x_S = z] & \text{if } j \neq i \\ 0 & \text{if } j = i. \end{cases}$$

A complex is called ℓ_∞-independent if $\|\Psi\|_\infty$ is bounded for all feasible configurations.

Definition 5.8.22 (ℓ_∞-independence)**.** Fix $q, d \in \mathbb{N}$ and let μ be a distribution over $[q]^d$. μ is called D-ℓ_∞-independent if for all feasible $S \subset [d]$ and $z \in \{0, 1\}^S$:

$$\|\Psi_{S\to z}\|_\infty \leq D.$$

We remark that in this regime, one typically thinks of q as fixed or small and the uniformity d as going to infinity. We show any sufficiently low-dimensional skeleton of an ℓ_∞-independent complex is a global HDX:

Proposition 5.8.23. *Let μ be a D-ℓ_∞-independent distribution over $[q]^d$ and $3 \le k \le d$. Then the k-skeleton $X_\mu^{\le k}$ satisfies:*

1. *$X_\mu^{\le k}$ is a $\frac{D+q}{d+3-k}$-two-sided local-spectral expander*
2. *$X_\mu^{\le k}$ is $\frac{k^2}{8} \cdot \frac{D+q}{d+3-k}$-global.*

We remark that (tighter) spectral variants of both items above are well known (see e.g. [24, 12]). Proposition 5.8.23 implies essentially every spin-system studied in the recent breakthrough line of work on approximate sampling through spectral independence admits optimal 1% agreement testers (taking $k \le \sqrt{\log(d)}$ skeletons).[15] See [232, Section 3] for an overview including distributions with the stochastic covering property, independent sets, various Ising/Potts models, list colorings, and more.

The proof of Proposition 5.8.23 relies on an intermediate notion of [200] called local-ℓ_∞-expansion.

Definition 5.8.24 ((λ, ∞)-expansion)**.** A simplicial complex X is called a (λ, ∞)-local expander if for every $\tau \in X$ with $|\tau| \le d - 2$:

$$\|A_\tau - \Pi_\tau\|_\infty \le \lambda$$

where A_τ is the weighted adjacency matrix of X_τ's 1-skeleton, and Π_τ is its corresponding stationary operator.

We require two lemmas regarding this notion. First, we note that up to dependence on the alphabet q (typically thought of as constant in this regime) ℓ_∞-expansion is essentially equivalent to ℓ_∞-independence.

[15] Note while spectral independence is weaker than what we require, almost all known methods for spin systems actually bound ℓ_∞-independence.

Lemma 5.8.25. *Let μ be a D-ℓ_∞-independent distribution over $[q]^d$. Then for every feasible $S \subset [d]$ and $x_S \in \{0,1\}^S$:*

$$\|A_{x_S} - \Pi_{x_S}\|_\infty \leq \frac{D+q}{d-|S|}.$$

Proof. We prove the case where $S = \emptyset$. The general statement follows from applying this argument to the links of X_μ. With this in mind, observe that by definition:

$$\begin{aligned}
&\|A_\emptyset - \Pi\|_\infty \\
=&\max_{(u,i)} \sum_{(v,j)\in X_{(u,i)}} \left| \frac{1}{d-1}\mathbb{P}[x_i = u | x_v = j] - \frac{1}{d}\mathbb{P}[x_i = u] \right| \\
\leq&\frac{1}{d}\max_{(u,i)} \left(\sum_{(v,j)\in X_{(u,i)}} |\mathbb{P}[x_i = u | x_j = v] - \mathbb{P}[x_i = u]| + \frac{1}{d-1}\mathbb{P}[x_i = u | x_j = v] \right) \\
\leq&\frac{q}{d} + \frac{1}{d}\max_{(u,i)} \sum_{(v,j)\in X_{(u,i)}} |\mathbb{P}[x_i = u | x_j = v] - \mathbb{P}[x_i = u]| \\
=&\frac{q}{d} + \frac{1}{d}\max_{(u,i)} \sum_{(v,j)\in X_{(u,i)}} |\mathbb{P}[x_i = u | x_j = v] - \mathbb{P}[x_i = u]| \\
\leq&\frac{D+q}{d}
\end{aligned}$$

□

Second, we'll need the slightly more involved fact from [200] that any ℓ_∞-local expander is global. We reproduce the proof here for completeness.

Lemma 5.8.26. *Let X be a d-uniform λ-ℓ_∞-local expander. Then X is $\frac{d^2}{8}\lambda$-global.*

Proof. Recall that it is sufficient to bound the infinity norm:

$$\|S_{\frac{d}{2},\frac{d}{2}} - \Pi_{\frac{d}{2},\frac{d}{2}}\|_\infty \leq \frac{d^2}{4}\lambda.$$

We prove a slightly stronger statement by induction. For any $\tau \in X$ and feasible i, j both

the ∞-norm and 1-norm are bounded by:

$$\|S^\tau_{i,j} - \Pi^\tau_{i,j}\|_1, \|S_{i,j} - \Pi_{i,j}\|_\infty \leq ij\lambda.$$

We induct on $i + j$. For the base case $i = j = 1$, first observe the ∞-norm is exactly the λ-ℓ_∞-expansion of the 1-skeleton. A classical consequence of Hölder duality states that for any Hölder conjugates (p, q) and operator M

$$\|M\|_p = \|M^*\|_q,$$

where M^* is M's adjoint. Since $S_{1,1} - \Pi_{1,1}$ is self-adjoint we therefore have

$$\|S^\tau_{0,0} - \Pi^\tau_{0,0}\|_1 = \|S^\tau_{0,0} - \Pi^\tau_{0,0}\|_\infty \leq \lambda$$

as desired.

Assume now by induction that, for some fixed $i + j > 0$, all $i' + j' < i + j$ and $\tau \in X$ satisfy

$$\|S^\tau_{i',j'} - \Pi^\tau_{i',j'}\|_\infty \leq i'j'\lambda.$$

Let $p \in \{1, \infty\}$ and let $\bar{p}$ denote the complement of p. We first argue we may assume without loss of generality that $i > 1$. This is again by Hölder duality since

$$\|S_{i,j} - \Pi_{i,j}\|_p = \|S_{j,i} - \Pi_{j,i}\|_{\bar{p}}.$$

Thus we are done if we show the result for both $p \in \{1, \infty\}$ just for the case $i \geq j$ (and thus $i > 1$).

Toward this end, fix any $f : X(j) \to \mathbb{R}$. As is typically the case, the idea is to

localize the input by first drawing an $(i-1)$-face, then a vertex from its link:

$$\|S_{i,j}f - \Pi_{i,j}f\|_p = \left\| \|(S_{i,j}f)|_s - (\Pi_{i,j}f)|_s\|_{p,v\in X_s(1)} \right\|_{p,s\in X(i-1)}$$

where for any $g \in X(i) \to \mathbb{R}$, $g|_s(v) = g(s \cup v)$ denotes the localization of g to the link of s. For concreteness, we write out the above explicitly for the $p = 1$ case in slightly less compact notation. The $p = \infty$ case is the same replacing expectations with absolute maxima:

$$\begin{aligned}
\|S_{i,j}f - \Pi_{i,j}f\|_p &= \mathop{\mathbb{E}}_{t\in X(i)}[|S_{i,j}f(t) - \Pi_{i,j}f(t)|] \\
&= \mathop{\mathbb{E}}_{s\in X(i-1)}\left[\mathop{\mathbb{E}}_{v\in X_s(1)}[|S_{i,j}f(s\cup v) - \Pi_{i,j}f(s\cup v)|]\right] \\
&= \mathop{\mathbb{E}}_{s\in X(i-1)}\left[\mathop{\mathbb{E}}_{v\in X_s(1)}[|S_{i,j}f|_s(v) - \Pi_{i,j}f|_s(v)|]\right] \\
&= \mathbb{E}_{s\in X(i-1)}\left[\mathbb{E}_{v\in X_s(1)}[|S_{i,j}f|_s(v) - \Pi_{i,j}f|_s(v)|]\right] \\
&= \mathbb{E}_{s\in X(i-1)}\left[\left|\mathbb{E}_{v\in X_s(1)}[|S_{i,j}f|_s(v) - \Pi_{i,j}f|_s(v)|]\right|\right] \\
&= \left\| \|(S_{i,j}f)|_s - (\Pi_{i,j}f)|_s\|_{p,v\in X_s(1)} \right\|_{p,s\in X(i-1)}.
\end{aligned}$$

We stick to the more compact norm notation for the remainder of the proof. The trick is now to observe that as a function of $X_s(1)$, $S_{i,j}f|_s$ is exactly $S^s_{1,j}f^s$, where $f^s : X_s(j) \to \mathbb{R}$ is the restriction $f^s(\tau) = f(\tau)$. Then by adding and subtracting the corresponding local stationary operator we have:

$$\begin{aligned}
\|S_{i,j}f - \Pi_{i,j}f\|_p &= \left\| \|S^s_{1,j}f^s - \Pi^s_{1,j}f^s + \Pi^s_{1,j}f^s - (\Pi_{i,j}f)|_s\|_{p,v\in X_s(1)} \right\|_{p,s\in X(i-1)} \\
&\le \left\| \|S^s_{1,j}f^s - \Pi^s_{1,j}f^s\|_{p,v\in X_s(1)} \right\|_{p,s\in X(i-1)}
\end{aligned}$$

$$+ \left|\left| \|\Pi^s_{1,j} f^s - (\Pi_{i,j} f)|_s\|_{p,v\in X_s(1)} \right|\right|_{p,s\in X(i-1)}$$

by the triangle inequality. The first term is now bounded by the inductive hypothesis applied in the link of s:

$$\left|\left| \|S^s_{1,j} f^s - \Pi^s_{1,j} f^s\|_{p,v\in X_s(1)} \right|\right|_{p,s\in X(i-1)} \leq \left|\left| j\lambda \|f^s\|_{p,X_s(j)} \right|\right|_{p,s\in X(i-1)} = j\|f\|_{p,X(j)}$$

Toward analyzing the second term, observe that $\Pi^s_{1,j} f^s = S_{i-1,j} f(s)$ and $(\Pi_{i,j} f)|_s(v) = \Pi_{i-1,j} f(s)$ so:

$$\left|\left| \|\Pi^s_{1,j} f^s - (\Pi_{i,j} f)|_s\|_{p,v\in X_s(0)} \right|\right|_{p,s\in X(i-1)} = \|S_{i-1,j} f - \Pi_{i-1,j} f\|_{p,s\in X(i-1)} \leq (i-1)j\lambda\|f\|_{p,X(j)}$$

by the inductive hypothesis. Altogether this gives

$$\|S_{i,j} f - \Pi_{i,j} f\|_p \leq \lambda j + \lambda(i-1)j = ij\lambda$$

as desired. □

The proof of Proposition 5.8.23 is now essentially immediate.

Proof of Proposition 5.8.23. The first fact is immediate from Lemma 5.8.25 and the fact that $\|A_\tau - \Pi_\tau\|_2 \leq \|A_\tau - \Pi_\tau\|_\infty$. The second fact is an immediate consequence of combining Lemma 5.8.25 and Lemma 5.8.26. □

We conclude the subsection by the observation that if a complex X is global, then so is its faces complex (Definition 5.5.7). Since the faces complex also inherits the two-sided expansion of the original complex up to factors in uniformity (c.f. the proof of Theorem 5.5.1), this shows that one can also take as examples faces complexes of any of the examples above.

Claim 5.8.27. Let X be a k-uniform and λ-global complex. Let ℓ be such that ℓ divides k, then the ℓ-faces complex F_X^ℓ is λ-global.

Proof. Let $m = \frac{k}{\ell}$ for notational convenience. Fix $t = \{s_1, s_2, \ldots, s_{\frac{m}{2}}\} \in F_X^\ell(\frac{m}{2})$. Observe that the set of neighbors of t with respect to the swap walk in $F^\ell X$ are exactly those t' such that $\cup t$ and $\cup t'$ are neighbors in the swap walk on $X\binom{k}{2}$ (here $\cup t = s_1 \cup s_2 \cup \cdots \cup s_{\frac{m}{2}}$ and similarly for t'). Moreover, one samples a neighbor of t by sampling a random neighbor τ of $\cup t$, and then sampling a uniform at random partition of τ to $t' = \{s'_1, s'_2, \ldots, s_{\frac{m}{2}}\}$ (the partition is chosen independent of t).

Thus one observes that the TV-distance between S_t and Π can be written as

$$\begin{aligned}\frac{1}{2}\sum_{t' \in F^\ell X\binom{m}{2}} |\mathop{\mathbb{P}}_{S_t(F^\ell X)}[t'] - \mathop{\mathbb{P}}_{\Pi}[t']| &= \frac{1}{2}\sum_{\tau \in X\binom{k}{2}} \sum_{t' : \cup t' = \tau} |\mathop{\mathbb{P}}_{S_t(F^\ell X)}[t'] - \mathop{\mathbb{P}}_{\Pi}[t']| \\ &= \frac{1}{2}\sum_{\tau \in X\binom{k}{2}} \sum_{t' : \cup t' = \tau} p|\mathop{\mathbb{P}}_{S_{\cup t}(X)}[\tau] - \mathop{\mathbb{P}}_{\Pi(X)}[\tau]|\end{aligned}$$

where p is one over the number of partitions of τ to t'. Obviously this is equal to

$$\frac{1}{2}\sum_{\tau \in X\binom{k}{2}} \mathop{\mathbb{P}}_{S_{\cup t}(X)}[\tau] - \mathop{\mathbb{P}}_{\Pi(X)}[\tau] = d_{TV}(S_{\cup t}(X), \Pi(X)) \leq \lambda$$

as desired. □

5.9 Analytic, Geometric, and Combinatorial Applications

In this section, we cover several brief applications of sampling and reverse hypercontractivity in well-studied combinatorial and analytic settings.

5.9.1 New Double Samplers

Double samplers are a strengthened notion of sampling introduced in [124, 121] that consist of two "stacked" samplers with additional local sampling properties under 'closure' of the top layer. These interesting objects have powerful applications in agreement testing [124] and list-decoding [121], and have even seen algorithmic use in the construction of space-efficient data structures for the heavy hitter problem [132].

Definition 5.9.1 (Double sampler, [121])**.** A double sampler consists of a triple (V_2, V_1, V_0), where V_0 is the ground set, V_1 is a collection of i-subsets of V_0 and V_2 is a collection of k-subsets of V_0, where $k > i \in \mathbb{N}$. We say that (V_2, V_1, V_0) is an $(\varepsilon, \beta, \varepsilon_0, \beta_0)$-*double sampler* if

- The inclusion graphs on (V_2, V_1) and (V_2, V_0) are (ε, β)-additive samplers and the inclusion graph on (V_1, V_0) is an $(\varepsilon+\varepsilon_0, \beta+\beta_0)$-additive sampler[16] (recall the inclusion graph is defined by connecting two subsets by an edge if one contains the other).
- For every $T \in V_2$, let $V_1(T) = \{S \in V_1 \ : \ S \subset T\}$ be the sets in V_1 that are contained in T. Let $G_{|T}$ be the bipartite inclusion graph connecting elements in T (viewed as elements in the ground set V_0) to subsets in $V_1(T)$. We require that for every $T \in V_2$, the graph $G_{|T}$ is an (β_0, ε_0)-additive sampler.

Double samplers are only known to arise from high dimensional expanders. In their original work, [124, 121] use the Ramanujan complexes [290] to construct explicit double samplers for all $(\varepsilon, \beta, \varepsilon_0, \beta_0)$:

Theorem 5.9.2 ([121, Theorem 2.11])**.** *For every $(\varepsilon, \beta, \varepsilon_0, \beta_0) > 0$, there exists an explicit*

[16]The weights over edges in (V_i, V_j) are the marginals of the probabilistic experiment where we first choose $T \in V_2$ (according to some given weight distribution Π), and then choose $v \in V_0$ and $S \in V_1$ such that $v \in S \subseteq T$, uniformly at random over all such pairs S, v.

infinite family of double samplers $\{(V_0^{(n)}, V_1^{(n)}, V_2^{(n)})\}$ *such that for all* n*:*

$$\frac{|V_1^{(n)}|}{|V_0^{(n)}|}, \frac{|V_2^{(n)}|}{|V_0^{(n)}|} = \exp\left(\text{poly}\left(\frac{1}{\beta}, \frac{1}{\beta_0}, \frac{1}{\varepsilon}, \frac{1}{\varepsilon_0}\right)\right).$$

One of the main open questions posed in [121] is to settle the overhead of double samplers. In particular, while it is known that exponential dependence on ε_i is necessary (even for standard samplers), it is plausible the dependence on the 'failure probability' β_i could be improved, perhaps even to match the corresponding polynomial dependence of optimal standard samplers. Since typical related applications (e.g. for low soundness agreement tests and PCPs [215, 128]) take ε to be constant (or at most some polylog in β), this latter dependence is where the crux of the problem lies.

To the best of our knowledge, there have been no lower bounds or improvements over Theorem 5.9.2 since [121]'s original work. Leveraging our concentration bounds and their corresponding optimality (see Section 5.11.1) we take a significant step toward resolving this problem: quasi-polynomial size double samplers exist and, under reasonable assumptions on the underlying complex, are the best possible.

Theorem 5.9.3 (Quasi-Polynomial Double Samplers)**.** *for every* $(\varepsilon, \beta, \varepsilon_0, \beta_0) > 0$*, there exists an explicit family of* (β, ε)*,*(β_0, ε_0) *double samplers* $\{(V_2^{(n)}, V_1^{(n)}, V_0^{(n)})\}$ *of size at most:*

$$\frac{|V_1^{(n)}|}{|V_0^{(n)}|}, \frac{|V_2^{(n)}|}{|V_0^{(n)}|} \leq \exp\left(\tilde{O}\left(\frac{\log^3 \frac{1}{\beta} \log^3 \frac{1}{\beta_0}}{\varepsilon^6 \varepsilon_0^6}\right)\right).$$

Moreover, for $\varepsilon \in (0, 0.01)$*, if the underlying complex family* $\{X_n\}$ *is* $\frac{\varepsilon}{3}$*-hitting, vertex-uniform, and non-contracting,*[17] *this is optimal up to polynomial factors in the exponent:*

$$\frac{|V_2^{(n)}|}{|V_0^{(n)}|} \geq \exp\left(\Omega\left(\frac{\log(\frac{1}{\beta}) \log(\frac{1}{\beta_0})}{\varepsilon^2 \varepsilon_0^2}\right)\right)$$

Note the above does not fully rule out the existence of simplicial poly-size double

[17]Here we mean 1) $\pi_0(v) = |X(1)|^{-1}$ for all $v \in X(1)$, and 2) $|X(i)| \geq |X(i-1)|$ for all $2 \leq i \leq d$.

samplers. While the assumption on regularity in the lower bound can be easily relaxed, it is unclear to what extent the assumption on hitting-set is an artifact of our proof technique (see Section 5.11.1 for details), though it is certainly a natural property for such objects to satisfy. Moreover, our bound strongly relies on the assumption that $|V_2| \geq \exp(d)|V_0|$ (implied by X being non-contracting), which is not generically true. Nevertheless, high dimensional expanders exhibit both good hitting set behavior (see Proposition 5.10.12) and at least exponential degree (see Proposition 5.9.6), so it is at least accurate to say double samplers arising from HDX are at best quasi-polynomial in β^{-1}.

The proof of Theorem 5.9.3 relies on two further results. The first, toward our upper bound, is Lubotzky, Samuels, and Vishne's [291] classical construction of Ramanujan complexes.

Theorem 5.9.4 (LSV-Complexes [291])**.** *For any $\lambda > 0$ and $d \in \mathbb{N}$, there exists an explicit infinite family of d-partite one-sided λ-local-spectral expanders $\{X^{(n)}, \Pi^{(n)}\}$ with degree $deg(X) \leq \lambda^{-O(d^2)}$. Moreover, for every $d \geq k > \ell \geq 0$, $\max\{\pi_k\} \leq o_n(1)$.*

The second, toward the lower bound, is the following optimality result we prove in Section 5.11.1.

Corollary 5.9.5 (Corollary of Theorem 5.11.1)**.** *For any $\varepsilon \in (0, 0.01)$, let X be a d-uniform, $\frac{\varepsilon}{3}$-hitting vertex-uniform complex and $i \leq k \leq d$. Then there exists a set $A \subset X(i)$ such that:*

$$\mathop{\mathbb{P}}_{s \in X(k)}[|U_{i,k} 1_A(s) - \mu_A| > \varepsilon] \geq \exp\left(O\left(\varepsilon^2 \frac{k}{i}\right)\right).$$

We are now ready to prove Theorem 5.9.3.

Proof of Theorem 5.9.3. We split into the upper and lower bounds.

Upper Bound.

Let $k = O\left(\frac{\log(1/\beta)\log(1/\beta_0)}{\varepsilon^2 \varepsilon_0^2}\right)$, $d = k\log(k)$, and $\{X_n\}$ be an explicit infinite family of d-uniform LSV-complexes with local-spectral expansion $\lambda = 2^{-\Omega(d)}$, and take $V_2 = X(k)$,

$V_1 = X(i)$, and $V_0 = X(1)$ for $i = \Theta(\frac{\log(1/\beta_0)}{\varepsilon_0^2})$. By Theorem 5.5.1, (V_2, V_1) is an (ε, β)-sampler for the right choice of constants. Moreover, every closure graph is simply the $(\frac{\log(1/\beta_0)}{\varepsilon_0^2})$-uniform complex on k vertices, which is negatively correlated and therefore also a (ε_0, β_0)-sampler by standard Chernoff-Hoeffding for negatively correlated variables. The overhead bound is then immediate from the degree of LSV complexes in Theorem 5.9.4.

Lower Bound.

Recall that by definition $V_0^{(n)} = X_n(1) = [n]$, $V_1^{(n)} = X_n(i)$, and $V_2^{(n)} = X_n(k)$ for some family of simplicial complexes $\{X_n\}$ and set sizes $1 < i < k$. By Corollary 5.9.5, for any $\varepsilon \in (0, 0.01)$, $(X(k), X(i))$ is at best an $(\varepsilon, e^{O(\varepsilon^2 \frac{k}{i})})$-sampler, therefore any such double sampler requires:

$$\frac{k}{i} \geq \Omega\left(\frac{\log(1/\beta)}{\varepsilon^2}\right).$$

On the other hand, it is also required for every $T \in X(k)$ that the induced graphs $G|_T = (V_2|_T, V_1|_T)$ are (ε_0, β_0)-samplers. By standard sampling lower bounds [87], this graph is at best a $(\varepsilon_0, e^{O(\varepsilon_0^2 i)})$-sampler, forcing

$$i \geq \Omega\left(\frac{\log(1/\beta_0)}{\varepsilon_0^2}\right).$$

Combining these bounds we have that for any valid choice of i the top uniformity k of an $(\varepsilon, \beta, \varepsilon_0, \beta_0)$-double sampler is at least

$$k \geq \Omega\left(\frac{\log(\frac{1}{\beta})\log(\frac{1}{\beta_0})}{\varepsilon^2 \varepsilon_0^2}\right).$$

It is therefore enough to argue that the overhead is at least exponential in the uniformity k:

$$\frac{|V_2|}{|V_0|} \geq 2^{\Omega(k)}.$$

This is true for any non-contracting complex. In particular, since $|X(k)| \geq |X(k/2)|$, the average degree of every $(k/2)$-face (and hence every vertex) is at least $2^{\Omega(k)}$. □

We remark that it may also be possible to construct double samplers with smaller overhead by removing the *simplicial* requirement from the definition (see e.g. the relaxed notion in [132]). It is likely, for instance, that 'Grassmannian HDX' [243, 171, 127] based on subspace structure lead to double-samplers with different parameters-overhead trade-offs.

5.9.2 A Degree Lower Bound for HDX

One of the most classical results in the study of expander graphs is the Alon-Boppana Theorem [307], roughly stating that any family of bounded-degree λ-expanders must have degree at least $\frac{2}{\lambda^2}$. In higher dimensions, the situation is less understood. Not only should degree scale in some way with the local expansion λ, but dependence on *dimension* becomes a critical parameter in application. In this section, we give (to our knowledge) the first systematic study of the degree of high dimensional expanders. In particular, we prove that various families of hyper-regular HDX require super-exponential degree and prove a threshold-phenomenon exists at the TD-barrier: there exist (reasonably balanced) constructions of 1-TD complexes achieving exponential degree at every level, while any (reasonably balanced) λ-TD complex must have degree $2^{-\Omega(k^2)}$ for any level $k \leq \sqrt{d}$.

Before applying our stronger concentration bounds, we first look at what one can infer about the degree of HDX directly from spectral methods. The following proposition, though not to our knowledge appearing anywhere in the literature, is the result of applying elementary spectral methods to the problem:

Proposition 5.9.6 (Spectral Lower Bound)**.** *Let X be a d-uniform $\{\lambda_i\}$-two-sided local-*

spectral expander. Then for every $k < d$, the degree of X's k-skeleton is at least:

$$\deg^{(k)}(X) \geq \frac{1}{(k-1)!} \prod_{i=1}^{k-1} \frac{1}{\lambda_i}$$

Proof. Observe the k-th degree of a vertex $v \in X(1)$ can be bounded by

$$\frac{1}{(k-1)!} \prod_{i=1}^{k-1} \min_{t \in X(i)} \{|X_t(1)|\}$$

simply by counting the number of ways a face $\{v_2, \ldots, v_{k-1}\} \in X_{v_1}(k)$ can be sampled by choosing $v_2 \sim X_{v_1}(1)$, $v_2 \in X_{\{v_1,v_2\}}(1)$ and correcting for the $\frac{1}{(k-1)!}$ possible orderings of the face. The result now follows from the elementary fact (see e.g. [162, Lemma A.7]) that any two-sided λ-expander has at least $\frac{1}{\lambda}$ vertices. □

We remark that one might reasonably think to use Alon-Boppana [307] (the classic 'optimal' trade-off between expansion and degree) to prove a version of the above for one-sided expansion. This does not work for high dimensional expanders, since their links may have low diameter.

Proposition 5.9.6 implies that any (hyper-regular) $o(\frac{1}{d})$-two-sided HDX must have super-exponential degree but is essentially trivial for complexes near the TD-barrier, under spectral independence, or even for arbitrarily strong partite one-sided HDX. In these cases, even if one is willing to look e.g. at polynomial size cutoffs (as is quite common in the literature) Proposition 5.9.6 only gives a slightly super-exponential lower bound:

$$\deg^{(k)}(X) \geq \exp(\Omega(k \log(k))).$$

Using concentration, we improve this bound to strongly super-exponential. We start with the partite case.

Theorem 5.9.7. *Fix $c \in [0,1]$ and let X be a d-uniform hyper-regular partite $2^{-\Omega(d^c)}$-HDX.*

Then either:

1. *X's 1-skeleton is dense:*[18]
$$\deg^1(X) \geq \frac{|X(1)|}{2}$$

2. *X has super-exponential degree:*
$$\deg^{(d)}(X) \geq \exp(\Omega(\max\{d^{2c}, d\}))$$

Second we prove an analogous bound for skeletons of TD and SI-complexes.

Theorem 5.9.8. *Let $\lambda, c \in [0,1)$, $\eta > 0$, and $d \in \mathbb{N}$. If X is a d-uniform hyper-regular λ-TD or η-SI complex then for any $k \leq d^c$:*

1. *X's 1-skeleton is dense:*
$$\deg^1(X) \geq \frac{|X(1)|}{2}$$

2. *X's k-skeleton has super-exponential degree:*
$$\deg^{(k)}(X) \geq \exp(\Omega_{\lambda,\eta}(\min\{k^2, k^{1/c}\}))$$

The latter setting above is tight in the sense that when $\lambda = 1$, that is *at* the TD-barrier, there exist (reasonably balanced) complexes where $\deg^{(k)}(X) = \exp(O(k))$ for all $k \leq d$. On the other hand, beyond the TD barrier the best known degree upper bound for a hyper-regular (or even reasonably balanced) HDX is $\exp(\exp(d))$ [159], leaving a substantial gap with the above. Nevertheless, Theorem 5.9.8 is the first to exhibit quantitative threshold behavior for degree at the TD-barrier.

Theorem 5.9.7 and Theorem 5.9.8 leave open two critical questions:

[18] While this condition may seem strange, note it is necessary to handle dense settings. Take, e.g., $\{0,1\}^n$, which is a hyper-regular 0-local-spectral expander whose $\sqrt{n}$-skeleton has degree $\exp(\sqrt{n}\log(n))$.

1. Can we prove super-exponential bounds for the *top* level of a λ-TD/η-SI complex?

2. Can the assumption of hyper-regularity be dropped?

As we will see in the proof of Theorem 5.9.8, the first question would be resolved if one could give *any* asymptotic improvement over exponential concentration for weak HDX. In other words, any bound of the form:

$$\mathop{\mathbb{P}}_{s \in X(d)} \left[\frac{1}{d} |s \cap A| - \mu > \frac{1}{2} \right] \leq \exp(\omega(\sqrt{d}))$$

for X and its links would imply a super-exponential degree lower bound. If, for instance, we have a bound of $\exp(-\Omega(d))$ holds, this would imply a degree lower bound of $\exp(\Omega(d^2))$ on the top level of X.

Regarding the second question, our proof technique extends to any complex whose links are reasonably 'balanced' in the sense that there should not be a $2^{-\Omega(\sqrt{d})}$ fraction of vertices making up an $\Omega(1)$ fraction of the mass. Unfortunately, many classical HDX constructions, e.g. the Ramanujan complexes of [291], actually are highly unbalanced in this sense. The links of these constructions essentially approximate an 'unbalanced product': they are partite complexes which are regular on each individual part, but many of the parts are vanishingly small and therefore make up a significant portion of the mass. While known constructions of this type all have minimum part size at least exponential in d (and therefore result in degree $\exp(d^2)$), it is unclear how to rule out a construction whose links look, e.g., like $\{0,1\}^d \times [2^d]$. Such a complex could potentially satisfy optimal inclusion sampling while maintaining exponential degree.

We leave the possible extension of our lower bounds to the top level of unbalanced complexes (or conversely the construction of highly locally unbalanced exponential-size HDX) as one of the main open problems suggested by this work:

Question 5.9.9 (Lower Bounds for Locally Unbalanced HDX)**.** Let $\{X_n\}$ be an infinite

family of bounded-degree HDX that are either λ-TD or η-SI. Is the degree of every (sufficiently large) X_n super-exponential:

$$\deg(X_n) \geq \exp(\omega(d))?$$

Question 5.9.9 is critical from the standpoint of applying HDX in other areas of theoretical computer science where degree dependence corresponds to the 'overhead' incurred by using the complex as a gadget. For instance, the lower bound in Theorem 5.9.3 can be achieved by HDX essentially if and only if Question 5.9.9 is false. Similarly, a variant of Question 5.9.9 where d may depend on n is critical to the study of agreement testers whose soundness scale with the *number of vertices* of the system. Constructing polynomial size testers with inverse polynomial soundness is the combinatorial core of the so-called *sliding scale conjecture* in PCP theory. If the above lower bound holds, HDX cannot give such objects: d must be taken to be $o(\log(n))$ to maintain polynomial size, and the resulting soundness of any HDX-based poly-size agreement tester would be at best $2^{-o(\log(n))}$, missing the inverse polynomial mark.

We now move to the proofs of Theorem 5.9.7 and Theorem 5.9.8 which are inspired by our *localization* technique for proving concentration that samples $s = \{v_1, v_1, \ldots, v_d\}$ by iteratively sampling $v_0 \in X$, then $v_2 \in X_{v_1}$, $v_3 \in X_{\{v_1,v_2\}}$... and analyzing the local concentration at each step. Upon further inspection, this technique yields the fact that the 1-skeleton of a complex with good sampling in 0 and 1-links must itself be an excellent sampler graph.

Lemma 5.9.10. *Fix $\varepsilon, \delta_1, \delta_2 > 0$. Let X be a d-uniform simplicial complex such that:*

1. *$(X(d), X(1))$ is a $(\frac{\varepsilon}{2}, \delta_1)$-additive sampler.*
2. *For every $v \in X(1)$, $(X_v(d), X_v(1))$ is an $(\frac{\varepsilon}{2}, \delta_2)$-additive sampler.*

Then the partite double cover of X's 1-skeleton is a $(\varepsilon, \frac{2\delta_1}{1-\delta_2})$-sampler.

Proof. Fix $A \subseteq X(1)$. At a high level, the proof follows from observing that if the base graph is a poor sampler, then after drawing the first vertex v_1 of our d-set $t = \{v_1, \ldots, v_d\}$, there is a non-trivial chance that the neighborhood of v_1 (i.e. $X_{v_1}(1)$) 'sees' A in the wrong proportion. However, since the remainder of the face $\{v_2, \ldots, v_d\} \in X_{v_1}(d)$, if $(X_{v_1}(d), X_{v_1}(1))$ is indeed a good sampler it will maintain this incorrect proportion from the first step and violate sampling of A in $(X(d), X(1))$.

We now formalize this argument. Given our assumptions we'd like to show:

$$\mathop{\mathbb{P}}_{v \in X(1)} \left[\mathop{\mathbb{P}}_{u \in X_v(1)} [A] > \mathbb{P}[A] + \varepsilon \right] \leq \frac{\delta_1}{1 - \delta_2}.$$

The lower tail follows from applying this on the complement set. Toward this end, denote by B the set of 'bad' vertices in A that over-sample the base graph:

$$B = \left\{ v \in X(1) \;\middle|\; \mathop{\mathbb{P}}_{u \in X_v(1)} [A] > \mathbb{P}[A] + \varepsilon \right\}.$$

We need to show $\mathbb{P}[B] \leq \frac{\delta_1}{1-\delta_2}$. We bound $\mathbb{P}[B]$ by relating it to the set of bad d-faces:

$$C = \left\{ t \in X(d) \;\middle|\; \left| \mathop{\mathbb{P}}_{u \in t} [A] - \mathbb{P}[A] \right| > \frac{\varepsilon}{2} \right\},$$

which has measure at most $\mathbb{P}[C] \leq \delta_1$ by Condition (2). The key to relate B and C is to observe we can sample a random d-face t by first sampling a vertex $v \in X(1)$, then sampling t from its link:

$$\mathbb{P}[C] = \mathop{\mathbb{E}}_{t \in X(d)} [\mathbf{1}_C(t)] = \mathop{\mathbb{E}}_{v \in X(1)} \left[\mathop{\mathbb{E}}_{t \in X_v(d)} [\mathbf{1}_C(t)] \right] \geq \mathbb{P}[B] \mathop{\mathbb{E}}_{v \in X(1)} \left[\mathop{\mathbb{E}}_{t \in X_v(d)} [\mathbf{1}_C(t)] \;\middle|\; v \in B \right].$$

When $v \in B$, we have $\mathbb{P}_{u \in X_v(1)}[A] \geq \mathbb{P}[A] + \varepsilon$ by assumption, but then by Condition (1)

$$\mathop{\mathbb{E}}_{v \in X(1)} \left[\mathop{\mathbb{E}}_{t \in X_v(d)} [\mathbf{1}_C(t)] \;\middle|\; v \in B \right] \geq \mathop{\mathbb{P}}_{t \in X_v(d)} \left[\mathbb{P}[A \mid t] > \mathop{\mathbb{P}}_{u \in X_v(1)} [A] - \frac{\varepsilon}{2} \right] \geq 1 - \delta_2.$$

Putting everything together we have

$$(1-\delta_2)\,\mathbb{P}[C] \le \mathbb{P}[B] \le \delta_1.$$

and $\mathbb{P}[C] \le \frac{\delta_1}{1-\delta_2}$ as desired. □

Combining this fact with standard degree lower bounds for sampler graphs [87] gives the following degree lower bound for the 1-skeletons of such complexes:

Corollary 5.9.11. *Let X be a d-uniform simplicial complex satisfying:*

1. *$(X(d), X(1))$ is a $(\frac{1}{10}, \delta_1)$-sampler*
2. *For all $v \in X(1)$, $(X_v(d), X_v(1))$ is a $(\frac{1}{10}, \delta_2)$-sampler.*

Then X's 1-skeleton has high max-degree:

$$\exists v \in X(1) : \deg^{(1)}(v) \ge \min\left\{\frac{|X(1)|}{2}, \frac{2(1-\delta_2)}{5\delta_1}\right\}$$

Proof. By Lemma 5.9.10 the underlying graph of X is a $(\frac{1}{5}, \delta = \frac{2\delta_1}{1-\delta_2})$-sampler. [87, Theorem 2] states the following relation on a graph $G = (L, R, E)$ that is an (ε, δ)-sampler of max degree $t \le \frac{|X(1)|}{2}$.

$$t|R| \ge \frac{|L|(1-2\varepsilon)}{2\delta}.$$

Here $L = R = X(1)$ and this inequality gives $t \ge \frac{0.8}{2\delta} = \frac{2(1-\delta_2)}{5\delta_1}$. □

Notice this bound is substantially stronger than what can be inferred from spectral expansion alone. For instance, when X is a $\lambda = \text{poly}(d^{-1})$-HDX, Alon-Boppana only implies a degree lower bound of $\Omega(\lambda^{-2}) = \text{poly}(d^{-1})$, whereas combining the above with Corollary 5.7.7 gives degree at least $\exp(-\sqrt{d})$. Given this fact, Theorem 5.9.7 and Theorem 5.9.8 follows by recursive application of Corollary 5.9.11 on each link of the HDX.

Proof of Theorem 5.9.7 and Theorem 5.9.8. Similar to the spectral argument, since X is hyper-regular the kth degree of any vertex $v_0 \in X(1)$ can be bounded by:

$$\deg^{(k)}(v_0) \geq \frac{1}{(k-1)!} \prod_{i=1}^{k-1} \min_{t \in X(i)} \left\{ \deg^{(1)}(X_t) \right\}.$$

It is therefore sufficient to bound the degree of the 1-skeleton of the links of X. We split the analysis into the partite and TD/SI cases.

Partite Case.

Let $k' = \max\{\sqrt{d}, d^c\}$. By Theorem 5.7.1, there exists a universal constant $c' > 0$ such that for every $t \in X^{\leq \frac{d}{2}}$, the link X_t satisfies:

1. $(X_t(d - |t|), X_t(1))$ is a $(\frac{1}{10}, \exp(-c'k'))$-sampler
2. For all $v \in X_t(1)$, $(X_{t\cup\{v\}}(d - |t| - 1), X_{t\cup\{v\}}(1))$ is a $(\frac{1}{10}, \exp(-c'k'))$-sampler.

Setting $t \in X(1)$, Corollary 5.9.11 implies

$$\deg^{(1)}(X) \geq \min\left\{ \frac{|X(1)|}{2}, \exp(-c'k') \right\}.$$

Assume $\exp(-c'k') \leq \frac{|X(1)|}{2}$ (else Condition (1) is satisfied and we are done). Under this assumption, we argue by induction that the max degree of any i-link is at least:

$$\forall i \leq \frac{d}{2} : \quad \min_{t \in X(i)} \{\deg^{(1)}(X_t)\} \geq \frac{1}{2^i} \exp(c'k'). \tag{5.27}$$

We remark the restriction on i is to enforce that links of this level are still $(\frac{1}{10}, \exp(-c'k'))$-samplers. Plugging this back into our bound on degree gives $\exp(-\Omega(k'^2)) \geq \exp(-\Omega(\max\{d, d^{2c}\}))$ as desired.

To prove Equation (5.27), observe the base case $i = 1$ is given by assumption. The inductive step simply follows from observing that the minimum number of vertices in any

i-link is exactly the minimum degree of the 1-skeleton of any $(i-1)$-link. Thus applying Corollary 5.9.11, we have:

$$\forall t \in X(i) : \deg^{(1)}(X_t) \geq \min\left\{\frac{|X_t(1)|}{2}, \exp(-c'k')\right\} \geq \frac{1}{2^i}\exp(c'k').$$

as desired.

TD/SI Case.

When X is λ-TD or η-SI, we'd like to run the same argument as above but run into a slight problem. We can still bound the kth degree of a vertex v by:

$$\deg^{(k)}(v) \geq \frac{1}{(k-1)!}\prod_{i=1}^{k-1}\min_{t\in X(i)}\left\{\deg^{(1)}(X_t)\right\},$$

but because X satisfies only exponential concentration, Equation (5.27) becomes

$$\forall i \leq \frac{d}{2} : \min_{t\in X(i)}\{\deg^{(1)}(X_t)\} \geq \frac{1}{2^i}\exp(c'\sqrt{d}).$$

This bound is now only non-trivial up to $i \leq O(\sqrt{d})$ levels of X, stopping us from proving super-exponential concentration for X's top level. Nevertheless, when $k = d^c$, plugging the above into the degree calculation gives $\exp(d) = \exp(k^{1/c})$ whenever $c \geq 1/2$, and $\exp(k\sqrt{d}) \geq \exp(k^2)$ whenever $c < 1/2$. □

5.9.3 Geometric Overlap

The geometric overlap property is a classical notion of high dimensional expansion (see e.g. [288]) which promises every embedding of the complex has a point in space hit by at least a constant fraction of the faces.

Definition 5.9.12. Let X be a d-uniform simplicial complex and let $k \leq d-1$. Let $c > 0$. We say that X has (k, c)-geometric overlap if for every $\rho : X(1) \to \mathbb{R}^k$, there exists a point

$q_0 \in \mathbb{R}^k$ such that

$$\mathop{\mathbb{P}}_{s \in X(d)} [q_0 \in conv(\rho(s))] \geq c.$$

Here $conv(\rho(s))$ is the convex hull of the points of $\rho(s) = \{\rho(v) \mid \ v \in s\}$. While we are not aware of any applications of geometric overlap in computer science, it is a popular topic in geometry and topology and related ideas (e.g. an overlap variant of the Borsuk-Ulam theorem) have seen recent use in the study of algorithmic stability [92].

Boros and Füredi [74] showed that the 2-dimensional complete complex has geometric overlap, a result which was later generalized to all dimensions by Bárány [50]. The notion was extended to general simplicial complexes by Gromov [182], who asked whether there are bounded degree complexes that have this property. Followup work by Fox, Gromov, Lafforgue, Naor, and Pach [153] gave an affirmative answer to this question both via random construction and through [290]'s Ramanujan complexes. Finally, a series of works [316, 140, 309] extended their ideas to general spectral high dimensional expanders by leveraging high dimensional expander-mixing properties. We note that Gromov also defined a closely related (but stricter) notion of topological overlap [182], which was later studied by [230, 141] among others. We focus only on the geometric setting here.

While prior works have succeeded in showing optimal geometric overlap for *random* bounded-degree complexes [153, Theorem 1.7], it has been open since Gromov's work to construct such a family *explicitly*. In this subsection, building on the approach of these prior works, we show sufficiently strong high dimensional expanders have optimal overlap, resolving this problem.

To state our result, it is useful to first understand the overlap properties of the complete complex. Let $c_{k,n}$ be the largest constant so that $\Delta_n(k+1)$ has $(k, c_{k,n})$-geometric overlap, and let $c(k) = \lim_{n\to\infty} c_{k,n}$. By [50] this limit exists and is positive. Similarly, let $c^p_{d,k,n}$ be the largest constant so that the complete d-partite complex with n-vertices on every side has $(k, c^p_{d,k,n})$-geometric overlap, and let $c^p_{d,k} = \lim_{n\to\infty} c^p_{d,k,n}$. Existence and

positivity of this limit follows from [312]. The former constant, $c(k)$, is known to be the best possible for bounded degree simplicial complexes [153, Theorem 1.6].

We show that every sufficiently strong (partite) high dimensional expanders has geometric overlap nearly matching the (partite) complete complex:

Theorem 5.9.13. *For every $\varepsilon > 0$ and $k < d \in \mathbb{N}$, there exist $\lambda, \lambda_p > 0$, and $n_0, n_{0,p} \in \mathbb{N}$ such that*

1. ***Two-sided case:*** *If X is d-uniform λ-two-sided high dimensional expander on at least n_0 vertices so that the measure on the vertices is uniform, then X has $(k, c_k - \varepsilon)$-geometric overlap.*

2. ***Partite case:*** *If X is d-partite λ_p-one-sided HDX on at least $n_{0,p}$ vertices so that the measure on the vertices is uniform, then X has $(k, c^p_{d,k} - \varepsilon)$-geometric overlap.*

We comment that this theorem applies to many of the complexes constructed in [290] (for the non-partite setting we need to take lower uniform skeletons).

The proof of Theorem 5.9.13 combines our concentration bounds with tools developed in [153]. We have not tried to find the optimal dependence on λ or on the degree of the complex. Finally before moving to the proof, we remark that the result really only relies on the weaker 'high-dimensional expander-mixing' type inequality which actually holds for the more general family of splitting-trees such as expander walks (see Section 5.10.1). Thus it is possible to obtain explicit complexes with geometric overlap approaching that of the complete (partite) complex through this more general family as well, though we omit the details.

Homogeneous tuples.

The core of [153]'s proof of geometric overlap is the following partitioning theorem. We follow their definitions and presentation precisely. A tuple of subsets $S_1, \ldots, S_{k+1} \subseteq \mathbb{R}^k$ is said to be *homogeneous* with respect to a point $q \in \mathbb{R}^k$ if either:

1. All simplices with one vertex in each of the sets $S_1, \ldots, S_{k+1}$ contain q.

2. None of these simplices contain q.

An *equipartition* of a finite set is a partition of the set into subsets whose sizes differ by at most one.

Theorem 5.9.14 ([153, Corollary 1.9])**.** *Let k be a positive integer and $\epsilon > 0$. There exists a positive integer $K = K(\epsilon, k) \geq k+1$ such that for any $K' \geq K$ the following statement holds. For any finite set $P \subseteq \mathbb{R}^k$ and for any point $q \in \mathbb{R}^k$, there is an equipartition $P = P_1 \cup \cdots \cup P_{K'}$ such that all but at most an ϵ-fraction of the $(k+1)$-tuples $P_{i_1}, \ldots, P_{i_{k+1}}$ are homogeneous with respect to q.*

Our proof adapts the idea in [153, Theorem 1.7].

Proof of Theorem 5.9.13. We split the proof into the two-sided and partite cases.

Two-sided Case.

Let $K = K(\varepsilon/16, k)$ be the constant promised in Theorem 5.9.14 and $K' \geq K$ be the smallest constant so that for any equipartition of the n-vertices to K' parts, the fraction of faces in the complete complex (over the same vertex set) that touch every part at most once is at least $1 - \frac{\varepsilon}{16}$.[19] Let $z_0 = \frac{1}{2K'}$ and let n_0 be such that for any $n \geq n_0$ and sets $A_1, A_2, \ldots, A_{k+1} \subset [n]$ of measure at least $\frac{1}{K'} - z_0$ it holds that

$$\mathop{\mathbb{P}}_{\{v_1, v_2, \ldots, v_{k+1}\} \in \Delta_n(k+1)} [\forall j = 0, 1, \ldots, d : \ v_j \in A_j] \geq \prod_{j=1}^{k+1} \mathbb{P}[A_j] - \frac{\varepsilon}{16}.$$

Let X be as in the statement and $\rho : X(1) \to \mathbb{R}^k$ an embedding. There is a point $q_0 \in \mathbb{R}^k$ so that q_0 is contained in $c_k - \frac{\varepsilon}{16}$ of the faces of $\Delta_n(k+1)$ embedded via ρ.

[19] One can verify as in [153, Theorem 1.7] that $K' = \max\{K, 16(d+1)/\varepsilon\}$ for large enough n.

Let $P_1, P_2, \ldots, P_{K'}$ be the partition we obtain from Theorem 5.9.14 with respect to ρ and the complete complex. Let

$$H := \left\{ \{P_{i_1}, P_{i_2}, \ldots, P_{i_{k+1}}\} \;\middle|\; \forall s \in E(P_{i_1}, P_{i_2}, \ldots, P_{i_{k+1}}),\ q_0 \in conv(\rho(s)) \right\}.$$

Here $E(P_{i_1}, P_{i_2}, \ldots, P_{i_{k+1}})$ are the faces $s \in X(k+1)$ such that for every $j = 1, \ldots, k+1$, $\left|s \cap P_{i_j}\right| = 1$. Recall that at most an $\frac{\varepsilon}{16}$-fraction of the tuples are not homogeneous and that at most $\frac{\varepsilon}{16}$ of the faces in $\Delta_n(k+1)$ touch a part in this partition more than once. Thus,

$$\sum_{(P_{i_j})_{j=1}^{k+1} \in H} \prod_{j=1}^{k+1} \mathbb{P}\left[P_{i_j}\right] \geq c_k - \frac{3\varepsilon}{16}. \tag{5.28}$$

On the other hand,

$$\mathop{\mathbb{P}}_{s \in X(k+1)} \left[q_0 \in conv(\rho(s))\right] \geq \sum_{\{P_{i_1}, P_{i_1}, \ldots, P_{i_{k+1}}\} \in H} \mathop{\mathbb{P}}_{s \in X(k+1)} \left[E(P_{i_1}, P_{i_2}, \ldots, P_{i_{k+1}})\right]. \tag{5.29}$$

By Proposition 5.7.2 (applied for indicators of the corresponding functions) for small enough $\lambda = \lambda(K', \varepsilon, z_0)$, for every $k+1$ parts $P_{i_1}, P_{i_2}, \ldots, P_{i_{k+1}}$ of size at least $\frac{1}{K'} - z_0$,

$$\mathop{\mathbb{P}}_{\{v_1, v_2, \ldots, v_{k+1}\} \in X(k+1)} \left[\forall j = 1, \ldots, k+1\ v_j \in P_{i_j}\right] \geq (1 - \frac{\varepsilon}{2}) \prod_{j=1}^{k+1} \mathbb{P}\left[P_{i_j}\right].$$

Combining (5.28) and (5.29) gives $\mathbb{P}_{s \in X(k+1)}\left[q_0 \in \rho(s)\right] \geq (1 - \frac{\varepsilon}{2})(c_k - \frac{\varepsilon}{4}) \geq c_k - \varepsilon$ as desired.

Partite Case.

The argument is similar. Let $K = K(\varepsilon/16, k)$ be the constant promised in Theorem 5.9.14. Let $K' \geq K$ be the smallest constant so that for any equipartition over K' parts, the fraction of faces in the complete partite complex that touch every part at most once is at least $1 - \frac{\varepsilon}{16}$.

Let $z = \frac{1}{2K'}$ be as before. As before we first identify the points of X with the complete partite complex with the same number of vertices in every side. Then we find a point q_0 such that $c^p_{d,k} - \frac{\varepsilon}{16}$ of the faces of the complete partite complex contain q_0. We find an equipartition $P_1, P_2, \ldots, P_{K'}$ of $\rho(X(1))$ to K' parts such that $1 - \frac{\varepsilon}{16}$ of the parts are homogeneous with respect to q_0.

We partition further every P_i according to the colors of the vertices to P_i^j for $j = 1, 2, \ldots, d$. Then we remove from the partition every subset P_i^j whose relative size is smaller than $\frac{\varepsilon}{16K' \cdot d^2}$. We note that by taking only the remaining sets in the partition the fraction of vertices we removed from every color is no more than $\frac{\varepsilon}{8}$ vertices in any side. This is because for every P_i, colors j that have that $\mathbb{P}\left[P_i^j\right] \leq \frac{\varepsilon}{16K' \cdot d^2}$ can account for no more than $\frac{\varepsilon}{16K' \cdot d}$ out of at least $\frac{1}{K'} - z = \frac{1}{2K'}$. Thus we have not removed more than $\frac{\varepsilon}{8d}$ vertices in total, or $\frac{\varepsilon}{8}$-fraction out of each color.

We now have subsets P_i^j that such that $\frac{\varepsilon}{16K'd^2} \leq \mathbb{P}\left[P_i^j\right] \leq \frac{3}{2K'}$. The proof now follows as in the two-sided case. □

5.9.4 Separating MLSI from Reverse Hypercontractivity

As discussed in Section 5.1, prior techniques for establishing reverse hypercontractivity relied on tensorization or bounded (Modified) Log-Sobolev constant (MLSI). To our knowledge, all known reverse-hypercontractive objects prior to this work also have bounded MLSI. In this section, we give the first separation between these analytic notions in the case where the leading hypercontractive constant C may be greater than 1 (or, alternatively, for the indicator variant). To start, recall the definition of MLSI.

Definition 5.9.15 (Modified Log-Sobolev Inequality)**.** Let M be a reversible markov chain with stationary distribution π_s. The modified log-sobolev constant of M is:

$$\rho_{\text{MLSI}} = \inf_{f>0} \frac{\langle f, (I - M) \log f \rangle}{2\text{Ent}_\pi(f^2)}$$

MLSI is classically used to bound the *mixing time* of its associated Markov chain,

$$T_{\text{mix}}(M,\varepsilon) = \min_t \max_{\pi_{init}} \|\pi_{init}^T M^t - \pi_s\|_{TV} \leq \varepsilon,$$

due to the following upper bound:

Fact 5.9.16. *The mixing time of a reversible Markov chain M with stationary distribution π is at most:*

$$T_{mix}(M,\varepsilon) \leq \frac{1}{\rho_{MLSI}} \log\left(\frac{\log \pi_*^{-1}}{\varepsilon}\right),$$

where π_* is the minimum stationary probability

High dimensional expanders (at least in the bounded degree setting), are *slow-mixing*, so they cannot have small MLSI constant. In particular, applying this fact to the Ramanujan complexes gives an infinite family of reverse-hypercontractive operators with at least inverse logarithmic dependence on the domain size.

Corollary 5.9.17. *There exist constants $\ell > 1$ and $C > 0$ and an infinite family $\{T_\rho^{(n)}\}$ of $(\frac{1}{\ell}, \frac{\ell}{1-\ell}, C)$-reverse hypercontractive operators with at most inverse logarithmic MLSI:*

$$\rho_{MLSI} \leq C_2 \frac{\log^3 \log(\pi_*)}{\log \pi_*}$$

Proof. Set $\rho = \frac{1}{2}$, let c be as in Theorem 5.6.3, and let $\{X^{(n)}\}$ be the family of $k = k(n)$-uniform cut-off coset complexes promised in Theorem 5.9.4 for $k = \log\log n$ and $\lambda = 2^{-k}$, and recall that

1. The degree of every vertex is at most $2^{C_2 k^3}$ for some universal constant $C_2 > 0$
2. The stationary distribution of $T_{1/2}$ on X is uniform

By Theorem 5.6.3, there exist (universal) constants C, ℓ such that the noise operator $T_{1/2}$ is $(\frac{1}{\ell}, \frac{\ell}{1-\ell}, C)$-reverse hypercontractive for every complex in the family. On the other hand,

we will argue that:

$$T_{\text{mix}}\left(T_\rho, \frac{1}{2}\right) \geq \min\left\{\frac{2^k}{4}, \frac{C_2' \log(n)}{k^3}\right\},$$

which combined with Fact 5.9.16 completes the result.

Define the ε-support of a distribution

$$\text{supp}_{\varepsilon}^{\geq}(\pi) \coloneqq \{x : \pi(x) \geq \varepsilon\}.$$

Setting $\varepsilon = \frac{1}{4n}$, observe that any distribution π with $|\text{supp}_{\varepsilon}^{\geq}(\pi)| < \frac{n}{4}$ has $\|\pi - \pi_s\|_{TV} \geq \frac{1}{2}$. We will argue this holds for $\pi_{\text{init}} T_{\frac{1}{2}}^t$ whenever π_{init} is any indicator and t is sufficiently small.

To see this, we separate out the bounded-degree and stationary components of $T_{\frac{1}{2}}$ as:

$$T_{\frac{1}{2}}^+ \coloneqq T_{\frac{1}{2}} - 2^{-k}\Pi$$

where Π is the stationary operator, and note that $T_{\frac{1}{2}} = T_{\frac{1}{2}}^+ + 2^{-k}\Pi$. Observe that after t steps, we can write any initial distribution π as:

$$\pi^T T_{\frac{1}{2}}^t = \pi^T (T_{\frac{1}{2}}^+ + 2^{-k}\Pi)^t = \pi^T (T_{\frac{1}{2}}^+)^t + (1 - (1 - 2^{-k})^t)\pi_s$$

For large enough k, as long as $t \leq 2^{-k/4}$, the latter term contributes strictly less than $\frac{1}{4}$ to each coordinate. On the other hand for π an indicator, the first term contributes positive weight to at most $deg(T_{\frac{1}{2}^+})^t$ coordinates, so setting t such that $deg(T_{\frac{1}{2}^+})^t \leq \frac{n}{4}$ completes the result. □

5.9.5 It Ain't Over Till It's Over

Friedgut and Kalai's 'It Ain't Over Till It's Over' Theorem is a classical result in social choice which states that even after taking a significant random restriction, any balanced function still has some uncertainty with high probability (in other words, it is

a tail bound on the random restriction operator). The precursor to this Theorem was a version for the noise operator proved in [302, Appendix C]. We prove the result holds for any LUS, which implies the same for any c-nice complex.

Theorem 5.9.18. *For any $\rho \in (0,1)$ and $\tau > 0$, let X be a d-uniform τ-LUS for d sufficiently large. Then there exists a constant $q = q(\rho, \tau)$ such that for any $\delta > 0$ and $f : X(d) \to [0,1]$ with density $\mu = \mathbb{E}[f]$:*

$$\mathbb{P}[T_\rho f > 1 - \delta] \leq c_{\mu,1}\delta^q$$

where $c_{\mu,1} = \left(\frac{8(1+\mu)^{q+1}}{(1-\mu)^{q+2}}\right)^{\frac{1}{q}}$, and

$$\mathbb{P}[T_\rho f < \delta] \leq c_{\mu,2}\delta^q$$

where $c_{\mu,2} = \left(\frac{8(2-\mu)^{q+1}}{\mu^{q+2}}\right)^{\frac{1}{q}}$.

Proof. We prove the former. The latter follows from considering $1 - f$. Let T of the set of elements whose neighborhoods are dense in f:

$$T \coloneqq \{x : T_\rho f(x) \geq 1 - \delta\},$$

and consider the indicator cut-off $h = \mathbf{1}_{f \leq \frac{1+\mu}{2}}$. We will argue that on the one hand the correlation of T and h cannot be large, since when $x \in T$, $y \sim_\rho x$ must be mostly above the cutoff. On the other hand, the correlation is at least some power of $\mu(T)$ by reverse hypercontractivity, so $\mu(T)$ must be small.

Formally, observe that for any $x \in T$, Markov's inequality gives that $T_\rho h(x) \leq \frac{2\delta}{1-\mu}$ so

$$\mathbb{E}[1_T T_\rho h] \leq \mu(T)\delta\frac{2}{1-\mu}.$$

On the other hand, we also have by Markov that $\mathbb{E}[h] \geq \frac{1-\mu}{1+\mu}$, so reverse hypercontractivity

for indicators (Theorem 5.6.4) implies there exists some ℓ such that:

$$\mathbb{E}[1_T T_\rho h] \geq \frac{1}{4}\left(\mu(T)\frac{1-\mu}{1+\mu}\right)^\ell$$

Combining the equations and setting $q = \ell - 1$ gives the desired result. □

We note we have made no attempt to optimize the constants, and it may be possible to improve dependence on μ. [302] combine a variant of Theorem 5.9.18 with the invariance principle to prove the classical form of the conjecture with respect to random restrictions. It is not clear what the correct form of this conjecture should be for HDX, or even for product spaces.[20] We leave this as an open question.

5.9.6 A Frankl–Rödl Theorem

The Frankl–Rödl Theorem [154] is a powerful result from extremal combinatorics that states that the independence number of the graph whose vertices are $\{0,1\}^n$ and whose edges are given by fixed weight intersection k (typically for some $k = \Theta(n)$) is at most $\exp(-\Omega(k))|V|$. Benabbas, Hatami, and Magen [61] proved a variant of this Theorem via reverse hypercontractivity on the cube, later leading to several applications in hardness of approximation [229]. Their method, which also bounds classical properties such as the chromatic and dominating set numbers, is based on the following claim:

Claim 5.9.19. Let $p_0, q_0 > 0$ and let $G = (V, E)$ be a weighted graph with minimum vertex weight q_0 such that for every $A, B \subset V$ of relative size at least p_0:

$$\mathbb{P}_{(s_1,s_2)\sim E}[s_1 \in A, s_2 \in B] > 0.$$

Then the maximal independent set in G has size at most $2p_0 + q_0$, the minimal dominating set has relative size at least $1 - (2p_0 + q_0)$, and the chromatic number is at least $\frac{1}{2p_0+q_0}$.

[20] We remark it should be possible to give a version of the result for suitably 'global' functions via tools of [246, 187, 39], but it is not clear such a notion would have significant qualitative meaning

As a corollary we have that the down-up walks in any of the discussed LUS-complexes in Section 5.6.3 have maximal independent sets of relative size at most $\exp(-\Omega(k))$, a chromatic number of at least $\exp(\Omega(k))$ and minimal dominating sets of size at least $1 - \exp(-\Omega(k))$. As discussed in the introduction, we remark that these graphs are not quite the correct analog of Frankl–Rödl, since they have edges with intersection up to some fixed γk instead of γk exactly. Benabbas, Hatami and Magen [61] handle this by analyzing the closeness of the exact intersection versus noise operator on the cube, using tools from Fourier analysis. It would be interesting to see if a variant of this result holds on HDX. If this could be shown generally, then our reverse hypercontractivity theorem would recover Frankl–Rödl up to constants in the exponent.

Proof of Claim 5.9.19. We prove the statement for independent sets. The chromatic number statement follows since every coloring is a partition to independent sets, and the statement for dominating sets follows since the complement of any dominating set is an independent set. Let $I \subseteq G$ be an independent set and assume toward contradiction that $\mathbb{P}[I] \geq 2p_0 + q_0$. We partition I into I_1, I_2 such that $\mathbb{P}[I_1], \mathbb{P}[I_2] \geq p_0$ (we can do so by greedily adding vertices to I_1 until its probability is between p_0 and $p_0 + q_0$). Then by assumption $\mathbb{P}_{(s_1,s_2)\sim E}[s_1 \in I_1, s_2 \in I_2] > 0$, which gives the desired contradiction. □

Remark 5.9.20. Although the graphs we discussed in this paper have self loops. The claim above continues to hold after removing these self loops, as is evident from the proof.

5.10 Codes and Splitting Trees

Error correcting codes, and in particular the powerful notions of *list-decoding* and *local-testability*, are among the earliest successful applications of inclusion samplers [213, 212, 215] and HDX [121, 10, 221, 315, 117]. In this section we explore the implications of our tools and related ideas over the weaker family of 'splitting trees' in this classical setting. Our main application is the first construction of constant rate codes over large

alphabets which simultaneously have 1) near-optimal distance 2) list-decodability and 3) local-testability. At the end of the section we introduce a (conjectural) HDX-based approach toward 'lossless' distance-amplification of LTCs that experiences no decay in soundness.

5.10.1 Splitting Trees

We start with a fairly substantial detour into the world of splitting trees, which are a weakening of high dimensional expanders introduced in [9] which only requires certain patterns of swap walks to expand. We will prove a high-dimensional expander mixing lemma and some basic sampling properties for general splitting trees that will be useful for our applications to coding theory.

Splitting Preliminaries

Given a binary tree T, let $\mathcal{L}(T)$ denote T's leaves and $\mathcal{I}(T)$ its internal nodes. We drop T from the notation when clear from context. For every internal node $v \in \mathcal{I}$, denote the left and right children of v by ℓ_u and r_u respectively.

Definition 5.10.1 (Ordered Binary Tree)**.** A d-uniform ordered binary tree is a pair (T, ρ) where T is a rooted binary tree with d leaves, and $\rho : T \to [d]$ is a labeling such that:

1. The label of every leaf $u \in \mathcal{L}$ is $\rho(u) = 1$.
2. For every interior $u \in \mathcal{I}$, it holds that $\rho(u) = \rho(\ell_u) + \rho(r_u)$.

We call such a ρ well-ordered with respect to T.

A complex is called splittable if it can be decomposed by an ordered tree where every non-leaf node corresponds to an expanding swap walk.

Definition 5.10.2 (Splitting tree)**.** Let $\gamma > 0$. A d-uniform γ-*splitting tree* is a triple (X, T, ρ) where X is a d-uniform simplicial complex and (T, ρ) is an ordered binary tree

such that for every $u \in \mathcal{I}(T)$:

$$\lambda(S_{\rho(\ell_u),\rho(r_u)}) \leq \gamma,$$

We will also work with the corresponding partite notion of splitting trees.

Definition 5.10.3 (Partite Ordered Binary Tree)**.** A d-uniform partite ordered binary tree is a pair (T, ρ) where T is a rooted binary tree with d leaves, and $\rho : T \to \mathbb{P}\{[d]\} \setminus \{\emptyset\}$ is a labeling such that:

1. $\rho|_{\mathcal{L}}$ is a bijection to the singletons of $[d]$.
2. For every interior $u \in \mathcal{I}$, it holds that $\rho(u) = \rho(\ell_u) \cup \rho(r_u)$.

We remark that when convenient, we may use a general set S of size $d+1$ as the domain of ρ (generally corresponding to a labeling of the "parts" of an associated d-uniform partite complex).

Definition 5.10.4 (Tuple splitting tree)**.** Let $\gamma > 0$. A d-uniform γ-*tuple splitting tree* is a triple (X, T, ρ) where X is a d-uniform simplicial complex and (T, ρ) is an ordered partite binary tree such that for every non-leaf vertex $u \in T$:

$$\lambda(S_{\rho(\ell_u),\rho(r_u)}) \leq \gamma$$

Since all swap walks on two-sided HDX expand, these give rise to a basic family of splitting trees.

Example 5.10.5 ([9, 109, 187])**.** Let $\lambda \geq 0$ and let X be a d-uniform two-sided λ-high dimensional expander. Then for any binary tree T and labeling ρ that satisfies the first and second item in Definition 5.10.2, (X, T, ρ) is a $d\lambda$-splitting tree.

An easy consequence of [109] is that partite one-sided HDX also give rise to tuple-splitting trees.

Example 5.10.6. Let $\lambda \geq 0$ and let X be a d-uniform two-sided λ-high dimensional expander. Then for any binary tree T and labeling ρ that satisfy the first and second item in Definition 5.10.4, (X, T, ρ) is a $d^2\lambda$-tuple splitting tree.

Finally, [10, Corollary 9.18] also prove that walks on expander graphs are splittable. Let us define the complex formally. Let $G = (V, E)$ be a graph and $k \in \mathbb{N}$. Let W_G be the following k-partite simplicial complex

$$W_G(1) = V \times [k],$$

$$W_G(k) = \{\{(v_1, 1), (v_2, 2), \ldots, (v_k, k)\} \mid \ (v_1, v_2, \ldots, v_k) \text{ is a walk in } G\}.$$

We choose a face in $W_G(k)$ by choosing a vertex $v_1 \in V$ according to the stationary distribution over vertices in G and then taking a $(k-1)$-length random walk on that vertex.

Example 5.10.7. Let G be a λ-expander and T, ρ an ordered partite tree. If every $u \in \mathcal{I}$ satisfies:

$$\max \rho(\ell_u) < \min \rho(r_u),$$

then (W_G, T, ρ) is a λ-splittable tree.

Operations on splitting trees

We will need several further operations on splitting trees for our analysis: a *pruning* method that allow us to analyze higher dimensionalT faces, and a *partitification* technique that allows us to reduce standard splitting trees to the tuple setting.

We start with the latter. Recall X's partitification is the complex

$$P(k) \coloneqq \left\{(s, \pi) \coloneqq \left\{(s_{\pi(1)}, 1), (s_{\pi(2)}, 2), \ldots, (s_{\pi(k)}, k)\right\} \mid \ s \in X(k), \pi \in S_k\right\},$$

This notion extends naturally to splitting trees.

Definition 5.10.8. Let (X, T, ρ) be a d-uniform λ-splitting tree, and $\rho' : T \to [d]$ a well-ordered labeling such that $|\rho'(u)| = \rho(u)$ for every $u \in T$. The *ρ'-partitification* of (X, T, ρ) is defined to be the tuple splitting tree (P, T, ρ').

Any un-ordered splitting tree can be embed into a tuple splitting tree by taking any partitification.

Claim 5.10.9. Let (X, T, ρ) be a λ-splitting tree. Every ρ'-partitification of (X, T, ρ) is a λ-tuple splitting tree.

Proof. It suffices to show for every internal node $u \in \mathcal{I}$ with $\rho'(\ell(u)) = A$ and $\rho'(r(u)) = B$ that $S_{A,B}(P)$ is a λ-bipartite expander. This follows from the observation that $S_{A,B}(P)$ is isomorphic to the bipartite graph with left vertices $L = X(|A|) \times [|A|!]$, right vertices $R = X(|B|) \times [|B|!]$, and edges $(s, i) \to (t, j)$ for all $i \in A, j \in B$, and (s, t) such that $s \sim t$ in $S_{\rho(\ell_u),\rho(r_u)}$.[21] It is a standard fact that such a bipartite construction inherits its expansion from $S_{\rho(\ell_u),\rho(r_u)}$, and is therefore a λ-bipartite expander. □

Finally, we note that tuple splitting trees can be pruned to create new splitting trees. Let (X, T, ρ) be a tuple splitting tree with k leaves. Let $u \in \mathcal{I}$ be an internal node whose label is $F = \rho(u)$. Let $P = P(X, F)$ be the partite complex with vertices $P(1) = X^{[k]\setminus F}(1) \cup X[F]$ and top level faces $t \cup \{s\}$ such that $t \cup s \in X(k)$, inheriting the corresponding densities. In other words, we replace the vertices of X in $X^F(1)$ with the set of faces $X[F]$.

Let $T(u)$ be the sub-tree of T rooted by u. Let $T' = (T \setminus T(u)) \cup \{u\}$. Let $\rho' : T' \to ([k] \setminus F) \cup \{F\}$ be the labeling that replaces the subset F with a single replacement index f:

$$\rho'(v) = \begin{cases} (\rho(v) \setminus F) \cup \{f\} & F \subseteq \rho(v) \\ \rho(v) & otherwise \end{cases}.$$

[21] Formally, the edge $(s, i) \to (t, j)$ also inherits the weight $\frac{\mathbb{P}(st)}{|A||B|}$ from $S_{\rho(\ell_u),\rho(r_u)}$.

Definition 5.10.10 (Pruning Trees)**.** Let (X, T, ρ) be a λ-tuple splitting tree with k leaves. Let $u \in T$ be a non-leaf node whose label is $F = \rho(u)$. The u-pruning of the tree is the triple $(P(X, F), T', \rho')$.

Claim 5.10.11. Let (X, T, ρ) be a λ-tuple splitting tree. Let $u \in T$ be a non-leaf node. Then the u-pruning is also a λ-splitting tree. $\square$

The proof follows directly from the definitions and we therefore omit it.

Finally, it will be convenient to introduce notation for *sequential pruning*, that is given a collection of disjoint color sets $\mathcal{F} = \{F_1, \ldots, F_m\}$, we denote the sequential pruning of X by $\mathcal{F}$ as $P(X, \mathcal{F}) := P(...P(X, F_1), F_2), F_m)$. Note that the resulting complex is independent of the order of $\mathcal{F}$, so $P(X, \mathcal{F})$ is well-defined. One can similarly define a sequentially pruned splitting tree as above. It is immediate from Claim 5.10.11 that these are also λ-tuple splitting.

Hitting set and Bias amplification

In this section, we prove two basic 'sampling-type' properties for splitting trees: hitting set and bias amplification. We start with the former. Recall a complex is (γ, i)-hitting if for any $A \subset X(0)$:

$$\mathop{\mathbb{P}}_{\sigma \in X(i)}[\sigma \subset A] \leq \mu(A)^i + \gamma$$

Proposition 5.10.12 (Hitting Set)**.** *Any depth D λ-(tuple) splitting tree with k leaves is $(3^D\lambda, k)$-hitting.*

Splitting trees also exhibit strong bias amplification, used to build balanced error correcting codes.

Proposition 5.10.13 (Bias-Amplification)**.** *Let (X, T, ρ) be a λ-tuple splitting tree for $\lambda < \frac{1}{16}$. For any $0 < \varepsilon < \frac{1}{4}$ and family of mean ε functions $\{f_i : X[i] \to \{\pm 1\}\}_{i\in[k]}$, $\prod f_i$ is a $\{\pm 1\}$-valued function with bias at most:*

$$\left| \mathop{\mathbb{E}}_{a\in X(k)} \left[\prod_{i=1}^{k} f_i(a_i) \right] \right| \leq \varepsilon^k + 2\lambda \tag{5.30}$$

While Proposition 5.10.12 and Proposition 5.10.13 are elementary, to the authors' knowledge they do not appear in the literature. Prior works on HDX (e.g. [124, 10]) use weaker variants of these lemmas which require super-exponential overhead to achieve a given distance or bias, while the above only requires quasi-polynomial overhead. We discuss this in greater detail in Section 5.10.

These results are immediate corollaries of a functional variant of [109]'s "high dimensional expander-mixing lemma" (HD-EML) extended to splitting trees. We give the statement here and include a proof in the appendix for completeness along with the proofs of Proposition 5.10.12 and Proposition 5.10.13.

Theorem 5.10.14 (high dimensional expander Mixing Lemma)**.** *Let (X, T, ρ) be a depth D λ-tuple splitting tree with k leaves. Denote by d_i the depth of the leaf labeled i. Then for any family of functions $\{f_i : X[i] \to \mathbb{R}\}_{i\in[k]}$:*

$$\left| \mathop{\mathbb{E}}_{a\in X(k)} \left[\prod_{i=1}^{k} f_i(a_i) \right] - \prod_{i=1}^{k} \mathop{\mathbb{E}}_{a_i \in X[i]} [f_i(a_i)] \right| \leq 3^D \lambda \prod_{i=1}^{k} \|f_i\|_{2^{d_i}}. \tag{5.31}$$

If (X, T, ρ) is instead a depth D standard λ-splitting tree, we take $\{f_i : X(1) \to \mathbb{R}\}_{i\in[k]}$ and have:

$$\left| \mathop{\mathbb{E}}_{a\in X(k),\pi\in S_k} \left[\prod_{i=1}^{k} f_i(a_{\pi(i)}) \right] - \prod_{i=1}^{k} \mathop{\mathbb{E}}_{a_i \in X(1)} [f_i(a_i)] \right| \leq 3^D \lambda \prod_{i=1}^{k} \|f_i\|_{2^{d_i}}. \tag{5.32}$$

In the future, we state our results only for the partite setting unless there is a

substantial difference in proof or parameters for the results. Otherwise, we collect analog statements for the unordered case at the end of Section 5.15 for completeness.

We close the subsection with a few remarks. First it should be noted that for the special case of expander walks, better hitting set and bias amplification methods are known (indeed one can achieve polynomial overhead vs our quasipolynomial bound above). Nevertheless, it is useful to have these lemmas on structures like HDX that admit stronger properties like agreement tests. We remark it is also actually possible to use the high dimensional expander mixing lemma for splitting trees directly to prove a different variant of Chernoff-Hoeffding by reduction to the complete complex. However, this results in exponentially worse parameters than our arguments in Section 5.5, and we will not need the result in the following.

5.10.2 Coding Preliminaries

A q-ary error correcting code C over a finite size-q alphabet Σ is a subset $C \subset \Sigma^n$, where $n \in \mathbb{N}$ is called the *block length*. We say a code is *linear* if $\Sigma = \mathbb{F}_p$ for prime power p and C is a subspace, and $\mathbb{F}_p$-*linear* if $\Sigma = \mathbb{F}_p^k$ for some $k \in \mathbb{N}$ and C is a subspace of $(\mathbb{F}_p^k)^n$.

The *rate* of a code C, which measures its overhead, is $r = \frac{\log_{|\Sigma|}(|C|)}{n}$, and its *distance*, which captures the codes error correction capability, is the minimum normalized hamming distance between any two codewords

$$d \coloneqq \min_{c_1, c_2 \in C} \left\{ \frac{|\{i \in [n] : c_1(i) \neq c_2(i)\}|}{n} \right\}.$$

We remark that for any $\mathbb{F}_p$-linear code, the distance is the minimum weight of a non-zero codeword. For ease of notation, we will call a q-ary code with rate r and distance d an $(r, d)_q$-code.

We will be interested in so-called "good" families of codes, which are infinite families with growing block length and constant rate and distance. We will focus in particular in

this section on building large distance $\mathbb{F}_2$-linear codes over $\Sigma = \mathbb{F}_2^k$. In this regime there is a classical upper bound due to McEliece, Rodemich, Rumsey, and Welch [295] stating that good q-ary codes have distance at most $1 - \frac{1}{q}$, or more exactly:

Theorem 5.10.15 (MRRW Bound [295]). *For all $\varepsilon > 0$ and $q \in \mathbb{N}$, any family of q-ary codes with distance $1 - \frac{1}{q} - \varepsilon$ has rate at most:*

$$r \leq O\left(\varepsilon^2 \log(1/\varepsilon)\right).$$

Given a code C, a *unique decoding algorithm* with radius $\gamma \in [0, 1/2]$ is a (possibly randomized) algorithm $Dec : \Sigma^n \to C$ which given any $w \in \Sigma^n$ such that $dist(w, C) \leq \gamma$ outputs the uniquely closest $y \in C$ to w. Note that unique decoding is only possible up to radius $\lfloor \frac{nd-1}{2n} \rfloor \leq 1/2$, even with large alphabet, as there may be no unique closest codeword beyond this point.

List-decoding is the natural extension of unique decoding beyond the $\frac{1}{2}$ barrier. A code is said to be list-decodable with radius t if there is an algorithm which outputs a *list* of all codewords within radius t of the given word. We first give the combinatorial definition, which simply bounds the number of codewords around any fixed point.

Definition 5.10.16 (Combinatorial List-decoding). For $\gamma \in (0, 1)$ and $L \in \mathbb{N}$, a code C is said to be (γ, L) list-decodable if for every $z \in Y$, there are at most L codewords $x \in C$ such that $dist(x, z) < \gamma$.

In this section, we will be interested in *efficient algorithms* for list-decoding, typically a much more challening task than the existential definition.

Definition 5.10.17 (Efficient List-decoding). For $\delta > 0$, $\gamma \in (0, 1)$ and $L \in \mathbb{N}$, a family of codes $\{C_n\}$ is said to be efficiently (γ, L) list-decodable with confidence δ if it is combinatorially (γ, L) list-decodable and there exists a (randomized) polynomial time[22]

[22]Here we mean polynomial in the block length.

algorithm DEC: $\Sigma^n \to P(C)$ such that for every word $w \in \Sigma^n$:

$$\text{DEC}(w) = \{c \in C : \text{dist}(c, w) < \gamma\}$$

except with probability δ.

Another classical approach to handling the high error regime is *local testability*. A code C is said to be locally testable if there is a (randomized, non-adaptive) algorithm $\mathcal{T}_C$ which on input of a word $w \in \Sigma^n$, queries constantly many symbols of w and rejects with high probability if it is far from the code. In practice, if one receives such a word, instead of trying to list decode they could simply request re-transmission. More formally, we consider the following standard notion of 'strong' local testability:

Definition 5.10.18 (Locally Testable Code)**.** For any $s > 0$ and $q \in \mathbb{N}$, a code C of blocklength n is said to be (q, s)-locally testable if there exists a function $\mathcal{T}_C : [n]^q \times \Sigma^q \to \{0, 1\}$ and a distribution $\mathcal{D}$ over $[n]^q$ such that on any word $w \in \Sigma^n$, the pair $(\mathcal{T}_C, \mathcal{D})$ (called the tester) satisfies:

1. **Completeness:** If $w \in C$, then $\mathbb{P}_{J \sim \mathcal{D}}[\mathcal{T}_C(J, w_J) = 1] = 1$
2. **Soundness:** $\mathbb{P}_{J \sim \mathcal{D}}[\mathcal{T}_C(J, w_J) = 0] \geq s \cdot \text{dist}(w, C)$

We call q the locality, and s the soundness.

Typically we will simply think of $\mathcal{T}_C$ as a randomized algorithm and write $\mathcal{T}_C(w)$ to denote its application on the word w, dropping the distribution of the tester when clear from context. We will work with testers that have one additional constraint: their queries should be marginally uniform.

Definition 5.10.19 (Uniform LTC)**.** $(\mathcal{T}, \mathcal{D})$ is called *uniform* if every marginal of $\mathcal{D}$ is uniform over $[n]$.

We remark that focusing on uniform LTCs is not particularly restrictive—since codes do not typically have 'preferred' coordinates, most natural constructions are uniform.

Finally, though not strictly necessary, we will assume throughout that all complexes in this section are "homogenous" subsets $X \subset [n]^k$, meaning that their marginal distribution over each part is uniform over $[n]$. We remark that in the context of high dimensional expanders, any partite 'type-regular' complex can be homogenized in this fashion while maintaining expansion and bounded degree (see [159] for details). As most known constructions of HDX are type-regular, this is not a particularly restrictive assumption, and it is also known to hold for other classical splitting trees such as expander walks [10].

5.10.3 The ABNNR Construction

In the previous section, we proved several 'agreement theorems' on simplicial complexes. These well-studied tests are actually very closely related to both local testability and list decoding of a particular family of codes called *direct product codes*, corresponding exactly to the set of 'global functions' discussed in Section 5.8. At a broader level, these are themselves a special case of a well-studied family of codes introduced in the seminal work Alon, Bruck, Naor, Naor and Roth [16] called the 'ABNNR'-construction. We focus on the case of $\mathbb{F}_2$ for simplicity, though our results extend naturally to larger fields.

Definition 5.10.20 (the ABNNR Construction)**.** Given a right k-regular bipartite graph $G = (L, R, E)$ and a base code $C \subseteq \mathbb{F}_2^L$, the "ABNNR-Encoding" of C with respect to G is the code $\mathrm{Im}(E_G(C))$ where E_G is an "encoding" map $E_G : \mathbb{F}_2^{|L|} \to (\mathbb{F}_2^k)^{|R|}$ defined by setting for each $c \in C$ and $v \in R$:

$$E_G(c)_v = c|_{N(v)}.$$

In other words, each right vertex concatenates the symbols of its neighbors to create a new code over R. One can check that in Section 5.8, the set of global functions

is exactly the ABNNR encoding over the inclusion graph $(X(j), X(k))$. For a general k-uniform complex X and $j \leq i \leq k$, we write $E_X^{(j,i)}$ to mean the ABNNR-encoding on the vertex inclusion graph $(X(i), X(j))$. In the special case where $i = k$ and $j = 1$, we write just E_X to denote the ABNNR-encoding over the inclusion graph $([n], X(k))$, where inclusion is viewed taking $X(k) \subseteq [n]^k$ as 'ordered tuples' of $[n]$, rather than as faces of a simplicial complex. This is a minor difference, but is a bit more convenient in the setting of amplifying base codes.

Recent years have seen a great deal of work on instantiating the ABNNR construction on high dimensional expanders and splitting trees [124, 121, 10, 221, 65, 222], in particular in the context of efficient list-decoding. Building on tools from these works, we give the first family of large alphabet codes with near-optimal distance that are simultaneously locally testable and list-decodable.

Theorem 5.10.21. *For all large enough $k \in \mathbb{N}$ and all $\varepsilon > 0$, there exists an explicit family of $\mathbb{F}_2$-linear ABNNR-Codes that are simultaneously:*

1. $(\varepsilon^{O(k)}, 1 - 2^{-k} - \varepsilon)_{2^k}$*-codes*
2. $(O(1), \frac{O(1)}{k \log(1/\varepsilon)})$*-locally testable.*
3. $(1 - 2^{-\Omega(k)}, 2^{O(k)})$*-efficiently list-decodable with confidence $2^{-\Omega(n-k)}$ whenever $\varepsilon \leq 2^{-\Omega(k)}$.*

We remark that very recently, Jeronimo, Srivastava, and Tulsiani [222] gave a direct list-decoding algorithm for the LTCs of [117] (which have some small constant distance $d \ll \frac{1}{2}$). To our knowledge, these are the only other examples of codes that are simultaneously c^3-LTC and efficiently list-decodable.

The remainder if this section is split into three parts focused respectively on distance amplification, local-testability, and list-decoding of the ABNNR construction, and a short fourth section devoted to a conjectural approach to removing the soundness

decay in Theorem 5.10.21. Each section is independent, containing both the corresponding component of Theorem 5.10.21 as well as more general results within the framework. To facilitate this structure we first give a brief proof overview of Theorem 5.10.21 that doubles as a roadmap for the section.

Proof Overview.

The high level idea behind Theorem 5.10.21 is simple: starting with a binary locally testable 'base code' C with distance near $\frac{1}{2}$, we will argue that the ABNNR-encoding of C on any k-partite splitting tree[23] has distance near $1 - 2^{-k}$ (Corollary 5.10.24), is list-decodable (Theorem 5.10.32), and maintains local testability (Theorem 5.10.25). Instantiating this framework on *expander walks* (the sparsest known family of splitting trees) gives codes with the claimed parameters. We give a brief description of each step.

Distance amplification on splitting trees is fairly elementary, and follows from the hitting-set lemma. This is a standard argument: one observes that the two encoded symbols $E(x), E(x')$ differ in any face which hits the vertex set $1_{x \neq x'}$. Since by assumption the base code has distance roughly $1/2$, the hitting set lemma promises this set is hit except with roughly 2^{-d} probability (see Corollary 5.10.24). Theorem 5.10.21 actually uses a slightly stronger distance amplification lemma for expander-walks and is given in Corollary 5.10.30. The weaker amplification lemma (in particular its bias variant) appears later in the proof as a sub-component of the list-decoding algorithm.

List-Decoding the ABNNR construction, despite being heavily studied on HDX and splitting trees, is surprisingly tricky. Roughly speaking there are two main issues. First, known list-decoding algorithms for this setting are actually either for an $\mathbb{F}_2$-valued variant of ABNNR called the *direct sum* construction [10, 221, 65], or only hold on *two-sided* HDX [10, 121] (which have exponentially worse rate). To handle this, we build a reduction from list-decoding the ABNNR-encoding on X to list-decoding the direct sum encodings

[23]Formally for list-decoding actually require a slightly stronger notion called 'complete' splittability, c.f. Definition 5.10.31.

of its projected sub-complexes X^F. The second issue with this strategy is that known list-decoding algorithms for direct sum [221, 65] actually requires the base code to be ε-*biased*, a stronger condition than simply having distance near $\frac{1}{2}$. We will discuss how to construct such base codes shortly. We instantiate this framework in Corollary 5.10.35 to prove the list-decoding component of Theorem 5.10.21.

Local Testability of the ABNNR construction, to our knowledge, has not been studied outside the special case of agreement testing. Our test follows a very simple strategy. We first check whether the given word f is a direct product (in the sense of Section 5.8) via a simple 2-query agreement test: pick a random vertex $v \in [n]$, and independently $s, s' \supset v$, and check whether f_s and $f_{s'}$ match on v. If this test passes, we then simulate the tester on the 'direct product decoding' of f. That is, for every location the base tester queries, we pick a random face including that vertex and feed the tester its value in the word f. We prove this procedure is sound in Theorem 5.10.25, and instantiate the process in Corollary 5.10.30 to prove the local testability component of Theorem 5.10.21.

The Base Code for our construction, as discussed above, must be an ε-biased c^3-LTC. c^3-LTCs with distance near $\frac{1}{2}$ are now known due to the breakthrough works of [117, 315] combined with LTC distance-amplification techniques of [268], but the latter amplification causes several core issues that make the tecnique largely useless for our setting: 1) the resulting codes make poly$(1/\varepsilon)$ queries, 2) may not be ε-biased, and 3) are heavily randomized. We remedy these issues via the ABNNR-construction itself. In particular, starting with the base LTC of [117], we encode it into $\log(1/\varepsilon)$-uniform walks on an expander graph using ABNNR, and then concatenate with the Hadamard code. This code is ε-biased roughly because concatenating with the perfectly balanced Hadamard code converts the good distance of the ABNNR-encoding into good bias, and the resulting code is locally testable using the Hadamard's classical local decoding and local testing algorithms. We formalize this step in Proposition 5.10.29

Finally we summarize the construction and proof below:

Code underlying Theorem 5.10.21:

1. Start with the base c^3-LTC C of [117]
2. Amplify C's distance via expander walks
3. Reduce alphabet and bias by concatenating with Hadamard
4. Re-amplify the resulting code via a second set of expander walks.

Proof of Theorem 5.10.21. Bias, rate, and local-testability are proven in Corollary 5.10.30. List-decodability is proved in Corollary 5.10.35. □

5.10.4 Distance Amplification

The ABNNR construction is one of the most classical distance amplification methods in coding theory. The standard result to this effect is that instantiating the framework on a (bounded degree) additive sampler amplifies distance while maintaining rate up to a constant factor.

Lemma 5.10.22 ([16])**.** *Let $G = (L, R, E)$ be a right k-regular bipartite graph such that (R, L, E) is an (α, β)-sampler. If $C \subset \mathbb{F}_2^{|L|}$ is a code of distance α and rate r, then $E_G(C)$ is a $(r\frac{|L|}{|R|}, 1-\beta)_{2^k}$-code.*

This observation was used in [124, 121] to give distance amplification for codes defined on HDX. Our sampling lemmas imply an exponential improvement in the rate-distance trade-off of such codes. We choose a fairly arbitrary setting of our parameters for simplicity, but any setup with good sampling works:

Corollary 5.10.23 (Distance Amplification on HDX)**.** *For any $\alpha, r, c > 0$, let C be a base code of distance α and rate r. Let X be a d-uniform c-nice complex. Then for any*

$i < j \leq d$ and $G = (X(i), X(j))$ the ABNNR code $E_X^{i,j}(C)$ is an $(r', d')_{|\Sigma|}$-code for

$$r' = r\frac{|X(i)|}{|X(j)|}, \qquad d' = 1 - \frac{1}{\alpha}\exp\left(\Omega_c\left(\alpha^2\frac{j}{i}\right)\right), \qquad |\Sigma| = 2^{\binom{j}{i}}.$$

Note that when $i = 1$, Corollary 5.10.23 has alphabet size $|\Sigma| = 2^j$ and distance $1 - \text{poly}(|\Sigma|^{-1})$, optimal up to polynomial factors by Theorem 5.10.15. However, in this special case, it is actually possible achieve optimal amplification directly via the hitting set lemma (Proposition 5.10.12).

Corollary 5.10.24 (Distance Amplification for Splitting Trees)**.** *For any $\alpha, r > 0$, let C be a base code of distance α and rate r. If X is a k-uniform λ-(tuple)-splitting tree of depth D, then $E_X(C)$ is an $(r', d')_{|\Sigma|}$-code for*

$$r' = r\frac{n}{|X(k)|}, \qquad d' = 1 - (1-\alpha)^k - 3^D\lambda, \qquad |\Sigma| = 2^k$$

Moreover when $\alpha = \frac{1}{2}(1-\varepsilon)$, then $d' \geq 1 - \frac{1}{2^k}(1+\varepsilon)^k - 3^D\lambda$.

Proof. Rate and alphabet size are immediate from construction. Let $x, x' \in C$ be any two distinct non-zero codewords. We wish to bound the distance between their encodings $E_X(x)$ and $E_X(x')$, which is exactly the hitting probability of $1_{x \neq x'}$ in the inclusion graph $([n], X(k))$. By the distance of C we are promised that $\mu(1_{x=x'}) \leq 1 - \alpha$, so Proposition 5.10.12 gives the desired result.[24] Plugging $\alpha = \frac{1}{2}(1-\varepsilon)$ into the resulting bound gives the final line. □

5.10.5 Local Testability of ABNNR

In this section, we prove the ABNNR encoding of a locally testable code is locally testable.

[24]Proposition 5.10.12 is formally for $(X(1), X(k))$, but because X is homogenous the bound carries over by simply setting each partite component of A to be $A_i = 1_{x=x'}$.

Theorem 5.10.25 (Local Testability of ABNNR)**.** *For any uniform (q, s)-LTC of block-length n and (k_L, k_R)-regular bipartite graph $G = (L, R, E)$ on $|L| = n$ vertices, $E_G(C)$ has a uniform $(q + 2, \frac{cs}{qk_R})$-local tester.*

Our proof relies on the following elementary agreement test. Given a word f, denote by $\mathcal{T}_G(f)$ the following process:

1. Pick a uniformly random vertex $v \in L$
2. Pick two uniformly random neighbors $w, w' \in N(v)$
3. Accept if and only if $f_w(v) = f_{w'}(v)$

In the agreement notation of the previous section, we'd write the word f as an ensemble $\mathcal{F} = \{f_w : N(w) \to \mathbb{F}_2\}_{w \in R}$ and denote the acceptance probability of this test as $Agree_0(\mathcal{F})$. We will move between these two viewpoints interchangeably. $\mathcal{T}_G$ is clearly complete. Its soundness follows from a general result of [166]:

Lemma 5.10.26 ([166, Lemma 5.2])**.** *There exists a universal constant $c \geq 1$ such that for all $\varepsilon > 0$ and any ensemble $\mathcal{F}$:*

$$Agree_0(\mathcal{F}) \geq 1 - \varepsilon \quad \Longrightarrow \quad dist(F, g_{maj}) \leq ck_R\varepsilon$$

where for $v \in L$ we define $g_{maj}(v) \coloneqq plurality_{w \in N(v)}\{f_w(v)\}$.[25]

With this in mind, we prove Theorem 5.10.25 via the following natural tester: given a word $f \in (\mathbb{F}_2^{k_R})^R$, run the DP-tester on w, and the base code tester on the plurality decoding of f if it passes:

1. Run $\mathcal{T}_G$ on f
2. For every $v \in L$ that $\mathcal{T}_C$ would query:

[25] Ties can be broken arbitrarily.

– Sample a random neighbor of v: $w \sim N(v)$

– Denote the 'decoded' plurality as $g_{\text{dec}}(v) = f_w(v)$

3. Run $\mathcal{T}_C$ on g_{dec}

Proof of Theorem 5.10.25. Query complexity and completeness are immediate from construction. The tester is uniform since the base code is uniform, and picking a uniformly random vertex $v \in L$ and random $w \in N(v)$ is uniformly distributed over R.

Toward analyzing soundness, fix any $\delta \in [0,1]$ and f some word at distance $dist(f, E_X(C)) = \delta$. We need to show the test *rejects* on f with probability at least $\delta c' \frac{s}{qk_R}$ for some $c' \geq 0$. Observe that we may assume $\mathcal{T}_G(f)$ passes with probability at least $1 - \delta \frac{1}{4c} \frac{s}{qk_R}$, since otherwise Step (1) rejects with probability at least $\delta \frac{1}{4c} \frac{s}{qk_R}$ and we are done. We claim that conditioned on this event, the following two properties hold:

1. g_{dec} matches g_{maj} with high probability:

$$\mathbb{P}_{S \sim \mathcal{T}_C}[\exists v \in S : g_{\text{dec}}(v) \neq g_{\text{maj}}(v)] \leq \frac{s}{4ck_R}\delta$$

2. g_{maj} is far from the base code:

$$dist(g_{\text{maj}}, C) \geq \frac{1}{2k_R}\delta.$$

Soundness is then immediate: as long as the base tester $\mathcal{T}_C$ receives the 'correct' symbols from g_{maj}, it rejects with probability $s \cdot dist(g_{\text{maj}}) \geq \frac{s}{2k_R}\delta$ by Item 2. Since the former occurs with probability at least $\frac{s}{4k_R}\delta$ by Item 1, the total rejection probability is at least $\frac{s}{2k_R}\delta - \frac{s}{4k_R}\delta = \frac{s}{4k_R}\delta$ as desired.

It is left to prove the claimed properties.

Proof of Item 1.

Observe that since the queries are marginally uniform, by a union bound it is enough to show that the decoding of a uniformly random vertex succeeds except with probability $\frac{s}{4cqk_R}\delta$. The proof is essentially a simpler variant of Claim 5.8.2. For a fixed $v \in L$ define the set of 'bad' neighbors as those whose labeling disagrees with the plurality decoding as:

$$B_v \coloneqq \{w \in N(v) : f_w(v) \neq g_{\text{maj}}(v)\}.$$

The failure probability of our single-query decoding procedure is exactly $\mathop{\mathbb{E}}_{v \in L}[B_v]$, so it is enough to show

$$\mathbb{E}[B_v] \leq \frac{s}{4cqk_R}\delta.$$

This follows from the fact that the test $\mathcal{T}_G$ can be written as an expectation over complete local distributions. In particular, observe that we can decompose $\mathcal{T}_G$ as:

$$Agree_0(f) = \mathop{\mathbb{P}}_{v \sim L, w, w' \sim N(v)}[f_w(v) = f_{w'}(v)] \geq 1 - \frac{s}{4cqk_R}\delta.$$

This implies that the expected probability that $f_w(v) \neq f_{w'}(v)$ in the local distribution is at most:

$$\mathop{\mathbb{E}}_{v \sim L}\left[\mathop{\mathbb{P}}_{w, w' \sim N(v)}[f_w(v) \neq f_{w'}(v)]\right] \leq \frac{s}{4cqk_R}\delta.$$

Since w and w' are independent neighbors of v, the inner term is at least

$$\mathop{\mathbb{P}}_{w, w' \sim N(v)}[f_w(v) \neq f_{w'}(v)] = 2\mathbb{E}[B_x]\mathbb{E}[\overline{B_v}] \geq \mathbb{E}[B_v]$$

since $\mathbb{E}[\overline{B_v}] \geq \frac{1}{2}$ by construction, which completes the proof.

Proof of Item 2.

The proof follows from the fact that conditioned on Step (1) (the agreement test) passing, $E_X(g_{\text{maj}})$ is close to a DP-codeword. Thus if g_{maj} itself is too close to C, it's

encoding will be close to $E_G(C)$ violating our original assumption on the distance. Formally, assume for the sake of contradiction that g_{maj} is not $\frac{\delta}{2k_R}$-far from C. Then there exists $x \in C$ such that $dist(g_{\text{maj}}, x) < \frac{\delta}{2k_R}$. Since G is bi-regular, this further implies a bound on the distance of the *encodings* of g_{maj} and x:

$$\begin{aligned} dist(E_G^k(g_{\text{maj}}), E_G^k(x)) &= \mathbb{P}_{w \sim R}[E_G(g_{\text{maj}})(w) \neq E_G(x)(w)] \\ &\leq k_R \mathbb{P}_{w \sim R, v \sim N(w)}[E_G(g_{\text{maj}})(w)_v \neq E_G(x)(w)_v] \\ &= k_R dist(g_{\text{maj}}, x) \\ &< \frac{\delta}{2}. \end{aligned}$$

where the second equality follows from the fact that since G is bi-regular, sampling a uniformly random vertex $v \in R$ and a random $s \in N(v)$ is equidistributed with sampling a random $w \in L$. Finally closeness of the encoded words contradicts the fact that f is δ-far from $E_G(C)$, since $E_G(x)$ is a codeword and by the triangle inequality:

$$dist(f, E_G(x)) \leq dist(f, E_G(g_{\text{maj}})) + dist(E_G(g_{\text{maj}}), E_G(x)) < \delta,$$

where we have used the fact that since $Agree_0(f) \geq 1 - \frac{s}{4cqk_R}$ by assumption, f is $\frac{s}{4q}\delta$-close to $E_G(g_{\text{maj}})$ by Lemma 5.10.26.

□

We remark that it is likely possible to reduce the query complexity of the above argument to just $q+1$ by re-using the first query of $\mathcal{T}_C$ for the agreement test and performing a more careful conditioned analysis on the latter passing. Since q for us is some small constant, the difference between $q+1$ and $q+2$ is negligible so we only give the simpler independent version above.

To prove Theorem 5.10.21, we need a good explicit family of binary linear LTCs

with near optimal distance to amplify via Theorem 5.10.25. In fact, as discussed in the proof overview we will need base codes with the stronger property of having small *bias*:

$$bias(C) \coloneqq \max_{c \in C \setminus \{0\}} \left\{ \left| \mathbb{E}_{i \in [n]} [(-1)^{c_i}] \right| \right\}.$$

Note that an ε-biased linear code always has distance at least $\frac{1-\varepsilon}{2}$ since for every $c \in C$, $\left| \mathbb{E}_{i \in [n]} \left[(-1)^{c_i} \right] \right| = 1 - 2\operatorname{dist}(c, 0)$. Our base codes start with the breakthrough construction of c^3-LTCs by Dinur, Evra, Livne, Lubotzky, and Moses [117]:

Theorem 5.10.27 ([117])**.** *There exists an explicit infinite family of binary, linear, uniform LTCs with constant rate, distance, locality, and soundness, and a linear time unique decoder up to constant radius.*

We remark that while [117] do not argue their testers are uniform, it is easy to check from their construction that this is the case. As discussed in the overview, it is possible to use the distance amplification techniques of [268] to achieve c^3-LTCs with distance near $1/2$ (even hitting the GV bound), but these codes may not be ε-biased and incur a heavy $\operatorname{poly}(\varepsilon^{-1})$ cost in the query complexity. We will take a different amplification strategy that avoids these issues almost entirely at the cost of a small polynomial blowup in rate. The idea is to concatenate our amplification procedure in Theorem 5.10.25 with the *Hadamard code*:

$$Had : \mathbb{F}_2^k \to \mathbb{F}_2^{2^k},\ \alpha \overset{Had}{\mapsto} \langle \alpha, \cdot \rangle$$

We denote the resulting code (the image of this map), by C_{Had}.

The Hadamard code comes with three properties critical to our construction. First, it is easy to check that the code is binary, linear, and perfectly balanced (0-biased). Second, it is a classical result that the Hadamard code is an excellent LTC:

Theorem 5.10.28 ([67])**.** *The Hadamard code is a uniform* $(3, 1)$*-LTC.*

Finally, it is well-known folklore (and an easy exercise) to show the Hadamard code is *locally decodable*. We will not go into the formal details of locally decodable codes, but it suffices to say there is a randomized algorithm $Dec : \mathbb{F}_2^{2^k} \times [k] \to \mathbb{F}_2$ which given a word y such that $d(y, C_{\text{Had}}) < \frac{1}{4}$ and a coordinate i, queries only 2 bits of y and outputs the correct decoding with probability proportional to its distance from the code:

$$\mathbb{P}[Dec(y, i) \neq x_i] \leq 2d(y, C_{\text{Had}}) \tag{5.33}$$

where $x = Had^{-1}(y)$ is the pre-image of the unique closest codeword to y.

Note that this is essentially the same property we used in Theorem 5.10.25 in decoding the ABNNR-encoding. We will show that concatenating with the Hadamard code similarly allows us to amplify the bias and distance of the base code while only incurring constant blow-up in query complexity and logarithmic loss in soundness.

Proposition 5.10.29 (The Base Code)**.** *For every $\varepsilon > 0$ there is an explicit infinite family of uniform, ε-biased, linear $(\text{poly}(\varepsilon), \frac{1}{2} - \varepsilon)_2$-codes that are $(O(1), \log^{-1}(1/\varepsilon)))$-locally testable. Moreover, there exists $t < 1$ such that these codes are uniquely decodable up to radius t in time $n \exp(\text{poly}(\varepsilon^{-1}))$.*

Proof. We apply Theorem 5.10.25 to [117]'s LTCs using the complex $X = W_G$ generated by length $k = O(\log(1/\varepsilon))$ random walks on any constant degree, regular $\frac{1}{2}$-spectral expander G (such expanders are known to exist for any n, see e.g. [15]). This results in a $\text{poly}(\varepsilon)$-rate code with distance at least $1 - \frac{\varepsilon}{2}$ by the classical expander hitting set lemma [198] which states that for any set $T \subset X(1)$ of measure ρ

$$\mathop{\mathbb{P}}_{(s_1,\ldots,s_k)\in X(k)}[|\{s_1, \ldots, s_k\} \cap T| = 0] \leq \rho \left(\frac{1+\rho}{2}\right)^{k-1}. \tag{5.34}$$

The argument is then as in Corollary 5.10.24. Finally by Theorem 5.10.25 this encoding has a uniform $(c, \frac{c'}{\log(1/\varepsilon)})$-local tester for some universal constants $c, c' > 0$.

The key step is now to concatenate this resulting family with the Hadamard code:

$$\mathcal{C} \coloneqq Had(E_X(C)),$$

where the Hadamard encoding is applied pointwise to each symbol in $E_X(C)$. This encoding maintains linearity and decreases the rate by a factor of at most $2^k = \text{poly}(\varepsilon)$. To compute the bias, observe that since the Hadamard code is perfectly balanced, every non-zero symbol of a word $w \in E_X(C)$ contributes 0 to the bias so in the worst case we have

$$bias(\mathcal{C}) \leq \max_{c \in E_X(C) \setminus \{0\}} \{wt(c) \cdot 0 + (1 - wt(c)) \cdot 1\} \leq 1 - d(E_X(C)) \leq \varepsilon,$$

where $wt(c)$ denotes the normalized hamming weight (fraction of non-zero elements in c).

We now argue that this concatenation has a uniform local tester. Write $q = c$ and $s = \frac{c'}{\log(1/\varepsilon)}$ and recall $E_X(C)$ has a uniform (q, s)-local tester (which we'll refer to as the "base tester") which requests only a single bit of each queried symbol. We'll denote these requests as pairs $(x, i) \in X(k-1) \times [k]$. Our testing procedure follows a similar form as in Theorem 5.10.25. Namely given a potential word y:

1. Draw a uniformly random $x \in X$, and run the Hadamard test on $y(x)$.

2. For each (x, i) requested by the base tester, run the Hadamard decoder on $y(x)$ at coordinate i.

3. Run the base tester on the output

The test is complete by construction, and makes at most $2q + 3$ queries (two for each run of the local decoder, and three for the initial Hadamard test). Soundness follows along the same lines as Theorem 5.10.25. We first argue that the word y must be close to a cartesian product of Hadamard codewords (else Step 1 rejects). We then argue that the

decoding succeeds with high probability, and the decoded word must be far from the base code (else y itself would be close to $\mathcal{C}$), so the base tester rejects with good probability.

Formally, let $y \in (\mathbb{F}_2^{2^k})^X$ satisfy $dist(y, \mathcal{C}) = \tau$. We argue that the above procedure rejects with probability at least $\frac{s\tau}{4q}$. To see this, first observe we can assume the word y is $1 - \frac{s\tau}{4q}$ close from being a product of Hadamard codewords. Otherwise Step 1 rejects with probability at least $\frac{s\tau}{4q}$ since

$$\mathbb{P}[\text{Step 1 Rejects}] \geq \mathbb{E}_x[dist(y(x), C_{Had})] = dist\left(y, \prod_X C_{Had}\right).$$

Fix some $y_{\text{had}} \in \prod_X C_{Had}$ which minimizes the distance from y and denote the set of 'corrupted' coordinates of y to be

$$B \coloneqq \left\{x \in X : dist\left(y(x), C_{\text{Had}}\right) \geq \frac{1}{4}\right\}.$$

Note that the measure of B is at most $\frac{s\tau}{q}$, since the closeness of y to $\prod C_{\text{Had}}$ is violated otherwise.

Define set of 'valid' decodings of y to be strings in $(\mathbb{F}_2^k)^X$ whose encodings match y_{had} on all uncorrupted coordinates:

$$Y_{\text{dec}} \coloneqq \left\{y_{\text{dec}} \in (\mathbb{F}_2^k)^X : \forall x \in \bar{B}, Had(y_{\text{dec}}(x)) = y_{\text{had}}\right\}.$$

Note that when we *re-encode* any $y_{\text{dec}} \in Y_{\text{dec}}$, it will be at distance at most $\frac{s\tau}{q}$ from y since:

$$\begin{aligned} dist(Had(y_{\text{dec}}), y) &\leq dist(Had(y_{\text{dec}}), y_{\text{had}}) + dist(Had(y_{\text{had}}), y) \\ &\leq (\mathbb{P}[B] \cdot \frac{1}{2} + \mathbb{P}[\bar{B}] \cdot 0) + \frac{s\tau}{4q} \\ &\leq \frac{s\tau}{q}. \end{aligned}$$

We now argue that Step 2 correctly decodes to some $y_{\text{dec}} \in Y_{\text{dec}}$ with high probability. Since decoding always 'succeeds' on coordinates in B (since any decoded value is accepted), the worst case is when $B = \emptyset$ and y_{dec} is the pre-image of the unique closest word in every coordinate. In this case, local decodability of the Hadamard code Equation (5.33) promises the probability any fixed query (x, i) fails to decode to $y_{dec}(x)_i$ is at most

$$\mathbb{E}_x[2dist(y(x), C_{Had})] \leq 2dist\left(y, \prod_X C_{\text{Had}}\right) \leq \frac{s\tau}{2q}.$$

Union bounding over all q queries, we decode to y_{dec} except with probability $\frac{s\tau}{2}$.

Finally, we argue any $y_{\text{dec}} \in Y_{\text{dec}}$ must be $\frac{3}{4}\tau$-far from $E_X(C)$, in which case the base tester rejects with probability at least $\frac{3}{4}s\tau - \frac{s\tau}{2} = \frac{s\tau}{4}$ by a union bound. This simply follows from the fact that if y_{dec} were $\frac{3}{4}\tau$-close to $E_X(C)$, then there exists $w \in E_X(C)$ such that

$$\begin{aligned} dist(Had(w), y) &\leq dist(Had(w), Had(y_{\text{dec}})) + dist(Had(y_{\text{dec}}), y) \\ &\leq dist(w, y) + \frac{s\tau}{q} \\ &\leq \tau \end{aligned}$$

where the second inequality comes from observing that the encodings of w and y_{dec} are the same on any symbol on which they agree, and can only improve in distance when they differ.[26]

It is left to show our concatenated family has an efficient unique decoder. It is a classical fact that concatenating codes with efficient unique decoding at radii t_{inner} and t_{outer} gives a code with an efficient unique decoder at radius $t_{\text{inner}}t_{\text{outer}}$ simply by running the inner decoder on every inner symbol, then running the outer decoder on the result [152]. In particular decoding at this distance is successful since if a true word m is corrupted

[26]Technically we've also used that $q \geq 4$ and $s \leq 1$ here, which follow from the parameters of [117].

in at most a $t_{\text{inner}}t_{\text{outer}}$ fraction of its coordinates, no more than a t_{outer} fraction of inner codewords can have a t_{inner} fraction of corruptions. Thus running the inner decoder results in the correct word except in at most a t_{outer} fraction of outer coordinates, which is then successfully decoded by the outer decoder.

We can decode over each Hadamard symbol up to radius $t' < \frac{1}{4}$ by brute force in $\exp(\text{poly}(\varepsilon^{-1}))$ time (and otherwise output some fixed word, say 0, beyond this radius). $E_X(C)$ can be efficiently decoded up to some constant radius t by taking the majority vote over each coordinate, then running the base decoder on C (see e.g. [121]). This process takes time $n\exp(\text{poly}(\varepsilon^{-1})) + O(n)$, for a total time of $n\exp(\text{poly}(\varepsilon^{-1}))$ as desired. □

Finally, we argue these base codes can be amplified to large alphabet LTCs with near optimal distance.

Corollary 5.10.30 (Large Distance c3-LTCs)**.** *For all large enough $k \in \mathbb{N}$ and all $\varepsilon > 0$, there exists an infinite family of $\mathbb{F}_2$-linear $(r, d)_{|\Sigma|}$-codes that are (q, s)-locally testable for:*

$$r = \varepsilon^{O(k)}, \quad d = 1 - 2^{-k} - \varepsilon, \quad |\Sigma| = 2^k, \quad q = O(1), \quad s = O\left(\frac{1}{k\log(\frac{1}{\varepsilon})}\right)$$

Proof. We start with the explicit family of base codes given in Proposition 5.10.29 with $q = O(1)$ queries, distance $\frac{1}{2} - \frac{\varepsilon}{4}$, and rate $\text{poly}(\varepsilon)$. For a given code C in the family of blocklength n, let X be the collection of length-k walks on any explicit degree $\frac{\varepsilon}{2}$-expander of degree $O(\varepsilon^{-2})$ on n vertices (such graphs exist for any large enough n [15]). The size of $X(k-1)$ is $O(\frac{n}{\varepsilon^{2k}})$, and the expander hitting-set lemma (Equation Equation (5.34)) ensures the distance of $E_X(C)$ is at least

$$1 - \left(\frac{1}{2} + \frac{\varepsilon}{2}\right)^k \geq 1 - 2^{-k} - \varepsilon.$$

Finally, since the base code's tester is uniform, local testability follows from Theorem 5.10.25.

□

It is worth comparing Theorem 5.10.25 to the distance amplification for LTCs of Kopparty, Meir, Ron-Zewi, and Saraf [268], who also analyzes local testability of distance-amplified codes via a related construction of Alon and Luby [19]. Their work largely focuses on achieving near-optimal *rate-distance* trade-off: they achieve the singleton bound over large alphabet codes and the GV-bound over $\mathbb{F}_2$. In comparison Theorem 5.10.25 has worse rate than their construction, but makes up for this with exponentially better alphabet size and local testability. Quantitatively, to achieve distance $1-\beta$, [268]'s codes require alphabet size $2^{\mathrm{poly}(1/\beta)}$ and lose $\mathrm{poly}(1/\beta)$ factors in the testing parameters, while ours have the optimal β^{-1}-dependence in the former and lose only *logarithmically* in the latter.

5.10.6 List-Decoding ABNNR

In this section, we show that the codes in Corollary 5.10.30 are efficiently list-decodable, completing the proof of Theorem 5.10.21. Efficient list-decoding of the ABNNR construction was first considered by Dinur, Harsha, Kaufman, Livni-Navon, and TaShma [121], who gave an algorithm for the construction instantiated on HDX. After [121]'s work, Alev, Jeronimo, Quintana, Srivastava, and Tulsiani [10] showed it possible to efficiently list decode a closely related construction called a *direct sum code* over some relaxed families of splitting trees (including HDX and expander walks). Using a reduction to direct sum, they also recover the parameters of [121]'s ABBNR-codes on HDX as well.

In this section, we give an efficient list-decoding algorithm for the ABNNR-construction on complexes satisfying a slightly strengthened variant of splittability that ensures the complex can be "split" at any coordinate. We refer to this notion as complete splittability.

Definition 5.10.31 (Complete Splittability)**.** A d-uniform partite complex X is called

completely λ-splittable if for all $i \in [d]$:

$$\lambda_2(S_{[1,i],[i+1,d]}) \leq \lambda$$

where we recall $S_{[1,i],[i+1,d]}$ is the 'swap walk' adjacency operator between $X[[1,i]]$ and $X[[i+1,d]]$.

Complete splittability sits in-between tuple-splitting trees and partite HDX, and is closely related to notions of splittability defined in more recent works on direct-sum decoding [221, 65] (in particular it is equivalent to the definition introduced in the former, see Section 5.17 for details). We can now state the main result of the subsection:

Theorem 5.10.32 (List Decoding on Completely Splittable Complexes)**.** *For any $t < 1/2$, let C_0 be a binary code with blocklength n, bias$(C_0) \leq 1-2t$, and which is unique-decodable to radius t in time $\mathcal{T}_0$. Let $X \subset [n]^k$ be a homogenous, completely λ-splittable complex, and $C = E_X(C_0)$ its ABNNR encoding. Let $d(C) = 1-\varepsilon$ and $\tau > 0$ be any value satisfying*

$$\tau \geq \max\left\{\sqrt{\varepsilon}, 4\sqrt{(1-2t)^i + 2\lambda}, \sqrt{2^{21}\lambda k^3}, 8\left(1-t\right)^{i/2}\right\}$$

for $i = \frac{k}{2} - \sqrt{k\log(\frac{2}{\tau})}$. There exists a randomized algorithm, which given $\tilde{y} \in (\mathbb{F}_2^k)^X$ recovers the list

$$\mathcal{L}_\tau(\tilde{y}) = \{y \in C : d(\tilde{y}, y) \leq 1-\tau\}$$

with probability $1 - \frac{\tau-\varepsilon}{\varepsilon(\tau^2-\varepsilon^2)}2^{-\Omega(n/t^2)}$ in time $\tilde{O}(\frac{\tau-\varepsilon}{\tau^2-\varepsilon^2}\exp(\log(1/t)\exp(k^3/\tau^2))(|X| + \mathcal{T}_0))$. Moreover, the number of such codewords is at most:

$$|\mathcal{L}_\tau(\tilde{y})| \leq \frac{\tau-\varepsilon}{\tau^2-\varepsilon^2}.$$

The proof of Theorem 5.10.32 goes via reduction to direct sum decoding. The *direct sum* encoding of a code C on a complex X simply concatenates the ABNNR-encoding

with $XOR : \mathbb{F}_2^k \to \mathbb{F}_2$, that is $dsum_X(C) \coloneqq XOR(E_X(C))$ (where XOR applied pointwise, similar to in the Hadamard concatenation in the prior section). Most of the work in proving Theorem 5.10.32 is already done via [221, 65]'s direct sum decoder for completely splittable complexes:

Theorem 5.10.33 (List Decoding Direct Sum [221, 65])**.** *For any $t < 1/2$, let C_0 be a binary code with blocklength n, $bias(C_0) \leq 1 - 2t$, and which is unique-decodable to radius t in time $\mathcal{T}_0$. Let $X \subset [n]^k$ be a homogenous, completely λ-splittable complex, and $C = dsum_X(C_0)$ be its direct sum encoding. Let $\varepsilon = bias(C)$ and $\beta > 0$ be such that*

$$\beta \geq \max\left\{\sqrt{\varepsilon}, \sqrt{2^{20}\lambda k^3}, 2\left(1-t\right)^{k/2}\right\}.$$

There exists a randomized algorithm which given $\tilde{y} \in \mathbb{F}_2^X$ recovers the list

$$\mathcal{L}_\beta^{X,(sum)}(\tilde{y}) = \left\{x \in C_0 : d(\tilde{y}, dsum_X(x)) \leq \frac{1}{2} - \beta\right\}$$

with probability $1 - \frac{1}{\varepsilon}2^{-\Omega(n/t^2)}$ in time $\tilde{O}(\exp(\log(t)\exp(\frac{k^3}{\beta^2}))|X| + \mathcal{T}_0)$. Moreover, the list has at most

$$|\mathcal{L}_\beta^{X,(sum)}(\tilde{y})| \leq \frac{1}{\varepsilon}$$

codewords.

We remark that the requirement on the unique decoder radius in [221, 65] is slightly more strict as stated, but one can easily modify parameters in their proof to achieve the above. We now prove Theorem 5.10.32 by reduction to Theorem 5.10.33.

Proof of Theorem 5.10.32. We appeal to a simple variant of the reduction of [10]. Given a potential word $\tilde{y} \in (\mathbb{F}_2^k)^X$ at distance $d(\tilde{y}, C) = 1 - \tau$, we claim the following process is an efficient list-decoder.

1. For every $F \subseteq [d]$, apply Theorem 5.10.33 to $XOR(\tilde{y}|_F)$ on $dsum_{X^F}(C)$ with $\beta = \tau/4$ to get:

$$\mathcal{L}^{pot}_{\beta}(\tilde{y}) \coloneqq \bigcup_{F \subseteq [d+1]} \mathcal{L}^{X^F,(sum)}_{\beta}(\tilde{y})$$

2. Prune any decodings that are far from $\tilde{y}$:

$$\mathcal{L}_{\tau}(\tilde{y}) \coloneqq \left\{ E_X(x) : x \in \mathcal{L}^{pot}_{\beta}(\tilde{y}),\ d(E_X(x), \tilde{y}) \leq 1 - \tau \right\}.$$

To prove this procedure indeed gives a valid list-decoding we will use the fact that any projection of a splittable complex, in particular each X^F above, is also splittable.

Claim 5.10.34. Let $X \subset [n]^d$ be a homogeneous, completely λ-splittable complex. Then for any $F \subseteq [d]$, X^F is homogeneous and completely λ-splittable.

The proof follows from the inheritance of expansion under projection, and is defered to Section 5.17.

We now prove the output list $\mathcal{L}_{\tau}$ is exactly the set of codewords at distance at most $1 - \tau$ from C. Recall that our base code is assumed to be at most $(1 - 2t)$-biased. By bias amplification lemma for splitting trees (Proposition 5.10.13) and Claim 5.10.34, we have that for any X^F, the bias of $dsum_{X^F}(C)$ is at most $(1 - 2t)^{|F|} + 2\lambda$. As such, our choice of $\beta = \tau/4$ satisfies the conditions of Theorem 5.10.33 whenever $i \leq |F| \leq k$ (for i as in the Theorem statement). Let $y \in E_X(C)$ be a codeword at distance at most $1 - \tau$ from $\tilde{y}$, then we have

$$\mathbb{P}[\tilde{y}_s = y_s] = \mathbb{E}_s \left[\mathbb{E}_{F \in \mathbb{F}_2^k} [\chi_F(\tilde{y}_s + y_s)] \right] \geq \tau,$$

where $\chi_F(y) = (-1)^{\langle F, y \rangle}$ are the standard characters. By averaging and Chernoff, there must exist $|F| \geq \frac{k}{2} - \sqrt{k \log(2/\tau))}$ such that

$$\mathbb{E}_s \left[\chi_F(\tilde{y}_s) \chi_F(y_s) \right] \geq \tau/2,$$

which is equivalent to the statement that the distance of $\tilde{y}|_F$ from $y|_F$ is at most $\frac{1}{2}-\frac{\tau}{4} = \frac{1}{2}-\beta$ in the direct sum encoding on X^F. Since X^F is homogeneous and completely λ-splittable, by Theorem 5.10.33 we have that the decoding of y appears in $\mathcal{L}_{pot}(\tilde{y})$ except with probability $\frac{1}{\varepsilon}\exp(n/t^2)$. By the Johnson bound (see e.g. [121][Theorem 2.15]), there can be at most $\frac{\tau-\varepsilon}{\tau^2-\varepsilon^2}$ such codewords y close to $\tilde{y}$, so union bounding over these gives that all valid codewords are in the list with the desired probability. Finally, all extraneous decodings are removed in the second step, which completes correctness of the procedure.

Running time follows immediately from the fact that we have run the direct sum decoder at most 2^k times, and the resulting potential list has at most $\exp(k)$ potential codewords by the Johnson bound and our choice of $\beta = \tau/4$, each of which can be pruned in $O(k|X|)$ time. □

Finally, we instantiate this result on expander walks to complete the proof of Theorem 5.10.21.

Corollary 5.10.35. *The family of codes in Corollary 5.10.30 is* $(1-2^{-\Omega(k)}, \frac{1}{2^{O(k)}})$ *efficiently list-decodable with confidence* $2^{-\Omega(n-k)}$.

Proof. Since [221] prove that walks on an λ-expander are homogeneous and completely λ-splittable, it is enough to ensure the parameter τ in Theorem 5.10.32 can be set to $2^{-\Omega(k)}$. Recalling that the base code of this construction is ε-biased and uniquely decodable in linear time up to constant radius, and that the encoding has distance at least $1-2^{-k}-\varepsilon$, as long as $\varepsilon \leq \exp(-\Omega(k))$ one can easily check it is possible to take $\tau = \exp(-\Omega(k))$ and $\frac{\tau-\varepsilon}{\tau^2-\varepsilon^2} = 2^{-O(k)}$ satisfies all conditions of Theorem 5.10.32. □

5.10.7 Toward Lossless Amplification from HDX?

Our expander-based construction experiences polylogarithmic decay in soundness with the alphabet size. This is exponentially better than prior methods, but it is plausible

the decay could be avoided entirely. In this section, we propose an approach toward such 'lossless' amplification via the ABNNR-Encoding on an HDX.

Instantiating Theorem 5.10.25 directly on an HDX does actually slightly improve soundness over Theorem 5.10.21, but the loss of $\frac{1}{k}$ in the dimension of the complex is inherent to our approach, no matter the instantiation. The reason is that our method only simulates the base tester a single time. As a result, the rejection probability will only scale with the distance of the decoded word from the *base* code. On the other hand, because X amplifies this distance by a factor of k, this results in an unavoidable loss in soundness.

There is a natural approach to fixing this issue: simulate the base tester "*k times in parallel*," analogous to Raz's parallel repetition theorem [327] and its de-randomized variants in agreement testing [215, 129, 110, 41]. Indeed on the complete complex, this approach can actually be carried out successfully:

Theorem 5.10.36 (Parallel LTC Amplification)**.** *Let C be a uniformly (q, s)-locally testable $(r, \frac{1-\varepsilon}{2})_2$-code of blocklength n and $X = \binom{[n]}{k}$. Then $E_X(C)$ is a uniformly (q', s')-locally testable $(r', d')_{|\Sigma|}$-code for:*

$$r' = \frac{r}{n^{k-1}}, \quad d' = 1 - \frac{1}{2^k}(1+\varepsilon)^k - o_n(1), \quad |\Sigma| = 2^k, \quad q' = q + 2, \quad s' = cs$$

where $c > 1$ is some universal constant (independent of k).

Proof. Rate and alphabet are immediate from construction, and distance follows from (near)-independence of $\binom{[n]}{k}$. For local testability, we rely on the following agreement theorem of Dinur and Steurer:

Claim 5.10.37 ([115]). There exists a two query agreement test on $\binom{[n]}{k}$ and a constant

$c \geq 1$ such that[27]

$$Agree_0(\mathcal{F}) \geq 1 - \varepsilon \quad \implies \quad dist(\mathcal{F}, g_{\text{maj}}) \leq c\varepsilon$$

With this in hand the proof is similar to Theorem 5.10.25. For k independent runs of the base tester $\mathcal{T}_C$, denote the requested symbols by $\{x_j^{(i)}\}_{j\in[q],i\in[k]}$ and write $s_j \coloneqq \{x_j^{(1)}, \ldots, x_j^{(k)}\} \in \binom{[n]}{k}$. Given a word f, our tester performs the following two-step procedure:

1. Run the Dinur-Steurer test on f
2. For every $i \in [q]$:
   - For each $j \in [k]$, denote the 'decoded' plurality function at $x_j^{(i)}$ as $g_{\text{dec}}(x_j^{(i)}) = f_{s_j}(x_j^{(i)})$
   - Run $\mathcal{T}_C$ on $g_{\text{dec}}(x_1^{(i)}), \ldots, g_{\text{dec}}(x_q^{(i)})$

The tester is complete and makes $q + 2$ queries by construction. The proof of soundness is very similar to Theorem 5.10.25, so we give a shortened explanation here. Assuming $dist(f, E_X(C)) = \delta$, we need to show our test rejects with probability $\Omega(\delta)$. Assume Step (1) passes with probability at least $1 - \frac{s\delta}{8cq}$ (else we are done). Since the base tester is marginally uniform and the k runs are independent, each face $s_j \in X$ is marginally uniform, and therefore by Claim 5.10.37 exactly matches plurality except with probability $\frac{s\delta}{8q}$. Union bounding over the q faces gives that g_{dec} matches g_{maj} on all queried points except with probability $\frac{s}{8}\delta$. By the same argument as in Theorem 5.10.25, g_{maj} must be at least $\frac{\delta}{2k}$-far from C, so conditioned on receiving the correct values, at least one of the base testers rejects with probability $1 - (1 - \frac{s\delta}{2k})^k \geq \frac{s\delta}{4}$ for a total rejection probability of $\frac{s\delta}{4} - \frac{s\delta}{8} = \frac{s\delta}{8}$ as desired. □

[27] We remark that formally Dinur and Steurer consider $[n]^k$, but their techniques transfer to this setting (formally, see e.g. [109] for the quoted result).

The issue with Theorem 5.10.36 is that the resulting code has very poor rate. It is natural to conjecture that, as in the single query variant, this can be resolved by replacing $\binom{[n]}{k}$ with a high dimensional expander. Indeed it is known that the agreement testing step (the analog of Claim 5.10.37) works in this setting [124, 109, 236]. Note that this is *not* the case for other splitting trees such as expander walks, which would lose a factor of k. The challenge then mainly lies in extending the parallel simulation. Below, we discuss the current challenges toward porting this second half of the proof to HDX.

The first difficulty in parallel simulation is that the HDX and base code must 'match' in the sense that there is a natural distribution over q-tuples of faces that corresponds to k valid q-query tests for the base code. Recently, Golowich [171] showed how to construct HDX from certain (modified) locally testable codes, and Dinur, Liu, and Zhang [127] built LTCs (of sub-constant rate) that sit on the triangles of an HDX. It is not clear a priori how to fit either of these constructions into our framework, but it is plausible they or modifications thereof do.

The second difficulty in simulating the parallel tester is that the k runs of the base test will be inherently correlated. There is a long history in the HDX literature of handling this type of behavior (e.g. [124, 109, 236, 177, 110, 41]), including the agreement tests we discuss in Section 5.8. Very recently, building upon tools in this work, [110, 41] prove that certain strong families of high dimensional expanders admit 1%-regime agreement tests, which are closely related to parallel repetition [215, 129]. As a start, it would be interesting to extend Theorem 5.10.36 even to dense complexes known to admit such tests, such as the Grassmann [215] or spherical buildings [110]. To summarize, we give the following informal conjecture:

Conjecture 5.10.38 (Parallel LTC Amplification)**.** *There exists a family of high dimensional expanders and locally testable codes which admit a "parallel" variant of Theorem 5.10.25, that is:*

1. *There is a q-step random walk or distribution on X generating k valid 'parallel' tests*

2. *The resulting parallel tester has no (asymptotic) decay in soundness.*

5.11 Concentration Lower Bounds

In this section we prove our results are essentially optimal with respect to their dependence on the sampling parameters and quantitative expansion of the complex. We further show one cannot hope to achieve our more involved applications such as reverse hypercontractivity using classical samplers such as expander walks.

5.11.1 Optimality for Inclusion Samplers

In Section 5.5.2, we proved that the if X is a high dimensional expander, then the inclusion graphs $(X(k), X(i))$ are roughly $(\varepsilon, \exp(-\varepsilon^2 \frac{k}{i}))$-additive samplers. Inspecting the proof of Theorem 5.5.1, it is easy to show that for any set A of density μ and strong enough local-spectral expansion $\lambda \ll \mu$, one can give the following improved bound:[28]

$$\mathop{\mathbb{P}}_{s \in X(k)} \left[\mathop{\mathbb{P}}_{r \in X(i)} [A \mid \ r \subseteq s] \geq \mathbb{P}[A] + \varepsilon \right] \leq \exp\left(-\Omega\left(\frac{\varepsilon^2}{\mathbb{P}[A]} \frac{k}{i} \right) \right).$$

In this section we prove that for most $i \leq k$, this is essentially optimal. Recall that a complex X is called (γ, i)-hitting if for any $A \subseteq X(1)$, $\mathbb{P}_{\sigma \in X(i)}[\sigma \subseteq A] \leq \mathbb{P}[A]^i + \gamma$. We say that X is (γ, i)-hitting for sets of size μ if $\mathbb{P}_{\sigma \in X(i)}[\sigma \subseteq A] \leq \mathbb{P}[A]^i + \gamma$ holds for any $A \subseteq X(1)$ of size $\mathbb{P}[A] = \mu$.

Theorem 5.11.1. *Let $\varepsilon \in (0, 0.1)$. Let $i < k$ be two integers satisfying $\frac{k}{i} \geq 3\varepsilon^{-2}$ and let $n \geq ke^k$. Let X be a k-uniform simplicial complex such that every vertex in X has the same probability $\frac{1}{n}$. If X is $(i, \frac{\varepsilon}{3})$-hitting, then $(X(k), X(i))$ is not an (ε, β)-additive sampler for $\beta < \exp(-O(\varepsilon^2 \frac{k}{i}))$.*

[28] Though formally when $i = \Omega(\frac{k}{\log k})$ one needs to assume ε is a large enough to subsume the $\frac{1}{\varepsilon}$ factor in front of the exponent.

In particular, for any $\mu \in (0, 0.2)$, if X is $(i, \frac{\min(\varepsilon,\mu)}{3})$-hitting for sets of size $\frac{\mu}{i}$ there exists $E \subseteq X(i)$ with $\frac{1}{2}\mu \leq \mathbb{P}[E] \leq \mu$ such that $\mathbb{P}_{s\in X(k)}\left[\mathbb{P}_{r\in X(i)}[E \mid r \subseteq s] > \mathbb{P}[E] + \varepsilon\right] \geq \exp(-O(\frac{\varepsilon^2}{\mu}\frac{k}{i}))$.

While the assumption that X is $(i, \frac{\varepsilon}{3})$-hitting is not completely general, it follows from $(X(k), X(i))$ being a modest sampler. We prove this at the end of the subsection.

Claim 5.11.2. Let $(X(k), X(i))$ be an $(\frac{1}{10i}, \frac{\varepsilon}{6i})$-additive sampler. Then X is a $(i, \frac{\varepsilon}{3})$-hitting for all sets of size at least $\frac{0.2}{i}$.

This results in the variant of the bound stated in the introduction.

Corollary 5.11.3. *Let $\varepsilon \in (0, 0.1)$. Let k and n be large enough (as a function of ε). Let X be a k-uniform simplicial complex such that every vertex in X has the same probability $\frac{1}{n}$. If $(X(k), X(i))$ is a $(\frac{1}{10i}, \frac{\varepsilon}{6i})$-additive sampler, then $(X(k), X(i))$ is not a (ε, β)-additive sampler for any $\beta < \exp(-O(\varepsilon^2\frac{k}{i}))$.*

The assumption that the vertices all have the same probability is also not general, see the discussion in the end of this subsection.

Proof of Theorem 5.11.1

For the rest of the section we restrict our attention to "one-sided" additive sampling, that is:

$$\mathop{\mathbb{P}}_{s}\left[\mathop{\mathbb{P}}_{v\sim s}[A \mid s] > \mathbb{P}[A] + \varepsilon\right].$$

This is no larger than the probability that $|\mathbb{P}_{v\sim s}[A \mid s] - \mathbb{P}[A]| < \varepsilon$ and is more convenient to handle.

Let us introduce some notation, similar to the notation in Section 5.7.1 (but for one-sided sampling in different levels of the complex). For a k-uniform simplicial complex, $i \leq k$ and constants $\mu, \varepsilon > 0$ let

$$\pi(X, k, i, \mu, \varepsilon) = \max\left\{\mathop{\mathbb{P}}_{s\in X(k)}\left[\mathop{\mathbb{P}}_{r\in X(i)}[A \mid r \subseteq s] > \mathbb{P}[A] + \varepsilon\right] \;\middle|\; A \subseteq X(i), \mathbb{P}[A] = \mu\right\}.$$

That is the best one-sided sampling bound for when restricting to sets of relative size μ.[29]

For general sampler graphs $G = (L, R, E)$ we use similar notation:

$$\pi(G, \varepsilon, \mu) = \max \left\{ \mathop{\mathbb{P}}_{v \in R} \left[\mathop{\mathbb{P}}_{u \in L} [A \mid \ v \sim u] > \mathbb{P}[A] + \varepsilon \right] \;\middle|\; A \subseteq L, \mathbb{P}[A] = \mu \right\}.$$

We also denote by Δ_n the complete complex on n vertices.

Proof of Theorem 5.11.1. We start by stating a general bound on *all* samplers (not just inclusion samplers), via reduction to the sampling properties of the complete complex.

Claim 5.11.4. Let $k, n > 0$. Let G be a bipartite graph such that $|L| = n$ and such that it is k-right regular. Then

$$\pi(G, \varepsilon, \mu) \geq \pi(\Delta_n, k, 1, \varepsilon, \mu).$$

We defer the proof, and first argue this implies Theorem 5.11.1.

Fix μ as in the theorem statement. By Claim 5.11.4 there exists $A \subseteq X(1)$ of measure $\mathbb{P}[A] = \frac{\mu}{i}$ such that

$$\mathop{\mathbb{P}}_{s \in X(k)} \left[\mathbb{P}[A \mid \ s] > \mathbb{P}[A] + \frac{10\varepsilon}{i} \right] \geq \pi(\Delta_n, k, 1, \frac{\mu}{i}, \frac{10\varepsilon}{i}).$$

Let $E = N_i(A)$ be the set of all i-faces that hit A and let $\mu' = \mathbb{P}[E]$. We claim that

$$\mu = i\,\mathbb{P}[A] \geq \mathbb{P}[E] \geq \frac{i}{2}\,\mathbb{P}[A] = \frac{\mu}{2}$$

The lower bound follows from the hitting property, namely:

$$\mathbb{P}[E] \geq 1 - (1 - \frac{\mu}{i})^i - \frac{\mu}{3} \geq \frac{\mu}{2} = \frac{i}{2}\,\mathbb{P}[A]$$

[29]Here let us assume for simplicity that there exists sets of this probability. We note that this assumption is justified because as n grows larger the possible sizes of sets, $\{\frac{k}{n}\}_{k=0}^{n}$, becomes dense in $[0, 1]$. The use of this claim later on is to find a set of size $\approx \mu$ anyway, so we neglect dealing with this minor point.

since $\mu < 0.2$.

The upper bound follows because the only way to sample a vertex in A is to sample an i-face in E, then subsample a vertex in A in that i-face. The probability of doing so (conditioned on landing in E) is at least $\frac{1}{i}$.

The proof will follow if we can show that

$$\mathop{\mathbb{P}}_{s\in X(k)}\left[\mathop{\mathbb{P}}_{r\in X(i)}\left[E \mid \ r \subseteq s\right] > \mathbb{P}\left[E\right] + \varepsilon\right] \geq \pi(\Delta_n, k, 1, \frac{\mu}{i}, \frac{10\varepsilon}{i}) \tag{5.35}$$

since Chernoff's bound is tight on the complete complex.

Lemma 5.11.5 (Reverse Chernoff, e.g. [261, Lemma 5.2])**.** *Let ε, μ and k be such that $\frac{\varepsilon'^2}{\mu'}k \geq 3$, and let $n \geq ke^k$, then $\pi(\Delta_n, k, 1, \mu', \varepsilon') \geq \exp(-10\frac{\varepsilon'^2}{\mu'}k)$.*[30]

Applying this lemma with $\mu' \coloneqq \frac{\mu}{i}$ and $\varepsilon' \coloneqq \frac{10\varepsilon}{i}$ and using the fact that $\frac{\varepsilon'^2}{\mu'}k \geq 3$ by assumption, implies

$$\mathop{\mathbb{P}}_{s\in X(k)}\left[\mathop{\mathbb{P}}_{r\in X(i)}\left[E \mid \ r \subseteq s\right] > \mathbb{P}\left[E\right] + \varepsilon\right] \geq \pi(\Delta_n, k, 1, \frac{\mu}{i}, \frac{10\varepsilon}{i}) \geq \exp(-O(\frac{\varepsilon^2}{\mu}\frac{k}{i})).$$

Toward proving (5.35) observe that for a fixed $s \in X(k)$, the probability of E inside s depends only on the probability of A inside s.

Assume that $\mathbb{P}[A|s] - \mathbb{P}[A] > \frac{10\varepsilon}{i}$. Then

$$\begin{aligned}
\mathop{\mathbb{P}}_{r\in X(i)}\left[E \mid \ r \subseteq s\right] &> 1 - (1 - \mathbb{P}\left[A \mid \ s\right])^i \\
&\geq 1 - (1 - \mathbb{P}[A] - \frac{10\varepsilon}{i})^i \\
&\geq 1 - (1 - \mathbb{P}[A])^i + 2\varepsilon \\
&\geq \mathbb{P}[E] + \varepsilon
\end{aligned}$$

[30]Technically the lemma in [261] is for sampling with replacement, instead of sampling without replacement (which is the complete complex case), but for $n \geq ke^k$ the two are close enough in TV-distance so we get essentially the same theorem.

where the first inequality is by observing the choice of $r \subseteq s$ subsamples i random vertices (without replacement). The probability of hitting A in the latter is then at least $1-(1-\mathbb{P}[A \mid\ s])^i$, the hitting probability with replacement. The third inequality is true when $i \geq 1$ and $\varepsilon < 0.1$.

Combining this with our original guarantee on sampling A itself we have

$$\begin{aligned}\pi(X,k,i,\mu',\varepsilon) &\geq \mathop{\mathbb{P}}_{s\in X(k)}\left[\mathop{\mathbb{P}}_{r\in X(i)}[E \mid\ r \subseteq s] > \mathbb{P}[E]+\varepsilon\right] \\ &\geq \mathop{\mathbb{P}}_{s\in X(k)}\left[\mathbb{P}[A \mid\ v\in S] \geq \mathbb{P}[A]+\frac{10\varepsilon}{i}\right] \\ &\geq \pi(\Delta_n,k,1,\frac{\mu}{i},\frac{10\varepsilon}{i})\end{aligned}$$

as desired. □

It remains to prove Claim 5.11.4. The proof is basically repeating the lower bound in [87]. The latter show any degree-k (ε,δ)-additive sampler satisfies $\delta \geq \exp(-\Omega(\varepsilon^2 k))$. Similar to our strategy above, their proof can be interpreted as showing $(\Delta_n(k),\Delta_n(1))$ containment graph is the optimal sampler in terms of right-degree, then reducing the general case to Δ_n. We give a refinement of their argument for sets of specific size, though our definition of samplers is less general than theirs.

Proof of Claim 5.11.4. Let D_μ be the uniform distribution over all sets $A \subseteq L$ of probability μ. Then

$$\pi(G,\varepsilon,\mu) \geq \mathop{\mathbb{P}}_{r\in R, A\sim D_\mu}\left[\mathop{\mathbb{P}}_{v\in L}[A \mid\ v\sim r] > \mu+\varepsilon\right].$$

Observe that since we randomize over A and G is right regular, the choice of $r \in R$ doesn't affect the inner probability. As such, we may fix any $r_0 \in R$ and instead consider

$$\mathop{\mathbb{P}}_{A\sim D_\mu}\left[\mathop{\mathbb{P}}_{v\in L}[A \mid\ v\sim r_0] > \mu+\varepsilon\right],$$

that is the probability (over A) that $A \cap r_0$ contains $\geq (\mu + \varepsilon)k$ elements.

We claim this is equal to $\pi(\Delta_n, k, 1, \varepsilon, \mu)$. To see this, assume without loss of generality that $L = \{1, 2, \ldots, n\}$ and $r_0 = \{1, 2, \ldots, k\}$. Since we have assumed the vertex weights are uniform, A is a uniform sample of μn points out of L. It is convenient to think of A as being sampled via the following procedure:

1. Draw a random permutation $\sigma \in S_n$
2. Set $A = \sigma(A_0)$, where $A_0 = \{1, 2, \ldots, \mu n\}$

Observe $|r_0 \cap A| \geq (\mu + \varepsilon)k$ if and only if $|\sigma^{-1}(r_0) \cap A_0| \geq (\mu + \varepsilon)k$. But note that the random variable $\sigma^{-1}(r_0)$ is exactly a random choice of a set $\Delta_n(k)$. Therefore, this is equal to

$$\mathop{\mathbb{P}}_{s \in \Delta_n(k)} \left[\mathop{\mathbb{P}}_{v \in [n]} [A_0 \mid \ v \in s] > \mu + \varepsilon \right].$$

Since every set of vertices of equal measure has the same sampling in Δ_n by symmetry, this is exactly $\pi(\Delta_n, k, 1, \varepsilon, \mu)$ as desired. □

Sampling implies hitting

We briefly prove Claim 5.11.2.

Proof of Claim 5.11.2. Let $A \subset X(1)$ be a set of size $\frac{\mu}{i}$ and $E \subseteq X(i)$ its set of neighbors. We need to show

$$\mathbb{P}[E] \geq 1 - (1 - \mathbb{P}[A])^i - \frac{\varepsilon}{3}.$$

Let $f : X(i) \to [0, 1]$ be $f(s) = \frac{|s \cap A|}{i}$ and observe that $\mathbb{E}_{s \in X(i)}[f(s)] = \frac{\mu}{i}$. Let B be the set of bad k-faces that over-samples f:

$$B = \left\{ t \in X(k) \;\middle|\; \mathop{\mathbb{E}}_{s \subseteq t} [f(s)] \geq \frac{\mu}{i} + \frac{1}{10i} \right\},$$

then by our sampling guarantee $\mathbb{P}[B] < \frac{\varepsilon}{6i}$.

Fix $t \notin B$. Then it holds that $\mathbb{P}_{v\in X(1)}[A \mid v \in t] \leq \frac{\mu+1}{i} \leq \frac{1.1}{i}$ because $\mathbb{P}_{v\in X(1)}[A \mid v \in t] = \mathbb{E}_{s\subseteq t}[f(s)]$. Moreover, by the hitting properties of the complete complex, we have that $\mathbb{P}[E \mid t] \geq 1 - (1 - \mathbb{P}[A \mid t])^i$ (using the fact that inside the complete complex, the sampling is negatively correlated). Thus

$$\begin{aligned}
\mathbb{P}[E] &= \mathop{\mathbb{E}}_{t}\left[\mathbb{P}[E \mid t]\right] \\
&\geq \mathbb{P}[\neg B] \mathop{\mathbb{E}}_{t\notin B}\left[\mathbb{P}[E \mid t]\right] \\
&\geq (1-\frac{\varepsilon}{6})(1 - \mathop{\mathbb{E}}_{t\notin B}\left[(1-\mathbb{P}[A \mid t])^i\right]) \\
&\geq (1-\frac{\varepsilon}{6})(1 - \mathop{\mathbb{E}}_{t\notin B}\left[(1-\mathbb{P}[A \mid t] - \frac{\varepsilon}{6i})^i\right]) \\
&\geq (1-\frac{\varepsilon}{6})(1 - (1-\mathbb{P}[A])^i - \frac{\varepsilon}{6}) \\
&= 1 - (1-\mathbb{P}[A])^i - \frac{\varepsilon}{3}.
\end{aligned}$$

□

Discussion

We conclude the section with a few remarks.

Remark 5.11.6.

1. Theorem 5.11.1 doesn't follow from the tight bounds depending on degree or on $\frac{|X(k)|}{|X(i)|}$ [87]. The degree of a k-face in $(X(k), X(i))$ is $\binom{k}{i}$, and the degree of a set in $X(i)$ can be arbitrarily large, as in the complete complex case. The lower bounds in [87], imply a lower of $\exp(-O(\binom{k}{i}))$ instead of $\exp(-O(\frac{k}{i}))$. In addition, the ratio $\frac{|X(k)|}{|X(i)|}$, which is another source for lower bounds in [87], can be arbitrarily large. That being said, we will *reduce* to the bound in [87] (only for small sets).

2. While most high dimensional expanders can be symmetrized to satisfy the assumed vertex uniformity [159], it is worth noting the condition can likely be relaxed to

a bound on $\max_{s \in X(k)} \mathbb{P}_{v \in X(1)}[v \in s]$ with some additional effort. Further, some assumption of this sort is necessary. To see this, take the complete k-partite complex X where $X[i] = \{x_i\}$ are singletons for all $i < k$, and $X[k] = \{y_1, y_2, \ldots, y_n\}$. That is $X(k) = \{\{x_0, x_1, \ldots, x_{k-1}\} \cup \{y_j\} \mid \ j = 1, 2, \ldots, n\}$ are all possible faces between the different parts. For any $f : X(1) \to [0,1]$ we note that $\mathbb{E}[f] = \frac{\sum_{i=1}^{k-1} f(x_0)}{k} + \frac{1}{k}\mathbb{E}_{y_i \in X(k)}[f(y_i)]$. Thus for every s, $|\mathbb{E}_{v \subseteq s}[f(v)] - \mathbb{E}[f]| \leq \frac{1}{k}$. In particular, for every $\varepsilon > \frac{1}{k}$, the inclusion graph is an $(\varepsilon, 0)$-sampler.

3. We note that a similar bound can be achieved by essentially the same proof assuming only $\frac{\mu}{3}$-hitting, albeit at the cost of worse dependence on ε. In particular, assuming only that X is $\Theta(1)$-hitting, one can still prove a lower bound against $(\varepsilon, \exp(-O(\varepsilon \frac{k}{i})))$-sampling.

5.11.2 High Dimensional Expansion

Having shown that the concentration bounds of Theorem 5.5.1 cannot be quantitatively improved, we turn to understanding the extent to which our requirements on the *expansion* of the underlying complex are necessary for strong concentration. We give two lower bounds to this effect. First, we argue that concentration is not implied by local-spectral expansion at or below the TD-barrier.

Proposition 5.11.7 (Lower Bounds at the TD-Barrier)**.** *For every $\beta < 1$ and $d \in \mathbb{N}$, there exists a family of 1-TD complexes $\{X_n\}$ such that $(X_n(d), X_n(1))$ is not a $(\frac{1}{2}, \beta)$-additive sampler.*

Thus in this sense our results giving concentration bounds for λ-TD complexes for any $\lambda < 1$ are tight. Note that by Lemma 5.4.15, the above also implies failure of sampling for these complexes for any (large enough) $k \leq d$ (this can also be shown directly by the same method).

The proof of Proposition 5.11.7 is based on the 'product' complexes of Golowich [170], whose construction we quickly describe. Given a weighted graph G defines the product-complex X_G as

$$X_G(d) = \left\{ \{(v_1, s_1), \ldots, (v_d, s_d)\} : \{v_i\}_{i\in[d]} \in E, \{s_i\}_{i\in[d]} \in \binom{[n]}{d} \right\}.$$

The measure of each face $\sigma = \{(v_1, s_1), \ldots, (v_d, s_d)\}$ corresponding to an edge $\{s, t\}$ is proportional to $w(s, t) \cdot f(j)$, where f is some weighting function dependent only on the number $v_i = s$.

Proof of Proposition 5.11.7. Let $G = K^{(1)}_{n/2} \amalg K^{(1)}_{n/2} \amalg \{e\}$, that is two disjoint copies of $K_{n/2}$ with an additional edge e passing between them, and take the weights of every edge to be uniform. [170] shows that there exists a weighting function f such that X_G is connected and has expansion $\frac{1}{d}$ in every $(d-2)$-link. By construction, the 1-skeleton of X is a $(d+1)$-cover of G with edges $(\{v_0, s_0\}, \{v_1, s_1\})$ whenever $\{v_0, v_1\}$ is an edge in E, and $s_0 \neq s_1$. With this in mind, define the set A to be the indicator of the first clique in each part:

$$A \coloneqq \{(v, s) : v \in K^{(1)}_{n/2}\}$$

A clearly has measure $\frac{1}{2}$ by construction, while on the other hand any $\sigma \in X(d)$ corresponding to an edge $e' \neq \{e\}$ is either entirely contained in or entirely misses A. Since all edges are evenly weighted, the probability that a d-face corresponds to $\{e\}$ is at most $O(n^{-2})$, which gives the desired bound for large enough n. □

The above proof extends immediately to a lower bound on sampling of X_G under any weight function and any graph G simply by taking A to be any (nearly) balanced subset in G with expansion (near) $\frac{1}{2}$. It is an easy exercise to show that choosing A randomly suffices (and namely that such a set always exists).

Our second lower bound studies a slightly different regime. 1-TD complexes typically

have only constant local-spectral expansion at low level links, so Proposition 5.11.7 does not, for instance, rule out showing $o_d(1)$-local-spectral HDX satisfy strong concentration. Using similar ideas to our degree lower bounds, we show at least some inverse polynomial local-spectral expansion is required to have strong concentration.

Proposition 5.11.8. *For any $c < \frac{1}{2}$ and large enough $d \in \mathbb{N}$, there exists a family $\{X_n\}$ of d^{-c}-two-sided local spectral expanders such that for any $\beta < 1$, $(X_n(d), X_n(1))$ is not a (d^{-c}, β)-additive sampler.*

The proof of Proposition 5.11.8 relies on another construction of Golowich [171] building inverse polynomial HDX from Cayley complexes.

Theorem 5.11.9 ([171])**.** *For every $c < \frac{1}{2}$ and large enough $d \in \mathbb{N}$, there exists a family of d-uniform, $2^{\Omega_d(\sqrt{\log(n)})}$-degree, $\frac{1}{d^c}$-two-sided local spectral expanders $\{X_n\}$ whose 1-skeletons are Cayley graphs over $\mathbb{F}_2^n$ with two-sided expansion $\lambda = \Theta(\frac{1}{d^c})$.*

We can now use Lemma 5.9.10 and the Cayley structure of X's 1-skeleton to prove Proposition 5.11.8.

Proof of Proposition 5.11.8. Since the 1-skeleton of X is a Cayley graph over $\mathbb{F}_2^n$, there is an eigenvector $g_0 : \mathbb{F}_2^n \to \{\pm 1\}$ whose eigenvalue is λ. Let us consider the indicator g of the set $A = \{v \in \mathbb{F}_2^n \mid \ g_0(v) = 1\}$ whose measure is $\mathbb{P}[A] = \frac{1}{2}$. This indicator is equal to $g = \frac{1}{2} + \frac{1}{2}g_0$. We note that for every v,

$$\mathop{\mathbb{P}}_{u \sim v}[A] = \mathop{\mathbb{E}}_{u \sim v}[g(u)] = \frac{1}{2} + \frac{1}{2} \mathop{\mathbb{E}}_{u \sim v}[g_0(u)].$$

The function g_0 is an eigenvalue so $\frac{1}{2}\mathbb{E}_{u \sim v}[g_0(u)] = \frac{\lambda}{2}g_0(v) = \pm\frac{\lambda}{2}$. Thus, $\mathbb{P}_{u \sim v}[A] - \mathbb{P}[A] = \pm\frac{\lambda}{2}$ so for every $\varepsilon \leq \frac{\lambda}{2}$ and any $\beta < 1$, the underlying graph *is not* a (ε, β)-sampler. By Lemma 5.9.10, this implies that either the containment graph is not a $(2\varepsilon, \beta)$-sampler, or the containment graph of some link of a vertex is not a $(2\varepsilon, \frac{1}{2})$-sampler. Since the vertex

links of $\{X_n\}$ also give an infinite family of $\frac{1}{d^c}$-two-sided local spectral expanders this implies the result. □

This counterexample poses the following question: For $\varepsilon \in [d^{-c}, d^{-c/2}]$ what is the worst λ which still promises that the containment graph is a $(\varepsilon, \exp(-d^c))$-sampler for some $c > 0$?

5.11.3 Expander-Walks

Finally we take a step back from concentration itself and look at necessity of *local* concentration for our applications (that is concentration of the links of X and in particular the locally nice property). We argue this is at least in some sense necessary: classical samplers such as expander walks which behave poorly under restriction fail reverse hypercontractivity. Recall the expander walk complexes W_G defined in Section 5.10.1.

Proposition 5.11.10. *For any $0 < \gamma < 1$ and $\lambda > 0$, there are infinitely many pairs $k, n \in \mathbb{N}$ with a corresponding λ-expanders $G = ([n], E)$ and balanced subset $A \subset W_G(k)$ satisfying:*

$$\mathbb{P}_{t,t' \sim UD_{k,\gamma k}} [t' \in X(1) \setminus A \text{ and } t \in A] \leq 2^{-\Omega_\gamma(k)}$$

Proof. Take k odd and n even. It suffices to take any regular graph G with $\text{girth}(G) > k$. Define A to be the set of walks whose center vertex is in $\{1, 2, \ldots, \frac{n}{2}\}$. Notice that A is indeed a balanced function, as every element is the center of the same number of random walks in a regular graph (see e.g. [10] for details).

On the other hand, it is easy to see A has very poor expansion due to the girth of G. In particular, the only way to leave A (resp. $X(1) \setminus A$) is if the walk re-samples all of $[k/2 + 1]$, or all of $[k] \setminus [k/2]$. If neither event occurs, there exist indices $i < \lceil \frac{k}{2} \rceil < j$ which were not re-sampled, and the girth of the graph ensures there is only one option for the $\lceil \frac{k}{2} \rceil$th element (namely its starting position). One can compute directly the probability either event occurs is at most $2^{-\Omega_\gamma(k)}$.

To complete the proof, we need an infinite family of regular λ-expanders for any $\lambda > 0$ with super-constant girth. Many such constructions exist in the literature, including the classical Ramanujan graphs of [289]. □

Note that for any fixed $\gamma, C, \ell > 0$, taking k in Proposition 5.11.10 sufficiently large means we can always find subsets $A, B \subset X(k)$ such that

$$\mathop{\mathbb{P}}_{t,t' \sim UD_{k,\gamma k}}[t \in A \wedge t' \in B] \leq C\,\mathbb{P}[A]^{\ell}\,\mathbb{P}[B]^{\ell},$$

violating reverse hypercontractivity for all parameter settings. We remark that the failure of expander walks in agreement testing is well known, and can be derived e.g. from [166].

Acknowledgements

We thank Irit Dinur, Tali Kaufman, and Shachar Lovett for many helpful discussions on high dimensional expanders, sampling, and agreement testing. We thank Swastik Kopparty and Madhur Tulsiani for helpful discussion of list-decoding and amplification techniques for locally testable codes. We thank Vedat Alev for helpful discussion on concentration of measure and for pointing the authors to the improved swap-walk bounds of [187], and Noam Lifshitz for pointers to work applying reverse hypercontractivity in extremal combinatorics. We thank Alex Lubotzky for pointing out the connection between high dimensional expanders and geometric overlap. Finally the authors thank the Simons Institute for the Theory of Computing for graciously hosting them for part of the duration of this work, and MH would further like to thank Irit Dinur and the Weizmann Institute for the same.

5.12 Concentration for the Complete Complex

In this section we show near-optimal concentration bounds for the complete complex.

Claim 5.12.1. Let X be the complete complex on n vertices, that is $X(k) = \binom{[n]}{k}$. Let $\alpha, \delta > 0$. Then for any $\ell \leq k$, the containment graph $G = (X(k), X(\ell))$ is an $(\alpha, \frac{4}{\alpha\delta}\exp(-\frac{\delta^2}{12}\alpha\lfloor\frac{k}{\ell}\rfloor), \delta)$-multiplicative sampler and $(\varepsilon, \frac{2}{\varepsilon}\exp(-\frac{\varepsilon^2}{8}\lfloor\frac{k}{\ell}\rfloor))$-additive sampler for any $\varepsilon > 0$.

Proof. Fix and $A \subseteq X(\ell)$ such that $\mathbb{P}[A] \geq \alpha$. We apply Lemma 5.5.9 in the following distribution. The distribution we use samples $(s_1, s_2, \ldots, s_m, t) \sim D$ such that $(s_1, s_2, \ldots, s_m)$ are independent and $t \in X(k)$ is a uniform face conditioned on containing all s_i. Obviously, (s_i, t) is a uniform pair of ℓ and k faces where $s_i \subseteq t$. Moreover, by Chernoff's bound for independent sampling

$$\mathbb{P}_{(s_1,s_2,\ldots,s_m,t)\sim D}\left[\left|\frac{|s_i \in A|}{m} - \mathbb{P}[A]\right| > \frac{\delta}{2}\mathbb{P}[A]\right] \leq 2\exp(\frac{\delta^2}{12}\alpha m).$$

The claim follows from Lemma 5.5.9. The additive bound follows by the same argument applying Hoeffding's inequality. □

Note that we do not restrict the number of vertices n in this claim. Since the k-skeleton of the complete complex is only an $\frac{1}{n-k}$-two sided spectral expander, this claim does not follow from Theorem 5.5.1.

5.13 Concentration for the Swap Complex

In this appendix we prove concentration of measure for the swap complex. Recall the swap complex $C = C_{\ell,k,n}$ is the $d = \lfloor\frac{k}{\ell}\rfloor$-uniform complex whose vertices are $C(0) = \binom{[n]}{\ell}$ and whose top level faces are $C(d) = \{\{s_1, s_2, \ldots, s_d\} \mid \forall i \neq j,\ s_i \cap s_j = \emptyset\}$ endowed with

the uniform distribution. We prove a Chernoff bound for the swap complex under the assumption that $n \geq \Omega(k\ell)$

Theorem 5.13.1 (Theorem 5.7.15 Restated)**.** *Let $C = C_{\ell,k,n}$ for $n \geq (k+1)\ell$, $f : C(0) \to [0,1]$ and $\varepsilon > 0$:*

1. ***Upper tail**: $\mathbb{P}[U_{1,\frac{k}{\ell}} f - \mathbb{E}[f] > \varepsilon] \leq e^{-c\varepsilon^2 \frac{k}{\ell}}$*
2. ***Lower tail**: $\mathbb{P}[U_{1,\frac{k}{\ell}} f - \mathbb{E}[f] < -Var] \leq e^{-c\varepsilon^2 \frac{k}{\ell}}$*

for some universal constant $c > 0$.

Proof. Recall it is enough to show that for any function 1_A, the variables $X_i = 1_A(v_i)$ generated by drawing a d-face from C are negatively correlated [133]. Toward this end, condition on an ℓ-set $v \in C(0)$ appearing in the face $s \in C(d)$. The conditional probability that a second ℓ-set $w \in C(0)$ appears in s is:

$$\mathop{\mathbb{P}}_{s \sim C(d)}[w \in s \mid v \in s] = \begin{cases} 0 & \text{if } |v \cap w| > 0 \\ \frac{\frac{k}{\ell}-1}{\binom{n-\ell}{\ell}} & \text{if } |v \cap w| = 0 \end{cases}$$

On the other hand the unconditioned probability that w appears in s is $\frac{k}{\ell\binom{n}{\ell}}$. Thus $C(d)$ is negatively correlated so long as

$$1 - \frac{\ell}{k} \leq \frac{\binom{n-\ell}{\ell}}{\binom{n}{\ell}}$$

The righthand side is exactly

$$\prod_{j=0}^{\ell-1} \left(1 - \frac{\ell}{n-j}\right) \geq \left(1 - \frac{\ell}{n-\ell}\right)^{\ell} \geq 1 - \frac{\ell^2}{n-\ell}$$

Thus $C(d)$ is negatively correlated as long as $\frac{\ell^2}{n-\ell} \leq \frac{\ell}{k}$, which holds for $n \geq (k+1)\ell$ $\square$

We conjecture that $C_{\ell,k,n}$ should actually satisfy a Chernoff bound whenever $n \geq ck$ for some sufficiently large constant $c > 1$. In fact, because the links of the swap complex are exactly Kneser graphs, this would follow from the local-to-global entropy contraction framework of [96, 186] if one could prove an optimal entropy contraction bound for Kneser graphs. While it is possible to derive some entropic bounds on Kneser graphs from their chi-squared mixing time [321], they do not seem to be strong enough to prove strong concentration.

5.14 Outstanding Proofs on Samplers and Concentration

5.14.1 Sampling

Claim 5.14.1 (Claim 5.4.8 Restated). Let $\beta, \delta > 0$, let $\delta' > \delta$ and $\alpha < \frac{\min\{\delta, 0.5\}}{1+\delta}$. Then for every (α, β, δ)-sampler $G = (L, R, E)$, it holds that $G_{op} \coloneqq (R, L, E)$ is a $(\frac{1-\alpha(1+\delta)}{\alpha(\delta'-\delta)}\beta, 2\alpha, \delta')$-sampler.

Proof of Claim 5.4.8. Let $A \subseteq L$ be such that $\mathbb{P}[A] \geq \frac{1-\alpha(1+\varepsilon)}{\alpha(\varepsilon'-\varepsilon)}\beta$. Let

$$B_S = \left\{ v \in R \,\middle|\, \mathop{\mathbb{P}}_{u \in L}[u \in A \mid \ u \sim v] < (1-\varepsilon')\,\mathbb{P}[A] \right\}$$

and let

$$B_B = \left\{ v \in R \,\middle|\, \mathop{\mathbb{P}}_{u \in L}[u \in A \mid \ u \sim v] > (1+\varepsilon')\,\mathbb{P}[A] \right\}.$$

Showing that $\mathbb{P}[B_B], \mathbb{P}[B_S] \leq \alpha$ will prove the claim. Let us begin with B_S. Assume toward contradiction that $\mathbb{P}[B_S] > \alpha$. We denote by $\mathbf{1}_A : L \to \{0,1\}$ and $\mathbf{1}_{B_S} : R \to \{0,1\}$ the indicators of A, B_S respectively. On the one hand

$$\mathop{\mathbb{E}}_{uv \in E}[\mathbf{1}_A(u)\mathbf{1}_{B_S}(v)] < \mathbb{P}[B_S]\,(1-\varepsilon')\,\mathbb{P}[A], \tag{5.36}$$

since the product of indicators is one iff $v \in B_S$, in which case, there are at most a $(1-\varepsilon')\,\mathbb{P}[A]$ fraction of $u \in A$ adjacent to v by the definition of B_S. On the other hand, by the sampling properties of (L, R, E) there is at most a β-fraction of $u \in L$ such that $\mathbb{P}_{v\in R}[v \in B_S \mid \ v \sim u] < (1-\varepsilon)\,\mathbb{P}[B_S]$. Hence

$$\mathop{\mathbb{E}}_{uv\in E}[\mathbf{1}_A(u)\mathbf{1}_{B_S}(v)] \geq \mathbb{P}[B_S]\,(1-\varepsilon)(\mathbb{P}[A]-\beta).^{31} \tag{5.37}$$

Combining (5.36) with (5.37) yields

$$\mathbb{P}[B_S]\,\mathbb{P}[A]\,(\varepsilon'-\varepsilon) < \beta\,\mathbb{P}[B_S]\,(1-\varepsilon)$$

or

$$\mathbb{P}[A] < \frac{1-\varepsilon}{\varepsilon'-\varepsilon}\beta \overset{(\alpha<\frac{\varepsilon}{1+\varepsilon})}{\leq} \frac{1-\alpha(1+\varepsilon)}{\alpha(\varepsilon'-\varepsilon)}\beta,$$

a contradiction to the lower bound on the size of A.

Let us bound B_B. Let $\mathbf{1}_{B_B} : R \to \{0,1\}$ be the indicator of B_B. Similar to (5.36) we have that

$$\mathop{\mathbb{E}}_{uv\in E}[\mathbf{1}_A(u)\mathbf{1}_{B_S}(v)] > \mathbb{P}[B_B]\,(1+\varepsilon')\,\mathbb{P}[A]\,. \tag{5.38}$$

By the sampling properties of (L, R, E) there is at most a β-fraction of $u \in L$ such that $\mathbb{P}_{v\in R}[v \in B_B \mid \ v \sim u] > (1+\varepsilon)\,\mathbb{P}[B_B]$. So similar to (5.37) we have that

$$\mathop{\mathbb{E}}_{uv\in E}[\mathbf{1}_A(u)\mathbf{1}_{B_S}(v)] \leq \beta + \mathbb{P}[B_B]\,(1+\varepsilon)(\mathbb{P}[A]-\beta). \tag{5.39}$$

Combining (5.38) with (5.39) yields

$$\mathbb{P}[B_B]\,\mathbb{P}[A]\,(\varepsilon-\varepsilon') < \beta\,(1-\mathbb{P}[B_B]\,(1+\varepsilon))$$

[31] $\mathbb{P}[A] > \beta$ by assumption.

or

$$\mathbb{P}[A] < \frac{1-\mathbb{P}[B_B](1+\varepsilon)}{\mathbb{P}[B_B](\varepsilon'-\varepsilon)}\beta \leq \frac{1-\alpha(1+\varepsilon)}{\alpha(\varepsilon'-\varepsilon)}\beta,$$

where the last inequality is because the function $x \mapsto \frac{1-(1+\varepsilon)x}{x(\varepsilon'-\varepsilon)}\beta$ is monotone increasing when $x \in [0, \frac{0.5}{1+\varepsilon}]$ and α is in this interval. This is a contradiction to the lower bound on the size of A. □

Claim 5.14.2 (Claim 5.4.9 Restated). Let $G = (L, R, E)$ be a bipartite graph.

1. If G is a (β, δ)-additive sampler then G is a $(C\delta, \beta, \frac{1}{C})$-multiplicative sampler for any $C > 1$.

2. If G is a (α, β, δ)-multiplicative sampler for $\alpha \leq \frac{1}{2}$. Then G is a (β, δ)-additive sampler, where $\delta = \max\{\delta, (1+\delta)(\alpha+p)\}$ and $p = \max_{v\in R}\mathbb{P}[v]$.

Proof of Claim 5.4.9. The first item follows immediately from the definition of multiplicative samplers. If $\mathbb{P}[A] \geq C\varepsilon$ then $\frac{1}{C}\mathbb{P}[A] \geq \varepsilon$ so by the promise of additive sampling it holds that

$$\begin{aligned}&\mathbb{P}_{v\in L}\left[\left|\mathbb{P}_{u\in R}[u \in A \mid \ u \sim v] - \mathbb{P}[A]\right| > \frac{1}{C}\mathbb{P}[A]\right]\\ \leq &\mathbb{P}_{v\in L}\left[\left|\mathbb{P}_{u\in R}[u \in A \mid \ u \sim v] - \mathbb{P}[A]\right| > \varepsilon\right] \leq \beta.\end{aligned}$$

We turn to the second item. Let $A \subseteq R$. For every $u \in L$ it holds that

$$\begin{aligned}\left|\mathbb{P}_{u\in R}[u \in A \mid \ u \sim v] - \mathbb{P}[A]\right| &= \left|1 - \mathbb{P}_{u\in R}[u \in A \mid \ u \sim v] - (1-\mathbb{P}[A])\right|\\ &= \left|\mathbb{P}_{u\in R}[u \in R\setminus A \mid \ u \sim v] - \mathbb{P}[R\setminus A]\right|,\end{aligned}$$

so we assume that without loss of generality $\mathbb{P}[A] \leq \frac{1}{2}$. We need to show that

$$\mathbb{P}_{v\in L}\left[\left|\mathbb{P}_{u\in R}[u \in A \mid \ u \sim v] - \mathbb{P}[A]\right| > \delta\right] \leq \beta. \tag{5.40}$$

If $\mathbb{P}[A] \geq \alpha$ then this holds from the multiplicative sampling guarantee. Hence it suffices to show that this inequality *always holds* for sets of relative size smaller that δ.

First we note that if $u \in L$ is such that $\mathbb{P}_{u \in R}[u \in A \mid u \sim v] < \mathbb{P}[A]$ then $|\mathbb{P}_{u \in R}[u \in A \mid u \sim v] - \mathbb{P}[A]| \leq \mathbb{P}[A] \leq \delta$ so

$$\mathop{\mathbb{P}}_{v \in L}\left[\left|\mathop{\mathbb{P}}_{u \in R}[u \in A \mid u \sim v] - \mathbb{P}[A]\right| > \delta\right] = \mathop{\mathbb{P}}_{v \in L}\left[\mathop{\mathbb{P}}_{u \in R}[u \in A \mid u \sim v] > \delta + \mathbb{P}[A]\right].$$

In this case we find a subset $B \supseteq A$ such that $\mathbb{P}[B] \leq \alpha + p$ (we can do so by adding vertices to A one-by-one which is where p comes into play). Then

$$\begin{aligned}
\mathop{\mathbb{P}}_{v \in L}\left[\mathop{\mathbb{P}}_{u \in R}[u \in A \mid u \sim v] > \delta + \mathbb{P}[A]\right] &\leq \mathop{\mathbb{P}}_{v \in L}\left[\mathop{\mathbb{P}}_{u \in R}[u \in B \mid u \sim v] > \delta + \mathbb{P}[A]\right] \\
&\leq \mathop{\mathbb{P}}_{v \in L}\left[\mathop{\mathbb{P}}_{u \in R}[u \in B \mid u \sim v] > (1+\varepsilon)\,\mathbb{P}[B]\right] \\
&\leq \beta.
\end{aligned}$$

The claim follows. □

One can also remove the dependence on δ in the second item.

Claim 5.14.3 (Claim 5.4.10 Restated). Let $\beta, \alpha_0 > 0$. If for every $\alpha > \alpha_0$ it holds that G is an $(\alpha, \beta, \frac{\alpha_0}{\sqrt{\alpha}})$-multiplicative sampler, then G is a$(\beta, 2(\alpha + p))$-additive sampler where $p = \max_{v \in R} \mathbb{P}[v]$.

Proof of Claim 5.4.10. The proof is the same as the proof of the second item in Claim 5.4.9. the only difference is that for *large* sets, we still have that

$$\mathop{\mathbb{P}}_{v \in L}\left[\left|\mathop{\mathbb{P}}_{u \in R}[u \in A \mid u \sim v] - \mathbb{P}[A]\right| > 2\alpha_0 + p\right] \leq \beta \tag{5.41}$$

because if $\mathbb{P}[A] = \alpha$ then $\frac{\alpha_0}{\sqrt{\alpha}}\alpha \leq \alpha_0$. □

Claim (Restatement of Claim 5.4.11). Let $G = (L, R, E)$ be an (ε, β)-additive sampler such that every $r \in R$ has degree at least k and every $v \in L$ has probability at most $\frac{1}{k}$. Assume that $\exp(-0.01\varepsilon^2 k)) < \frac{1}{4}$. Then G is also a $(4\beta, 2\varepsilon)$-function additive sampler.

Proof of Claim 5.4.11. Fix $f : L \to [0, 1]$. We prove a $(2\beta, 2\varepsilon)$ upper tail bound. The lower bound is similar. Let $A \sim \mathcal{P}(L)$ be a random subset where every vertex is inserted into A independently with probability $p_v = f(v)$. Since G is a sampler, for every possible outcome A we have $\mathbb{P}_{r\in R}\left[\mathbb{P}_{v\sim r}\left[v \in A\right] - \mathbb{P}\left[A\right] > \varepsilon\right] < \beta$. Let $\mathbf{1}(A, r)$ be the indicator for this event. In particular, we have that $\mathbb{E}_{A,r\sim R}\left[\mathbf{1}(A, r)\right] \leq \beta$. By Markov, the fraction of $r_0 \in R$ such that $\mathbb{E}_A\left[\mathbf{1}(A, r_0)\right] > \frac{1}{2}$ is at most 2β. On the other hand, we will show that if

$$\mathop{\mathbb{E}}_{v\sim r_0}\left[f(v)\right] \geq \mathop{\mathbb{E}}_{v\in L}\left[f(v)\right] + 2\varepsilon$$

then $\mathbb{E}_A\left[\mathbf{1}(A, r_0)\right] > \frac{1}{2}$. Indeed, fix such an r_0. Observe that if $\mathbb{P}_{v\in L}\left[A\right] < \mathbb{E}_{v\in L}\left[f\right] + \frac{1}{2}\varepsilon$ and $\mathbb{P}_{v\sim r_0}\left[A\right] > \mathbb{E}_{v\sim r_0}\left[f(v)\right] - \frac{1}{2}\varepsilon$, then

$$\mathop{\mathbb{P}}_{v\sim r_0}\left[A\right] > \mathop{\mathbb{E}}_{v\sim r_0}\left[f(v)\right] - \frac{1}{2}\varepsilon \geq \mathop{\mathbb{E}}_{v\in L}\left[f(v)\right] + \frac{3}{2}\varepsilon \geq \mathbb{P}\left[A\right] + \varepsilon.$$

It is direct to check that $\mathbb{E}_A\left[\mathbb{P}\left[A\right]\right] = \mathbb{E}[f]$ and that $\mathbb{E}_A\left[\mathbb{P}_{v\sim r_0}\left[A\right]\right] = \mathbb{E}_{v\sim r_0}\left[f(v)\right]$. Hoeffding's inequality therefore bounds the probability that either $\mathbb{P}_{v\in L}\left[A\right] < \mathbb{E}_{v\in L}\left[f\right] + \frac{1}{2}\varepsilon$ or $\mathbb{P}_{v\sim r_0}\left[A\right] > \mathbb{E}_{v\sim r_0}\left[f(v)\right] - \frac{1}{2}\varepsilon$ by $2\exp(-0.01\varepsilon^2 k) < \frac{1}{2}$ and $\mathbb{E}_A\left[\mathbf{1}(A, r_0)\right] > \frac{1}{2}$ follows as desired.

□

Finally we prove the basic Chebyshev-type sampling bound from expansion.

Claim (Restatement of Claim 5.7.3). Let $G = (L, R, E)$ be a λ-bipartite expander and let A be its bipartite adjacency operator. Let $f : L \to [0, 1]$ be a function with $\mathbb{E}[f] \geq \mu$ and

let $0 < \varepsilon < \mu$. Then

$$\mathbb{P}_{u \in R}[Af(u) < \mu - \varepsilon] < \frac{\lambda^2 \mu}{\varepsilon^2}.$$

Proof of Claim 5.7.3. Dinur and Kaufman observe the following bound

$$\mathbb{P}_{u \sim R}[|Af(u) - \mu| > \varepsilon] < \frac{\lambda^2 Var(f)}{\varepsilon^2}.$$

Let $f^{\perp} = f - \mu$. Then $Af(v) - \mu = Af^{\perp}$. By Chebyshev's inequality, $\mathbb{P}[T] \leq \frac{\mathbb{E}[(Af^{\perp})^2]}{\varepsilon^2}$ and by λ-expansion $\mathbb{E}[(Af^{\perp})^2] \leq \lambda^2 \mathbb{E}[(f^{\perp})^2] = \lambda^2 Var(f)$. The claim follows. Now by the above we can write

$$\mathbb{P}[Af(u) < \mu - \varepsilon] \leq \mathbb{P}[|Af(u) - \mathbb{E}[f]| > \mathbb{E}[f] - \mu + \varepsilon] \leq \frac{\lambda^2 Var(f)}{(\varepsilon + (\mathbb{E}[f] - \mu))^2}.$$

Noting that $Var(f) \leq \mathbb{E}[f] = \mu + (\mathbb{E}[f] - \mu)$ and denoting $x = \mathbb{E}[f] - \mu$, this is at most $\frac{\lambda^2(\mu + x)}{(\varepsilon + x)^2} \leq \frac{\lambda^2 \mu}{\varepsilon^2}$. □

5.14.2 Concentration

Lemma 5.14.4 (Lemma 5.4.15 Restated)**.** *Let X be a d-uniform simplicial complex, $k \leq d$, and $f : X(k) \to [0,1]$ a function satisfying*

1. ***Upper Tail:*** $\mathbb{P}_{X(k)}[f - \mathbb{E}[f] > t] \leq f_{up}(t)$
2. ***Lower Tail:*** $\mathbb{P}_{X(k)}[f - \mathbb{E}[f] < -t] \leq f_{low}(t)$.

for some functions $f_{up}, f_{low} : \mathbb{R}_+ \to [0,1]$. Then the d-lift $U_{k,d}f : X(d) \to \mathbb{R}$ satisfies:

1. ***Upper Tail:*** $\mathbb{P}_{X(d)}[U_{k,d}f - \mathbb{E}[f] > t] \leq f_{up}(\frac{t}{2})(1 - \pi_{low}^{d,k,f}(\frac{t}{2}))^{-1}$
2. ***Lower Tail:*** $\mathbb{P}_{X(d)}[U_{k,d}f - \mathbb{E}[f] < -t] \leq f_{low}(\frac{t}{2})(1 - \pi_{up}^{d,k,f}(\frac{t}{2}))^{-1}$.

Proof. We give the argument for the upper tail. The argument for the lower tail is analogous. Assume for the sake of contradiction that

$$\mathop{\mathbb{P}}_{X(d)}[U_{k,d}f - \mu > 2t] > f_{up}(t)\left(1 - \pi_{low}^{f,k,d}(t)\right)^{-1}.$$

Since a k-face can be drawn by first drawing a d-face, then a uniformly random k-subface, we can use concentration of the complete complex to derive a contradiction. In particular, let E_1 denote the event that $U_{k,d}f - \mu > 2t$. Then:

$$\begin{aligned}
\mathop{\mathbb{P}}_{\{v_0,\ldots,v_k\}\in X(k)}[U_{k,d}f - \mu > t] &\geq \mathbb{P}[E_1] \mathop{\mathbb{P}}_{r\subset s}[f(r) - \mu > t \mid s \in E_1] \\
&\geq \mathbb{P}[E_1](1 - \mathop{\mathbb{P}}_{r\subset s}[f(r) - \mathbb{E}[f] < t \mid s \in E_1]) \\
&\geq \mathbb{P}[E_1](1 - \mathop{\mathbb{P}}_{r\subset s}[f(r) - U_{k,d} < -t \mid s \in E_1]) \\
&> f_{up}(t)
\end{aligned}$$

□

Lemma 5.14.5 (Lemma 5.4.16 Restated)**.** *Let X be a d-uniform simplicial complex and $k \leq d$. Assume there exist functions $f_{up}(t,\nu)$ and $f_{low}(t,\nu)$ such that any ν-Lipschitz $f : X(d) \to \mathbb{R}$ satisfies:*

1. ***Upper Tail**: $\mathbb{P}[f - \mathbb{E}[f] > t] \leq f_{up}(t,\nu)$*
2. ***Lower Tail**: $\mathbb{P}[f - \mathbb{E}[f] < -t] \leq f_{low}(t,\nu)$.*

Then any function $f' : X(k) \to \mathbb{R}$ with ν-bounded difference satisfies:

1. ***Upper Tail**: $\mathbb{P}[f' - \mathbb{E}[f'] > t] \leq f_{up}(\frac{t}{2}, \frac{k}{d}\nu) + e^{-\frac{t^2}{4\nu}}$*
2. ***Lower Tail**: $\mathbb{P}[f' - \mathbb{E}[f'] < -t] \leq f_{low}(\frac{t}{2}, \frac{k}{d}\nu) + e^{-\frac{t^2}{4\nu}}$.*

Proof. Let $f : X(k) \to \mathbb{R}$ be a function with ν-bounded difference and g its d-lift $g = U_{k,d}f$. Since a face $r \in X(k)$ can equivalently be sampled by first drawing a face $s \in X(d)$, then subsampling r uniformly from s, we can bound the concentration of f as:

$$\begin{aligned}\mathbb{P}[f(r) - \mathbb{E}[f] \geq t] &\leq \mathop{\mathbb{P}}_{r\subset s}[g(s) - \mathbb{E}[f] \geq t/2 \vee f(r) - g(s) \geq t/2] \\ &\leq \mathop{\mathbb{P}}_{r\subset s}[g(s) - \mathbb{E}[g] \geq t/2] + \mathop{\mathbb{P}}_{r\subset s}\left[f(r) - \mathop{\mathbb{E}}_{r\subset s}[f(r)] \geq t/2\right]\end{aligned}$$

where we have used the fact that $\mathbb{E}[g] = \mathbb{E}[f]$. Informally, the idea is then to argue the first term is small due to exponential concentration of $X(d)$ and the fact that the lift g is itself more concentrated than f, and the second term is small by subgaussian concentration of the complete complex.

We first argue that if f has ν-bounded difference, g is $\frac{k}{d}\nu$-Lipschitz. Given $s \in X(d)$ and $i \in [d]$, let $s^{(i)}$ denote a copy of s with the ith coordinate re-sampled. Expanding ν_g we have:

$$\begin{aligned}\nu_g &\leq \sum_{i=1}^{d} \mathop{\mathbb{E}}_{s,s^{(i)}}\left[\left(g(s) - g(s^{(i)})\right)_+^2\right] \\ &= \sum_{i=1}^{d} \mathop{\mathbb{E}}_{s,s^{(i)}}\left[\left(\mathop{\mathbb{E}}_{t\subset s}[f] - \mathop{\mathbb{E}}_{t\subset s^{(i)}}[f]\right)_+^2\right] \\ &= \frac{1}{\binom{d}{k}^2}\sum_{i=1}^{d} \mathop{\mathbb{E}}_{s,s^{(i)}}\left[\left(\sum_{T\subseteq[d]} f(s_T) - f(s_T^{(i)})\right)_+^2\right]\end{aligned}$$

where we recall s_T is the k-face consisting of the elements of s indexed by T. Observe that the inner term is non-zero only when $i \in T$, since otherwise $s_T = s_T^{(i)}$. Thus by Cauchy-Schwarz, we have:

$$\nu_f \leq \frac{\binom{d-1}{k-1}}{\binom{d}{k}^2}\sum_{i=1}^{d} \mathop{\mathbb{E}}_{s,s^{(i)}}\left[\sum_{T\subseteq[d]}\left(f(s_T) - f(s_T^{(i)})\right)^2\right]$$

$$\leq \frac{\binom{d-1}{k-1}}{\binom{d}{k}^2} \sum_{i=1}^{d} \mathop{\mathbb{E}}_{s,s^{(i)}} \left[\sum_{i \in T \subseteq [d]} \frac{\nu}{k} \right]$$
$$= \frac{k}{d}\nu$$

where we have used the fact that $s(T)$ and $s'_{(i)}(T)$ are neighboring k-faces in the down-up walk on level k. We can now bound the first term of our upper tail by assumption as

$$\mathop{\mathbb{P}}_{r \subset s}[g(s) - \mathbb{E}[g] \geq t/2] \leq f_{up}\left(\frac{t}{2}, \frac{k}{d}\nu\right).$$

For the second term, observe that since f has ν-bounded difference as a function of $X(k)$, the restriction of f to $\binom{s}{k} \subset X(k)$ also has ν-bounded difference (since any edge (r, r') in the down-up walk restricted to s is also an edge in the walk on $X(k)$). Thus using subgaussian concentration for Lipschitz functions on the complete complex (see e.g. [102, Corollary 2]) we also have

$$\mathop{\mathbb{P}}_{r \subset s}\left[f(r) - \mathop{\mathbb{E}}_{r \subset s}[f(r)] \geq t/2\right] \leq e^{-\frac{t^2}{4\nu}}.$$

Combining the two gives the result. □

5.15 High Dimensional Expander-Mixing Lemma

In this section, we give the proof of our variant of the high-dimensional expander mixing lemma and its various corollaries.

Theorem 5.15.1 (high dimensional expander Mixing Lemma (Theorem 5.10.14 Restated))**.** *Let (X, T, ρ) be a depth D λ-tuple splitting tree with k leaves. Denote by d_i the depth of the leaf labeled i. Then for any family of functions $\{f_i : X[i] \to \mathbb{R}\}_{i \in [k]}$:*

$$\left| \mathop{\mathbb{E}}_{a \in X(k-1)} \left[\prod_{i=1}^{k} f_i(a_i) \right] - \prod_{i=1}^{k} \mathop{\mathbb{E}}_{a_i \in X[i]} [f_i(a_i)] \right| \leq 3^D \lambda \prod_{i=1}^{k} \|f_i\|_{2^{d_i}}.$$

If (X, T, ρ) is instead a depth D standard λ-splitting tree, we take $\{f_i : X(1) \to \mathbb{R}\}_{i\in[k]}$ and have:

$$\left| \mathop{\mathbb{E}}_{a\in X(k-1),\pi\in S_k} \left[\prod_{i=1}^{k} f_i(a_{\pi(i)}) \right] - \prod_{i=1}^{k} \mathop{\mathbb{E}}_{a_i\in X(1)} [f_i(a_i)] \right| \leq 3^D \lambda \prod_{i=1}^{k} \|f_i\|_{2^{d_i}}.$$

Proof. We prove the partite case directly, then argue Equation (5.32) follows from partitification. It is convenient to introduce the following notation. Let $\bar{f} = (f_1, f_2, \ldots, f_k)$ such that $f_i : X[i] \to \mathbb{R}$. Let $S \subseteq [k]$. Let (T, ρ) be an ordered tree. Without loss of generality we identify every leaf $u \in \mathcal{L}$ with its label $\rho(u)$.

- Denote by $\bar{f}_S$ the sub-tuple containing only functions whose index is in S.
- Let $\pi_{\bar{f}_S} : X[S] \to \mathbb{R}$ be the product of functions in $\bar{f}_S$, that is $\pi_{\bar{f}_S}(s) = \prod_{i\in S} f_i(a_i)$ (where $s = \{a_i\}_{i\in S}$). We note that when S_1, S_2 are disjoint then $\pi_{\bar{f}_{S_1}} \cdot \pi_{\bar{f}_{S_2}} = \pi_{\bar{f}_{S_1\cup S_2}}$.
- Let $E_{\bar{f}_S} = \prod_{i\in S} \mathbb{E}[f_i]$.
- Let T_L and T_R be the sub-trees rooted by the left and right children of the root of T, respectively. Let L, R be the leaves of T_L, T_R respectively.
- For T, T_L, T_R and tuple of functions $\bar{f}$ we denote by

$$\mathcal{E}(\bar{f}_{[k]}, T) = \prod_{i=1}^{k} \|f_i\|_{2^{d_i}},\ \mathcal{E}(\bar{f}_L, T_L) = \prod_{i\in L} \|f_i\|_{2^{d_i-1}},\ \mathcal{E}(\bar{f}_R, T_R) = \prod_{i\in R} \|f_i\|_{2^{d_i-1}}.$$

In this notation we need to show:

$$\left| \mathbb{E}[\pi_{\bar{f}_{[k]}}] - E_{\bar{f}_{[k]}} \right| \leq 3^D \lambda \mathcal{E}(\bar{f}_{[k]}, T).$$

The proof is by induction on the depth D. The base case (depth 1) is trivial. Assume that the theorem holds for any set of functions on any tree of depth $\leq D - 1$. By

adding and subtracting $\mathbb{E}[\pi_{\bar{f}_L}] \cdot \mathbb{E}[\pi_{\bar{f}_R}]$, observe that

$$\left|\mathbb{E}[\pi_{\bar{f}_{[k]}}] - E_{\bar{f}_{[k]}}\right| \leq \underbrace{\left|\mathbb{E}[\pi_{\bar{f}_{[k]}}] - \mathbb{E}[\pi_{\bar{f}_L}] \cdot \mathbb{E}[\pi_{\bar{f}_R}]\right|}_{\text{I}} + \underbrace{\left|\mathbb{E}[\pi_{\bar{f}_L}] \cdot \mathbb{E}[\pi_{\bar{f}_R}] - E_{\bar{f}_{[k]}}\right|}_{\text{II}} \tag{5.42}$$

by the triangle inequality. We bound I and II separately, starting with I. Since the swap walk $S_{L,R}$ is a λ-bipartite expander, applying (5.6) to $\pi_{\bar{f}_L}, \pi_{\bar{f}_R}$ we get that the first term in the right-hand side of (5.42) is

$$\left|\mathbb{E}[\pi_{\bar{f}_{[k]}}] - \mathbb{E}[\pi_{\bar{f}_L}]\,\mathbb{E}[\pi_{\bar{f}_R}]\right| = \left|\mathbb{E}[\pi_{\bar{f}_L} \cdot \pi_{\bar{f}_R}] - \mathbb{E}[\pi_{\bar{f}_L}]\,\mathbb{E}[\pi_{\bar{f}_R}]\right| \leq \lambda \|\pi_{\bar{f}_L}\|_2 \|\pi_{\bar{f}_R}\|_2 \leq \lambda \prod_i \|f_i\|_{2^{d_i}}$$

where we have used the fact that

$$\|\pi_{\bar{f}_L}\|_2 \leq \prod_{i \in L} \|f_i\|_{2^{d_i}}, \quad \text{and} \quad \|\pi_{\bar{f}_R}\|_2 \leq \prod_{i \in R} \|f_i\|_{2^{d_i}} \tag{5.43}$$

by repeatedly applying Cauchy-Schwarz along the splitting trees of L and R respectively.

It remains to bound II. By the triangle inequality, we can factor this term into components dependent on the left and right subtrees:

$$\begin{aligned}\left|\mathbb{E}[\pi_{\bar{f}_R}] \cdot \mathbb{E}[\pi_{\bar{f}_L}] - E_{\bar{f}_{[k]}}\right| &\leq \left|\mathbb{E}[\pi_{\bar{f}_L}] \cdot \mathbb{E}[\pi_{\bar{f}_R}] - \mathbb{E}[\pi_{\bar{f}_L}] \cdot E_{\bar{f}_R}\right| + \left|\mathbb{E}[\pi_{\bar{f}_L}] \cdot E_{\bar{f}_R} - E_{\bar{f}_{[k]}}\right| \\ &= \left|\mathbb{E}[\pi_{\bar{f}_L}]\right| \left|\mathbb{E}[\pi_{\bar{f}_R}] - E_{\bar{f}_R}\right| + \left|E_{\bar{f}_R}\right| \left|\mathbb{E}[\pi_{\bar{f}_L}] - E_{\bar{f}_L}\right|.\end{aligned} \tag{5.44}$$

Since $(X^L, T_L, \rho|_{T_L}), (X^R, T_R, \rho|_{T_R})$ are both λ-tuple splitting trees of depth $\leq D-1$,[32] the inductive hypothesis implies

$$\left|\mathbb{E}[\pi_{\bar{f}_R}] - E_{\bar{f}_R}\right| \leq 3^{D-1}\lambda \mathcal{E}(\bar{f}_R, T_R), \quad \text{and} \quad \left|\mathbb{E}[\pi_{\bar{f}_L}] - E_{\bar{f}_L}\right| \leq 3^{D-1}\lambda \mathcal{E}(\bar{f}_L, T_L)$$

[32]Recall that X^L, X^R are the simplicial sub-complexes induced by the faces in $X[L]$ and $X[R]$ respectively.

so we can upper bound (5.44) by

$$3^{D-1}\lambda\left(\left|\mathbb{E}[\pi_{\bar{f}_L}]\right|\mathcal{E}(\bar{f}_R,T_R)+\left|E_{\bar{f}_R}\right|\mathcal{E}(\bar{f}_L,T_L)\right).$$

Finally, observe that

$$\left|E_{\bar{f}_R}\right|=\prod_{i\in R}|\mathbb{E}[f_i]|\leq\prod_{i\in R}\|f_i\|_1\leq\prod_{i\in R}\|f_i\|_{2^{d_i}}$$

by monotonicity of expectation norms and similarly

$$\left|\mathbb{E}[\pi_{\bar{f}_L}]\right|\leq\|\pi_{\bar{f}_L}\|_2\leq\prod_{i\in L}\|f_i\|_{2^{d_i}}$$

by (5.43). Altogether, this gives an upper bound of

$$\mathrm{I}+\mathrm{II}\leq\lambda\mathcal{E}(\bar{f}_{[k]},T)+2\cdot3^{D-1}\lambda\mathcal{E}(\bar{f}_{[k]},T)\leq\lambda3^D\mathcal{E}(\bar{f}_{[k]},T)$$

as desired.

For the unordered case, take an arbitrary partitification (X',T,ρ') as in Definition 5.10.8. (X',T,ρ') is a λ-tuple splitting tree by Claim 5.10.9. Define $\tilde{f}_i:X[\rho'(u_i)]\to\mathbb{R}$ by $\tilde{f}_i((v,j))=f_i(v)$. Viewed as random variables, $\tilde{f}_i$ and f_i are equidistributed, so $\mathbb{E}[f_i]=\mathbb{E}[\tilde{f}_i]$ and $\|f_i\|_{2^{d_i}}=\|\tilde{f}_i\|_{2^{d_i}}$. Moreover, by construction

$$\mathop{\mathbb{E}}_{a\in X(k-1),\pi\in S_k}\left[\prod_{i=1}^k f_i(a_{\pi(i)})\right]=\mathop{\mathbb{E}}_{(a,\pi)\in X'(k-1)}\left[\prod_{i=1}^k\tilde{f}_i(a_{\pi(i)},\pi(i))\right],$$

so can write

$$\left|\mathop{\mathbb{E}}_{a\in X(k),\pi\in S_k}\left[\prod_{i=1}^k f_i(a_{\pi(i)})\right]-\prod_{i=1}^k\mathop{\mathbb{E}}_{a_i\in X(1)}[f_i(a_i)]\right|$$

$$
\begin{aligned}
&= \left| \mathop{\mathbb{E}}_{(a,\pi)\in X'(k)} \left[\prod_{i=1}^{k} \tilde{f}_i((a,\pi)_i) \right] - \prod_{i=1}^{k} \mathop{\mathbb{E}}_{(a,\pi)_i \in X'(i)} \left[\tilde{f}_i(a_i, i) \right] \right| \\
&\leq 3^D \lambda \prod_{i=1}^{k} \|\tilde{f}_i\|_{2^{d_i}} \\
&= 3^D \lambda \prod_{i=1}^{k} \|f_i\|_{2^{d_i}}
\end{aligned}
$$

as desired. □

We now record a few useful corollaries from which hitting set and bias amplification are essentially immediate.

Corollary 5.15.2. *Let X be as in Theorem 5.10.14 and $A_i \subseteq X[i]$. Then*

$$
\left| \mathop{\mathbb{P}}_{a\in X(k-1)} \left[\bigwedge_{i=1}^{k} a_i \in A_i \right] - \prod_{i=1}^{k} \mathbb{P}\left[A_i\right] \right| \leq 3^D \lambda \prod_{i=1}^{k} \mathbb{P}\left[A_i\right]^{2^{-d_i}} \leq 3^D \lambda. \tag{5.45}
$$

Corollary 5.15.3. *Let (X, T, ρ) be a λ-tuple splitting tree. Let $u_1, u_2, \ldots, u_\ell \in T$ be nodes such that for every $i \neq j$, u_i is not an ancestor of u_j, and let d_i be the depth of u_i in T. Let $f_1, f_2, \ldots, f_\ell$ be functions such that $f_i : X(\rho(u_i)) \to \mathbb{R}$. Then*

$$
\left| \mathop{\mathbb{E}}_{a\in X(k-1)} \left[\prod_{i=1}^{\ell} f_i(a_{\rho(u_i)}) \right] - \prod_{i=1}^{\ell} \mathbb{E}[f_i(a_{\rho(u_i)})] \right| \leq 3^D \lambda \prod_{i=1}^{\ell} \|f_i\|_{2^{d_i}}, \tag{5.46}
$$

where $D = \max\{d_i\}$.

Proof. We sequentially prune the tuple-splitting tree using nodes $u_1, u_2, \ldots, u_\ell$ as in Definition 5.10.10. If there are nodes of depth greater than D we also prune their depth D ancestor. By Claim 5.10.11, this results in a pruned λ-tuple splitting tree (X', T', ρ') where by construction $u_1, u_2, \ldots, u_\ell$ are leaves in T' corresponding to $X'[i] = X[\rho(u_i)]$ (we note X' may have other parts not associated with these leaves as well). Setting $f_j = 1$ for the latter, applying Theorem 5.10.14 to (X', T', ρ') gives the desired bound.

□

Proposition 5.15.4 (Hitting Set (Proposition 5.10.12 Restated))**.** *Any depth D λ-(tuple) splitting tree with k leaves is $(3^D\lambda, k)$-hitting.*

Proof. Any subset $A \subset X(1)$ can be divided into partite components $A = A_1 \amalg \ldots \amalg A_k$ where $A_i \subset X[i]$. Corollary 5.15.2 then implies

$$\mathop{\mathbb{P}}_{\sigma \in X(k)}[\sigma \subset A] \leq \prod_{i=1}^{k} \mu(A_i) + 3^D\lambda \leq \mu(A)^k + 3^D\lambda$$

where the righthand inequality follows from the fact that $\frac{1}{k}\sum_{i=1}^{k} \mu(A_i) = \mu(A)$. □

Proposition 5.15.5 (Bias-Amplification (Proposition 5.10.13 Restated))**.** *Let (X, T, ρ) be a λ-tuple splitting tree for $\lambda < \frac{1}{16}$. For any $0 < \varepsilon < \frac{1}{4}$ and family of mean ε functions $\{f_i : X[i] \to \{\pm 1\}\}_{i \in [k]}$, $\prod f_i$ is a $\{\pm 1\}$-valued function with bias at most:*

$$\left| \mathop{\mathbb{E}}_{a \in X(k-1)} \left[\prod_{i=1}^{k} f_i(a_i) \right] \right| \leq \varepsilon^k + 2\lambda \tag{5.47}$$

Proof. The proof is similar to Theorem 5.10.14, and we adopt the same notational conventions. The base case $k = 1$ is trivial. For the inductive step, we have

$$\begin{aligned} |\mathbb{E}[\pi_{\bar{f}_{[k]}}]| &\leq \left| \mathbb{E}[\pi_{\bar{f}_{[k]}}] - \mathbb{E}[\pi_{\bar{f}_L}] \cdot \mathbb{E}[\pi_{\bar{f}_R}] \right| + \mathbb{E}[\pi_{\bar{f}_L}] \cdot \mathbb{E}[\pi_{\bar{f}_R}] \\ &\leq \lambda + (\varepsilon^{|L|} + 2\lambda)(\varepsilon^{|R|} + 2\lambda) \\ &\leq \varepsilon^k + 2\lambda \end{aligned}$$

by our assumptions on λ and ε. □

Finally we list for completeness the unordered analogs of these results, which follow immediately from partitification as in Theorem 5.10.14.

Corollary 5.15.6 (Expander-Mixing (Corollary 5.15.2))**.** *Let X be a depth D λ-splitting*

tree. Then for any $A_1, \ldots, A_k \subset X(1)$:

$$\left| \mathop{\mathbb{P}}_{a \in X(k), \pi \in S_k} \left[\bigwedge_{i=1}^{k} a_i \in A_i \right] - \prod_{i=1}^{k} \mathbb{P}\left[A_i\right] \right| \leq 3^D \lambda \prod_{i=1}^{k} \mathbb{P}\left[A_i\right]^{2^{-d_i}} \leq 3^D \lambda. \quad \square \tag{5.48}$$

Corollary 5.15.7 (Extended HD-EML (Corollary 5.15.3))**.** *Let (X, T, ρ) be a λ-splitting tree, $u_1, u_2, \ldots, u_\ell \in T$ nodes such that for every $i \neq j$, u_i is not an ancestor of u_j, and let d_i be the depth of u_i in T. Let $f_1, f_2, \ldots, f_\ell$ be functions such that $f_i : X(\rho(u_i)) \to \mathbb{R}$. Then*

$$\left| \mathop{\mathbb{E}}_{a \in X(k-1), \pi \in S_k} \left[\prod_{i=1}^{\ell} f_i(a_{\rho(u_i)}) \right] - \prod_{i=1}^{\ell} \mathbb{E}[f_i(a_{\rho(u_i)})] \right| \leq 3^D \lambda \prod_{i=1}^{\ell} \|f_i\|_{2^{d_i}}, \tag{5.49}$$

where $D = \max\{d_i\}$.

Corollary 5.15.8 (Hitting Set (Proposition 5.10.12))**.** *Any depth D λ splitting tree with k leaves is $(3^D\lambda, k)$-hitting.*

Corollary 5.15.9 (Bias-Amplification (Proposition 5.10.13))**.** *Let (X, T, ρ) be a λ splitting tree for $\lambda < \frac{1}{16}$. For any $0 < \varepsilon < \frac{1}{4}$ and family of mean ε functions $\{f_i : X(1) \to \{\pm 1\}\}_{i \in [k]}$, $\prod f_i$ is a $\{\pm 1\}$-valued function with bias at most:*

$$\left| \mathop{\mathbb{E}}_{a \in X(k-1), \pi \in S_k} \left[\prod_{i=1}^{k} f_i(a_{\pi(i)}) \right] \right| \leq \varepsilon^k + 2\lambda \tag{5.50}$$

5.16 Reverse Hypercontractivity: From Boolean to All Functions

We now give a generic reduction from reverse hypercontractivity for general functions to the Boolean case.

Theorem (Restatement of Theorem 5.6.12)**.** *For every $\ell \geq 1$ and $\varepsilon > 0$ there exists $\kappa' \geq \left(\frac{1}{5}\right)^{2\ell(1+\varepsilon)} \left(\frac{1}{18+\frac{12}{\varepsilon}}\right)^{2(\ell-1)}$ such that the following holds. Let V be a finite probability*

space and D a monotone linear operator such that for every $A, B \subseteq V$,

$$\langle 1_A, D1_B \rangle \geq \kappa \, \mathbb{P}[A]^{\ell} \, \mathbb{P}[B]^{\ell}. \tag{5.51}$$

Then for every two arbitrary *functions $f_1, f_2 : V \to \mathbb{R}_{\geq 0}$ it holds that*

$$\langle f_1, Df_2 \rangle \geq \kappa\kappa' \|f_1\|_{\frac{1}{\ell(1+\varepsilon)}} \|f_2\|_{\frac{1}{\ell(1+\varepsilon)}}. \tag{5.52}$$

That is, D is $(\frac{1}{\ell(1+\varepsilon)}, \frac{1}{1-\ell(1+\varepsilon)}, \kappa\kappa')$-reverse hypercontractive.

Before moving directly to the proof, we give a brief overview of the main idea. Starting with arbitrary f_1, f_2, our goal is to build discrete approximations of the f_i that remain close to the original functions in expectation and don't take too many unique values. Once this is the case, we can write the functions as a weighted sum over indicators for each value, and apply the boolean result without too much loss.

The key to the proof really lies in these approximations, which we build by discretizing the functions to powers of 2 and applying a careful iterative zero-ing procedure to components that fall too far from the mean. Naively this approach seems problematic, as f_i is arbitrary and may not have a well-behaved tail to truncate. We handle this by exploiting the fact that no function can have more than a small constant fraction of its mass beyond its expectation times (an appropriate power of) its $(1 \to (1+\varepsilon))$-norm.

More formally, the proof of Theorem 5.6.12 is split into two intermediate reductions, corresponding to the 'approximation step' and the 'indicator step' above. The first of these is a reduction from general reverse hypercontractivity to a slightly richer set of functions that are *balanced* and *discrete*.

Definition 5.16.1 (Nice Functions)**.** For any $\alpha \in [0, 1]$, and $\beta \in [1, \infty)$, we say the function $f : V \to \mathbb{R}_{\geq 0}$ is (α, β)-nice if it is:

1. **Balanced:** all non-zero outputs of f satisfy

$$\alpha\mathbb{E}[f] \leq f(x) \leq \beta\mathbb{E}[f].$$

2. **Discrete:** the range of f is entirely contained in $\{0\} \cup \{2^j : j \in \mathbb{Z}\}$.

To state the reduction, we require a setting of parameters α and β based on the $(1\to(1+\varepsilon))$-norm of the functions in question due to the tail truncation strategy described above. In particular, for a fixed non-negative function f_i, define

$$\eta_i = \frac{\mathbb{E}[f_i^{1+\varepsilon}]}{\mathbb{E}[f_i]^{1+\varepsilon}} = \|f_i\|_{1\to(1+\varepsilon)}^{1+\varepsilon}.$$

With this in mind, we state our two intermediate reductions and prove Theorem 5.6.12.

Proposition 5.16.2 (From General to Nice)**.** *Let $\ell, \varepsilon, \kappa, V$, and D be as in Theorem 5.6.12. Suppose that for any two functions $f_1, f_2 : V \to \mathbb{R}_{\geq 0}$ that are $(\frac{1}{16\eta_i}, (25\eta_i)^{\frac{2(1+\varepsilon)}{\varepsilon}})$-nice*

$$\langle f_1^{\ell(1+\varepsilon)}, D f_2^{\ell(1+\varepsilon)} \rangle \geq \kappa\, \mathbb{E}[f_1]^{\ell(1+\varepsilon)}\, \mathbb{E}[f_2]^{\ell(1+\varepsilon)}. \tag{5.53}$$

Then for every two functions $f_1, f_2 : V \to \mathbb{R}_{\geq 0}$ it holds that

$$\langle f_1^{\ell(1+\varepsilon)}, D f_2^{\ell(1+\varepsilon)} \rangle \geq \left(\frac{1}{5}\right)^{2\ell(1+\varepsilon)} \kappa\, \mathbb{E}[f_1]^{\ell(1+\varepsilon)}\, \mathbb{E}[f_2]^{\ell(1+\varepsilon)}. \tag{5.54}$$

The 'indicator step' then reduces reverse hypercontractivity for nice functions to the Boolean case.

Proposition 5.16.3 (From Nice to Boolean)**.** *Let ℓ, ε, V, and D be as in Theorem 5.6.12. If for every $A, B \subseteq V$,*

$$\langle 1_A, D 1_B \rangle \geq \kappa\, \mathbb{P}[A]^{\ell}\, \mathbb{P}[B]^{\ell}, \tag{5.55}$$

then every two $(\frac{1}{16\eta_i}, (25\eta_i)^{\frac{2(1+\varepsilon)}{\varepsilon}})$*-nice functions* $f_1, f_2 : V \to \mathbb{R}_{\geq 0}$ *satisfy*

$$\langle f_1^{\ell(1+\varepsilon)}, D f_2^{\ell(1+\varepsilon)} \rangle \geq \left(\frac{1}{18 + \frac{12}{\varepsilon}}\right)^{2(\ell-1)} \kappa \, \mathbb{E}[f_1]^{\ell(1+\varepsilon)} \, \mathbb{E}[f_2]^{\ell(1+\varepsilon)}. \tag{5.56}$$

The proof of Theorem 5.6.12 is now essentially immediate.

Proof of Theorem 5.6.12. We have by assumption that all $A, B \subseteq V$ satisfy $\langle 1_A, D1_B \rangle \geq \kappa \, \mathbb{P}[A]^{\ell} \, \mathbb{P}[B]^{\ell}$, so we may apply Proposition 5.16.3 to get that every two $(\frac{1}{16\eta_i}, (25\eta_i)^{\frac{2(1+\varepsilon)}{\varepsilon}})$-nice functions $f_1, f_2 : V \to \mathbb{R}_{\geq 0}$ satisfy

$$\langle f_1^{\ell(1+\varepsilon)}, D f_2^{\ell(1+\varepsilon)} \rangle \geq \left(\frac{1}{18 + \frac{12}{\varepsilon}}\right)^{2(\ell-1)} \kappa \, \mathbb{E}[f_1]^{\ell(1+\varepsilon)} \, \mathbb{E}[f_2]^{\ell(1+\varepsilon)}.$$

Applying Proposition 5.16.2 then implies *arbitrary* non-negative g_1, g_2 satisfy

$$\langle g_1^{\ell(1+\varepsilon)}, D g_2^{\ell(1+\varepsilon)} \rangle \geq \left(\frac{1}{5}\right)^{2\ell(1+\varepsilon)} \left(\frac{1}{18 + \frac{12}{\varepsilon}}\right)^{2(\ell-1)} \kappa \, \mathbb{E}[g_1]^{\ell(1+\varepsilon)} \, \mathbb{E}[g_2]^{\ell(1+\varepsilon)}.$$

Finally, for f_1, f_2 arbitrary non-negative functions applying the above to $g_i = f_i^{\frac{1}{\ell(1+\varepsilon)}}$ gives the result. □

We now move to the formal proofs.

Proof of Proposition 5.16.2. Given arbitrary $f_1, f_2 : V \to \mathbb{R}_{\geq 0}$ we will construct corresponding f_1', f_2' satisfying

1. $0 \leq f_i' \leq f_i$
2. $\mathbb{E}[f_i'] \geq \frac{1}{5}\mathbb{E}[f]$
3. f_i' is $(\frac{1}{16\eta_i}, (25\eta_i)^{\frac{2(1+\varepsilon)}{\varepsilon}})$-nice

Combining the above with monotonicity of D and reverse hypercontractive inequality for nice functions gives

$$\begin{aligned}
\langle f_1, Df_2\rangle &\geq \langle f_1', Df_2'\rangle && \text{(Item (1) and Monotonicity)}\\
&\geq \kappa\mathbb{E}[f_1']^{\ell(1+\varepsilon)}\mathbb{E}[f_2']^{\ell(1+\varepsilon)} && \text{(Item (3) and Equation (5.53))}\\
&\geq \left(\frac{1}{5}\right)^{2\ell(1+\varepsilon)} \kappa\mathbb{E}[f_1']^{\ell(1+\varepsilon)}\mathbb{E}[f_2']^{\ell(1+\varepsilon)} && \text{(Item (2))}
\end{aligned}$$

as desired.

We construct f_i' in two separate steps: a simple discretization procedure, and a more involved balancing procedure that iteratively zeroes out unbalanced parts of the function. Starting with the former, given a non-negative f define its *rounding down* $\tilde{f}$ by setting for $j \in \mathbb{Z}$:

$$\tilde{f}(x) = \begin{cases} 2^j & \text{for } f(x) \in [2^j, 2^{j+1}) \\ 0 & \text{if } f(x) = 0. \end{cases}$$

Note that, by construction, $\frac{f_i}{2} \leq \tilde{f}_i \leq f_i$.

We now define an iterative balancing procedure outputting a sequence of functions $\tilde{f}_i = f_i^{(0)} \geq f_i^{(1)} \geq \ldots \geq f_i^{(k_i)} = f_i'$ for $k_i \in \mathbb{N}$ some finite stopping time. For these (soon-to-be-defined) functions, define $\eta_i^{(j)} = \frac{\mathbb{E}[(f_i^{(j)})^{1+\varepsilon}]}{\mathbb{E}[(f_i^{(j)})]^{1+\varepsilon}}$ and denote the set of 'balanced' inputs of $f_i^{(j)}$ as:

$$B_i^{(j)} = \left\{ x \;\middle|\; \frac{1}{4\eta_i^{(j)}}\,\mathbb{E}[f_i^{(j)}] \leq f_i^{(j)}(x) \leq \left(4\eta_i^{(j)}\right)^{\frac{2(1+\varepsilon)}{\varepsilon}}\,\mathbb{E}[f_i^{(j)}] \right\}.$$

If $f_i^{(0)}$ is already $(\frac{1}{16\eta_i}, (25\eta_i)^{\frac{2(1+\varepsilon)}{\varepsilon}})$-nice, set $k_i = 0$ and output $f_i' = \tilde{f}_i$. Otherwise for $j = 1, 2, \ldots$ define $f_i^{(j)}$ by zeroing out all non-balanced inputs:

$$f_i^{(j)} := f_i^{(j-1)} \cdot \mathbf{1}_{B_i^{(j-1)}},$$

and define the stopping time k_i to be the first index j such that $\eta_i^{(j)} \geq \frac{1}{4}\eta_i^{(j-1)}$. This procedure terminates (i.e. k_i is finite) since each $\eta_i^{(j)} \geq 1$ by convexity. Every step the process doesn't stop decreases $\eta_i^{(j)}$ by a factor of 4, so there can be a total of $O(\log(\eta_i^{(0)}))$ steps.

It is immediate from definition that $0 \leq f_i' \leq f_i$, so we just need to show Item 2 and Item 3. The key is to observe that our balancing process cannot remove too much mass in each step:

$$\mathbb{E}[f_i^{(j-1)}] \geq \mathbb{E}[f_i^{(j)}] \geq \left(1 - \frac{1}{2\eta_i^{(j-1)}}\right) \mathbb{E}[f_i^{(j-1)}]. \tag{5.57}$$

We defer the proof, and first show Item 2 and Item 3 given this assumption.

Proof of Item 2.

The proof is essentially immediate by the exponential decay of $\eta_i^{(j)}$. Namely by iterated application of (5.57) we have

$$\frac{\mathbb{E}[f_i^{(k)}]}{\mathbb{E}[f_i^{(0)}]} \geq \prod_{j=0}^{k-1} \left(1 - \frac{1}{2\eta_i^{(k-1-j)}}\right) \geq \prod_{j=0}^{\infty} \left(1 - \frac{1}{2 \cdot 4^j}\right) \geq \frac{2}{5}$$

since by construction $\eta_i^{(j)} > 4^{k-1-j}\eta_i^{(k-1)} \geq 4^{k-1-j}$. Combining with the fact $\tilde{f}_i = f_i^{(0)} \geq \frac{f_i}{2}$ gives the result.

Proof of Item 3.

Recall that for any j we have by construction the following 'pseudo'-balance condition:

$$\forall f_i^{(j)}(x) \neq 0 : \frac{1}{4\eta_i^{(j-1)}} \mathbb{E}[f_i^{(j-1)}] \leq f_i^{(j)}(x) \leq \left(4\eta_i^{(j-1)}\right)^{\frac{2(1+\varepsilon)}{\varepsilon}} \mathbb{E}[f_i^{(j-1)}].$$

When $j = k_i$ is the stopping time, observe that by Equation (5.57) and construction respectively we have

1. $\mathbb{E}[f_i^{(k_i)}] \leq \mathbb{E}[f_i^{(k_i-1)}] \leq 2\mathbb{E}[f_i^{(k_i)}]$

2. $4\eta_i^{(k_i)} \geq \eta_i^{(k_i-1)}$.

Substituting into the above gives

$$\frac{1}{16\eta_i^{(k_i)}}\,\mathbb{E}[f_i^{(k_i)}] \leq f_i^{(k_i)}(x) \leq \left(25\eta_i^{(k_i)}\right)^{\frac{2(1+\varepsilon)}{\varepsilon}}\,\mathbb{E}[f_i^{(k_i)}]$$

as desired.

Proof of (5.57)**.**

The lefthand inequality is by construction. Toward the righthand, fix $f = f_i^{(j-1)}$ and $\eta = \eta_i^{(j-1)}$ for notational simplicity and define the set of 'terrible' inputs zeroed in the jth step as:

$$T_1 \coloneqq \left\{x \;\middle|\; f(x) < \frac{1}{4\eta}\,\mathbb{E}[f]\right\} \qquad\qquad T_2 \coloneqq \left\{x \;\middle|\; f(x) > (4\eta)^{\frac{2(1+\varepsilon)}{\varepsilon}}\,\mathbb{E}[f]\right\}$$

Since $T_1 \cup T_2 = V \setminus B_i^{(j-1)}$ and $f_i^{(j)} = \mathbf{1}_{B_i^{(j-1)}} f$, it is enough to show that

$$\mathbb{E}[f \cdot \mathbf{1}_{T_1}], \mathbb{E}[f \cdot \mathbf{1}_{T_2}] \leq \frac{1}{4\eta}\,\mathbb{E}[f]. \tag{5.58}$$

It is obvious that $\mathbb{E}[f \cdot \mathbf{1}_{T_1}] \leq \frac{1}{4\eta}\,\mathbb{E}[f]$ since f is upper bounded by this value within T_1. To bound the upper tail we use Hölder and Markov:

$$\begin{aligned}
\mathbb{E}[f \cdot \mathbf{1}_{T_2}] &\leq \|f\|_{1+\varepsilon}\,\mathbb{P}\left[f \geq (4\eta)^{\frac{2(1+\varepsilon)}{\varepsilon}}\,\mathbb{E}[f]\right]^{\frac{\varepsilon}{1+\varepsilon}} && \text{(Hölder's inequality)} \\
&\leq \|f\|_{1+\varepsilon}\frac{1}{16\eta^2} && \text{(Markov's inequality)} \\
&\leq \left(\frac{1}{4}\eta^{-\frac{\varepsilon}{1+\varepsilon}}\right)\frac{1}{4\eta}\,\mathbb{E}[f] && (\|f\|_{1+\varepsilon} = \eta^{\frac{1}{1+\varepsilon}}\,\mathbb{E}[f]) \\
&\leq \frac{1}{4\eta}\,\mathbb{E}[f] && (\eta^{-\frac{\varepsilon}{1+\varepsilon}} \leq 1)
\end{aligned}$$

as desired. □

Finally we prove reverse hypercontractivity for 'nice' functions (Proposition 5.16.3).

Proof of Proposition 5.16.3. Since f_i is $(\frac{1}{16\eta_i}, (25\eta_i)^{\frac{2(1+\varepsilon)}{\varepsilon}})$-nice, it can attain at most

$$n_i = \left(3 + \frac{2}{\varepsilon}\right) \log \eta_i + \frac{10}{\varepsilon} + 15$$

non-zero values. With this in mind, decompose $f_i = \sum_{j \in \mathbb{Z}} 2^j \mathbf{1}_{f_i(x)=2^j}$ into level sets and observe that, by linearity of D, we can decompose the inner product itself into boolean sub-components and apply our assumed reverse hypercontractive inequality (Equation Equation (5.56)):

$$\begin{aligned}
\langle f_1^{(1+\varepsilon)\ell}, D f_2^{(1+\varepsilon)\ell} \rangle &= \sum_{i,j \in \mathbb{Z}} 2^{(1+\varepsilon)\ell j} 2^{(1+\varepsilon)\ell i} \langle \mathbf{1}_{f_1(x)=2^j}, D\mathbf{1}_{f_2(x)=2^i} \rangle \\
&\geq \kappa \sum_{i,j \in \mathbb{Z}} 2^{(1+\varepsilon)\ell j} 2^{(1+\varepsilon)\ell i} \mathbb{P}\left[f_1(x) = 2^j\right]^\ell \mathbb{P}\left[f_2(x) = 2^j\right]^\ell \\
&= \kappa \left(\sum_{j \in \mathbb{Z}} (2^{(1+\varepsilon)j} \mathbb{P}\left[f_1(x) = 2^j\right])^\ell \right) \cdot \left(\sum_{j \in \mathbb{Z}} (2^{(1+\varepsilon)j} \mathbb{P}\left[f_2(x) = 2^j\right])^\ell \right).
\end{aligned}$$

Now since the sums have only n_1, n_2 non-zero terms respectively, applying $(\ell, \frac{\ell}{\ell-1})$-Hölder's inequality lower bounds this quantity by

$$\begin{aligned}
&\kappa (n_1 n_2)^{1-\ell} \left(\sum_{j \in \mathbb{Z}} 2^{(1+\varepsilon)j} \mathbb{P}\left[f_1(x) = 2^j\right] \right)^\ell \cdot \left(\sum_{j \in \mathbb{Z}} 2^{(1+\varepsilon)j} \mathbb{P}\left[f_1(x) = 2^j\right] \right)^\ell \\
=&\kappa n_1^{1-\ell} \mathbb{E}[f_1^{1+\varepsilon}]^\ell \cdot n_2^{1-\ell} \mathbb{E}[f_2^{1+\varepsilon}]^\ell \\
\geq&\kappa \cdot (n_1^{1-\ell} \eta_1^\ell) \cdot (n_2^{1-\ell} \eta_2^\ell) \cdot \mathbb{E}[f_1]^{\ell(1+\varepsilon)} \mathbb{E}[f_2]^{\ell(1+\varepsilon)}.
\end{aligned}$$

It can be checked directly that

$$\min_{\eta_i \geq 1} n_i \eta_i \geq \frac{1}{18 + \frac{12}{\varepsilon}}$$

which gives the desired result. □

5.17 Splittability

In this section we discuss the relations between our notion of complete splittability, the splittability notion of [221], and the "τ-sampling" gaurantee of [132], and the inheritance of splittability under projection. We start by defining the latter two notions.

Definition 5.17.1 (JST-Splittability [221])**.** $X \subset [n]^d$ is said to be λ-JST-splittable if for all $1 \leq a \leq t \leq b \leq d$:

$$\lambda_2(S_{[a,t],[t+1,b]}) \leq \lambda$$

In their later work, [132] introduced the weaker notion of τ-sampling which suffices for list-decoding.

Definition 5.17.2. $X \subset [n]^d$ is said to be τ-sampling if for $t \in [d]$, $S \subset [n]$ and $W \subset [n]^t$:

$$\operatorname*{Cov}_{w \sim X}\left(1[w_{t+1} \in S], 1[w_{[t]} \in W]\right) \leq \tau.$$

Before arguing relations between these notions, we give the key lemma that underlies all proofs in this section: the expansion of swap walks is inherited under projection.

Lemma 5.17.3 (Swap Walk Inheritence)**.** *Let X be a partite d-uniform complex, and A and B disjoint subsets of $[d]$, and F any subset of $[d]$. Then:*

$$\lambda_2(S_{A\cap F, B\cap F}) \leq \lambda_2(S_{A,B})$$

Proof. By the variational characterization it is enough to show

$$\max_{\mathbb{E}[f^F]=0} \frac{\|S_{A\cap F, B\cap F} f^F\|}{\|f^F\|_2} \leq \lambda_2(S_{A,B})$$

Given any such function $f^F : X[A \cap F] \to \mathbb{R}$ define its extension to $X[A]$ by $f(x) \coloneqq f^F(x_F)$. Here $x_F = \{v \in x : col(v) \in F\}$ is the projection into X^F. and observe that this extension satisfies:

1. $\mathbb{E}_{X[A]}[f] = \mathbb{E}_{X[A\cap F]}[f^F] = 0$
2. $\|f\|_{2,X[A]} = \|f^F\|_{2,X[A\cap F]}$.

The result then follows from convexity, in particular:

$$\begin{aligned}
\|S_{A\cap F,B\cap F} f^F\|_2 &= \mathop{\mathbb{E}}_{y\sim X[A\cap F]}\left[\left(\mathop{\mathbb{E}}_{x\in X(d)}[f^F(x_{B\cap F})|x_{A\cap F} = y]\right)^2\right]^{\frac{1}{2}} \\
&= \mathop{\mathbb{E}}_{y\sim X[A\cap F]}\left[\left(\mathop{\mathbb{E}}_{x\in X(d)}[f(x_B)|x_{A\cap F} = y]\right)^2\right]^{\frac{1}{2}} \\
&= \mathop{\mathbb{E}}_{y\sim X[A\cap F]}\left[\left(\mathop{\mathbb{E}}_{y'\sim X_y[A\setminus F]}\left[\mathop{\mathbb{E}}_{x\in X(d)}[f(x_B|x_{A\cap F} = y \wedge x_{A\setminus F} = y']\right]\right)^2\right]^{\frac{1}{2}} \\
&\leq \mathop{\mathbb{E}}_{y\sim X[A]}\left[\left(\mathop{\mathbb{E}}_{x\in X(d)}[f(x_B|x_A = y]\right)^2\right]^{\frac{1}{2}} \\
&= \|S_{A,B} f\|_2 \\
&\leq \lambda_2(S_{A,B})\|f\|_2 \\
&= \gamma\|f^F\|_2
\end{aligned}$$

□

With this in mind, we prove the following relations between complete splittability, JST-splittability, and τ-sampling.

Proposition 5.17.4. *$X \subset [n]^t$ is λ-JST-splittable if and only if it is completely λ-splittable. Moreover any completely λ-splittable complex is λ-sampling.*

Proof. If X is λ-JST-splittable it is clearly completely λ-splittable setting $a = 1, b = d$. If

X is completely λ-splittable, then for any $1 \leq a \leq i \leq b \leq d$ we have:

$$S_{[a,i],[i+1,b]} \leq S_{[1,i],[i+1,d]} \leq \lambda$$

as desired. Finally, if X is completely λ-splittable, then for any $S \subset [n]$ and $X \subset [n]^t$ write $1_S \coloneqq \mathbf{1}[w_{t+1} \in S]$ and $1_X \coloneqq \mathbf{1}[w_{[t]} \in X]$

$$\begin{aligned}\mathrm{Cov}_{w\sim X}(1_S, 1_X) &= \mathop{\mathbb{E}}_{y\sim X^{t+1}}[1_S(y) S_{t+1,[t]} 1_X(y)] - \mathbb{E}[1_S][E][1_X] \\ &\leq \lambda_2(S_{t+1,[t]})\|1_S\|\|1_X\| \\ &\leq \lambda\end{aligned}$$

where we have used that by Hölder duality and Lemma 5.17.3 $\lambda_2(S_{t+1,[t]}) = \lambda_2(S_{[t],t+1}) \leq \lambda_2(S_{[t],t+1}) \leq \lambda$. □

Finally, we show that both complete splittability and standard tuple splittability are inherited under projections. Recall that given a d-partite complex X and $F \subseteq [d]$, X^F is the projection of X given by drawing a d-face from X, and projecting the resulting face onto the parts in F.

Proposition 5.17.5 (Claim 5.10.34 Restated)**.** *Let $X \subset [n]^d$ be a homogeneous, completely λ-splittable complex. Then for any $F \subseteq [d]$ X^F is homogeneous and completely λ-splittable.*

Proof. The proof is essentially immediate from definition and Lemma 5.17.3. Recall X is homogenous if the projection onto each part is uniform on $[n]$. Since projecting onto X^F and then a coordinate $i \in F$ is the same as just projecting onto i, homogeneity is inherited. Toward splittability, write $F = \{i_1, \ldots, i_{|F|}\}$ where $i_j < i_{j+1}$. Then for any i_j, applying Lemma 5.17.3 gives

$$\lambda_2(S^{X^F}_{\{i_1,\ldots,i_j\},\{i_{j+1},\ldots,i_{|F|}\}}) \leq \lambda_2(S^{X}_{\{i_1,\ldots,i_j\},\{i_{j+1},\ldots,i_{|F|}\}})$$

$$\leq \lambda_2(S_{[i_j],[i_j+1,d]})$$
$$\leq \lambda$$

where we have used the fact that $\lambda_2(S^{X^F}_{\{i_1,\ldots,i_j\},\{i_{j+1},\ldots,i_{|F|}\}}) = \lambda_2(S^{X}_{\{i_1,\ldots,i_j\},\{i_{j+1},\ldots,i_{|F|}\}})$ (indeed they are the same operator). □

Finally, while not strictly necessary for the results in this work, we show for completeness that inheritance of splittability also holds in the setting of general tuple-splitting trees as well.

Proposition 5.17.6 (Projected Splitting Trees)**.** *Let X be any d-uniform partite complex with a λ-tuple splitting tree of depth D. For every $F \subseteq [d]$, X^F also has a λ-tuple splitting tree of depth at most D.*

The proof of Proposition 5.17.6 relies on a simple inductive algorithm for projecting partite orderered binary trees (T, ρ) onto a subset of their coordinates. In particular, given such a tree and a subset $F \subseteq [d]$, we define the projection operation via the following inductive algorithm that modifies the tree T in place starting from a specific node $v \in T$.

In other words, Algorithm 1 is the result of intersecting ρ with F and deleting the resulting nodes with empty or repeated labelings. We define the F-projection of a tree by applying this process at the root.

Definition 5.17.7 (Projected Trees)**.** Let (T, ρ) be a d-uniform partite ordered binary tree and $F \subset [d+1]$ any coordinate subset. The F-projection of (T, ρ), denoted (T^F, ρ^F) is given by

$$T^F = \text{Project}(T, \rho, F, \text{root}(T)), \qquad \rho^F = (\rho \cap F)|_{T^F}.$$

We are now ready to prove Proposition 5.17.6.

Proof of Proposition 5.17.6. We will show that (X^F, T^F, ρ^F) is a λ-tuple splitting tree of depth at most D. We first argue that (X^F, T^F, ρ^F) is a partite ordered binary tree of

Algorithm 1. Project(T, ρ, F, v)

Output: Projected partite ordered binary tree (T_v^F, ρ_v^F) rooted at v
if v is a leaf **then**
    return
if $\rho(\ell_v) \cap F = \emptyset$ **then**
    Delete$(T(\ell_v))$
    Merge(v, r_v) ▷ Merge(v,r_v) deletes the node v and replaces r_v as the respective child of v's parent.
    Project(T, ρ, F, r_v)
else if $\rho(r_v) \cap F = \emptyset$ **then**
    Delete$(T(r_v))$
    Merge(v, ℓ_v)
    Project(T, ρ, F, ℓ_v)
else
    Project(T, ρ, F, r_v)
    Project(T, ρ, F, ℓ_v)

depth at most D. First, observe that every operation in Algorithm 1 maps binary trees to binary trees, and cannot increase depth. In particular, the only modification of the tree structure occurs when a branch is deleted and the relevant root is contracted with its other child, maintaining the invariant that every internal node has one parent and two children, and that the depth is at most D. To see that the leaves are in bijection with F, observe that all leaves labeled by F survive the projection, all leaves labeled by $[d] \setminus F$ are deleted, and no new leaves are introduced by construction. The first two claims hold since a node v is deleted if and only if either $\rho(v) \cap F = \emptyset$, or the labeling is repeated later down the tree. Finally, to see that the children of any internal node in T^F partition its labeling in ρ^F, observe that this property holds trivially for T under the intersected labeling $\rho \cap F$ (albeit the partition may be trivial), and that this invariant is preserved by every operation in Algorithm 1.

Finally we argue the projected tree inherits the splittability of X^F. By Lemma 5.17.3, we have that for any internal node u:

$$\lambda_2(S_{\rho^F(\ell_u),\rho^F(r_u)}) = \lambda_2(S_{\rho(\ell_u)\cap F,\rho(r_u)\cap F}) \leq \lambda_2(S_{\rho(\ell_u),\rho(r_u)}) \leq \lambda$$

as desired. □

This chapter, in full, has been submitted for publication of the material as it may appear in Foundations of Computer Science 2024. Dikstein, Yotam; Hopkins, Max. "Chernoff-Hoeffding and Reverse Hypercontractivity on High Dimensional Expanders". The dissertation author was a primary investigator and author of this material.

Chapter 6

Explicit Lower Bounds Against $\Omega(n)$-Rounds of Sum-of-Squares

6.1 Introduction

The Sum-of-Squares (SoS) semi-definite programming (SDP) hierarchy is one of the most powerful and widely studied algorithmic frameworks for approximating constraint satisfaction problems (CSPs) in theoretical computer science, yet very little is known about the structure of instances that are *hard* for the paradigm. Indeed, while it has long been known that *random* instances of CSPs are hard for Sum-of-Squares [180, 340, 354, 45, 90, 269], there are essentially no *explicit* constructions of hard instances better than brute force search [119, 355, 322]. Leveraging recent breakthroughs in locally testable [117, 281] and quantum low-density parity-check (qLDPC) codes [315, 277], we resolve this problem, giving the first explicit family of highly unsatisfiable CSPs that cannot be refuted by $\Omega(n)$-rounds of Sum-of-Squares.

Theorem 6.1.1 (Main Result: Explicit 3-XOR Instances Hard for SoS)**.** *There exist constants $\mu_1, \mu_2 \in (0, 1)$ and an infinite family of 3-XOR instances constructable in deterministic polynomial time such that:*

1. *No assignment satisfies more than a $1 - \mu_1$ fraction of constraints*

2. *No instance can be refuted by $\mu_2 n$ levels of the corresponding Sum-of-Squares SDP*

Relaxation.

Though Theorem 6.1.1 only exhibits an 'integrality gap' of 1 v.s $1-\mu_1$ (meaning the instance are $(1-\mu_1)$-satisfiable but *look* fully satisfiable to SoS), combined with standard PCP-like reductions in the SoS hierarchy this gap can be amplified to $1-\varepsilon$ v.s $\frac{1}{2}+\varepsilon$ for any $\varepsilon > 0$ [354, 119], which matches the hardness of random 3-XOR instances up to imperfect completeness [180, 340].[1] In fact, it is worth noting that Theorem 6.1.1 is the first explicit family of CSPs to even beat more than $O(\log(n))$ levels of the SoS hierarchy, which can be done either by unique neighbor expanders [322, 17] or (up to lower order factors) simply by brute force search [355]. While explicit constructions against $\Omega(n)$-rounds of SoS were known in proof complexity (e.g. Tseitin formulas [178], knapsack [179]), these examples do not lead to inapproximability since their satisfiability is not bounded away from 1.

Thus, at a high level, Theorem 6.1.1 provides the first example of an approximation problem with short witnesses of unsatisfiability that cannot be captured by the Sum-of-Squares proof system, settling (in the negative) the completeness of SoS in this setting. Furthermore, it is worth noting that 3-XOR is not somehow 'special' in this sense. As observed in [119] (who showed an analogous result for $O(\sqrt{\log(n)})$-levels of SoS), Theorem 6.1.1 also gives explicit hard instances across many types of CSPs by standard reduction techniques [354], including instances with optimal integrality gaps for CSPs with approximation resistant predicates based on pairwise independent subgroups [90, 119].

6.1.1 High Dimensional Small-Set Expanders

Theorem 6.1.1 is based on a new form of *high dimensional expansion* (HDX), a nascent area of computer science and math that has already seen an impressive array of breakthrough results across areas such as coding theory [221, 117, 315, 281], approximate sampling [239, 25, 11, 24], approximation algorithms [9, 38], analysis of boolean functions

[1] Indeed one can see such a gap is essentially optimal, as a random assignment to any 3-XOR instance will satisfy 1/2 the constraints in expectation.

[111, 39, 187], agreement testing [124, 109], and, recently, Sum-of-Squares lower bounds [119]. While most of these works consider notions of expansion on *hypergraphs* (often called *simplicial complexes* in this setting), we take inspiration from recent breakthroughs on LTCs [117, 281] and quantum codes [315, 277] and consider expansion on the more general class of *chain complexes*:

$$X : \mathbb{F}_2^{X(0)} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{X(1)} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{X(2)}.$$

Here $X(0)$, $X(1)$, and $X(2)$ are sets, δ_0 and δ_1 are linear maps (called the *co-boundary operators*), ∂_2 and ∂_1 are their transposes (called the *boundary operators*), and both satisfy $\partial_1\partial_2 = 0$, $\delta_1\delta_0 = 0$.

Chain complexes admit a natural analog of boundary (edge) expansion in graphs called high-dimensional (co)-boundary expansion [283]. To see this, we first note an important inherent structural property of chain complexes: any function $f \in \mathrm{im}(\delta_0)$ (called a *co-boundary*) satisfies $|\delta_1 f| = 0$. A complex is called a *ρ-co-boundary expander* essentially when this is the only obstruction to $|\delta_1 f|$ being large:

$$\forall f \in \mathbb{F}_2^{X(1)} : |\delta_1 f| \geq \rho \cdot d(f, \mathrm{im}(\delta_0)).$$

For intuition, it is worth briefly discussing why this generalizes boundary expansion on graphs. Any graph $G = (V, E)$ (or indeed hypergraph, see Section 6.4.2) can be written as a chain complex:

$$X : \mathbb{F}_2^{\emptyset} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{V} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{E},$$

where $\delta_0 f(v) = f(\emptyset)$, $\delta_1 f((u, v)) = f(u) \oplus f(v)$, and it is easily checked that $\delta_1\delta_0 = 0$. Notice that in this setting the only co-boundaries are $\mathrm{im}(\delta_0) = \{\emptyset, V\}$, and furthermore that for any $S \subset V$ and $e \in E$, the value of $\delta_1 1_S$ on e is 1 iff e crosses the cut defined by S. This implies the ratio $\frac{|\delta_1 1_S|}{d(1_S, \mathrm{im}(\delta_0))} = \frac{E(S, V\setminus S)}{\min\{|S|, |V\setminus S|\}}$, which is just the standard boundary

expansion of G!

Unfortunately, while standard boundary expansion on (random) graphs has been quite useful for proving SoS lower bounds in the past [60, 180, 340], high dimensional co-boundary expansion seems to be too strong a notion for this setting: good (co)-boundary expanders are not known to exist (even probabilistically), and their structure is prohibitively restrictive in other senses as well.[2] We avoid these issues by introducing a simple relaxation of boundary expansion to *small-sets*:

Definition 6.1.2 (Small-set (Co)-Boundary Expansion)**.** We call X a (ρ_1, ρ_2)-small-set boundary expander if the weight of any 'small' function $f \in \mathbb{F}_2^{X(1)}$ satisfying $|f| \leq \rho_1 |X(1)|$ expands:

$$|\partial_1 f| \geq \rho_2 \cdot d(f, \mathrm{im}(\partial_2)).$$

Similarly, X is a (ρ_1, ρ_2)-small-set co-boundary expander if all $f \in \mathbb{F}_2^{X(1)}$ s.t. $|f| \leq \rho_1 |X(1)|$ satisfy:

$$|\delta_1 f| \geq \rho_2 \cdot d(f, \mathrm{im}(\delta_0)).$$

We call X a (ρ_1, ρ_2)-small-set HDX (SS-HDX) if it satisfies both the above conditions.

Small-set (co)-boundary expansion is a direct generalization of small-set expansion on graphs, a notion that lies at the heart of many problems in hardness of approximation. A close variant of the above definition in the co-boundary direction was first considered on simplicial complexes in [230, 141, 235, 237] to construct co-systolic expanders (a different weakening of co-boundary expansion). In our case, it is critical the notion hold in both directions, which requires moving away from (known) simplicial constructions.

In the next section, we will show how SS-HDX naturally lead to hard instances of XOR for Sum-of-Squares (largely following a similar result of Dinur, Filmus, Harsha, and

[2] We'll discuss this issue in Section 6.2, but in brief co-boundary expansion implies $\ker(\delta_1) = \mathrm{im}(\delta_0)$. Like [119], our instances will rely on a function in $\ker(\delta_1) \setminus \mathrm{im}(\delta_0)$ to enforce global structure on the CSP that cannot be detected through local algorithms like Sum-of-Squares.

Tulsiani [119] for the LSV complex [290]), giving the first connection between hardness of approximation and *high dimensional* small-set expanders. Thus, Theorem 6.1.1 boils down to constructing an infinite family of SS-HDX on a growing number of vertices, each of which can be constructed in deterministic polynomial time. While this may seem hopelessly strong, a weaker variant of these requirements was very recently achieved in breakthrough constructions of qLPDC codes by [315, 277]. Indeed, it turns out these known constructions are already enough: we show Leverrier and Zémor's [277] recent qLDPC codes are in fact small-set HDX as well.[3]

Theorem 6.1.3 (Small-Set HDX Exist (informal Theorem 6.8.1))**.** *There exist constants* $\rho_1, \rho_2 \in (0,1)$ *and an explicit (polynomial time constructable) infinite family of bounded-degree*[4] *(3-term) chain complexes* $\{X_i\}$ *satisfying:*

1. X_i *has non-trivial 'co-homology,' i.e.* $\mathrm{im}(\delta_0) \neq \ker(\delta_1)$
2. X_i *is a* (ρ_1, ρ_2)*-SS HDX.*

The guarantees of Theorem 6.1.3 are stronger than those originally proved by Leverrier and Zémor [277] (see Section 6.3.1 for discussion), and give the strongest known form of bi-directional high dimensional expansion to date.[5] Indeed if one could remove the small-set requirement[6] or prove similar bounds for a 5-term chain complex, it would resolve the qLTC conjecture [230, 136, 281], a major open problem in quantum computation.

[3]It is interesting here to observe the qualitative parallel with a classical relation between small-set expanders and locally testable codes of [46, 174], though to our knowledge there is no quantitative connection between the results.

[4]A complex is bounded degree roughly if each element in $X(i)$ only has constantly many neighbors with respect to the boundary and co-boundary operators. See Section 6.4 for an exact definition.

[5]In fact it's worth mentioning we actually prove a stronger guarantee regarding *local* functions. See Remark 6.8.9 and discussion in Section 6.3.1.

[6]Though it is worth noting one must be careful that the dimension of the cohomology stays large, which requires weakening the expansion guarantee to a related notion called (co)-systolic expansion (the correct notion for qLTC regardless) [136].

6.2 Proof Overview

We now overview the constructions and proof techniques underlying our main result (Theorem 6.1.1). Broadly speaking, this breaks into two main steps:

1. Show any SS-HDX implies a hard instance of 3-XOR
2. Construct an explicit infinite family of SS-HDX.

To start, it will be useful to cover some basic background on CSPs, Sum-of-Squares, and chain complexes in a bit more detail. A more formal treatment is given in Section 6.4 and Section 6.7.

6.2.1 Background

In this work, we study the limitations of the Sum-of-Squares proof system for refuting MAX-k-XOR, a widely studied class of constraint satisfaction problems (CSPs). An instance of MAX-k-XOR $\mathcal{I}$ consists of a set of variables $\{x_i\}_{i\in[n]}$ and constraints $\{C_i\}_{i\in[m]}$, where each C_i is a boolean function of the form:

$$C_i(x) = \mathbf{1}\left\{x_{i_1} \oplus \ldots \oplus x_{i_j} = b_i\right\},$$

where $j = j(i) \leq k$ and $\{i_1, \ldots, i_j\} \subset [n]$. If all constraints have exactly k variables, we say $\mathcal{I}$ is an instance of k-XOR. We will usually omit the indicator $\mathbf{1}$ from notation when clear from context. The *value* of $\mathcal{I}$ is the maximum fraction of constraints that can be satisfied by any assignment, and we say $\mathcal{I}$ is $(1-\mu)$-*satisfiable* if there exists an assignment satisfying at least a $(1-\mu)$ fraction of constraints. We call an infinite family of instances $\{\mathcal{I}_i\}$ *explicit* if each instance can be constructed in deterministic polynomial time in the number of variables.

The Sum-of-Squares semi-definite programming hierarchy is a powerful algorithmic framework for approximating the value of any CSP (or more generally for solving con-

strained polynomial optimization problems). The hierarchy consists of *rounds* or *levels* of progressively stronger SDP relaxations (see Section 6.4). For the moment, it is enough to know that the round-t SoS relaxation is local[7] in the sense that it ranges over subsets of variables of size at most t. We will cover more details on the SoS framework as they arise.

Finally, it will be useful to have some basic terminology corresponding to chain complexes. Recall that a chain complex is a sequence $X : \mathbb{F}_2^{X(0)} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{X(1)} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{X(2)}$ such that $\partial_1\partial_2 = 0,\ \delta_1\delta_0 = 0$. Functions in the image of ∂_2 and δ_0 are called *boundaries* and *co-boundaries* respectively, and are denoted:

$$\mathrm{im}(\partial_2) = B_1, \quad \mathrm{im}(\delta_0) = B^1.$$

Functions in the kernel of ∂_1 and δ_1 are called *cycles* and *co-cycles* respectively, and are denoted:

$$\ker(\partial_1) = Z_1, \quad \ker(\delta_1) = Z^1.$$

The structure of a chain complex promises that $B_1 \subset Z_1 \subset \mathbb{F}_2^{X(1)}$ and $B^1 \subset Z^1 \subset \mathbb{F}_2^{X(1)}$. This leads to notions of *homology* and *co-homology* given by (co)-cycles mod (co)-boundary and respectively denoted:

$$H_1 = Z_1/B_1, \quad H^1 = Z^1/B^1,$$

where G/H denotes the quotient group. A complex has non-trivial co-homology if $B^1 \neq Z^1$.

6.2.2 From SS-HDX to Hardness

With notation out of the way, we can now discuss how to transform an expanding chain complex into a hard instance of 3-XOR. Before we give an informal theorem statement to this effect, it is instructive to overview how one even relates a CSP to a chain complex at all. To this end, let's first recall the classical construction of CSPs (also frequently seen

[7]We note the relaxation does have (low-degree) global consistency checks, so it is not fully a local algorithm in this sense.

in coding theory) based upon a bipartite graph $B = (L, R, E)$. In this setting, elements in L correspond to variables $\{x_v\}_{v \in L}$, and elements in R correspond to the set of constraints $\{C_r\}_{r \in R}$. Fixing some assignment $\beta \in \{0,1\}^R$ to constraints, the XOR instance classically associated with the graph B is characterized by ensuring the (mod 2) sum across neighbors of each $r \in R$ is given by $\beta(r)$:

$$C_r \coloneqq \left\{ \sum_{v \in N(r)} x_v = \beta(r) \pmod 2 \right\}. \tag{6.1}$$

In prior hardness constructions, B is typically picked at random in order to satisfy strong expansion properties, while β is typically chosen at random to ensure un-satisfiability (see e.g. [180, 340, 269]). While it is sometimes possible to de-randomize the choice of B and retain good inapproximability guarantees, no de-randomization of β better than brute force search over $\log(n)$-size instances was known up until this point.

The basic form of our XOR instances from chain complexes is actually very similar to Equation (6.1) (indeed they can be viewed as a special instantiation of this framework). Recall that a chain complex is a sequence:

$$X : \mathbb{F}_2^{X(0)} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{X(1)} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{X(2)},$$

and in particular that the co-boundary operator $\delta_0 : \mathbb{F}_2^{X(0)} \to \mathbb{F}_2^{X(1)}$ is a linear map. To define an instance of XOR on X, we simply move to the *graph representation* of δ_0. Namely, recall that any linear operator mapping from $\mathbb{F}_2^{X(0)}$ to $\mathbb{F}_2^{X(1)}$ can be written as an $(|X(1)| \times |X(0)|)$-dimensional matrix over $\mathbb{F}_2$. We can think of this matrix as the bipartite adjacency matrix of a graph on left vertex set $L = X(0)$ and right vertex set $R = X(1)$. Thus given a function $\beta \in \mathbb{F}_2^{X(1)}$, we construct the associated XOR instance, denoted $\mathcal{I}_{X,\beta}$

as in Equation (6.1) by adding the constraint for each $r \in X(1)$:

$$C_r \coloneqq \left\{ \sum_{\substack{v \in X(0): \\ e_r^T \delta_0 e_v = 1}} x_v = \beta(r) \pmod 2 \right\}, \tag{6.2}$$

where $e_v \in \mathbb{F}_2^{X(0)}$ and $e_r \in \mathbb{F}_2^{X(1)}$ are the standard basis vectors associated to $v \in X(0)$ and $r \in X(1)$. Note that $e_r^T \delta_0 \in \mathbb{F}_2^{X(0)}$ is just the list of neighbors of r, so this is indeed an instantiation of the standard bipartite framework. We note that this construction also generalizes the recent approach of [119] who built XOR instances via a 3-dimensional simplicial complex (4-uniform hypergraph) by letting *triangles* correspond to constraints, and *edges* correspond to variables. This is exactly the result of the above construction when applied to the natural chain complex associated with a 3-dimensional simplicial complex (see Section 6.4.2 for further details).

So far, we have not used the fact that δ_0 is part of a chain complex, or even the fact that the higher dimensional component $X(2)$ exists at all. This structure comes into play in the choice of β. Notice that by construction, the instance corresponding to X and a choice of β is satisfiable exactly when β is a co-boundary. Following the framework laid out in [119], the idea is to choose $\beta \in Z^1 \setminus B^1$, a function which is a co-cycle, but not a co-boundary. On a sufficiently expanding complex, this choice induces *global structure* on the XOR instance that cannot be captured by *local views* of the complex, where both the homology and co-homology look trivial. Since Sum-of-Squares only looks over local views in this sense, this leads to the following direct translation between SS-HDX and hard instances of XOR.

Theorem 6.2.1 (SS-HDX $\implies$ Hard XOR Instance (Informal Theorem 6.6.4))**.** *Let* $X : \mathbb{F}_2^{X(0)} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{X(1)} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{X(2)}$ *be an SS-HDX with non-trivial co-homology. Then there exist* $\mu_1, \mu_2 \in (0,1)$ *such that for any* $\beta \in Z^1 \setminus B^1$*, the associated XOR instance* $\mathcal{I}_{X,\beta}$ *satisfies:*

1. *Soundness: $\mathcal{I}_{X,\beta}$ is at most $(1-\mu_1)$-satisfiable,*

2. *Completeness: $\mathcal{I}_{X,\beta}$ cannot be refuted by $\mu_2|X(0)|$ levels of the SoS hierarchy.*

Before moving on to the construction of SS-HDX, let's discuss how small-set expansion implies soundness and completeness for these instances. Soundness, the simpler of the two, intuitively comes from the fact that small-set co-boundary expansion promises that any element in $Z^1 \setminus B^1$ must be *far from the co-boundary*.[8] Recall that by construction, the instance $\mathcal{I}_{X,\beta}$ is satisfiable exactly when $\beta \in \mathbb{F}_2^{X(1)}$ is a co-boundary. Intuitively one might then expect that functions which are far from the co-boundary would therefore be far from satisfiable. Indeed this intuition holds true—it is easy to show this robust version of the statement holds for small-set co-boundary expanders, and therefore that our instances are far from satisfiable as well.

Completeness is somewhat trickier and, unlike soundness, does actually require the full power of small-set boundary expansion. We stated earlier that the completeness of our instances, much like those of [119], comes from the fact that the global structure of (co)-homology cannot be detected through local views of the complex. This is formalized by observing that small-set boundary expansion can be equivalently re-stated as the following *isoperimetric inequality* (see Lemma 6.5.4): "small, minimal[9] functions have large boundaries." Largely following [119] (who use a much weaker isoperimetric inequality for the LSV complex due to Gromov [181]), the idea is then to combine this fact with the classical arguments of Ben-Sasson and Wigderson [60] to show that the width[10] of any refutation of $\mathcal{I}_{X,\beta}$ in the $\oplus$-resolution proof system[11] is large. Since Schoenebeck [340]

[8] It is worth noting that this property, called *co-systolic distance*, is quite well studied. Indeed as we will soon discuss it is exactly the property needed (in both directions) to build good qLDPC codes [315], and was also used directly by [119] to prove soundness of their 3-XOR instances by the same argument stated here.

[9] A function $f \in \mathbb{F}_2^{X(1)}$ is said to be minimal if adding any boundary can only increase its size (Hamming weight).

[10] The width of a refutation is the largest number of variables appearing in any equation.

[11] In this proof system, one is allowed to combine linear equations (equivalently XOR constraints) $\ell_1 = b_1$ and $\ell_2 = b_2$ to derive the equation $\ell_1 \oplus \ell_2 = b_1 \oplus b_2$. A refutation is a proof based on this rule deriving a

showed any such bound transfers to a completeness lower bound for Sum-of-Squares, this completes the proof.

In slightly more detail, a refutation in the $\oplus$-resolution system can be viewed as an (in-degree two) DAG where leaves correspond to the original XOR constraints, internal nodes correspond to the XOR of their two parents (as in the $\oplus$-resolution derivation rule), and the root derives the contradiction $0 = 1$. Recall that each element $s \in X(1)$ corresponds to a constraint in our XOR instance. Following [119], the idea is to assign a function in $h_v \in \mathbb{F}_2^{X(1)}$ for each node v in the DAG that tracks which XOR constraints are being used at that node. The *boundary* of this function, $\partial_1 h_v \in \mathbb{F}_2^{X(0)}$, is exactly the set of variables appearing in the equation corresponding to node v. Thus lower bounding the width of the refutation boils down to finding a node with large boundary.

This is where small-set boundary expansion (namely the isoperimetric formulation) finally comes into play. In particular, the corresponding inequality states that it is enough to find a node v of 'medium' weight:[12] small enough that one can apply the inequality, but large enough to result in a large boundary. This can be done by fairly standard potential arguments (see e.g. [60, 119]) where one sets of up a potential function tracking this weight throughout the DAG, and argues that the leaves have small potential, the root has large potential, and that potential is sub-additive. This implies the existence of an interior node with medium potential and completes the proof. The details are given in Section 6.6.

Finally, before moving on to overviewing our construction of SS-HDX, we note that except in very special cases (e.g. the simplicial complexes considered in [119]), the CSPs given by Equation (6.2) (and therefore also Theorem 6.2.1) are actually instance of MAX-k-XOR, not 3-XOR, where k is given by the maximum degree of the complex. As it turns out, this is not a significant issue because the SS-HDX we construct in the next section are *bounded degree*, meaning not only that every constraint in the XOR has

contradiction ($0 = 1$), which is equivalent in our setting to showing the XOR instance is unsatisfiable.

[12]We note that weight here is not just the standard Hamming weight, but must take into account distance from the boundary as well. See Section 6.6.

a constant number of variables, but also that every variable only appears in a constant number of constraints. This observation allows us to move to hard instances of 3-XOR by standard NP-reduction type arguments within the SoS hierarchy [340, 354] while only losing constant factors in the soundness and levels of hardness for SoS.

6.2.3 Constructing SS-HDX

Now that we know how to transform an expanding chain complex into a hard instance of 3-XOR, we turn our attention to the construction of such complexes. Our method relies on recent breakthroughs on LTCs [117, 281] and quantum LDPC codes [315, 277]. As such, we'll split this section into three parts: a review of the connection between quantum LDPC codes and expanding chain complexes, the recent qLDPC construction of Leverrier and Zémor [277], and our proof of small-set (co)-boundary expansion.

Quantum LDPC Codes and Chain Complexes

A classical error correcting code is a method of encoding k classical bits into $n > k$ classical bits such that it is possible to recover the original bit string even if the encoded string becomes corrupted. We will consider *linear codes*, which are defined by a linear operator $M : \mathbb{F}_2^n \to \mathbb{F}_2^{n-k}$ called the parity check matrix,[13] where the corresponding code $\mathcal{C} \coloneqq \ker M$.

Similar to the classical setting, a *quantum* code encodes quantum bits into a larger number of quantum bits, but is resistent to two types of corruption: the X-type errors (bit flips) and the Z-type errors (phase flips). In this work, we will focus on a popular notion of quantum codes called *CSS-codes* [86, 347], which come with the benefit of having an entirely classical interpretation. In particular, a length n CSS-code is made up of two classical codes $\mathcal{C}_0 \coloneqq \ker M_0 \subset \mathbb{F}_2^n$ and $\mathcal{C}_1 \coloneqq \ker M_1 \subset \mathbb{F}_2^n$ such that $\mathcal{C}_0^\perp \subset \mathcal{C}_1$, or equivalently

[13] We note the parity check matrix is traditionally denoted by 'H,' but this conflicts with the notation for homology.

$M_1 \cdot M_0^T = 0$.[14] The dimension of the code is defined as $k = \dim\mathcal{C}_0 - \dim\mathcal{C}_1^\perp$, and its *distance* (which measures how much corruption it can handle) is defined as $d = \min(d_x, d_z)$ where

$$d_x = \min_{v \in \mathcal{C}_0 \setminus \mathcal{C}_1^\perp} |v|, d_z = \min_{v \in \mathcal{C}_1 \setminus \mathcal{C}_0^\perp} |v|$$

and d_x (d_z) is called the X-distance (Z-distance). The quantum low-density parity-check (LDPC) conjecture, recently resolved by [315], states that there exists a family of quantum CSS codes with linear dimension and distance, $k = \Theta(n)$ and $d = \Theta(n)$, where M_0 and M_1 have at most some constant number of ones in any row or column (and thus are 'low-density' parity check matrices).

Since we are promised by definition that $M_1 \cdot M_0^T = 0$, it is easy to see that any CSS-code induces the following chain complex:

$$X : \mathbb{F}_2^{m_0} \underset{M_0}{\overset{M_0^T}{\rightleftarrows}} \mathbb{F}_2^n \underset{M_1^T}{\overset{M_1}{\rightleftarrows}} \mathbb{F}_2^{m_1},$$

where $m_i = \dim\left(\mathrm{im}(M_i)\right)$. Indeed the same holds in reverse as well, given a chain complex

$$X : \mathbb{F}_2^{X(0)} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{X(1)} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{X(2)},$$

one obtains a quantum CSS code by letting $M_0 \coloneqq \partial_1$, and $M_1 \coloneqq \delta_1$.

In fact, it turns out this equivalence between quantum CSS codes and chain complexes runs deeper: all of the discussed properties (e.g. distance, LDPC) have analogs in the homological language we developed in the previous section. The classical codes $\mathcal{C}_0$ and $\mathcal{C}_1$, for instance, correspond to the cycles and co-cycles of the chain complex ($\mathcal{C}_0 = Z_1, \mathcal{C}_1 = Z^1$), while the dual codes $\mathcal{C}_0^\perp$ and $\mathcal{C}_1^\perp$ correspond to the co-boundaries and boundaries ($\mathcal{C}_0^\perp = B^1, \mathcal{C}_1^\perp = B_1$). The dimension of the code k corresponds to the

[14] Here $\mathcal{C}_0^\perp$ denotes the dual code, consisting of all elements orthogonal to $\mathcal{C}_0$. This code is generated by the transpose of the parity check matrix M_0^T.

dimension of the co-homology ($k = \dim H^1$), and the maximum degree of the complex corresponds to the maximum density of the parity check codes (so the bounded-degree and LDPC conditions are equivalent). Finally, the X-distance and Z-distance of the code correspond to what is known as the *(co)-systolic distance* of the chain complex, the minimum weight of any (co)-cycle that is not a (co)-boundary:

$$d_x = \min_{v \in \mathcal{C}_0 \setminus \mathcal{C}_1^\perp} |v| = \min_{v \in Z_1 \setminus B_1} |v|,$$
$$d_z = \min_{v \in \mathcal{C}_1 \setminus \mathcal{C}_0^\perp} |v| = \min_{v \in Z^1 \setminus B^1} |v|.$$

In [315] and [277], the authors construct two different explicit families of good quantum LDPC codes. This partially solves our problem since the codes correspond to a family of bounded-degree chain complexes with non-trivial co-homology and linear co-systolic distance (which is enough to imply soundness of our XOR construction). We will show these complexes in fact satisfy the stronger small-set (co)-boundary expansion condition, which as discussed in the previous section further implies completeness and (up to reduction to 3-XOR) finishes the proof of Theorem 6.1.1.

Leverrier and Zémor's qLDPC Codes

Before discussing the proof, we need to overview the original construction of [277]. A significantly more detailed description of the construction and its associated components is given in Section 6.7 and Section 6.8.

Leverrier and Zémor's qLDPC codes are based on a classical object called a *Tanner code* [351]. Given an n_0-regular graph $\mathcal{G} = (V, E)$ and a linear code C of length n_0, the Tanner code $T(\mathcal{G}, C) \subset \mathbb{F}_2^E$ is

$$\{c \in \mathbb{F}_2^E : \forall v \in V, c|_{E(v)} \in C\},$$

where $c|_{E(v)} \in \mathbb{F}_2^{n_0}$ is the vector formed by the values on the edges incident to v. Tanner codes have long been used in coding theory. The main insight of [277] was to observe that one can construct a quantum CSS code via two Tanner codes coming from a higher-dimensional object called the *left-right Cayley complex*, recently developed in [117] to construct c3-LTCs.

The left-right Cayley complex corresponding to a group G and two sets of generators $A = A^{-1}$ and $B = B^{-1}$ consists of a vertex set $V = G$, edges given by (left) Cayley graph $C(G, A)$ and (right) Cayley graph $C(G, B)$, and higher-dimensional 'squares' of the form $\{g, ag, gb, agb\}$ for $g \in G, a \in A, b \in B$. More formally, [277] consider the *double cover* of this complex where:

- The vertices are $V = V_0 \cup V_1$ where $V_0 = G \times \{0\}$ and $V_1 = G \times \{1\}$.
- The 'A-edges' and 'B-edges' are respectively:

$$E_A = \{\{(g, 0), (ag, 1)\} : g \in G, a \in A\}, E_B = \{\{(g, 0), (gb, 1)\} : g \in G, b \in B\}.$$

- The squares are

$$F = \{\{(g, 0), (ag, 1), (gb, 1), (agb, 0)\} : g \in G, a \in A, b \in B\}.$$

Notice each square contains exactly two vertices in V_0 and two vertices in V_1. This allows us to think of each square as an *edge* between two vertices in V_0 (or V_1) and to define corresponding graphs $\mathcal{G}_0^{\square} = (V_0, F)$ and $\mathcal{G}_1^{\square} = (V_1, F)$. The local view around each vertex in $(g, i) \in \mathcal{G}_i^{\square}$ then corresponds to the squares $\{(g, i), (ag, 1 - i), (gb, 1 - i), (agb, i)\}$ for $a \in A, b \in B$. Assuming $|A| = |B| = \Delta$ for some constant Δ, we will always think about these local views as square matrices with rows indexed by A and columns indexed by B.

Leverrier and Zémor [277] observed that the Tanner codes associated to these

graphs, $\mathcal{C}_0 = T(\mathcal{G}_0^{\square}, C_0^{\perp})$ and $\mathcal{C}_1 = T(\mathcal{G}_1^{\square}, C_1^{\perp})$, give a quantum CSS code (i.e. satisfy $\mathcal{C}_0^{\perp} \subset \mathcal{C}_1$) whenever the associated local codes $C_0 = C_A \otimes C_B$ and $C_1 = C_A^{\perp} \otimes C_B^{\perp}$ are tensors[15] of linear codes $C_A \subseteq \mathbb{F}_2^A$ and $C_B \subseteq \mathbb{F}_2^B$. Furthermore, they showed that whenever $C_A, C_B, C_A^{\perp}, C_B^{\perp}$ have linear distance and the codes $C_1^{\perp} = C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B$, and $C_0^{\perp} = C_A^{\perp} \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B^{\perp}$ satisfy certain robustness properties (see Section 6.7.5), then the associated quantum code has linear distance. [277] complete their construction by showing random base codes C_A, C_B satisfy these properties with high probability. Note that because these base codes are constant size, this final step can be brute-forced to maintain explicitness of the construction.

Proving Small-Set (Co)-Boundary Expansion

With [277]'s construction in hand, we can now sketch the proof of small-set (co)-boundary expansion. As mentioned previously, all other major requirements (e.g. non-trivial homology, bounded-degree) already follow from the fact that the complex corresponds to a good qLDPC code. We will focus here on proving small-set *co*-boundary expansion in particular, but we note that small-set boundary expansion follows the same argument by symmetry of [277]'s construction.

With this in mind, recall that small-set co-boundary expansion can equivalently be phrased as an isoperimetric inequality for small, minimal functions (see Lemma 6.5.4). In particular, to show small-set co-boundary expansion for the chain complex

$$X : \mathbb{F}_2^{m_0} \xrightarrow{\delta_0 := M_0^T} \mathbb{F}_2^n \xrightarrow{\delta_1 := M_1} \mathbb{F}_2^{m_1},$$

it is enough to show there exist constants $\rho_1, \rho_2 \in (0, 1)$ such that any minimal $x \in \mathbb{F}_2^n$ with weight $|x| \leq \rho_1 n$ has large boundary: $|\delta_1 x| \geq \rho_2 |x|$. We proceed by contradiction. Assuming $|\delta_1 x| < \rho_2 |x|$, we will show x is not minimal by finding $y \in B^1$ such that

[15]The tensor code $C_A \otimes C_B$ is the set of matrices whose rows are given by elements of C_B and columns are given by elements of C_A.

$|x + y| < |x|$.

The proof of this fact largely follows the technique of [277] for proving the weaker co-systolic distance property. The main difference is that while [277] only consider functions $x \in \mathbb{F}_2^n$ that are co-cycles, we consider arbitrary functions. In particular, recall that the co-cycles in our construction correspond to codewords in the Tanner code $T(\mathcal{G}_1^{\square}, C_1^{\perp})$, or equivalently to functions $x \in \mathbb{F}_2^n$ whose 'local view' around each vertex $(g, 1) \in V_1$ is given by a codeword of $C_1^{\perp}$. Since our functions do not a priori have this structure, we will need to track the set of 'violations' coming from local views that are not codewords (this essentially corresponds to where $\delta_1 x$ is non-zero).

To this end, recall x is a bit string indexed by the squares of the double-covered Cayley complex, and let $S \subset V_1$ denote the set of vertices incident to any square in x. We partition S into three parts: the violated vertices S_v, the normal vertices S_n, and the exceptional vertices S_e. A vertex is *violated* if the local view of x around the vertex does not form a codeword in $C_1^{\perp}$. When the local view does form a codeword, if the codeword has weight less than $w := \Delta^{3/2-\epsilon}$ we call it *normal*, and otherwise call it *exceptional*. This weight-based distinction comes from the robustness condition of the local tensor code. We cover this in detail in Section 6.7.5, but for the moment it is sufficient to think of robustness as a structural condition forcing codewords with weight less than w to be zero outside of a small number of rows and columns. In particular, this promises that each column (respectively row) in the local view of a normal vertex is at most $O(\Delta^{1/2-\epsilon})$ away from a codeword in C_A (respectively C_B).

Following [277], our goal is now to find a vertex $v \in V_0$ that shares $\Omega(\Delta)$ columns or rows with S_n. As long as S_e and S_v are not too large compared to S_n, robustness of the code then implies the local view of v is within $O(\Delta^{3/2+\epsilon})$ of a codeword $c \in C_A \otimes C_B$, but also has total weight $\Omega(\Delta^2)$.[16] This means we can construct a vector $y \in B^1$ by defining y to be c on the local view of v and 0 everywhere else. Since $x + y$ and x match outside

[16]We note C_A and C_B can be chosen to have linear distance to ensure this.

the local view (where x has weight $\Omega(\Delta^2)$ and $x+y$ has weight $O(\Delta^{3/2+\epsilon})$), this implies $|x+y| < |x|$ as desired.

It therefore remains to find such a vertex $v \in V_0$, which is the main technical component of the proof. Let $T \subset V_0$ be the vertices that share at least one 'heavy' column or row with a normal vertex (that is one with many 1s). One can equivalently think of this as an edge between V_0 and V_1 that is 'heavy' in the sense that it is contained in many squares in x. The idea is then to show that there are many such heavy edges passing between S and T. Using expansion of the underlying graph and our assumption $|\delta_1 x| < \rho_2 |x|$, one can prove that T, S_e, and S_v are small compared to S_n. This implies that a typical vertex in T has not just one, but $\Omega(\Delta)$ heavy edges to S_n, which in turn corresponds to sharing $\Omega(\Delta)$ rows and columns with normal vertices and completes the proof.

6.3 Discussion

6.3.1 Related Work

Sum-of-Squares Lower Bounds:.

At a conceptual level, our work fits into a long line of research on the limitations of Sum-of-Squares and related proof systems (e.g. Nullstellensatz [53, 178], Polynomial Calculus [99, 85]), and LP/SDP hierarchies (e.g. Sherali-Adams [91, 294, 62], Lovász-Schrijver [7, 341, 164]). Most relevant to our setting is the line of work on Sum-of-Squares lower bounds initiated by Grigoriev [180] (and later independently Schoenebeck [340]), who used boundary expansion to prove random 3-XOR instances cannot be refuted by $\Omega(n)$ levels of SoS. This lead to a number of works improving integrality gaps for more general classes of random k-CSPs [354, 45, 90, 269] along with a number of other combinatorial optimization problems by reduction [340, 354].

In a sense, these prior works on SoS lower bounds for random instances can be viewed as increasingly strong and general formulations of the statement: 'Sum-of-Squares

fails to capture the probabilistic method.' In contrast, Dinur, Filmus, Harsha, and Tulsiani [119] recently exhibited the first *explicit* families of CSPs hard for Sum-of-Squares based on an algebraic, highly structured family of objects called Ramanujan (or LSV) complexes [290], suggesting a new paradigm of hardness for structured instances. Due to the poor systolic expansion of the Ramanujan complex, [119]'s bounds only hold up to $O(\sqrt{\log(n)})$ rounds of SoS as compared to $\Omega(n)$ levels for random instances. Nevertheless, the authors conjectured it might be possible to use such anti-random objects to fool $\Omega(n)$ levels as well. Our work can be viewed as a confirmation of this general hypothesis: anti-random structure (in particular certain *algebraic* structure) is indeed as hard as random for Sum-of-Squares.

High Dimensional Expansion:.

High dimensional expansion in the form we consider (i.e. topological expansion) was originally introduced by Linial and Meshulam [283] to study the vanishing of cohomology on random simplicial complexes, and independently by Gromov [182] to study the topological overlapping principle. The closest notion of expansion in the literature to our SS-HDX is the stronger simplicial co-isoperimetric inequality for small, *locally* minimal[17] sets of Kaufman, Kazhdan, and Lubotzky [230], used to construct bounded degree co-systolic expanders (another weakening of co-boundary expansion that replaces distance from B^1 with distance from Z^1). A similar strategy was recently employed by Lin and Hsieh to construct c3-LTCs [281] and (conditional) qLDPC codes [282]. It is worth noting that this stronger condition actually holds for our construction as well (see Remark 6.8.9).

Quantum Codes and LTCs:.

Quantum LDPC and locally testable codes have long been known to share a close connection with topological notions of high dimensional expansion (see e.g. [136]). Indeed it was qLDPC constructions based on the Ramanujan complex [142, 244] that first broke

[17]A function is locally minimal if its weight cannot be decreased by adding the image of any standard basis vector $\partial_2(e_v)$. Any minimal function is also locally minimal (and the converse does not in general hold), so this is a strictly stronger notion of expansion than we study.

the $\sqrt{n}$ distance barrier and started the race to good qLDPCs [142, 80, 314, 244, 193, 315, 219, 277]. As discussed in Section 6.2.3, qLDPC codes satisfy a weaker variant of expansion called (co)-systolic distance, but must do so *in both directions*. This is in strong contrast to typical constructions in the HDX literature which, due to the inherent asymmetry of simplicial complexes, typically have very poor boundary expansion (indeed this is also why we avoid simplicial complexes in this work). Such a guarantee was only recently achieved by Panteleev and Kalachev [315] using refined products of chain complexes, and very recently simplified through a more geometric lens by Leverrier and Zémor [277]. Since small-set (co)-boundary expansion is a stronger notion than (co)-systolic distance (see Section 6.5), our analysis provides the strongest form of two-sided topological expansion to date. Further, this stronger form of two-sided expansion gives some hope for a positive resolution of the famous qLTC conjecture, which in some sense requires another notion of lying somewhere between basic (co)-systolic distance and (co)-boundary expansion, but requires weaker expansion for all sets rather than the strongest notion for small sets as in SS-HDX [136].

6.3.2 Further Directions

Strongly Explicit Instances:.

The XOR instances we construct may be built in polynomial time, but are not necessarily *strongly* explicit because it is unclear whether the target function can be computed locally in polylog time. Subsequent to this work, Golowich and Kaufman [173] resolved this issue by a clever trick planting the all 1s function as a co-systole in a variant of [277] and [315]'s balanced-product type construction.

Improved Integrality Gaps:.

We prove the existence of an explicit family of 3-XOR instances with a constant integrality gap of 1 v.s $1-\mu$ for 3-XOR, which falls short of reaching the 1 v.s $\frac{1}{2}+\varepsilon$ gap exhibited by random instances [180, 340]. While standard reductions in the SoS hierarchy

can improve our gap to arbitrarily close ($1-\varepsilon$ v.s $\frac{1}{2}+\varepsilon$), perfect completeness is lost in the process. The same issue was observed in [119]'s original explicit construction from the Ramanujan complex. They asked whether it is possible to bypass imperfect completeness by giving a direct construction with co-systolic distance at least $\frac{1}{2}-\varepsilon$. This remains a natural open question in our setting as well—can one directly construct a small-set boundary expander with co-systolic distance $\frac{1}{2}-\varepsilon$? This would lead to a 1 v.s $\frac{1}{2}+\varepsilon$ gap for MAX-k-XOR. Another natural question is whether such a bound can be transferred to 3-XOR without losing factors in the soundness. Our current reduction loses a factor in k, but we have made no attempt to optimize this step (since any constant gap is sufficient to amplify with PCP techniques if one is okay with imperfect completeness).

Hardness Beyond XOR:.

Many of the best integrality gaps known for combinatorial optimization problems (e.g. maximum independent set, chromatic number) are proved by reduction from k-CSPs [354]. Unfortunately, such reductions are often *randomized*, so they do not imply explicit hard instances even when combined with our XOR construction. This raises a natural question: can we build *explicit* reductions from k-CSPs to classical combinatorial problems such as maximum independent set? Combined with our construction, this could lead to new families of hard instances for many well-studied combinatorial optimization problems. On a related note, it is worth observing that these reductions usually rely on CSPs with better integrality gaps than k-XOR. For instance, it is not hard to see that while random instances of k-XOR only exhibit a 1 v.s $1/2+\varepsilon$ integrality gap, more constrained k-CSPs (e.g. constraints of the form $Ax=b$ for some matrix $A \in \mathbb{F}_2^{d\times k}$) can lead to much larger integrality gaps up to 1 v.s $\frac{2k}{2^k}+\varepsilon$ [354]. Can we use high dimensional expanders to recover explicit k-CSPs matching these bounds?

Small-Set HDX and Hardness of Approximation:.

Small set expansion plays a fundamental role in hardness of approximation, ranging from use as a computational hardness assumption itself [325], to its pivotal use in the proof of the 2-2 games conjecture [253, 126, 125, 47, 252, 255] and recent converse use for algorithms for unique games [37, 38]. This work gives the first application of *high dimensional* small-set expansion to hardness of approximation, raising the natural question: does this high dimensional variant have a broader role to play in the field as well?

6.4 Preliminaries I: SS-HDX to Hardness

We now cover the preliminary definitions required to understand our general translation of expanding chain complexes into hard instance of 3-XOR, including basics on Sum-of-Squares, chain complexes, and traditional notions of high dimensional expansion. Background required for the HDX construction itself (e.g. on left-right Cayley complexes, robust tensor codes, etc.) is postponed to Section 6.7.

6.4.1 Sum of Squares and Refutations

The Sum-of-Squares Semidefinite Programming Hierarchy is a powerful method for approximately solving constrained polynomial optimization problems, and is in particular the strongest known algorithmic framework for approximating CSPs. In brief, the SoS heirarchy presents a series of successively stronger SDP relaxations of a problem, where the 'round-t' relaxation optimizes over t-local views and runs in time $n^{O(t)}$. We refer the reader to [49, 150] for general information on the SoS hierarchy.

In this work, we focus in particular on the SoS relaxations of MAX-k-XOR, the family of CSPs on n variables $\{x_1, \ldots, x_n\}$ and m constraints $\{C_i\}_{i\in[m]}$ of the form:

$$x_{i_1} \oplus \ldots \oplus x_{i_j} = z_i,$$

where $z_i \in \{0,1\}$, $\{i_1, \ldots, i_j\} \subset [n]$, and $j = j(i) \leq k$. Let $T_i \subset [n]$ denote the set of variables appearing in the ith constraint. Then the round-t SoS SDP relaxation for MAX-k-XOR can be written as:

Round-t SoS Relaxation of MAX-k-XOR:

Input: variables $\{v_S\}_{S \in \binom{[n]}{\leq t}}$

Maximize: $\frac{1}{2} + \frac{1}{2m} \sum_{i=1}^{m} (-1)^{z_i} \langle v_{T_i}, v_\emptyset \rangle$

Constraint to:

1. $\forall S_1 \oplus S_2 = S_3 \oplus S_4, |S_i| \leq t : \langle v_{S_1}, v_{S_2} \rangle = \langle v_{S_3}, v_{S_4} \rangle$

2. $\forall S, |S| \leq t : \|v_S\|_2 = 1$

We refer to the maximum obtained by this SDP as the *value* of the round-t relaxation, and say an infinite family of instances of MAX-k-XOR is hard for (or cannot be refuted by) t rounds of Sum of Squares if there exists a constant μ such that every instance is at most $(1-\mu)$-satisfiable, but the round-t SDP relaxation has value 1. In other words, t-rounds of the SoS hierarchy cannot distinguish between completely satisfiable and $(1-\mu)$-satisfiable instances—this is often said to induce an *integrality gap* for the problem of size $\frac{1}{1-\mu}$.

Rather than working directly with the Sum-of-Squares SDP relaxations, we prove our hardness results through a fruitful connection with refutation complexity due to Schoenebeck [340] and Tulsiani [354]. More formally, following [119] we will use a proof system called $\oplus$-resolution where, given a system of linear equations Λ over $\mathbb{F}_2$, we may derive new equations by mod 2 summation:

$$\{\ell_1 = b_1\}, \{\ell_2 = b_2\} \implies \ell_1 \oplus \ell_2 = b_1 \oplus b_2.$$

A *refutation* in this system is a derivation that $0 = 1$, and in our setting corresponds to a proof that the XOR instance given by Λ is unsatisfiable. Schoenebeck [340] and Tulsiani

[354] showed that any system without a short refutation has a matching SoS lower bound.

Theorem 6.4.1 ([340, Lemma 13] (as stated in [119]))**.** *Let Λ be a system of linear equations in n variables over $\mathbb{F}_2$. If all refutations of Λ have an equation using at least $2t$ variables, then the round-t SoS Relaxation of Λ has value 1.*

6.4.2 Chain Complexes

While previous works constructing hard instances of CSPs rely on structure coming from graphs (e.g. [180, 340]) or hypergraphs [119], we take inspiration from recent work on c3-LTCs [117, 281] and qLDPC codes [315, 277] and instead study a more general set of objects called *chain complexes*.

Definition 6.4.2 (Chain Complex)**.** Let $X(0)$, $X(1)$, and $X(2)$ be sets, and $\partial_2 : \mathbb{F}_2^{X(2)} \to \mathbb{F}_2^{X(1)}$, $\partial_1 : \mathbb{F}_2^{X(1)} \to \mathbb{F}_2^{X(0)}$ linear maps. The sequence

$$X : \mathbb{F}_2^{X(0)} \xleftarrow{\partial_1} \mathbb{F}_2^{X(1)} \xleftarrow{\partial_2} \mathbb{F}_2^{X(2)}$$

is called a (3-term) chain complex if $\partial_1 \partial_2 = 0$.

For the sake of intuition, let's take a moment to see why chain complexes are indeed a generalization of hypergraphs. Given an r-uniform hypergraph $H \subseteq \binom{[n]}{r}$, let $X(i) \subset \binom{[n]}{i}$ denote any i-set contained in some r-set in H. H then induces an $(r+1)$-term chain complex:[18]

$$X : \mathbb{F}_2^{X(0)} \xleftarrow{\partial_1} \mathbb{F}_2^{X(1)} \xleftarrow{\partial_2} \dots \xleftarrow{\partial_r} \mathbb{F}_2^{X(r)},$$

where $\partial_i f(x)$ is given by summing $f \pmod 2$ over x's 'boundary:'

$$\forall f \in \mathbb{F}_2^{X(i)} : \partial_i f(x) = \sum_{y \in X(i) : y \supset x} f(y). \tag{6.3}$$

[18]Note $X(0)$ is defined to be the empty set, and that our indexing is off by 1 from the usual notation in topology.

For instance, when x is a vertex, $\partial_2 f(x)$ averages over all edges containing x. As such, ∂ is usually called the *boundary operator*, and it can be checked without too much difficulty that $\partial_{i-1}\partial_i = 0$ (e.g. for $r = 3$, this follows by noting a vertex is incident to either 0 or 2 edges of any given triangle).

In fact, the boundary operators can actually always be seen to have a similar form to Equation (6.3), even on a generic chain complex. This follows from passing to the matrix representation as discussed in Section 6.1. Namely, we may view our 3-term chain complex as a pair of bipartite graphs $B_0 = (X(0), X(1), E_1)$ and $B_1 = (X(1), X(2), E_2)$, whose bipartite adjacency matrices are given by the matrix representations of ∂_1 and ∂_2 respectively (in the standard basis). In this setting, it is easy to see that ∂_1 and ∂_2 are also given by mod 2 summation over neighbors on these underlying bipartite graphs:

$$\forall f \in \mathbb{F}_2^{X(1)} : \partial_1 f(x) = \sum_{y \in X(1):(x,y)\in E_1} f(y) \pmod 2$$
$$\forall f \in \mathbb{F}_2^{X(2)} : \partial_2 f(y) = \sum_{z \in X(2):(y,z)\in E_2} f(z) \pmod 2,$$

where we have assumed for simplicity that ∂_1 and ∂_2 are *non-degenerate* in the sense that every row and column have at least one 1.[19] All complexes we study are non-degenerate, so we make this assumption throughout.

In matrix form, it is also easy to see that the transpose operators of ∂, called the *co-boundary operators* and denoted $\delta_0 \coloneqq \partial_1^T$ and $\delta_1 \coloneqq \partial_2^T$, also form a chain complex in the opposite direction. As a result, we will usually write our chain complexes in the following form:

$$X : \mathbb{F}_2^{X(0)} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{X(1)} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{X(2)}.$$

We call elements of $\mathbb{F}_2^{X(i)}$ *i-chains*, and note $\mathbb{F}_2^{X(i)}$ is often written as "C_i" in the literature. We avoid this notation since it conflicts with classical notation for codes used later in the

[19] In a graph, for instance, non-degeneracy corresponds to have no free-floating (degree 0) vertices.

paper.

Finally, before moving on to expansion on chain complexes, we cover two further concepts that will control important parameters of our corresponding XOR instaces: *maximum degree* and *explicitness*.

Definition 6.4.3 (Maximum Degree)**.** The maximum degree of a chain complex $X : \mathbb{F}_2^{X(0)} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{X(1)} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{X(2)}$ is the maximum Hamming weight[20] across rows and columns of ∂_1 and ∂_2.

In the bipartite graph view, this is simply the maximum vertex degree across both graphs. We call an infinite family of chain complexes *bounded degree* if there exists some constant $d \in \mathbb{N}$ such that all complexes in the family have maximum degree at most d.

Finally, in this work we will be interested in infinite families of chain complexes (and their associated XOR instances), so we need to define a notion of computational complexity over these objects. We will follow the standard notions used for expander families, and call a family of complexes *explicit* if its elements can be constructed in deterministic polynomial time (this is often called *mildly explicit*, but the difference is not particularly important in our setting).

Definition 6.4.4 (Explicit Chain Complexes)**.** We call an infinite family of chain complexes $\{X_i\}$ explicit if there exists a determinstic algorithm computing each X_i in time polynomial in $|X_i(0) \cup X_i(1) \cup X_i(2)|$.

All complexes studied in this work will be bounded-degree, in which case this notion may equivalently be defined looking only at the size of $X_i(0)$. This corresponds correctly to the standard notion of complexity for the associated k-CSP family where $|X_i(0)|$ gives the number of variables.

[20]The Hamming weight of binary vector v, denoted $|v|$, counts the number of entries with a 1.

6.4.3 Homology and High Dimensional Expansion

High dimensional expansion is a generalization of expansion in graphs originally introduced by Linial and Meshulam [283] (and later independently by Gromov [182]) to study the vanishing of homology in simplicial complexes. In this section we cover the basics of homology and introduce Linial and Meshulam's original notion of *(co)-boundary expansion*. These notions (or modifications thereof) will play an important role in our CSP construction.

Following standard notation, we call functions in the kernel of ∂_i *cycles*, and functions in the kernel of δ_i *co-cycles*, denoted:

$$Z_i = \ker(\partial_i), \quad Z^i = \ker(\delta_i).$$

Since $\delta^2 = \partial^2 = 0$, notice that $\mathrm{im}(\partial_{i+1})$ are always cycles, and $\mathrm{im}(\delta_{i-1})$ are always co-cycles. We call functions in these classes *boundaries* and *co-boundaries* respectively, denoted:

$$B_i = \mathrm{im}(\partial_{i+1}), \quad B^i = \mathrm{im}(\delta_{i-1}).$$

The *homology* and *co-homology* of the chain complex correspond to (co)-cycles mod (co)-boundary:

$$H_i = Z_i/B_i, \quad H^i = Z^i/B^i,$$

where G/H denotes the quotient group. The notions of cycles and boundaries can be used to define a natural generalization of expander graphs to chain complexes called *(Co)-boundary expansion*.

Definition 6.4.5 ((Co)-Boundary Expansion)**.** We call $X : \mathbb{F}_2^{X(0)} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{X(1)} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{X(2)}$ a ρ-boundary expander if the weight of any element in $\mathbb{F}_2^{X(1)} \setminus B_1$ is proportional to its

distance from the boundary:

$$\forall f \in \mathbb{F}_2^{X(1)} \setminus B_1 : \frac{|\partial_1 f|}{d(f, B_1)} \geq \rho,$$

where $d(f, B_1) = \min_{b \in B_1} |f + b|$. Similarly, X is an ρ-co-boundary expander if:

$$\forall f \in \mathbb{F}_2^{X(1)} \setminus B^1 : \frac{|\delta_1 f|}{d(f, B^1)} \geq \rho.$$

Since this definition may seem un-motivated at first glance, let's again take a look at the case of a graph $G = (V, E)$ which induces the (3-term) chain complex:

$$X : \mathbb{F}_2^{\emptyset} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{V} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{E}.$$

It is not hard to see that the co-boundary expansion of this chain is exactly Cheeger's constant:

$$h(G) \coloneqq \min_{S \neq V, \emptyset} \left\{ \frac{E(S, V \setminus S)}{\min\{|S|, |V \setminus S|\}} \right\},$$

where $E(S, V \setminus S)$ is the standard notation for the size of the edge boundary between S and the rest of the graph. This connection follows from noting that the only co-boundaries on this chain are V and $\emptyset$, and that $|\delta_1 1_S|$ exactly counts the edge-boundary of S, so in particular we have:

$$\frac{|\delta_1 1_S|}{d(1_S, B^1)} = \frac{E(S, V \setminus S)}{\min\{|S|, |V \setminus S|\}}.$$

6.5 Small Set Boundary Expansion

(Co)-boundary expansion is a very strong property, and unconditional construction of bounded degree (co)-boundary expanders is still a major open question in topological high dimensional expansion. Furthermore, (co)-boundary expansion actually implies the vanishing of (co)-homology. This is an issue in and of itself in our setting, since as discussed

in Section 6.2, our CSP construction rests crucially on the associated chain complex having non-trivial co-homology. With this in mind, we introduce a new notion of high dimensional expansion which requires boundary expansion to hold *only over small sets*.

Definition 6.5.1 (Small-Set (Co)-Boundary Expansion)**.** We call $X : \mathbb{F}_2^{X(0)} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{X(1)} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{X(2)}$ a (ρ_1, ρ_2)-small-set boundary expander if the weight of small chains in $\mathbb{F}_2^{X(1)}$ is proportional to their distance from the boundary:

$$\forall f \in \mathbb{F}_2^{X(1)} \setminus B_1, |f| \leq \rho_1 |X(1)| : \frac{|\partial_1 f|}{d(f, B_1)} \geq \rho_2.$$

Similarly, X is a (ρ_1, ρ_2)-small-set co-boundary expander if:

$$\forall f \in \mathbb{F}_2^{X(1)} \setminus B^1, |f| \leq \rho_1 |X(1)| : \frac{|\delta_1 f|}{d(f, B^1)} \geq \rho_2$$

We call X a (ρ_1, ρ_2)-small-set HDX if it is both a (ρ_1, ρ_2)-small-set boundary and (ρ_1, ρ_2)-small-set co-boundary expander.

Just like standard co-boundary expansion is a higher-order analog of Cheeger's constant (edge-expansion) in graphs, small-set co-boundary expansion is the natural analog of small-set expansion on graphs. As discussed in Section 6.3.1, various strengthened notions of (unidirectional) high-dimensional small-set expansion have been considered on both simplicial [230, 141, 235, 237] and chain complexes [281, 282], but this basic generalization seems to be missing from the literature. In this work we show how small-set HDX can be transformed into explicit hard CSP instances for linear levels of Sum-of-Squares. Given the general prominence of small-set expansion throughout hardness of approximation (see e.g. [325, 255]), we expect SS-HDX may have many further applications in the area.

Before moving on, it will be useful to observe two important implications of a complex satisfying small-set (co)-boundary expansion. First, while the notion does not

require the vanishing of (co)-homology like standard boundary expansion, it does still imply a strong restriction on the structure of elements in $Z_1 \setminus B_1$: they must be large.

Lemma 6.5.2 (Small-Set (Co)-Boundary Expansion $\to$ (Co)-Systolic Distance)**.** *If X is a (ρ_1, ρ_2)-small-set boundary expander, then all chains $f \in Z_1 \setminus B_1$ are large:*

$$\min_{f \in Z_1 \setminus B_1} \{|f|\} > \rho_1 |X(1)|. \tag{6.4}$$

Similarly, if X is a (ρ_1, ρ_2)-small-set co-boundary expander, then all chains $f \in Z^1 \setminus B^1$ are large:

$$\min_{f \in Z^1 \setminus B^1} \{|f|\} > \rho_1 |X(1)|. \tag{6.5}$$

Proof. We prove the first statement only, the second follows similarly. Assume $h_1 \in Z_1 \setminus B_1$ satisfies $|h_1| \leq \rho_1 |X(1)|$. Since h_1 is a cycle, we have $\partial_1 h_1 = 0$, but then by small-set boundary expansion we have $d(h_1, B_1) = 0$, so $h_1 \in B_1$ giving the desired contradiction. □

We say complexes satisfying Equation (6.4) have *systolic distance* ρ_1, and complexes satisfying Equation (6.5) have *co-systolic distance* ρ_1. As discussed in Section 6.2, these properties were recently crucial to the construction of good qLDPC codes [315], and were also used by [119] to prove the soundness of their 3-XOR construction. Indeed it is worth noting that bounded co-systolic distance is actually enough for soundness in our construction as well, we only truly need the full power of small-set boundary expansion in one direction.

Second, we will crucially rely on a standard connection between boundary expansion and a concept known as an *isoperimetric inequality*, which relates the size of an object to the size of its boundary.[21] In particular, it is well known that boundary expansion is actually equivalent to an isoperimetric inequality for *minimal* chains (see e.g. [231]).

[21] For example the isoperimetric inequality on $\mathbb{R}^2$ says the length (boundary) of any closed curve is at least $2\sqrt{\pi}$ times the square root of its area.

Definition 6.5.3 (Minimal Chains)**.** A function $h \in \mathbb{F}_2^{X(1)}$ is called minimal if $\forall b \in B_1$, $|h + b| \geq |h|$.

A similar equivalence holds for small-set boundary expansion as well, and will be crucial for the completeness of our CSP instances: X is a small-set boundary expander if and only if small, minimal chains in X satisfy an isoperimetric inequality.

Lemma 6.5.4 (Small-Set (Co)-Boundary $\leftrightarrow$ (Co)-Isoperimetric Inequality)**.** *Let X be a (ρ_1, ρ_2)-small-set boundary expander. Then for any $h \in \mathbb{F}_2^{X(1)}$ satisfying:*

1. *h is small: $|h| \leq \rho_1 |X(1)|$*
2. *h is minimal: $\forall b \in B_1 : |h + b| \geq |h|$*

the boundary $\partial_1 h$ must be large relative to h:

$$|\partial_1 h| \geq \rho_2 |h| . \tag{6.6}$$

Conversely if Equation (6.6) *holds for any small minimal chain, then X is a (ρ_1, ρ_2)-small-set boundary expander.*

Proof. We start with the forward direction. Since $|h| \leq \rho_1 |X(1)|$ and h is minimal, by small-set boundary expansion we have that:

$$|\partial_1 h| \geq \rho_2 d(h, B) = \rho_2 \min_{b \in B} \{|h + b|\} = \rho_2 |h| .$$

The converse implication is similar. Let $h \in \mathbb{F}_2^{X(1)}$ be a small chain satisfying $|h| \leq \rho_1 |X(1)|$, and let $b \in B_1$ be the boundary minimizing $|h + b|$. Then by isoperimetry of $h + b$, we have:

$$|\partial_1 h| = |\partial_1 (h + b)| \geq \rho_2 |h + b| = \rho_2 d(h, B_1)$$

as desired. □

We note the same result holds for co-boundary expansion by the same proof. Isoperimetry (combined with good systolic distance) will be crucial for showing completeness of our XOR instances, replacing the use of Gromov's filling inequality in [119].

6.6 From Expansion to Hardness

We now show how to translate any family of expanding, bounded-degree 3-term chain complexes with non-trivial cohomology into hard instances of 3-XOR for $\Omega(n)$-levels of Sum-of-Squares.

Theorem 6.6.1. *Let $\{X_i\}$ be an explicit family of chain complexes of maximum degree $k \in \mathbb{N}$ and $\mu, \rho_1, \rho_2 \in (0,1)$ constants such that:*

1. *H^1 is non-trivial,*
2. *X has μ-co-systolic distance,*
3. *X is a (ρ_1, ρ_2)-small-set boundary expander.*

Then there exist constants $\mu_1, \mu_2 \in (0,1)$ depending only on k, μ, ρ_1, and ρ_2 and an explicit family of MAX-3-XOR instances $\{\mathcal{I}_i\}$ on n_i variables such that:

1. *Every instance is at most $(1-\mu_1)$-satisfiable,*
2. *No instance can be refuted by $\mu_2 n_i$ levels of the SoS hierarchy.*

Moreover if the complex has degree lower bounded by 3, $\{\mathcal{I}_i\}$ are instances of 3-XOR.

Theorem 6.6.1 is actually proved mainly by associating an instance of MAX-k-XOR to every complex X_i in the family. Moving to 3-XOR can then be done through standard NP-reduction arguments within the SoS hierarchy.[22] Thus the main challenge is to build hard instances of MAX-k-XOR from our complexes. We'll start by overviewing

[22]Though one must be careful that the number of variables does not blow up in the reduction, as we discuss later in the section.

our construction, which is a generalization of [119]'s 3-XOR construction from simplicial complexes to generic chain complexes.

Construction:.

It will be convenient to phrase our construction in the bipartite graph formulation discussed in Section 6.4. Recall that any chain complex $X : \mathbb{F}_2^{X(0)} \underset{\partial_1}{\overset{\delta_0}{\rightleftarrows}} \mathbb{F}_2^{X(1)} \underset{\partial_2}{\overset{\delta_1}{\rightleftarrows}} \mathbb{F}_2^{X(2)}$ may be written as a pair of bipartite graphs $B_1 = (X(0), X(1), E_1)$ and $B_2 = (X(1), X(2), E_2)$ where E_1 and E_2 are uniquely determined by the matrix representations of the boundary operators. Assuming our complex has non-trivial co-homology, let $\beta \in Z^1 \setminus B^1$.[23] Our associated CSP $\mathcal{I}_{X,\beta}$ is given by adding for every $y \in X(1)$ the constraint:

$$C_y \coloneqq \left\{ \sum_{x \in X(0):(x,y)\in E_1} x \pmod 2 = \beta(x) \right\}.$$

Since the choice of $\beta \in Z^1 \setminus B^1$ will not matter, in what follows we will drop it from the notation and just write $\mathcal{I}_X$. We make two observations about $\mathcal{I}_X$ before moving on. First, let's confirm $\mathcal{I}_X$ is indeed an instance of MAX-k-XOR.

Observation 6.6.2. If X has maximum degree k, then $\mathcal{I}_X$ is an instance of MAX-k-XOR.

Proof. This follows immediately from the chain complex having maximum degree k, as every $y \in X(1)$ then has at most k neighbors in $X(0)$ (i.e. that there are at most k elements x such that $(x, y) \in E_1$). □

Second, we observe that our instances have at most a linear number of constraints.

Observation 6.6.3. If X has maximum degree k, then $\mathcal{I}_X$ has at most $k|X(0)|$ constraints.

Proof. Since our complex is non-degenerate and degree at most k, we have that $|X(1)| \leq k|X(0)|$. $\mathcal{I}_X$ has $|X(1)|$ constraints by construction. □

[23] Note that β can be found in polynomial time by standard linear algebraic techniques.

As a result, any explicit infinite family of bounded degree chain complexes with non-trivial cohomology induces an explicit infinite family of MAX-k-XOR instances with linearly many constraints for some constant $k \in \mathbb{N}$. The main work in proving Theorem 6.6.1 therefore boils down to proving that the instances $\mathcal{I}_X$ are *sound* (at most $(1-\mu)$-satisfiable), and *complete* (look satisfiable to SoS).

Theorem 6.6.4. *Let X be a chain complex of maximum degree k and $\mu, \rho_1, \rho_2 \in (0,1)$ constants such that:*

1. *H^1 is non-trivial,*
2. *X has μ-co-systolic distance,*
3. *X is a (ρ_1, ρ_2)-small-set boundary expander.*

Then $\mathcal{I}_X$ is an instance of MAX-k-CSP on $|X(0)|$ variables satisfying:

1. *Soundness: $\mathcal{I}_X$ is at most $(1-\mu)$-satisfiable,*
2. *Completeness: $\mathcal{I}_X$ cannot be refuted by $\left(\frac{\rho_1\rho_2}{4k}|X(0)|\right)$-levels of the SoS hierarchy.*

We'll break the proof of Theorem 6.6.4 into two parts, corresponding to soundness and completeness.

Soundness:.

The soundness of our construction can be proved with no further background, and is a direct generalization of arguments in [119] from simplicial complexes to general chain complexes.

Proof of Soundess (Theorem 6.6.4). Recall that our constraints are defined by some function $\beta \in Z^1 \setminus B^1$. Let $f \in \mathbb{F}_2^{X(0)}$ be a potential assignment to variables in our instance. For any constraint $y \in X(1)$, we can check if f satisfies y by evaluating $(\beta + \delta_0 f)(y)$:

$$(\beta + \delta_0 f)(y) = \beta(y) + \sum_{(x,y)\in E_1} f(x).$$

In other words, the Hamming weight $|\beta + \delta_0 f|$ exactly corresponds to the number of violated constraints in our instance. The key is now to observe that since $\beta \in Z^1 \setminus B^1$, $\beta + \delta_0 f$ also lies in $Z^1 \setminus B^1$. Since X has μ-co-systolic distance, we have $|\beta + \delta_0 f| \geq \mu |X(1)|$, so any assignment to variables must violate at least a μ fraction of constraints as desired. $\square$

Completeness:.

Proving the completeness of Theorem 6.6.4 requires a bit more setup. As discussed in Section 6.4, we appeal to the general paradigm of Grigoriev [180], Schoenebeck [340], and Tulsiani [354] relating refutation width with Sum-of-Squares completeness. Our lower bound on the refutation width of $\mathcal{I}_X$ can be viewed in some sense as a mix of the classical strategy of Ben-Sasson and Wigderson [60] (who used traditional boundary expansion on graphs to show lower bounds against refuting Tseiten formulas) and the recent argument of [119] using Gromov's filling inequality on the Ramanujan complex. We mostly follow the exposition given in the latter.

We will consider refutations in the $\oplus$-resolution proof system, in which two linear equations $\ell_1 = b_1$ and $\ell_2 = b_2$ can be added to derive $\ell_1 \oplus \ell_2 = b_1 \oplus b_2$. By Theorem 6.4.1, it is enough to prove that any refutation of the linear equations corresponding to $\mathcal{I}_X$ has width at least $\frac{\rho_1 \rho_2}{2k} |X(0)|$, where width measures the largest number of variables appearing in any equation in the refutation. A refutation in the $\oplus$-resolution proof system can be modeled as a DAG where leaves correspond to linear equations (our XOR constraints), internal nodes have two incoming edges and correspond to the XOR of their parents, and the root derives the contradiction $0 = 1$.

To track the number of variables at each step, we follow the strategy of [119] and associate to each node v of the DAG a function $h_v \in \mathbb{F}_2^{X(1)}$ and value $b_v \in \{0, 1\}$ as follows. Since each leaf in the refutation corresponds to one of our XOR constraints, assign the leaf corresponding to $s \in X(1)$ the indicator $1_s \in \mathbb{F}_2^{X(1)}$ and value $\beta(s) \in \mathbb{F}_2$ (where we recall $\beta \in Z^1 \setminus B^1$ was the chain used to define our constraint values). The function and

value assigned to each internal node v with parents v_1, v_2 is then defined recursively to be the (mod 2) sum of its parents:

$$h_v = h_{v_1} \oplus h_{v_2}, \text{ and } \beta_v = \beta_{v_1} \oplus \beta_{v_2}.$$

Notice that by construction, $\partial_1 h_v$ exactly corresponds to the variables appearing in the linear equation at node v. This means we can bound the width of the refutation by identifying some node v in the refutation whose associated function h_v has large boundary.

To this end, following [119]'s high dimensional variant of [60]'s original technique we define the following potential function across nodes in our refutation:

$$\kappa(v) \coloneqq \min_{b \in B_1} |h_v + b| .$$

Our goal will be to find a node in the refutation whose potential is large, but still small enough that we can apply small-set boundary expansion. Namely, if we can find v such that $\frac{\rho_1}{2}|X(1)| \leq \kappa(v) \leq \rho_1|X(1)|$, then by our isoperimteric inequality for small sets (Lemma 6.5.4) we have:

$$|\partial h_v| = |\partial(h_v + b)| \geq \rho_2 |h_v + b| \geq \frac{\rho_1\rho_2}{2}|X(1)| \geq \frac{\rho_1\rho_2}{2k}|X(0)|$$

which would give the desired bound on refutation width. With this in mind, we can finally prove completeness.

Proof of completeness (Theorem 6.6.4). As discussed above, it is sufficient to prove that any refutation has width at least $\frac{\rho_1\rho_2}{2k}|X(0)|$, and that this can be done by finding a node v with potential $\frac{\rho_1}{2}|X(1)| \leq \kappa(v) \leq \rho_1|X(1)|$. The proof follows the classical strategy of [60]. Namely it is enough to show the following three properties:

1. The root node has large potential: $\kappa(r) > \rho_1|X(1)|$

2. The leaves have small potential: $\kappa(s) \leq 1$

3. The potential function is sub-additive: $\kappa(v) \leq \kappa(v_1) + \kappa(v_2)$.

As long as these hold, getting from the leaf potential of (at most) 1 to the root potential of $\kappa(r) > \rho_1|X(1)|$ requires passing through some internal node v with $\frac{\rho_1}{2}|X(1)| \leq \kappa(v) \leq \rho_1|X(1)|$ as desired.

It is left to prove the three properties, which follow from similar analysis as in [119] for the Ramanujan complex. The second and third properties are essentially immediate. Leaves are given by the indicator function of elements $s \in X(1)$, which are at most distance one from $\vec{0} \in B_1$ (the all 0s function). Sub-additivity follows from the triangle inequality. For a node v with parents v_1 and v_2, let b_1 and b_2 be boundaries minimizing $d(h_{v_1}, B_1)$ and $d(h_{v_2}, B_1)$, then we have:

$$\kappa(v_1) + \kappa(v_2) = |h_{v_1} + b_1| + |h_{v_2} + b_2| \geq |h_{v_1} + b_1 + h_{v_2} + b_2| = |h_v + b_1 + b_2| \geq \kappa(v).$$

For the first property, we argue the root node r must satisfy $h_r \in Z_1 \setminus B_1$. If this is the case we are done by the fact that our complex has good co-systolic distance by Lemma 6.5.2:

$$\kappa(h_r) = \min_{b \in B_1} |h_r + b| > \rho_1|X(1)|,$$

since any $h_r + b \in Z_1 \setminus B_1$ as well. To see that $h_r \in Z_1 \setminus B_1$, first note that since the root node in our refutation corresponds to the equation $0 = 1$, we must have $\partial_1 h_r = 0$ and therefore $h_r \in Z_1$. To complete the proof we therefore only need to show $h_r \notin B_1$, which follows from the fact that $b_r = 1$ for the root node. Namely, notice that for any node v we have $b_v = \langle \beta, h_v \rangle$ by construction (since we are just summing mod 2 over the constraints), and in particular that $\langle \beta, h_r \rangle = 1$. On the other hand, if $h_r \in B_1$, then by definition there

exists $f \in \mathbb{F}_2^{X(2)}$ such that $h_r = \partial_2 f$ and since $\delta_1 = \partial_2^T$ we have

$$\langle \beta, h_r \rangle = \langle \beta, \partial_2 f \rangle = \langle \delta_1 \beta, f \rangle = 0$$

since $\beta \in Z^1$. Thus h_r is in Z_1 but not B_1, which completes the proof. □

We are now one step away from proving Theorem 6.6.1; we just need to show how to move from a hard instance of MAX-k-XOR to a hard instance of 3-XOR. Such a reduction is fairly standard within the SoS literature, but we'll include the proof for completeness. To do so, we'll need to introduce a second way to characterize completeness of an instance for t rounds of SoS through an object called a *pseudo-expectation*. Given a set of variables $\{x_i\}_{i \in [n]}$ and $d \in \mathbb{N}$, let $\text{poly}_{\mathbb{R}}(\{x_i\}, d)$ denote the set of degree at most d polynomials in $\mathbb{R}[x_1, \ldots, x_n]$. For our purposes, it is enough to think of a degree $2t$ pseudo-expectation as an operator $\tilde{\mathbb{E}} : \text{poly}_{\mathbb{R}}(\{x_i\}, 2t) \to \mathbb{R}$ that 'pretends' to be an expectation in the following four ways:

1. Scaling: $\tilde{\mathbb{E}}[1] = 1$

2. Linearity:

$$\forall a, b \in \mathbb{R}, p(x), q(x) \in \text{poly}_{\mathbb{R}}(\{x_i\}, 2t) : \quad \tilde{\mathbb{E}}[ap(x) + bq(x)] = a\tilde{\mathbb{E}}[p(x)] + b\tilde{\mathbb{E}}[q(x)]$$

3. Positivity of Squares:

$$\forall q(x) \in \text{poly}_{\mathbb{R}}(\{x_i\}, t) : \quad \tilde{\mathbb{E}}[q(x)^2] \geq 0$$

4. Booleanity:

$$\forall j \in [n], p(x) \in \text{poly}_{\mathbb{R}}(\{x_i\}, 2t-2) : \quad \tilde{\mathbb{E}}[x_j^2 p(x)] = \tilde{\mathbb{E}}[p(x)].$$

With this in mind, let $\mathcal{I}$ be an instance of XOR on n variables $\{x_1, \ldots, x_n\}$. It will be convenient to express constraints in $C_i \in \mathcal{I}$ multiplicatively as:

$$C_i \coloneqq \left\{x_{i_1} \ldots x_{i_j} = b_i\right\}$$

where $b_i \in \{-1, 1\}$ and assignments now range over $\{-1, 1\}^n$. Let $C_i(x)$ be shorthand for the lefthand product of variables in the constraint, and $|C_i|$ denote the degree of $C_i(x)$. It turns out (see e.g. [150]) that completeness of $\mathcal{I}$ against t levels of Sum-of-Squares is equivalent to the existence of a degree $2t$ pseudo-expectation which respects every constraint $C_i \in \mathcal{I}$ in the following strong sense:

$$\forall p(x) \in \mathrm{poly}_{\mathbb{R}}(\{x\}, 2t - |C_i|) : \quad \tilde{\mathbb{E}}[C_i(x)p(x)] = b_i \tilde{\mathbb{E}}[p(x)]. \tag{6.7}$$

With this in mind, we can finally put everything together and prove Theorem 6.6.1.

Proof of Theorem 6.6.1. We'll start by constructing an explicit family of hard instances of MAX-k-XOR, then reduce to 3-XOR through the above machinery. By Theorem 6.6.4, every complex X_i in our family corresponds to an instance $\mathcal{I}_{X_i}$ of MAX-k-XOR on $n_i = |X_i(0)|$ vertices and $m_i \leq k|X(0)|$ constraints that is at most $(1 - \mu)$-satisfiable but cannot be refuted by the $\frac{\rho_1 \rho_2}{4k} n_i$-level SoS relaxation. Furthermore each instance $\mathcal{I}_{X_i}$ can be constructed in $\mathrm{poly}(n_i)$ time. This follows immediately from the fact that $\{X_i\}$ itself is explicit (and bounded degree), and that finding some $\beta \in Z_1 \setminus B_1$ can be done in polynomial time by basic linear algebra over dimension $O(n_i)$ vector spaces.

It is left to argue that we can use $\mathcal{I}_{X_i}$ to construct a corresponding instance of 3-XOR that remains hard for Sum-of-Squares. We will use the following simple approach: given a clause with more than 3 variables, split it into two clauses of about half the size whose product is the original clause. More formally, given a constraint $C_i \coloneqq \left\{x_{i_1} \ldots x_{i_j} = b_i\right\}$,

we apply the transformation:

$$C_i \to \left\{ C_i^{(0)} \coloneqq \{x_{i_1} \dots x_{i_{\lfloor j/2 \rfloor}} y_i = b_i\}, \quad C_i^{(1)} \coloneqq \{x_{i_{\lfloor j/2 \rfloor + 1}} \dots x_{i_j} y_i = 1\} \right\} \tag{6.8}$$

where y_i is a newly introduced 'dummy' variable. Given a generic instance of MAX-k-XOR $\mathcal{I}_k$, let $\Phi(\mathcal{I}_k)$ denote the CSP resulting from applying the above transformation to every constraint with more than 3 variables. We will argue that $\Phi(\mathcal{I}_k)$ has about half as many variables per clause as the original instance, but maintains soundness and completeness up to constant factors.

Claim 6.6.5. Let $\mathcal{I}_k$ be an instance of MAX-k-XOR for $k \geq 4$ on n variables and m constraints such that:

1. $\mathcal{I}_k$ is at most $(1-\mu)$-satisfiable,
2. $\mathcal{I}_k$ cannot be refuted by t rounds of Sum-of-Squares.

Then $\Phi(\mathcal{I}_k)$ is an instance of MAX-j-XOR for $j = \lceil k/2 \rceil + 1$ on at most $n + m$ variables and $2m$ constraints satisfying:

1. $\mathcal{I}_k$ is at most $(1-\mu/2)$-satisfiable,
2. $\mathcal{I}_k$ cannot be refuted by $\frac{2t}{k}$ rounds of Sum-of-Squares.

Let's first show Claim 6.6.5 completes the proof of our main theorem. Starting from our MAX-k-XOR instance $\mathcal{I}_{X_i}$, Claim 6.6.5 shows that $\Phi^{\lceil \log(k) \rceil}(\mathcal{I}_{X_i})$ is an instance of MAX-3-XOR on $O_k(n_i)$ variables that is at most $(1 - \Omega_k(\mu))$-satisfiable but cannot be refuted by $\Omega_k(n_i)$ rounds of Sum-of-Squares. This follows from the fact that the original (and all transformed instances) have $m \leq O_k(n_i)$ constraints.[24] Finally, if the original instance had no constraints with fewer than 3 variables (which occurs if the original

[24]We note that it is possible to improve the dependence on k by slightly more involved analysis, but since k is just a constant we choose to work with iterated applications of the above for simplicity of exposition.

complex has degree lower bounded by 3), $\Phi^{\lceil \log(k) \rceil}(\mathcal{I}_{X_i})$ is an instance of 3-XOR. With this in mind, it is left to prove the claim.

Proof of Claim 6.6.5. The fact that $\Phi(\mathcal{I}_k)$ is an instance of MAX-j-XOR for $j = \lceil k/2 \rceil + 1$ on at most $n + m$ variables and at most $2m$ constraints is immediate from construction. The main interest lies in proving soundness and completeness of the instance.

Soundness: Soundness of $\Phi(\mathcal{I}_k)$ follows from observing that since $y_i^2 = 1$, $C_i = C_i^{(0)} \cdot C_i^{(1)}$. Namely by the soundness of the original instance, any assignment of variables to $\Phi(\mathcal{I}_k)$ must fail at least a μ fraction of original constraints C_i (since these have no dependence on the new dummy variables). If $C_i = C_i^{(0)} \cdot C_i^{(1)}$ is violated it must be the case that either $C_i^{(0)}$ or $C_i^{(1)}$ is violated, so any assignment of variables to our transformed CSP $\Phi(\mathcal{I}_k)$ must still violate at least a $\mu/2$ fraction of its constraints.

Completeness: Given a degree $2t$ pseudo-expectation $\tilde{\mathbb{E}}$ satisfying the constraints of $\mathcal{I}_k$ (in the sense of Equation (6.7)), we must construct a new pseudo-expectation $\tilde{\mathbb{E}}_\Phi$ on the variables of $\Phi(\mathcal{I}_k)$ satisfying the transformed constraints. Given a polynomial $p(x, y) \in \mathbb{R}[\{x_i\}, \{y_j\}]$, let $p(x, 1) \in \mathbb{R}[\{x_i\}]$ denote the result of setting each y variable to 1. The idea is to observe that each dummy variable y_i in the new instance can really be thought of as a 'stand-in' for the product $b_i x_{i_1} \dots x_{i_{\lfloor j/2 \rfloor}} = b_i C_i^{(0)}(x, 1)$ in the sense that replacing each y_i with $b_i C_i^{(0)}(x, 1)$ simply returns the original instance. This suggests a natural strategy for defining our new pseudo-expectation $\tilde{\mathbb{E}}_\Phi$: just replace y_i with $b_i C_i^{(0)}(x, 1)$.[25]

Formally, this takes a bit of work. Let $S \subseteq [m]$ denote the set of indices on which we transformed our original instance, $\{y_j\}_{j \in S}$ denote the newly introduced variables, and $T : \mathbb{R}[\{x_i\}_{i \in [n]}, \{y_j\}_{j \in S}] \to \mathbb{R}[x_1, \dots, x_n]$ denote the map which independently replaces

[25]We thank Sam Hopkins for suggesting this general approach.

each occurrence of y_j with $b_j C_j^{(0)}(x,1)$ (and leaves variables in $\{x_i\}$ unchanged). It is an elementary exercise to show that T satisfies the following useful properties:

1. T is (additively) linear:

$$T(az(x,y)) = aT(z(x,y)) \text{ and } T(z_1(x,y)+z_2(x,y)) = T(z_1(x,y))+T(z_2(x,y))$$

2. T is (multiplicatively) linear:

$$T(z_1(x,y)z_2(x,y)) = T(z_1(x,y))T(z_2(x,y))$$

3. T does not substantially blow up degree:

$$Deg(T(z(x,y))) \leq \lfloor k/2 \rfloor Deg(z(x,y)).$$

With this in mind, define the value of our new pseudo-expectation on any degree at most $\frac{2t}{\lfloor k/2 \rfloor}$ polynomial $z(x,y) \in \mathbb{R}[\{x_i\}_{i\in[n]}, \{y_j\}_{j\in S}]$ as:

$$\tilde{\mathbb{E}}_\Phi[z(x,y)] \coloneqq \tilde{\mathbb{E}}[T(z(x,y))]$$

which is well-defined by the third property. It is an easy exercise to check that $\tilde{\mathbb{E}}_\Phi$ remains a pseudo-expectation, as the linearity of T ensures scaling, linearity, positivity of squares, and booleanity are all inherited from $\tilde{\mathbb{E}}$. Thus it is left to check that $\tilde{\mathbb{E}}_\Phi$ satisfies every constraint $C_i^{(j)} \in \Phi(\mathcal{I}_k)$ in the sense of Equation (6.7). To see this, first observe that

$$\begin{aligned}\tilde{\mathbb{E}}_\Phi[C_i^{(j)}(x,y)z(x,y)] &= \tilde{\mathbb{E}}[T(C_i^{(j)}(x,y)z(x,y))] \\ &= \tilde{\mathbb{E}}[T(C_i^{(j)}(x,y))T(z(x,y))].\end{aligned}$$

Taking a closer look at $T(C_i^{(j)}(x,y))$, we have by definition that:

$$T(C_i^{(j)}(x,y)) = \begin{cases} b_i C_i(x) & \text{if } j = 1 \\ b_i (C_i^{(0)}(x,1))^2 & \text{if } j = 0. \end{cases}$$

Breaking into case analysis, we then have for $j = 1$:

$$\begin{aligned} \tilde{\mathbb{E}}[T(C_i^{(1)}(x,y))T(z(x,y))] &= \tilde{\mathbb{E}}[b_i C_i(x) T(z(x,y))] \\ &= \tilde{\mathbb{E}}[T(z(x,y))] \\ &= \tilde{\mathbb{E}}_{\Phi}[z(x,y)] \end{aligned}$$

and for $j = 0$ that:

$$\begin{aligned} \tilde{\mathbb{E}}[T(C_i^{(0)}(x,y))T(z(x,y))] &= \tilde{\mathbb{E}}[b_i (C_i^{(0)}(x,1))^2 T(z(x,y))] \\ &= b_i \tilde{\mathbb{E}}[T(z(x,y))] \\ &= b_i \tilde{\mathbb{E}}_{\Phi}[z(x,y)] \end{aligned}$$

which match the form of the constraints given in Equation (6.8) as desired. □

□

6.7 Preliminaries II: Constructing SS-HDX

We now cover the tools necessary for constructing our small-set HDX, including background on basic expander graphs, left-right Cayley complexes, error correcting codes, Tanner codes, and tensor codes. We closely follow the discussion in [277] who largely cover the same background material.

6.7.1 Expander Graphs

The main building block of Leverrier and Zémor's qLDPC codes are a ubiquitous class of graphs in computer science called *spectral expanders*. Let $\mathcal{G} = (V, E)$ be an undirected Δ-regular (multi)-graph on n vertices, and define $\lambda(\mathcal{G}) := \max\{|\lambda_2|, |\lambda_n|\}$ where $\Delta = \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ are the eigenvalues of the adjacency matrix of $\mathcal{G}$. We say $\mathcal{G}$ is a λ-spectral expander if $\lambda(\mathcal{G}) \leq \lambda$, and call it *Ramanujan* if $\lambda(\mathcal{G}) \leq 2\sqrt{\Delta - 1}$, which is the optimal expansion for infinite families of fixed degree [14].

We will rely on spectral expanders for two main reasons. First, as we will discuss in the following section, infinite families of these objects are well-known not only to exist, but to be explicitly constructable (see e.g. [298]). Second, spectral expansion provides a useful proxy for edge-expansion in the sense that for any $S, T \subseteq V$, there cannot be too many edges passing between S and T. This is classically known as the *expander-mixing lemma*, and likely first appeared in [18]:

Lemma 6.7.1 (Expander mixing lemma)**.** *Let $\mathcal{G}$ be a Δ-regular graph. Then for any subset $S, T \subset V(G)$ we have*

$$|E(S,T)| \leq \frac{\Delta}{|V|}|S||T| + \lambda(\mathcal{G})\sqrt{|S||T|}.$$

When $|S|$ and $|T|$ are small compared with $|V|$, we will think of $\lambda(\mathcal{G})\sqrt{|S||T|}$ as the main term and $\frac{\Delta}{|V|}|S||T|$ as the error term (we note this is the opposite of how the lemma is often applied).

It will also be important for us that the expander mixing lemma holds for *double covers* of a spectral expanders with a small modification. The double cover $\mathcal{G}' = (V', E')$ of a graph $\mathcal{G} = (V, E)$ has vertex set $V' = V_0 \cup V_1$, for $V_0 = V \times \{0\}$ and $V_1 = V \times \{1\}$, and edge set $E' = \{\{(v, 0), (w, 1)\} : v, w \in V, \{v, w\} \in E\}$. The expander mixing lemma applies for double covered graphs when $S \subset V_0, T \subset V_1$.

Lemma 6.7.2 (Expander mixing lemma for double covered graph)**.** *Let $\mathcal{G}$ be a Δ-regular graph and $\mathcal{G}'$ be its double cover. Then for any subset $S \subset V_0(\mathcal{G}'), T \subset V_1(\mathcal{G}')$ we have*

$$|E(S,T)| \leq \frac{\Delta}{|V(\mathcal{G})|}|S||T| + \lambda(\mathcal{G})\sqrt{|S||T|}.$$

This can be shown easily by projecting S and T back to the original graph.

6.7.2 Left-Right Cayley Complexes

While expansion is a useful property in its own right, our arguments require higher dimensional structure. The key lies in an object called the *left-right Cayley complex* introduced in [117] to build c3-LTCs. A left-right Cayley complex is determined by a group G and two sets of generators $A = A^{-1}$ and $B = B^{-1}$. The complex consists of vertices, A-edges, B-edges, and squares as follows:

- The vertices are $V^0 = G$.
- The A-edges are E^0_A and the B-edges are E^0_B where

$$E^0_A = \{\{g, ag\} : g \in G, a \in A\}, E^0_B = \{\{g, gb\} : g \in G, b \in B\}.$$

- The squares are

$$F^0 = \{\{g, ag, gb, agb\} : g \in G, a \in A, b \in B\}.$$

The main criterion for choosing G, A, and B is to ensure the Cayley graphs $Cay(G, A)$ and $Cay(G, B)$ are good expanders, and in particular are Ramanujan. Besides this, for simplicity we further assume two technical conditions as in [117]: that $|A| =$

$|B| = \Delta$, and the so-called *total no-conjugacy* condition

$$\forall a \in A, b \in B, g \in G, ag \neq gb.$$

The total no-conjugacy condition ensures squares are non-degenerate (contain exactly 4 distinct vertices), and that each vertex is incident to exactly k^2 squares [117, Claim 3.7]. Leveraging classical results of Morgenstern [298] and Lubotzky, Samuels, and Vishne [290], [117] show that explicit families of left-right Cayley complexes exist for infinitely many degrees.

Theorem 6.7.3 ([117, Claim 6.7])**.** *There exists an infinite sequence of degrees $\Delta = q + 1$ (where q is an odd prime power) such that for each fixed Δ there exists an explicit infinite family of left-right Cayley complexes with $G_i = \mathrm{PSL}_2(q^i)$ and generator sets A_i and B_i such that $|A_i| = |B_i| = \Delta$, $Cay(G_i, A_i)$ and $Cay(G_i, B_i)$ are Ramanujan, and A_i,B_i satisfy the total no-conjugacy condition.*

As in [277], we will use the *double cover* of the left-right Cayley complex, defined as:

- The vertices are $V = V_0 \cup V_1$ where $V_0 = G \times \{0\}$ and $V_1 = G \times \{1\}$.
- The A-edges are E_A and the B-edges are E_B where

$$E_A = \{\{(g, 0), (ag, 1)\} : g \in G, a \in A\}, E_B = \{\{(g, 0), (gb, 1)\} : g \in G, b \in B\}.$$

- The squares are

$$F = \{\{(g, 0), (ag, 1), (gb, 1), (agb, 0)\} : g \in G, a \in A, b \in B\}.$$

Note that every square in the original left-right Cayley complex corresponds to two squares

in the double cover, and therefore that the double cover has a total of $\frac{\Delta^2|G|}{2}$ squares. Since we will only use the double cover in our arguments, from now on the term "square" will *always refer to these double-covered squares*, not the squares in the original Cayley complex.

Following [277]'s notation, we will mainly think of the double-covered complex as represented by the following graphs. First, we'll define a graph that captures the vertices and edge-structure of the Cayley complex: $\mathcal{G}^{\cup} = (V, E_A \cup E_B)$. Second, we'll define graphs[26] $\mathcal{G}_0^{\square} = (V_0, E_0^{\square})$ and $\mathcal{G}_1^{\square} = (V_1, E_1^{\square})$ capturing squares in the double cover, where

$$E_i^{\square} = \{\{(g,i),(agb,i)\} : g \in G, a \in A, b \in B\}$$

for $i \in \{0,1\}$. Notice that the edges in these graphs have a one-to-one correspondence with the double-covered squares, namely that $E_i^{\square} \cong F$ for $i = 0,1$ through the following identifications:

$$\{(g,0),(agb,0)\} \leftrightarrow \{(g,0),(ag,1),(gb,1),(agb,0)\}$$

and

$$\{(g,1),(agb,1)\} \leftrightarrow \{(g,1),(ag,0),(gb,0),(agb,1)\}.$$

These identifications will be particularly important in the proof of small-set (co)-boundary expansion as we move between the squares of our complex and their associated graph representations.

Finally, it will be important to observe that these graphs inherit the spectral properties of $Cay(G,A)$ and $Cay(G,B)$. Namely that when the latter are Ramanujan, $\mathcal{G}^{\cup}, \mathcal{G}_0^{\square}, \mathcal{G}_1^{\square}$ are also very good expanders.

Lemma 6.7.4 ([277, Lemma 4])**.** *If $Cay(G,A), Cay(G,B)$ are Ramanujan graphs, then $\lambda(\mathcal{G}_0^{\square}) \leq 4\Delta$, $\lambda(\mathcal{G}_1^{\square}) \leq 4\Delta$, and $\mathcal{G}^{\cup}$ is the double cover of a $4\sqrt{\Delta}$-spectral expander.*

We note this is not exactly the statement given in [277], but the proof is the same.

[26]We note these may technically be multi-graphs as in [277], but this has no effect on our arguments.

6.7.3 Error Correcting Codes

A classical (n, k, d)-error correcting (erasure) code is a method for encoding a string of k classical bits into $n > k$ classical bits such that one can recover the original string even when up to $d-1$ bits of the encoded string are erased. More formally, we will consider the standard setting of *linear codes*, where the encoded space is a linear subspace $\mathcal{C} \subset \mathbb{F}_2^n$. Here n is the *length* of the code, $k \coloneqq \dim(\mathcal{C})$ is its *dimension*, and the minimum weight of any element (also called codeword) of $\mathcal{C}$, $d \coloneqq \min_{c \in \mathcal{C}}\{|c|\}$, is called its *distance*.[27] One can check that in a linear code of distance d, it is indeed possible to uniquely correct up to $d-1$ errors. Finally, the ratio $r \coloneqq \frac{k}{n}$ is called the *rate* of the code, and measures the overhead from the original to encoded space. We will typically be interested in families of codes that have constant rate and linear distance.

One of the main reasons to use linear codes is that there are nice linear algebraic ways of describing the objects. In particular, the linear subspace (code) $\mathcal{C}$ is typically described either by a *parity-check* matrix, or a *generator* matrix. In particular, one can always find a parity-check matrix $M : \mathbb{F}_2^n \to \mathbb{F}_2^{n-k}$ whose kernel is the code in question ($\mathcal{C} := \ker M \subset \mathbb{F}_2^n$), and likewise a generator matrix $M' : \mathbb{F}_2^k \to \mathbb{F}_2^n$ whose image gives the code ($\mathcal{C} := \operatorname{im} M' \subset \mathbb{F}_2^n$). When clear from context, we sometime abuse notation and write $\mathcal{C}$ to mean the parity check matrix of $\mathcal{C}$.

6.7.4 Tanner Codes

The *Tanner construction* (or *tanner code*) [351] is a classical strategy in coding theory to build a linear code out of a 'large' regular graph and a 'small' *local* code that sits on the neighborhood of each vertex. Crucially, when the underlying graph is an expander, it is often the case that the Tanner code inherits desirable properties from the small code.

More formally, let $\mathcal{G} = (V, E)$ be a Δ-regular graph and $E(v)$ denote the set of

[27] We note that this is similar to the distance operator $d(\cdot, \cdot)$ used to define co-boundary expansion. Indeed the distance of a code $\mathcal{C}$ is just $d(\emptyset, \mathcal{C})$. We will abuse notation slightly to match standard coding theory notation and write this as $d(\mathcal{C})$ throughout.

edges incident to any $v \in V$. Assume an identification of $\mathbb{F}_2^{E(v)}$ with $\mathbb{F}_2^{\Delta}$ for each $v \in V$, which we call the *local view* of v. Given a local code C_0 with length Δ, the Tanner code $T(\mathcal{G}, C_0) \subset \mathbb{F}_2^E$ is given by

$$\{c \in \mathbb{F}_2^E : \forall v \in V, c|_{E(v)} \in C_0\},$$

where $c|_{E(v)} \in \mathbb{F}_2^{\Delta}$ is the vector formed by the values of c on the local view of v.

It will be convenient for us to view the Tanner construction through its parity check matrix, which will make up the co-boundary operators of our chain complex. If our local code C_0 has parity check matrix M_0 and rate r_0, the parity check matrix of the Tanner code $T(\mathcal{G}, C)$ is given by the composition:

$$\mathbb{F}_2^E \to \mathbb{F}_2^{V \times \Delta} \to \mathbb{F}_2^{V \times (1-r_0)\Delta}$$

where the first map copies the value on the edge to each local view of the vertices, and the second map applies M_0 to each local view independently for each vertex. We will sometimes refer to this parity check matrix as the *Tanner map*.

6.7.5 Robust Tensor Codes Against Puncture

The properties of our Tanner maps are highly dependent on the local code used to instantiate them. Following [277], we use a special type of local code called a *tensor code*. We closely follow the discussion of these objects given in [277].

Recall that the generators of our left-right Cayley complex A and B have size Δ. We will consider codes on $\mathbb{F}_2^{A \times B}$ with tensor product structures. Namely, given two linear codes $C_A \subset \mathbb{F}_2^A, C_B \subset \mathbb{F}_2^B$, we define the *tensor code* $C_A \otimes C_B$ to be the set of $\Delta \times \Delta$ matrices M where each column vector $(M_{ab})_{a \in A}$ belongs to C_A and each row vector $(M_{ab})_{b \in B}$ belongs to C_B. We define the *dual tensor code* to be the sum $C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B$,

where $C_A \otimes \mathbb{F}_2^B$ are the $\Delta \times \Delta$ matrices whose columns belong to C_A, and $\mathbb{F}_2^A \otimes C_B$ are the $\Delta \times \Delta$ matrices whose rows belong to C_B. The following claims about the dimension and distance of these codes are standard and easy to verify:

1. $\dim(C_A \otimes C_B) = \dim(C_A)\dim(C_B)$
2. $d(C_A \otimes C_B) = d(C_A)d(C_B)$
3. $\dim(C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B) = \Delta\dim(C_A) + \Delta\dim(C_B) - \dim(C_A)\dim(C_B)$
4. $d(C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B) = \min(d(C_A), d(C_B))$.

To ensure our Tanner maps have the right properties, we will actually require our local tensor codes to have a stronger property called robustness. One can think of robustness as a generalization of distance of usual linear codes to the context of tensor codes, or as we will soon see, as a sort of robust testability property.

Definition 6.7.5 (Robust [277, Definition 5])**.** Let $C_A \subset \mathbb{F}_2^A, C_B \subset \mathbb{F}_2^B$ be codes of length Δ of distance d_A and d_B respectively. We say the dual tensor code $C = C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B$ is w-robust if for every codeword $c \in C$ with Hamming weight $|c| < w$, there exist $A' \subset A, B' \subset B, |A'| \leq |c|/d_B, |B'| \leq |c|/d_A$, such that $c_{ab} = 0$ for any $a \notin A'$ and $b \notin B'$.

Leverrier and Zémor [277] prove that robust tensor codes satisfy a useful small-set robust testability property.

Lemma 6.7.6 ([277], Proposition 6)**.** *Let $C_A \subset \mathbb{F}_2^A, C_B \subset \mathbb{F}_2^B$ be codes of length Δ of distance d_A and d_B respectively. If the dual tensor code $C = C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B$ is w-robust with $w \leq d_A d_B/2$, then any word x close to both the column and row code is also close to the tensor code. More explicitly, if $d(x, C_A \otimes \mathbb{F}_2^B) + d(x, \mathbb{F}_2^A \otimes C_B) < w$ then:*

$$d(x, C_A \otimes C_B) \leq \frac{3}{2}\left(d(x, C_A \otimes \mathbb{F}_2^B) + d(x, \mathbb{F}_2^A \otimes C_B)\right).$$

In fact, [277] need a slightly stronger condition than just robustness of the code: it needs to remain robust even after the removal of a small set of rows and columns. Conceptually, this is similar to the idea of smooth codes [131] where the code maintains nice properties even after the removal of a small number of variables or checks. Given a code $C_A \subset \mathbb{F}_2^A$ and $A' \subset A$, let $C_{A'} \subset \mathbb{F}_2^{A'}$ denote the *puncture code* which is the restriction of all codewords in C_A to the coordinates in A' (more precisely, $C_{A'} = \{(c_a)_{a \in A'} : (c_a)_{a \in A} \in C_A\}$).

Definition 6.7.7 (Robust against puncture [277, Definition 7])**.** Given linear codes $C_A \subset \mathbb{F}_2^A, C_B \subset \mathbb{F}_2^B$, we say the dual tensor code $C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B$ is w-robust with p-resistance to puncture if for any $w' \leq p$ and $A' \subset A$ and $B' \subset B$ such that $|A'| = |B'| = \Delta - w'$, the dual tensor code $C_{A'} \otimes \mathbb{F}_2^{B'} + \mathbb{F}_2^{A'} \otimes C_{B'}$ is w-robust.

Extending prior work of [315], [277] show random tensor codes are robust against puncture.

Theorem 6.7.8 ([277, Theorem 8])**.** *Let $0 < r_A < 1$ and $0 < r_B < 1$. Let $0 < \epsilon < 1/2$ and $1/2 + \epsilon < \gamma < 1$. Let C_A be a random code obtained from a random uniform $r_A\Delta \times \Delta$ generator matrix, and let C_B be a random code obtained from a random uniform $(1 - r_B)\Delta \times \Delta$ parity-check matrix. With probability tending to 1 when Δ goes to infinity, the dual tensor code*

$$C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B$$

is $\Delta^{3/2-\epsilon}$-robust with Δ^γ-resistance to puncturing.

Because the dual of a random code is again a random code, this implies both $C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B$ and $C_A^\perp \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B^\perp$ are robust against puncture with high probability.

Corollary 6.7.9 ([277, Theorem 17])**.** *Fix $r \in (0, 1/2)$, $\epsilon \in (0, 1/2)$, $\gamma \in (1/2 + \epsilon, 1)$ and $\delta > 0$ satisfying $-\delta \log \delta - (1 - \delta) \log(1 - \delta) < r$. when Δ is large enough, there exist codes C_A and C_B of length Δ such that*

1. $\dim C_A = \lfloor r\Delta \rfloor$ *and* $\dim C_B = \Delta - \dim C_A$

2. *The distances of* $C_A, C_B, C_A^\perp, C_B^\perp$ *are all at least* $\delta\Delta$

3. *Both dual tensor codes* $C_0^\perp = (C_A \otimes C_B)^\perp$ *and* $C_1^\perp = (C_A^\perp \otimes C_B^\perp)^\perp$ *are* $\Delta^{3/2-\epsilon}$*-robust with* Δ^γ*-resistance to puncturing*

4. $C_A, C_B, C_A^\perp$*, and* $C_B^\perp$ *have generator matrices where every row and column have at least two ones.*

We note that this is not exactly the statement of [277, Theorem 17], who prove the first three conditions occur with probability going to 1 as Δ becomes large when C_A and C_B are generated as in Theorem 6.7.8. The fourth item is not included in [277], but also occurs under this distribution with high probability by fairly standard arguments. We give the proof in the appendix for completeness.

6.8 Constructing Small-Set HDX

We are finally ready to construct a family of 3-term chain complexes with small-set boundary and co-boundary expansion.

Theorem 6.8.1. *There exists an explicit infinite family of chain complexes* $\{X_i\}$ *and constants* $d \in \mathbb{N}$ *and* $\rho_1, \rho_2 \in (0, 1)$ *such that each* X_i *satisfies:*

1. X_i *has maximum degree* d *and minimum degree at least* 3

2. X_i *has non-trivial co-homology* H^1

3. X_i *is a* (ρ_1, ρ_2)*-small-set HDX.*

Combined with Theorem 6.6.1 which transforms SS-HDX into hard instances of 3-XOR, this completes the proof of our main theorem.

Proof of Theorem 6.1.1. Theorem 6.6.1 gives the desired explicit family of 3-XOR instances as long as it is provided an explicit family of chain complexes with bounded maximum degree, minimum degree at least 3, non-trivial co-homology, and which are (ρ_1, ρ_2)-small-set boundary expanders with μ-co-systolic distance for some set of constants $\mu, \rho_1, \rho_2 \in (0, 1)$. Since any (ρ_1, ρ_2)-small-set co-boundary expander has ρ_1-co-systolic distance (Lemma 6.5.2), Theorem 6.8.1 provides an explicit family of chain complexes matching these conditions with $\mu = \rho_1$. □

As discussed, Theorem 6.8.1 is proved via Leverrier and Zémor's [277] recent construction of good qLPDC codes. They show the associated 3-term chain complex has linear systolic and co-systolic distance. Our contribution is to observe that the same construction actually satisfies the stronger small-set boundary and co-boundary expansion conditions. We note that while we only show this property for Leverrier and Zémor's [277] simplified construction, similar arguments likely hold for Panteleev and Kalachev's [315] original good qLDPC codes as well.

Construction:.

We first describe Leverrier and Zémor's construction, which is based upon Tanner maps (parity-check matrices of Tanner codes). To start, we'll first need to describe the underlying graphs and local codes of these maps. Recall the explicit family of left-right Cayley complexes promised by Theorem 6.7.3 and for any fixed complex X_i in the family let the group $G = G_i$ and generator sets $A = A_i, B = B_i$ be as in the theorem. The graphs underlying our Tanner maps will be the 'square graphs' $\mathcal{G}_0^{\square}$ and $\mathcal{G}_1^{\square}$, which we recall have

- Vertices $V_i = G \times \{i\}$,
- Edges $E_i^{\square} = \{\{(g, i), (agb, i)\} : g \in G, a \in A, b \in B\}$

for $i \in \{0, 1\}$ respectively. It bears repeating that edges in these graphs are in one-to-one correspondence with squares of the double covered Cayley complex via the following

identifications:

$$\{(g,i),(agb,i)\} \leftrightarrow \{(g,i),(ag,1-i),(gb,1-i),(agb,i)\}.$$

We will frequently refer to edges in $\mathcal{G}_i^{\square}$ as squares due to this connection.

Since the square graphs $\mathcal{G}_i^{\square}$ are Δ^2-regular, we can define a Tanner map by combining them with any length Δ^2 local code. This role will be played by the robust dual tensor codes promised by Corollary 6.7.9. Namely, letting $C_A : \mathbb{F}_2^{\Delta} \to \mathbb{F}_2^{r\Delta}$ and $C_B : \mathbb{F}_2^{\Delta} \to \mathbb{F}_2^{(1-r)\Delta}$ be as in Corollary 6.7.9 for some choice of r, ε, and γ, our local codes will be $C_0^{\perp}$ and $C_1^{\perp}$ where $C_0 = C_A \otimes C_B$ and $C_1 = C_A^{\perp} \otimes C_B^{\perp}$.[28]

Combining these graphs and local codes gives the Tanner maps $\mathcal{C}_0 = T(\mathcal{G}_0^{\square}, C_0^{\perp}) : \mathbb{F}_2^n \to \mathbb{F}_2^m$ and $\mathcal{C}_1 = T(\mathcal{G}_1^{\square}, C_1^{\perp}) : \mathbb{F}_2^n \to \mathbb{F}_2^m$, where $n = |F| = \Delta^2|G|/2$ is the number of squares, and $m = r(1-r)\Delta^2|V_0| = r(1-r)\Delta^2|G|$ comes from the fact that the both dual tensor codes have dimension $(1 - r(1-r))\Delta^2$. Associating the edges of $\mathcal{G}_i^{\square}$ with squares in the discussed manner, one can check that $\mathcal{C}_1\mathcal{C}_0^T = 0$ (see [277, Section 4.1]) and therefore that these maps define a chain complex:

$$X : \mathbb{F}_2^m \xrightarrow{\delta_0 := \mathcal{C}_0^T} \mathbb{F}_2^n \xrightarrow{\delta_1 := \mathcal{C}_1} \mathbb{F}_2^m. \tag{6.9}$$

Moreover, this process gives an explicit family of chain complexes $\{X_i\}$ by choosing G_i, A_i, and B_i as in the explicit family of left-right Cayley complexes promised by Theorem 6.7.3, and computing C_A, C_B with the desired properties by brute force search over all pairs of length Δ codes C_A, C_B of dimensions $r\Delta$ and $(1-r)\Delta$ respectively.[29]

This completes the construction. We now move to showing that X has the three desired properties: bounded-degree, non-trivial co-homology, and small-set (co)-boundary

[28] Note we are assuming for simplicity that $r\Delta$ and $(1-r)\Delta$ are integer valued, but these can be replaced with $\lfloor r\Delta \rfloor$ and $\Delta - \lfloor r\Delta \rfloor$ without substantially affecting the proof (see [277]).

[29] Note that since Δ is a constant with respect to our infinite family, brute force search only requires $O(1)$ time here.

expansion.

X has (upper) bounded-degree:.

By definition X is bounded-degree if and only if the parity-check matrices of our two Tanner codes have a bounded number of ones in every row and column. By the nature of the Tanner code construction the support of any row or column is at most twice the degree of the underlying graph. Since our graphs are of degree Δ^2 (a constant with respect to the family), the resulting complex is bounded-degree as desired.

X has (lower) bounded-degree:.

Recall we are promised that $C_A, C_B, C_A^\perp$, and $C_B^\perp$ have generator matrices where every row and column have at least two ones. This implies that the tensor codes $C_A \otimes C_B$ and $C_A^\perp \otimes C_B^\perp$ can be taken to have generator matrices with at least four ones in each row and column. Since these correspond to the parity check matrices of $C_0^\perp = (C_A \otimes C_B)^\perp$ and $C_1^\perp = (C_A^\perp \otimes C_B^\perp)^\perp$ respectively, it can be easily checked that the parity check matrices of the associated Tanner codes $T(\mathcal{G}_0^\square, C_0^\perp)$ and $T(\mathcal{G}_1^\square, C_1^\perp)$ also have at least four ones in every row and column.

H^1 is non-trivial:.

This follows immediately from dimensionality arguments. In particular, notice that $\dim Z^1 \geq n - m$, whereas $\dim B^1 \leq m$. As a result we have $\dim H^1 \geq n - 2m = (1/2 - 2r(1-r))\Delta^2|G|$ which is > 0 whenever $r \neq 1/2$.

X is a small-set (co)-boundary expander:.

It is left to show our complexes are small-set (co)-boundary expanders. In what follows we show the co-boundary expansion case. Since the construction is symmetric, a similar proof gives small-set boundary expansion. For convenience, we first re-formulate the problem as the following technical theorem. Note that this is the analog of [277, Theorem 1] where co-systolic distance is replaced with small-set co-boundary expansion.

We follow their notation when possible for consistency.

Theorem 6.8.2. *Fix $\epsilon \in (0, 1/2)$, $\gamma \in (1/2 + \epsilon, 1)$ and $\delta > 0$. For any fixed large enough Δ, if the linear codes C_A and C_B have minimum distance at least $\delta\Delta$ and if the dual tensor code $C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B = C_1^\perp$ is w-robust with p-resistance to puncturing for $w = \Delta^{3/2-\epsilon/2}$ and $p = \Delta^\gamma$,[30] then the chain complex in Equation (6.9):*

$$X : \mathbb{F}_2^m \xrightarrow{\delta_0 := \mathcal{C}_0^T} \mathbb{F}_2^n \xrightarrow{\delta_1 := \mathcal{C}_1} \mathbb{F}_2^m,$$

satisfies the following isoperimetric inequality for small, minimal chains:

$$\forall x \in \mathbb{F}_2^n \text{ s.t. } x \text{ is minimal and } |x| \leq \rho_1 n : |\delta_1 x| \geq \rho_2 |x|,$$

where $\rho_1 = \frac{\delta}{6\Delta^{3/2+\epsilon}}$, $\rho_2 = \frac{56}{\Delta^{3-2\epsilon}}$.

Recall that this isoperimetric condition is equivalent to (ρ_1, ρ_2)-small-set coboundary expansion (Lemma 6.5.4), so this indeed proves the desired property. The proof of Theorem 6.8.2 closely follows the analogous proof in [277] for systolic distance. The main difference is that we must track an additional set of elements consisting of vertices in $\mathcal{G}_0^\square$ corresponding to violated constraints. Since [277] only need to consider $x \in \mathbb{F}_2^n$ that are true codewords, this is not a relevant consideration in their result. We note that throughout we set our coefficients to match those in [277] for ease of comparison.

Proof of Theorem 6.8.2. We assume $x \neq 0$, as the theorem holds trivially otherwise. We proceed by contradiction. Assuming $|\delta_1 x| < \rho_2 |x|$, we will show there exists $y \in B_1$ such that $|x + y| < |x|$, contradicting minimality of x.

We first lay out some relevant notation. Thinking of x as a subset of $E_1^\square$ (the edge set of $\mathcal{G}_1^\square$), we will consider the edge-induced subgraph $\mathcal{G}_{1,x}^\square \subset \mathcal{G}_1^\square$ and denote its vertex set

[30] We note that the value of w here is slightly different than in [277]. This corrects a small error in the application of robust testability (Lemma 6.7.6) in the original work.

by $S \subset V_1$. Recall that each vertex $(g,1) \in V_1$ has a corresponding local view made up of $|A| \cdot |B|$ incident squares, which we'll denote by:

$$\ell(g,1) \coloneqq \{\{(g,1),(ag,0),(gb,0),(agb,1)\} : a \in A, b \in B\}.$$

Thinking of x now as a set of squares, let $x|_{\ell(g,1)} \in \mathbb{F}_2^{A \times B}$ denote the restriction of x to the local view of $(g,1)$, and recall that x is a co-cycle exactly when these local views correspond to codewords in $C_1^\perp = C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B$.

Since x is arbitrary in our setting (unlike [277] who only consider co-cycles) we will partition the vertices of our induced subgraph into three parts: $S = S_v \cup S_n \cup S_e$. First, let $S_v \subset S$ denote the set of *violated vertices* $(g,1) \in V_1$ whose local views $x|_{\ell(g,1)}$ do *not* form codewords in $C_1^\perp$. Following [277], we split the remaining vertices in $S \setminus S_v$ into two parts based upon their degree in the induced subgraph $\mathcal{G}_{1,x}^\square$: the *normal vertices* S_n with degree less than $w_2 := \Delta^{3/2-\epsilon}$, and the *exceptional vertices* S_e with degree at least w_2. The intuition behind this strategy is that because $C_1^\perp$ is $(w > w_2)$-robust, the codewords associated to vertices in S_n have particularly nice structure: they are zero outside of a small set of at most $w_2/(\delta\Delta)$ rows and columns. This implies that any column (respectively row) is close to a codeword in C_A (respectively C_B) which will eventually help us apply small-set robust testability (Lemma 6.7.6) to prove x is close to a co-boundary (and is therefore non-minimal).

To find such a co-boundary, we'll first need to look to the other side of the complex. Broadly speaking, the idea (which is the same as in [277]) is to find a vertex $v \subset V_0$ whose local view shares many (heavy) rows and columns with local views of vertices in S_n. One can then apply robustness to see that the value of x on this local view is close to a codeword in $C_A \otimes C_B$ which can easily be translated to the desired co-boundary.

More formally, let $E_x \subset \mathcal{G}^\cup$ denote the set of edges incident to the squares in x,[31]

[31] In particular for any square $\{(g,0),(ag,1),(gb,1),(agb,0)\} \in x$, add its four edges $\{(g,0),(ag,1)\}$,

and call an edge *heavy* if it is incident to at least $\delta\Delta - \Delta^{1/2-\epsilon}/\delta$ squares in x. We will consider the set of vertices $T \subset V_0$ which are adjacent to $S_n \subset V_1$ through a heavy edge in the graph $\mathcal{G}^{\cup}$. Given $v \in T$, note that every heavy edge with an element in S_n corresponds to a row or column that is shared in their local view (and is therefore close to a codeword of C_B or C_A respectively). The goal is therefore to show that there exists a vertex in T that is adjacent to many elements in S_n through heavy edges, while simultaneously adjacent to few 'bad' vertices in S_e and S_v. This will allow us to apply robustness against puncture to find a co-boundary that reduces the weight of x. We formalize these statements below in the following two claims.

Claim 6.8.3 (Modification of [277, Claim 13]). There exist $h_1 \geq \Omega(\Delta)$, $d_1 \leq O(\Delta^{1/2+\epsilon})$, and $v \in T$ such that v is incident to at least h_1 heavy edges and adjacent to at most d_1 vertices of $S_e \cup S_v$.

Claim 6.8.4 (Summary of paragraph following [277, Claim 13]). For all sufficiently large[32] Δ, if there exists a vertex $v \in V_0$ incident to $h_1 \geq \Omega(\Delta)$ heavy edges and at most $d_1 \leq O(\Delta^{1/2+\epsilon})$ vertices of $S_e \cup S_v$, then we can find a vector $y \in B_1$ such that $|x + y| < |x|$.

Together, Claim 6.8.3 and Claim 6.8.4 complete the proof of Theorem 6.8.2, as they promise the existence of some $y \in B_1$ such that $|x + y| < |x|$, violating minimality of x. Thus it is left to prove the claims. While Claim 6.8.4 follows largely from arguments in [277], it is helpful to present first to motivate the more technical proof of Claim 6.8.3.

Proof of Claim 6.8.4. Recall we are given an element $v \in V_0$ which is incident to at least $h_1 \geq \Omega(\Delta)$ heavy edges and adjacent to at most $d_1 \leq O(\Delta^{1/2+\varepsilon})$ vertices in $S_e \cup S_v$. We consider the local view of x around v (considered as an element of $\mathcal{G}_0^{\square}$), denoted $x_v \in \mathbb{F}_2^{A\times B}$ here for notational simplicity. Because at most d_1 vertices adjacent to v in $\mathcal{G}^{\cup}$ are exceptional or violated (as considered in $\mathcal{G}_{1,x}^{\square}$), one can find $A' \subset A, B' \subset B$ with

$\{(g,0),(bg,1)\}$, $\{(agb,0),(ag,1)\}$, and $\{(agb,0),(gb,1)\}$ to E_x.

[32] Here we mean in terms of r, ε, and γ, so Δ remains constant with respect to the infinite family.

$|A'| = |B'| \geq \Delta - d_1$, such that A' and B' are indexed by either normal vertices, or vertices not in S. Furthermore, since $d_1 \leq \Delta^\gamma$ for large enough Δ, we also have by robustness to puncture that the restricted dual tensor code $(C_1^\perp)' \coloneqq C_{A'} \otimes \mathbb{F}_2^{B'} + \mathbb{F}_2^{A'} \otimes C_{B'}$ is w-robust.

Let x'_v be the restriction of x_v in $A' \times B'$. Recall each column (row) of the local view of a normal vertex is at most $w_2/(\delta\Delta)$ away from a codeword by w-robustness. Then since each column (row) of x'_v is indexed by either a normal vertex or a vertex whose local view is all zero (i.e. not in S), every column (respectively row) of x'_v is at most $w_2/(\delta\Delta) = \Delta^{1/2-\epsilon}/\delta$ away from a codeword in $C_{A'}$ (respectively $C_{B'}$). Since there are at most Δ rows and columns, this means that x'_v is at most $\Delta^{3/2-\varepsilon}/\delta$ away from either $C_{A'} \otimes \mathbb{F}_2^{B'}$ or $\mathbb{F}_2^{A'} \otimes C_{B'}$, and moreover that:

$$d(x'_v, C_{A'} \otimes \mathbb{F}_2^{B'}) + d(x'_v, \mathbb{F}_2^{A'} \otimes C_{B'}) \leq 2\Delta^{3/2-\varepsilon}/\delta \leq w$$

for sufficiently large Δ. Because $(C_1^\perp)'$ is w-robust, we can apply small-set robust testability (Lemma 6.7.6) to infer that x'_v is close to some codeword $c' \in C_{A'} \otimes C_{B'}$:

$$d(x'_v, c') \leq \frac{3}{2}\left(d(x'_v, C_{A'} \otimes \mathbb{F}_2^{B'}) + d(x'_v, \mathbb{F}_2^{A'} \otimes C_{B'})\right) \leq 3\frac{\Delta^{3/2-\epsilon}}{\delta}.$$

Finally, since the total number of punctured rows and columns is less than the code distance for large enough Δ, we can extend c' uniquely to a codeword $c \in C_A \otimes C_B$. Taking into account the rows and columns added in this process, the distance from x_v to c then becomes at most $d(x_v, c) \leq d(x'_v, c') + 2d_1\Delta \leq O(\Delta^{3/2+\epsilon}) < o(\Delta^2)$ since $\varepsilon < 1/2$.

On the other hand, because v is incident to $\Omega(\Delta)$ heavy edges, the weight $|x_v| = \Theta(\Delta^2)$. Thus for large enough Δ, it must be the case that flipping c strictly reduces the weight of x. More precisely, set y to be c on the local view x_v and 0 elsewhere, then we have $|x + y| < |x|$. Since $c \in C_A \otimes C_B = C_0$, $y \in B^1$ is indeed a co-boundary which completes the proof. $\square$

The only thing left is to show that our main technical claim actually holds, the existence of a vertex with many heavy edges that is adjacent to few violated or exceptional vertices. The proof technique is similar to that of [277, Claim 13], and mostly boils down to proving that S_v and S_e are small compared to S_n.

Proof of Claim 6.8.3. We split the proof into the following three claims. First, we claim T is non-empty.

Claim 6.8.5. $|T| > 0$.

With this in mind, let $\alpha, \beta = \Theta(1)$ be constants to be set later in the proof. Following [277], we claim that a reasonable fraction of T is incident to many heavy edges:

Claim 6.8.6 ([277, Claim 12]). At least an $\alpha/2$ fraction of vertices in T are incident to at least $h_1 = \alpha\Delta$ heavy edges,

and further that at most some smaller fraction is adjacent to greater than d_1 violated and exceptional vertices:

Claim 6.8.7 ([277, Paragraph between Claim 4.10 and Claim 4.11]). At most an $\alpha/4$ fraction of vertices in T are incident to more than $d_1 = \frac{4\beta}{\alpha}\Delta^{1/2+\epsilon}$ vertices of $S_e \cup S_v$.

Combining these claims implies at least an $\alpha/4$ fraction of vertices satisfy the requirements of Claim 6.8.3. Since T is non-empty, this must apply to at least one $v \in T$ which gives the desired result.

The key to proving all three claims lies in showing that the number of vertices in $S_e \cup S_v$ is small compared to S_n. We will show that S_v can be upper bounded by taking ρ_2 sufficiently small, and S_e can be upper bounded by the expander mixing lemma as in [277].

Lemma 6.8.8 (Modification of [277, Claim 6])**.** *The number of exceptional and violated vertices is at most*

$$|S_e \cup S_v| \leq \frac{64}{\Delta^{1-2\epsilon}}|S|. \tag{6.10}$$

On the other hand, the number of normal vertices is at least

$$|S_n| \geq (1 - \frac{64}{\Delta^{1-2\epsilon}})|S|. \tag{6.11}$$

Proof. The latter fact follows immediately from the former and recalling that S_e, S_v, and S_n partition S. We now show $|S_v|$ is small. Note that by assumption we have that

$$|S_v| \leq |\delta_1 x| < \rho_2 |x|,$$

since δ_1 is the parity-check matrix of $\mathcal{C}_1$ and every violated vertex corresponds to at least one violated constraint in $\mathcal{C}_1$. Because V_1 has degree Δ^2 in $\mathcal{G}_1^{\square}$ (i.e. each vertex sits in Δ^2 squares), we also have $|x| \leq \Delta^2|S|/2$. Altogether this gives

$$|S_v| < \rho_2\Delta^2|S|/2 = 28|S|/\Delta^{1-2\epsilon}$$

for our choice of ρ_2.

Now we show $|S_e|$ is small. The degree of each non-violated vertex is at least $\delta\Delta$ because the local view corresponds to a non-zero codeword in $C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B$. This implies $|S_n| + |S_e| \leq \frac{2|x|}{\delta\Delta}$. Combining this with our bound on $|S_v|$ gives

$$|S| = |S_n| + |S_e| + |S_v| \leq (\rho_2 + \frac{2}{\delta\Delta})|x| \leq \frac{4}{\delta\Delta}|x| \tag{6.12}$$

where the second inequality holds for large enough Δ (recalling that $\rho_2 = O(\Delta^{-3+2\epsilon})$). Applying the expander mixing lemma to $E(S_e, S)$, we then obtain

$$\begin{aligned} |E(S_e, S)| &\leq \frac{\Delta^2}{|V_1|}|S_e||S| + 4\Delta\sqrt{|S_e||S|} \\ &\leq \frac{4\Delta}{\delta|V_1|}|x||S_e| + 4\Delta\sqrt{|S_e||S|} \end{aligned}$$

$$= \frac{1}{3}\Delta^{3/2-\epsilon}|S_e| + 4\Delta\sqrt{|S_e||S|}$$

where we have used the assumption that $|x| \leq \frac{\delta n}{6\Delta^{3/2+\epsilon}}$ and the fact that $|V_1| = |G| = 2n/\Delta^2$. On the other hand, by definition of exceptional vertices we have that $|E(S_e, S)| \geq \Delta^{3/2-\epsilon}|S_e|$. Combining the inequalities we obtain $|S_e| \leq 36|S|/\Delta^{1-2\epsilon}$, and plugging in our bound on $|S_v|$ then gives $|S_e \cup S_v| \leq 64|S|/\Delta^{1-2\epsilon}$ as desired. □

Finally, we prove Claim 6.8.5, Claim 6.8.6, and Claim 6.8.7, completing the result. The latter two follow essentially as in [277] (replacing S_e with $S_e \cup S_v$), but we give the proofs here for completeness.

Proof of Claim 6.8.5. We wish to prove T is non-empty. First, recall that since $x \neq 0$ by assumption, $|S| > 0$. By Lemma 6.8.8, we then have $|S_n| > 0$ as well. We now argue that every vertex in S_n is incident to at least one heavy edge. Since S_n is non-empty, this implies T is non-empty as desired.

To see each vertex in S_n has a heavy edge, recall the local view of each normal vertex is a codeword in $C_A \otimes \mathbb{F}_2^B + \mathbb{F}_2^A \otimes C_B$ with weight less than $w = \Delta^{3/2-\epsilon}$. Because the dual tensor code is w-robust, each column (respectively row) is within $\Delta^{1/2-\epsilon}/\delta$ of a codeword in C_A (respectively C_B). Since these codes all have distance at least $\delta\Delta$, there must be a row or column with at least $\delta\Delta - \Delta^{1/2-\epsilon}/\delta$ ones which exactly corresponds to a heavy edge. We note this fact also implies the total number of heavy edges is at least $|S_n|$, which will be useful later on. □

Proof of Claim 6.8.6. Now that we have confirmed the existence of T, we want to show it is incident to many heavy edges. To do so, we'll argue that T is small compared to the number of heavy edges.

To start, we show that $|T| \leq \frac{64}{\delta^2\Delta}|S|$. The proof is the same as [277, Claim 11], but we give it here for completeness. First, note that by the expander mixing lemma on $\mathcal{G}^{\cup}$

(which is the double cover of a $4\sqrt{\Delta}$-spectral expander) we have:

$$\begin{aligned}|E(S,T)| &\leq \frac{2\Delta}{|G|}|S||T| + 4\sqrt{\Delta}\sqrt{|S||T|} \\ &\leq \frac{2\Delta^{1/2-\varepsilon}}{3}|T| + 4\sqrt{\Delta}\sqrt{|S||T|}\end{aligned}$$

where as in Lemma 6.8.8 we have again used the fact that

$$|S| \leq \frac{4}{\delta\Delta}\,|x| \leq \frac{2}{3\Delta^{5/2+\varepsilon}}n = \frac{1}{3\Delta^{1/2+\varepsilon}}|G|.$$

On the other hand, since each vertex $v \in T$ is incident to at least one heavy edge e by definition, v (and e) are contained in at least $\delta\Delta - \Delta^{1/2-\epsilon}/\delta$ squares in x. Since each of these contains an additional (unique) edge incident to v, we also have the following lower bound

$$|E(S,T)| \geq (\delta\Delta - \Delta^{1/2-\epsilon}/\delta)|T|.$$

Combining these inequalities one can check that $|T| \leq \frac{64}{\delta^2\Delta}|S|$ for large enough Δ as desired.

With this in hand, recall from the proof of Claim 6.8.5 that the total number of heavy edges in E_x is at least $|S_n| \geq (1 - \frac{64}{\Delta^{1-2\epsilon}})|S|$ (where the inequality is given by Lemma 6.8.8). Together, this implies the average number of heavy edges incident to a vertex in T is at least:

$$\frac{|S_n|}{|T|} \geq \frac{\delta^2\Delta}{64}\left(1 - \frac{64}{\Delta^{1-2\epsilon}}\right) =: 2\alpha\Delta. \tag{6.13}$$

Finally given that the average degree is at least $2\alpha\Delta$, we want to show there is some fraction of vertices with degree $\geq \alpha\Delta$. This is immediate from recalling that the maximum degree of $\mathcal{G}^{\cup}$ (and thus T) is 2Δ, which implies at least an $\alpha/2$ fraction of vertices in T are incident to at least $\alpha\Delta$ heavy edges as desired. □

Proof of Claim 6.8.7. Finally, we want to show there are few edges between T and $S_e \cup S_v$.

This follows from the fact that both sets are small, and the underlying graph $\mathcal{G}^{\cup}$ is the double cover of a $4\sqrt{\Delta}$-expander on $|G| = \frac{2n}{\Delta^2}$ vertices. In particular, combining the expander mixing lemma with our bounds from Lemma 6.8.8 gives:

$$\begin{aligned} E(S_e \cup S_v, T) &\leq \frac{2\Delta}{|G|}|S_e \cup S_v||T| + 4\sqrt{\Delta}\sqrt{|T||S_e \cup S_v|} \\ &\leq \frac{128\Delta^{2\varepsilon}}{|G|}|S||T| + 32\Delta^{\varepsilon}\sqrt{|T||S|}. \end{aligned}$$

Recall that $|S| \leq \frac{1}{3\Delta^{1/2+\varepsilon}}|G|$. Further, since each normal vertex is adjacent to T and the degree of T is at most 2Δ, we have $(1 - \frac{64}{\Delta^{1-2\epsilon}})|S| \leq |S_n| \leq 2\Delta|T|$, and thus for large enough Δ that $|S| \leq 4\Delta|T|$. Altogether we therefore have:

$$E(S_e \cup S_v, T) \leq \frac{128}{3\Delta^{1/2-\varepsilon}}|T| + 64\Delta^{1/2+\varepsilon}|T| \leq \beta\Delta^{1/2+\epsilon}|T|$$

where $\beta = 64 + \frac{128}{3\Delta}$. As a result, at most an $\alpha/4$ fraction of vertices in T are incident to more than $d_1 = \frac{4\beta}{\alpha}\Delta^{1/2+\epsilon}$ vertices of $S_e \cup S_v$ as desired, which completes the proof of Claim 6.8.3 and Theorem 6.8.2 in turn. □

□

□

Putting everything together, we now prove the existence of an explicit family of SS-HDX.

Proof of Theorem 6.8.1. Fix any $r \in (0, 1/2)$, $\varepsilon \in (0, 1/2)$, $\gamma \in (1/2 + \varepsilon, 1)$, and $\delta \in (0, 1)$ satisfying $-\delta \log \delta - (1 - \delta)\log(1 - \delta) < r$, and let $\Delta = \Delta(r, \varepsilon/2, \gamma, \delta) \in \mathbb{N}$ be sufficiently large that the guarantees of Corollary 6.7.9 and Theorem 6.8.2 are met. Brute forcing over pairs of length Δ codes C_A, C_B of dimensions $r\Delta$ and $(1-r)\Delta$ respectively, Corollary 6.7.9 promises we can find in $O_\Delta(1)$ time codes C_A, C_B such that:

1. $\dim C_A = \lfloor r\Delta \rfloor$ and $\dim C_B = \Delta - \dim C_A$,

2. The distances of $C_A, C_B, C_A^\perp, C_B^\perp$ are all at least $\delta\Delta$,

3. Both dual tensor codes $C_0^\perp = (C_A \otimes C_B)^\perp$ and $C_1^\perp = (C_A^\perp \otimes C_B^\perp)^\perp$ are $\Delta^{3/2-\epsilon/2}$-robust with Δ^γ-resistance to puncturing.

4. $C_A, C_B, C_A^\perp$, and $C_B^\perp$ have generator matrices where every row and column have at least two ones.

Following the construction and the discussion earlier this section, the Tanner maps resulting from these codes and the explicit left-right Cayley complexes of [117] give an explicit family of chain complexes with degree between 3 and $2\Delta^2$ and non-trivial cohomology. Furthermore each individual complex in the family satisfies the requirements of Theorem 6.8.2 in both directions, so by symmetry the complexes are (ρ_1, ρ_2)-small-set HDX for $\rho_1 = \frac{\delta}{6\Delta^{3/2+\epsilon}}$ and $\rho_2 = \frac{56}{\Delta^{3-2\epsilon}}$. This concludes the proof of Theorem 6.8.1. □

Remark 6.8.9. We note that the proof of Theorem 6.8.2 actually gives a stronger guarantee than small-set (co)-boundary expansion. In particular, because the boundary y that reduces the weight of x is supported on a local view of a single vertex, the result actually gives an isoperimetric inequality for the broader class of small, locally minimal functions:

$$\forall x \in \mathbb{F}_2^n \text{ s.t. } x \text{ is locally minimal and } |x| \leq \rho_1 n : |\delta_1 x| \geq \rho_2 |x|,$$

where x is locally minimal if $|x| \leq |x + \delta_0(e_v)|$ for all basis vectors $e_v \in \mathbb{F}_2^m$. As discussed in Section 6.3.1, this stronger isoperimetric inequality has seen prior use in the topological HDX literature [230, 141, 235, 237] as well as in recent work on c3-LTCs [281] and qLDPC codes [282].

Acknowledgements

The authors thank Noah Fleming, Sam Hopkins, and Russell Impagliazzo for helpful discussion on reductions within the Sum-of-Squares hierarchy, Amy Kanne for helpful discussions on qLDPC codes and [277], and Tali Kaufman for many fruitful discussions on high dimensional expansion. The authors also thank Sam Hopkins, Shachar Lovett, and Anthony Ostuni for helpful comments on an earlier version of the manuscript.

6.9 Existence of Good Base Codes

In this section we prove Corollary 6.7.9, the existence of base codes C_A and C_B with the properties needed for our SS-HDX construction in Theorem 6.8.1. We restate the result here for convenience.

Corollary 6.9.1. *Fix $r \in (0, 1/2)$, $\epsilon \in (0, 1/2)$, $\gamma \in (1/2 + \epsilon, 1)$ and $\delta > 0$ satisfying $-\delta \log \delta - (1-\delta)\log(1-\delta) < r$. when Δ is large enough, there exist codes C_A and C_B of length Δ such that*

1. $\dim C_A = \lfloor r\Delta \rfloor$ *and* $\dim C_B = \Delta - \dim C_A$
2. *The distances of $C_A, C_B, C_A^\perp, C_B^\perp$ are all at least $\delta\Delta$*
3. *Both dual tensor codes $C_0^\perp = (C_A \otimes C_B)^\perp$ and $C_1^\perp = (C_A^\perp \otimes C_B^\perp)^\perp$ are $\Delta^{3/2-\epsilon}$-robust with Δ^γ-resistance to puncturing*
4. *$C_A, C_B, C_A^\perp$, and $C_B^\perp$ have generator matrices where every row and column have at least two ones.*

Proof. We assume for notational simplicity that $r\Delta$ and $(1-r)\Delta$ are integral (the proof is essentially the same without this assumption). We will argue that all four properties are satisfied with probability going to one (as Δ becomes large) under some distribution

for the generation of C_A and C_B. By a union bound, a pair satisfying all properties must then exist for large enough Δ.

Consider the distribution over codes C_A and C_B given by generating C_A by a uniformly random $r\Delta \times \Delta$ generator matrix, and $C_B^\perp$ from an independent uniformly random $r\Delta \times \Delta$ generator matrix. Leverrier and Zémor [277] prove that the first three conditions occur with probability going to one under this distribution (see [277, Theorem 17]), so we need only show the last condition holds.

This follows easily from a few basic observations. Let $r_0 \in (0, 1)$ be any constant. First, observe that conditioned on being full rank, a uniformly random $r_0\Delta \times \Delta$ generator matrix corresponds to a uniformly random subspace of dimension $r_0\Delta$, and furthermore that such a matrix is full rank with probability going to 1 as Δ grows large. Second, note that by a Chernoff and union bound, the probability this random generator matrix has any row or column with less than two ones also quickly goes to zero. This implies that for any fixed r_0, as Δ grows large the probability that a random subspace of dimension $r_0\Delta$ has a generator matrix satisfying condition 4 goes to 1.

Since C_A and $C_B^\perp$ are generated by uniformly random $r\Delta \times \Delta$ generator matrices, they clearly satisfy condition 4 with high probability. The trick is then simply to notice that (conditioned on full rank), $C_A^\perp$ and C_B are uniformly random subspaces of dimension $(1 - r)\Delta$, and therefore also satisfy condition 4 with probability going to one by the above observation. □

This chapter, in full, is based on the material as it appears in Foundations of Computer Science 2022. Hopkins, Max; Lin, Ting-Chun. "Explicit Lower Bounds Against $\Omega(n)$-Rounds of Sum-of-Squares". The dissertation author was a primary investigator and author of this material.

Chapter 7

High Dimensional Expanders: Eigenstripping, Pseudorandomness, and Unique Games

7.1 Introduction

Since their introduction by Kaufman and Mass [234] in 2016, higher order random walks (HD-walks) on high dimensional expanders (HDX) have seen an explosion of research and application throughout theoretical computer science, perhaps most famously in approximate sampling [25, 11, 24, 95, 96, 94, 146, 218, 286, 66], but also in CSP-approximation [9], error correction [121, 112, 220, 221], and agreement testing [124, 109, 236]. These breakthroughs, while evidence of the importance of HD-walks, tend to have a fairly narrow focus in their analysis of the walks themselves—most rely only on proving (often one-sided) bounds on spectral expansion. On the other hand, there are many reasons to study combinatorial and spectral structure beyond the second eigenvalue. Such results are often useful, for instance, when designing graph algorithms (e.g. techniques relying on threshold rank [48, 189, 311]), and have even shown up in recent breakthroughs in hardness of approximation [253, 126, 125, 47, 252, 255], where such an analysis of the Johnson and Grassmann graphs (themselves HD-walks) proved crucial for the resolution of the 2-2 Games Conjecture. Despite so many interesting connections and the recent flurry of work

on HDX, these structures remain relatively unexplored and poorly understood.

Building on work of [239, 111], we make progress on this problem, building a combinatorial characterization of the spectral structure of HD-walks on *two-sided local-spectral expanders*, a variant of HDX introduced by Dinur and Kaufman [124] to generalize the Johnson and Grassmann schemes[1] (and their corresponding agreement tests). We show how our characterization leads to a new understanding of local-to-global algorithms on HDX, including the introduction of a novel spectral parameter generalizing threshold rank, and give a concrete application to efficient algorithms for Unique Games on HD-walks. While the structural theorems we develop lie in a different regime than the one needed to recover hardness results like the proof of 2-2 Games, we open the door for future work connecting HDX and hardness of approximation.

7.1.1 Contributions

We start with an informal overview of our contributions, broken down into three main sections. We give a more thorough exposition of these results in Section 7.2 along with some relevant background, and overview their proofs in Section 7.3. The remainder of the paper is devoted to background, proof details, and further discussion.

Eigenstripping and ST-Rank:.

We develop the interplay of spectral and combinatorial structure on HD-walks, and introduce a spectral parameter called *Stripped Threshold Rank* (ST-Rank) that generalizes threshold rank and controls the performance of local algorithms on such graphs. In more detail, we prove that the spectrum of any HD-walk can be decomposed into small, disjoint intervals we call "eigenstrips," where the corresponding eigenvectors in each strip share the same combinatorial structure.

[1]The Johnson Scheme consists of matrices indexed by k-sets of $[n]$ which depend only on intersection size. HD-walks on the complete complex are exactly the non-negative elements of the Johnson Scheme. The Grassmann scheme is an analogous object over vector spaces.

Theorem 7.1.1 (Eigenstripping (informal))**.** *Let M be an HD-walk on k-sets of an γ-HDX. Then the spectrum of M lies in $k+1$ eigenstrips of width $O_{k,M}(\gamma^{1/2})$,[2] each corresponding combinatorially to a level in the underlying HDX.*[3]

The ST-Rank of an HD-walk (Definition 7.2.3) then measures the number of spectral strips with eigenvalues above some smallness threshold. This generalizes the concept of standard threshold rank, which measures the *total* number of such eigenvalues, to coarser eigendecompositions. Threshold rank itself is well known to control the performance of many graph algorithms [28, 48, 189, 311, 209]. At their core, these techniques often boil down to various methods of enumerating over eigenvectors with large eigenvalues. On structured objects like HD-walks, we argue one should instead enumerate over *strips* of eigenvectors with matching combinatorial structure, and therefore that *ST-rank* is the relevant parameter. Since many important objects share similar combinatorial and spectral characteristics to HD-walks (e.g. noisy hypercube, q-ary hypercube...), we expect ST-Rank to have far-reaching applications beyond those studied in this work.

Edge-expansion in HD-walks:.

The combinatorial and spectral machinery we develop allows us to characterize a fundamental graph property on HD-walks: edge-expansion. Given a graph $G_M = (V, E)$ (corresponding to a random walk M), the edge expansion of a subset $S \subseteq V$ measures the expected probability of leaving S after a single application of M. It is hard to overstate the importance of edge-expansion throughout theoretical computer science. While the most widely used characterizations of expansion are for families where all sets expand (expander graphs) or all *small* sets expand (small set expanders), more involved characterizations for objects such as the Johnson and Grassmann graphs have seen recent use for both hardness of approximation [253, 126, 125, 47, 255] and algorithms [37]. We prove that HD-walks

[2]This was recently improved to $O_k(\gamma)$ by Zhang [363].

[3]There are $k+1$ rather than k strips since 0 is a dimension of the complex (corresponding to the empty set).

exhibit similar expansion characteristics to the Johnson graphs in two key aspects: the expansion of *locally-structured* sets called links, and the *structure* of non-expanding sets.

Applying our machinery to the former reveals a close connection between local expansion and ST-rank:

Theorem 7.1.2 (Local Expansion vs. Eigenstripping (informal))**.** *The (non)-expansion of any i-link*[4] *is almost exactly the eigenvalue corresponding to the ith level's eigenstrip.*

This connection is particularly useful in the construction and analysis of spectral 'local-to-global' graph algorithms, where it helps tie performance guarantees to ST-rank rather than threshold rank. Surprisingly, to our knowledge, Theorem 7.1.2 is novel even for the Johnson graphs (though in this case it follows easily from known results). In fact, this special case alone already gives a better understanding of recent local-to-global algorithm for unique games [37].

A corollary of this result, on the other hand, extends a very well known fact on the Johnson graphs: *locally structured sets expand poorly.* This raises a natural question: are *all* (small) non-expanding sets explained by local structure? Before answering, we address a subtle point: what exactly does it mean to be "explained" by local structure? There are a number of reasonable ways to formalize this notion, each with its own use. Following [37], we resolve an ℓ_2-*variant* of this conjecture stating:

Theorem 7.1.3 (Characterizing Expansion in HD-walks (informal ℓ_2-variant))**.** *Any non-expanding set in an HD-walk has* ***high variance*** *across low-level links.*

Similar results used in hardness of approximation [253, 126, 125, 47, 252, 255], on the other hand, rely on an ℓ_∞-variant that replaces variance with *maximum.* While our bound is exactly tight in the former regime, it (necessarily) loses a factor in the latter in cases where there is a significant gap between ℓ_2 and ℓ_∞ structure. Proving a tight bound

[4]Links are grouped by level of the complex. The set of links at each level give increasingly finer decompositions of the HDX into local parts. See Section 7.2 for details.

directly on the ℓ_∞-variant for HD-walks remains an important open question given their close connections to the structures used in hardness of approximation.

Playing Unique Games:.

Unique Games are a well-studied class of 2-CSP that underlie a central open question in computational complexity and algorithms called the *Unique Games Conjecture* (UGC). The UGC stipulates that distinguishing between almost satisfiable (value $\geq 1-\epsilon$) and highly unsatisfiable (value $\leq \epsilon$) instances of Unique Games is NP-hard. It implies optimal hardness of approximation for a host of optimization problems such as Max-Cut [251], Vertex Cover [257], and in fact all CSPs [323], but is still not known to be true or false. In a recent breakthrough, Khot, Minzer, and Safra [255] (following a long line of work [253, 126, 125, 47, 252]) used an ℓ_∞ characterization of expansion on the Grassmann graphs to prove a weaker variant known as the 2-2 Games Conjecture: it is NP-hard to distinguish $(\frac{1}{2}-\varepsilon)$-satisfiable instances from ϵ-satisfiable instances of unique games. On the other hand, in contrast to hard CSPs such as 3SAT, a long line of research establishes sub-exponential time approximation algorithms for unique games [28] and furthermore shows that approximating unique games is easy on various restricted classes of graphs such as expanders [293], perturbed random graphs [266], certifiable small-set expanders and Johnson graphs [37].

While our ℓ_2-characterization of expansion may not be as useful for hardness of approximation as the ℓ_∞-variant, it gets its chance to shine in the latter context: *algorithms*. For most classes of constraint graphs, the best known algorithms for unique games depend on the threshold rank [28, 48, 189] of the graph. Using a sum-of-squares variant of our ℓ_2-characterization and the connection between local expansion and stripped eigenvalues, we build a local-to-global algorithm for unique games on HD-walks (based off the recent paradigm of [37]) that depends instead on ST-rank.

Theorem 7.1.4 (Playing Unique Games on HD-walks (informal))**.** *For any $\varepsilon \in (0, .01)$,*

there exists an algorithm $\mathcal{A}$ with the following guarantee. If $\mathcal{I}$ is a $(1-\varepsilon)$-satisfiable instance of unique games whose constraint graph is an HD-walk M[5] on k-sets of an HDX, then:

1. *$\mathcal{A}$ outputs a poly(k^{-r}, ε)-satisfying solution.*
2. *$\mathcal{A}$ runs in time at most $N^{\tilde{O}(\binom{k}{r}\varepsilon^{-1})}$*

where $r = R_{1-O(\epsilon)}(M)$ is the ST-rank of M above threshold $1 - O(\epsilon)$, and there are N total k-sets.

When $k = O(1)$, the ST-rank above any threshold τ is substantially smaller than the threshold rank above τ for (non-expanding) HD-walks, and Theorem 7.1.4 therefore obtains a strict improvement over previous results based on threshold rank for this family. In many cases, such as sparsifications of (constant-level) Johnson graphs or related objects like slices of the noisy q-ary hypercube, our algorithm reduces the best-known runtime from nearly-exponential to polynomial. Besides its independent algorithmic interest, the result carries a number of connections to other open problems both in and outside of the study of unique games. The algorithm sheds some (limited) light,[6] for instance, on requisite structure for candidate hardness reductions that use direct product testing [256] or potential attempts to use agreement tests based on local-spectral expanders [124]. It gives some hope for better unique games algorithms on graphs like the hypercube [2] as well, which can be viewed as an HD-walk on level $k = O(\log N)$ on a weaker (one-sided) notion of high dimensional expansion than we study. Outside of unique games the result has some further connections to error correcting codes, where approximation algorithms for general CSPs on HDX underlie recent breakthroughs in efficient decoding for locally testable codes [220, 221]. In fact, algorithms specifically for unique games over expanders

[5]Note that self-adjoint random walks can be equivalently viewed as undirected weighted graphs, and therefore also as underlying constraint graphs for unique games.

[6]The result certainly does not rule out reductions using such structure, but states that one must be careful in doing so that the resulting constraint graphs fall outside the parameter regime for our algorithm.

have already seen similar use for efficient decoding of direct-product codes based on HDX structure [121].

7.2 Our Results

We now move to a more in-depth exposition of our main results. First, however, we cover some background regarding local-spectral expanders and higher order random walks. For a full treatment of these and related objects, see Section 7.5. Local-spectral expansion is a robust notion of local connectivity on **pure simplicial complexes** introduced by Dinur and Kaufman [124]. A d-dimensional pure simplicial complex is the downward closure of a d-uniform hypergraph, that is a collection:

$$X = X(0) \cup \ldots \cup X(d)$$

where $X(d) \subseteq \binom{[n]}{d}$ is a d-uniform hypergraph, and $X(i)$ consists of all $\tau \in \binom{[n]}{i}$ such that $\tau \subset T$ for some $T \in X(d)$ (and $X(0) = \{\emptyset\}$). We note that this notation is off-by-one from much of the HDX literature, where $X(i)$ is instead given by sets of size $i+1$, a notation rooted in the topological view of simplicial complexes. The former definition is more common in combinatorial works (e.g. the sampling literature [24]), and more natural for our purposes.

Simplicial complexes come equipped with natural local structure called **links**. For every i and $\tau \in X(i)$, the link of τ is the restriction of the complex to faces containing τ, that is:

$$X_\tau = \{\sigma \ : \ \sigma \cap \tau = \emptyset, \ \sigma \cup \tau \in X\}.$$

We call the link of an i-face an i-link, and when clear from context, will also use X_τ to denote the set of faces at a given level (e.g. in $X(d)$) containing τ. Following Dinur and Kaufman [124], we say a complex is a **two-sided γ-local-spectral expander** if the

graph underlying every link is a γ-spectral expander (that is all non-trivial eigenvalues are smaller than γ in absolute value). Finally, it is worth noting that we actually study the more general set of **weighted pure simplicial complexes** (see Section 7.5), but for simplicity restrict to the unweighted case for the moment as it requires essentially no modification.

Higher order random walks [234] are analogues of the walk associated with standard graphs (given by the normalized adjacency matrix) that moves from vertex to vertex via an edge. In a higher order random walk, one applies a similar process at any level of the complex—moving for instance between two edges via a triangle, or two triangles via a pyramid. We call walks between k-faces of a complex **k-dimensional HD-walks** (see Definition 7.5.7 for formal definition), and study a broad set of walks that capture important structures such as sparsifications of the Johnson and Grassmann schemes.[7] For simplicity, throughout the introduction we will often focus on two natural classes of HD-walks which see the most use in the literature: the **canonical walks** N_k^i which walk between k-faces via a neighboring $(k+i)$-face, and the **partial-swap walks** S_k^i which do the same but only move between k-faces of fixed intersection size $k-i$. While this latter class may seem less natural at first, notice that on the complete complex they are exactly the well-studied Johnson graph $J(n,k,k-i)$ (the graph on $\binom{[n]}{k}$ where $(v,w) \in E$ iff $|v \cap w| = k-i$).

7.2.1 Eigenstripping and ST-Rank

With background out of the way, we start our results in earnest with a more in-depth discussion of eigenstripping and ST-rank: the spectral and combinatorial structure of HD-walks. We prove that the spectra of k-dimensional HD-walks lie in $k+1$ tightly concentrated "eigenstrips", where the ith strip corresponds combinatorially to functions

[7] We note that the Grassmann scheme comes from applying HD-walks to the Grassmann poset, which is not a simplicial complex.

lifted from the ith level of the complex by averaging.

Theorem 7.2.1 (Spectrum of HD-Walks (Informal Theorem 7.6.2 + Corollary 7.7.6 + Proposition 7.7.11))**.** *Let M be an HD-walk on the kth level of a two-sided γ-local-spectral expander. Then the spectra of M is highly concentrated in $k+1$ strips:*

$$Spec(M) \in \{1\} \cup \bigcup_{j=1}^{k} [\lambda_i(M) - e, \lambda_i(M) + e]$$

where the $\lambda_i(M)$ are decreasing constants depending only on M, and the error term satisfies $e \leq O_{k,M}(\sqrt{\gamma})$. Moreover, the span of eigenvectors with eigenvalues in the strip $\lambda_i(M) \pm e$ are (approximately) functions in $X(i)$ lifted to $X(k)$ by averaging.

Recently, Zhang [363] provided a quantitative improvement of Theorem 7.2.1 such that $e \leq O_k(\gamma)$. Theorem 7.2.1 can be seen as a marriage (and generalization) of previous results studying N_k^1 of Kaufman and Oppenheim [239], and Dikstein, Dinur, Filmus, and Harsha (DDFH) [111]. The former prove that N_k^1 exhibits spectral eigenstripping, while the latter introduce the corresponding combinatorial structure but lack the machinery to tie it directly to these strips. We fill in the gap by proving a general linear algebraic theorem of independent interest.

Theorem 7.2.2 (Approximate Eigendecompositions Imply Eigenstripping (Informal Theorem 7.6.2))**.** *Let M be a self-adjoint operator over an inner product space V, and $V = V^1 \oplus \ldots \oplus V^k$ a decomposition satisfying $\forall 1 \leq i \leq k, f_i \in V^i$:*

$$\|Mf_i - \lambda_i f_i\| \leq c_i \|f_i\|$$

for some family of constants $(\{\lambda_i\}_{i=1}^k, \{c_i\}_{i=1}^k)$. Then as long as the c_i are sufficiently

small, the spectra of M is concentrated around each λ_i:

$$Spec(M) \subseteq \bigcup_{i=1}^{k} [\lambda_i - e, \lambda_i + e] = I_{\lambda_i},$$

where $e = O_{k,\lambda}\left(\sqrt{\max_i\{c_i\}}\right)$. This was recently improved to $e \leq O_k(\max_i\{c_i\})$ by Zhang [363].

Theorem 7.2.1 then follows from proving that DDFH's decomposition is an approximate eigendecomposition for all HD-walks (DDFH only show this holds for N_k^1).

In the analysis of graphs, it is often useful to bound the number of eigenvalues above some smallness threshold. This parameter, called the **threshold rank**, often plays a key role, for instance, in graph algorithms, including many methods for generic 2CSP approximation [28, 48, 189, 311, 209]. Painted in broad strokes, these algorithms usually boil down to some method of eigenvalue enumeration, whether performed directly or implicitly in the proof of key structural lemmas. Given the spectral structure of HD-walks (or generally any operator with a natural approximate eigendecomposition), we argue that it is generally more natural to enumerate over *strips* with large eigenvalues instead. This motivates a natural spectral complexity measure we call **stripped threshold rank**.

Definition 7.2.3 (Stripped Threshold Rank)**.** Let M be a linear operator over a vector space V with decomposition $V = \bigoplus_i V^i$ denoted $\mathscr{D}$, where each V^i is the span of some set of eigenvectors. Given $\delta \in \mathbb{R}$, the stripped threshold rank (ST-Rank) with respect to δ and $\mathscr{D}$ is:

$$R_\delta(M, \mathscr{D}) = \left|\{V^i \in \mathscr{D} : \exists f \in V^i, Mf = \lambda f, \lambda > \delta\}\right|.$$

In this work, $\mathscr{D}$ will always correspond to the eigenstrips given by Theorem 7.2.1 so we drop it from the notation.

The ST-Rank of any (non-expanding) k-dimensional HD-walk is always substantially smaller than its standard threshold rank (which is at least poly(n), and often as large as

$n^{\Omega(k)}$). As a general algorithmic paradigm, if one can leverage the combinatorial structure of eigenstrips to replace eigenvalue enumeration, it is possible to achieve substantially better performance on structured graphs. In Section 7.2.3, we discuss a concrete instantiation of this framework to unique games, where we gain substantial improvements over state of the art algorithms [48, 28] over such constraint graphs.

Before moving on to such results, however, it is natural to ask whether we can give a finer-grained characterization of ST-rank for HD-walks. In Section 7.7 we show how to explicitly compute the approximate eigenvalue $\lambda_i(M)$ corresponding to each strip based upon the structure of M. While the full result requires further background, its specification to canonical and partial-swap walks has a nice combinatorial interpretation. Recall that the canonical walk N_k^i walks between k-faces through a shared $(k+i)$-face. We say it has **depth** $i/(k+i)$ since it traverses i of the $k+i$ relevant levels of the complex. Similarly, we say the partial-swap walk S_k^i has depth i/k, since it swaps i out of k elements (up to factors in γ, this can also be viewed as walking between k faces via a shared $k-i$ face, where one traverses i out of the k relevant levels). We prove that the stripped eigenvalues corresponding to the canonical and partial-swap walks decay exponentially fast with a base rate dependent on depth.

Theorem 7.2.4 (ST-rank of HD-walks (Corollary 7.7.7 + Corollary 7.7.9))**.** *Let M be a canonical or partial-swap walk of depth $0 \leq \beta \leq 1$ on a sufficiently strong two-sided local-spectral expander. Then the eigenvalues corresponding to the eigenstrips of M decay exponentially fast:*

$$\lambda_i(M) \leq e^{-\beta i}.$$

The ST-Rank $R_\delta(M)$ is then at most:

$$R_\delta(M) \leq \frac{\ln\left(\frac{1}{\delta}\right)}{\beta}$$

In other words, walks that reach deep into the complex (e.g. $N_k^{k/2}$) have constant ST-Rank, whereas shallow walks like N_k^1 have ST-rank $\Omega(k)$.

7.2.2 Characterizing (non)-Expansion in HD-Walks

We now move to our main structural application, characterizing edge expansion in HD-walks. Edge expansion is a fundamental combinatorial property of graphs with applications across many areas of theoretical computer science, including (as we will soon discuss) both hardness and algorithms for unique games. For a subset of vertices $S \subseteq V$, edge expansion measures the (normalized) fraction of edges which leave S.

Definition 7.2.5 (Edge Expansion)**.** Given a graph $G(V, E)$, the edge expansion of a subset $S \subset V$ is:

$$\phi(G, S) = \frac{E(S, V \setminus S)}{E(S, V)}$$

where $E(S, T)$ counts the number of edges crossing from S to T (double-counting edges in the intersection). When convenient, we denote non-expansion, $1 - \phi(G, S)$, as $\bar{\phi}(G, S)$, and drop G from the notation when clear from context.

It is often useful to characterize exactly which sets in a graph expand. The two most widely used characterizations are when all[8] sets expand (expanders), or when all *small* sets expand (small-set expanders). However, many important structures fall outside such a simple characterization. The Johnson and Grassmann graphs, for instance, are well known to have small non-expanding sets. Characterizing the structure of non-expansion in these graphs was crucial not only for the resolution of the 2-2 Games Conjecture [255], but also for recent algorithms for unique games [37].

Using the spectral machinery developed in the previous section, we give a tight characterization of expansion in HD-walks. As discussed, we focus mainly on two key aspects: the expansion of *links*, and the structure of non-expanding sets. We start with

[8]By this we really mean all sets of size at most $|V|/2$.

the former, where links have long been the prototypical example of small, non-expanding sets for the Johnson graphs. In fact, we prove this stems from a much stronger connection between local expansion and spectral structure in HD-walks.

Theorem 7.2.6 (Local (non)-Expansion = Global Spectrum (Informal Theorem 7.9.2))**.** *Let M be a k-dimensional HD-walk on a sufficiently strong d-dimensional two-sided γ-local-spectral expander with $d > k$,*[9] *and let $\lambda_i(M)$ be approximate eigenvalues corresponding to M's $k+1$ eigenstrips. For all $0 \le i \le k$ and $\tau \in X(i)$, let X_τ denote the set of k-faces that contain τ. Then X_τ is expanding if and only if $\lambda_i(M)$ is small:*

$$\bar{\phi}(X_\tau) \in \lambda_i(M) \pm O_k(\gamma).$$

Since HD-walks (like the Johnson graphs) are generally poor spectral expanders, Theorem 7.2.6 implies the existence of small, non-expanding sets (namely 1-links[10]). Our stronger characterization of local expansion is of independent interest beyond this simple corollary, however, due to its important connections to local-to-global algorithms and ST-rank. In particular, essentially all work on high-dimensional expanders relies on a strategy known as the *local-to-global paradigm*, where a global property (e.g. mixing, agreement testing, etc.) is reduced to examining a local version on links. These arguments often reduce to showing that, in some relevant sense, interaction *between* links is minimal. Theorem 7.2.6 offers exactly such a statement for levels of the complex which correspond to eigenspaces with large eigenvalues—since links at these levels are non-expanding, they don't have much interaction. This gives a holistic local-to-global approach for spectral graph algorithms on HDX, since the eigenspaces corresponding to *small* eigenvalues are

[9]The lower bound on non-expansion still holds for $d = k$, which is sufficient for most applications of this result. It's also worth noting that the condition $d > k$ itself is generally trivially satisfied, since all known two-sided local-spectral expanders are in fact cutoffs of larger complexes.

[10]Since we generally consider the regime where $|X(1)| \gg k$, a basic averaging argument gives that there must exist a $o(1)$-size 1-link. One may also note that in a γ-local-spectral expander, no 1-link can have more than $O_k(\gamma)$ mass, so all links are small if $\gamma \le o(1)$.

traditionally easy to handle through other means. Later we will see an algorithmic application of this approach to unique games, where the non-expansion of links allows us to patch together local solutions, and the connection to global spectra ties performance of the algorithm to ST-rank.

Returning to the topic of expansion, Theorem 7.2.6 shows that as long as M is not a spectral expander, there always exist small, non-expanding sets in the form of links. This raises a natural question: are *all* non-expanding sets explained by links? Before answering, let's spend a little time formalizing this. Given a subset $S \subset X(k)$, let $L_{S,i}$ be a function on i-faces measuring the deviation of S from its average across links at level i:

$$\forall \tau \in X(i) : L_{S,i}(\tau) = \mathop{\mathbb{E}}_{X_\tau}[\mathbb{1}_S] - \mathbb{E}[\mathbb{1}_S].$$

The statement "non-expansion is explained by links" then really amounts to saying that we expect $L_{S,i}$ to be far from 0 in some sense if S is non-expanding. There are a number of reasonable ways to formalize this notion, each with its own use. Following [37], we mainly focus on an ℓ_2-variant that is particularly useful for designing local-to-global algorithms: if S is non-expanding, then $\|L_{S,i}\|_2^2$ (or equivalently the *variance* of S across i-links) must be large. It is perhaps easier to think about this result in terms of its contrapositive. Call a set ℓ_2**-pseudorandom** if its variance across i-links is *small*.

Definition 7.2.7 (ℓ_2-Pseudorandom Sets)**.** Let X be a pure simplicial complex. We call a subset $S \subset X(k)$ (ε, ℓ)-ℓ_2-pseudorandom if

$$\forall i \leq \ell : \|L_{S,i}\|_2^2 \leq \varepsilon \mathbb{E}[\mathbb{1}_S]$$

We remark briefly on the choice of normalization by $\mathbb{E}[\mathbb{1}_S]$ on the right. While mostly a formality, we will see this is in fact a natural choice both in comparison to the ℓ_∞-variant (which differs in some natural sense by at least a factor of $\mathbb{E}[\mathbb{1}_S]$), and when

considering expansion which is itself normalized by $\mathbb{E}[\mathbb{1}_S]$. Indeed with this in mind, we prove that (ε, ℓ)-ℓ_2-pseudorandom sets expand near-perfectly.

Theorem 7.2.8 (ℓ_2-Pseudorandom Sets Expand (Informal Theorem 7.9.5))**.** *Given a k-dimensional HD-walk M on a sufficiently strong two-sided γ-local-spectral expander and small constants $\alpha, \delta, \varepsilon > 0$, we have that any $(\varepsilon, R_\delta(M))$-pseudorandom set S of density α expands near-perfectly:*

$$\phi(S) \geq 1 - \alpha - \delta - O_k(\varepsilon) - O_k(\gamma)$$

We note that (the formal version of) this result is exactly tight (in the limit of $\gamma \to 0$). Passing back to the original (now contrapositive) form of the statement, Theorem 7.2.8 immediately implies that any non-expanding set must have large variance across links at a level determined by the ST-rank of M. On the algorithmic side, this informally translates to the statement that most interesting structure lies on low-level links, which is crucial to any local-to-global algorithm.

Before formalizing this algorithmic intuition in the case of unique games, it's worth taking a moment to consider the implication of Theorem 7.2.8 to a different formalization of this problem with recent applications to hardness of approximation: the ℓ_∞-variant. In this characterization, the (squared) ℓ_2-norm of $L_{S,i}$ is replaced with its maximum. In other words, the ℓ_∞-variant characterization posits that every non-expanding set must be highly concentrated in some *individual* link. We prove that Theorem 7.2.8 actually holds in this regime as well by a simple reduction. If we analogously define ℓ_∞-pseudorandom sets (see Definition 7.8.5), it is not hard to show that any ℓ_∞-pseudorandom set must also be ℓ_2-pseudorandom. This results in a tight version of Theorem 7.2.8 for the ℓ_∞-regime, but only when $\max(L_{S,i})$ is close to $\frac{1}{\mathbb{E}[\mathbb{1}_S]}\|L_{S,i}\|^2$. Practically, the interesting regime in hardness of approximation is when these two quantities are far apart. In this case our ℓ_∞ to ℓ_2 reduction necessarily loses important factors, so we cannot recover any results from the hardness of approximation literature.

Nevertheless, it is worth stating that as an immediate corollary of our ℓ_2-based analysis we get an ℓ_∞-characterization of expansion in HD-walks: any non-expanding set must be non-trivially concentrated in a link.

Corollary 7.2.9 (Non-expanding Sets Correlate with Links (Informal Corollary 7.9.8))**.** *Let M be a k-dimensional HD-walk on a sufficiently strong two-sided γ-local-spectral expander. Then if $S \subset X(k)$ is a set of density α and expansion at most:*

$$\phi(S) < 1 - \alpha - \delta - O_k(\gamma)$$

for some $\delta > 0$, S must be non-trivially correlated with some i-link for $1 \leq i \leq r = R_{\delta/2}(M)$:

$$\exists 1 \leq i \leq r, \tau \in X(i) : \mathop{\mathbb{E}}_{X_\tau}[\mathbb{1}_S] \geq \alpha + \frac{c_{\delta,r}}{\binom{k}{i}}$$

where $c_{\delta,r}$ depends only on δ and the ST-rank r.

Proving a k-independent ℓ_∞-characterization for HD-walks, either directly or via a stronger k-dependent reduction to the ℓ_2-regime, remains an interesting open problem. Such a characterization is known on the Johnson graphs (and a number of related objects [245, 147]), and may shed light on similar structure in the Grassmann graphs used to prove the 2-2 Games Conjecture [255]. We discuss some additional subtleties in this direction in Section 7.2.4.

7.2.3 Playing Unique Games

One motivation for studying spectral structure and non-expansion in HD-walks stems from a simple class of 2-CSPs known as unique games, a central object of study in hardness-of-approximation since Khot's introduction of the Unique Games Conjecture (UGC) [250] nearly 20 years ago. We study affine unique games which are known to be as hard as unique games [251]:

Definition 7.2.10 (Affine Unique Games)**.** An instance $I = (G, \mathcal{S})$ of affine unique games over alphabet $\Sigma = \{0, \ldots, m-1\}$ is a weighted undirected graph $G(V, E)$, and set of affine shifts $\mathcal{S} = \{s_{uv} \in \Sigma\}_{(u,v)\in E}$. The value of I, $\mathrm{val}(I)$, is the maximum fraction of satisfied constraints over all possible assignments Σ^V:

$$\max_{X\in\Sigma^V} \mathop{\mathbb{P}}_{(u,v)\sim E}[X_u - X_v = s_{uv} (\mathrm{mod}\ |\Sigma|)],$$

where edges are drawn corresponding to their weight. For an individual assignment X, we refer to this expectation as $\mathrm{val}_I(X)$. Any weighted undirected graph G is uniquely associated with a self-adjoint random-walk matrix M where $M(u, v) = \mathbb{P}_E[(u, v)]/\sum_w \mathbb{P}_E[(u, w)]$ (see Section 7.14.1). We will usually view the constraint graph in this manner instead, referring to unique games instances over random-walks as $I = (M, \Pi)$.

Informally, the UGC states that for sufficiently small constants ε, δ, there exists an alphabet size such that distinguishing between instances of unique games with value $1 - \varepsilon$ and δ is NP-hard. A positive resolution to the UGC would resolve the hardness-of-approximation of many important combinatorial optimization problems, including CSPs [323], vertex-cover [257], and a host of others such as [250, 251, 188, 98, 258, 260].

Building on the recent framework of Bafna, Barak, Kothari, Schramm, and Steurer [37], we show how our structural theorems combine with the notion of ST-rank to give efficient algorithms for unique games over HD-walks.

Theorem 7.2.11 (Playing Unique Games on HD-walks (Informal Theorem 7.10.1))**.** *For any $\varepsilon \in (0, .01)$, there exists an algorithm $\mathcal{A}$ with the following guarantee. If $\mathcal{I} = (M, \mathcal{S})$ is an instance of affine unique games with value at least $1 - \varepsilon$ over M, a complete k-dimensional HD-walk on a d-dimensional two-sided γ-local-spectral expander with $\gamma \ll o_k(1)$ and $d > k$,*[11] *then $\mathcal{A}$ outputs a $\mathrm{poly}(\tau)$-satisfying assignment in time $|X(k)|^{\mathrm{poly}(1/\tau)}$, where*

[11] Formally, we note γ also has dependence on M (see Theorem 7.10.1), though this can be removed in special cases like the Johnson scheme. We also note again that $d > k$ is generally a trivial condition for strong two-sided local-spectral expanders.

$\tau = \left(\frac{\varepsilon}{\binom{k}{r(\varepsilon)}}\right)$ *and* $r(\varepsilon) = R_{1-O(\varepsilon)}(M)$ *is the ST-rank of* M*.*

It is worth giving a quantitative comparison of this result to the best known algorithms based on threshold rank [28, 48, 189]. These algorithms run in time roughly exponential in the $(1-O(\epsilon))$-threshold rank of G to get similar soundness guarantees when $k = O(1)$.[12] Due to the poor threshold rank of HD-walks, this generally amounts to nearly exponential time $(2^{|X(k)|^{\mathrm{poly}(\epsilon)}})$, whereas our algorithm runs in time roughly $|X(k)|^{2^{O(k)}}$. In the regime of constant k, Theorem 7.2.11 therefore gives a polynomial time algorithm for such instances achieving an exponential improvement over threshold-rank based algorithms. We describe a few examples of such families below.

Since Theorem 7.2.11 can be a bit hard to interpret without additional knowledge of the high dimensional expansion literature, we end our discussion of algorithms for unique games with a few concrete examples. Perhaps the most basic example generalizing the Johnson graphs that fits into our framework are slices of the q-ary noisy hypercube. For constant-level slices, we improve over BRS from nearly exponential to polynomial. Further, the result is robust in the sense that it continues to hold even when the underlying complex is perturbed. This results in algorithms, for instance, for dense random sparsifications of these slices, or slices taken from a negatively correlated distribution over $\mathbb{Z}_q^n$ (or more generally slices of any spectrally-independent spin-system [24]). Another interesting class of graphs we see significant speed-ups on are algebraic sparsifications of the Johnson graphs stemming from bounded-degree constructions of two-sided local-spectral expanders such as (cutoffs of) Lubotsky, Samuels, and Vishne's [290] Ramanujan complexes, or Kaufman and Oppenheim's [238] coset complex expanders. The resulting speedups on this class of graphs is somewhat more surprising than the above, since they exhibit substantially different structure in other aspects (they are, for instance, of bounded degree unlike the Johnson graphs).

[12] This corresponds to the setting in which the dimension of the HD-walk is fixed, and the number of vertices in the complex grows.

7.2.4 Connections to Hardness and the Grassmann Graphs

We finish the discussion of our results by taking a deeper look at connections with recent progress on the UGC, and argue that our framework opens an avenue for further progress. The resolution of the 2-2 Games Conjecture [255] hinged on a characterization of non-expanding sets on the Grassmann graph not dissimilar to what we have shown for two-sided local-spectral expanders. While we have focused above on HDX which are simplicial complexes, our work extends to a broader set of objects introduced by DDFH [111] called expanding posets which includes class of objects includes expanding subsets of the Grassmann poset we call q-eposets (we refer the reader to [111] for definitions and further discussion).

Theorem 7.2.8 and Corollary 7.2.9 extend naturally to HD-walks on q-eposets. We state the latter result here since it follows without too much difficulty from the same arguments as in this paper, but the full details (and further generalizations to expanding posets) will appear in a companion paper.

Corollary 7.2.12 (Non-expansion in q-eposet)**.** *Let (X, Π) be a two-sided γ-q-eposet with γ sufficiently small, M a k-dimensional HD-walk on (X, Π). Then if $S \subset X(k)$ is a set of density α and expansion:*

$$\phi(S) < 1 - \alpha - O_{q,k}(\gamma) - \delta$$

for some $\delta > 0$ and $r = R_{\delta/2}(M)$, S must be non-trivially correlated with some i-link for $1 \leq i \leq r$:

$$\exists 1 \leq i \leq r, \tau \in X(i) : \mathop{\mathbb{E}}_{X_\tau}[\mathbb{1}_S] \geq \alpha + \frac{c_{\delta,r}}{\binom{k}{i}_q}$$

where $c_{\delta,r}$ depends only on δ and the ST-rank r, and $\binom{k}{i}_q$ is the standard q-binomial coefficient:

$$\binom{k}{i}_q = \prod_{j=0}^{i-1} \frac{q^{k-j} - 1}{q^{i-j} - 1},$$

Since the Grassmann graphs are simply partial-swap walks on the Grassmann poset,[13] Corollary 7.2.12 provides a direct connection to the proof of the 2-2 Games Conjecture [255]. Unfortunately, due to the dependence on k, this result is too weak to recover the proof.

The dependence of both Corollary 7.2.9 and Corollary 7.2.12 on k is a subtle but important point, so we'll finish the section by discussing it in a bit more detail. The particular dependence we get in the ℓ_2-regime, $\binom{k}{i}$ for the Johnson and $\binom{k}{i}_q$ for the Grassmann, is tight and can be understood by examining the variance of i-links. While i-links are the prototypical example of an $\Omega(1)$-pseudorandom set in the ℓ_∞ regime, one can show that they are actually about $1/\binom{k}{i}$-pseudorandom in the ℓ_2-regime (or $1/\binom{k}{i}_q$ for the Grassmann). Since links are non-expanding, any bound in the ℓ_2-regime must have matching dependence on k to make up for this fact. Indeed, one can use the same argument to show that a k-independent bound cannot exist in ℓ_2-regime (even if we relax dependence on pseudorandomness, see Proposition 7.9.7). As a result, any ℓ_∞ to ℓ_2 reduction like ours (which has no dependence on k) will always result in a final bound depending on k.

Finally, it's worth noting that while the ℓ_∞-regime escapes this particular issue since links are $\Omega(1)$-pseudorandom (and indeed that for the Johnson graphs, a k-independent bound is known [252]), there is an additional consideration for the Grassmann graphs: there exist small, non-expanding ℓ_∞-pseudorandom sets [125]. The proof of the 2-2 Games Conjecture therefore relies on a finer-grained definition of local structure than links (called zoom-in zoom-outs) [255]. While we cannot hope to apply exactly the same techniques to analyze this variant, we view our method's generality and simplicity as evidence that a deeper understanding of higher order random walks may be key to further progress on the UGC.

[13] Seeing that they are HD-walks is non-trivial, and follows from the q-analog of work in [9].

7.3 Proof Overview

In this section we give a proof overview of our main results. Throughout the section we will assume the complex is endowed with a uniform distribution, and will (usually) ignore error terms in the spectral parameter γ. The full details for weighted complexes and a careful treatment of the error terms is given in the main body along with a number of further generalizations.

7.3.1 Eigenstripping and the HD-Level-Set Decomposition

We start with a discussion of the techniques underlying Theorem 7.2.1 (Eigenstripping). At its core, this (and indeed all of) our results rely upon now-standard machinery for working on simplicial complexes called the averaging operators.

Definition 7.3.1 (The Averaging Operators)**.** Let X be a d-dimensional pure simplicial complex. For any $0 \leq k \leq d$, denote the space of functions $f : X(k) \to \mathbb{R}$ by C_k. The "Up" operator lifts $f \in C_k$ to $U_k f \in C_{k+1}$:

$$\forall y \in X(k+1) : U_k f(y) = \frac{1}{k+1} \sum_{x \in X(k) : x \subset y} f(x).$$

The "Down" operator lowers functions $f \in C_{k+1}$ to $D_{k+1} f \in C_k$:

$$\forall x \in X(k) : D_{k+1} f(x) = \frac{1}{n-k} \sum_{y \in X(k+1) : y \supset x} f(y).$$

Since it is often useful to compose these operators, we will use the shorthand $U_i^k = U_{k-1} \dots U_i$ and $D_i^k = D_{i+1} \dots D_k$ to denote the composed averaging operators which raise and lower functions between C_i and C_k by averaging.

In fact, the averaging operators are crucial even to *defining* higher order random walks. We discuss this definition in greater detail in Section 7.5, and for now settle for

noting that a basic class of HD-walks are given simply by composing U and D. For instance, the canonical walk N_k^i is the composition $D_k^{k+i}U_k^{k+i}$, and the partial swap walks S_k^i turn out to be an affine combinations of the N_k^i [9].

The proof of Theorem 7.2.1 relies on a useful decomposition for functions on simplicial complexes based upon the averaging operators we call the *HD-level-set Decomposition* recently introduced by Dikstein, Dinur, Filmus, and Harsha [111]. The idea is to break functions into components coming from each level of the complex, lifted by averaging to the top level.

Theorem 7.3.2 (HD-Level-Set Decomposition, Theorem 8.2 [111])**.** *Let X be a d-dimensional two-sided γ-local-spectral expander, $\gamma < \frac{1}{d}$, $0 \leq k \leq d$, and let:*

$$H^0 = C_0, H^i = Ker(D_i), V_k^i = U_i^k H^i.$$

Then:

$$C_k = V_k^0 \oplus \ldots \oplus V_k^k.$$

In other words, every $f \in C_k$ has a unique decomposition $f = f_0 + \ldots + f_k$ such that $f_i = U_i^k g_i$ for $g_i \in Ker(D_i)$.

The HD-Level-Set Decomposition is particularly useful not only for its rigid combinatorial structure, but also its spectral properties. Namely, one familiar with the Johnson Scheme might notice that when X is the complete complex, this decomposition exactly gives the eigenspaces of the Johnson Scheme. DDFH gave an approximate extension of this result to the basic "upper" walk $D_{k+1}U_k$, proving that the V_k^i are approximate eigenspaces.

The proof of Theorem 7.2.1 follows from combining an extension of this result to all HD-walks (see Corollary 7.7.6) with Theorem 7.2.2, which states that any approximate eigendecomposition strictly controls the spectrum of the underlying operator. However, since the proof of Theorem 7.2.2 is purely linear algebraic (and is essentially irrelevant

for understanding our core results on expansion and unique games), we defer detailed discussion of the result to Section 7.6.

7.3.2 Characterizing Expansion

We now take a look at how this combinatorial understanding of the spectra of higher order random walks allows us to characterize their edge-expansion. We start with a standard observation, the expansion of a set can be written as the inner product of its indicator function. In other words, if X is a simplicial complex and M is an HD-walk on $X(k)$, the expansion of any $S \subset X(k)$ with respect to M may be written as:

$$\phi(M, \mathbb{1}_S) = 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \langle \mathbb{1}_S, M\mathbb{1}_S \rangle.$$

The key is then to notice that by the bilinearity of inner products, we can expand the right-hand side in terms of the HD-Level-Set Decomposition. In particular, writing $\mathbb{1}_S = \sum_{i=0}^{k} \mathbb{1}_{S,i}$ for $\mathbb{1}_{S,i} \in V_k^i$, we have:

$$\phi(M, \mathbb{1}_S) = 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=0}^{k} \langle \mathbb{1}_S, M\mathbb{1}_{S,i} \rangle.$$

Finally, since we know each V_k^i is approximately an eigenstrip concentrated around some λ_i, up to additive error in γ we can simplify this to:

$$\phi(M, \mathbb{1}_S) \approx 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=0}^{k} \lambda_i \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle. \tag{7.1}$$

Thus we see that understanding the expansion of S comes down to the interplay between its projection onto each level of the complex and their corresponding approximate eigenvalues. Our characterization of expansion combines this observation with the combinatorial and spectral structure of the HD-Level-Set Decomposition. We devote the rest of the section to sketching the proofs of Theorem 7.2.6 (local expansion vs global spectra), Theorem 7.2.8

(pseudorandom sets expand), and our reduction from the ℓ_∞-variant to ℓ_2-variant.

Proof sketch of Theorem 7.2.6:.

Recall that Theorem 7.2.6 shows a tight inverse relation between the expansion of links and the spectra of M. Given $\tau \in X(j)$, we wish to examine the indicator function $\mathbb{1}_{X_\tau}$ of the link X_τ. First, notice that since $\mathbb{1}_{X_\tau} = \binom{k}{j} U_j^k \mathbb{1}_\tau$, it's easy to see that $\mathbb{1}_{X_\tau} \in V_k^0 \oplus \ldots \oplus V_k^j$. We prove something stronger, that $\mathbb{1}_{X_\tau}$ in fact lies almost entirely in V_k^j. In particular, we show that for every $i \neq j$, $\mathbb{1}_{X_\tau}$ has almost no projection onto V_k^i:

$$\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,i} \rangle \leq O_k(\gamma) \mathbb{E}[\mathbb{1}_{X_\tau}].$$

Since the HD-Level-Set Decomposition is approximately orthogonal [111] (more generally, this is true of any approximate eigendecomposition, see Lemma 7.6.3), this implies that the mass on level j is around $(1 \pm O_k(\gamma))\mathbb{E}[\mathbb{1}_{X_\tau}]$, and plugging these observations into Equation (7.1) we get:

$$\begin{aligned}
\phi(M, \mathbb{1}_S) &\approx 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=0}^{k} \lambda_i \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle \\
&\approx 1 - \frac{\lambda_i}{\mathbb{E}[\mathbb{1}_S]} \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle \\
&\approx 1 - \lambda_i
\end{aligned}$$

where we have ignored some $O_k(\gamma)$ error terms.

Proof sketch of Theorem 7.2.8:.

Proving that pseudorandom functions expand is a bit more involved, but still at its core revolves around the analysis of Equation (7.1). In particular, since the approximate eigenvalues of M monotonically decrease (i.e. $\forall i \geq j$, $\lambda_i(M) \leq \lambda_j(M)$) then for any $\delta > 0$

we can simplify Equation (7.1) to:

$$\phi(M, \mathbb{1}_S) \lesssim 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=0}^{R_\delta(M)-1} \lambda_i \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle + \delta, \tag{7.2}$$

where we recall $R_\delta(M)$ is the ST-rank, denoting the number of eigenstrips of M with eigenvalues greater than δ. Using Theorem 7.2.1, it is possible to show (see Proposition 7.7.11) that essentially all HD-walks of interest satisfy this property. With this in hand, proving Theorem 7.2.8 boils down to characterizing the projection of a set S onto low levels of the complex by its behavior on links.

Theorem 7.3.3 (Pseudorandomness bounds low-level weight (Informal Theorem 7.8.7))**.** *Let X be a sufficiently strong d-dimensional γ-local-spectral expander. For any $i \leq k \leq d$ let $S \subset X(k)$ be a (ε, i)-ℓ_2-pseudorandom set. Then for all $j \leq i$:*

$$\langle \mathbb{1}_S, \mathbb{1}_{S,j} \rangle \lesssim \binom{k}{i} \varepsilon \mathbb{E}[\mathbb{1}_S]$$

Theorem 7.2.8 follows immediately from plugging this result into Equation (7.2). Perhaps surprisingly, Theorem 7.3.3 itself follows without too much difficulty from combining analysis of the averaging operators and our spectral analysis of the HD-Level-Set Decomposition (namely Theorem 7.2.1). In particular, given a set $S \subset X(k)$, the idea is to examine the lowered indicator function $D_j^k \mathbb{1}_S$, which exactly gives the expectation of S over j-links, that is for any $\tau \in X(j)$:

$$D_j^k \mathbb{1}_S(\tau) = \mathop{\mathbb{E}}_{X_\tau}[\mathbb{1}_S].$$

Proving Theorem 7.2.8 then corresponds to lower bounding $\mathrm{Var}(D_j^k \mathbb{1}_S)$ by some function of the projection of S onto level j. In fact, it is possible to exactly understand this connection. The idea is to reduce the problem to a spectral analysis of HD-walks using the adjointness

of D_j^k and U_j^k (see e.g. [111]):

$$\begin{aligned}\mathrm{Var}(D_j^k \mathbb{1}_S) &= \langle D_j^k \mathbb{1}_S, D_j^k \mathbb{1}_S\rangle - \mathbb{E}[D_j^k \mathbb{1}_S]^2 \\ &= \langle \mathbb{1}_S, U_j^k D_j^k \mathbb{1}_S\rangle - \mathbb{E}[\mathbb{1}_S]^2 \\ &= \sum_{\ell=0}^{k} \langle \mathbb{1}_S, U_j^k D_j^k \mathbb{1}_{S,\ell}\rangle - \mathbb{E}[\mathbb{1}_S]^2 \\ &= \sum_{\ell=1}^{k} \langle \mathbb{1}_S, U_j^k D_j^k \mathbb{1}_{S,\ell}\rangle.\end{aligned}$$

Since $U_j^k D_j^k$ is an HD-walk, Theorem 7.2.1 implies that $U_j^k D_j^k \mathbb{1}_{S,\ell} \approx \lambda_\ell \mathbb{1}_{S,\ell}$ for some approximate eigenvalue λ_ℓ. In the formal version of Theorem 7.2.1 we compute the approximate eigenvalues of such walks, and in this case a simple computation shows that $\lambda_\ell = \frac{\binom{j}{\ell}}{\binom{k}{\ell}}$. As a result, we get a tight connection between $\mathrm{Var}(D_j^k \mathbb{1}_S)$ and the projection of S onto low levels of the complex:

$$\mathrm{Var}(D_j^k \mathbb{1}_S) \approx \sum_{\ell=1}^{k} \frac{\binom{j}{\ell}}{\binom{k}{\ell}} \langle \mathbb{1}_S, \mathbb{1}_{S,\ell}\rangle$$

where we have ignored factors in γ. The proof then follows from noting that the right-hand side is approximately lower bounded by $\frac{1}{\binom{k}{j}}\langle \mathbb{1}_S, \mathbb{1}_{S,j}\rangle$ by near orthogonality of the HD-Level-Set decomposition [111].

ℓ_∞ to ℓ_2 reduction:.

We end with a simple observation that generalizes our results to the ℓ_∞-variant characterization via reduction. The idea is that the 2-norm may be re-written as a (normalized) expectation over a modified distribution:

$$\frac{1}{\mathbb{E}[\mathbb{1}_S]}\mathrm{Var}(D_j^k \mathbb{1}_S) = \mathop{\mathbb{E}}_{\Pi_S}[D_j^k \mathbb{1}_S] - \mathbb{E}[\mathbb{1}_S] \leq \max(D_j^k \mathbb{1}_S - \mathbb{E}[\mathbb{1}_S])$$

where Π_S is the distribution on $X(j)$ naturally induced by restricting Π to the support of S (see Section 7.8 for further details). It immediately follows that any (ε, ℓ)-ℓ_∞-pseudorandom set is also (ε, ℓ)-ℓ_2-pseudorandom.

7.3.3 Playing Unique Games

We end the section with a discussion of Theorem 7.2.11, an application of our structural theorems to algorithms for unique games. We follow the algorithmic framework of [37] for playing unique games on Johnson graphs, a special case of HD-walks on the complete complex. We generalize their algorithm and its analysis to HD-walks over all (sufficiently strong) two-sided local spectral expanders by abstracting the algorithm into the broader local-to-global paradigm for HDX:

1. For some r, break the UG instance down into sub-instances over r-links of the complex.

2. Solve the local sub-instances and patch together the solutions into a good global solution.[14]

For their underlying algorithmic machinery, BBKSS rely heavily on a well-studied paradigm known as the Sum-of-Squares (SoS) semidefinite programming hierarchy. We won't go into too much detail about this paradigm here (see Section 7.10.1 for details), and for the moment it suffices to say that analysing SoS algorithms relies on porting the proofs of relevant inequalities (specific to the problem at hand) into the Sum-of-Squares proof system, a restricted proof system for proving polynomial inequalities. We leverage the BBKSS framework to solve unique games on HDX by observing that their algorithm's analysis can be generalized to rely on two core structural properties true of the graphs underlying HD-walks:

[14]This is only informally speaking, as the actual algorithm is iterative and repeats these two steps many times to obtain a good solution for the whole graph

1. A low-degree Sum-of-Squares proof that non-expanding sets have high variance in size across links.

2. For every small enough ϵ, the existence of some $r = r(\varepsilon)$ such that:

   (a) The $(r+1)$-st largest (distinct) stripped-eigenvalue of the constraint graph is small:

   $$\lambda_r \leq 1 - \Omega(\varepsilon)$$

   (b) The expansion of any r-link is small as well:

   $$\forall \tau \in X(r) : \Phi(X_\tau) \leq O(\varepsilon).$$

In essence, properties (1) and (2a) ensure that there are good local solutions at level r, and property (2b) ensures that these solutions can be patched together without too much loss. Given the above properties, the proof of Theorem 7.2.11 follows the BBKSS analysis framework, but is more technical as properties that hold for Johnson graphs generally only hold approximately for HD-walks.

The novelty in our analysis lies mostly in proving these two properties. Luckily, we have already done most of the work! Property (1) is simply an SoS variant of Theorem 7.2.8, which we prove by developing SoS versions of the now standard machinery for the HD-Level-Set Decomposition from [111]. The proof then follows essentially as discussed in Section 7.3.2. The second property is slightly more subtle. This parameter, found in [37] by direct computation on the Johnson graphs, determines both the soundness and runtime of their algorithm. In fact, our framework completely demystifies its existence: since λ_r and the expansion of r-links are inversely correlated by Theorem 7.2.6, r is exactly the ST-Rank of the underlying constraint graph! As a result, using these properties and generalizing the analysis of BBKSS results in Theorem 7.2.11: an algorithm for unique games on HD-walks with approximation and runtime guarantees dependent on ST-Rank.

Organization

In Section 7.4 we discuss related work on HD-walks, Unique Games, and CSP approximation. In Section 7.5 we give requisite background and notation. In Section 7.6 we prove that approximate eigendecompositions tightly control spectral structure. In Section 7.7, we show that the combinatorial decomposition of [111] is an approximate eigendecomposition for any HD-walk and compute its corresponding approximate spectra, thereby determining the spectral structure of HD-walks. In Section 7.8 we prove Theorem 7.3.3, showing how pseudorandomness controls projections onto any HD-walks eigendecomposition. In Section 7.9, we use this fact to give a tight characterization of edge expansion in HD-walks. Finally, in Section 7.10 we give a sum-of-squares variant of our results and show they imply efficient algorithms for unique games over HD-walks.

7.4 Related Work

Higher Order Random Walks:.

The spectral structure of higher order random walks has seen significant study in recent years, starting with the work of Kaufman and Oppenheim [239] who proved bounds on the spectra of N_k^1 on one-sided local-spectral expanders. Their result lead not only to the resolution of the Mihail-Vazirani conjecture [25], but to a number of further breakthroughs in sampling algorithms via a small but consequential improvement on their bound by Alev and Lau [11]. The spectral structure of N_k^1 on the stronger two-sided local-spectral expanders was further studied by DDFH [111] who introduced the HD-Level-Set Decomposition, and Kaufman and Oppenheim [239] who introduced a distinct approximate eigendecomposition with the benefit of orthogonality (though this came at the cost of additional combinatorial complexity). In recent work, Kaufman and Sharakanski [242] claim that these two decompositions are equivalent on sufficiently strong two-sided γ-local-spectral expanders, but their proof relies on [239, Theorem 5.10] which

has a non-trivial error. Indeed, it is possible to construct arbitrarily strong two-sided local-spectral expanders for which the HD-Level-Set Decomposition is not orthogonal (see Section 7.13), so their result cannot hold.[15] Finally, Alev, Jeronimo, and Tulsiani [9] showed that the HD-Level-Set Decomposition is an approximate eigendecomposition (in a weaker sense than we require) for general HD-walks, a result we strengthen in Section 7.7. For further information on these prior works and their applications, the interested reader should see [8].

Unique Games:.

The study of unique games has played a central role in hardness-of-approximation since Khot's [250] introduction of the Unique Games Conjecture. One line of work towards *refuting* the UGC focuses on building efficient algorithms for unique games for restricted classes of constraint graphs based off of spectral or spectrally-related properties; these include works employing spectral expansion [29, 293], threshold rank [28, 48, 189, 265], hypercontractivity [44], and certified small-set-expansion or characterized non-expansion [37]. Our work continues to expand this direction with polynomial-time algorithms for (affine) unique games over HD-walks and the introduction of ST-Rank. On the other hand, recent work towards *proving* the UGC has focused on characterizing non-expanding sets in structures such as the Grassmann [255, 253, 126, 125, 47, 252, 255] and Shortcode [47, 255] graphs. Our spectral framework based on HD-walks and the HD-Level-Set Decomposition provides a more general method to approach this direction than previous Fourier analytic machinery.

CSPs on HDXs:.

Finally, it is worth noting a related, recent vein of work connecting high dimensional expansion, Sum of Squares, and CSP-approximation. In particular, Alev, Jeronimo, and Tulsiani [9] recently showed that for $k > 2$, certain natural k-CSP's on two-sided local-

[15] It is worth noting that the main results of [239, 242] are unaffected by this error, as an approximate version of [239, Theorem 5.10] remains true and is sufficient for their purposes.

spectral expanders can be efficiently approximated by Sum of Squares. Conversely, Dinur, Filmus, Harsha, and Tulsiani [119] later used cosystolic expanders (a stronger variant) to build explicit instances of 3-XOR that are hard for SoS. Both these works focus on k-CSPs for $k \geq 3$, where the variables of the CSP are level 1 of the complex and the constraints are defined using the higher levels of the complex; therefore these works do not encompass unique games. At a high level, the techniques of [9] are based on [48], and they generalize the BRS algorithm for 2-CSPs to k-CSPs. While these works are not directly related to ours since their definition do not encompass unique games, we see a similar pattern where high dimensional expanding structure is useful both for hardness of and algorithms for CSP-approximation.

7.5 Preliminaries and Notation

7.5.1 Local-Spectral Expanders and Higher Order Random Walks

We now overview the theory of local-spectral expanders and higher order random walks in more formality than our brief treatment in Section 7.2.

Two-Sided Local-Spectral Expanders

Two-sided local-spectral expanders are a generalization of spectral expander graphs to weighted, uniform hypergraphs, which we will think of as simplicial complexes.

Definition 7.5.1 (Weighted, Pure Simplicial Complex)**.** A d-dimensional, pure simplicial complex X on n vertices is a subset of $\binom{[n]}{d}$. We will think of X as the downward closure of these sets, and in particular define the level $X(i)$ as:

$$X(i) = \left\{ s \in \binom{[n]}{i} \;\middle|\; \exists t \in X, s \subseteq t \right\}.$$

We call the elements of $X(i)$ i-faces.[16] A simplicial complex is *weighted* if its top level faces are endowed with a distribution Π. This induces a distribution over each $X(i)$ by downward closure:

$$\Pi_i(x) = \frac{1}{i+1} \sum_{y \in X(i+1): y \supset x} \Pi_{i+1}(y), \tag{7.3}$$

where $\Pi_d = \Pi$.

Two-sided local-spectral expanders are based upon a phenomenon called *local-to-global* structure, which looks to propogate information on local neighborhoods of a simplicial complex called *links* to the entire complex.

Definition 7.5.2 (Link)**.** Given a weighted, pure simplicial complex (X, Π), the **link** of an i-face $s \in X(i)$ is the sub-complex containing s, i.e.

$$X_s = \{t \setminus s \in X \mid t \supseteq s\}.$$

Π induces a distribution over X_s by normalizing over top-level faces which we denote by Π_s. When considering a function on the k-th level of a complex, we also use X_s to denote the k-faces which contain s as long as it is clear from context, and refer to X_s as an i-link if $s \in X(i)$.

Two-sided local-spectral expansion simply posits that the graph underlying every link[17] must be a two-sided spectral expander.

Definition 7.5.3 (Local-spectral expansion)**.** A weighted, pure simplicial complex (X, Π) is a two-sided γ-local-spectral expander if for every $i \leq d-2$ and every face $s \in X(i)$, the underlying graph of X_s is a two-sided γ-spectral expander.[18]

[16] We differ here from much of the HDX literature where an i-face is often defined to have $i+1$ elements. Since our work is mostly combinatorial rather than topological or geometric, defining an i-face to have i elements ends up being the more natural choice.

[17] The underlying graph of a simplicial complex X is its 1-skeleton $(X(1), X(2))$.

[18] A weighted graph $G(V, E)$ with edge weights Π_E is a two-sided γ-spectral expander if the vertex-edge-vertex random walk with transition probabilities proportional to Π_E has second largest eigenvalue in absolute value at most γ.

Higher Order Random Walks

Weighted simplicial complexes admit a natural generalization of the standard vertex-edge-vertex walk on graphs known as *higher order random walks* (HD-walks). The basic idea is simple: starting at some k-set $S \subset X(k)$, pick at random a set $T \in X(k+1)$ such that $T \supset S$, and then return to $X(k)$ by selecting some $S' \subset T$. Let the space of functions $f : X(k) \to \mathbb{R}$ be denoted by C_k. Formally, higher order random walks are a composition of two averaging operators: the "Up" operator which lifts a function $f \in C_k$ to $U_k f \in C_{k+1}$:

$$\forall y \in X(k+1) : U_k f(y) = \frac{1}{k+1} \sum_{x \in X(k) : x \subset y} f(x),$$

and the "Down" operator which lowers a function $f \in C_{k+1}$ to $D_{k+1} f \in C_k$:

$$\forall x \in X(k) : D_{k+1} f(x) = \frac{1}{k+1} \sum_{y \in X(k+1) : y \supset x} \frac{\Pi_{k+1}(y)}{\Pi_k(x)} f(y).$$

These operators exist for each level of the complex, and composing them gives a basic set of higher order random walks we call *pure* (following [9]).

Definition 7.5.4 (k-Dimensional Pure Walk)**.** Given a weighted, simplicial complex (X, Π), a k-dimensional pure walk $Y : C_k \to C_k$ on (X, Π) (of height $h(Y)$) is a composition:

$$Y = Z_{2h(Y)} \circ \cdots \circ Z_1,$$

where each Z_i is a copy of D or U.

We call an affine combination[19] of pure walks which start and end on $X(k)$ a *k-dimensional HD-walk*.

Definition 7.5.5 (HD-walk)**.** Let (X, Π) be a pure, weighted simplicial complex. Let $\mathcal{Y}$

[19] An affine combination is a linear combination whose coefficients sum to 1.

be a family of pure walks $Y : C_k \to C_k$ on (X, Π). We call an affine combination

$$M = \sum_{Y \in \mathcal{Y}} \alpha_Y Y$$

a k-dimensional HD-walk on (X, Π) as long as it is self-adjoint and remains a valid walk (i.e. has non-negative transition probabilities).

Previous work on HD-walks mainly focuses on two natural classes: canonical walks (introduced in [234, 124]), and partial-swap walks (introduced in [9, 109]).

Definition 7.5.6 (Canonical Walk)**.** Given a d-dimensional weighted, pure simplicial complex (X, Π), and parameters $k + j \leq d$, the canonical walk N_k^j is:

$$N_k^j = D_k^{k+j} U_k^{k+j},$$

where $U_\ell^k = U_{k-1} \dots U_\ell$, and $D_\ell^k = D_{\ell+1} \dots D_k$.

In other words, the canonical walk N_k^j takes j steps up and down the complex via the averaging operators. Partial-swap walks are a similar process, but after ascending the complex, we restrict to returning to faces with a given intersection from the starting point.

Definition 7.5.7 (Partial-Swap walk)**.** The partial-swap walk S_k^j is the restriction of N_k^j to faces with intersection $k - j$. In other words, if $|s \cap s'| \neq k - j, S_k^j(s, s') = 0$, and otherwise $S_k^j(s, s') = \alpha_s N_k^j(s, s')$, where

$$\alpha_s = \left(\sum_{s' : |s \cap s'| = k-j} N_k(s, s') \right)^{-1}$$

is the appropriate normalization factor.

It is not hard to see that partial-swap walk S_k^t on the complete complex $J(n,d)$ (all d-subsets of $[n]$ endowed with the uniform distribution) is exactly the Johnson graph $J(n,k,k-t)$. While it is not immediately obvious that the partial-swap walks are HD-walks, Alev, Jeronimo, and Tulsiani [9] showed this is the case by expressing them as an alternating hypergeometric sum of canonical walks.

Expansion of HD-Walks

In this work, we study the combinatorial edge expansion of HD-walks, a fundamental property of graphs with strong connections to many areas of theoretical computer science, including both hardness and algorithms for unique games. Given a weighted graph $G = ((V,E),(\Pi_V,\Pi_E))$ where Π_V is a distribution over vertices, and Π_E is a set of non-negative edge weights, the expansion of a subset $S \subset V$ is the average edge-weight leaving S.

Definition 7.5.8 (Weighted Edge Expansion)**.** Given a weighted, directed graph $G = ((V,E),(\Pi_V,\Pi_E))$, the weighted edge expansion of a subset $S \subset V$ is:

$$\phi(G,S) = \mathop{\mathbb{E}}_{v\sim\Pi_V|_S}\left[E(v,V\setminus S)\right],$$

where

$$E(v,V\setminus S) = \sum_{(v,y)\in E: y\in V\setminus S} \Pi_E((v,y))$$

is the total weight of edges between vertex v and the subset $V \setminus S$, and $\Pi_V|_S$ is the re-normalized restriction of Π_V to S. In the context of a k-dimensional HD-Walk M on a weighted simplicial complex (X,Π), we will always have $V = X(k)$, $\Pi_V = \Pi_k$, and E, Π_E given by M. Thus when clear from context, we will simply write $\phi(S)$.

Edge expansion in a weighted graph is closely related to the spectral structure of

its adjacency matrix. Given a set $S \subset V$ of density $\alpha = \mathbb{E}[\mathbb{1}_S]$, we may write

$$\phi(G, S) = 1 - \frac{1}{\alpha}\langle \mathbb{1}_S, A_G \mathbb{1}_S \rangle_{\Pi_V},$$

where A_G is the adjacency matrix with weights given by Π_E, and $\langle f, g \rangle_{\Pi_V}$ is the expectation of fg over Π_V. When considering such an inner product over a weighted simplicial complex (X, Π), the associated distribution will always be Π_k, so we will drop it from the corresponding notation. Notice that the right-hand side of this equivalence may be further broken down via a spectral decomposition of $\mathbb{1}_S$ with respect to A_G. Thus to understand the edge-expansion of HD-walks, it is crucial to understand the structure of their spectra.

7.6 Approximate Eigendecompositions and Eigenstripping

In this section we prove a general linear algebraic result concerning the spectra of operators that admit an approximate eigendecomposition: their spectra lies tightly concentrated around the decomposition's approximate eigenvalues. Before giving the formal result, we formalize the concept of approximate eigendecompositions.

Definition 7.6.1. Let M be an operator over an inner product space V. We call $V = V^1 \oplus \ldots \oplus V^k$ a $(\{\lambda_i\}_{i=1}^k, \{c_i\}_{i=1}^k)$-approximate eigendecomposition if for all i and $v_i \in V^i$, the following holds:

$$\|Mv_i - \lambda_i v_i\| \leq c_i \|v_i\|.$$

As long as the c_i are sufficiently small, we prove each V^i (loosely) corresponds to an eigenstrip, the span of eigenvectors with eigenvalue closely concentrated around λ_i.

Theorem 7.6.2 (Eigenstripping)**.** *Let M be a self-adjoint operator over an inner product space V, and $V = V^1 \oplus \ldots \oplus V^k$ a $(\{\lambda_i\}_{i=1}^k, \{c_i\}_{i=1}^k)$-approximate eigendecomposition. Let $c_{\max} = \max_i\{c_i\}$, $\lambda_{dif} = \min_{i,j}\{|\lambda_i - \lambda_j|\}$, and $\lambda_{ratio} = \frac{\max_i\{|\lambda_i|\}}{\lambda_{dif}^{1/2}}$. Then as long as $c_{\max}$ is*

sufficiently small:

$$c_{\max} \leq \frac{\lambda_{dif}}{4k},$$

the spectra of M is concentrated around each λ_i:

$$Spec(M) \subseteq \bigcup_{i=1}^{k} [\lambda_i - e, \lambda_i + e] = I_{\lambda_i},$$

where $e = O\left(k \cdot \lambda_{ratio} \cdot c_{max}^{1/2}\right)$.

This result was recently improved by Zhang [363] to have dependence $e \leq O(\sqrt{k}c_{\max})$, removing the dependence on the approximate eigenvalues altogether. This follows from a simple modification to our proof which we will note below. Finally, it is also worth mentioning that a version of Theorem 7.6.2 holds with no assumption on $c_{\max}$, but the assumption substantially simplifies the bounds and is sufficient for our purposes.

Before proving Theorem 7.6.2, we note a useful property of approximate eigendecompositions of self-adjoint operators: they are approximately orthogonal.

Lemma 7.6.3. *Let M be a self-adjoint operator over an inner product space V. Further, let $V = V^1 \oplus \ldots \oplus V^k$ be a $(\{\lambda_i\}_{i=1}^k, \{c_i\}_{i=1}^k)$-approximate eigen-decomposition. Then for $i \neq j$, V^i and V^j are nearly orthogonal. That is, for any $v_i \in V^i$ and $v_j \in V^j$:*

$$|\langle v_i, v_j \rangle| \leq \frac{c_i + c_j}{|\lambda_i - \lambda_j|} \|v_i\| \|v_j\|.$$

Proof. This follows from the fact that M is self-adjoint, and V^i and V^j are approximate eigenspaces. In particular, notice that for any $v_i \in V^i$ and $v_j \in V^j$ we can bound the interval in which $\langle Mv_i, v_j \rangle = \langle v_i, Mv_j \rangle$ lies by Cauchy-Schwarz:

$$\langle Mv_i, v_j \rangle \in \lambda_i \langle v_i, v_j \rangle \pm c_i \|v_i\| \|v_j\|$$

and

$$\langle v_i, Mv_j \rangle \in \lambda_j \langle v_i, v_j \rangle \pm c_j \|v_i\| \|v_j\|.$$

Since these terms are equal, the right-hand intervals must overlap. As a result we get:

$$|(\lambda_i - \lambda_j)\langle v_i, v_j \rangle| \leq (c_i + c_j)\|v_i\| \|v_j\|,$$

as desired. □

Using Lemma 7.6.3, we can modify [239, Theorem 5.9] to prove Theorem 7.6.2. Given an eigenvalue μ of M, the idea is to find a probability distribution over $[k]$ for which the expectation of $|\mu - \lambda_i|$ is small, where $i \in [k]$ is sampled from the aforementioned distribution.

Proof. The proof follows mostly along the lines of [239, Theorem 5.9], modifying where necessary due to lack of orthogonality. Let ϕ be an eigenvector of M with eigenvalue μ. Our goal is to prove the existence of some λ_i such that $|\mu - \lambda_i|$ is small. To do this, we appeal to an averaging argument. In particular, denoting the component of ϕ in V^i by ϕ_i, we bound the expectation of $|\mu - \lambda_i|^2$ over a distribution P_ϕ given by the (normalized) squared norms $\|\phi_i\|^2$:

$$\mathop{\mathbb{E}}_{i \sim P_\phi} \left[|\mu - \lambda_i|^2 \right] = \frac{1}{\sum_{j=1}^{k} \|\phi_j\|^2} \sum_{i=1}^{k} |\mu - \lambda_i|^2 \|\phi_i\|^2. \tag{7.4}$$

If we can upper bound this expectation by some value c, then by averaging there must exist λ_i such that $|\mu - \lambda_i| \leq \sqrt{c}$, and thus the spectra of M must lie in strips $\lambda_i \pm \sqrt{c}$. To upper bound Equation (7.4), consider the result of pushing the outer summation inside

the norm:

$$\sum_{i=1}^{k} |\mu - \lambda_i|^2 \|\phi_i\|^2 = \|\sum_{i=1}^{k} (\mu - \lambda_i)\phi_i\|^2 - \sum_{1 \le i \ne j \le k} (\mu - \lambda_i)(\mu - \lambda_j) \langle \phi_i, \phi_j \rangle . \tag{7.5}$$

We will separately bound the two resulting terms, the former by the fact that the ϕ_i are approximate eigenvectors, and the latter by their approximate orthogonality. We start with the former, which follows by a simple application of Cauchy-Schwarz:

$$\begin{aligned}
\|\sum_{i=1}^{k} (\mu - \lambda_i)\phi_i\|^2 &= \|\mu\phi - \sum_{i=1}^{k} \lambda_i \phi_i\|^2 \\
&= \|M\phi - \sum_{i=1}^{k} \lambda_i \phi_i\|^2 \\
&= \|\sum_{i=1}^{k} (M\phi_i - \lambda_i \phi_i)\|^2 \\
&\le k \sum_{i=1}^{k} \|(M\phi_i - \lambda_i \phi_i)\|^2 \\
&\le k c_{\max}^2 \sum_{i=1}^{k} \|\phi_i\|^2.
\end{aligned}$$

The latter takes a bit more effort. Let $\lambda_{\max}$ be $\max_i\{|\lambda_i|\}$, then by Lemma 7.6.3 we have:

$$\begin{aligned}
\left| \sum_{1 \le i \ne j \le k} (\mu - \lambda_i)(\mu - \lambda_j) \langle \phi_i, \phi_j \rangle \right| &\le \sum_{1 \le i \ne j \le k} |\mu - \lambda_i||\mu - \lambda_j| \frac{c_i + c_j}{|\lambda_i - \lambda_j|} \|\phi_i\| \|\phi_j\| \\
&\le 2c_{\max} \lambda_{\text{dif}}^{-1} (\lambda_{\max} + \|M\|)^2 \left(\sum_{i=1}^{k} \|\phi_i\| \right)^2 \\
&\le 2kc_{\max} \lambda_{\text{dif}}^{-1} (\lambda_{\max} + \|M\|)^2 \sum_{i=1}^{k} \|\phi_i\|^2
\end{aligned}$$

Since we'd like our bound to depend only on λ_i and c_i, we must further bound $\|M\|$ which

will follow similarly from approximate orthogonality. Let v be a unit eigenvector with eigenvalue $\|M\|$ and v_i be v's component on V^i, then we have:

$$\begin{aligned}
\|M\| &= \|Mv\| \\
&= \|\sum_{i=1}^{k} Mv_i - \lambda_i v_i + \lambda_i v_i\| \\
&\leq \sum_{i=1}^{k} (\lambda_i + c_i)\|v_i\| \\
&\leq (\lambda_{\max} + c_{\max}) \sum_{i=1}^{k} \|v_i\| \\
&\leq (\lambda_{\max} + c_{\max}) \sqrt{k \sum_{i=1}^{k} \|v_i\|^2} \\
&\leq (\lambda_{\max} + c_{\max}) \sqrt{2k}.
\end{aligned}$$

where the last step follows from Lemma 7.6.3 and our assumption on $c_{\max}$:

$$\begin{aligned}
\sum_{i=1}^{k} \|v_i\|^2 &= \|v\|^2 + \sum_{1 \leq i \neq j \leq k} \langle v_i, v_j \rangle \\
&\leq 1 + \frac{2c_{\max}}{\lambda_{\text{dif}}} \sum_{1 \leq i \neq j \leq k} \|v_i\| \|v_j\| \\
&\leq 1 + \frac{2c_{\max}}{\lambda_{\text{dif}}} \left(\sum_{i=1}^{k} \|v_i\| \right)^2 \\
&\leq 1 + \frac{2kc_{\max}}{\lambda_{\text{dif}}} \sum_{i=1}^{k} \|v_i\|^2 \\
&\leq 1 + \frac{1}{2} \sum_{i=1}^{k} \|v_i\|^2.
\end{aligned}$$

Together, these bounds imply the existence of some $\lambda_{i'}$ such that:

$$|\mu - \lambda_{i'}| \leq \sqrt{kc_{\max} \left(c_{\max} + 2\lambda_{\text{dif}}^{-1} \left(\lambda_{\max} + (\lambda_{\max} + c_{\max}) \sqrt{2k} \right)^2 \right)},$$

which implies the desired result when accounting for our assumption on $c_{\max}$. □

Zhang's [363] improvement to our proof came from the observation that the analysis of the latter term can be simplified by a recursive strategy. In particular, this term can instead be upper bounded by $\frac{2c_{\max}k}{\lambda_{\text{dif}}} \sum_{i=1}^{k} |\mu - \lambda_i|^2 \|\phi_i\|^2$. With the appropriate assumption on $c_{\max}$, plugging this back into Equation (7.5) gives the desired result.

In either case, notice that if $c_{\max}$ is sufficiently small, the intervals I_{λ_i} are disjoint. As a result, each V^i corresponds to an *eigenstrip* W^i:

$$W^i = \text{Span}\left\{\phi : \ M\phi = \mu\phi, \mu \in I_{\lambda_i}\right\}.$$

The approximate eigenspaces V^i are closely related to the resulting eigenstrips. Indeed, it is possible to show that most of the weight of a function in V^i must lie on W^i, though we will not need this result in what follows. Previous works [239, 242] make stronger claims for the specific case of the HD-Level-Set Decomposition, most notably that V^i and W^i are in fact equivalent on sufficiently strong two-sided local-spectral expanders. Unfortunately, these results are based off of [239, Theorem 5.10], whose proof has a non-trivial error we discuss further in Section 7.13. Indeed, were their proof correct, it would imply (due to the generality of their argument) that $V^i = W^i$ for any approximate eigendecomposition. However, it is easy to see this cannot be the case by considering a diagonal 2×2 matrix with an approximate eigendecomposition given by a slight rotation of the standard basis vectors in $\mathbb{R}^2$.

7.7 The Spectra of HD-walks

We now show that the HD-Level-Set Decomposition is an approximate eigendecomposition for any HD-Walk. Combined with Theorem 7.6.2, this proves Theorem 7.2.1, that the spectrum of any k-dimensional HD walk is tightly concentrated in $k+1$ eigenstrips. As

a result, we give explicit bounds on the spectra of HD-walks, paying special attention to the canonical and partial-swap walks. Finally, we show that the approximate eigenvalues (and thus the values in their corresponding eigenstrips) of the HD-Level-Set Decomposition decrease monotonically for a broad class of HD-Walks we call *complete* walks which, to our knowledge, encompass all walks used in the literature. As we will see in the following section, such decay is crucial for understanding edge expansion.

To start, we recall the definition of pure and HD-walks along with introducing some useful notation.

Definition 7.7.1 (k-Dimensional Pure Walk)**.** Given a weighted, simplicial complex (X, Π), a k-dimensional pure walk $Y : C_k \to C_k$ on (X, Π) is a composition:

$$Y = Z_{2h(Y)} \circ \cdots \circ Z_1,$$

where each Z_i is a copy of D or U, and $h(Y)$ is the *height* of the walk, measuring the total number of down (or up) operators.

Definition 7.7.2 (k-Dimensional HD-Walk)**.** Given a weighted, simplicial complex (X, Π), a k-dimensional HD-walk on (X, Π) is an affine combination of pure walks

$$M = \sum_{Y \in \mathcal{Y}} \alpha_Y Y$$

which is self-adjoint and gives a valid walk on (X, Π) (i.e. has non-negative transition probabilities). We say the height of M, $h(M)$, is the maximal height of any Y with a non-zero coefficient, and say the weight of M, $w(M)$, is the one norm of the α_Y (namely, $w(M) = \sum |\alpha_Y|$).

Our proofs in this section rely mainly on a useful observation of [111], who show that the up and down operators on two-sided γ-local-spectral expanders satisfy the following

relation:

$$\|D_{i+1}U_i - \frac{1}{i+1}I - \frac{i}{i+1}U_{i-1}D_i\| \leq \gamma. \tag{7.6}$$

This fact leads to a particularly useful structural lemma showing the effect of flipping D through multiple U operators.

Lemma 7.7.3 (Claim 8.8 [111]). *Let (X, Π) be a d-dimensional γ-local-spectral expander. Then for all $j < k < d$:*

$$\|D_{k+1}U_{k-j}^{k+1} - \frac{j+1}{k+1}U_{k-j}^{k} - \frac{k-j}{k+1}U_{k-j-1}^{k}D_{k-j}\| \leq \frac{(j+1)(2k-j+2)}{2(k+1)}\gamma.$$

One crucial application of Lemma 7.7.3 lies in understanding the relation between $\|f_i\|$, and $\|g_i\|$, where $f_i = U_i^k g_i$.

Lemma 7.7.4 (Lemmas 8.10, 8.13, Theorem 4.6 [111]). *Let (X, Π) be a d-dimensional γ-local-spectral expander with $\gamma < 1/d$, $f \in C_k$ a function with HD-Level-Set Decomposition $f_0 + \ldots + f_k$. Then for all $0 \leq \ell \leq k \leq d$:*

$$\|f_\ell\|^2 = \frac{1}{\binom{k}{\ell}}\left(1 \pm c_1(k,\ell)\gamma\right)\|g_\ell\|^2,$$

where $c_1(k,\ell) = O(k^2\binom{k}{\ell})$.

In Section 7.12, we prove a stronger version of both Lemma 7.7.3 and Lemma 7.7.4 for $\gamma \leq 2^{-\Omega(k)}$ where the dependence on the first order term γ is polynomial rather than exponential in k. However, since this only provides a substantial improvement for a small range of γ, we use the simpler versions from [111] throughout the body of the paper. Using Lemma 7.7.3 and Lemma 7.7.4, an inductive argument shows that the HD-Level-Set Decomposition is an approximate eigendecomposition. We show this first for the basic

case of a pure walk, and then note that the general result follows immediately from the triangle inequality.

Proposition 7.7.5. *Let (X, Π) be a two-sided γ-local-spectral expander with $\gamma \leq 2^{-\Omega(k)}$ and $Y : C_k \to C_k$ a pure walk:*

$$Y = Z_{2h(Y)} \circ \cdots \circ Z_1.$$

Let $i_1 \leq \ldots \leq i_{h(Y)}$ denote the $h(Y)$ indices at which Z_i is a down operator. Then for all $0 \leq \ell \leq k, f \in V_k^\ell$:

$$\|Yf - \prod_{s=1}^{h(Y)} \left(1 - \frac{\ell}{\max\{\ell, i_s - 2s + k + 1\}}\right) f\| \leq O\left(\gamma h(Y)(k + h(Y)) \binom{k}{\ell} \|f\|\right).$$

Proof. We prove a slightly stronger statement to simplify the induction. For $b > 0$, let $Y_j^b : C_\ell \to C_{\ell+b}$ denote an unbalanced walk with j down operators, and $j + b$ up operators. If Y_j^b has down operators in positions $i_1 \leq \ldots \leq i_j$ and $g_\ell \in H^\ell$, we claim:

$$\|Y_j^b g_\ell - \prod_{s=1}^{j} \left(1 - \frac{\ell}{\max\{\ell, i_s - 2s + \ell + 1\}}\right) U_\ell^{b+\ell} g_\ell\| \leq \gamma j(b + j)\|g_\ell\|. \tag{7.7}$$

Notice that since $f \in V_k^\ell$ may be written as $U_\ell^k g_\ell$ for $g_\ell \in H^\ell$, then we may write Yf as $Y_{h(Y)}^{k-\ell} g_\ell$ where $Y_{h(Y)}^{k-\ell}$ has down operators in positions $i_1 + k - \ell \leq \ldots \leq i_j + k - \ell$. Combining Equation (7.7) with Lemma 7.7.4 then implies the result.

We prove Equation (7.7) by induction. The base case $j = 0$ is trivial. Assume the inductive hypothesis holds for all $Y_i^b, i < j$. Notice first that if $i_1 = 1$, we are done since $g_\ell \in H^\ell$, and

$$\prod_{s=1}^{j} \left(1 - \frac{\ell}{\max\{\ell, i_s - 2s + \ell + 1\}}\right) Y_0^b g_\ell = 0,$$

as $i_s - 2s + \ell + 1 = \ell$ for $s = 1$. Otherwise, it must be the case that one or more copies of the up operator appear before the first down operator, and we may therefore apply

Lemma 7.7.3 to get:

$$Y_j^b g_\ell = \left(\frac{i_1 - 1}{i_1 + \ell - 1}\right) Y_{j-1}^b g_\ell + \Gamma g_\ell,$$

where we can (loosely) bound the spectral norm of Γ by

$$\|\Gamma\| \leq (b + j)\gamma$$

since at worst the first down operator D passes through $b + j$ up operators. By the form of Lemma 7.7.3, Y_{j-1}^b has down operators at indices $i_2 - 2 \leq \ldots \leq i_j - 2$. Then by the fact that $i_1 + \ell - 1 > \ell$ and the inductive hypothesis:

$$\begin{aligned}
Y_j^b g_\ell &= \left(\frac{i_1 - 1}{\max\{\ell, i_1 + \ell - 1\}} \prod_{s=1}^{j-1} \frac{i_{s+1} - 2s - 1}{\max\{\ell, i_{s+1} - 2s + \ell - 1\}}\right) Y_0^b g_\ell + \frac{i_1 - 1}{i_1 + \ell - 1} h + \Gamma g_\ell \\
&= \left(\frac{i_1 - 1}{\max\{\ell, i_1 + \ell - 1\}} \prod_{s=2}^{j} \frac{i_s - 2s + 1}{\max\{\ell, i_s - 2s + \ell + 1\}}\right) Y_0^b g_\ell + \frac{i_1 - 1}{i_1 + \ell - 1} h + \Gamma g_\ell \\
&= \left(\prod_{s=1}^{j} \frac{i_s - 2s + 1}{\max\{\ell, i_s - 2s + \ell + 1\}}\right) Y_0^b g_\ell + \frac{i_1 - 1}{i_1 + \ell - 1} h + \Gamma g_\ell,
\end{aligned}$$

where $\|h\| \leq \gamma(j-1)(b+j-1)\|g_\ell\|$ and we have used the (vacuous) fact that $\max\{\ell, i_1 + \ell - 1\} = i_1 + \ell - 1$. Finally, we can bound the norm of the right-hand error term by:

$$\begin{aligned}
\left\|\frac{i_1 - 1}{i_1 + \ell - 1} h + \Gamma g_\ell\right\| &\leq \|h\| + \|\Gamma\|\|g_\ell\| \\
&\leq (j-1)(b+j-1)\|g_\ell\| + (b+j)\|g_\ell\| \\
&\leq j(b+j)\|g_\ell\|
\end{aligned}$$

as desired. □

Since HD-walks are simply affine combinations of pure walks, the triangle inequality immediately implies the result carries over to this more general setting.

Corollary 7.7.6. *Let (X, Π) be a two-sided γ-local-spectral expander with $\gamma \leq 2^{-\Omega(k)}$ and $M = \sum_i \alpha_i Y_i$ a k-dimensional HD-walk on (X, Π). Then for all $0 \leq \ell \leq k, f \in V_k^\ell$:*

$$\|Mf - \lambda_\ell(M)f\| \leq O\left(\gamma w(M)h(M)(k + h(M))\binom{k}{\ell}\|f\|\right),$$

where

$$\lambda_\ell(M) = \sum \alpha_i \lambda_\ell(Y_i)$$

and $\lambda_\ell(Y_i)$ is the approximate eigenvalue of Y_i given in Proposition 7.7.5.

It is worth noting that the resulting approximate eigenvalues in Corollary 7.7.6 are exactly the eigenvalues of M when considered on a sequentially differential poset with $\vec{\delta}_i = i/(i+1)$. We discuss this generalization in more depth and give tighter bounds on the approximate spectra in our upcoming companion paper. It should be noted that this result is similar to one appearing in [9], where a weaker notion of approximate eigenspaces based on the quadratic form $\langle f, Mf\rangle$ is analyzed. Plugging Corollary 7.7.6 into Theorem 7.6.2, we immediately get that for small enough γ the true spectra of HD-walks lie in strips around each $\lambda_i(M)$, and thus that that the approximate eigenvalues of the HD-Level-Set Decomposition and the spectra of HD-walks are essentially interchangeable.

For concreteness, we now turn our attention to computing the approximate eigenvalues (and thereby the true spectra) of the canonical and swap walks.

Corollary 7.7.7 (Spectrum of Canonical Walks)**.** *Let (X, Π) be a d-dimensional γ-local-spectral expander with γ satisfying $\gamma \leq 2^{-\Omega(k+j)}$, $k + j \leq d$, and $f_\ell \in V_k^\ell$. Then:*

$$\|N_k^j f_\ell - \frac{\binom{k}{\ell}}{\binom{k+j}{\ell}} f_\ell\| \leq c(k, \ell, j)\|f_\ell\|,$$

where $c(k,\ell,j) = O\left(\gamma j(j+k)\binom{k}{\ell}\right)$. *Moreover:*

$$Spec(N_k^j) = \{1\} \cup \bigcup_{j=1}^{k} \left[\frac{\binom{k}{\ell}}{\binom{k+j}{\ell}} \pm 2^{O(j+k)}\sqrt{\gamma} \right].$$

Proof. By Proposition 7.7.5, N_k^j is an $\left(\{\lambda_\ell\}_{\ell=0}^k, \{c(k,\ell,j)\}_{\ell=0}^k\right)$-approximate eigendecomposition for

$$\lambda_\ell = \prod_{s=1}^{j} \left(1 - \frac{\ell}{\max\{k-2s+i_s+1, \ell\}}\right),$$

where $i_1 \leq \ldots \leq i_s$ denote the indices of down operators. By the definition of N_k^j we have $i_s = j+s$, and therefore

$$\lambda_\ell = \prod_{s=1}^{j} \left(1 - \frac{\ell}{k-s+j+1}\right) = \frac{\binom{k}{\ell}}{\binom{k+j}{\ell}}$$

as desired. The bounds on $\mathrm{Spec}(N_k^j)$ follow immediately from plugging the above into Theorem 7.6.2. □

A priori, it is not obvious how to bound the spectra of the partial-swap walks, or indeed even that they are HD-walks. However, Alev, Jeronimo, and Tulsiani [9] proved that partial-swap walks may be written as an alternating hypergeometric sum of canonical walks.

Proposition 7.7.8 (Corollary 4.13 [9])**.** *Let* (X,Π) *be a two-sided* γ*-local-spectral expander with* $\gamma < 1/k$*. Then for* $0 \leq j \leq k$*:*

$$S_k^j = \frac{1}{\binom{k}{k-j}} \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} \binom{k+i}{i} N_k^i.$$

As a result, we can use Corollary 7.7.7 to bound their approximate eigenvalues and true spectrum.

Corollary 7.7.9. *Let X be d-dimensional two-sided γ-local-spectral expander, $\gamma < 2^{-\Omega(k)}$, $k + j \leq d$, and $f_\ell \in V_k^\ell$. Then:*

$$\|S_k^j f_\ell - \frac{\binom{k-j}{\ell}}{\binom{k}{\ell}} f_\ell\| \leq c(k)\|f_\ell\|,$$

where $c(k) = \gamma 2^{O(k)}$. Moreover,

$$Spec(S_k^j) = \{1\} \cup \bigcup_{j=1}^{k} \left[\frac{\binom{k-j}{\ell}}{\binom{k}{\ell}} \pm 2^{O(k)}\sqrt{\gamma} \right].$$

Proof. By Corollary 7.7.6, $\bigoplus_{\ell=0}^{k} V_k^\ell$ is a $\left(\{\lambda_\ell\}_{\ell=0}^k, \{c'(k,\ell,j)\}_{\ell=0}^k\right)$-approximate eigendecomposition for S_k^j with

$$\begin{aligned}
\lambda_\ell &= \frac{1}{\binom{k}{k-j}} \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} \binom{k+i}{i} \lambda_\ell(N_k^i) \\
&= \frac{1}{\binom{k}{k-j}} \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} \binom{k+i}{i} \frac{\binom{k}{\ell}}{\binom{k+i}{\ell}} \\
&= \frac{1}{\binom{k}{k-j}} \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} \binom{k-\ell+i}{i} \\
&= \frac{\binom{k-\ell}{j}}{\binom{k}{k-j}} \\
&= \frac{\binom{k-j}{\ell}}{\binom{k}{\ell}},
\end{aligned}$$

and

$$c'(k,\ell,j) = \gamma 2^{O(k)}.$$

This latter fact follows from noting that

$$\|\vec{\alpha}\|_1 = \sum_{i=0}^{j} \binom{j}{i}\binom{k+i}{i} \leq 2^{2j+k}$$

where $\vec{\alpha}$ consists of the hypergeometric coefficients of Proposition 7.7.8. The bounds on $\mathrm{Spec}(S_k^j)$ then follow from Theorem 7.6.2. □

Together, Corollary 7.7.7 and Corollary 7.7.9 prove Theorem 7.2.4 (assuming γ is sufficiently small).

In Theorem 7.2.1, we mentioned that approximate eigenvalues $\lambda_i(M)$ are monotonically decreasing for any HD-walk. In fact, to prove this we will need to restrict our original definition of HD-walks slightly, requiring that our walks are always well-defined on the complete complex.

Definition 7.7.10 (Complete HD-Walk)**.** Let (X, Π) be a weighted, pure simplicial complex and $M = \sum_{Y \in \mathcal{Y}} \alpha_Y Y$ an HD-walk on (X, Π). We call M *complete* if for all $n \in \mathbb{N}$ there exist $n_0 > n$ and d such that $\sum_{Y \in \mathcal{Y}} \alpha_Y Y$ is also an HD-walk when taken to be over $J(n_0, d)$.

To our knowledge, all walks considered in the literature (pure, canonical, partial-swap) are complete. We can prove that the eigenstrips of complete HD-walks corresponding to the HD-Level-Set Decomposition exhibit eigenvalue decay by noting that the approximate eigenvalues of Corollary 7.7.6 are independent of the underlying complex.

Proposition 7.7.11. *Let (X, Π) be a two-sided γ-local-spectral expander, $M = \sum_{Y \in \mathcal{Y}} \alpha_Y Y$ a complete HD-walk over (X, Π), and γ small enough to apply the conditions of Theorem 7.6.2. Then for all $0 \leq i < j \leq k$,*

$$\lambda_i(M) \geq \lambda_j(M)$$

Proof. The proof follows from two observations. First, recall from Corollary 7.7.6 that $\lambda_i(M)$ is independent of the underlying complex. Second, any HD-walk on the complete

complex can be written as a non-negative sum of partial-swap walks, which satisfy the monotonic decrease property. Let $n \in \mathbb{N}$ be any parameter such that applying $\sum_{Y \in \mathcal{Y}} \alpha_Y Y$ to $J(n,d)$ results in a valid walk (i.e. a non-negative matrix). By the symmetry of $J(n,d)$, the transition probabilities of this walk depends only on size of intersection, and it may thus be written as some convex combination of partial-swap walks:

$$M = \sum_{Y \in \mathcal{Y}} \alpha_Y Y = \sum_{i=0}^{k} \beta_i S_k^i.$$

Since these walks are equivalent over $J(n,d)$, their spectra must match. Then by Theorem 7.6.2, it must be the case that for every $1 \leq \ell \leq k$ and n sufficiently large, the intersection of $\sum_{Y \in \mathcal{Y}} \alpha_Y \lambda_\ell(Y) \pm O(1/n)$ and $\sum_{Y \in \mathcal{Y}} \beta_i \lambda_\ell(S_k^i) \pm O(1/n)$ is non-empty. Since we may take n arbitrarily large, this implies the two quantities are in fact equivalent. Finally, by Corollary 7.7.9 $\lambda_\ell(S_k^i)$ decreases monotonically in ℓ for all i, which implies that the $\lambda_i(M) = \sum \alpha_Y \lambda_i(Y) = \sum \beta \lambda_i(S_k^j)$ decrease monotonically as desired. □

7.8 Pseudorandomness and the HD-Level-Set Decomposition

Now that we have examined the spectral structure of the HD-Level-Set Decomposition, we turn to understanding its combinatorial characteristics. In this section, we give a combinatorial characterization how arbitrary functions project onto the HD-Level-Set decomposition, proving in particular a generalization of Theorem 7.3.3: pseudorandom sets have bounded projection onto corresponding levels of the complex. We discuss two variants of pseudorandomness, an ℓ_2-variant stating that the variance across links is small, and an ℓ_∞-variant stating that the maximum (or ℓ_∞-norm for arbitrary functions) across links is small. We focus mainly on giving an exact analysis for the ℓ_2-case, as this forms the structural core of our algorithm for unique games in Section 7.10. We further discuss the implications of our ℓ_2 analysis to the ℓ_∞-variant by reduction.

As such, we'll start by analyzing the ℓ_2-variant. First, let's extend our definition of ℓ_2-pseudorandomness from sets (boolean functions) to arbitrary functions.

Definition 7.8.1 (ℓ_2-Pseudorandom functions)**.** A function $f \in C_k$ is $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_2$-pseudorandom if its variance across i-links is small for all $1 \leq i \leq \ell$:

$$Var(D_i^k f) \leq \varepsilon_i |\mathbb{E}[f]|$$

As mentioned in Section 7.3, the key to connecting $\mathrm{Var}(D_i^k f)$ and the HD-Level-Set Decomposition is to notice that by the adjointness of D and U, we can reduce the problem to analyzing the spectral structure of HD-walks. The proof then follows immediately from arguments in the previous section.

Theorem 7.8.2. *Let (X, Π) be a γ-local-spectral expander with $\gamma \leq 2^{-\Omega(k)}$ and let $f \in C_k$ have HD-Level-Set Decomposition $f = f_0 + \ldots + f_k$. Then for any $\ell \leq k$, the $Var(D_\ell^k f)$ is controlled by its projection onto $V_k^0 \oplus \ldots \oplus V_k^\ell$ in the following sense:*

$$Var(D_\ell^k f) \in \sum_{j=1}^{\ell} \frac{\binom{\ell}{j}}{\binom{k}{j}} \langle f, f_j \rangle \pm c\gamma \langle f, f \rangle$$

where $c \leq 2^{O(k)}$.

Proof. To start, notice that since $\langle D_k^\ell f, D_k^\ell f \rangle = \langle U_\ell^k D_\ell^k f, f \rangle$ it is enough to analyze the application of the HD-walk $U_\ell^k D_\ell^k$ to f. By Proposition 7.7.5, we know that each f_j is an approximate eigenvector satisfying:

$$\|U_\ell^k D_\ell^k f_j - \frac{\binom{\ell}{j}}{\binom{k}{j}} f_j\| \leq c_1 \gamma \|f\|,$$

where $c_1 \leq 2^{O(k)}$. Combining these observations gives:

$$\left\langle D_\ell^k f, D_\ell^k f \right\rangle = \left\langle f, U_\ell^k D_\ell^k f \right\rangle$$

$$= \sum_{j=0}^{k} \langle f, U_\ell^k D_\ell^k f_j \rangle$$
$$\in \sum_{j=0}^{\ell} \frac{\binom{\ell}{j}}{\binom{k}{j}} \langle f, f_j \rangle \pm c\gamma \langle f, f \rangle$$

where all constants $c \leq 2^{O(k)}$, and noting that $\langle f, f_0 \rangle = \mathbb{E}[f]^2$ completes the result. $\square$

Theorem 7.3.3 follows immediately from combining this result with approximate orthogonality of the HD-Level-Set Decomposition.

Lemma 7.8.3 (DDFH Theorem 4.6)**.** *Let (X, Π) be a γ-local-spectral expander with $\gamma \leq 2^{-\Omega(k)}$ and let $f \in C_k$ have HD-Level-Set Decomposition $f = f_0 + \ldots + f_k$. Then for all $i \neq j$:*

$$|\langle f_i, f_j \rangle| \leq c_1 \gamma \langle f, f \rangle$$

where $c_1 \leq 2^{O(k)}$.

Corollary 7.8.4. *Let (X, Π) be a γ-local-spectral expander with $\gamma \leq 2^{-\Omega(k)}$ and let $f \in C_k$ be an $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_2$-pseudorandom function. Then for any $1 \leq i \leq \ell$:*

$$|\langle f, f_i \rangle| \leq \binom{k}{i} \varepsilon_i |\mathbb{E}[f]| + c\gamma \langle f, f \rangle$$

where $c \leq 2^{O(k)}$.

Proof. By Lemma 7.8.3, for all j we have $\langle f, f_j \rangle \geq -c\gamma \langle f, f \rangle$ for some $c \leq 2^{O(k)}$. Then by Theorem 7.8.2, for all $0 \leq i \leq k$ the variance of $D_i^k f$ is lower bounded by the projection onto f_i:

$$\mathrm{Var}(D_i^k f) \geq \frac{1}{\binom{k}{i}} \langle f, f_i \rangle - c_2 \gamma \langle f, f \rangle,$$

where $c_2 \leq 2^{O(k)}$. Finally, since f is $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_2$-pseudorandom, we have by definition that for all $1 \leq i \leq \ell$, $\mathrm{Var}(D_i^k f) \leq \varepsilon_i |\mathbb{E}[f]|$. Therefore isolating $\langle f, f_i \rangle$ gives the desired

upper bound:

$$\begin{aligned}\langle f, f_i\rangle &\leq \binom{k}{i}\mathrm{Var}(D_i^k f) + c_3\gamma\langle f, f\rangle \\ &\leq \binom{k}{i}\varepsilon_i|\mathbb{E}[f]| + c_3\gamma\langle f, f\rangle\end{aligned}$$

where $c_3 \leq 2^{O(k)}$. Finally, we already noted that $\langle f, f_i\rangle \geq -c\gamma\langle f, f\rangle$ for some $c \leq 2^{O(k)}$, which gives the desired lower bound and completes the proof. □

It is worth noting that both Theorem 7.8.2 and Corollary 7.8.4 are tight. This is obvious for the former which is a near-equality, and the latter is clearly tight for subsets like i-links which project (almost) entirely onto level i (we'll prove this formally in the next section).

We'll also see in Section 7.9 that Corollary 7.8.4 leads to a tight ℓ_2-characterization of edge-expansion stating that any non-expanding set must have high variance across links. Since an ℓ_∞-variant of this result for the Grassmann graphs was recently crucial for the resolution of the 2-2 Games Conjecture [255], it is natural to discuss what our results imply for this regime. First, let's formalize ℓ_∞-pseudorandomness for arbitrary functions.

Definition 7.8.5 (ℓ_∞-Pseudorandom functions)**.** A function $f \in C_k$ is $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_\infty$-pseudorandom if for all $1 \leq i \leq \ell$ its local expectation is close to its global expectation:

$$\left\|D_i^k f - \mathbb{E}[f]\right\|_\infty \leq \varepsilon_i.$$

We will prove via reduction to the ℓ_2-variant that a version of Corollary 7.8.4 holds in this regime as well. We proceed in two steps. First, we show that ℓ_∞-pseudorandom functions are also ℓ_2-pseudorandom assuming a weak local-consistency property.

Definition 7.8.6. Let (X, Π) be a weighted, pure simplicial complex. We say a function $f \in C_k$ has ℓ-local constant sign if:

1. $\mathbb{E}[f] \neq 0$,

2. $\forall s \in X(\ell)$ s.t. $\underset{X_s}{\mathbb{E}}[f] \neq 0 : \text{sign}\left(\underset{X_s}{\mathbb{E}}[f]\right) = \text{sign}\left(\mathbb{E}[f]\right)$.

Second, we'll reduce to the case of locally constant sign by noting that we can always shift a function to satisfy this property. With these definitions in hand, we can now state the ℓ_∞-variant of Theorem 7.3.3:

Theorem 7.8.7. *Let (X, Π) be a γ-local-spectral expander with $\gamma \leq 2^{-\Omega(k)}$ and let $f \in C_k$ have HD-Level-Set Decomposition $f = f_0 + \ldots + f_k$. If f is $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_\infty$-pseudorandom, then for all $1 \leq i \leq \ell$:*

$$|\langle f, f_i \rangle| \leq \left(\binom{k}{i} + c(k)\gamma\right)\varepsilon_i^2 + c(k)\gamma\|f\|^2,$$

where $c(k) \leq 2^{O(k)}$, and if f has i-local constant sign:

$$|\langle f, f_i \rangle| \leq \binom{k}{i}\varepsilon_i|\mathbb{E}[f]| + c(k)\gamma\|f\|^2.$$

In dealing with expansion, we will mainly be interested in boolean-valued functions, which always have locally-constant sign and satisfy $\langle f, f \rangle = \mathbb{E}[f]$. Thus in the boolean case we have:

$$\langle f, f_i \rangle \leq \left(\binom{k}{i}\varepsilon_i + 2^{O(k)}\gamma\right)\mathbb{E}[f],$$

which is particularly useful since the expansion of f may be written as $1 - \frac{1}{\mathbb{E}[f]}\langle f, Mf \rangle$. In the case f is non-negative, we can also replace the ℓ_∞ norm with maximum in our definition of pseudorandomness, which is important to show non-expanding sets are locally *denser* than expected. Finally, we note that one can improve the dependence on γ by pushing exponential dependence on k to the second order γ^2 term via more careful analysis of error propagation. However since the analysis is complicated and only gives a substantial improvement for a small range of relevant γ, we relegate such discussion to Section 7.12.

We now move to the proof of Theorem 7.8.7, starting with our generic ℓ_∞ to ℓ_2 reduction for functions with locally-constant sign.

Lemma 7.8.8. *Let (X, Π) be a weighted, pure simplicial complex, and $f \in C_k$ a $(\varepsilon_1, \dots, \varepsilon_\ell)$-$\ell_\infty$-pseudorandom function with i-local constant sign for any $i \leq \ell$. Then f is also $(\varepsilon_1, \dots, \varepsilon_\ell)$-$\ell_2$-pseudorandom*

Proof. For ease of notation, let $\Pi_k(X_s)$ be shorthand for $\sum_{t \in X_s} \Pi_k(t)$ (i.e. the normalization factor for the above restricted expectation). The trick is to notice that since f has locally constant sign, we may rewrite $\|D_i^k f\|_2^2$ as an expectation over a related distribution P_i:

$$\begin{aligned}
\frac{1}{\mathbb{E}[f]}\langle D_i^k f, D_i^k f\rangle &= \sum_{s \in X(i)} \Pi_i(s) \left(\frac{1}{\mathbb{E}[f]} \sum_{t \in X_s} \frac{\Pi_k(t) f(t)}{\Pi_k(X_s)} \right) D_i^k f(s) \\
&= \sum_{s \in X(i)} \left(\frac{1}{\mathbb{E}[f]} \sum_{t \in X_s} \frac{\Pi_k(t) f(t)}{\binom{k}{i}} \right) D_i^k f(s) \\
&= \mathop{\mathbb{E}}_{P_i}[D_i^k f],
\end{aligned}$$

where we have used the fact that $\Pi_k(X_s) = \binom{k}{i}\Pi_i(s)$ by Equation (7.3). To understand $P_i(s)$ more intuitively, consider the special case when f is non-negative. Here, Π and f induce a distribution P_k over $X(k)$, where

$$P_k(t) = \frac{\Pi_k(t) f(t)}{\mathbb{E}[f]}.$$

P_k then induces the distribution P_i on $X(i)$ via the following process: draw a face $t \in X(k)$ from P_k, and then choose a i-face $s \subset t$ uniformly at random. Replacing the non-negativity of f with the conditions in the theorem statement still leaves $P_i(s)$ a valid distribution, albeit one with a less intuitive description.

The result then follows from an averaging argument:

$$\left|\frac{1}{\mathbb{E}[f]}\mathrm{Var}(D_i^k f)\right| = \left|\mathop{\mathbb{E}}_{P_i}[D_i^k f] - \mathbb{E}[f]\right| \leq \|D_i^k f - \mathbb{E}[f]\|_\infty$$

We note that when $\mathbb{E}[f] > 0$, the ℓ_∞-norm may be replaced with maximum in the above. $\square$

To complete the proof of Theorem 7.8.7, we reduce to Theorem 7.8.2 by noting that any function may be shifted to have locally constant sign and applying our reduction.

Proof of Theorem 7.8.7. Note that the latter bound for functions with locally constant sign is immediate since, by Lemma 7.8.8, f is also $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_2$-pseudorandom and therefore satisfies the guarantees of Corollary 7.8.4.

For the former, assume for simplicity that $\mathbb{E}[f] \geq 0$ (the negative case follows from a similar argument) and consider the shifted function $f' = f + (\varepsilon_i - \mathbb{E}[f])\mathbb{1}$. Notice that as long as $\varepsilon_i > 0$, f' has positive expectation over all i-links and non-zero expectation, and further that

$$f' = f'_0 + f_1 + \ldots + f_k,$$

where $f'_0 = f_0 + (\varepsilon_i - \mathbb{E}[f])\mathbb{1}$ and $f = \sum f_i$ is the original HD-Level-Set Decomposition of f. Since adding a constant has no effect on the ℓ_∞-pseudorandomness, f' remains $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_\infty$-pseudorandom and, by our ℓ_∞ to ℓ_2 reduction, $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_2$-pseudorandom as well. Applying Corollary 7.8.4 then gives:

$$\begin{aligned}
&\langle f + (\varepsilon_i - \mathbb{E}[f])\mathbb{1}, f_i\rangle \\
\leq &\binom{k}{i}\varepsilon_i \mathbb{E}[f + (\varepsilon_i - \mathbb{E}[f])\mathbb{1}] + c\gamma\langle f + (\varepsilon_i - \mathbb{E}[f])\mathbb{1}, f + (\varepsilon_i - \mathbb{E}[f])\mathbb{1}\rangle \\
\leq &\left(\binom{k}{i} + c\gamma\right)\varepsilon_i^2 + c\gamma\langle f, f\rangle
\end{aligned}$$

where we have used the fact that f_i is orthogonal to $\mathbb{1}$ for all $i > 0$. We are left to deal with the case that $\varepsilon_i = 0$, which follows from a limiting argument applying the above to

any $\varepsilon > 0$. $\square$

7.9 Expansion of HD-walks

In this section we characterize the edge expansion of HD-walks, proving Theorem 7.2.6, Theorem 7.2.8, and Corollary 7.2.9. As a reminder, these results focus on two key aspects of edge-expansion: the expansion of links, and the structure of non-expanding sets. We'll start with the former, but first let's recall the definition of edge-expansion (specified to HD-walks for simplicity).

Definition 7.9.1 (Weighted Edge Expansion)**.** Given a weighted simplicial complex (X, Π), a k-dimensional HD-Walk M over (X, Π), and a subset $S \subset X(k)$, the weighted edge expansion of S is

$$\phi(S) = \mathop{\mathbb{E}}_{v \sim \Pi_k|_S} \left[M(v, X(k) \setminus S) \right],$$

where

$$M(v, X(k) \setminus S) = \sum_{y \in X(k) \setminus S} M(v, y)$$

and $M(v, y)$ is the transition probability from v to y.

We start by proving Theorem 7.2.6: that the expansion of i-links is (up to $O(\gamma)$ error) exactly controlled by the eigenvalue of the ith eigenstrip.

Theorem 7.9.2 (Local Expansion vs Global Spectra)**.** *Let (X, Π) be a d-dimensional two-sided γ-local-spectral expander with $\gamma \leq 2^{-\Omega(k)}$, and M a k-dimensional, complete HD-walk with $k < d$. Then for all $0 \leq i \leq k$ and $\tau \in X(i)$:*

$$\phi(X_\tau) \in 1 - \lambda_i(M) \pm c\gamma,$$

where $c \leq w(M)h(M)^2 2^{O(k)}$.

The main idea behind Theorem 7.9.2 is simply to show that the indicator function of any i-link always lies almost entirely in V_k^i (or equivalently, almost entirely in the ith eigenstrip W_k^i).

Lemma 7.9.3. *Let (X, Π) be a d-dimensional two-sided γ-local-spectral expander with $\gamma \leq 2^{-\Omega(k)}$. Then for all $0 \leq i \leq k < d$ and $\tau \in X(i)$, 1_{X_τ} lies almost entirely in V_k^i. That is for all $j \neq i$:*

$$\langle 1_{X_\tau}, 1_{X_\tau, j} \rangle \leq c\gamma \langle 1_{X_\tau}, 1_{X_\tau} \rangle$$

where $c \leq 2^{O(k)}$.

Proof. It is enough to analyze the expansion of X_τ with respect to N_k^1 since the quantity can both be analyzed directly, and expressed in terms of X_τ's HD-Level-Set decomposition. For the direct analysis, recall that N_k^1 describes the process of moving from a k-face σ to a $(k+1)$-face $T = \sigma \cup \{v\}$, then back to a k-face $\sigma' \subset T$. Crucially, the latter step is performed uniformly at random. Applying N_k^1 to any element in X_τ, the probability of returning to X_τ is exactly the probability that we remove an element in $T \setminus \{\tau\}$ in the final step, which gives:

$$\bar{\phi}(1_{X_\tau}) = \frac{k+1-i}{k+1}.$$

On the other hand, we may also expand out $\bar{\phi}(1_{X_\tau})$ in terms of 1_{X_τ}'s HD-Level-Set decomposition. Using the fact that $1_{X_\tau} = \binom{k}{i} U_i^k 1_\tau \in V_k^0 \oplus \ldots \oplus V_k^i$, we may write:

$$\begin{aligned}
\bar{\phi}(1_{X_\tau}) &= \frac{1}{\langle 1_{X_\tau}, 1_{X_\tau} \rangle} \sum_{j=0}^{i} \langle 1_{X_\tau}, N_k^1 1_{X_\tau, j} \rangle \\
&= \frac{1}{\langle 1_{X_\tau}, 1_{X_\tau} \rangle} \sum_{j=0}^{i} \frac{k+1-j}{k} \langle 1_{X_\tau}, 1_{X_\tau, j} \rangle + \frac{1}{\langle 1_{X_\tau}, 1_{X_\tau} \rangle} \sum_{s=0}^{i} \langle \mathbb{1}_{X_\tau}, \Gamma_s \rangle,
\end{aligned}$$

where $\|\Gamma_s\| \leq 2^{O(k)}\gamma \|\mathbb{1}_{X_\tau}\|$ by Corollary 7.7.6 and the fact that $\|\mathbb{1}_{X_\tau, s}\| \leq (1 + 2^{O(k)}\gamma)\|\mathbb{1}_{X_\tau}\|$.

Finally, applying Cauchy-Schwarz to the error term gives:

$$\bar{\phi}(1_{X_\tau}) \in \frac{1}{\langle 1_{X_\tau}, 1_{X_\tau}\rangle} \sum_{j=0}^{i} \frac{k+1-j}{k} \langle 1_{X_\tau}, 1_{X_\tau,j}\rangle \pm c\gamma \tag{7.8}$$

for $c \leq 2^{O(k)}\gamma$. Recall by approximate orthogonality (Lemma 7.8.3), the projections $\langle 1_{X_\tau}, 1_{X_\tau,j}\rangle$ cannot be too negative, that is $\langle 1_{X_\tau}, 1_{X_\tau,j}\rangle \geq -c\gamma\langle 1_{X_\tau}, 1_{X_\tau}\rangle$ for some $c \leq 2^{O(k)}$. Then if there exists some $j \neq i$ such that $\langle 1_{X_\tau}, 1_{X_\tau,j}\rangle > c_2\gamma\langle 1_{X_\tau}, 1_{X_\tau}\rangle$ for large enough $c_2 \leq 2^{O(k)}$, the LHS of Equation (7.8) lies strictly above $\frac{k+1-i}{k+1}$, giving the desired contradiction. □

Since the HD-Level-Set decomposition of 1_{X_τ} has no dependence on the walk in question, Theorem 7.9.2 follows almost immediately.

Proof of Theorem 7.9.2. The expansion of X_τ may be written as:

$$\begin{aligned}\phi(X_\tau) &= 1 - \frac{1}{\alpha}\langle \mathbb{1}_{X_\tau}, M\mathbb{1}_{X_\tau}\rangle \\ &= 1 - \frac{1}{\alpha}\sum_{s=0}^{i}\langle \mathbb{1}_{X_\tau}, M\mathbb{1}_{X_\tau,s}\rangle,\end{aligned}$$

where α is the density of X_τ and $\mathbb{1}_{X_\tau,s} \in V_k^s$. By Corollary 7.7.6 we can simplify the sum up to an error term:

$$\begin{aligned}\phi(X_\tau) &= 1 - \frac{1}{\alpha}\sum_{s=0}^{i}\langle \mathbb{1}_{X_\tau}, M\mathbb{1}_{X_\tau,s}\rangle \\ &= 1 - \frac{1}{\alpha}\sum_{s=0}^{i}\lambda_j(M)\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,s}\rangle + \frac{1}{\alpha}\sum_{s=0}^{i}\langle \mathbb{1}_{X_\tau}, \Gamma_s\rangle,\end{aligned}$$

where by the fact that $\|\mathbb{1}_{X_\tau,s}\| \leq (1 + 2^{O(k)}\gamma)\|\mathbb{1}_{X_\tau}\|$ we have

$$\|\Gamma_s\| \leq O\left(w(M)h(M)(h(M)+k)\binom{k}{s}\gamma\|\mathbb{1}_{X_\tau}\|\right).$$

By Lemma 7.9.3 for all $s \neq i$ we can absorb $\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,s} \rangle$ into the error term which gives:

$$\phi(X_\tau) \in 1 - \frac{1}{\alpha}\lambda_i(M)\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,i} \rangle \pm c\gamma$$

for $c \leq w(M)h(M)^2 2^{O(k)}$. Finally, by Lemma 7.9.3 and approximate orthogonality, we know that $\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,i} \rangle$ is very close to α:

$$(1 - c_1\gamma)\alpha \leq \langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau,i} \rangle \leq (1 + c_2\gamma)\alpha$$

where $c_1, c_2 \leq 2^{O(k)}$. Combining this with the above completes the proof. Note that the proof falls through for $d = k$ since we cannot analyze N_k^1 on such a complex. The upper bound, however, still holds in this case simply by expanding out the inner product and bounding the inner summation using $\lambda_i(M)$. □

Theorem 7.9.2 will form one of two core pieces of our algorithm with unique games. The fact that links corresponding to bad eigenvalues have poor expansion will allow us to patch together good local solutions into a global solution without seeing too much interference. Moreover, close connection between local expansion and stripped eignenvalues will result in the performance of our algorithm being tied directly to ST-rank (which we recall here for convenience).

Definition 7.9.4 (Stripped Threshold Rank)**.** Let (X, Π) be a two-sided γ-local-spectral expander and M a k-dimensional HD-walk with γ small enough that the HD-Level-Set Decomposition has a corresponding decomposition of disjoint eigenstrips $C_k = \bigoplus W_k^i$.[20] The ST-Rank of M with respect to δ is the number of strips containing an eigenvector

[20]It should be noted that when the HD-Level-Set Decomposition has spaces with the same approximate eigenvalue, their corresponding eigenstrips technically must be merged. However, since this detail has no effect on our arguments, we ignore it in what follows.

with eigenvalue at least δ:

$$R_\delta(M) = |\{W_k^i : \exists f \in V^i, Mf = \lambda f, \lambda > \delta\}|.$$

We often write just R_δ when M is clear from context.

A basic corollary of Theorem 7.9.2 is that, like the Johnson graphs, links are small, non-expanding sets (at least when $|X(1)| \gg k$ or γ is small). The second core piece of our algorithm for unique games relies on a certain converse to this result: that *all* non-expanding sets are explained by links. As discussed in Section 7.1 and Section 7.2, we consider two regimes for this problem. The first, which we call the ℓ_2-variant, claims that any non-expanding set must have high *variance* over links—this regime is useful for constructing algorithms for unique games, as we'll show in the next section. The second is the ℓ_∞-variant, which claims that any non-expanding set must have a high *maximum* over links—this regime is useful in hardness of approximation. We examine both regimes through their contrapositive: that both ℓ_2/ℓ_∞-pseudorandom sets expand near-perfectly.

Theorem 7.9.5. *Let (X, Π) be a two-sided γ-local-spectral expander, M a k-dimensional, complete HD-walk, and let γ be small enough that the eigenstrip intervals of Theorem 7.6.2 are disjoint. For any $\delta > 0$, let $r = R_\delta(M) - 1$. Then the expansion of a set $S \subset X(k)$ of density α is at least:*

$$\phi(S) \geq 1 - \alpha - (1-\alpha)\delta - c\gamma - \sum_{i=1}^{r} (\lambda_i(M) - \delta)\binom{k}{i}\varepsilon_i,$$

where $\lambda_i(M)$ is the approximate eigenvalue given by Corollary 7.7.6, S is either $(\varepsilon_1, \dots, \varepsilon_r)$-$\ell_2$-pseudorandom or $(\varepsilon_1, \dots, \varepsilon_r)$-$\ell_\infty$-pseudorandom, and $c \leq w(M)h(M)^2 2^{O(k)}$.

Proof. Recall that the expansion of S may be written as:

$$\phi(S) = 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]}\langle \mathbb{1}_S, M\mathbb{1}_S \rangle.$$

Decomposing $\mathbb{1}_S = \mathbb{1}_{S,0} + \ldots + \mathbb{1}_{S,k}$ by the HD-Level-Set Decomposition, we have:

$$\begin{aligned}\phi(S) &= 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=0}^{k} \langle \mathbb{1}_S, M\mathbb{1}_{S,i} \rangle \\ &= 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=0}^{k} \lambda_i(M) \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle + \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=1}^{k} \langle \mathbb{1}_S, \Gamma_i \rangle\end{aligned}$$

where by Corollary 7.7.6 $\|\Gamma_i\| \leq O\left(w(M)h(M)(h(M)+k)\binom{k}{i}\gamma \|\mathbb{1}_{S,i}\|\right)$. Using Cauchy-Schwarz and the fact that $\|\mathbb{1}_{S,i}\| \leq (1 + 2^{O(k)}\gamma)\|\mathbb{1}_S\|$ (this follows from approximate orthogonality, see [111, Corollary 8.13]) we can simplify this to

$$\phi(S) \geq 1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=0}^{k} \lambda_i(M) \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle - e\gamma,$$

where $e \leq w(M)h(M)^2 2^{O(k)}$. Since M is a complete walk, we know the $\lambda_i(M)$ decrease monotonically and as long as γ is sufficiently small, correspond to the eigenvalues in strip W^i as well. Thus we may write:

$$\begin{aligned}\phi(S) &\geq 1 - e\gamma - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=0}^{r} \lambda_i(M) \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle - \frac{\delta}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=r+1}^{k} \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle \\ &= 1 - e_2\gamma - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=1}^{r} \lambda_i(M) \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle - \delta \left(1 - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=0}^{r} \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle\right) \\ &= 1 - e_2\gamma - \delta - \frac{1}{\mathbb{E}[\mathbb{1}_S]} \sum_{i=1}^{r} (\lambda_i(M) - \delta) \langle \mathbb{1}_S, \mathbb{1}_{S,i} \rangle\end{aligned}$$

Finally, recalling that $\mathbb{1}_{S,i} = \mathbb{E}[\mathbb{1}_S]\vec{1}$ and applying Theorem 7.8.7 gives the ℓ_2-variant result, and the ℓ_∞-variant follows immediately from our ℓ_∞ to ℓ_2 reduction (Lemma 7.8.8). $\square$

We now turn to the discussion of a surprisingly subtle point: the tightness of

Theorem 7.9.5. There are two main parameters of interest: the pseudorandomness parameter ε, and the level of the HD-walk k. We'll first prove that in both the ℓ_2 and ℓ_∞ regimes, if we fix the dependence on ε to be linear, our bound is exactly tight.

Proposition 7.9.6. *Let $X = J(n,d)$ be the complete complex, $2k - t \leq d$, $m|n$, and B_m be the the set of all k-faces $\binom{[n/m]}{k}$. Then for any t, B_m witnesses the tightness of Theorem 7.9.5 with respect to S_k^{k-t} as $n, m \to \infty$.*

Proof. First, note that it is enough to examine only the ℓ_∞ bound, since the function in question has locally-constant sign the tightness of the ℓ_2 bound follows from Lemma 7.8.8.

By direct computation, it is not hard to show that the expansion of B_m with respect to S_k^{k-t} is:

$$\phi(B_m) = 1 - \frac{\binom{\frac{n}{m}-k}{k-t}}{\binom{n-k}{k-t}} = 1 - \frac{m^t}{m^k} + O_{k,m}(1/n)$$

On the other hand, we can directly compute that B_m is $(\varepsilon_1, \ldots, \varepsilon_k)$-pseudorandom, where:

$$\varepsilon_i \leq \frac{\binom{\frac{n}{m}-i}{k-i}}{\binom{n-i}{k-i}}$$

Since the complete complex is a two-sided $O(1/n)$-local-spectral expander [111], for large enough n Theorem 7.9.5 gives the bound:

$$\begin{aligned}
\phi(B_m) &\geq 1 - \sum_{i=0}^{t} \binom{t}{i} \frac{\binom{\frac{n}{m}-i}{k-i}}{\binom{n-i}{k-i}} - O_{k,m}(1/n) \\
&= 1 - \frac{\binom{n/m}{k}}{\binom{n}{k}} \sum_{i=0}^{t} \binom{t}{i} \frac{\binom{n}{i}}{\binom{n/m}{i}} - O_{k,m}(1/n) \\
&\geq 1 - m^{-k} \sum_{i=0}^{t} \binom{t}{i} m^i - O_{k,m}(1/n) \\
&= 1 - \frac{(m+1)^t}{m^k} - O_{k,m}(1/n)
\end{aligned}$$

Thus we see that for large n, the bound is tight up to the leading term in m. $\square$

However, Proposition 7.9.6 does not preclude a bound with better (or even no) dependence on k. Indeed, proving that such a bound holds in the ℓ_∞-regime for the Johnson graphs was an important stepping stone in the proof of the 2-2 Games Conjecture [252]. On the other hand, we can actually prove that a k-independent bound is *impossible* in the ℓ_2-regime. The intuition behind the difference can be summarized by examining the behavior of these two variants on a link. In the ℓ_∞ regime, links are of course $\Omega(1)$-pseudorandom. On the other hand, somewhat counter-intuitively, links are actually $O(\binom{k}{i}^{-1})$-pseudorandom in the ℓ_2-regime. Since links tend to have poor expansion, the dependence on k in our bound has to make up for the fact that links are $O(\binom{k}{i}^{-1})$-ℓ_2-pseudorandom (this also explains the particular dependence on $\binom{k}{i}$).

Proposition 7.9.7. *For every $\ell \in \mathbb{N}$ and $c_1(\ell), c_2(\ell) > 0$, there exist $n \gg k \gg \ell$ and an $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_2$-pseudorandom subset S of $S_k^{k/2}$ on $J(n, 2k)$[21] of density α satisfying:*

$$\phi(S) < 1 - \alpha - \lambda_{\ell+1}(S_k^{k/2}) - c_1\varepsilon^{c_2}$$

Proof. The result follows from letting S be any ℓ-link of the complete complex $J(n, 2k)$ for sufficiently large n,k. In particular, notice that the expansion of S with respect to $S_k^{k/2}$ is then:

$$\phi(S) \leq 1 - 2^{-\ell} + O_k\left(\frac{1}{n}\right).$$

On the other hand, note that $\lambda_{\ell+1}(S_k^{k/2}) \in 2^{-\ell-1} \pm O_k(\frac{1}{n})$. Since one can also check that S is $(\varepsilon_1, \ldots, \varepsilon_\ell)$-$\ell_2$-pseudorandom for $\varepsilon_i \leq \binom{k}{i}^{-1} + O_k\left(\frac{1}{n}\right)$, we for large enough $n \gg k$ that:

$$1 - \alpha - \lambda_{\ell+1}(S_k^{k/2}) - c_1\varepsilon^{c_2} \geq 1 - 2^{-\ell-1} - O_k\left(\frac{1}{n}\right) - o_k(1) > \phi(S)$$

[21] Note that this is exactly the Johnson Graph $J(n, k, k/2)$ mentioned in Section 7.2.

as desired. □

In other words, no k-independent version of Theorem 7.9.5 exists for the ℓ_2-variant, as such a result would violate the upper bound in Proposition 7.9.7. While this doesn't directly rule out a reduction from ℓ_∞ to ℓ_2 that proves a k-independent bound for the former (the reduction itself would need to be *k-dependent* in this case), it's known to be impossible in our framework which encompasses the Grassmann where such a result is known to be false [125].

The difference between a k-independent bound and the regime we consider is most stark when examining the contrapositive of Theorem 7.9.5, which states that non-expanding sets must be concentrated inside links.

Corollary 7.9.8. *Let (X, Π) be a two-sided γ-local-spectral expander, M a k-dimensional, complete HD-walk, and let γ be small enough to satisfy the requirements of Theorem 7.6.2. Then for any $\delta > 0$, if $S \subset X(k)$ is a set of density α and expansion:*

$$\phi(S) < 1 - \alpha - (1-\alpha)\delta - c\gamma$$

for $c \leq w(M)h(M)^2 2^{O(k)}$, then S is non-trivially correlated with an i-link for $1 \leq i \leq R_{\delta/2}$:

$$\exists 1 \leq i \leq R_{\delta/2}, \tau \in X(i) : \mathop{\mathbb{E}}_{X_\tau}[\mathbb{1}_S] \geq \alpha + \frac{\delta}{c_2 R_{\delta/2}\binom{k}{i}\lambda_i(M)},$$

where $c_2 > 2$ is some small absolute constant.

Notice that the excess correlation implied by Corollary 7.9.8 decays as k grows large (except in the case of very deep walks like $S_k^{k-O(1)}$); this is one of the main obstructions to using results like Corollary 7.9.8 for hardness of unique games. On the other hand, we now turn our attention to *algorithms* for unique games, where the ℓ_2-regime gets its chance to shine.

7.10 Playing Unique Games on HD-Walks

Following the recent algorithmic framework of [37], we show how to translate our spectral machinery and combinatorial characterization of non-expansion into a polynomial time algorithm for unique games over HD-walks whose finer-grained runtime and approximation guarantees depend on ST-rank. Before we dive into the theorems let's start with some notation for this section.

Notation:.

We briefly comment on the use of $\mathcal{I} = (M, \mathcal{S})$ to denote an affine UG instance over a random-walk M, since unique games are formally defined over graphs, not random walks. In particular, it is well known that every HD-walk M over $X(k)$ uniquely corresponds to an undirected weighted graph $G_M = (X(k), E)$ (see Section 7.14.1). Thus by $\mathcal{I} = (M, \mathcal{S})$, we really mean that the constraints $\mathcal{S}$ are over the edges of G and the value is calculated according to the distribution over edges E. We will use the $\widetilde{O}(\cdot)$ notation in this section to hide log factors, that is, $\widetilde{O}(f)$ denotes $O(f \log^c f)$ for any constant $c > 0$.

With notation out of the way, let us describe the main theorem:

Theorem 7.10.1. *For any $\varepsilon \in (0, .01)$, there exists an algorithm $\mathcal{A}$ with the following guarantee. Let (X, Π) be a d-dimensional two-sided γ-local spectral expander and M be a k-dimensional complete HD-walk over X such that $\gamma \leq w(M)^{-1} 2^{-\Omega(k+h(M))}$ and $d > k$. Let $\mathcal{I} = (M, \mathcal{S})$ be an instance of affine unique games over M with value at least $1 - \varepsilon$. Then $\mathcal{A}$ outputs an $\Omega\left(\frac{\varepsilon^3}{\binom{k}{r(2\varepsilon)}^2}\right)$-satisfying assignment in time $|X(k)|^{\widetilde{O}\left(\binom{k}{r(2\varepsilon)}^{\frac{1}{\varepsilon}}\right)}$, where $r(\varepsilon) = R_{1-16\varepsilon}(M)$ is the ST-rank of M.*

Since the HDX literature focuses mostly on canonical and partial-swap walks, we also give the specification of Theorem 7.10.1 to this class for concreteness. Here we see that the finer-grained performance of our algorithm depends on the *depth* of the walk.

Corollary 7.10.2. *For any $\varepsilon \in (0, .01)$, there exists an algorithm $\mathcal{A}$ with the following guarantee. Let (X, Π) be a d-dimensional two-sided γ-local spectral expander and M be a k-dimensional canonical or partial-swap walk of depth $\beta \in [0, 1]$ over X such that $\gamma \leq w(M)^{-1}2^{-\Omega(k+h(M))}$ and $d > k$. Let $\mathcal{I} = (M, \mathcal{S})$ be an instance of affine unique games over M with value at least $1 - \varepsilon$. Then $\mathcal{A}$ outputs an $\Omega\left(\frac{\varepsilon^3}{\binom{k}{c\frac{\varepsilon}{\beta}}^2}\right)$-satisfying assignment in time $|X(k)|^{\widetilde{O}\left(\binom{k}{c\varepsilon/\beta}\frac{1}{\varepsilon}\right)}$ for some absolute constant $c > 0$.*

To prove Theorem 7.10.1, we follow the general SoS algorithmic paradigm introduced by BBKSS for Johnson graphs (partial-swap walks on the complete complex), described in Section 7.10.2 for completeness. As discussed in Section 7.3.3 we abstract BBKSS' analysis to rely on two core structural properties true of the graphs underlying HD-walks:

1. There exists a low-degree Sum-of-Squares proof that non-expanding sets have high variance in size across links.

2. For every small enough ϵ, there exists a parameter $r = r(\varepsilon)$ such that:

   (a) The $(r+1)$-st largest (distinct) stripped-eigenvalue of G is small:

   $$\lambda_r \leq 1 - \Omega(\varepsilon)$$

   (b) The expansion of any r-link is small as well:

   $$\forall \tau \in X(s) : \Phi(X_\tau) \leq O(\varepsilon).$$

Recall that we have already proved property (1) in Theorem 7.9.5, albeit not in the low-degree SoS proof system. In Section 7.10.4 we modify the proof of Theorem 7.9.5 to show that it has a degree 2 SoS proof. For property (2), it is clear that the existence of $r(\epsilon)$ is an inherent consequence of Theorem 7.9.2, and furthermore it is exactly the

$(1 - O(\epsilon))$ ST-Rank of the HD-walk. Analyzing the generalized BBKSS algorithm using these properties combined with a few more technicalities (described in Section 7.10.4) then gives an efficient algorithm for unique games over HD-walks on two-sided local spectral expanders.

7.10.1 Background for Unique Games and SoS

Proving Theorem 7.10.1 from the ground up requires substantial background in the SoS framework. However, since we mostly rely on a number of higher level results from [37] for the SoS side of our work, we cover here only background necessary to understand our methods, and refer the reader to the surveys of [49, 151, 324] and additionally Sections 1, 2, and A of [37] for more information.

The Sum of Squares framework is a method for approximating polynomial optimization problems through semi-definite programming relaxations. In particular, given the problem:UG

$$\text{Maximize } p \in \mathbb{R}[x_1, \ldots, x_n] \text{ constraint to } \{q_i = 0\}_{i=1}^m,$$

for $q_i \in \mathbb{R}[x_1, \ldots, x_n]$, the Degree-$D$ Sum of Squares semidefinite programming relaxation outputs in time $n^{O(D)}$ a *pseudoexpectation operator* $\widetilde{\mathbb{E}} : \text{poly}_{\mathbb{R}}(n, D) \to \mathbb{R}$ over polynomials in $\mathbb{R}[x_1, \ldots, x_n]$ of degree at most D satisfying:

1. Scaling: $\widetilde{\mathbb{E}}[1] = 1$
2. Linearity: $\widetilde{\mathbb{E}}[af(x) + bg(x)] = a\,\widetilde{\mathbb{E}}[f(x)] + b\,\widetilde{\mathbb{E}}[g(x)]$
3. Non-negativity (for squares): $\widetilde{\mathbb{E}}[f(x)^2] \geq 0$
4. Program constraints: $\widetilde{\mathbb{E}}[f(x)q_i(x)] = 0$

5. Optimality: $\widetilde{\mathbb{E}}[p(x)] \geq \max_x \{p(x) : \{q_i = 0\}_{i=1}^m\}$

Note that the first four properties give the definition of a pseudoexpectation (under constraints $\{q_i = 0\}_{i=1}^m$), whereas the fifth is promised by the SoS relaxation. The pseudoexpectation operator can equivalently be defined as a weighted expectation over a "pseudodistribution" μ satisfying similar properties (and analogous to an actual distribution). We will sometimes use the pseudodistribution view below when convenient and refer to $\widetilde{\mathbb{E}}[\cdot]$ (or $\widetilde{\mathbb{E}}_\mu[\cdot]$ in this case) as pseudomoments of the pseudodistribution (see [49] for a more detailed exposition).

A Degree-D Sum of Squares proof of a polynomial inequality $f(x) \leq g(x)$ (where f, g are polynomials of degree at most D) is a method for ensuring the inequality continues to hold over any degree-D pseudoexpectation. In particular, given constraints $\{q_i = 0\}_{i=1}^m$, a degree-D sum of squares proof of $f \leq g$, denoted by:

$$\{q_i = 0\}_{i=1}^m \vdash_D f \leq g,$$

is a certificate of the form $g(x) = f(x) + \sum s(x)^2 + \sum_i t(x)q_i(x)$ where all terms have degree at most D. Notice that properties 2, 3, and 4 then immediately imply $\widetilde{\mathbb{E}}[f(x)] \leq \widetilde{\mathbb{E}}[g(x)]$.

Conditioning is a standard algorithmic technique in the SoS paradigm used to improve the value of independently sampling a solution from the output of an SoS semidefinite relaxation (see e.g. [48, 37]). Given a degree D pseudodistribution μ, and a degree $< D$ sum of squares polynomial $s(x)$, we can define a new pseudodistribution μ' by *conditioning* on $s(x)$ as follows:

$$\widetilde{\mathbb{E}}_{\mu'}[f(x)|s(x)] = \frac{\widetilde{\mathbb{E}}_\mu[f(x)s(x)]}{\widetilde{\mathbb{E}}_\mu[s(x)]},$$

for all polynomials $f(x)$ of degree $\leq D - \deg(s)$. We have that μ' is a valid pseudodistribution of degree $D - \deg(s)$ that satisfies the axioms satisfied by μ[22]. In an algorithmic context, this is often used to restrict the pseudoexpectation to some partial solution.

7.10.2 The Algorithm

We now present our algorithm for solving Unique Games on HD-walks. We follow the overall framework of BBKSS, which is based on the Sum-of-Squares semidefinite programming (SDP) relaxation paradigm and its view as optimizing over pseudoexpectation operators. The unique games problem (Definition 7.2.10) can be written as a polynomial optimization problem. In particular, given an instance I of unique games with alphabet Σ and constraints $\mathcal{S}$ over $G(V, E)$ (that are of the form $X_u - X_v = s_{uv} \pmod k$), consider the following quadratic optimization problem $\mathcal{A}_I$ over variables $\{X_{v,s}\}_{V \times \Sigma}$ that computes $\text{val}(I)$:

$$\begin{aligned}
\text{Maximize:} \quad & \mathop{\mathbb{E}}_{(u,v)\sim E} \left[\sum_{s \in \Sigma} X_{u,s} X_{v,\pi_{uv}(s)} \right] & \\
\text{Constraint to:} \quad & X_{v,s}^2 = X_{v,s} & \forall v \in V, s \in \Sigma \\
& X_{v,a} X_{v,b} = 0 & \forall v \in V, a \neq b \in \Sigma \\
& \sum_{s \in \Sigma} X_{v,s} = 1 & \forall v \in V
\end{aligned}$$

The variables $X_{v,a}$ are 0/1 indicators that vertex $v \in V$ takes the value $a \in \Sigma$, E is a distribution over the edges of the weighted graph G, the constraints are $\pi_{uv}(s) = s - s_{uv} \pmod k$, and the objective function maximizes the fraction of constraints satisfied. We will work with the Degree-D Sum of Squares relaxation of this program, which outputs a degree-D pseudodistribution μ and corresponding pseudoexpectation operator $\widetilde{\mathbb{E}}_\mu : X^{\leq D} \to \mathbb{R}$, where $X^{\leq D}$ is the set of all monomials in the X variables up to degree D,

[22] Technically only the axioms that were of degree $\leq D - \deg(s)$

and $\widetilde{\mathbb{E}}_\mu$ satisfies the above equality constraints as axioms. The *value* of $\widetilde{\mathbb{E}}_\mu$, with respect to the instance I is denoted by $\mathrm{val}_\mu(I) = \widetilde{\mathbb{E}}_\mu[\mathrm{val}_I(X)] = \widetilde{\mathbb{E}}_\mu[\mathop{\mathbb{E}}_{(u,v)\sim E}[\sum_{s\in\Sigma} X_{u,s}X_{v,\pi_{uv}(s)}]]$. This operator is obtained in time $|V|^{O(D)}$ such that $\widetilde{\mathbb{E}}_\mu[\mathrm{val}_I(X)] \geq \mathrm{val}(I)$.

We now describe the rounding algorithm that takes a pseudodistribution μ over an almost-satisfiable UG instance I and produces a high value assignment.

Condition&Round:.

We start with a basic sub-routine which will then be used in the final algorithm for HD-walks. This subroutine takes a pseudodistribution μ for an affine unique games instance $(G(V,E),\Pi)$ on alphabet Σ and outputs an assignment $x \in \Sigma^V$ via the following process:

1. Sample a vertex $v \in V$ uniformly at random, and condition μ on event $X_{v,0} = 1$ to get the conditioned pseudodistribution $\mu|(X_{v,0} = 1)$.

2. Sample a solution $x \in \Sigma^V$ by independently sampling a label for every vertex $w \neq v$ from its marginal distribution in $\mu|(X_{v,0} = 1)$: $\mathbb{P}[x_w = s] = \widetilde{\mathbb{E}}_\mu[X_{w,s}|X_{v,0} = 1]$.

Note that $\widetilde{\mathbb{E}}_\mu[\cdot|X_{v,0} = 1]$ is by definition conditioning μ on the polynomial $X_{v,0}$, hence $\widetilde{\mathbb{E}}_\mu[X_{w,s}|X_{v,0} = 1] = \widetilde{\mathbb{E}}_\mu[X_{w,s}X_{v,0}]/\widetilde{\mathbb{E}}_\mu[X_{v,0}]$. Following [37], we define the term Condition&Round-value (abbreviated to CR-val) of an instance I with respect to a pseudodistribution μ:

Definition 7.10.3 (Condition&Round value)**.** The *CR-Value* of the instance I with respect to a pseudodistribution μ is the expected value of the solution output by Condition&Round on instance I and pseudodistribution μ, denoted CR-$\mathrm{Val}_\mu(I)$. We drop the subscript μ when clear from context.

Before we describe the main algorithm, an iterative framework for applying Condition&Round, we need to introduce a simple operation on pseudodistributions for affine unique games that allows for ease of analysis:

Symmetrization:.

Symmetrization is an operation on pseudodistributions introduced in [37] to take advantage of the symmetric structure of affine unique games. The idea is to average the pseudoexpectation operator over shifts $s \in \Sigma$. Formally, given a degree D pseudodistribution μ, define the symmetrized pseudodistribution μ_{sym} via its pseudoexpectation operator as follows. For all degree $\leq D$ monomials:

$$\widetilde{\mathbb{E}}_{\mu_{sym}}[X_{u_1,a_1} \dots X_{u_t,a_t}] = \frac{1}{|\Sigma|} \sum_{s \in \Sigma} \widetilde{\mathbb{E}}_{\mu}[X_{u_1,a_1+s} \dots X_{u_t,a_t+s}].$$

We will call a pseudodistribution shift-symmetric if it is invariant under this operation. If μ is a degree D pseudodistribution that satisfies the unique games axioms $\mathcal{A}_I$ it is easy to verify that the symmetrized pseudodistribution μ_{sym} is also a valid degree D pseudodistribution satisfying $\mathcal{A}_I$ with $\mathrm{val}(\mu) = \mathrm{val}(\mu_{sym})$. Furthermore symmetrization can be performed in time subquadratic in the description of $\widetilde{\mathbb{E}}_{\mu}$. As a result, we can perform this operation essentially for free inside our algorithm and therefore assume throughout that we are working with a shift-symmetric pseudodistribution.

We are now ready to describe the main algorithm which is called Iterated Condition&Round. The algorithm follows the strategy presented in [37, Algorithm 6.1], differing mainly in that the parameter $r(\varepsilon)$ satisfying their second condition has been replaced with the $(1 - O(\epsilon))$-ST-Rank of the underlying constraint graph.

Iterated Condition&Round:.

The full algorithm builds a solution by iteratively applying Condition&Round to links. Let M be a complete k-dimensional HD-walk over a d-dimensional two-sided γ-local spectral expander (X, Π) and $G_M = (X(k), E)$ be the corresponding undirected weighted graph on vertex set $X(k)$. Let $I = (M, \mathcal{S})$ be an instance of affine unique games over alphabet Σ with $\mathrm{val}(I) \geq 1 - \varepsilon$ and $r = R_{1-16\varepsilon}(M)$. Further, given a subset $H \subset X(k)$, let I_H denote the restriction of the instance I to the subgraph vertex-induced by H. Given

a subroutine for finding a link with high CR-value (see Proposition 7.10.5), the following process returns an $\Omega_{\varepsilon,r,k}(1)$ satisfying assignment.

1. Let $\delta(\varepsilon) := \Omega\left(\frac{\varepsilon}{\binom{k}{r}}\right)$. Solve the Degree-$D = \widetilde{O}\left(1/\delta(\varepsilon)\right)$ SoS SDP relaxation of unique games, and symmetrize the resulting pseudodistribution to get μ_0. Set $j = 1$.

2. Let $\mathrm{Dif}(j) = \widetilde{\mathbb{E}}_{\mu_0}[val_I(x)] - \widetilde{\mathbb{E}}_{\mu_{j-1}}[val_I(x)]$. While $\mathrm{Dif}(j) \leq \varepsilon$:

   (a) Find an r-link X_τ such that the CR-Value of $I|_{X_\tau}$ is at least $\delta(\varepsilon + \mathrm{Dif}(j))$[23].

   (b) Let S_j be the subgraph of X_τ induced by the vertices in $X(k)$ which have not yet been assigned a value in any partial assignment f_i, $i \leq j$, and perform Condition&Round on S_j to get partial assignment f_j.

   (c) Create a new pseudodistribution μ_j by making the marginal distribution over assigned vertices uniform and independent of others, i.e. for all degree $\leq D$ monomials let $\widetilde{\mathbb{E}}_{\mu_j}$ be:

$$\widetilde{\mathop{\mathbb{E}}_{\mu_j}}[X_{h_1,a_1} \dots X_{h_t,a_t} X_{u_1,b_1} \dots X_{u_m,b_m}] = \frac{1}{|\Sigma|^t} \widetilde{\mathop{\mathbb{E}}_{\mu_{j-1}}}[X_{u_1,b_1} \dots X_{u_m,b_m}],$$

   where $h_i \in S_j$ and $u_i \in X(k) \setminus S_j$. Increment $j \leftarrow j + 1$.

It is worth noting that the Condition&Round subroutine, and thus the entire Iterated Condition&Round algorithm, can be derandomized by standard techniques like the method of conditional expectations [37].

7.10.3 Analysis of Algorithm 7.10.2

BBKSS' analysis of Iterated Condition&Round algorithm relies heavily on analysing a quantity called the Approximate shift-partition potential defined therein. For completeness we include the definitions of the potential functions used in Section 7.14.2. For the

[23]When γ, k, d satisfy certain inequalities then such a link is guaranteed to exist by Proposition 7.10.5. Therefore we can find it by enumerating over all links $\tau \in X(r)$ and computing CR-val(X_τ)

purpose of this section, the potential $\Phi^I_{\beta,\nu}(X)$ can be thought of as a low-degree polynomial (with degree $= \widetilde{O}(1/\nu)$) in the variables of $\mathcal{A}_I$ ($X = (\ldots, X_{v,a}, \ldots)$ where $v \in V$ and $a \in \Sigma$). Given a pseudodistribution μ for instance I, the pseudoexpectation of the potential will be denoted by $\Phi^I_{\beta,\nu}(\mu) = \widetilde{\mathbb{E}}_\mu[\Phi^I_{\beta,\nu}(X)]$. β, ν are some parameters in $[0,1]$ that control the degree of $\Phi^I_{\beta,\nu}(X)$ and the value of the assignment we finally obtain via Condition&Round.

The framework developed by BBKSS for analyzing Iterated Condition&Round can be described as follows. First, they prove that the Condition&Round subroutine, when run on a pseudodistribution μ for a unique games instance I, returns a high value assignment whenever $\Phi^I_{\beta,\nu}(\mu)$ (Definition 7.14.2) is high:

Theorem 7.10.4 (BBKSS Theorem 3.3)**.** *Let $I = (G, \mathcal{S})$ be an affine instance of Unique Games over graph $G = (V, E)$ and the alphabet Σ. Let $\beta, \nu \in [0,1]$ and μ be a degree-$\widetilde{O}(1/\nu)$ shift-symmetric pseudodistribution satisfying the unique games axioms $\mathcal{A}_I$ (Section 7.10.2). If $\Phi^I_{\beta,\nu}(\mu) \geq \delta$, then on input μ, the Condition&Round Algorithm runs in time* $\mathrm{poly}(|V(G)|)$ *and returns an assignment of expected value at least $(\delta - \nu)(\beta - \nu)$ for I.*

Using this rounding algorithm as a subroutine the analysis of Iterated Condition&Round in BBKSS proceeds in the following way:

1. For every $I = (G, \Pi)$, where G is a Johnson graph, using the structure theorem for non-expanding sets of G and the properties of the potential function, there exists a link X_τ of G such that the potential induced on C is large: $\Phi^{I|_{X_\tau}}_{\beta,\nu}(\mu) \geq \mathrm{poly}(\delta)$, where $I|_{X_\tau}$ denotes the UG instance induced on the subgraph X_τ. Using the rounding theorem 7.10.4 we get that the Condition&Round value on this link is high.

2. Iterative Condition&Round: Using iterations one can patch together solutions on different links, and since each link is a non-expanding set we get a good solution for the whole graph.

We show that this framework extends to all UG instances on HD-walks. We use the analysis of Condition&Round (Theorem 7.10.4) in a blackbox way. We start with the analysis of Point 1, which relies on the technical machinery developed in the previous sections for analysing the non-expanding sets of HD-walks. Using this, we show that there always exists a link with high potential and therefore with high Condition&Round value.

Proposition 7.10.5. *Let M be a k-dimensional complete HD-walk on a d-dimensional, two-sided γ-spectral expander with $\gamma \leq w(M)^{-1}2^{-\Omega(h(M)+k)}$ and $d > k$, and I be an affine unique games instance over M with value at least $1-\eta$ where $1/2^k \leq \eta < .02$. Let $r = r(\eta) = R_{1-16\eta}(M) - 1$, and assume $r \leq k/2$. Then given a degree-$\widetilde{O}\left(\frac{1}{\eta}\binom{k}{r}\right)$ pseudodistribution satisfying the axioms $\mathcal{A}_I$, we can find in time $|X(k)|^4$ an r-link X_τ with CR-Value$(X_\tau) \geq \Omega\left(\frac{\eta}{\binom{k}{r}}\right)$.*

The proof of this lemma is similar in nature to the proof of the analogous lemma in BBKSS, albeit with all of our technical machinery is plugged in. In particular we use a sum-of-squares version of Theorem 7.8.2, the relation between eigenvalues and expansion of links (Theorem 7.9.2), and a new property analyzing the expansion of vertices within a link (Lemma 7.10.12). We elaborate on this in Section 7.10.4.

Given the ability to find a link with high CR-value, we can use BBKSS analysis of Iterated Condition&Round ([37, Lemma 6.12]) as a blackbox to conclude it produces an assignment satisfying a large fraction of the constraints for the *whole graph*.

Lemma 7.10.6 (Lemma 6.12 [37])**.** *Let $\varepsilon \in (0, .01)$, $\delta : [0,1] \to [0,1]$ be any function, and $\delta_{min} = \min_{\delta(\eta)\in[\varepsilon,2\varepsilon]}$. Let G be a weighted graph[24] and I be any affine unique games instance on G with alphabet size $|\Sigma| \geq \Omega\left(\frac{1}{\delta_{min}}\right)$ and value at least $1-\varepsilon$.*

Suppose we have a subroutine $\mathcal{A}$ which, for any $\varepsilon \leq \eta \leq 2\varepsilon$, given as input a shift-symmetric degree-D pseudodistribution μ satisfying $\mathcal{A}_I$ with $\widetilde{\mathbb{E}}_\mu[val_I(x)] \geq 1-\eta$ returns a vertex-induced subgraph H such that:

[24]BBKSS only state the result for regular graphs, but their proof works for weighted graphs too, where the value, expansion etc are measured appropriately according to the edge-weights.

1. *The CR-Value of I_H is at least $\delta(\eta)$.*

2. *The expansion of H is $O(\eta)$.*

Then if $\mathcal{A}$ runs in time $T(\mathcal{A})$, Iterated Condition&Round[25] *outputs a solution for I satisfying an $\Omega(\delta_{min}^2\varepsilon)$-fraction of edges of G in time $|V(G)|(T(\mathcal{A}) + |V(G)|^{O(D)})$.*

With these results in hand, the final observation to prove Theorem 7.10.1 lies in the relation between the local expansion of links and global spectra of eigenstrips we proved in Section 7.9. Namely, since the rth strip is by definition the last one with eigenvalue worse than $1 - O(\varepsilon)$, the expansion of r-links is at most $O(\varepsilon)$ by Theorem 7.9.2. We therefore meet the conditions of Lemma 7.10.6, and can use Iterated Condition&Round to output a good global assignment. We now prove Theorem 7.10.1 assuming Proposition 7.10.5 and Lemma 7.10.6 hold.

Proof of Theorem 7.10.1. To start, we first prove that we may assume without loss of generality both that $\varepsilon \geq 1/2^k$, and that $r(2\varepsilon) \leq k/2$ (we will need these properties to meet the conditions of Proposition 7.10.5 and Lemma 7.10.6). This relies on the following claim, the proof of which we defer to Section 7.14.4.

Claim 7.10.7. Let M be a k-dimensional complete HD-walk on a d-dimensional, two-sided γ-local-spectral expander with $d > k$ and $\gamma \leq w(M)^{-1}2^{-\Omega(h(M)+k)}$. If the expected laziness of M (i.e. $\mathbb{E}_{\Pi_k}[\mathbb{1}_v^T M \mathbb{1}_v]$) is at least .1, then the spectral gap of M is at least $\Omega(1/k)$ and $\lambda_{k/2}(M) \leq .68$.

With this in mind, note that we can always assume the expected laziness $\mathbb{E}[\mathbb{1}_v^T M \mathbb{1}_v]$ is at most 1/10. This follows from the fact that the expected laziness of M exactly corresponds to the probability of drawing a self-edge on the corresponding graph G_M (see Section 7.14.1), and in an affine unique game, every self edge is either always or never satisfiable. Since our game has value at least .99, in a walk with more than .1 laziness, at

[25] We note that the version of Iterated Condition&Round analyzed in [37] differs slightly from the one we present. Our version simplifies their algorithm a bit but the analysis is exactly the same.

least .09 of the weight on self-edges must be satisfiable, and therefore *every* assignment will satisfy our approximation guarantee. Assuming then that $\mathbb{E}[\mathbb{1}_v^T M \mathbb{1}_v] \leq .1$, Claim 7.10.7 implies that the spectral gap of M is at least $\Omega(1/k)$, in which case standard algorithms for unique games on expander graphs (e.g. [293]) give the desired result. Similarly, we have $\lambda_{k/2} \leq .68 \leq 1 - 32\varepsilon$ by assumption, so we may further assume $r(2\varepsilon) \leq k/2$ as desired.

Now that we have $1/2^k \leq \varepsilon < .02$ and $r \leq k/2$, we are in position to solve the degree-D SoS relaxation of the unique games integer program $\mathcal{A}_I$ for $D = \widetilde{O}(\frac{1}{\varepsilon}\binom{k}{r(2\varepsilon)})$, and apply Proposition 7.10.5 to build a sub-route that satisfies Lemma 7.10.6. Namely, we have that for any $\varepsilon \leq \eta \leq 2\varepsilon$ and pseudodistribution of value at least $1 - \eta$, Proposition 7.10.5 finds a link X_τ with high Condition&Round value:

$$\text{CR-val}(X_\tau) \geq \Omega\left(\frac{\eta}{\binom{k}{r(\eta)}}\right).$$

Further, note that by our assumptions on γ and the fact that M is complete, Proposition 7.7.11 gives that the eigenvalues corresponding to each eigenstrip strictly decrease, and therefore by definition of ST-Rank that the approximate eigenvalue corresponding to the rth eigenstrip is at least $1 - O(\eta)$. By Theorem 7.9.2, this implies that X_τ has poor expansion:

$$\Phi(X_\tau) \leq O(\eta) + w(M)h(M)^2 2^{O(k)}\gamma \leq O(\eta)$$

since $\eta \geq \varepsilon \geq w(M)2^{O(h(M)+k)}\gamma$ by assumption. As a result, this sub-routine satisfies the conditions of Lemma 7.10.6 with δ set to $\delta(\eta) = \Omega\left(\frac{\eta}{\binom{k}{r(\eta)}}\right)$. The only catch is that we need the alphabet Σ to satisfy $|\Sigma| \leq \Omega\left(\frac{1}{\delta_{\min}}\right)$. This can be assumed without loss of generality, since a random solution satisfies a $1/|\Sigma|$ fraction of constraints in expectation (and can be easily derandomized), which satisfies our approximation guarantee if $|\Sigma| \leq O\left(\frac{1}{\delta_{\min}}\right)$. As a result, applying Lemma 7.10.6 outputs a $\Omega(\delta_{\min}^2\varepsilon)$-satisfying solution for $\delta_{\min} = \min_{\eta\in[\varepsilon,2\varepsilon]}\left\{\frac{\varepsilon}{\binom{k}{r(\eta)}}\right\} \geq \Omega\left(\frac{\varepsilon}{\binom{k}{r(2\varepsilon)}}\right)$. The running time guarantee follows

from noting that the sub-routine promised by Proposition 7.10.5 runs in time $|V(G)|^4$, so Lemma 7.10.6 then runs in time $|V(G)|^{O(D)}$ with $|V(G)| = |X(k)|$. Since solving the original Degree-D SoS relaxation also takes time only $|V(G)|^{O(D)}$, we get the desired result. □

7.10.4 Proof of Proposition 7.10.5

Now that we have given the algorithmic background, we prove the main technical lemma behind the analysis. The broad outline of the proof is as follows:

1. We first prove a structure theorem (Theorem 7.10.8) for non-expanding sets of HD-walks. We show an SoS proof of the fact that every non-expanding set must have large variance of size when restricted to links of the complex.

2. Using the structure theorem, in Proposition 7.10.10 we show that given a pseudodistribution μ with objective value $1 - \eta$ for unique games over an HD-walk M, one can find a link X_τ with high *global* shift-partition potential (Definition 7.14.3).

3. In the final step (Lemma 7.10.11), we relate the global shift-partition potential to the shift-partition potential on the subgraph induced by X_τ[26]: we show that $\Phi^{I_\tau}_{\beta,\nu}(\mu)$ (Definition 7.14.4) is large. By the rounding theorem (Theorem 7.10.4), we then conclude that the expected value of the Condition&Round algorithm, when performed on X_τ (CR-$val_\mu(X_\tau)$) must be high.

This proof structure is similar to the analogous Lemma 6.9 of [37]. Most of our technical work goes into proving points 1 and 3 above, and point 2 turns out to be straightforward given BBKSS.

We now turn to proving our structure theorem in the low-degree SoS proof system. We will reprove the lemmas for expansion for pseudorandom sets in HD-walks proved

[26]This is just the potential measured on the sub-instance I when restricted to the induced subgraph given by the link X_τ.

in Section 7.8 and Section 7.9, but this time carefully making sure that they are in the low-degree SoS proof system.

Theorem 7.10.8 (SoS Structure Theorem for HD-Walks)**.** *Let (X, Π) be a two-sided γ-local-spectral expander and M be a k-dimensional, complete HD-walk on X with $\gamma \leq w(M)^{-1}h(M)^{-2}2^{-\Omega(k)}$. Then for any $f \in C_k$ and any $0 \leq \ell \leq k/2$, the expansion[27] of f with respect to M is large:*

$$\vdash_2 \ \langle f, (I - M)f\rangle \geq (1 - \lambda_{\ell+1}) \left((1 - c_1\gamma)(\mathbb{E}\, f + B(f)) - \binom{k}{\ell} \langle D_k^\ell f, D_k^\ell f\rangle \right),$$

where $B(f) = \mathbb{E}[f^2 - f]$ represents the Booleanity constraints and $c_1 \leq 2^{O(k)}$.

The proof of Theorem 7.10.8 relies on a number of SoS variants of properties of the HD-Level-Set Decomposition used in the previous sections, namely approximate orthogonality and the relation between $\|f_i\|$ and $\|g_i\|$.

Lemma 7.10.9. *Let (X, Π) be a two-sided γ-local-spectral expander with $\gamma \leq 2^{-\Omega(k)}$ and $f \in C_k$. Then for $f_i = U_i^k g_i \in V_k^i$:*

$$\vdash_2 \langle f_i, f_i\rangle \in \left(\frac{1}{\binom{k}{i}} \pm c_1\gamma \right) \langle g_i, g_i\rangle,$$

where $c_1 \leq O(k^2)$. Further, weak variants of approximate orthogonality also have degree 2 proofs:

$$\vdash_2 \langle f_i, f_i\rangle \leq (1 + c_2\gamma)\langle f, f\rangle,$$

and

$$\vdash_2 \langle f, f_i\rangle \geq -c_3\gamma\langle f, f\rangle,$$

where $c_2, c_3 \leq 2^{O(k)}$

[27] This notion of expansion varies slightly from our previous definition, coinciding (up to normalization) for Boolean functions.

The proof of Lemma 7.10.9 follows from combining arguments in [111] with standard SoS tricks. For completeness, we give the proof in Section 7.14.3. With Lemma 7.10.9 in hand, Theorem 7.10.8 follows similarly to Theorem 7.8.7 and Theorem 7.9.5.

Proof of Theorem 7.10.8. As in the proof of Theorem 7.9.5, it is sufficient to bound the weight of f onto low levels of the HD-Level-Set decomposition to prove that the expansion of f is small. First, note that given the function f, the low-level decomposition functions f_i's are explicit linear functions of the coordinates of f [111]. We will thus show that Lemma 7.8.2 has an SoS proof, that is the following relation between $\|D_\ell^k f\|^2$ and f's weight holds:

$$\vdash_2 \quad \sum_{j=0}^{\ell} \langle f, f_j \rangle \leq \binom{k}{\ell} \langle D_k^\ell f, D_k^\ell f \rangle + 2^{O(k)} \gamma \langle f, f \rangle. \tag{7.9}$$

Before proving Equation (7.9), we check that if the above inequality holds then the theorem statement follows. To see this, first recall from Section 7.7 that $Mf_i = \lambda_i f_i + \Gamma g_i$ where Γ is a matrix with spectral norm $\leq w(M)h(M)^2 2^{O(k)}\gamma$ which we call C for convenience. It will be useful to have an SoS upper bound on $\langle f, \Gamma g_i \rangle$, which we will use multiple times throughout this proof. First by an SoS version of Cauchy-Schwarz we get that for all real constants $\zeta > 0$:

$$\vdash_2 \langle f, \Gamma g_i \rangle \leq \frac{\zeta}{2} \langle f, f \rangle + \frac{1}{2\zeta} \langle \Gamma g_i, \Gamma g_i \rangle.$$

Now substituting $\zeta = C$ and simplifying using the spectral norm bounds on Γ and Lemma 7.10.9 we get:

$$\begin{aligned}
\langle f, \Gamma g_i \rangle &\leq \frac{C}{2} \langle f, f \rangle + \frac{1}{2C} \langle \Gamma g_i, \Gamma g_i \rangle \\
&\leq \frac{C}{2} \langle f, f \rangle + \frac{C}{2} \langle g_i, g_i \rangle \\
&\leq \frac{C}{2} \langle f, f \rangle + \frac{C}{2} \left(\binom{k}{i} + O(\gamma) \right) \langle f_i, f_i \rangle,
\end{aligned}$$

$$\leq c_1\gamma\langle f, f\rangle,$$

where all inequalities are degree 2 SoS inequalities and $c_1 \leq w(M)h(M)^2 2^{O(k)}$. With this in hand and assuming Equation (7.9), the result follows from expanding out $\langle f, (I-M)f\rangle$ and applying Lemma 7.10.9:

$$\begin{aligned}
\langle f, (I-M)f\rangle &= \langle f, f\rangle - \sum_{i=0}^{k} \lambda_i \langle f, f_i\rangle - \sum_{i=0}^{k} \langle f, \Gamma g_i\rangle \\
&\geq (1 - c_1\gamma)\langle f, f\rangle - \sum_{i=0}^{\ell} \lambda_i \langle f, f_i\rangle - \sum_{i=\ell+1}^{k} \lambda_i \langle f, f_i\rangle \\
&\geq (1 - c_2\gamma)\langle f, f\rangle - \sum_{i=0}^{\ell} \langle f, f_i\rangle - \lambda_{\ell+1} \sum_{i=\ell+1}^{k} \langle f, f_i\rangle \\
&= (1 - c_2\gamma)\langle f, f\rangle - \sum_{i=0}^{\ell} \langle f, f_i\rangle - \lambda_{\ell+1}\langle f, f\rangle + \lambda_{\ell+1} \sum_{i=0}^{\ell} \langle f, f_i\rangle \\
&= (1 - \lambda_{\ell+1}) \left((1 - c_3\gamma) \langle f, f\rangle - \sum_{i=0}^{\ell} \langle f, f_i\rangle \right) \\
&\geq (1 - \lambda_{\ell+1}) \left((1 - c_4\gamma) \langle f, f\rangle - \binom{k}{\ell} \langle D_k^\ell, D_k^\ell\rangle \right) \\
&= (1 - \lambda_{\ell+1}) \left((1 - c_4\gamma) \mathbb{E}[f] + (1 - c_4\gamma) B(f) - \binom{k}{\ell} \langle D_k^\ell, D_k^\ell\rangle \right)
\end{aligned}$$

where $c_1, c_2, c_3, c_4 \leq w(M)h(M)^2 2^{O(k)}$, and $B(f) = \mathbb{E}[f^2 - f]$.

It remains to prove Equation (7.9). This follows from a similar modification of Lemma 7.8.2. Notice that by the adjointness of D and U, it is enough to analyze the walk $U_\ell^k D_\ell^k$:

$$\langle D_\ell^k f, D_\ell^k f\rangle = \langle f, U_\ell^k D_\ell^k f\rangle$$

Using Proposition 7.7.5 and the assumption $\ell \leq k/2$, we can decompose the righthand side as:

$$\binom{k}{\ell} \langle f, U_\ell^k D_\ell^k f\rangle = \sum_{j=0}^{\ell} \binom{k-j}{\ell-j} \langle f, f_j\rangle + \sum_{j=0}^{k} \langle f, \Gamma g_j\rangle$$

where $\|\Gamma\| \leq w(M)h(M)^2 2^{O(k)}\gamma := C$. Noting that $\binom{k-j}{\ell-j}$ is at least 1 for $0 \leq j \leq \ell$, we can apply the upper bound on $\langle f, \Gamma g_j \rangle$ proved above and Lemma 7.10.9 to get:

$$\begin{aligned}\binom{k}{\ell}\langle D_\ell^k f, D_\ell^k f\rangle &\geq \sum_{j=0}^{\ell}\langle f, f_j\rangle - c_5\gamma\langle f, f\rangle - \sum_{j=0}^{k}\langle f, \Gamma g_j\rangle \\ &\geq \sum_{j=0}^{\ell}\langle f, f_j\rangle - c_6\gamma\langle f, f\rangle\end{aligned}$$

where all the constants c_i's are less than $w(M)h(M)^2 2^{O(k)}$ and all inequalities are degree-2 SoS. □

It is worth giving a brief qualitative comparison of this result to a similar version for the Johnson graphs in [37]. In particular, Theorem 7.10.8 not only gives a tighter bound (by a factor of $\exp(r)$), but perhaps more importantly shows how viewing the problem from the framework of high dimensional expansion demystifies the original Fourier analytic proof. This understanding allows us to extend the structural result well beyond the Johnson graphs to all HD-walks. In fact, this result also holds for the more general class of expanding posets [111] (albeit with different parameters). We leave their discussion to future work.

Given this structure theorem, we use the BBKSS' framework to show that the global potential (Definition 7.14.3) on a link is high. This follows because of the properties of the potential function and the pseudodistribution over $(1-\eta)$-satisfying assignments. In general, the potential function $\Phi^I_{\beta,\nu}(\mu)$ turns out to be an average taken over non-expanding partitions of the graph. Using our structure theorem, we get that every non-expanding partition must have a large variance across links, and therefore there must exist a link X_τ such that the partition is large even when restricted to X_τ. As a result, there must be be a link X_τ where the partitions are large on average too, which corresponds to a quantity called the global potential restricted to X_τ, denoted by $\Phi^I_{\beta,\nu}(\mu|_\tau)$, being large. The proof is essentially the same as [37, Lemma 6.9], but we include it in Section 7.14.5

for completeness.

Proposition 7.10.10. *Let M be a k-dimensional complete HD-walk on a d-dimensional, two-sided γ-local-spectral expander satisfying $\gamma w(M)2^{O(h(M)+k)} \leq 1/2^k < \eta < 0.02$ and $d > k$, and I be an affine unique games instance over M with value at least $1-\eta$. Let $r(\eta) = R_{1-16\eta}(M) - 1$, and assume $r \leq k/2$. Let $\beta = 19\eta$ and $\nu = \frac{\eta}{56\binom{k}{r}}$. Then given a degree-$\widetilde{O}(1/\nu)$ pseudodistribution μ satisfying the axioms $\mathcal{A}_I$, there exists an r-link X_τ such that the global potential restricted to X_τ is large:*

$$\Phi^I_{\beta,\nu}(\mu|_\tau) \geq \frac{1}{4\binom{k}{r}}.$$

In the next lemma, we will relate the global potential to the potential induced on X_τ. In particular, we'll show that since the global potential on X_τ is high (Proposition 7.10.10), the potential induced on X_τ, $\Phi^{I_\tau}_{\beta,\nu}(\mu)$, must also be high. Since this is just the usual potential function when applied to the sub instance $I|_\tau$ corresponding to the subgraph induced by X_τ (Definition 7.14.4), we can then apply the rounding theorem of BBKSS Theorem 7.10.4 to surmise that the CR-value of X_τ is high.

Lemma 7.10.11. *Assume the conditions of Proposition 7.10.10 hold. Let X_τ be an r-link for $r = R_{1-16\eta}(M) - 1$ with high global potential: $\Phi^I_{\beta,\nu}(\mu|_\tau) \geq \frac{1}{4\binom{k}{r}}$, for $\beta = 19\eta$ and $\nu = \frac{\eta}{56\binom{k}{r}}$. Then the potential induced on X_τ is also high:*

$$\Phi^{I|_\tau}_{\eta,\nu}(\mu) \geq \frac{1}{8\binom{k}{r}}.$$

We prove Lemma 7.10.11 in Section 7.14.5, but it's worth pausing to discuss the proof, especially in how it differs from the analogous result in BBKSS [37, Claim 6.11]. Recall that the main idea behind Lemma 7.10.11 is to relate the global potential on X_τ (bounded in Proposition 7.10.10) to the potential induced on X_τ itself (which implies high CR-Value by Theorem 7.10.4). In a bit more detail, the global potential on a link depends

on the "value" of the vertices in the link (which measures the value of an assignment at the vertex), but is measured with respect to its neighbors across the *entire graph*. On the other hand, the induced potential on the link depends on the value of vertices when measured only with respect to the edges *inside* the link (see Definitions 7.14.3,7.14.4 for exact details). It is possible to relate the two potentials by observing that the internal-value of a vertex (value with respect to only the neighbors inside the link) can decrease by at most an additive factor equal to its edge-expansion inside the link (i.e. the fraction of edges incident on the vertex that leave the set). By leveraging machinery developed in Section 7.7 and some additional properties, we show not only that the expansion of links is small (by Theorem 7.9.2), but in fact that this holds approximately vertex by vertex: for most vertices in the link, only an $O(\eta)$-fraction of their edges are outgoing. We prove the following claim formally in Section 7.14.5:

Lemma 7.10.12. *Let M be a k-dimensional HD-walk on d-dimensional two-sided γ-local-spectral expander satisfying $\gamma \leq w(M)^{-1}2^{-\Omega(h(M)+k)}$ and $d > k$. Then for every i-link X_τ, the deviation of the random variable $\phi_{X_\tau}(v)$ ($v \sim X_\tau$) is small:*

$$\mathop{\mathbb{E}}_{v\sim X_\tau}[|\phi_{X_\tau}(v) - \phi(X_\tau)]|] \leq \frac{1}{2^{11k}},$$

where $\phi_{X_\tau}(v)$ denotes the fraction of edges incident on $v \in X_\tau$ that leave X_τ.

BBKSS use an exact version of this statement for the Johnson (that expansion holds vertex-by-vertex) to prove their analogous version of Lemma 7.10.11. We relax the conditions required for their proof and show that the approximate statement state above is sufficient in Section 7.14.5.

Finally, we complete the section by using Proposition 7.10.10 and Lemma 7.10.11 to prove Proposition 7.10.5 (and thereby Theorem 7.10.1 as well). We restate the proposition here for convenience:

Proposition 7.10.13 (Restatement of Proposition 7.10.5)**.** *Let M be a k-dimensional complete HD-walk on a d-dimensional, two-sided γ-spectral expander with $\gamma \leq w(M)^{-1}2^{-\Omega(h(M)+k)}$ and $d > k$, and I be an affine unique games instance over M with value at least $1-\eta$ where $1/2^k \leq \eta < .02$. Let $r = r(\eta) = R_{1-16\eta}(M) - 1$, and assume $r \leq k/2$. Then given a degree-$\widetilde{O}\left(\frac{1}{\eta}\binom{k}{r}\right)$ pseudodistribution μ satisfying the axioms $\mathcal{A}_I$, we can find in time $|X(k)|^4$ an r-link X_τ with CR-$\mathrm{val}_\mu(X_\tau) \geq \Omega\left(\frac{\eta}{\binom{k}{r}}\right)$.*

Proof. Let $\beta = 19\eta$ and $\nu = \frac{\eta}{56\binom{k}{r}}$ as in Proposition 7.10.10. By Proposition 7.10.10, there exists an r-link X_τ with high global potential:

$$\Phi^I_{\beta,\nu}(\mu|_\tau) \geq \frac{1}{4\binom{k}{r}}.$$

By Lemma 7.10.11, this implies that X_τ also has high induced potential:

$$\Phi^{I|_\tau}_{\eta,\nu}(\mu) \geq \frac{1}{8\binom{k}{r}}.$$

As mentioned previously, this is just the usual potential function on the instance induced by X_τ (Definition 7.14.4), so we may apply the rounding theorem of BBKSS Theorem 7.10.4 to bound X_τ's CR-Value. In particular, since we have set ν such that $(\eta - \nu) \geq \Omega(\eta)$ and $(\Phi^{I|_\tau}_{\eta,\nu}(\mu) - \nu) \geq \Omega\left(\frac{1}{\binom{k}{r}}\right)$, Theorem 7.10.4 implies that rounding μ, which is a degree $\widetilde{O}(1/\nu) = \widetilde{O}\left(\frac{1}{\eta}\binom{k}{r}\right)$ pseudodistribution as required, would give an assignment with large -value on X_τ:

$$\text{CR-}\,\mathrm{val}_\mu(X_\tau) \geq (\eta - \nu)(\Phi^{I|_\tau}_{\eta,\nu}(\mu) - \nu) \geq \Omega\left(\frac{\eta}{\binom{k}{r}}\right).$$

Now that we have shown that there exists an r-link X_τ with large CR-value, it remains to show that we can efficiently find such a link. We can do a brute-force enumeration over all r-links in time $|X(r)|$, compute every CR-value in time $|X(r)|^3$, and therefore in time $|X(r)|^4$ find a link with large CR-value. Since $|X(k)| \geq |X(r)|$ is a standard consequence

of X being a local-spectral expander [111] the proposition follows. □

7.11 Acknowledgements

This work stemmed in part from collaboration at the Simons Institute for the Theory of Computing, Berkeley during the 2019 summer cluster: "Error-Correcting Codes and High-Dimensional Expansion". We thank the Simons institute for their hospitality and the organizers of the cluster for creating such an opportunity. Further, the authors would like to thank Sankeerth Rao and Yotam Dikstein for useful discussions in the early stages of this work, and Ella Sharakanski for discussion regarding HD-walk decompositions. We thank Madhu Sudan and Sam Hopkins for helpful comments on initial drafts of this work. Finally, we additionally owe many thanks to Sam Hopkins for discussions on the Sum of Squares framework, and to Vedat Alev for his insights on the spectral structure of higher order random walks.

7.12 Proof of Lemma 7.7.3

In this section, we prove a strengthening of the main technical lemma of DDFH Section 8 [111, Claim 8.8], which allows for better control of error propagation.

Lemma 7.12.1 (Strengthened Claim 8.8 [111])**.** *Let (X, Π) be a d-dimensional two-sided γ-local-spectral expander. Then for all $j < k < d$:*

$$D_{k+1}U_{k-j}^{k+1} - \frac{j+1}{k+1}U_{k-j}^{k} - \frac{k-j}{k+1}U_{k-j-1}^{k}D_{k-j} = \sum_{i=-1}^{j-1}\frac{k-i}{k+1}U_{k-1-i}^{k}\Gamma_i U_{k-j}^{k-1-i}$$

where $\|\Gamma_i\| \leq \gamma$.

Proof. The proof follows by a simple induction. The base cases, $j = 0$ and $k < d$, follow

immediately from Equation (7.6). For the inductive step, consider:

$$D^{k+1}U^{k+1}_{k-(j+1)} = \left(D^{k+1}U^{k+1}_{k-j} - \frac{j+1}{k+1}U^{k}_{k-j} - \frac{k-j}{k+1}U^{k}_{k-j}D_{k-j}\right)U_{k-j-1} + \frac{j+1}{k+1}U^{k}_{k-j-1} + \frac{k-j}{k+1}U^{k}_{k-j-1}D_{k-j}U_{k-j-1}$$

By the inductive hypothesis, the first term on the RHS may be written as:

$$\left(D^{k+1}U^{k+1}_{k-j} - \frac{j+1}{k+1}U^{k}_{k-j} - \frac{k-j}{k+1}U^{k}_{k-j}D_{k-j}\right)U_{k-j-1} = \sum_{i=-1}^{j-1}\frac{k-i}{k+1}U^{k}_{k-i-1}\Gamma_i U^{k-1-i}_{k-j-1},$$

where $\|\Gamma_i\| \leq \gamma$. For the latter term, consider flipping DU and UD. By Equation (7.6) we have:

$$\frac{k-j}{k+1}U^{k}_{k-j-1}D_{k-j}U_{k-j-1} = U^{k}_{k-j-1}\left(\frac{1}{k+1}I + \frac{k-j-1}{k+1}U_{k-j-2}D_{k-j-1} + \frac{k-j}{k+1}\Gamma_j\right),$$

for some Γ_j satisfying $\|\Gamma_j\| \leq \gamma$. Combining these observations yields the desired result:

$$\begin{aligned}
&D^{k+1}U^{k+1}_{k-(j+1)} - \frac{(j+1)+1}{k+1}U^{k}_{k-j+1} + \frac{k-(j+1)}{k+1}U^{k}_{k-(j+1)-1}D_{k-j+1} \\
&= D^{k+1}U^{k+1}_{k-(j+1)} - \frac{j+1}{k+1}U^{k}_{k-j-1} - U^{k}_{k-j-1}\left(\frac{1}{k+1}I - \frac{k-j-1}{k+1}U_{k-j-2}D_{k-j-1}\right) \\
&= \left(\sum_{i=-1}^{j-1}\frac{k-i}{k+1}U^{k}_{k-i-1}\Gamma_i U^{k-1-i}_{k-j-1}\right) + \frac{k-j}{k+1}U^{k}_{k-(j+1)}\Gamma_j \\
&= \sum_{i=-1}^{j}\frac{k-i}{k+1}U^{k}_{k-1-i}\Gamma_i U^{k-1-i}_{k-(j+1)}.
\end{aligned}$$

□

We now show how to use this strengthened result to prove tighter bounds on the quadratic form $\langle f, N_k^j f\rangle$ which implies a stronger version of Lemma 7.7.4 as an immediate corollary. This improvement mainly matters in the regime where $\gamma \leq 2^{-ck}$ for c a small constant.

Proposition 7.12.2. *Let (X, Π) be a d-dimensional γ-local-spectral expander with γ satisfying $\gamma \leq 2^{-\Omega(k+j)}$, $k + j \leq d$, and $f_\ell \in V_k^\ell$. Then:*

$$\langle f_\ell, N_k^j f_\ell \rangle = \frac{\binom{k}{\ell}}{\binom{k+j}{\ell}} \left(1 \pm \frac{j(j+2k+2\ell+3)}{4}\gamma \pm c_3(k,j,\ell)\gamma^2\right) \langle f_\ell, f_\ell \rangle$$

where $c_3(k,j,\ell) = O((k+j)^3\binom{k+j}{\ell})$.

Proof. We proceed by induction on j. We will prove a slightly stronger statement for the base-case $j = 1$:

$$\langle f_\ell, D_{k+1}U_k f_\ell \rangle = \left(\frac{k+1-\ell}{k+1} \pm \frac{(k-\ell+1)(k+\ell+2)}{2(k+1)}\gamma \pm c_2(k,\ell)\gamma^2\right) \langle f_\ell, f_\ell \rangle,$$

where $c_2(k,\ell) = O(k^3\binom{k}{\ell})$. Recall that f_ℓ may be expressed as $U_\ell^k g_\ell$, for $g_\ell \in H^\ell$. For notational convenience, we write $f_\ell^i = U_\ell^i g_\ell$. Then we may expand the inner product based on Lemma 7.7.3, and simplify based on applying the naive bounds on N_k^i given by Corollary 7.7.6:

$$\begin{aligned}
&\langle f_\ell, D_{k+1}U_k f_\ell \rangle \\
=&\langle f_\ell, D_{k+1}U_\ell^{k+1} g_\ell \rangle \\
=&\frac{k-\ell+1}{k+1}\langle f_\ell, f_\ell \rangle + \sum_{i=-1}^{k-\ell-1} \langle f_\ell, \frac{k-i}{k+1} U_{k-1-i}^k \Gamma_i U_\ell^{k-1-i} g_\ell \rangle \\
=&\frac{k-\ell+1}{k+1}\langle f_\ell, f_\ell \rangle + \sum_{i=-1}^{k-\ell-1} \frac{k-i}{k+1} \langle N_{k-i-1}^{i+1} f_\ell^{k-1-i}, \Gamma_i f_\ell^{k-1-i} \rangle \\
=&\frac{k-\ell+1}{k+1}\langle f_\ell, f_\ell \rangle + \sum_{i=-1}^{k-\ell-1} \frac{k-i}{k+1} \frac{\binom{k-i-1}{\ell}}{\binom{k}{\ell}} \langle f_\ell^{k-1-i}, \Gamma_i f_\ell^{k-1-i} \rangle + \sum_{i=-1}^{k-\ell-1} \frac{k-i}{k+1} \langle h_i, \Gamma_i f_\ell^{k-1-i} \rangle
\end{aligned}$$

where $\|h_i\| \leq \gamma(k-\ell)(i+1)\|g_\ell\|$. We now apply Cauchy-Schwarz, and Lemma 7.7.4 to

collect terms in $\langle f_\ell, f_\ell \rangle$:

$$\begin{aligned}
&\langle f_\ell, D_{k+1} U_k f_\ell \rangle \\
&\quad = \frac{k-\ell+1}{k+1} \langle f_\ell, f_\ell \rangle \pm \gamma \sum_{i=-1}^{k-\ell-1} \frac{k-i}{k+1} \frac{\binom{k-i-1}{\ell}}{\binom{k}{\ell}} \langle f_\ell^{k-1-i}, f_\ell^{k-1-i} \rangle \pm a_1(k,\ell)\gamma^2 \langle g_\ell, g_\ell \rangle \\
&\quad = \frac{k-\ell+1}{k+1} \langle f_\ell, f_\ell \rangle \pm \gamma \sum_{i=-1}^{k-j-1} \frac{k-i}{k+1} \frac{1}{\binom{k}{j}} \langle g_\ell, g_\ell \rangle \pm a_2(k,\ell)\gamma^2 \langle g_\ell, g_\ell \rangle \\
&\quad = \frac{k-\ell+1}{k+1} \langle f_\ell, f_\ell \rangle \pm \gamma \sum_{i=-1}^{k-j-1} \frac{k-i}{k+1} \frac{\langle f_\ell, f_\ell \rangle}{(1 - c_1(k,\ell)\gamma)} \pm a_2(k,\ell)\gamma^2 \frac{\langle f_\ell, f_\ell \rangle}{(1 - c_1(k,\ell)\gamma)} \\
&\quad = \frac{k-\ell+1}{k+1} \left(1 \pm \frac{(k+\ell+2)}{2} \gamma \pm a_3(k,\ell)\gamma^2 \right) \langle f_\ell, f_\ell \rangle
\end{aligned}$$

where the final step comes from a Taylor expansion assuming γ sufficiently small, and $a_3(k,\ell) = O(k^3 \binom{k}{\ell})$.

The inductive step follows from noting that the canonical walk essentially acts like a product of upper walks from lower levels in the following sense:

$$\begin{aligned}
\langle f_\ell, N_k^j f_\ell \rangle &= \langle U_k^{k+j-1} f_\ell, D_{k+j} U_k^{k+j} f_\ell \rangle \\
&= \langle U_k^{k+j-1} f_\ell, N_{k+j-1}^1 (U_k^{k+j-1} f_\ell) \rangle.
\end{aligned}$$

Thus by the base-case and inductive hypothesis we get:

$$\begin{aligned}
\langle f_\ell, N_k^j f_\ell \rangle &= \langle U_k^{k+j-1} f_\ell, N_{k+j-1}^1 (U_k^{k+j-1} f_\ell) \rangle \\
&= \left(\frac{k+j-\ell}{k+j} \pm \frac{(k+j-\ell)(k+j+\ell+1)}{2(k+j)} \gamma \pm c_2(k+j-1,\ell)\gamma^2 \right) \langle f_\ell, N_k^{j-1} f_\ell \rangle \\
&= \frac{\binom{k}{\ell}}{\binom{k+j}{\ell}} \left(1 \pm \frac{(k+j+\ell+1)}{2} \gamma \pm \frac{k+j}{k+j-\ell} c_2(k+j-1,\ell)\gamma^2 \right) \\
&\qquad \cdot \left(1 \pm \frac{(j-1)(j+2k+2\ell+2)}{4} \gamma \pm c_3(k,j-1,\ell)\gamma^2 \right) \langle f_\ell, f_\ell \rangle \\
&= \frac{\binom{k}{\ell}}{\binom{k+j}{\ell}} \left(1 \pm \frac{j(j+2k+2\ell+3)}{4} \gamma \pm c_3(k,j,\ell)\gamma^2 \right) \langle f_\ell, f_\ell \rangle,
\end{aligned}$$

□

Notice that this immediately implies a stronger version of Lemma 7.7.4, since $\langle U_\ell^k g_\ell, U_\ell^k g_\ell \rangle = \langle N_\ell^{k-\ell} g_\ell, g_\ell \rangle$. Finally, we conjecture that a stronger result is true, and the error dependence on γ should in fact be $\exp(-\text{poly}(k)\gamma)$. Proving this would require a more careful and involved analysis of how the error term propogates.

7.13 Orthogonality and the HD-Level-Set Decomposition

In this section we discuss in a bit more depth the error in [239, Theorem 5.10], and further show by direct counter-example that its implication [242] that the HD-Level-Set is orthogonal does not hold. In [239], Kaufman and Oppenheim analyze an approximate eidgendecomposition of the upper walk N_k^1 for two-sided local-spectral expanders. They prove a specialized version of Theorem 7.6.2 for this case, and in particular that for sufficiently strong two-sided local-spectral expanders, the spectra of N_k^1 is divided into strips concentrated around the approximate eigenvalues of their decomposition. They call the span of each strip W^i, and note that the W^i form an orthogonal decomposition of the space. Let V^i be the space in the original approximate eigendecomposition corresonding to strip W^i. Kaufman and Oppenheim claim in [239, Theorem 5.10] that the W^i are closely related to the original approximate decomposition in the following sense:

$$\forall \phi \in C_k : \|P_{W^i}\phi\| \leq c\|P_{V^i}\phi\|$$

for some constant $c > 0$, where P_{W^i} and P_{V^i} are projection operators. Unfortunately, this relation cannot hold, as it implies [242] that the HD-Level-Set Decomposition is orthogonal for sufficiently strong two-sided local-spectral expanders, which we will show below is false by direct example. In slightly greater detail, the issue in the argument is the following.

The authors show that for any $j \neq i$:

$$\|P_{W^j} P_{V^i}\| \leq c',$$

for some small constant c', and then claim that this fact implies for any $\phi \in C_k$:

$$\|P_{W^j} P_{V^i} \phi\| \leq c' \|P_{V^j} \phi\|.$$

Unfortunately, this is not true—the righthand side should read P_{V^i} rather than P_{V^j} for the relation to hold, but this makes it impossible to compare $P_{W^i}\phi$ solely to $P_{V^i}\phi$.

We now move to showing that for any $\gamma > 0$, there exists a two-sided γ-local-spectral expander such that the HD-Level-Set Decomposition is not orthogonal, which implies [239, Theorem 5.10] cannot hold by arguments of [242].

Proposition 7.13.1. *For any $\gamma > 0$, there exists a two-sided γ-local-spectral expander such that the HD-Level-Set Decomposition is not orthogonal.*

Proof. Our construction is based off of a slight modification of the complete complex $J(n, 3)$. In particular, we consider the uniform distribution Π over triangles $X = \binom{[n]}{3} \setminus (123)$. It is not hard to see through direct computation that (X, Π) is a two-sided $O(1/n)$-local-spectral expander. Recall that the link of 1, $U_1^3 \mathbb{1}_1$, lies in $V_3^0 \oplus V_3^1$. Our goal is to prove the existence of a function $f = Ug \in V_3^2$ such that the inner product:

$$\langle U_1^3 \mathbb{1}_1, f \rangle \propto \sum_{(1xy) \in X} g(1x) + g(1y) + g(xy) \tag{7.10}$$

is non-zero. To do this, we first simplify the above expression assuming $g \in \mathrm{Ker}(D_2)$, which we recall implies the following relations:

$$\forall y \in [n] : \sum_{(xy) \in X(2)} \Pi_2(xy) g(xy) = 0.$$

In particular, summing over all $y \in [n]$ gives

$$\sum_{(xy)\in X(2)} \Pi_2(xy)g(xy) = 0.$$

Notice further that by definition of Π_2, we have $\Pi_2(12) = \Pi_2(13) = \Pi_2(23) = \frac{n-3}{3\binom{n}{3}-3}$, and otherwise $\Pi_2(xy) = \frac{n-2}{3\binom{n}{3}-3}$. We then may write:

$$\sum_{\substack{(1x)\in X(2):\\ x\notin[3]}} g(1x) = -\frac{n-3}{n-2}\left(g(12)+g(13)\right),$$

$$\sum_{\substack{(xy)\in X(2):\\ (xy)\notin[3]\times[3]}} g(xy) = -\frac{n-3}{n-2}(g(12)+g(13)+g(23)).$$

Plugging this into Equation (7.10), the inner product drastically simplifies to depend only on $g(23)$. To see this, we separate the inner product into two terms and deal with each separately:

$$\sum_{(1xy)\in X} g(1x)+g(1y)+g(xy) = \left(\sum_{(1xy)\in X} g(1x)+g(1y)\right) + \sum_{(1xy)\in X} g(xy)$$

We start with the former. Notice that each face $(1z)$ in this term is counted exactly the number of times it appears in a triangle in X, and further that this is exactly how Π_2 is defined. Thus we have:

$$\left(\sum_{(1xy)\in X} g(1x)+g(1y)\right) \propto \sum_{(1x)\in X(2)} \Pi_2(1x)g(1x) = 0.$$

It is left to analyze the latter term. Since (123) is not in our complex, we may write:

$$\begin{aligned}\sum_{(1xy)\in X} g(xy) &= \left(\sum_{\substack{(xy)\in X(2):\\ (xy)\notin[3]\times[3]}} g(xy)\right) - \left(\sum_{\substack{(1x)\in X(2):\\ x\notin[3]}} g(1x)\right) \\ &= -\frac{n-3}{n-2}(g(12)+g(13)+g(23)) + \frac{n-3}{n-2}(g(12)+g(13)) \\ &= -\frac{n-3}{n-2}g(23).\end{aligned}$$

Thus it remains to show that there exists $g \in \mathrm{Ker}(D_2)$ such that $g(23) \neq 0$. Note that the kernel of D_2 is exactly the space of solutions to the underdetermined linear system of equations given by $D_2 g(i) = 0$ for all $1 \leq i \leq n$. Thus we can check if a solution exists with $g(23) = c$ for $c \neq 0$ by ensuring that this constraint is linearly independent of the $D_2 g(i)$. This can be checked through a direct but tedious computation that we leave to the reader. □

7.14 Unique Games

7.14.1 Random-walks and Weighted graphs

Unique games are defined on weighted, *undirected* constraint graphs, unlike most of the walks we analyze in the previous section. However, it is not hard to see that every self-adjoint random walk corresponds to some underlying undirected graph. In particular, recall that given a weighted, undirected graph $G(V, E)$, G induces a random walk on vertices where the transition probability from $x \in V$ to $y \in V$ is given by the normalized weights:

$$P(x, y) = \frac{W(s, t)}{\sum_{v\in N(s)} W(s, v)}.$$

In fact, *every* self-adjoint walk can be described as such a process.

Lemma 7.14.1. *Let M be a k-dimensional HD-walk on a weighted simplicial complex (X, Π). Define G_M to be the graph whose vertex set is $X(k)$ and whose edge set consists of any pair (s,t) such that $M(s,t) \neq 0$. Further, endow the edges of G_M with weight:*

$$W(s,t) = \Pi_k(s)M(s,t).$$

Then M is the random walk induced by G_M.

Proof. The crucial observation for this proof is the following implication due to M being self adjoint:

$$\forall s,t \in X(k) : \Pi_k(s)M(s,t) = \Pi_k(t)M(t,s).$$

Given this fact, let $P_G(s,t)$ denote the transition probability of the induced walk on G. We have:

$$\begin{aligned} P_G(s,t) &= \frac{\Pi_k(s)M(s,t)}{\sum_{v \in N(s)}} \Pi_k(s)M(s,t) \\ &= \frac{M(s,t)}{\sum_{v \in N(s)}} M(s,t) \\ &= M(s,t) \end{aligned}$$

□

7.14.2 The Shift-Partition Potential

The analysis of the Iterated Condition&Round algorithm relies on analysing a quantity called the Approximate shift-partition potential which was defined in BBKSS. For completeness we include the definitions of the various potential functions used in this section.

For analysing the Condition&Round subroutine, define a low-degree polynomial $\Phi_{\beta,\nu} : \Sigma^V \times \Sigma^V \to [0,\infty)$ called the "approximate shift partition potential":

Definition 7.14.2 (Approximate Shift-Partition Potential)**.** For any $\nu, \beta \in (0,1)$, and two assignments $X, X' \in \Sigma^V$ define the *approximate shift-partition potential* to be the quantity

$$\Phi_{\beta,\nu}(X,X') = \sum_{s\in\Sigma} \mathbb{E}\, u \mathbb{1}[X_u - X'_u = s] \cdot p_{\beta,\nu}(\mathrm{val}_u(X))^2,$$

for $p_{\beta,\nu}(x)$ the degree-$\widetilde{O}(1/\nu)$ polynomial which SoS-certifiably approximates the indicator $\mathbb{1}[x \geq \beta]$ for $x \in [0,1]$ up to an error of ν. The exact definition of $p_{\beta,\nu}$ and the properties required of it are described in more detail in Section 7 of BBKSS and we omit them here.

The approximate shift-partition potential on a pseudodistribution μ of degree at least $O(\deg(\Phi_{\beta,\nu}))$ is defined as:

$$\Phi_{\beta,\nu}(\mu) = \widetilde{\mathbb{E}}_{\mu}[\Phi_{\beta,\nu}(X,X')].$$

For analysing the Iterated Condition&Round algorithm we further need to define the potential when restricted to a subgraph which in our case will be a link of the complex.

Definition 7.14.3 (Global shift-potential restricted to links)**.** Let $I = (M,\Pi)$ be a UG instance on a complex X where M is a complete random walk on $X(k)$ with stationary distribution π_k. For any $\nu, \beta \in (0,1)$ and a link X_τ of the complex X, define the *approximate global shift-partition potential restricted to* X_τ to be the quantity:

$$\Phi_{\beta,\nu}(X,X')|_\tau = \sum_{s\in\Sigma} \mathop{\mathbb{E}}_{u\sim\pi_k|\tau} [\mathbb{1}[X_u - X'_u = s] \cdot p_{\beta,\nu}(\mathrm{val}_u(X))]^2,$$

where $\mathrm{val}_u(X) = \mathbb{E}\,(u,v) \sim M \mathbb{1}[X \text{ satisfies } (u,v))$, and as before $p_{\beta,\nu}(x)$ is the degree-$\widetilde{O}(1/\nu)$ polynomial that ν-approximates the indicator function $\mathbb{1}[x \geq \beta]$. The global potential restricted to X_τ with respect to a pseudodistribution μ will be denoted by

$\Phi_{\beta,\nu}(\mu|_\tau)$:

$$\Phi_{\beta,\nu}(\mu|_\tau) = \widetilde{\mathbb{E}}_{\mu}[\Phi_{\beta,\nu}(X, X')|_\tau].$$

Note that the global shift-partition potential measures the size of the global partition inside X_τ, and although the expectation is taken only over the vertices $u \in X_\tau$, $\mathrm{val}_u(X)$ is a function of *all* the edges in M that are incident on u, not just the edges in X_τ. We will also need to consider the potential induced on a link, which is defined as applying the potential function (Definition 7.14.2) to the graph induced by X_τ:

Definition 7.14.4 (Induced Shift-Potential on a Link)**.** Let $I = (M, \Pi)$ be a UG instance on a complex X where M is a complete random walk on $X(k)$ with stationary distribution π_k. For any $\nu, \beta \in (0, 1)$ and a link X_τ of the complex X, define the *approximate induced shift-partition potential on* X_τ to be the quantity:

$$\Phi^\tau_{\beta,\nu}(X, X') = \sum_{s\in\Sigma} \mathop{\mathbb{E}}_{u\sim\pi_k|_\tau} [\mathbb{1}[X_u - X'_u = s] \cdot p_{\beta,\nu}(\mathrm{val}^\tau_u(X))]^2,$$

where $\mathrm{val}^\tau_u(X) = \mathbb{E}\,(u, v) \sim M|_\tau \mathbb{1}[X \text{ satisfies } (u, v))$. The induced potential on X_τ with respect to a pseudodistribution μ will be denoted by $\Phi^\tau_{\beta,\nu}(\mu)$:

$$\Phi^\tau_{\beta,\nu}(\mu) = \widetilde{\mathbb{E}}_{\mu}[\Phi^\tau_{\beta,\nu}(X, X')].$$

Note that in the above definition the value of a vertex $u \in X_\tau$ is measured only with respect to the edges incident on u that lie *inside* the link X_τ.

7.14.3 Sum of Squares and the HD-Level-Set Decomposition

This section is devoted to proving Lemma 7.10.9 which we separate into two parts. First, we examine the relation between $\|f_i\|$ and $\|g_i\|$.

Lemma 7.14.5 (Restated Lemma 7.10.9 (Part 1))**.** *Let* (X, Π) *be a* d*-dimensional two-sided*

γ-local-spectral expander. Then for any $f_i = U_i^k g_i \in V_k^i$:

$$\vdash_2 \langle f_i, f_i \rangle \in \left(\frac{1}{\binom{k}{i}} \pm \frac{(k-i)(k+1)}{i+2} \gamma \right) \langle g_i, g_i \rangle$$

Proof. It is sufficient to prove the following equality:

$$\langle f_i, f_i \rangle = \frac{1}{\binom{k}{i}} \langle g_i, g_i \rangle + \langle g_i, \Gamma g_i \rangle, \tag{7.11}$$

where $\|\Gamma\| < \frac{(k-i)(k+1)}{i+2}\gamma$. To see why, note that:

$$\langle g_i, \Gamma g_i \rangle = \left\langle g_i, \frac{\Gamma + \Gamma^*}{2} g_i \right\rangle$$

where Γ^* is the adjoint of Γ. Since $\frac{\Gamma+\Gamma^*}{2}$ is self-adjoint and $\|\Gamma^*\| = \|\Gamma\|$, we have both:

1. $\vdash_2 \left\langle g_i, \frac{\Gamma+\Gamma^*}{2} g_i \right\rangle \leq \frac{(k-i)(k+1)}{i+2}\gamma \langle g_i, g_i \rangle$
2. $\vdash_2 \left\langle g_i, \frac{\Gamma+\Gamma^*}{2} g_i \right\rangle \geq -\frac{(k-i)(k+1)}{i+2}\gamma \langle g_i, g_i \rangle$

as desired. To prove Equation (7.11), we induct on k. The base case $k = i$ is trivial. Assume $k > i$, then we may write:

$$\langle f_i, f_i \rangle = \langle U_i^{k-1} g_i, D_k U_i^k g_i \rangle$$

In order to apply the inductive hypothesis, we recall the machinery of [111, Claim 8.8] for pushing D_k through U_i^k. In particular:

$$\| D_k U_i^k - \frac{k-i}{k} U_i^{k-1} - \frac{i}{k} U_{i-1}^{k-1} D_i \| \leq (k-i)\gamma.$$

Since g_i lies in the kernel of D_i, combining this with our initial observation gives:

$$\langle U_i^{k-1} g_i, D_k U_i^k g_i \rangle = \frac{k-i}{k} \langle U_i^{k-1} g_i, U_i^{k-1} g_i \rangle + \langle g_i, \Gamma g_i \rangle,$$

where $\|\Gamma\| \leq (k-i)\gamma$. Since $U_i^{k-1} g_i \in V_i^{k-1}$. Applying the inductive hypothesis, we see that:

$$\begin{aligned}\langle f_i, f_i \rangle &= \frac{k-i}{k} \left(\frac{1}{\binom{k-1}{i}} \langle g_i, g_i \rangle + \langle g_i, \Gamma' g_i \rangle \right) + \langle g_i, \Gamma g_i \rangle \\ &= \frac{1}{\binom{k}{i}} \langle g_i, g_i \rangle + \langle g_i, \left(\frac{k-i}{k} \Gamma' + \Gamma \right) g_i \rangle,\end{aligned}$$

where:

$$\| \frac{k-i}{k} \Gamma' + \Gamma \| \leq \frac{(k-i)(k+1)}{i+2} \gamma$$

by the triangle inequality and inductive hypothesis. □

Second, we prove that a version of approximate orthogonality of the HD-Level-Set Decomposition has a low-degree SoS proof.

Lemma 7.14.6 (Restated Lemma 7.10.9 (Part 2))**.** *let (X, Π) be a two-sided γ-local-spectral expaner with $\gamma \leq 2^{-\Omega(k)}$ and $f \in C_k$. Then for any $f_i = U_i^k g_i \in V_k^i$ and $f_j = U_j^k g_j \in V_k^j$ we have:*

$$\vdash_2 \langle f_i, f_i \rangle \leq (1 + c_1 \gamma) \langle f, f \rangle,$$

and

$$\vdash_2 \langle f, f_i \rangle \geq -c_2 \gamma \langle f, f \rangle,$$

where $c_1, c_2 \leq 2^{O(k)}$

Proof. First, note that by Lemma 7.7.3 and the fact that $g_j \in \mathrm{Ker}(D_j)$, we have that:

$$\langle f_i, f_j \rangle = \langle g_i, \Gamma g_j \rangle$$

where $\|\Gamma\| \leq 2^{O(k)}$. We can bound the latter by an SoS version of Cauchy Schwarz. In particular, we have that:

$$\begin{aligned}
\langle f_i, f_j \rangle &= \langle g_i, \Gamma g_i \rangle \\
&\leq \frac{\|\Gamma\|}{2}\langle g_i, g_i \rangle + \frac{1}{2\|\Gamma\|}\langle \Gamma g_j, \Gamma g_j \rangle \\
&\leq \frac{\|\Gamma\|}{2}\langle g_i, g_i \rangle + \frac{\|\Gamma\|}{2}\langle g_j, g_j \rangle \\
&\leq c\gamma(\langle f_i, f_i \rangle + \langle f_j, f_j \rangle)
\end{aligned}$$

where all inequalities are degree 2 SoS and $c \leq 2^{O(k)}$. Further, notice that that by applying the same argument to $\langle -f_i, f_j \rangle$ we get that $\langle f_i, f_j \rangle$ is bounded above and below:

$$\vdash_2 \quad -c\gamma(\langle f_i, f_i \rangle + \langle f_j, f_j \rangle) \leq \langle f_i, f_j \rangle \leq c\gamma(\langle f_i, f_i \rangle + \langle f_j, f_j \rangle).$$

We now apply this fact directly to prove the two desired inequalities. First, we have

$$\begin{aligned}
\langle f, f \rangle &= \sum_{\ell=0}^{k} \langle f_\ell, f_\ell \rangle + \sum_{\ell \neq m} \langle f_\ell, f_m \rangle \\
&\geq \sum_{\ell=0}^{k} \langle f_\ell, f_\ell \rangle - \sum_{\ell=0}^{k} kc\gamma \langle f_\ell, f_\ell \rangle \\
&= (1 - kc\gamma) \sum_{\ell=0}^{k} \langle f_\ell, f_\ell \rangle \\
&\geq (1 - c_1\gamma)\langle f_i, f_i \rangle,
\end{aligned}$$

where $c_1 \leq 2^{O(k)}$. For small enough γ, Taylor expanding $(1 - c_1\gamma)^{-1}$ gives the desired result. We can now apply the first inequality to easily prove the second inequality:

$$\langle f, f_i \rangle = \langle f_i, f_i \rangle + \sum_{j \neq i} \langle f_i, f_j \rangle$$

$$\geq (1 - ck\gamma)\langle f_i, f_i\rangle - \sum_{j\neq i} c\gamma\langle f_j, f_j\rangle$$

$$\geq -ck\gamma(1 + c_2\gamma)\langle f, f\rangle \geq -c_3\langle f, f\rangle$$

for $c_2, c_3 \leq 2^{O(k)}$. □

7.14.4 Remaining Proofs from Section 7.10.3

Here we prove Claim 7.10.7, restated for convenience.

Claim 7.14.7 (Restated Claim 7.10.7). Let M be a k-dimensional complete HD-walk on a d-dimensional, two-sided γ-local-spectral expander with $d > k$ and $\gamma \leq w(M)^{-1}2^{-\Omega(h(M)+k)}$ with stationary distribution π. If the total laziness of M $\mathbb{E}[\mathbb{1}_v^T M \mathbb{1}_v]$ is at least .1, then the spectral gap of M is at least $\Omega(1/k)$ and $\lambda_{k/2}(M) \leq .68$.

Proof. To see this, note that any walk can be (approximately) decomposed by Lemma 7.7.3 into an affine combination of pure walks of the form $(U_{k-1}D_k)^i$, that is:

$$M = \sum_{i=0}^{h(M)} \alpha_i (U_{k-1}D_k)^i + \Gamma,$$

where $\|\Gamma\| \leq w(M)2^{O((M)}$ and $(U_{k-1}D_k)^0 = I$. Notice that for any $(U_{k-1}D_k)^i$ for $i > 0$, the probability of returning to any given k-face is at most γ. This follows from the fact that the probability of returning to any particular face in the final up step (given that the previous step is at $\sigma \in X(k-1)$)) is at most $|\Pi_{\sigma,1}|_\infty \leq \gamma$. This last inequality follows from (X, Π) being a two-sided γ-local-spectral expander satisfying $d > k$, as the link (X_σ, Π_σ) (at level k) is then a γ-spectral expander, which are easily shown to satisfy this property.

Since the total laziness of M is at most .1, by our assumption on γ the above analysis implies that the coefficient α_0 cannot be too large, say at most 1/5 (since the remaining parts cannot contribute much to the laziness). We can use this fact to bound the spectral gap of M by the same trick used in Proposition 7.7.11: transferring M over

to the complete complex. Since the approximate eigenvalues of M are independent of the underlying complex and $(UD)^i$ has a vanishing lazy component on $J(n,d)$ as n goes to infinity, it must be the case that in lazy component of M on the Johnson complex is also small. Since M can be written as a convex combination of partial-swap walks, all of which have spectral gap at least $1/k$ (save for the identity which carries at most a $1/5$ of the weight), we get that $1-\lambda_2(M) \geq \Omega(1/k)$ as desired. The bound on $\lambda_{k/2}$ follows similarly noting that all (non-identity) partial-swap walks satisfy $\lambda_{k/2} \leq 1/2$. □

7.14.5 Remaining Proofs from Section 7.10.4

In this section we will prove the claims that were stated without proof in Section 7.10.4. We will use the exact definition of the potential functions from Section 7.14.2 and also the propertites of the polynomials $p_{\beta,\nu}(Y)$ that SoS certifiably ν-approximate the indicator function $\mathbb{1}[Y \geq \beta]$. See Section 7 of BBKSS for a detailed overview about these polynomials and their properties.

Proposition 7.14.8 (Restated Proposition 7.10.10)**.** *Let M be a k-dimensional complete HD-walk on a d-dimensional, two-sided γ-local-spectral expander satisfying $\gamma w(M) 2^{O(h(M)+k)} \leq 1/2^k < \eta < 0.02$ and $d > k$, and I be an affine unique games instance over M with value at least $1-\eta$. Let $r(\eta) = R_{1-16\eta}(M) - 1$, and assume $r \leq k/2$. Let $\beta = 19\eta$ and $\nu = \frac{\eta}{56\binom{k}{r}}$. Then given a degree-$\widetilde{O}(1/\nu)$ pseudodistribution μ satisfying the axioms $\mathcal{A}_I$, there exists an r-link X_τ such that the global potential restricted to X_τ is large:*

$$\Phi^I_{\beta,\nu}(\mu|_\tau) \geq \frac{1}{4\binom{k}{r}}.$$

Proof of Proposition 7.10.10. Before diving into the details of the proof we give a proof overview and some high level intuition. Given two UG assignments X, X', BBKSS define a partition of the graph G into $|\Sigma|$ parts and call this the approximate shift-partition of G with respect to X, X'. The partition is given by the functions $f_s^{X,X'}$ for $s \in \Sigma$, where

$f_s^{X,X'}(u)$ roughly equals the indicator function of whether u belongs to part s. Concretely, we have that $f_s^{X,X'}(u) = \mathbb{1}[X_u - X'_u = s] \cdot p_{\beta,\nu}(\mathrm{val}_u(X))$, that is, u belongs to part s if $X_u - X'_u = s$ and the value of u measured with respect to assignment X is large. Note that any vertex can only belong to one part s but some vertices may not belong to any part. BBKSS prove that for any instance of affine unique games and any two UG assignments X, X' that have large value, the functions $f_s^{X,X'}$ satisfy three properties: a) they include a large number of vertices in total, b) they form a non-expanding partition of G, that is, the fraction of edges that go across parts is small (roughly at most edges violated by X, X'), and finally c) the functions f_s are approximately Boolean-valued on average. These properties are proved in Section 4 of their paper. They then apply the structure theorem for non-expanding sets of G to conclude that there exists a subgraph H where the size of the partition restricted to H ($\sum_s \mathbb{E}_{u\sim H}[f_s^{X,X'}(u)]^2$) must be large. The crucial observation then is that the pseudoexpectation of the restricted size on H is by definition the global potential on H. Therefore by taking pseudoexpectation over $X, X' \sim \mu$ where μ is a pseudodistribution with high value, they conclude that there exists a subgraph with large potential. We will use the same proof strategy to conclude that the global potential on a link is large, using the SoS structure theorem (Theorem 7.10.8) along the way.

Let us elaborate on the properties (proved in BBKSS Section 4) of the approximate shift-partition given by the functions $f_s^{X,X'}$. We will drop the superscript X, X' for convenience and throughout the proof the pseudoexpectation is taken with respect to $X, X' \sim \mu$. Since μ is a degree $\widetilde{O}(1/\nu)$ pseudodistribution with value at least $1 - \eta$ we have that:

1. The approximate shift-partition includes most of the vertices in pseudoexpectation over μ:
$$\widetilde{\mathbb{E}}_{\mu}\left[\sum_{s\in\Sigma} \mathbb{E}_u[f_s]\right] \geq 1 - \frac{\eta}{1-\beta-\nu} - \nu$$

2. The approximate shift-partition is non-expanding[28] in pseudoexpectation over μ:

$$\widetilde{\mathbb{E}}_{\mu}\left[\sum_{s\in\Sigma}\langle f_s,(I-A_G)f_s\rangle\right]\leq 2\eta+2\frac{\eta}{1-\beta-\nu}+2\nu$$

3. The functions f_s are approximately Boolean-valued in pseudoexpectation over μ:

$$\widetilde{\mathbb{E}}_{\mu}[\sum_{s}B(f_s)]=\widetilde{\mathbb{E}}_{\mu}\left[\sum_{s\in\Sigma}\mathbb{E}_{u}[f_s-f_s^2]\right]\leq\frac{\eta}{1-\beta-\nu}+\nu$$

Since the functions f_s are non-expanding (property (2) above), this allows us to use our ℓ_2 characterization of non-expansion, to get a lower bound on the size of f_s restricted to r-links. We will now use the key observation that $\mathbb{E}_{X_\tau}[f_s]^2$ is exactly $D_i^k f_s(\tau)^2$. In particular, applying Theorem 7.10.8 to the function f_s with ℓ set to $r(\eta)=R_{1-16\eta}(M)$, gives:

$$\langle f_s,(I-M)f_s\rangle\geq(1-\lambda_{r+1}(M))\left((1-c_1\gamma)\,\mathbb{E}[f_s]-\binom{k}{r}\langle D_r^k f_s,D_r^k f_s\rangle+(1-c_1\gamma)B(f_s)\right),$$

where $B(f)=\mathbb{E}_{u\sim X_k}[f(u)-f(u)^2]$. Now using the key observation that $D_r^k f_s$ is exactly $\mathbb{E}_{X_\tau}[f_s]$ and noting that $1-\lambda_{r+1}\geq 16\eta$ by definition, we can simplify the above to:

$$\frac{1}{16\eta}\langle f_s,(I-M)f_s\rangle\geq\left((1-c_1\gamma)\,\mathbb{E}[f_s]-\binom{k}{r}\mathop{\mathbb{E}}_{r\sim X(r)}[\mathop{\mathbb{E}}_{X_\tau}[f_s]^2]+(1-c_1\gamma)B(f_s)\right).$$

Taking a sum over s and then a pseudoexpectation over μ yields:

$$\begin{aligned}&\frac{1}{16\eta}\widetilde{\mathbb{E}}_{\mu}[\sum_{s\in\Sigma}\langle f_s,(I-M)f_s\rangle]\\&\geq\widetilde{\mathbb{E}}_{\mu}\left[\sum_{s\in\Sigma}\left((1-c_1\gamma)\mathbb{E}[f_s]-\binom{k}{r}\mathop{\mathbb{E}}_{\tau\in X(r)}\left[\mathop{\mathbb{E}}_{X_\tau}[f_s]^2\right]+(1-c_1\gamma)B(f_s)\right)\right]\end{aligned}$$

[28]If the functions $f_s(u)$ were exactly Boolean-valued and further exactly partitioned the graph, then the term $\sum_s\langle f_s,(I-A_G)f_s\rangle$ is equal to the fraction of edges in G that have endpoints in different parts. Therefore this expression measures the non-expansion of a partition.

$$=(1-c_1\gamma)\,\widetilde{\mathbb{E}}_{\mu}[\sum_{s\in\Sigma}\mathbb{E}[f_s]] - \binom{k}{r}\mathop{\mathbb{E}}_{\tau\in X(r)}\left[\widetilde{\mathbb{E}}_{\mu}\left[\sum_{s\in\Sigma}\mathop{\mathbb{E}}_{X_\tau}[f_s]^2\right]\right]$$
$$+(1-c_1\gamma)\,\widetilde{\mathbb{E}}_{\mu}[\sum_{s\in\Sigma}B(f_s)],$$

where $c_1 \leq 2^{O(k)}$.

We will now use the observation that the global potential restricted to X_τ exactly corresponds to the second term above! Recall that the global potential (Definition 7.14.3) over an r-link X_τ is given by:

$$\Phi^I_{\beta,\nu}(\mu|_\tau) = \widetilde{\mathop{\mathbb{E}}}_{X,X'\sim\mu}[\sum_{s\in\Sigma}\mathop{\mathbb{E}}_{u\sim X_\tau}[f_s^{X,X'}(u)]^2].$$

For convenience of notation we will drop the superscript I in the potential notation henceforth. Re-arranging gives a lower-bound on the global potential averaged across r-links:

$$\mathop{\mathbb{E}}_{\tau\in X(r)}[\Phi_{\beta,\nu}(\mu|_\tau)] \geq \frac{1}{\binom{k}{r}}\Bigg((1-c_1\gamma)\,\widetilde{\mathbb{E}}_{\mu}[\sum_{s}\mathbb{E}[f_s]] + (1-c_1\gamma)\,\widetilde{\mathbb{E}}_{\mu}[\sum_{s}B(f_s)] \tag{7.12}$$
$$-\frac{1}{16\eta}\,\widetilde{\mathbb{E}}_{\mu}[\sum_{s}\langle f_s,(I-M)f_s\rangle]\Bigg). \tag{7.13}$$

Plugging in the properties of the functions f_s discussed at the start of the proof, this gives the following bound on the global potential:

$$\mathop{\mathbb{E}}_{\tau\in X(r)}[\Phi_{\beta,\nu}(\mu|_\tau)] \geq \frac{1}{\binom{k}{r}}\left(1-c_1\gamma-2\frac{\eta}{1-\beta-\nu}-2\nu-\frac{1}{16\eta}\left(2\eta+2\frac{\eta}{1-\beta-\nu}+2\nu\right)\right).$$

Setting $\beta = 19\eta$ and $\nu = \frac{\eta}{56\binom{k}{r}}$ and recalling our assumptions on γ and η, this may be simplified by direct computation to:

$$\mathop{\mathbb{E}}_{\tau\in X(r)}[\Phi_{\beta,\nu}(\mu|_\tau)] \geq \frac{1}{4\binom{k}{r}}.$$

As a result, we see by averaging that there must exist some r-link with high potential:

$$\exists \tau \in X(r) : \Phi_{\beta,\nu}(\mu|_\tau) \geq \frac{1}{4\binom{k}{r}}, \tag{7.14}$$

as desired. □

We will now prove Lemma 7.10.11.

Lemma 7.14.9 (Lemma 7.10.11 restated)**.** *Assume the conditions of Proposition 7.10.10 hold. Let X_τ be an r-link for $r = R_{1-16\eta}(M)-1$ with high global potential: $\Phi^I_{\beta,\nu}(\mu|_\tau) \geq \frac{1}{4\binom{k}{r}}$, for $\beta = 19\eta$ and $\nu = \frac{\eta}{56\binom{k}{r}}$. Then the potential induced on X_τ is also high:*

$$\Phi^{I|_\tau}_{\eta,\nu}(\mu) \geq \frac{1}{8\binom{k}{r}}.$$

As discussed in Section 7.10, proving Lemma 7.10.11 requires that a strong expansion property holds for links: expansion must be similar vertex-by-vertex. We now formalize this property and prove it holds for HD-walks.

Definition 7.14.10. The edge-expansion of a vertex v in a set $S \subseteq X(k)$ with respect to a random walk operator M is given by,

$$\phi_S(M, v) = 1 - \mathbb{1}_v^T M \mathbb{1}_S.$$

The edge-expansion of S with respect to M is the average edge-expansion taken over vertices in S:

$$\phi(M, S) = \mathbb{E}\, v \sim S \phi_S(M, v),$$

where $v \sim S$ is the distribution π_k conditioned on S. As before, we drop M from the notation when clear from context.

We prove that the expansion of vertices in a link cannot vary much from the link's overall expansion (which is small by Theorem 7.9.2).

Lemma 7.14.11 (Restatement of Lemma 7.10.12)**.** *Let M be a k-dimensional HD-walk on d-dimensional two-sided γ-local-spectral expander satisfying $\gamma \leq w(M)^{-1}2^{-\Omega(h(M)+k)}$ and $d > k$. Then for every i-link X_τ, the deviation of the random variable $\phi_{X_\tau}(v)$ ($v \sim X_\tau$) is small:*

$$\mathop{\mathbb{E}}_{v\sim X_\tau}[|\phi_{X_\tau}(v) - \phi(X_\tau)]|] \leq \frac{1}{2^{11k}}.$$

We prove this claim by reducing to simpler HD-walks we can analyze directly. With that in mind, let's first prove that a stronger point-wise result holds for any product of "lower walks" $(U_{k-1}D_k)^i$.

Proposition 7.14.12. *Let M be a k-dimensional HD-walk on d-dimensional two-sided γ-local-spectral expander satisfying $\gamma \leq w(M)^{-1}2^{-\Omega(h(M)+k)}$ and $d > k$. For every i-link X_τ, walk $(U_{k-1}D_k)^t$, and k-face $v \in X_\tau$, the expansion of v in X_τ is almost exactly $\phi(X_\tau)$:*

$$|\phi_{X_\tau}((U_{k-1}D_k)^t, v) - \phi((U_{k-1}D_k)^t, X_\tau)| \leq t\gamma.$$

Proof. It is more convenient to analyze the non-expansion at v, $\overline{\phi}_{X_\tau}((U_{k-1}D_k)^t, v) = 1 - \phi_{X_\tau}((U_{k-1}D_k)^t, v)$, which for brevity we will denote by $\overline{\phi}(v)$. Below we will prove that for every $v \in X_\tau$, $\overline{\phi}(v) \in [c, c + t\gamma]$, for some fixed constant c. This immediately implies the result. Since by definition the non-expansion of X_τ denoted $\overline{\phi}((U_{k-1}D_k)^t, X_\tau)$, is an average over $\overline{\phi}(v)$, we get that:

$$|\phi(v) - \phi((U_{k-1}D_k)^t, X_\tau)| = |\overline{\phi}(v) - \overline{\phi}((U_{k-1}D_k)^t, X_\tau)| \leq t\gamma,$$

as desired.

Therefore let us now prove that $\overline{\phi}(v) \in [c, c + t\gamma]$ for all $v \in X_\tau$. First, notice that

the non-expansion at v is lower bounded by the probability that the walk does not remove any element from τ in any down-step. Since the down step is uniformly random, this probability is the same across all $v \in X_\tau$. Denote this probability by c.

Let us now upper bound $\overline{\phi}(v)$. The probability that the walk returns to X_τ is equal to c plus the probability that starting from v, the walk leaves the link X_τ at some intermediate (down) step but ends back in X_τ regardless. We will show that the latter probability is at most $t\gamma$.

Consider the first down-step where the walk leaves X_τ, removing an element $w \in \tau$. To end up in the link of τ, the walk needs to add w back in one of the future up steps. The probability of this occurring at an up step from any $\sigma \in X(k-1)$ is exactly $\Pi_{\sigma,1}(w)$ by definition (where $\Pi_{\sigma,1}(w) = 0$ if $w \notin X_\sigma$). Since (X, Π) is a two-sided γ-local-spectral expander, (X_σ, Π_σ) is a standard γ-spectral expander. It is a standard result that such graphs cannot have any vertex of weight greater than γ, thus $\Pi_{\sigma,1}(w) \leq \gamma$. By a union bound the probability that w is added back in any of the future up steps is therefore at most $t\gamma$. Hence the probability that the walk returns to X_τ given that it exited in an intermediate step is at most $t\gamma$, so we get that $\overline{\phi}(v) \in [c, c + t\gamma]$. □

It's worth noting that the vertex-by-vertex expansion property actually holds exactly for the canonical walks. However, we can only reduce to canonical walks when the height of M is small, an issue we avoid by analyzing lower walks. We now prove Lemma 7.10.12 by reducing general M to this case.

Proof of Lemma 7.10.12. By repeated application of Lemma 7.7.3, any k-dimensional HD-walk M on a two-sided γ-local-spectral expander with $k > 0$ can be written as a linear combination of lower products $(U_{k-1}D_k)^i$ up to $O(\gamma)$ error in the following sense:

$$M = \sum_{i=0}^{h(M)} c_i (U_{k-1}D_k)^i + \Gamma,$$

where $(U_{k-1}D_k)^0$ denotes the identity matrix and $\|\Gamma\| \leq w(M)2^{O(h(M))}\gamma$. For simplicity of notation, we write $C \coloneqq \|\Gamma\|$, which is in turn $\leq 1/2^{13k}$ for γ chosen to be small enough.

We also know by Lemma 7.14.12 that for all i there exists a constant u_i such that for all vertices $v \in X_\tau$, $\phi_{X_\tau}((U_{k-1}D_k)^i, v) = u_i \pm h(M)\gamma$. Hence we get that,

$$1 - \phi_{X_\tau}(M, v) = \sum_{i=0}^{j} c_i \mathbb{1}_v^T (U_{k-1}D_k)^i \mathbb{1}_{X_\tau} + \mathbb{1}_v^T \Gamma \mathbb{1}_{X_\tau} = \sum_i c_i u_i + \mathbb{1}_v^T \Gamma \mathbb{1}_{X_\tau} \pm w(M)h(M)\gamma.$$

Let $\sum_i c_i u_i = u$. We will show that in expectation over $v \in X_\tau$, $|1 - \phi_{X_\tau}(M, v) - u| \leq |\mathbb{1}_v^T \Gamma \mathbb{1}_{X_\tau}| + w(M)h(M)\gamma$ is small. Define err as the error vector $\Gamma \mathbb{1}_{X_\tau}$. We will bound the 1-norm of err when restricted to coordinates in X_τ. Define the vector s to be $\text{sign}(err(v))$ for $v \in X_\tau$ and 0 otherwise. First note that,

$$\langle s, err \rangle = \sum_{v \in X_\tau} \mathop{\mathbb{P}}_{\pi_k}[v] \cdot |err(v)|.$$

Applying Cauchy-Schwarz on the LHS we get that,

$$\begin{aligned}
\sum_{v \in X_\tau} \Pi_k(v)|err(v)| &\leq \|s\| \cdot \|err\| \\
&= \|\mathbb{1}_{X_\tau}\| \cdot \|\Gamma \mathbb{1}_{X_\tau}\| \\
&\leq \|\mathbb{1}_{X_\tau}\| \cdot C\|\mathbb{1}_{X_\tau}\| \\
&= C\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau} \rangle_{\pi_k}.
\end{aligned}$$

So we get that,

$$\frac{\sum_{v \in X_\tau} \Pi_k(v)|err(v)|}{\langle \mathbb{1}_{X_\tau}, \mathbb{1}_{X_\tau} \rangle_{\pi_k}} = \mathop{\mathbb{E}}_{v \sim X_\tau}[|err(v)|] \leq C.$$

Using the fact that $\mathbb{E}\, v \sim X_\tau \phi_{X_\tau}(M, v) = \phi(M, X_\tau)$, we can conclude with a simple trick:

$$|1 - \phi(M, X_\tau) - u| = |\,\mathbb{E}\, v \sim X_\tau 1 - \phi_{X_\tau}(M, v) - u|$$

$$\leq \mathbb{E}\, v \sim X_\tau |err(v)| + w(M)h(M)\gamma$$
$$\leq 2C.$$

Since the average deviation of $(1 - \phi_{M,\tau}(v))$ from u is small, the average deviation from the mean $(1 - \phi_M(X_\tau))$ should also be small by triangle inequality:

$$\mathbb{E}\, v \sim X_\tau |\phi_{X_\tau}(M, v) - \phi(M, X_\tau)| \leq \mathbb{E}\, v \sim X_\tau |1 - \phi_{X_\tau}(M, v) - u| + C \leq 3C \leq \frac{1}{2^{11k}}.$$

□

We are now ready to prove Lemma 7.10.11/Lemma 7.14.9.

Proof of Lemma 7.10.11. We will use Lemma 7.10.12 to relate the global potential on X_τ to the potential induced on X_τ. As a reminder, we proved that for every i-link X_τ, the deviation of the random variable $\phi_{X_\tau}(M, u)$ $(u \sim X_\tau)$ is small:

$$\mathop{\mathbb{E}}_{u \sim X_\tau} [|\phi_{X_\tau}(M, u) - \phi(M, X_\tau)]|] \leq \frac{1}{2^{11k}}.$$

Let us partition X_τ into two sets, G and $\overline{G}$, where G contains all vertices u for which $\phi_{M,X_\tau}(u)$ is close to $\phi_M(X_\tau)$. We will henceforth drop the subscripts M, X_τ. Formally we define:

$$G = \left\{ u \in X_\tau \mid |\phi(u) - \phi(X_\tau)| \leq \frac{1}{2^k}. \right\}$$

Now note that by Markov's inequality on Lemma 7.10.12 we have that $\mathbb{P}_{u \sim X_\tau}[u \notin G] := c \leq 1/2^{10k}$. For an assignment X, let $Z_{u,s}$ denote $\mathbb{1}[X_u - X'_u = s]$. Expanding out the definition of the global potential on X_τ (Definition 7.14.3), for any two assignments

X, X' we get:

$$\begin{aligned}
\Phi_{\beta,\nu}(X,X')|_\tau &= \sum_s \mathop{\mathbb{E}}_{u\sim X_\tau}[Z_{u,s}p_{\beta,\nu}(\mathrm{val}_u(X))] \\
&= \sum_s ((1-c)\mathop{\mathbb{E}}_{u\sim G}[Z_{u,s}p_{\beta,\nu}(\mathrm{val}_u(X))] + c\mathop{\mathbb{E}}_{u\sim \overline{G}}[Z_{u,s}p_{\beta,\nu}(\mathrm{val}_u(X))])^2 \\
&\leq \sum_s (1-c)^2 \mathop{\mathbb{E}}_{u\sim G}[Z_{u,s}p_{\beta,\nu}(\mathrm{val}_u(X))]^2 + \sum_s c(1-c)\mathop{\mathbb{E}}_{u\sim \overline{G}}[Z_{u,s}] \\
&+ \sum_s c^2 \mathop{\mathbb{E}}_{u\sim \overline{G}}[Z_{u,s}]^2 \\
&\leq \sum_s (1-c)^2 \mathop{\mathbb{E}}_{u\sim G}[Z_{u,s}p_{\beta,\nu}(\mathrm{val}_u(X))]^2 + c(1-c) + c^2,
\end{aligned}$$

where $c \leq 1/2^{10k}$ and we used the fact that $p_{\beta,\nu}(y) \leq 1$, when $y \in [0,1]$, $Z_{u,s} \in [0,1]$ and $\sum_s Z_{u,s} = 1$.

To relate the two potentials we will relate the quantities $\mathrm{val}_u(X)$ and $\mathrm{val}_u^\tau(X)$ for $u \in G$, where $\mathrm{val}_u(X)$ denotes the fraction of edges incident on u that are satisfied by X, and further $\mathrm{val}_u^\tau(X)$ denotes the fraction of edges incident on u that lie inside X_τ and are satisfied by X. We will use the fact that vertices in G have small expansion. When $r = R_{1-16\eta}(M)$, by Theorem 7.9.2 the expansion of the r-link X_τ is at most $1 - \lambda_{r+1} + w(M)h(M)^2 2^{O(k)}\gamma$ which is less than 17η because of the way the parameters have been set up. Therefore for all vertices u in G, $\phi(u) \leq 18\eta$ since $\eta > 1/2^k$. This implies that for any assignment $X \in \Sigma^V$ and any $u \in G$:

$$\mathbb{1}[\mathrm{val}_u^\tau(X) \geq \beta - 18\eta] \geq \mathbb{1}[\mathrm{val}_u(X) \geq \beta],$$

Furthermore by the properties of the polynomials $p_{\beta,\nu}(Y)$, since $\nu < \eta$,

$$p_{\beta-18\eta,\nu}(\mathrm{val}_u^\tau(X)) + \nu \geq p_{\beta,\nu}(\mathrm{val}_u(X)) - \nu.$$

Finally expanding out the definition of the potential induced on X_τ (Definition 7.14.4), for any two assignments X, X' we get:

$$\begin{aligned}
\Phi^\tau_{\beta-18\eta,\nu}(X,X') &= \sum_{s\in\Sigma} \mathop{\mathbb{E}}_{u\sim X_\tau} [Z_{u,s}\cdot p_{\beta-18\eta,\nu}(\mathrm{val}^\tau_u(X))]^2 \\
&= \sum_{s\in\Sigma} \mathop{\mathbb{E}}_{u\sim X_\tau} [Z_{u,s}\cdot (p_{\beta,\nu}(\mathrm{val}_u(X)) - 2\nu)]^2 \\
&\geq \left(\sum_{s\in\Sigma} \mathop{\mathbb{E}}_{u\sim X_\tau} [Z_{u,s}\cdot (p_{\beta,\nu}(\mathrm{val}_u(X)))]^2\right) - 4\nu \\
&\geq \sum_{s\in\Sigma} (1-c)^2 \mathop{\mathbb{E}}_{u\sim G} [Z_{u,s} p_{\beta,\nu}(\mathrm{val}_u(X))]^2 - 4\nu \\
&\geq \Phi_{\beta,\nu}(X,X')|_\tau - c(1-c) - c^2 - 4\nu, \qquad (7.15)
\end{aligned}$$

where we again used the facts that $Z_{u,s}\in[0,1]$, $\sum_s Z_{u,s}=1$ and $p_{\beta,\nu}(y)\in[0,1]$ when $y\in[0,1]$. We also have that $c\leq \frac{1}{2^{10k}}\leq \frac{\eta}{56\binom{k}{r}}$ since $\eta > 1/2^k$ and therefore $c\leq\nu$ implying that: $\Phi^\tau_{\beta-18\eta,\nu}(X,X')\geq \Phi_{\beta,\nu}(X,X')|_\tau - 6\nu$.

Now note that all the inequalities above are sum-of-squares inequalities of degree at most $2\deg(p) = \widetilde{O}(1/\nu)$. We can therefore relate the two potentials when measured with respect to μ, which is a degree-$\widetilde{O}(1/\nu)$ pseudodistribution, by applying the pseudoexpectation operator $\widetilde{\mathbb{E}}_\mu$ to Equation (7.15) above:

$$\Phi^\tau_{\beta-18\eta,\nu}(\mu) = \widetilde{\mathop{\mathbb{E}}_{\mu}}[\Phi^\tau_{\beta-18\eta,\nu}(X,X')] \geq \widetilde{\mathop{\mathbb{E}}_{\mu}}[\Phi_{\beta,\nu}(X,X')|_\tau] - 6\nu = \Phi_{\beta,\nu}(\mu|_\tau) - 6\nu.$$

By the conditions of the lemma, $\Phi_{\beta,\nu}(\mu|_\tau) = 1/4\binom{k}{r}$, $\beta = 19\eta$ and $\nu = \frac{\eta}{56\binom{k}{r}}$ we get $\Phi^\tau_{\eta,\nu}(\mu)\geq \frac{1}{8\binom{k}{r}}$. This completes the proof of the lemma. □

This chapter, in full, is based on the material as it appears in the Symposium on Discrete Algorithms 2022. Bafna, Mitali; Hopkins, Max; Kaufman, Tali; Lovett, Shachar. "High Dimensional Expanders: Eigenstripping, Pseudorandomness, and Unique Games".

The dissertation author was a primary investigator and author of this material.

Chapter 8

Sampling Equilibria: Fast No-Regret Learning in Structured Games

8.1 Introduction

Online learning and equilibrium computation in games has long played a major role in our understanding of human behavior and general multi-agent systems, with applications ranging all the way from politics [55, 270] and national defense [350] to complexity theory [104, 334] and machine learning [155, 156, 26]. Perhaps the most celebrated line of work in this area is the introduction and analysis of randomized weighted-majority (RWM) and its 'mixed' variant (Optimistic) Hedge [285, 155, 326, 103]. These powerful algorithms allow players to engage in repeated gameplay *without regret*, in the sense that the overall loss experienced by any player is not much more than that of the best fixed strategy, even against an arbitrary, adaptive adversary. Such a guarantee is not only powerful in its own right, but is also known to converge quickly to equilibria when performed by all players in repeated rounds of play [88].

Randomized weighted majority is a surprisingly simple algorithm given its powerful guarantees. In each round of a repeated game, a player following RWM samples a strategy s with probability proportional to its (exponentiated) historical loss $\ell(s)$:

$$\mathbb{P}[\text{Player chooses } s] \propto \beta^{\ell(s)} \tag{8.1}$$

for some specified 'learning rate' $\beta \in (0, 1)$. RWM is also well studied in the setting where the player 'plays' the distribution itself (typically called a mixed strategy), and experiences its expected loss. This variant, called *Hedge*, is perhaps the best studied algorithm in all of learning in games [155, 88].

Unfortunately, while RWM and Hedge are *statistically* optimal [285, 155], they come with an inherent computational barrier: both techniques crucially rely on tracking a distribution over all possible actions. Since the number of actions is typically exponential in the relevant parameters of the game (e.g. in the famous Colonel Blotto problem), this seems to render both Hedge and RWM completely infeasible.

It turns out, however, that this intuition is not entirely correct. In many important settings the distributions that arise from RWM are *highly structured*, and while it still may not be possible to efficiently *output* the distribution itself as in Hedge, it is sometimes possible to efficiently *sample* from it. It is known, for instance, that RWM can be implemented in *polylogarithmic* time when actions are given by the k-edges of a complete hypergraph and rewards decompose linearly over vertices [361]. This raises an important question:

When is it possible to efficiently sample in Randomized Weighted Majority?

Toward this end, we introduce a natural generalization of the complete hypergraph setting we call *linear hypergraph games*, where actions are given by k-edges of an *arbitrary* hypergraph, and the reward of any edge similarly decomposes as a sum over individual reward functions on its vertices (see Section 8.2.2 for more detail). This simple definition captures a surprising number of settings studied in the literature including resource allocation problems like Colonel Blotto [72], along with other widely-studied settings such as congestion [332], security [350, 4, 348, 43], and basic dueling games [211, 4].

In this work, we show it is indeed possible to efficiently (approximately) sample from RWM over several important subclasses of linear hypergraph games including Colonel

Blotto and its variants, matroid congestion [356], matroid security [350, 4, 348, 43], and basic dueling games [211, 4]. This leads to the first algorithms for no-regret learning in these settings that are *polylogarithmic* in the size of the state space, and thereby the first polylogarithmic time algorithms for (approximate) equilibrium computation. On top of giving an exponential improvement over prior results, this also constitutes the first efficient algorithm for equilibrium computation whatsoever in several more involved settings such as dice games, Colonel Blotto with multiple resources, and for multiplayer and general-sum variants of all games we consider.

Our techniques are largely based on two main paradigms: dynamic programming, and Monte Carlo Markov Chains (MCMC). Generalizing seminal work on learning k-sets and other structured concepts [361, 267], we show that the distributions arising from RWM on linear hypergraph games correspond to well-studied structure in approximate sampling and statistical physics called *external fields*. In resource allocation games like Colonel Blotto that are played over (ordered) fixed-size partitions of n, we exploit this structure to build a dynamic program that approximately computes the normalization factor of Equation (8.1) (often called the partition function). On the other hand, in settings like matroid congestion and security, we rely on deep results from the MCMC-sampling literature showing that any hypergraph that is a sufficiently good high dimensional expander can be sampled under arbitrary external fields [25, 13, 22]. To the authors' knowledge, these are the first applications of approximate sampling techniques to game theory.

8.1.1 Results

We briefly review the theory of games, equilibria, and no-regret learning before discussing our results in more formality. Games are mathematical objects that model (possibly non-cooperative) interaction between rational agents. A (simultaneous) game consists of a set of actions A_i for each player, and reward functions R_i mapping action tuples to rewards (real numbers). Players seek to maximize their own reward, and optimal play is

typically characterized by Nash equilibria: randomized strategies such that no player can improve by deviating. By the historic result of [306], every finite game has at least one NE. As they are not always efficiently computable [104], one often instead hopes to understand weaker notions such as Coarse Correlated Equilibria (CCE), where the strategies of different players are chosen in coordination with one another (see Section 8.2.1).

There is a deep connection between equilibrium computation and no-regret learning in games. We consider the typical adaptive online setting in which, in each round, a learner chooses an action and receives an adversarially selected loss that may depend on the learner's previous actions (see [88, Chapter 4]). An algorithm is said to have "no-regret" when the expected loss suffered by the learner in sequential rounds grows sublinearly compared to the loss of the best fixed action in hindsight. No-regret learning is itself a powerful tool, as it allows for optimal play against sub-optimal opponents (unlike equilibria which only model the setting where all agents play optimally). Furthermore, it is well known that any no-regret algorithm[1] leads to approximate equilibrium computation with similar runtime simply by simulating the algorithm for all players for sufficiently many rounds. RWM, for instance, is well-known to satisfy the following (optimal) regret guarantee.

Lemma 8.1.1 (RWM is No-Regret [88, Lemma 4.1])**.** *The regret of RWM over T rounds, N actions, and with rewards in $[-L_{\max}, L_{\max}]$, satisfies*

$$\mathrm{Reg}_T \leq O\left(L_{\max}\sqrt{T\log(N)}\right),$$

against any adaptive adversary (with high probability).

In fact, it is important to note in our setting that essentially all guarantees of RWM also hold in the approximate regime, where the learner only δ-approximately samples from

[1]Formally we may need to require that a players strategy depends only on the opponents history and not their own [88]. This is satisfied by all algorithms considered in this work.

the distribution in Equation (8.1) in each round (in Total Variation distance). We call such algorithms δ-approximate RWM (δ-RWM). It is not hard to show that δ-RWM also satisfies the above regret guarantees for small enough δ (see Lemma 8.2.10). We now cover four of the main settings in which we give new algorithms for no-regret learning and equilibrium computation through efficient implementatin of δ-RWM: Colonel Blotto, Matroid Security, Matroid Congestion, and Dueling games. We note that all results are given in the algebraic computation model for simplicity (where algebraic operations such as addition and subtraction are considered to be unit time), but can easily be moved to the standard bit model with no substantial loss in running time (see Section 8.6).

The Colonel Blotto Game

The Colonel Blotto game was originally described by Borel in 1921 [72] and formalizes how warring colonels should distribute soldiers over different battlefields. In the most general version of this game, two colonels have n_1 and n_2 soldiers that they must assign to k different battlefields, each with a non-negative integer weight. A colonel wins a battle (receiving its weight in reward) if they assign more armies to that battle than their opponent. Each colonel seeks to maximize the total weight of battles won in a single assignment.

Despite its breadth of applications and simplicity to state, the first polynomial time algorithm to compute optimal strategies for this game was only developed recently in [4]. This breakthrough result deservedly received significant media attention [216, 296], but struggled to see any practical use due to an infeasible $O(n^{13}k^{14})$ running time (where $n = \max\{n_1, n_2\}$). To this day, this is the only known algorithm to provably compute exact optimal strategies for the (discrete) Colonel Blotto game with arbitrary parameters in polynomial time. Though some progress has been made towards more practical algorithms in different settings [56], even these methods cannot handle parameters beyond a few hundred troops [358].

Indeed, solving the Colonel Blotto problem is now only more relevant than it was in 1921, with practical applications in a large swath of market competitions including advertising and auctions [331], budget allocation [272], elections [275], and even ecological modeling [169]. We give the first no-regret learning algorithm for the Colonel Blotto games under the most general setting [271], where rewards are heterogeneous across battles and players and different players are allowed different troop capacities. Moreover, our algorithm runs in time *polylogarithmic* in the state space, making it extremely efficient in the regime where $n \gg k$ (i.e. there are many more troops than battlefields).

Theorem 8.1.2 (Blotto without Regret (Informal Corollary 8.4.9))**.** *In a Colonel Blotto game, for a player with n soldiers, k battlefields, and maximum reward bounded by $L_{\max}$, δ-RWM can be implemented over T rounds of play in time:*

$$\widetilde{O}\left(T^3 L_{\max} k \log(n) \delta^{-1}\right),$$

and is no-regret. In the regime where $n = O(k^2)$, we give a faster algorithm running in time $\widetilde{O}(Tnk)$.

Theorem 8.1.2 is the first no-regret algorithm for Colonel Blotto in online adaptive settings, and also gives the fastest known algorithms to compute (approximate) Nash equilibria provided the game is zero-sum, and approximate coarse correlated equilibrium in general sum settings with many players. We state the theorem for the two-player zero-sum setting here.

Corollary 8.1.3 (Equilibrium Computation for Blotto (informal Corollary 8.4.10))**.** *Let $n = \max(n_1, n_2)$, where n_1, n_2 are the soldier counts for the two player Colonel Blotto game. Let $L_{\max}$ be maximum reward of the game. There exists an algorithm to compute an ϵ-approximate Nash equilibrium for the two-player Colonel Blotto Game in time*

$$\widetilde{O}\left(L_{\max}^7 k^4 \log^4(n) \epsilon^{-6}\right)$$

with high probability. When $n = O(k^2)$, we give a faster algorithm running in time $\widetilde{O}(nk^2 L_{\max}^2 \epsilon^{-2})$.

Not only is this algorithm exponentially faster than any prior work in most relevant scenarios (namely when $n \gg k$), it is also the first known method for computing CCE for *multiplayer* Blotto at all. Even more generally, our algorithm extends to a number of other variants of Blotto (or 'resource allocation' problems) such as Dice games and settings with multiple types of troops known as the *Multi-resource Colonel Blotto* problem [56] (though in this latter setting we lose the logarithmic dependence on n). We cover these further applications in Section 8.4.3 and Section 8.4.4.

Congestion Games

Another natural example is a *congestion game*, a class introduced by Rosenthal [332] to model resource competition among greedy players. In a congestion game, m players compete to select from a set of n resources and receive rewards depending on how many players chose a particular resource. Classical examples of congestion games include routing traffic (pick the least congested route) and variants of the famed El Farol Bar Problem [32] (players aim to choose a bar that is neither too under nor over-crowded).

Unlike Blotto, equilibrium computation is known to be hard for congestion games, namely (PPAD $\cap$ PLS)-complete [36]. However, this can be circumvented when the underlying strategy spaces are sufficiently combinatorially structured. It has long been known, for instance, that a Nash equilibrium can be found in time $\widetilde{O}(m^2 nqk)$ via iterated best-response when all strategies are given by the bases of a rank-k matroid[2] over n resources of q types [1]. We show matroid congestion games are similarly well-behaved under RWM, and provide a near-optimal no-regret algorithm in both a computational and statistical sense.

[2]Matroid bases can be thought of as a generalization of the combinatorial properties enjoyed by spanning trees, see Section 8.3.1 for details.

Theorem 8.1.4 (Congestion without Regret (informal Corollary 8.3.10))**.** *Let* $\mathcal{I} = \{\{A_i\}_{i=1}^m, c\}$ *be a congestion game over a size-*n *ground set* Ω *with* q *resource types where each* A_i *is the set of bases of a rank-*k *matroid. Then* δ*-RWM can be implemented for* T *rounds in time*

$$\widetilde{O}\left(kT\log(n)\log(\delta^{-1})(q+kmT)\right),$$

and is no-regret.

To our knowledge, this is the first efficient no-regret algorithm for matroid congestion. Moreover, in the setting where there are polylog(n) resource types, the algorithm leads to exponentially faster (approximate) equilibrium computation than the typical best response strategy (albeit for CCE rather than Nash).

Corollary 8.1.5 (Equilibrium Computation for Congestion (informal Corollary 8.3.11))**.** *Let* $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ *be a congestion game over a size-*n *ground set* Ω *with* q *resource types where each* A_i *is the set of bases of a rank-*k *matroid. Then there exists an algorithm to compute an* ε*-CCE in time*

$$\widetilde{O}\left(m^2 L_{\max}^4 k^4 \log^3(n)\epsilon^{-4} + qmL_{\max}^2 k^2 \log^2(n)\epsilon^{-2}\right)$$

with high probability.

Security Games

While slightly less intuitive, games modeling security also fit within the resource allocation paradigm. *Security games* are a basic two-player setting modeling the behavior of a limited-resource player defending n targets, and an adversarial attacker. Each target in the game has a cost to defend, and a "k-resource" defender may choose a (possibly restricted) k-set to defend. Similarly, each target has a cost to attack, and the attacker chooses a single element, receiving a reward depending on whether or not the selected

target was defended by the opponent. Depending on the cost/reward structure, security games model several real-world scenarios, ranging from allocating defensive resources at military checkpoints to choosing a path to transmit critical resources (in the latter the attacker actually *wins* if they attack a 'defended' node). Indeed, security games have actually seen significant use in critical real-life infrastructure such as checkpoint placements at LAX and US Coast Guard and Federal Air Marshal Service patrol schedules [350].

Given their practical importance, it is no surprise equilibrium computation is well-studied in the security game setting [350, 4, 348, 43], and polynomial time algorithms are known in several settings, notably including when allocation constraints are given by matroid bases [348, 43]. Unfortunately, as is the case in previous work on Blotto, known algorithms are not practically useful and have large polynomial factors in the number of targets. We take a major step toward resolving this issue by showing δ-RMW can be implemented in time *polylogarithmic* in n, an exponential improvement over prior techniques [350, 348, 43].

Theorem 8.1.6 (Security without Regret (Informal Corollary 8.3.13))**.** *Let $\mathcal{I}$ a security game over the bases of a rank-k matroid over n targets with q distinct defender costs. Then δ-RMW can be implemented for T rounds in time*

$$\widetilde{O}(kT\log(n)\log(\delta^{-1})(q+T)),$$

and is no-regret.

Corollary 8.1.7 (Equilibrium Computation for Security (Informal Corollary 8.3.14))**.** *Let $\mathcal{I}$ be a security game over the bases of a rank-k matroid over n targets with q distinct attacker and defender costs. Then it is possible to compute an ε-CCE in time*

$$\widetilde{O}\left(L_{\max}^4 k^3\log^3(n)\varepsilon^{-4} + qL_{\max}^2 k^2\log^2(n)\epsilon^{-2}\right).$$

If the game is zero-sum, the resulting strategy is ε-Nash.

Dueling Games

Finally, we make a slight departure from the resource allocation framework to consider the popular class of *dueling games* studied in [211, 4]. Dueling games model competitive optimization between two players over a randomized set of events. We will focus our attention on one of the simplest dueling games called *ranking duel* (also known as the 'Search Engine game') where two players compete to rank n elements over a known distribution μ, and win a round if they rank $x \sim \mu$ higher than the opponent. This classically models the problem of search engines competing to optimize a page ranking given a known distribution over searches.

Equilibrium computation is well-studied in dueling games, and algorithms are known in a few settings via a mix of bilinear embedding techniques and reduction to non-competitive optimization [211, 4]. As in previous settings, however, the algorithms are too slow to be of practical use. In contrast, we focus our attention only on the basic ranking duel, but give both a faster algorithm and a novel no-regret guarantee over the original space.

Theorem 8.1.8 (Dueling without Regret (Informal Corollary 8.3.19))**.** *Let $\mathcal{I}$ be an instance of ranking duel. Then δ-RMW can be implemented for T rounds in time*

$$\widetilde{O}(T^2 n^7 \log(\delta^{-1})),$$

and is no-regret.

Corollary 8.1.9 (Equilibrium Computation for Ranking Duel (Informal Corollary 8.3.20))**.** *Let $\mathcal{I}$ be an instance of ranking duel. Then it is possible to compute an ε-CCE in time*

$$\widetilde{O}(n^9 \varepsilon^{-4}).$$

If the game is zero-sum, the resulting strategy is ε-Nash.

While a running time of $O(n^9)$ can hardly be claimed as practical, the broader technique used in this result has the eventual possibility of running in near-linear time. We discuss this further in Section 8.1.3.

8.1.2 Techniques

At their core, our results all stem from the ability to approximately sample distributions arising from randomized weighted majority on various linear hypergraph games. Recall that RWM maintains a mixed strategy, which we denote as the RWM distribution, whose probabilities are proportional to their (exponentiated) total historical loss (negative reward):

$$\forall x \in A_i : \mathbb{P}(x) \propto \beta^{\ell^T(x)} \tag{8.2}$$

where $\ell^T(x)$ is the total loss experienced by pure strategy x up to round T. As discussed earlier in the section, we typically cannot hope to maintain this distribution explicitly, but it may still be possible to sample from it in polylogarithmic time. Furthermore, while sampling such a distribution exactly is a challenging task (and very few such algorithms are known), *approximate* sampling is perfectly sufficient in our setting. Indeed, our approximate variant δ-RWM satisfies essentially the same guarantees as RWM itself.

Lemma 8.1.10 (δ-RWM is No-Regret (Lemma 8.2.10))**.** *δ-RWM over N actions has*

$$\mathrm{Reg}(T) \leq O\left(L_{\max}\sqrt{T \log N} + \delta L_{\max} T\right)$$

expected regret, where $L_{\max}$ is the maximum loss experienced by any action.

As discussed at the start of Section 8.1, no-regret algorithms like RWM are classically used to compute equilibria of the base game by simulating repeated play across all players. While much of the current literature centers around the Hedge algorithm that 'plays' an

entire mixed strategy in each round, classical (and therefore approximate) RWM still leads to equilibrium computation with high probability, by the classic result that no-regret implies equilibrium computation [156].

Lemma 8.1.11 (Approximate RWM $\to$ Equilibria (Informal Corollary 8.2.11))**.** *Let $\mathcal{I}$ be an m-player game where each player has at most N strategies. Let $\{(x_1^{(t)}, \dots, x_m^{(t)})\}_{t=1}^T$ be the strategies arising from T rounds of δ-approximate RWM. There exist universal constants $C > 0$ such that for $T = C \cdot L_{\max}^2 \varepsilon^{-2} \cdot \log(N)$ rounds, and approximation parameter $\delta \leq \epsilon/(CL_{\max})$, these strategies constitute an ε-CCE with high probability (Nash if the game is two-player zero-sum).*

As a result, efficient no-regret learning and equilibrium computation truly reduces to the existence of an efficient approximate sampling scheme for distributions arising in the execution of (δ-approximate) RWM. Of course, this is easier said than done. While approximate sampling is easier than its exact variant, it is still a challenging problem, even over structured domains. Using a mixture of novel sampling techniques and reductions to known methods in the literature, we show it is indeed possible to efficiently sample from RWM across a wide variety of structured games. Our strategies fall into two main paradigms: dynamic programming (DP), and Monte Carlo Markov Chains (MCMC).

Sampling via Dynamic Programming

We start with the former: sampling in basic resource allocation settings via dynamic programming. At its most general, resource allocation problems are played over (possibly constrained) fixed size partitions of n. The discrete Colonel Blotto game on n troops and k battlefields is the simplest example of this problem, where the strategy space corresponds to the set of all k-size (ordered) partitions of n (i.e. assignments $x_1, \dots, x_k$ such that $\sum x_i = n$). In this section, we will focus only on the Colonel Blotto problem—general resource allocation follows from very similar arguments (see Section 8.4 for details).

Our goal is now to design an algorithm for approximately sampling distributions over strategies of the Colonel Blotto game that arise from RWM. In this setting, it will actually be easier to solve an equivalent problem, computing the normalizing factor of Equation (8.2), otherwise known as the *partition function*:

$$f_k(n) = \sum_{x_1+\ldots+x_k=n} \beta^{\ell(x)} = \sum_{x_1+\ldots+x_k=n} \prod_{h=1}^{k} \beta^{\ell_h(x_h)},$$

where $\ell_h(x_h)$ is the historical losses from the h-th battlefield over previous rounds of play if one were to place x_h soldiers on that battlefield. Notice that once we know the value of $f_{k'}(n')$ for all $k' \leq k$ and $n' \leq n$, it is actually possible to exactly sample from Equation (8.2) (and therefore implement RWM). In particular, one does this simply by sequentially sampling the number of troops to put in each battlefield conditional on prior choices in the following manner:

$$\mathbb{P}[x_1 = y] \propto \beta^{\ell_1(y)} \cdot f_{k-1}(n-y) \text{ for the first battlefield,}$$
$$\mathbb{P}\left[x_{h+1} = y | x_{1\cdots h}\right] \propto \beta^{\ell_{h+1}(y)} \cdot f_{k-h-1}\left(n - \left(\sum_{j=1}^{h} x_j\right) - y\right) \text{ for the remaining battlefields.}$$

One can easily check the joint distribution arising from this procedure is exactly the RWM distribution.

Thus we have reduced our problem to computing the partition functions $f_{k'}(n')$. This can be done by a simple dynamic programming argument, and in particular by noticing that:

$$f_{k'}(n') = \sum_{i=0}^{n'} \beta^{\ell_{k'}(i)} \cdot f_{k'-1}\left(n'-i\right). \tag{8.3}$$

Since filling each entry $f_{k'}(n')$ takes time at most $O(n)$ given that f_{h-1} is pre-computed,

we can fill the entire DP table in time $O(n^2k)$.[3]

While this procedure already gives the first no-regret learning algorithm for Blotto in the adversarial setting and by far the fastest known equilibrium computation, one can still hope to do much better. Indeed, it is known that there exist ε-Nash Equilibrium with support that is *logarithmic* in the size of the state-space [284], so there is hope in building a *polylogarithmic* time algorithm (equivalently, a polynomial time algorithm in the description complexity of the equilibria). We show this is indeed possible by building an approximation scheme for the above DP. The key is to observe that the partition functions $f_{k'}(n')$ are bounded and monotonic. Roughly speaking, this means f can be approximated within multiplicative $(1 \pm \varepsilon)$ factors by a piece-wise function with only $\text{poly}(k \log(n)/\varepsilon)$ pieces (which is polylogarithmic in the size of the state space).

By carefully computing and maintaining approximate versions of the partition function, we can run a modified variant of the same dynamic program that computes approximations for all nk partition functions $f_{k'}(n')$ (despite their sizes, these can indeed be presented in only $\text{poly}(k \log(n))$ bits due to being piece-wise). Once we have approximately computed the partition functions, it is easy to show that a similar sampling scheme as discussed for the exact case gives an efficient approximate sampling scheme for RWM running in $\text{poly}(k \log(n)/\varepsilon)$ time. Combined with Lemma 8.1.10 and Lemma 8.1.11, this results in the first polylogarithmic time algorithm for no-regret learning and (approximate) equilibrium computation for Colonel Blotto (Theorem 8.1.2 and Corollary 8.1.3), as well as for a number of related resource allocation variants discussed later in the paper (e.g. multi-resource Blotto and Dice games).

MCMC-methods

While dynamic programming is a powerful algorithmic method for structured computation, there are many combinatorial settings common to games we cannot hope to

[3]We note that this can actually be improved to near-linear in n using the Fast Fourier Transform.

handle via such techniques. Building an analogous exact-counting based DP for games over bipartite matchings or matroids, for instance, would give efficient algorithms for classical #P-hard problems such as the permanent and counting matroid bases [101]. On the other hand, we do actually know of approximation algorithms for these problems based on a powerful tool called MCMC-sampling [223, 25].

MCMC-sampling is an elegant method for approximately sampling from a distribution μ with exponential size support usually traced back to Ulam and Von Neumann in the 1940s (see e.g. [135]). The idea is simple. Imagine we can construct a Markov chain (random process) M satisfying the following three conditions:

1. The stationary distribution of M is μ
2. A single step of M can be implemented efficiently
3. M converges quickly to its stationary distribution.

Approximate sampling would then simply boil down to finding a starting configuration and running the chain until it is within δ of stationary (this typically takes around $O(\log(N/\delta))$ samples for a good chain).

Unsurprisingly, while MCMC-sampling itself is a simple technique, the design and analysis of Markov chains is a difficult task, and general recipes for their construction are known in very few scenarios. One particularly well-studied setting in the literature that arises from simulation problems in statistical physics are *external fields*. Given a hypergraph $\Omega \subset \binom{[n]}{k}$, the distribution arising from external field $w \in \mathbb{R}_+^n$ simply assigns each k-set a probability proportional to the product of its fields:

$$\Omega^w(s) \ \propto \prod_{v \in s} w(v).$$

External fields often correspond to particularly natural problems, and are well-studied in the literature. In a recent breakthrough series of works, for instance, it was shown that

approximate sampling under external fields is possible whenever the underlying state-space is a good enough high dimensional expander [239, 11, 25, 22].[4]

This is particularly relevant to our setting since it is a simple observation that the distributions arising from RWM on a linear hypergraph game are exactly given by the application of an external field over the action space.

Observation 8.1.12 (RWM $\rightarrow$ External Fields (Informal Observation 8.3.3)). Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player linear hypergraph game. Then for any A_i and any round of play, Player i's RWM distribution can be written as the application of an external field w to A_i.

As a result, no-regret learning and equilibrium computation are possible in any linear hypergraph game whose state space can be sampled under arbitrary external fields. As a short detour, it is worth noting that the result in the previous section can be phrased as a slight refinement of this statement. At a technical level, our resource allocation algorithm simply corresponds to an efficient approximate sampling scheme for fixed-size partitions of n under *monotonic* external fields (corresponding to the fact that assigning more troops to a battlefield always results in at least as many victories).

Many well-studied games in the literature have state spaces where efficient approximate sampling schemes under external fields exist. In this work we focus mostly on games played on matroids (e.g. matroid congestion, security), and dueling games arising from bipartite matchings such as ranking duel. Both settings have well-known sampling schemes over external fields [25, 102, 223], which leads to our results for Congestion, Security, and Dueling games (Theorems 8.1.4,8.1.6, 8.1.8 and Corollaries 8.1.5,8.1.7, 8.1.9 respectively).

8.1.3 Discussion

In this work, we present two potential paradigms for learning in games via approximate sampling. In this section we touch on the pros and cons of each method, their

[4]More formally, when the space satisfies a property known as 'fractional log-concavity.'

likelihood to generalize beyond the settings considered in this work, and natural open problems.

Dynamic Programming vs MCMC-sampling

Broadly speaking, the DP and MCMC approaches we develop in this work seem to be largely incomparable. Dynamic programming works well in relatively unconstrained resource allocation problems, where recursive structure allows for inductive computation of the partition function. On the other hand, typical MCMC methods (which are usually *local* in nature) actually fail drastically in this sort of setting due to the need for global coordination. One natural example of this issue appears in the Colonel Blotto game. Imagine a scenario where Colonel A has k more troops than Colonel B, then there always exists a configuration where A wins every battle by assigning one more troop in each battlefield than B. Finding this sort of optimum, however, requires coordinated planning across battlefields. Typical MCMC methods like Glauber dynamics (see Section 8.3) only look at a few battlefields at a time, and therefore struggle to converge to such solutions. Simulations confirm this intuition—even for small n and k Glauber dynamics seem to exhibit very poor mixing on distributions arising in RWM.

On the other hand, as we mentioned in the previous section, our dynamic programming approach has a significant issue in any setting with non-trivial combinatorial structure. In particular, because the underlying method relies on exactly computing the partition function, constructing any such method for a problem like matroid games is #P-hard. On the other hand, local chains such as the Glauber Dynamics mix extremely fast in these settings, providing near-optimal algorithms.

Of course, neither of these arguments rules out either approach. It is possible there exist successful MCMC methods for Blotto that are more *global* in nature—indeed the main insight leading to the resolution of approximate permanent was exactly such a Markov chain that avoided these issues [223]. On the other hand, there may exist DP-based

approaches that do not go through computing the partition function. Understanding in which scenarios these two or other potential sampling methods may apply remains an interesting and important open problem if we wish to extend efficient learnability in games beyond the few structured settings considered in this work.

Further Open Problems

Our work gives the first no-regret learning guarantees and polylogarithmic equilibrium computation for several well-studied settings in game theory, but there is still much to be done. Perhaps the most obvious open directions involve improving the computational efficiency (and therefore practicality) of our algorithms. The polynomial dependencies of our algorithm would be universally improved if we can show that the δ-approximate *optimistic* variant of RWM achieves $\widetilde{O}(1)$ regret in games like its deterministic counter-part, Optimistic Hedge [103], even for polynomial approximation δ.

Question 8.1.13 (Optimistic-RWM)**.** Does δ-RWM with weights $\{w_i^{(t)}\}_{i,t}$ and optimistic updates
$w_i^{(t+1)} \leftarrow w_i^{(t)} \cdot \beta^{2\ell_i^{(t)} - \ell_i^{(t-1)}}$ achieve polylog(T) regret in games (even for $\delta = (\text{polylog}\, N)^{-1}$)?

These techniques are well known to give a substantial improvement in the exact setting [103], but their analysis is subtle and may be nontrivial to adapt to the δ-approximate sampling variant we need for efficient computation.

Similarly, our algorithm for dueling games (while faster than prior work in the worst-case), is not practical at $O(n^9)$ running time. One interesting question is whether the MCMC-sampling technique can be improved in this setting using the fact that the weights arising from RWM are not arbitrary, but exhibit *monotonic* structure (in the sense that ranking a page higher is always better). This actually corresponds to a well-studied problem in the sampling and geometry of polynomials literature (monotone permanent [79]), but giving an improved sampling algorithm over the JSV-chain [223] remains an interesting open problem.

Question 8.1.14 (Sampling with Monotone Weights)**.** Can a perfect matching in a complete bipartite graph with n nodes under monotone external fields be sampled in faster than $O(n^7)$ time? In near-linear time?

There is certainly hope in this direction, as recent years have seen many breakthroughs towards near-optimal MCMC methods, including similar linear time guarantees on problems that once seemed infeasible [25, 102, 96].

Another natural direction is to try to strengthen the type of equilibria we compute in multiplayer and general-sum games. Foremost in this direction are the so-called Correlated Equilibria (CE), a substantially stronger notion than CCE which allows a player to switch strategies even after they receive instructions from the coordinator. It was recently shown that a variant of Optimistic Hedge converges quickly to CE in multiplayer, general-sum games [21]. It is an open question whether an approximate, sampled variant could do the same. We pose this as the following,

Question 8.1.15 (Correlated Equilibria with RWM)**.** Is there a variant of δ-approximate RWM that converges to CE and remains efficiently samplable? A variant that achieves $\widetilde{O}(T^{-1})$ convergence rate?

Finally, we end with a concrete direction toward answering our original question: *when can one efficiently sample from RWM?* We rely in part of this work on a series of breakthroughs in the approximate sampling and high dimensional expansion literatures [239, 11, 25, 22] leading to a sufficient condition called fractional log-concavity [13, 22] for sampling hypergraphs under arbitrary external fields (generalizing an earlier result for matroids [25]). This is in fact a stronger guarantee than we actually need to ensure efficient sampling for RWM. Not only are the fields we study typically additionally structured (e.g. monotonic), but we are also okay with some amount of decay in the mixing time depending (logarithmically) on the field size. Is there a characterization of such objects in terms of geometry of polynomials or high dimensional expansion?

Question 8.1.16. Is there a general condition on hypergraphs (e.g. in terms of high dimensional expansion, geometry of polynomials) that allows for approximate sampling under external fields with polylogarithmic dependence on the worst field size? What about under structural constraints (e.g. monotonicity)?

8.1.4 Further Related Work

No-regret learning with structured loss

Online learning over exponentially large classes with structured losses has been considered previously in other contexts (e.g. [263, 100, 196, 267, 34, 359]). Much of this work considers the combinatorial bandit setting [89], which typically competes against a non-adaptive adversary, but has restricted information. This work introduces the notion of sampling from the MWU distribution in structured games by constructing a linear embedding and performing MWU explicitly on each dimension (their ComBand algorithm). We note two distinctions with our current work. First, since the ComBand algorithm focuses on the partial information setting, the regret guarantee of ComBand has worse dependence on dimension than MWU. Furthermore, the natural embedding of a linear hypergraph game into binary vectors (which maps a k-set to its corresponding weight k vector in $\{0,1\}^{\Omega}$) typically has dimension linear in n, leading to exponentially worse runtime compared to our techniques in all of our settings except ranking duel.

There are two additional works which also consider efficient implementation of RWM [361, 196], but only for the very special settings of k-sets and permutations (which are generalized by our framework). Also of note is the later work of [267], who built a new hedge-based algorithm for these settings called component-hedge that also gives efficient online learning in a few additional cases (e.g. for spanning trees).

One may also consider a kernelized approach to sampling in RWM [144]. We note that the kernel approach (which in particular implies the ability to compute the partition function) cannot be applied efficient in many settings (e.g. matroids, where this problem is

known to be #P-hard). Indeed, it is worth noting that the ability to exactly compute the partition function implies the ability to exactly sample efficiently (at least in self-reducible settings), so in this sense requiring efficient kernel computation is strictly stronger than our approach. Further, since the standard embedding of our setting into a 0/1-polytope has dimension that is linear in n (or more accurately the number of vertices in the hypergraph), a naive application of a kernelized method leads to bounds that are linear rather than polylogarithmic in n.

Computing equilibra for Colonel Blotto

The Colonel Blotto game is one of the most well studied problems in algorithmic game theory—we restrict our attention here to some of the most notable and relevant results. As mentioned previously, the first known algorithm to compute exact Nash equilibria strategies for discrete Colonel Blotto was introduced in [4], who consider games that are asymmetric across battles and across players (allowing different troop capacity and rewards across battles and players). This work remains the only known algorithm for exact equilibrium computation that is polynomial in the number of troops and battlefields, though follow-up work gave a more practical (but potentially exponential time) algorithm [56].

Due to the difficulty of understanding the discrete version, a number of works have also considered Colonel Blotto's continuous relaxation. It should be noted that the equilibria in the continuous version do not apply to the discrete case [319]. The works by [184, 185, 274, 352, 271] consider the case that troop counts are identical (symmetric) for both players. Later, the symmetric case was also studied when Colonels have different values for battles [271]. On the other hand, when the troop counts of the two players differ, constructing/computing equilibria becomes more complicated. In [331], the author constructs equilibrium strategies explicitly in the case that the rewards for each battlefield are the same (homogeneous). The authors of [292] consider Blotto with heterogeneous

rewards and asymmetric troops counts (but with only two battles). In [342], they consider more than two battles but with strict assumptions on the battle weights. More recently, the authors in [319] present an efficient algorithm to compute approximate Nash equilibria in the two-player continuous Colonel Blotto game with asymmetric troop and battle values.

There is also a breadth of work that constructs strategies for approximate and exact equilibrium under constrained parameter settings of Borel's two-player discrete version. Beginning with [191], the author constructs optimal strategies explicitly when the troop counts and battle rewards are identical. In [358], the authors give an algorithm to compute equilibria with fixed approximation (decaying with the number of battles). They also give an algorithm to compute the best-response strategy to a given distribution over soldiers in each battlefield using dynamic programming. In [70, 352], the authors describe equilibria in the symmetric case where the number of soldiers is the same for both players. Moreover, [70] introduces the multiplayer variant. In [357], the authors construct equilbria under particular conditions for an extension of the Colonel Blotto game that accounts for pre-allocations and resource effectiveness.

8.2 Preliminaries

All throughout the paper, for integers $a \leq b$ we denote by $[a, b]$ the set $\{a, \ldots, b\}$ and shorthand $[n] = [1, n]$. We use the notations $\widetilde{O}(f)$ to hide polylogarithmic dependencies on the argument. Given a finite set Ω, we denote by $\Delta(\Omega)$ the (convex) polytope of all distributions defined on Ω. We denote by 2^{Ω} as the power set of Ω, i.e. the set of all subsets of Ω. Given two finite sets Ω_1, Ω_2, we denote by $\Omega_1 \times \Omega_2$ as the Cartesian product of the two sets, i.e. $(x, y) \in \Omega_1 \times \Omega_2$ if $x \in \Omega_1$ and $y \in \Omega_2$. We will use $\binom{\Omega}{k}$ to denote all size-k subsets of the ground set Ω. Given two integers $k, n \in \mathbb{Z}^+$, we will use $P_k(n)$ to represent the set of ordered size-k partition of n.

8.2.1 Game Theory

Definition 8.2.1 (Multiplayer Simultaneous Game)**.** An m-player **Simultaneous Game** is a tuple $\{\{A_i\}_{i=1}^m, \{R_i\}_{i=1}^m\}$ where A_i denotes the finite set of actions available for the i-th player and $R_i : A_1 \times \cdots \times A_m \mapsto \mathbb{R}$ denotes the reward function for the i-th player.

Given a set of actions $a_1, \ldots, a_m$, we often write a_{-i} to represent the combined action tuples without a_i, i.e. $(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_m)$, and $R_i(a_i, a_{-i}) = R_i(a_1, \ldots, a_m)$ where we have abused notation in the input ordering to R_i for simplicity of notation.

In a game, a player can choose to play an action, often called a pure strategy, or to draw randomly from a *mixed strategy* given by a probability distribution over the set of available actions.

Definition 8.2.2 (Mixed Strategy)**.** Let $\{\{A_i\}_{i=1}^m, \{R_i\}_{i=1}^m\}$ be an m-player simultaneous game. For the i-th player, the set of mixed strategies are all possible probability distributions over the actions A_i. Let $\mathbf{s}_i \in \Delta(A_i)$ be the mixed strategy chosen by the i-th player. Then, the expected reward received by the i-th player is given by $\mathbb{E}_{a_1 \sim \mathbf{s}_1, \ldots, a_m \sim \mathbf{s}_m}[R_i(a_1, \ldots, a_i, \ldots, a_m)]$.

We will also make use of the following notion of a joint strategy.

Definition 8.2.3 (Joint Strategy)**.** A joint strategy is a distribution $\sigma \in \Delta(A_1 \times \cdots \times A_m)$. If players were to participate in a joint strategy, then a central coordinator samples an action tuple $a = (a_1, \ldots, a_m) \sim \sigma$, and each player then plays the action a_i correspondingly. As a result, the expected reward of the i-th player is given by $\mathbb{E}_{a \sim \sigma} R_i(a)$.

For a set of actions $a_i^{(t)}$ for $i \in [m]$, and $t \in [T]$, we will often write $\frac{1}{T}\sum_{t=1}^T a_i^{(t)}$ as the mixed strategy of player i such that action $a_i^{(t)}$ is played with probability $1/T$, and $\frac{1}{T}\sum_{t=1}^T a_1^{(t)} \otimes \cdots \otimes a_m^{(t)}$ as the joint mixed strategy such that the action tuple $(a_1^{(t)}, \ldots, a_m^{(t)})$ is played with probability $1/T$.

It is well known that if all players play a game optimally, the resulting strategy tuples compose of a *Nash Equilibrium* of the game.

Definition 8.2.4 (Nash Equilibrium)**.** In an m-player game $\{\{A_i\}_{i=1}^m, \{R_i\}_{i=1}^m\}$, a tuple of mixed strategies $(s_1, \ldots, s_m)$ composes an ϵ-Nash Equilibria (ε-NE) if for all $i \in [m]$ it satisfies:

$$\mathop{\mathbb{E}}_{s_1,\ldots,s_m}[R_i(s_1, \ldots, s_i, \ldots, s_m)] \geq \sup_{s' \in \Delta(A_i)} \mathop{\mathbb{E}}_{s_1,\ldots,s_m}[R_i(s_1, \ldots, s', \ldots, s_m)] - \epsilon,$$

where the mixed strategies s_i for $i \in [m]$ are mutually independent.

In multiplayer and general-sum games, computating Nash equilibria is challenging. In fact, this problem is known to be complete for PPAD [104], a complexity class containing many other computationally hard problems. A standard and arguably more realistic goal is to find the so-called Coarse Correlated Equilibria (CCEs) of the multi-player game, a relaxation of Nash Equilibrium introduced by Aumann [35].

In a CCE, all players together sample from a joint mixed strategy (in contrast to NE where players *independently* sample from their own mixed strategy). Although a player i cannot benefit from switching to any single action s_i' before the joint strategy is sampled, once a strategy s_i is sampled from a CCE distribution (becoming known to each player), a player may improve their outcome by deviating (using the fact that her strategy is correlated with other players'). Thus, CCE apply to situations where a player must commit to their strategy up front and are unable to deviate after sampling.

Definition 8.2.5. Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player game. An **ϵ-approximate coarse correlated equilibrium** (ϵ-CCE) is a joint mixed strategy $\sigma \in \Delta(A_1 \times \cdots \times A_m)$ that satisfies:

$$\forall i \in [m], \text{ and actions } a_i' \in A_i : \quad \mathbb{E}_{a \sim \sigma} R_i(a) \geq \mathbb{E}_{a \sim \sigma} R_i(a_i', a_{-i}) - \epsilon.$$

8.2.2 Linear Hypergraph Game

Given a ground set of vertices Ω, a hypergraph H is a collection of subsets of Ω called hyperedges. If all hyperedges of H are in $\binom{\Omega}{k}$, the graph is called k-uniform. In this work we study a special class of games whose reward functions can be 'decomposed' based on the underlying structure of the game's action space. More formally, we consider games played over k-uniform hypergraphs whose rewards *decompose linearly* over vertices. We denote this class of games as Linear Hypergraph Games.

Definition 8.2.6 (Linear Hypergraph Games)**.** Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player game. We call $\mathcal{I}$ a **linear hypergraph game** if for all $i \in [m]$ there is a groundset Ω_i and parameter $k_i \in \mathbb{N}$ such that $A_i \subset \binom{\Omega_i}{k_i}$ and a 'vertex-wise' reward function $R_i^{\Omega_i} : \Omega_i \times A_{-i}$ such that for all $t_1, \ldots, t_n \in A_1 \times \ldots \times A_n$

$$R_i(t_i, t_{-i}) = \sum_{v \in t_i} R_i^{\Omega}(v, t_{-i}).$$

In other words, each element v in the ground set Ω_i has a certain reward with respect to any choice of the opponents, and the reward of a k-set is simply the sum of its individual rewards. Many important games that are well-studied in the game theory literature falls under this category, e.g. Colonel Blotto Games, Security Games, Congestion Games, Dueling Games, etc. In fact, it should be noted a similar notion has been studied in the online learning setting in [267], who develop an efficient no-regret algorithm called Component Hedge for linear losses over basic structures such as the complete hypergraph, truncated permutations, and spanning trees.

8.2.3 No-Regret Learning in Games

We consider the framework of No-Regret Learning in Games (see [88, 103] and references therein). In this framework, a game is iterated with one or more players implementing a no-regret learning algorithm to adaptively choose strategies. At the t-th

round of the game, each player selects a mixed strategy $\mathbf{s}_i^{(t)}$, and samples the action $a_i^{(t)} \sim \mathbf{s}_i^{(t)}$, where the choice of $\mathbf{s}_i^{(t)}$ depends only on $a_j^{(t')}$ for $j \in [m]$ and $t' < t$.

The goal for each player is to optimize her regret, defined as the following.

Definition 8.2.7 (Regret)**.** At the T-th round of the game, the regret for the i-th player is defined as

$$\mathrm{Reg}_{T,i} := \max_{a^* \in A_i} \sum_{t=1}^{T} R_i(a_1^{(t)}, \cdots, a_{i-1}^{(t)}, a^*, a_{i+1}^{(t)}, \cdots, a_m^{(t)}) - \sum_{t=1}^{T} R_i(a_1^{(t)}, \cdots, a_m^{(t)}).$$

It is classical result that if all players follow no-regret learning strategies, the overall dynamics quickly converge to a Nash or Coarse-Correlated Equilibria (CCEs) of the game (see e.g. [156, 88, 310]).

Theorem 8.2.8 (No-Regret Implies Equilibrium Computation [156])**.** *Suppose m players are playing under the No-Regret Learning in Games framework. Let $\sigma^* := \frac{1}{T}\sum_{t=1}^{T} a_1^{(t)} \otimes \cdots \otimes a_m^{(t)}$ be the average mixed joint strategies played by the i-th player over T rounds. Then, σ^* forms an $T^{-1}\max\left(\mathrm{Reg}_{T,1}, \cdots, \mathrm{Reg}_{T,m}\right)$-approximate CCE of the game, where $R_i^{(T)}$ is the regret for the i-th player at the T-th round. When $m = 2$ and the game is zero-sum, the mixed strategies $\left(\frac{1}{T}\sum_{t=1}^{T} a_1^{(t)}, \frac{1}{T}\sum_{t=1}^{T} a_2^{(t)}\right)$ constitute a $T^{-1}\max\left(\mathrm{Reg}_{T,1}, \mathrm{Reg}_{T,2}\right)$-approximate Nash Equilibrium.*

8.2.4 Randomized Weighted Majority Algorithm

As in online learning, no-regret learning in games studies the regret of a player against the opponents' strategies in repeated play with respect to the best single strategy in hindsight.

One of the most frequently used tools in no-regret learning is the randomized weighted majority (RWM) algorithm. For player $i \in [m]$, RWM maintains the mixed strategies from $\Delta(A_i)$ as follows: at the first round, it chooses uniformly among the actions.

At the $(T+1)$-st round, a cumulative reward is computed for each action $x \in A_i$

$$r^{(T+1)}(x) = \sum_{t=1}^{T} R_i(s_1^{(t)}, \cdots, s_{i-1}^{(t)}, x, s_{i+1}^{(t)}, \cdots, s_m^{(t)}),$$

and RWM chooses the mixed strategy $RM^{(T+1)}(\beta)$ (which we refer to as the RWM distribution) such that

$$\mathbb{P}\left[RM^{(T+1)}(\beta) = x\right] \propto \beta^{-r^{(T+1)}(x)}. \tag{8.4}$$

It is well known that if any player samples according to $RM^{(T+1)}(\beta)$ in each round, her expected regret will be bounded in the worst case by $O_{\beta,N}(\sqrt{T})$.

In games whose action sets are exponentially large, exactly sampling from the the RWM distribution may be intractable in relevant cases. Nonetheless, we show that similar regret bounds hold even when one *approximately* samples the RWM distributions in each round (the proof is given in Appendix).

Definition 8.2.9 (Approximate Sampling)**.** We say a randomized algorithm $\mathcal{A}$ with output space Ω δ-approximately samples a distribution μ over Ω if the output of $\mathcal{A}$ is δ-close to μ in TV-distance.

We call any strategy that δ-approximately samples from $RM^{(t)}(\beta)$ in each of T rounds of repeated play *δ-approximate RWM*, and denote this class of algorithms by δ-RWM$_\beta^T$. It is not hard to show that δ-RWM has near-optimal regret in the adversarial setting (see Section 8.5).

Lemma 8.2.10 (δ-RWM is No-Regret)**.** *Let $\mathcal{I}$ be an m-player game with at most N actions and $L_{\max}$ reward. If the i-th player follows δ-RWM$_\beta^T$ in T rounds of play with learning rate $\beta = 1 - \sqrt{\log(N)/T}$ and approximation factor $\delta \leq \sqrt{\log(N)/T}$,*[5] *then for*

[5]We are assuming $T \gg \log N$. Otherwise, the regret bound becomes $L_{\max}T$, which can be achieved by any arbitrary sequence of choices.

any $\eta > 0$ they experience regret at most

$$\mathrm{Reg}_{T,i} \leq O\left(L_{\max}\sqrt{T}\left(\sqrt{\log N} + \sqrt{\log(1/\eta)}\right)\right)$$

with probability at least $1 - \eta$.

As an immediate corollary, for any game, if we can approximate sample from the RWM distribution efficiently, we immediately get an efficient no-regret learner. In addition, connecting it with Theorem 8.2.8, we also obtain the following corollary for equilibrium computation with RWM.

Corollary 8.2.11 (Equilibrium Computation with δ-RWM)**.** *Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player where $|A_i| \leq N$ for each $i \in [m]$ and reward bounded by $L_{\max}$. Suppose the game is played repeatedly for T rounds. If there is an algorithm which can perform δ-RWM$_\beta^T$ for each player i in time $f_{\mathcal{I}}(T, \delta)$, then there exists an algorithm which computes an ϵ-CCE of $\mathcal{I}$ (Nash if the game is 2-player zero-sum) with probability at least $1 - \eta$. Moreover, the algorithm runs in time*

$$O\left(m \cdot f_{\mathcal{I}}\left(C \cdot L_{\max}\epsilon^{-2}\log(Nm/\eta), \frac{\epsilon}{C \cdot L_{\max}}\right)\right).$$

for some universal constant $C > 0$.

In Sections 8.3 and 8.4, we develop two different types of methods of approximate sampling from the RWM distribution in many well-studied games and discuss their implications. Before moving on, however, it is convenient to briefly discuss one computational consideration that frequently occurs in efficient implementation of RWM. In particular, it will often be the case that our algorithm needs to deal with piece-wise constant functions that map from $[0, n]$ to $\mathbb{R}$ (e.g. reward functions in Blotto for each battlefield are 2-piecewise in this sense). To represent such functions, we will use the following data structure that we refer as a *succinct representation*.

Definition 8.2.12 (Succinct Descriptions of piecewise constant functions)**.** Let $f : [0, n] \mapsto \mathbb{R}$ be a q-piecewise constant function. The succinct description of f, denoted as D_f, consists of q tuples of the form $(a_i, b_i, y_i) \in (\mathbb{Z}^+, \mathbb{Z}^+, \mathbb{R})$ such that for all $x \in [a_i, b_i]$, $f(x) = y_i$ and the intervals $\{[a_1, b_1] \cdots [a_q, b_q]\}$ partition $[0, n]$.

We will often write $|D_f|$ to denote the number of intervals contained in the succinct description. Finally, note that assuming access to succinct descriptions does not lose much generality, as given query access to a standard representation for the monotonic piece-wise function in question (e.g. in the RAM model), it is easy to build a succinct description in time $q \log(n)$ by binary search.

8.3 Playing Games via MCMC-Sampling

In this section, we develop the connections between linear hypergraph games, the RWM distribution, and efficient sampling. In doing so, we unlock access to powerful tools from the sampling literature for the first time in the context of games. This allows for a number of immediate applications including the first no-regret algorithms for well-studied settings such as matroids.

With this in mind, let's first recall the basic framework of Monte Carlo Markov Chains: a powerful tool for approximately sampling from large spaces like the RWM distribution. More formally, consider the following problem: given a distribution π over a large state space A, we'd like to *approximately* sample a state from π in polylog($|A|$) time. MCMC-sampling is an elegant approach to this problem in which one defines a Markov chain M on A satisfying the following three conditions:

1. The stationary distribution of M is π
2. A single step of M can be implemented efficiently
3. M converges quickly to its stationary distribution.

As long as these three conditions hold, it is possible to efficiently sample from π up to any desired accuracy simply by running the Markov chain from any starting position a few steps and outputting the resulting state. More formally, recall that the *mixing time* of a Markov chain M is the number of steps until the resulting distribution is close in TV-distance to π:

Definition 8.3.1 (Mixing Time)**.** The mixing time of a Markov chain M is the worst-case number of steps until the total variation distance of M is close to its stationary measure:

$$T(M, \delta) \coloneqq \min_{t \in \mathbb{N}} : \forall \pi_s,\ TV(M^t \pi_s, \pi) \leq \delta.$$

Thus one only needs to run the chain $T(M, \delta)$ times (from any starting position) in order to δ-approximately sample from π.

While MCMC-sampling is a promising approach, designing efficient Markov chains is typically a challenging task. However, in structured settings such as linear hypergraph games, the distributions arising from RWM seem to be more conducive to the approach. In fact, they correspond to well-studied structure in the approximate sampling literature called *external fields*.

Definition 8.3.2 (External Field)**.** Let π be a distribution over a k-uniform hypergraph $H \subset \binom{\Omega}{k}$. The distribution given by π 'under external field w' for $w \in \mathbb{R}_+^\Omega$ has measure proportional to the product of w across each k-set:

$$\pi^w(s) \ \propto \ \pi(s) \prod_{v \in s} w(v).$$

When π is uniform over H, we often just write H^w for π^w.

It is a simple observation that the distribution arising from RWM (or indeed any reasonable variant) is exactly given by the application of an external field to the state space.

Observation 8.3.3 (RWM $\to$ External Fields). Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player linear hypergraph game. Then for any $i \in [m]$, $RM^{(T)}(\beta)$ can be written as the application of an external field w to A_i such that A_i^w has minimum probability at most $\frac{\beta^{-2L_{\max}T}}{|A_i|}$.

Proof. Assume $i = 1$ without loss of generality (for simplicity of notation). Recall that RWM operates at round $T+1$ by exponentiating the total loss over the previous T rounds:

$$\mathbb{P}\left[RM^{(T)}(\beta) = a\right] \propto \beta^{-\sum_{j=1}^{T} R_1(a, s^{(j)})}$$

where $s^{(1)}, \ldots, s^{(T)} \in A_{-1}$ are the historical strategies played by players $\{2, \ldots, n\}$ in rounds one through T. For simplicity of notation, let $\ell^{(T)}(a) \coloneqq -\sum_{t=1}^{T} R(a, s^{(t)})$ be the total loss. We can similarly define this quantity for any element of the ground set $v \in \Omega$ as:

$$\ell^{(T)}(v) = -\sum_{t=1}^{T} R_1^{\Omega}(v, s^{(t)}).$$

Switching the summations, linearity promises we can express $\ell_T(s)$ as a sum over $\ell_T(v)$:

$$\ell_T(a) = \sum_{v \in a} \ell_T(v).$$

As a result, the RWM distribution is proportional to the product of (exponentiated) total loss for each vertex:

$$\mathbb{P}(a) \;\propto\; \prod_{v \in a} \beta^{\ell^{(T)}(v)}.$$

This is exactly A_1 under the external field $w \in \mathbb{R}_+^{\Omega}$ where $w(v) = \beta^{\ell^{(T)}(v)}$. Since the distribution started uniformly over A_1 and is update by at most $\beta^{L_{\max}}$ in each step, the minimum probability is at worst $\frac{\beta^{-2L_{\max}}}{|A_1|}$. □

As an immediate corollary, we get an efficient no-regret algorithm for any linear hypergraph game whose state space can be approximately sampled under arbitrary external

fields, a well-studied problem in approximate sampling. In the remainder of the section, we'll show how such results lead to new efficient no-regret learning algorithms for many well-studied settings in game theory.

8.3.1 Glauber Dynamics and Fractionally Log-Concave Games

The past few years have seen major advances in approximate sampling various combinatorial objects under external fields [25, 24, 13, 22]. The recent breakthroughs have largely been driven by new analysis techniques for a simple local Markov chain arising from the study of Ising models in statistical physics called the *Glauber Dynamics*. Starting from a state $\sigma \in H \subset \binom{\Omega}{k}$, the (single-site) Glauber Dynamics for a distribution π over A_i are given by the following two-stage procedure:

1. "Down-Step:" Remove a vertex v uniformly at random from σ.
2. "Up-Step:" Sample from π conditional on $\sigma \setminus v$.

It is not hard to show that π is the stationary distribution of this process. The first major breakthrough towards rapid mixing of Glauber Dynamics was due to Anari, Liu, Oveis-Gharan, and Vinzant [25], who used tools developed in the high dimensional expansion literature [239] to prove rapid-mixing of Glauber Dynamics on a broad class of combinatorial objects called *matroids*.

Definition 8.3.4 (Matroids)**.** Let Ω be a ground set and $\mathcal{J}$ a family of subsets of Ω. $(\Omega, \mathcal{J})$ is called a matroid if it satisfies

1. **Non-emptiness:** $\mathcal{J}$ contains at least one subset.
2. **Downward-closure:** For all $S \in \mathcal{J}$ and $S' \subset S$, $S' \in \mathcal{J}$
3. **Exchange-property:** For all $S, S' \in \mathcal{J}$ s.t. $|S| > |S'|$, there exists $x \in S$ s.t. $S' \cup x \in \mathcal{J}$.

An element $S \in \mathcal{J}$ is a **basis** if it is maximal, and the **rank** of the matroid (denoted $r(\mathcal{J})$) is the size of its largest basis.

Note that the bases of a matroid make up an $r(\mathcal{J})$-uniform hypergraph over vertex set Ω. These objects are perhaps best thought of as generalizing the combinatorial structure seen in spanning trees (which form the bases of a 'graphic' matroid). ALOV's [25] major breakthrough was to prove rapid mixing of Glauber Dynamics on matroid bases, a problem known as the Mihail-Vazirani Conjecture (this result was later optimized by Cryan, Guo, and Mousa [102]). Since matroids maintain their structure under external fields (see e.g. [22]), this leads to the following MCMC-algorithm for sampling matroid bases under arbitrary external fields.

Theorem 8.3.5 (Glauber Dynamics on Matroids [22, Theorem 5])**.** *Let H be the set of bases of a rank-k matroid $(\Omega, \mathcal{J})$. Let $w \in \mathbb{R}_+^\Omega$ be any external field. Then, the single-step Glauber Dynamics on H^w has mixing time*

$$T(GD, \delta) \leq O\left(k \log\left(\frac{\log(|\Omega|/w_*)}{\delta}\right)\right).$$

A substantial amount of progress has been made since ALOV and CGM's works. In fact, recently Anari, Jain, Koehler, Pham, and Vuong [22] introduced an even more general class of hypergraphs that can be sampled under arbitrary external fields called fractionally log-concave hypergraphs.[6] All of our results extend to this setting, but to our knowledge matroids already capture most settings of interest in game theory so we focus just on this case for concreteness.

Since matroids are typically exponential size in their rank, we will need implicit access in order to build efficient algorithms. This is typically done through various types of oracle access to the independent sets. For simplicity of presentation, we will assume

[6]Formally this requires a slight generalization known as the q-step Glauber Dynamics.

access to a *contraction oracle*, a standard operation on matroids that restricts the object to independent sets containing some fixed $S \in \mathcal{J}$.

Definition 8.3.6 (Contraction Oracle)**.** Let $(\Omega, \mathcal{J})$ be a rank-k matroid. A rank-r contraction oracle inputs an independent set $S \in \mathcal{J}$ of size r, and outputs (query access to) the contracted matroid $(\Omega, \mathcal{J}_S)$, where

$$\mathcal{J}_S = \{T : T \cup S \in \mathcal{J}\}.$$

Crucially we will only use rank-$(k-1)$ contraction oracles on matroids with rank k (and thus drop rank from the notation below). This can always be implemented in $|\Omega|$ applications of a standard independence oracle (which decides given $S \subseteq \Omega$ whether $S \in \mathcal{J}$), but can often be implemented much more efficiently. For instance, it is easy to see in the case of uniform (unconstrainted) matroids, this can be implemented in $O(k)$ time simply by removing each element of S from the list.

Before stating the main guarantees for no-regret learning and equilibrium computation, it will be useful to discuss a few finer-grained properties typical to games on matroids that help parameterize related computation complexities. First, while not strictly necessary, it will be convenient to restrict our attention to games where the action sets of all players are given by k_i-uniform hypergraphs on some shared groundset Ω, i.e. $A_i \subseteq \binom{\Omega}{k_i}$ for all i. Given such a game $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$, we will typically write $k_{-i} = \max_{j \neq i} k_j$ to denote the maximum support of any viable opponent strategy.

Second, it will be useful to introduce an important property of congestion and security games we call *collision-sensitivity*: the vertex-wise reward of an element $v \in \Omega$ only changes if v is also selected by another opponent.

Definition 8.3.7 (Collision-sensitive Games)**.** Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player linear hypergraph game where $A_i \subseteq 2^\Omega$. We call the rewards of player i 'collision-sensitive' if for

all $v \in \Omega$, the vertex-wise reward of v only changes if another opponent also selects v:

$$\forall v \in \Omega \text{ and } s, s' \in A_{-i} \text{ s.t. } v \notin s, s' : R_i^{\Omega}(v, s) = R_i^{\Omega}(v, s').$$

We will write $\mathrm{NC}_i : \Omega \mapsto \mathbb{R}$ as the function specifying the i-th player's no-collision reward function for each vertex, i.e. $\mathrm{NC}_i(x) = R_i^{\Omega}(x, s)$ for $x \notin s$. We say a collision-sensitive reward has support q if the no-collision reward function NC_i for each player takes on at most q values across all vertices $v \in \Omega$.

In a sense, collision-sensitivity can be thought of as an independence criterion on the vertices: roughly speaking, actions taken on v do not effect actions taken on w for $w \neq v$. With these definitions out of the way, we can now state our main guarantees for no-regret learning on matroids.

Theorem 8.3.8 (RWM on Matroids)**.** *Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player linear hypergraph game on a size-n ground set Ω. If A_i is collision-sensitive with support q, then it is possible to implement δ-RWM$_\beta^T$ in time*

$$O\left(k_i T(CO + q\log(n) + mk_{-i}T\log(n))\log\left(\frac{k_i\log(n) + L_{\max}T\log(\beta^{-1})}{\delta}\right)\right),$$

assuming access to a q-piecewise succinct description of NC_i encoded under an ordering of Ω and a contraction oracle matching the same ordering.

The proof of Theorem 8.3.8 is not particularly interesting beyond combining Observation 8.3.3 and Theorem 8.3.5 and involves mostly tedious implementation details of Glauber Dynamics on matroids. We give these details in Section 8.7 for completeness.

Before moving on, we briefly note this result is nearly tight in many of the main parameters. For instance, the dependence on k,m, and n is $\tilde{O}(mk\log(n))$,[7] which in many

[7]This may increase when the support of other players is non-constant. E.g. if all players are playing bases of k-uniform matroids, we require $mk^2\log(n)$

cases (e.g. uniform matroid) is the number of bits required even to express a set of pure strategies for each player. The bound is also linear in q, which is easily seen to be necessary since one needs to know the q distinct values in order to sample.

Many games in the literature satisfy the conditions of Theorem 8.3.8. We'll end this subsection by giving a few concrete examples. Perhaps the most well-studied variant of these games is a popular setting called *congestion games*. Congestion games are a natural model for resource competition where m-players compete to share n resources and receive rewards dependent on the number of players sharing the same resource.

Definition 8.3.9 (Congestion Game)**.** Given a ground set Ω, an m-player congestion game on Ω consists of a collection $\{A_i\}_{i=1}^m$ and a reward function $c : \Omega \times [m] \to \mathbb{R}$ where each $A_i \subseteq 2^\Omega$ are the strategies of player i, and the reward on the actions $(s_1, \ldots, s_n)$ is given by:

$$R_i(s_i, s_{-i}) = \sum_{e \in s_i} c(e, |e(s)|)$$

where $e(s) = \{s_i : e \in s_i\}$.

Congestion games are particularly well-studied on matroid bases, which are the only structure on which best response is known to converge to Nash in polynomial time. However, to our knowledge Theorem 8.3.8 provides the first no-regret algorithm for congestion games.

Corollary 8.3.10 (Matroid Congestion without Regret)**.** *Let $\mathcal{I} = \{\{A_i\}_{i=1}^m, c\}$ be a congestion game where each A_i is the set of bases of a rank-k_i matroid on a ground set Ω of size n satisfying $\max_{i \in [m]} k_i = k$. Suppose $\mathrm{NC}(e) = c(e, 1)$, the no collision reward function, is q-piecewise under some ordering of Ω. Then there is a no-regret learning algorithm for $\mathcal{I}$ with regret:*

$$\mathrm{Reg}_T \leq O\left(L_{\max}\sqrt{T} \cdot \left(\sqrt{k \log(n)} + \sqrt{\log(1/\eta)}\right)\right)$$

with probability at least $1 - \eta$ *that runs in time*

$$O\left(kT(CO + q\log(n) + Tmk\log(n))\log\left(L_{\max}Tk\log(n)\right)\right)$$

assuming access to a q-piecewise succinct description of NC *encoded in some ordering of* Ω *and a contraction oracle matching this ordering.*

Proof. It is enough to argue the game is linear, as the result then follows immediately from Theorem 8.3.8. Denote by s be the strategy tuples chosen by the players. Recall that the reward of any strategy $s_i \in A_i \subset \binom{\Omega}{k_i}$ in the congestion game is given by:

$$R(s_i, s_{-i}) = \sum_{e \in s_i} c(e, |e(s)|).$$

c can easily be extended into the desired vertex-wise reward function, so we are done. □

Corollary 8.3.11 (Equilibrium Computation for Matroid Congestion)**.** *Let* $\mathcal{I} = \{\{A_i\}_{i=1}^m, c\}$ *be a congestion game where each* A_i *is the set of bases of a rank-*k_i *matroid on a shared ground set* Ω *of size* n *satisfying* $\max_{i\in[m]} k_i = k$*. Suppose* $\mathrm{NC}(e) = c(e, 1)$ *is q-piecewise under some ordering of* Ω*. Then it is possible to compute an* ε*-CCE with probability at least* $1 - \eta$ *in time*

$$O\Bigg(mL_{\max}^2k^2\log(mn/\eta)\varepsilon^{-2}(CO + q\log(n) + mL_{\max}^2k^2\log^2(mn/\eta)\varepsilon^{-2}) \\ * \log\left(L_{\max}k\log(mn/\eta)/\varepsilon\right)\Bigg).$$

We note that these results also easily generalizes matroid congestion over any FLC, unlike the best response strategy for computing Nash. Furthermore, we note that Hedge is actually known to converge to better equilibria [262] than original techniques based on best response, which gives this approach an additional potential advantage.

Another setting particularly well-suited to matroids are *security games*, which model a variety of attack/defense scenarios.

Definition 8.3.12 (Security Game)**.** A security game $\mathcal{I} = (A_d, A_a, \{r, \zeta, c, \rho\})$ over ground set Ω consists of defender actions $A_d \subseteq 2^\Omega$, attacker actions $A_a = \Omega$, and reward/cost functions $r, \zeta, c, \rho : \Omega \to \mathbb{R}$. Let $S \in A_d$, $i \in A_a$ be the actions taken by the defender and the attacker respectively. The reward matrices are given by:

$$R_d(S,i) = \begin{cases} r(i) & \text{if } i \in S \\ c(i) & \text{else,} \end{cases} \quad \text{and} \quad R_a(S,i) = \begin{cases} \zeta(i) & \text{if } i \in S \\ \rho(i) & \text{else.} \end{cases}$$

Security games can model a couple natural settings dependent on the choice of parameters. One basic setting is where the defender has k security resources to defend a set of n targets, and 'wins' if the attacker chooses a defended target. On the other hand, the model also captures the complement of this game where the defender chooses k targets to distribute key resources, and the attacker wins if they intercept this distribution (pick one of the k marked targets). Security games have broad applicability in practice, and indeed have been used in cases such as assigning security checkpoints at LAX [350].

Security games are inherently linear in their natural representation and thus admit efficient no-regret algorithms when the defender's state-space is a matroid (simulating RWM for the attacker is trivial as it corresponds to a size-$|\Omega|$ multinomial distribution).

Corollary 8.3.13 (Security without Regret)**.** *Let $\mathcal{I} = (A_d, A_a, \{r, \zeta, c, \rho\})$ be a security game where A_d are the bases of a rank-k matroid on the ground set Ω and $A_a = \Omega$. Suppose $c, \rho : \Omega \mapsto \mathbb{R}$ are q-piecewise under some ordering of Ω. Then there exists a no-regret learning algorithm for $\mathcal{I}$ with regret:*

$$\text{Reg}_T \leq O\left(L_{\max}\sqrt{T} \cdot \left(\sqrt{k\log(n)} + \sqrt{\log(1/\eta)}\right)\right)$$

with probability at least $1-\eta$ *that runs in time*

$$O\left(kT(CO + q\log(n) + T\log(n))\log\left(k\log(n)TL_{\max}\right)\right),$$

assuming access to q-piecewise succinct descriptions for c *and* ρ *encoded in some ordering of* Ω *and a contraction oracle matching this ordering.*

Proof. Note that the attacker's strategy consists of bases of just rank-1 matroid so implementing RWM for her is trivial. We focus on the implementation for the defender side. Again it is enough to show the game is linear. By definition, we have

$$R_D(S,i) = \sum_{j\in S} R_D^{\Omega}(j,i)$$

where $R_D^{\Omega} : [n]\times[n] \to \mathbb{R}$ is

$$R_D^{\Omega}(i,j) = \begin{cases} r(i) & \text{if } i = j \\ c(i) & \text{else.} \end{cases}$$

□

As an immediate corollary, we also get fast equilibrium computation.

Corollary 8.3.14 (Equilibrium Computation for Security)**.** *Let* $\mathcal{I} = (A_d, A_a, \{r, \zeta, c, \rho\})$ *be a security game where* A_d *are the bases of a rank-k matroid on the ground set* Ω *and* $A_a = \Omega$*. Suppose* $c, \rho : \Omega \mapsto \mathbb{R}$ *are q-piecewise under some ordering of* Ω*. Then it is possible to compute an* ε*-CCE (Nash if the game is zero-sum) with probability at least* $1-\eta$ *in time*

$$O\left(L_{\max}^2 k^2 \log(n/\eta)\varepsilon^{-2}(CO + q\log(n) + L_{\max}^2 k\varepsilon^{-2}\log^2(n/\eta))\log\left(k\log(n/\eta)\epsilon^{-1}L_{\max}\right)\right),$$

assuming access to an q-piecewise succinct description for c and ρ encoded under some ordering of Ω and a contraction oracle matching this ordering.

We note that this result easily generalizes to settings with multiple attackers or an attacker who chooses targets corresponding to a matroid basis.

8.3.2 Dueling Games and the JSV Chain

Matroids (or more generally FLC's) are not the only type of constrained state space that can be sampled under arbitrary external fields. Indeed, long before these results Jerrum, Sinclair, and Vigoda [223] famously proved (in work on approximating the permanent) that bipartite matchings have this property as well. We give an improved version of their result due to Bezáková, Štefankovič, Vazirani, and Vigoda [63].

Theorem 8.3.15 (JSV Chain)**.** *Let $(K_{n,n}, w)$ be an edge-weighted complete bipartite graph, and consider the distribution over perfect matchings given by:*

$$Pr(M) \propto \prod_{e \in M} w_e.$$

It is possible to δ-approximately sample from this distribution in $\widetilde{O}(n^7 \log \frac{1}{\delta w_{\min}})$ time, where $w_{\min}$ is the minimum weight.

Note one can phrase this result as a sampling algorithm for permutations over external fields, where the state-space is viewed as a subset of $[n]^n$. Like matroids, bipartite matchings are very natural objects and underlie a fair number of well-studied games. In this section, we focus on the setting of *dueling games*. Dueling games model two player competitive optimization over a shared ground set.

Definition 8.3.16 (Dueling Games)**.** A dueling game $\mathcal{I} = (\Omega, \mu, A_1, A_2)$ consists of a set Ω, a distribution μ over Ω, and strategy spaces $A_1, A_2 \subset \mathbb{R}_+^{\Omega}$. The reward matrices are

given by the probability of ranking $x \sim \mu$ higher than the opponent:

$$R_1(s,t) = Pr_{x\sim\mu}[s(x) > t(x)] - Pr_{x\sim\mu}[t(x) > s(x)],$$

and likewise:

$$R_2(s,t) = Pr_{x\sim\mu}[t(x) > s(x)] - Pr_{x\sim\mu}[s(x) > t(x)].$$

There is no known polynomial time algorithm for computing equilibria of general dueling games. We will give a general algorithm for a class of dueling games we call *unrestricted*.

Definition 8.3.17 (Unrestricted Dueling Games)**.** A dueling game $\mathcal{I} = (\Omega, \mu, A_1, A_2)$ is called un-restricted if there exist subsets $S_1, S_2 \subset \mathbb{R}$ with $|S_1| = |S_2| = |\Omega|$ such that A_1 (respectively A_2) consists of all possible assignments of Ω to S_1 (respectively S_2).

It is not hard to see that unrestricted dueling games are linear over perfect matchings in a complete bipartite graph. As a result, we can use the JSV-chain to simulate optimistic hedge in polynomial time.

Theorem 8.3.18 (Sampling Unrestricted Dueling Games)**.** *Let $\mathcal{I} = (\Omega, \mu, A_1, A_2)$ be an unrestricted dueling game where $|\Omega| = n$. Then it is possible to implement δ-RMW_β^T in time $\widetilde{O}(T^2 n^7 \log(1/\delta))$.*

Proof. We focus on player 1. The result is analogous for player 2. Strategies in an unrestricted dueling game correspond to perfect matchings in the complete bipartite graph $K_{n,n}$, where the LHS corresponds to elements of Ω, and the RHS corresponds to elements in S_1. To fit into our prior framework of linearity and external fields, one may view these perfect matchings as elements of E^n (where E is the edge set of $K_{n,n}$). Recall that the reward is given by:

$$R_1(s,t) = Pr_{x\sim\mu}[s(x) > t(x)] - Pr_{x\sim\mu}[t(x) > s(x)]$$

It is not hard to see this is linear over the edges of matching:

$$R_1(s,t) = \sum_{e \in s} R_1^{\Omega}(e,t)$$

where $R_1^{\Omega} : E \times A_2 \to \mathbb{R}$ is given by:

$$R_1^{\Omega}(\{v,w\},t) = \begin{cases} \mu(v) & \text{if } w > t(v) \\ -\mu(v) & \text{if } w < t(v) \\ 0 & \text{else} \end{cases}$$

As a result, Observation 8.3.3 implies that RWM is given by the application of an external field over the edges of perfect bipartite matchings, which is exactly the distribution considered in Theorem 8.3.15. All that is left is to efficiently build access to the weights of the underlying bipartite graph, which is a small onetime cost that is asymptotically dominated by even the mixing time of the JSV chain. As a result, it is enough to run Theorem 8.3.15 T times, which gives the resulting runtime bound. □

We note that this result can easily be generalized to a slightly larger class of games where $|S_1|$ and $|S_2|$ may be larger than Ω, and specific edges in the bipartite representation may be disallowed (i.e. we might add the constraint that $x \in \Omega$ can never be given rank 1). Such strategies correspond to sampling matchings on a generic bipartite graph (rather than $K_{n,n}$), and no-regret learning can can also be performed by the JSV-chain.

Finally, we'll look at a classic dueling game that fit into the unrestricted framework: ranking duel. Ranking duel (or the 'search engine game') is a game where two players compete to choose the best ranking of n items. One of these items is pulled from a known distribution, and the player who ranked it higher wins.

Ranking duel is an unrestricted game where $S_1, S_2 = [n]$ and the action spaces $A_1, A_2 = \mathcal{S}_n$, i.e. permutations of n. As a result Theorem 8.3.18 immediately implies an

efficient algorithm for sampling in δ-RWM.

Corollary 8.3.19 (Ranking Duel without Regret)**.** *Let $\mathcal{I} = ([n], \mu, \mathcal{S}_n, \mathcal{S}_n)$ be an instance of ranking duel. Then there exists an algorithm with regret:*

$$\mathrm{Reg}_T \leq O\left(\sqrt{T} \cdot \left(\sqrt{k \log(n)} + \sqrt{\log(1/\eta)}\right)\right)$$

with probability at least $1 - \eta$ that runs in time $\widetilde{O}(T^2 n^7)$.

As a corollary we get the fastest known equilibrium computation for ranking duel,

Corollary 8.3.20 (Equilibrium Computation for Ranking Duel)**.** *Let $\mathcal{I} = ([n], \mu, S_n, S_n)$ be an instance of ranking duel. Then there exists an algorithm computing an ε-CCE (Nash if the game is zero-sum) with probability at least $1 - \eta$ in time $\tilde{O}(n^9 \log(1/\eta)/\varepsilon^4)$*

Unfortunately, while the JSV-chain is an improvement over previous extended linear programming approaches to dueling games [211, 4], n^9 can hardly be called a practical running time. In fact, it should be noted there is a faster known no-regret algorithm for perfect bipartite matchings called PermELearn that runs in $O(Tn^4)$ time.

Thus Theorem 8.3.18 is perhaps more interesting from the perspective of the method than the result itself. Designing faster and simpler Markov chains for sampling bipartite matchings has long been a favorite open problem in the sampling community. Our setting gives a nice intermediate version of this problem, as the matching problems arising from unrestricted dueling games have particularly nice structure. In particular, they correspond to *monotonic weightings*, in the sense that for every fixed vertex v on the LHS on the graph, the edge weight of $w(v, i) \geq w(v, j)$ if $i \geq j$. Matchings with monotonic weights are actually well-studied in the literature, including the resolution of the monotone column permanent conjecture [79] and rapid mixing of the switch chain[8] for binary monotonic

[8]The switch chain is in essence the 2-step Glauber Dynamics on the view of matchings as a subset of E^n.

weights [134]. However despite these related results, a fast algorithm for sampling general bipartite matchings with monotonic weights remains an interesting open problem, and the application to practical no-regret algorithms for dueling games gives yet another motivation for its study.

8.4 Playing Games via DP-Sampling

While MCMC-sampling is a powerful tool, standard techniques like Glauber Dynamics may not perform well in settings that require global coordination across coordinates. In this section, we develop a new sampling technique toward this end based on Dynamic Programming, taking advantage of the fact that many settings of interest, such as Colonel Blotto, additionally exhibit certain recursive structure. In particular, we consider a large class of problems called *Resource Allocation Games* that broadly generalize the Colonel Blotto game.

Resource Allocation Game..

In a resource allocation game, each player assigns fungible items to some number of battlefields. Namely, for the i-th player, the action space A_i is the set of ordered size k_i partition of n_i for $k_i, n_i \in \mathbb{Z}^+$. [9] One can see the action space is indeed a hypergraph where the vertices correspond to pairs (h, x) interpreted as "assigning x items to the h-th battlefield", and a strategy is simply a subset of vertices $(1, x_1), \cdots, (k, x_k)$ satisfying $\sum_{h=1}^{k} x_h = n$. Let the variables $x_{i,h}$ denote the number of items assigned by the i-th player to the h-th battlefields and $A_{-i} := A_1 \times \cdots \times A_{i-1} \times A_{i+1} \times \cdots \times A_m$ denote the set of action tuples from players other than player i. The reward structure of a resource allocation game is defined by a set of battlefield reward functions $r_{i,h} : [0, n] \times A_{-i} \mapsto \mathbb{R}$ for each player $i \in [m]$ and battlefield $h \in [k_i]$. Let $\mathbf{a}_{-i} \in A_{-i}$ be the actions picked by

[9] While we define actions as all k-partitions of $[n]$, our algorithm also applies to action spaces that are subsets of these partitions with arbitrary assignment constraints on each battlefield (i.e., of the form, at most m items can be assigned to battle j).

players other than player i. The total reward received by the i-th player is given by the sum of rewards over individual battles:

$$R_i(x_{i,1}, \cdots, x_{i,h}, \mathbf{a}_{-i}) = \sum_{h=1}^{k_i} r_{i,h}(x_{i,h}, \mathbf{a}_{-i}).$$

Additionally, let $r_{i,h,\mathbf{a}_{-i}} : [0, n] \mapsto \mathbb{R}$ be the restriction of $r_{i,h}$ after fixing the strategies of the other players i.e. $r_{i,h,\mathbf{a}_{-i}}(x) = r_{i,h}(x, \mathbf{a}_{-i})$. We say the resource allocation game is q-piecewise and monotonic if for every $i \in [m], h \in [k_i]$, and $\mathbf{a}_{-i} \in A_{-i}$, $r_{i,h,\mathbf{a}_{-i}}$ is a q-piecewise constant and monotonically increasing function.

In the remainder of this section, we discuss how one achieves no-regret learning for the first player (the algorithm for the other players is analogous). For this purpose, we will drop the subscript indicating the player number and let $n = n_1, k = k_1, r_h = r_{1,h}, \mathbf{a} = \mathbf{a}_{-1}$.

RWM Distributions..

To achieve no-regret learning, we will need to approximately sample from the distributions arising from running the RWM algorithm on Resource Allocation Games. Assume that the players have played the game for T rounds. Let $a^{(t)} \in A_{-1}$ be the action tuples observed from all but the first player at the t-th round. For an action $x \in A_1$ that assigns x_h items to the h-th battlefield, RWM will set its weight $w_T(x)$ as

$$w_T(x) = \beta^{\sum_{h=1}^{k} \sum_{t=1}^{T} r_h(x_h, a^{(t)})}, \tag{8.5}$$

where $\beta \in [1/2, 1)$ is the learning rate. For simplicity, we will define the cumulative battlefield reward function (negated to simplify the syntax appearing after)

$$\ell_h^{(T+1)}(x_h) = -\sum_{t=1}^{T} r_j(x_h; a^{(t)}), \tag{8.6}$$

so the weight for strategy x can alternatively be written as $\prod_{h=1}^{k} \beta^{\ell_h^{(t+1)}(x_h)}$. We remark that, if the resource allocation game is monotonic and q piecewise, $\ell_j^{(T+1)}(\cdot)$ is also monotonically increasing and a $((q-1)\cdot t+1)$-piecewise constant function. Though the domain of ℓ_j is of size $n+1$, the property allows us to represent it succinctly in space $O(qt)$, which is critical when we try to design algorithms whose runtime depends on n polylogarithmically. In this section, we focus on how one could design algorithms to efficiently sample from $RM^{(T)}(\beta)$ approximately given succinct descriptions of the functions $r_{h,\mathbf{a}^{(t)}}$ for all $h \in [k]$ and $t \in T$ (recall that $r_{h,\mathbf{a}^{(t)}}$ is the restriction of the battlefield reward function r_h after fixing the other players' actions). This allows us to implement δ-RWM in polylogarithmic time.

Theorem 8.4.1 (RWM in Resource Allocation Game)**.** *Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player monotonic, q-piecewise resource allocation game where $A_1 = P_k(n)$ for $n, k \in \mathbb{Z}^+$. Suppose the reward of the first player is bounded by $L_{\max}$. Then it is possible to implement δ-RWM$_\beta^T$ in time*

$$O(Tk)\cdot\left(\min\left(T^2 L_{\max}\log(1/\beta)/\delta\cdot\zeta_1\left(\log T+\log\zeta_2\right)\cdot\log n\,,n\log n\right)+\min(Tq,n)\right).$$

where $\zeta_1 := \min\left(L_{\max}\log(1/\beta)/\delta, q\right), \zeta_2 := \max\left(L_{\max}\log(1/\beta)/\delta, q\right)$, assuming access to a Tq-piecewise succinct description of $\ell_h^{(t)}$ defined in Equation (8.6) for all $h \in [k], t \in [T]$.

As corollaries of the above theorem, we obtain efficient no-regret learners and algorithms for computing CCEs in Resource Allocation Games.

In the remainder of this section, we present the sampling algorithm whose analysis leads to the proof of Theorem 8.4.1, and discuss a number of applications to games in the literature including Colonel Blotto and its variants.

8.4.1 Sampling via estimation of partition function

We will focus for the moment on how one could sample from the RWM distribution just in round T. For that purpose, we often omit the superscript T for the function $\ell_h^{(T)}$

(Equation (8.6)) for simplicity. To sample with a dynamic program, we define the partition function for an RWM distribution in resource allocation.

Partition Function.

The partition function $f_h : [0] \cup \mathbb{Z}^+ \mapsto \mathbb{R}^+$ for $h \in [k]$ is defined as the sum of partial weights of all strategies that allocate y soldiers in the subgame induced on the first h battles. Namely,

$$f_h(y) = \sum_{x_1+\cdots+x_h=y} \prod_{i=1}^{h} \beta^{\ell_i(x_i)}. \tag{8.7}$$

It has long been known that efficient algorithms for computing the partition function of a self-reducible problem imply efficient (approximate) samplers for the problem's solution space [224]. As one can see, computing the partition function f_h in our setting simply corresponds to counting the number of (weighted) size h partitions of y, which is exactly such a self-reducible problem. Consequently, if one has the values for the partition functions precomputed, one can use them to sample from the RWM distribution efficiently. We provide the detailed sampling procedure below for completeness.

In particular, this is done by sampling the number of items to put in each battlefield sequentially, conditioned appropriately on prior choices. One puts $x_1 \in \{0, \ldots, n\}$ soldier to the first battlefield with probability

$$\mathbb{P}[x_1 = y] \propto \beta^{\ell_1(y)} \cdot f_{k-1}(n-y). \tag{8.8}$$

To sample from the $(h+1)$-th battlefield conditioned on the fact that one has put $x_1, \ldots, x_h$ soldiers in battles $1 \ldots h$, it is enough to sample according to the distribution

$$\mathbb{P}\left[x_{h+1} = y | x_{1\ldots h}\right] \propto \beta^{\ell_{h+1}(y)} \cdot f_{k-h-1}\left(n - \left(\sum_{j=1}^{h} x_j\right) - y\right). \tag{8.9}$$

The probabilities defined according to Equations (8.8), (8.9) yields exactly the RWM distribution, but computing the partition function exactly can be quite costly. For this purpose, we consider the notion of δ-*approximations* (multiplicative) of functions.

Definition 8.4.2 (δ-approximation)**.** Given $f : [0, n] \mapsto \mathbb{R}^+$ and $\hat{f} : [0, n] \mapsto \mathbb{R}^+$, we say $\hat{f}$ is a δ-approximation of f if for all $x \in [0, n]$ we have

$$\Big(1 - \delta\Big) f(x) \leq \hat{f}(x) \leq \Big(1 + \delta\Big) f(x).$$

Fortunately, with $\delta/(Ck)$-approximations of the partition functions for some sufficiently large constant C, one can still perform δ-*approximate* sampling from the RWM distribution (See proof of Lemma 8.4.3).

Another important observation for achieving efficient (approximate) sampling is that ℓ_h, the reward function for each battlefield, is a piece-wise constant function. Hence, further optimization is possible if the approximations used for each partition function is also piece-wise constant (and this is indeed the case as we will see shortly). The pseudo-code for the sampling algorithm that takes as input the succinct descriptions of the (approximate) partition functions and the reward functions is given below.

Algorithm 2. Partition-Sampling

Require: Succinct descriptions $D_{\hat{f}_h}$ for $h \in [k]$; succinct description D_{g_h} of the function $g_h(x) = \beta^{\ell_h(x)}$.

Initialize the number of unused soldiers $u = n$.

Initialize an empty assignment description S.

for $h = 1 \dots k$ **do**

Compute the succinct description D_{κ_h} of the function $\kappa_h : [0, u] \mapsto \mathbb{R}^+$ defined as

$$\kappa_h(i) := g_h(i) \cdot \hat{f}_{k-h}(u - i). \tag{8.10}$$

$\triangleright$ $|D_{\kappa_h}| = |D_{g_h}| + |D_{\hat{f}_h}|$ and D_{κ_h} can be computed in time linear with respect to the description length.

{*Compute intervals*}

for $(a_i, b_i, y_i) \in D_{\kappa_h}$ **do**

Compute the cumulative weight of constant intervals.

$$\nu_i := y_i \cdot (b_i - a_i + 1). \tag{8.11}$$

{*Sample an interval*}

Sample $j \in 1 \dots |D_{\kappa_h}|$ according to the weight vector ν.

{*Sample soldiers used in battle h*}

Sample z_h uniformly for $\{a_j, a_j + 1, \dots b_j - 1, b_j\}$ where $[a_j, b_j]$ is the j-th constant interval in D_{κ_h}.

Add z_h to the strategy S description.

$u \leftarrow u - z_h$.

return S.

Lemma 8.4.3. *Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player monotonic, q-piecewise resource allocation game where $A_i = P_{k_i}(n_i)$ for $n_i, k_i \in \mathbb{Z}^+$. At the t-th round, for each $h \in [k]$, let $\hat{f}_h$ be $\delta/(2k)$-approximations of the partition function defined in Equation (8.7), and let $g_h(x) = \beta^{\ell_h(x)}$. Assume one is given the succinct descriptions $D_{\hat{f}_h}$ and D_{g_h}. Then, there exists an algorithm **Partition-Sampling** which performs δ-approximate sampling from $RM^{(T)}(\beta)$ in time*

$$k \cdot O\left(p + \min\left(Tq, n\right) + \log n\right)$$

where $p := \max |D_{\hat{f}_h}|$.

Proof. If we were to perform the sampling process with f_h instead of $\hat{f}_h$, we would get exactly the distribution $RM^{(t)}(\beta)$. This follows from repeated applications of Bayes' rule and that we are sampling the correct conditional distribution. Namely,

$$\mathbb{P}\left[x_1 = y_1, \ldots, x_k = y_k\right] = \prod_{h=1}^{k} \mathbb{P}\left[x_h = y_h \mid x_{h'} = y_{h'}, \forall h' < h\right]$$

To make sure that we are performing δ-approximate sampling overall, it suffices if we perform δ/k approximate sampling for each conditional distribution (each battlefield). To perform exact sampling, one needs to compute the weight

$$\kappa_h^*(i) := g_h(i) \cdot f_{k-h-1}(u - i).$$

Since for every $h \in [k]$ we have $\hat{f}_h$ are $\delta/(2k)$ approximations of f_h, we also have κ_h is $\delta/(2k)$ approximations of κ_h^*, which implies that

$$(1 - \delta/(2k)) \sum_{i=0}^{n} \kappa_h^*(i) \leq \sum_{i=0}^{n} \kappa_h(i) \leq (1 + \delta/(2k)) \sum_{i=0}^{n} \kappa_h^*(i).$$

It then follows that the distribution defined by κ_h^* and κ_h differs by at most δ in total variation distance.

To analyze the runtime, we note that for each battlefield, we first compute the succinct description D_{κ_h} defined in Equation (8.10). Since it is the point-wise multiplication between g_h, which is $\min(Tq, n)$-piecewise constant, and $\hat{f}_{k-h-1}$, which is p-piecewise constant, κ_h will be $p+\min(Tq, n)$ piecewise constant. To construct the succinct description of D_{κ_h}, one maintains two pointers $a = 0, b = u$ and keeps track of the interval from D_{g_h} that a is in and the interval from $D_{\hat{f}_h}$ that b is in. Then, one shifts a forward and b backward to seek for constant intervals of D_{κ_h}. It is easy to see the runtime of the construction is linear with respect to $|D_{\kappa_h}|$. Then, computing ν_i requires scanning through the succinct description of D_{κ_h} once. After that, we first sample from a multinomial distribution with support at most p, which takes time $O(p)$. Then, we sample from a uniform distribution with support at most n, which takes time $O(\log n)$. Adding everything together then gives our final runtime. □

8.4.2 Computing the Partition Function

We now move to showing how to (approximately) compute the partition function. As a warmup, we will first show how this can be done exactly via dynamic programming. In particular, we want to fill a $k \times n$ table such that the (h, y) entry corresponds to the value $f_h(y)$.

Proposition 8.4.4. *The values $f_h(y)$ for all $h \in [k]$ and $y \in [0, n]$ can be computed in time $O(nk \log n)$.*

Proof. Notice that we have the following recursion

$$f_h(y) = \sum_{x=0}^{y} \beta^{\ell_h(x)} \cdot f_{h-1}(y - x). \tag{8.12}$$

f_h is exactly the convolution of $\beta^{\ell_h(\cdot)}$ and f_{h-1}. Using Discrete Fast Fourier Transform, f_h can be evaluated in time $O(n \log n)$ ([81]). Hence, in total, the entire DP table can be filled in time $O(nk \log n)$.

□

Next, we will discuss how one can develop faster algorithms when n is substantially larger than k, T, and $L_{\max}$. Our main technical result is an algorithm to pre-compute the partition functions "approximately" whose runtime depends on n *polylogarithmically.*

Proposition 8.4.5. *There exists an algorithm* ***Approx-DP*** *which constructs $\hat{f}_1, \cdots, \hat{f}_k$ such that $\hat{f}_i$ is a δ-approximation of the partition function f_i pointwisely, and runs in time*

$$O\left(kT^2 L_{\max} \log(1/\beta)/\delta \cdot \zeta_1 \left(\log T + \log \zeta_2\right) \cdot \log n\right).$$

where $\zeta_1 := \min\left(L_{\max} \log(1/\beta)/\delta, q\right), \zeta_2 := \max\left(L_{\max} \log(1/\beta)/\delta, q\right)$.

This seems a bit surprising as there are in total $k \cdot (n+1)$ values that we need to pre-compute ($f_h(y)$ for all $h \in [k]$ and $y \in [0, n]$). However, notice that we are only interested in computing approximations to these values. And, as each f_h is itself a monotonically increasing function, we can approximate it with a sufficiently simple piece-wise function.

Fact 8.4.6. *Given a monotonically increasing function $f : [0, n] \mapsto \mathbb{R}^+$, it can be δ-approximated by a function that is d-piecewise constant where*

$$d = \Theta(\log(\max_x f(x) / \min_x f(x))/\delta).$$

Hence, the algorithm **Approx-DP** does not need to output the entire $k \times (n+1)$ tables specifying the partition functions. Rather it can just construct the succinct descriptions of a series of functions $\hat{f}_1, \cdots, \hat{f}_k$ such that $\hat{f}_i$ is a δ-approximation of f_i. By Claim 8.4.6, we can indeed find such $\hat{f}_i$ that are $\Theta(\log(\beta^{TL_{\max}})/\delta) = \Theta(TL_{\max} \log(1/\beta)/\delta)$ piecewise constant. As the first building block of the **Approx-DP**, we demonstrate the routine which, given query access to an unknown monotonically increasing function, constructs a piecewise constant approximation of the function.

Lemma 8.4.7. *Given a monotonically increasing function $f : [0, n] \mapsto \mathbb{R}^+$ and query access to f, there exists an algorithm* ***Piecewise-Approximate*** *which outputs a piecewise function $\hat{f}$ satisfying that*

- *$\hat{f}$ is d-piecewise constant for $d = \Theta(\log(\max_x f(x)/\min_x f(x))/\delta)$.*
- *$(1-\delta) \cdot f(x) \leq \hat{f}(x) \leq f(x)$.*
- *the algorithm runs in time $O(\log n \cdot d \cdot Q)$.*

where Q is the cost of making a single query to f.

Proof. The algorithm proceeds by iteratively finding the longest interval from a starting point such that the function values at the endpoints are within a $(1+\delta)$-factor of each other, and then letting $\hat{f}$ on this interval be the constant function given by the value of f on the right endpoint. By the monotonicity of f, it is easy to see that $\hat{f}$ is indeed a δ-approximation of f on this interval. The algorithm then repeats this process starting from the next value on which $\hat{f}$ is not yet defined and repeats until $\hat{f}$ has been defined on the entire domain.

Since the function values of f increase by at least a factor of $(1+\delta)$ between each interval and the last, the total number of intervals is at most $d = \Theta(\log(\max_x f(x)/\min_x f(x))/\delta)$. □

Algorithm 3. Piecewise-Approximate

Require: Query access to $f : [0, n] \mapsto \mathbb{R}^+$; Approximation accuracy δ.

1: $D \leftarrow \{\}, a \leftarrow 0$.
2: **while** $a \leq n$ **do**
3: Binary Search to find the largest b such that $f(b) \leq f(a) \cdot (1+\delta)$.
4: Add $(a, b, f(a))$ to D. (setting $\hat{f}(x)$ to $f(a)$ for all $a \leq x \leq b$.)
5: $a \leftarrow b + 1$.
6: **return** D.

With this in mind the construction of the series of piecewise constant approximation functions $\hat{f}_1, \cdots, \hat{f}_k$ becomes clear: one initializes $\hat{f}_1 =$ **Piecewise-Approximate**(f_1) and then defines $\hat{f}_h$ recursively as $\hat{f}_h =$ **Piecewise-Approximate**(f'_h) where $f'_h(y) = \sum_{x=0}^{y} \hat{f}_{h-1}(x) \cdot g_h(y-x)$. Recall that the routine **Piecewise-Approximate** requires query access to the input function. Hence, we need to show how one could implement query access to f_1 and f'_h for $h \in [k]$ efficiently (independent of n). The former is easy given the succinct description of f_1 since $f_1 = \beta^{\ell_1(x)}$ is a $T \cdot q$ piecewise constant function. To realize the latter, we use the fact that f'_h is the convolution of $\hat{f}_h$ and g_h which are both piecewise constant functions. In particular, we give the following routine which efficiently implements query access to convolutions of piecewise constant functions.

Algorithm 4. Convolution-Query

Require: Succinct descriptions D_f,D_g of two piece-wise constant functions f, g; Query point x.

1: Preprocess D_f to extend each tuple into the form (a_i, b_i, y_i, s_i) where $s_i := \sum_{j<i}(b_j - a_j + 1) \cdot y_j$.

2: Let

$$F(z) = \begin{cases} \sum_{i=0}^{z} f(i) \text{ for } i \geq 0\,, \\ 0 \text{ otherwise.} \end{cases}$$

3: $res \leftarrow 0$.

4: **for** each interval $(a_i, b_i, y_i) \in D_g$ **do**

5: $\quad res \mathrel{+}= (F(x - a_i) - F(x - b_i - 1)) \cdot y_i$.

6: **return** res.

We note that the pre-processing incurs a one-time cost for each new function f and does not need to be performed for different queries with respect to the same function f. We also note $F(z)$ can be evaluated by first binary search the first index j such that for

$(a_j, b_j, y_j, s_j) \in D_f$ we have $b_j \geq z$. Then use the equality $F(z) = s_j + (z - a_j) \cdot y_j$.

Lemma 8.4.8. *Assume one is given the succinct description D_f, D_g of two functions $f, g : [0, n] \mapsto \mathbb{R}^+$. Let $(f \star g)(x) = \sum_{i=0}^{x} f(i) \cdot g(x - i)$. Suppose $|D_f| = p_f$ and $|D_g| = p_g$ with $p_f < p_g$. There exists an algorithm **Convolution-Query** that takes $O(p_f + p_g)$ time to preprocess D_f, D_g and then takes $O(p_f \cdot \log p_g)$ time to return query access $(f \star g)(x)$ for each $x \in [0, n]$.*

Proof. In the pre-processing step, for each tuple $(a_i, b_i, y_i) \in D_f$, we add an extra number s_i which denotes the prefix-sum of all elements before the interval $[a_i, b_i]$. This can be done easily by scanning through the tuples of succinct description in order in one pass. Then, if one want to query the prefix sum $F(z)$, one can just find out which interval z falls into by binary search and then computes in constant time with s_i.

Then, using the fact that g is piecewise constant, we can rewrite the convolution query $(f \star g)(x)$ as

$$\sum_{i=0}^{x} f(x-i) \cdot g(i) = \sum_{i=1}^{|D_g|} y_i \cdot \left(\sum_{j=x-b_i-1}^{x-a_i} f(j) \right) = \sum_{i=1}^{|D_g|} y_i \cdot \Big(F(x-a_i) - F(x-b_i-1) \Big).$$

where (a_i, b_i, y_i) are tuples in D_g. Since evaluating each query to F takes at most $O(\log |D_f|)$ time (for binary search), the above expression can be evaluated in time $O(|D_g| \log |D_f|)$. □

We are now ready to present the pseudocode and analysis of **Approx-DP**, whose analysis then lead to the proof of Proposition 8.4.5.

Algorithm 5. Approx-DP

Require: For each $h \in [k]$, succinct description D_{g_h} of the function $g_h(x) = \beta^{-\ell_h(x)}$; number of goods n; number of battlefields k; approximation error δ.

1: Initialize $\hat{f}_1$ as g_1.

$$D_{\hat{f}_1} \leftarrow \textbf{Piecewise-Approximate}\left(g_1, \delta/(4k)\right).$$

2: **for** $h = 2 \cdots k$ **do**

3: Consider the function

$$f'_h(y) = \sum_{x=0}^{y} \hat{f}_{h-1}(x) \cdot g_h(y-x) = \sum_{x=0}^{y} \hat{f}_{h-1}(x) \cdot \beta^{\ell_h(y-x)}. \tag{8.13}$$

4: Set $\hat{f}_h$ to be the piecewise approximation of f'_h.

$$D_{\hat{f}_h} \leftarrow \textbf{Piecewise-Approximate}\left(f'_h, \delta/(4k)\right).$$

5: **return** $D_{\hat{f}_h}$ for $h \in [k]$.

We note that the query access to g_1 is by simply reading from the succinct description D_{g_1}. Further, query access to $f'_h(y)$ is implemented by routine **Convolution-Query**$(D_{\hat{f}_{h-1}}, D_{g_h}, y)$ (see the Appendix for its pseudocode and runtime).

Proof of Proposition 8.4.5. We will show via induction that each $\hat{f}_h$ is monotonically increasing and $\hat{f}_h$ an $(h \cdot \delta/k)$-approximation of the original function f_h. Consider the function f'_h defined in Equation (8.13). Since both $\hat{f}_{h-1}(x)$ and $\beta^{-\ell_h(x)}$ are monotonically increasing, their convolution f'_h is also monotonically increasing. Besides, by our inductive hypothesis, $\hat{f}_{h-1}$ is an $((h-1) \cdot \delta/k)$-approximation of f_{h-1}, implying that f'_h is an $((h-1) \cdot \delta/k)$-approximation of f_h. By Lemma 8.4.7, $\hat{f}_h$ is a (δ/k)-approximation of f'_h,

and consequently a $(h \cdot \delta/k)$-approximation of f_h.

The runtime of the algorithm is dominated by the $(k-1)$ times we call the routine **Piecewise-Approximate** in Line 4. Notice that $\max_x f_h(x)/\min_x f_h(y)$ can be at most $\beta^{-\ell(x)}$ where $\ell(x)$ is at most $TL_{\max}$. As we have argued, each f'_h is always a δ-approximation of f_h, which implies that $\max_x f'_h(x)/\min_x f'_h(y)$ is at most $\exp(2 \cdot \log(1/\beta)TL_{\max})$. Therefore, each $\hat{f}_h$ is a $O(\log(1/\beta)TL_{\max})$ constant function.

Piecewise-Approximate uses the routine **Convolution-Query** as its query access to the input function f'_h. By Lemma 8.4.8, **Convolution-Query** incurs a one-time cost of $O(TL_{\max}\log(1/\beta)/\delta + Tq)$ to preprocess the succinct descriptions of $g_h(x) = \beta^{\ell_h(x)}$ and $\hat{f}_{h-1}$. Then, each query takes time $O\left(T \cdot \zeta_1 \cdot (\log T + \log \zeta_2)\right)$. where $\zeta_1 := \min\left(L_{\max}\log(1/\beta)/\delta, q\right), \zeta_2 := \max\left(L_{\max}\log(1/\beta)/\delta, q\right)$. By Lemma 8.4.7, each call to **Piecewise-Approximate** then takes time

$$O\left(T^2 L_{\max}\log(1/\beta)/\delta \cdot \zeta_1 \left(\log T + \log \zeta_2\right) \cdot \log n\right).$$

The overall runtime just multiplies the entire expression by k. □

Combining Lemmas 8.4.3,8.4.4, and Lemma 8.4.5, we then obtain an efficient algorithm for δ-RWM in Resource Allocation Games, which concludes the proof of Theorem 8.4.1.

8.4.3 Applications of the meta algorithm

In this section, we describe the main applications of our sampling algorithm.

Colonel Blotto Game.

A well-studied example of the resource allocation game is the Colonel Blotto Game. In the game, m players try to assign $\{n_i\}_{i=1}^m$ troops to k different battlefields. For the i-th battlefield, the player who places more soldiers wins the battle and earn a reward of $w_i \in \mathbb{Z}^+$

(ties are broken e.g. lexicographically).[10] This can be viewed as a resource allocation game where the reward function r_j is simply the threshold function $r_j(x) = w_j \cdot \mathbb{1}(x > y)$ where y is the maximum number of soldiers placed by the other players. It is easy to see that r_j is monotonically increasing and 2-piecewise constant. Hence, Theorem 8.4.1 immediately gives an efficient no-regret learning algorithm for the Colonel Blotto Game.

Corollary 8.4.9 (Colonel Blotto without Regret)**.** *Let $\mathcal{I}$ be an m-player Colonel Blotto Game where the i-th player tries to assign n_i soldiers to k battlefields satisfying $n_i \leq n$. Then there is a no-regret learning algorithm for $\mathcal{I}$ with regret:*

$$\mathrm{Reg}_T \leq O\left(L_{\max}\sqrt{T} \cdot \left(\sqrt{k\log(n)} + \sqrt{\log(1/\eta)}\right)\right)$$

with probability at least $1-\eta$ that runs in time

$$Tk \cdot O\left(\min\left(n\log n, T^2 L_{\max}\log(TL_{\max})\log n\right) + m\right).$$

Proof. By Lemma 8.2.10, to achieve no-regret learning in T rounds, we simply need to perform δ-approximate sampling from the RWM distributions with learning rate $\log(1/\beta)$, where $\delta = \sqrt{k\log n/T}$ and $\beta = 1 - \sqrt{k\log n/T}$. For $T \geq C \cdot k\log n$ for a sufficiently large constant, we have $\log(1/\beta) = O(\sqrt{k\log n/T})$. Hence, $\log(1/\beta)/\delta = O(1)$. To show that we can perform the sampling process efficiently, we will apply Theorem 8.4.1 with $q = 2$ since Colonel Blotto is a resource allocation game whose reward function is always 2-piecewise. To do so, we need to construct and maintain the succinct descriptions of the cumulative battlefield reward function $\ell_h^{(t)}$ required by the sampling algorithm. Let $a_{j,h}^{(t)}$ be the number of soldiers that the j-th player assigns to the h-th battlefield at the t-th

[10] In the zero-sum variant the other player also loses w_i.

round. Then, for the first player, we essentially have

$$\ell_h^{(t)}(x) = w_h \cdot \sum_{t'=1}^{t-1} \mathbb{1}\left\{x > \max_{j=2\cdots m} a_{j,h}^{(t')}\right\}.$$

After observing the assignments from other players at round t, one first computes the maximum $\nu_h^{(t)} = \max_{j=2\cdots m} a_{j,h}^{(t)}$, which can be done in $O(m)$ time. After that, one essentially adds $\ell_h^{(t-1)}$ with the threshold function $\mathbb{1}\left\{x > \nu_h^{(t)}\right\}$, which takes time $O(\min(T, n))$, which is strictly dominated by the sampling time. Overall, the update just adds an additive factor of $O(Tkm)$ in total. □

This immediately gives an algorithm for computing ϵ-approximate CCEs in the Colonel Blotto Game (or Nash Equilibria when $m = 2$). Namely, we simulate no-regret playing for all m players simultaneously for $T = C \cdot kL_{\max}^2 \epsilon^{-2} \log(mn/\eta)$ many rounds.

Corollary 8.4.10 (Equilibrium Computation for Colonel Blotto Games)**.** *Let $\mathcal{I}$ be an m-player Colonel Blotto Game where the i-th player tries to assign n_i soldiers to k battlefields satisfying $n_i \leq n$. There exists an algorithm to compute an ϵ-approximate CCE (Nash if the game is two-player zero-sum) with probability at least $1 - \eta$ in time*

$$m \cdot O\Bigg(mk^2\epsilon^{-2}L_{\max}^2 \log(mn/\eta) \min\Bigg(nk^2 L_{\max}^2 \epsilon^{-2} \log(m/\eta) \cdot \log^2 n\,,\; k^4 L_{\max}^7 \epsilon^{-6} \log^4(n) \\ \cdot \log^3(m/\eta) \cdot \log\left(kL_{\max}\epsilon^{-1}\log(mn/\eta)\right)\Bigg)\Bigg).$$

In the regime where n is far greater than k, ϵ, and $L_{\max}$, this gives an exponential improvement over prior algorithms [4] (at the cost of being approximate rather than exact).

Dice Game.

Dice Games are a randomized variant of Blotto first proposed in [105] where m-players[11] construct and roll dice, with the highest roller winning the game. These games are another natural instance of resource allocation. More formally, in the dice game each of m-playes has a k_i-sided die and n_i points to distribute. The i-th player builds their die by assigning a number of dots to each face such that the sum is exactly n_i (where $k_i, n_i \in \mathbb{Z}^+$). Note that the action space A_i is then exactly the set of ordered partitions $P_{k_i}(n_i)$. The rewards are determined by "rolling" the m dice simultaneously; the player with the highest roll wins. In other words, for each player a face is selected uniformly at random and independently, and the player with more dots on the chosen faces wins and earns a reward of 1. The reward function is given by the expected reward of this process.

Let $x_{i,h}$ be the number of dots that the i-th player placed on the h-th face of her die for $i \in [m]$ and $h \in k_i$, and denote by X_i the random variable representing the number of dots obtained by player i after rolling. The expected utility for the first player is given by

$$\begin{aligned}
\mathbb{P}[X_1 > \max(X_2, \cdots, X_m)] &= \frac{1}{k_1} \sum_{h=1}^{k_1} \mathbb{P}[x_{1,h} > \max(X_2, \cdots, X_m)] \\
&= \frac{1}{k_1} \sum_{h=1}^{k_1} \prod_{i=2}^{m} \mathbb{P}[x_{1,h} > X_i]. \\
&= \frac{1}{k_1} \sum_{h=1}^{k_1} \prod_{i=2}^{m} \frac{1}{k_i} \sum_{h'=1}^{k_i} \mathbb{1}(x_{1,h} > x_{i,h'}).
\end{aligned}$$

Let $k = \max_i k_i$. It is not hard to see that $\frac{1}{k_i} \sum_{h'=1}^{k_i} \mathbb{1}(x_{1,h} > x_{i,h'})$ is an $O(k)$-piecewise monotonic function and $\prod_{i=2}^{m} \frac{1}{k_i} \sum_{h'=1}^{k_i} \mathbb{1}(x_{1,h} > x_{i,h'})$ is an $O(mk)$ piecewise monotonic function. Hence, this is indeed a $O(mk)$-piecewise monotonic resource allocation game. Applying our meta-algorithm immediately gives the following results on no-regret learning

[11]The original dice game is defined for two players. This can be nonetheless generalized naturally to an m-player setting.

and equilibrium computation in dice games.

Corollary 8.4.11 (No-regret Learning in Dice Games)**.** *Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m player dice game such that $\max_i k_i \le k$ and $\max n_i \le n$. Then there is a no-regret learning algorithm for $\mathcal{I}$ with regret*

$$\mathrm{Reg}_T \le O\left(L_{\max}\sqrt{T} \cdot \left(\sqrt{k\log(n)} + \sqrt{\log(1/\eta)}\right)\right)$$

with probability at least $1-\eta$, and runs in time

$$Tk \cdot O\left(\min\left(T^2 \cdot \log(Tm) \cdot \log n, n\log n\right) + Tkm\right).$$

Proof. Similar to the proof of Theorem 8.4.9, we will apply Theorem 8.4.1 with $q := mk$ and $L_{\max} = 1$. The main step is to maintain the succinct description of $\ell_h^{(t)}$. In dice games, the cumulative reward functions $\ell_h^{(t)}$ for all faces are identical and take the form

$$\ell^{(t)}(x) = \sum_{t'=1}^{t-1} \frac{1}{k_1} \prod_{i=2}^{m} \frac{1}{k_i} \sum_{h=1}^{k_i} \mathbb{1}(x > x_{i,h}^{(t')}).$$

After observing the actions $x_{i,h}^{(t)}$ for $i \in [m]$, $h \in [k_i]$ at the t-th round, we first compute the functions $\nu_i^{(t)}(x) = \frac{1}{k_i}\sum_{h=1}^{k_i} \mathbb{1}(x > x_{i,h}^{(t)})$, which are all at most k-piecewise. This requires sorting $x_{i,1}^{(t)}, \cdots x_{i,k_i}^{(t)}$, which takes time at most $O(k\log k)$. As a result, constructing all $\nu_2^{(t)}, \cdots \nu_m^{(t)}$ takes time in total $O(mk\log k)$. Then, we will point-wisely multiply all $\nu_i^{(t)}$ together. One can proceed in a divide and conquer manner: in the first pass, multiply together $\nu_i^{(t)}$ in groups of two, in the second pass, multiply together $\nu_i^{(t)}$ in groups of four and continues until all $\nu_i^{(t)}$ are multiplied together. There will be $\log m$ passes and the computation cost for each pass is at most $O(mk)$. Hence, the process takes time $O(mk\log m)$ in total. Then, we add the resulting function to $\ell^{(t-1)}$, which incurs another cost of $O(Tkm)$. Hence, in total, (assuming $\log m < T$), it takes time $O(Tkm)$ to update

$\ell^{(t)}$ at one round. □

Corollary 8.4.12 (Equilibrium for Dice Games)**.** *Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be a dice game with $n = \max(n_1, \ldots, n_m)$, and $k = \max(k_1, \ldots, k_m)$. There exists an algorithm to compute an ϵ-CCE (Nash if the game is two-player zero-sum) with probability at least $1 - \eta$ in time*

$$m \cdot O\Bigg(\min\left(nk^2\epsilon^{-2}\log(m/\eta)\cdot\log^2 n, k^4\epsilon^{-6}\log^4(n)\cdot\log^3(m/\eta)\cdot\log(k\epsilon^{-1}m\log(n/\eta))\right) + mk^4\epsilon^{-4}\log^2(mn/\eta)\Bigg).$$

8.4.4 Multi-resource allocation games

A natural generalization of the resource allocation game is when each player has multiple resource types. This occurs naturally in many settings: a Colonel in Blotto, for instance, might have access to multiple unit types including troops, tanks, and planes (this variant was introduced in [56]). One would expect that the reward functions should vary depending on which types of units the Colonel chooses.

More formally, in the multi-resource allocation game, the i-th player has B_i types of fungible items. We denote by $n_{i,b}$ the number of type-b items that the i-th player possesses. Her strategy is an allocation of these items to k_i battlefields. We denote $X_{i,b,h}$ as the number of type b items that the i-th player assigns to the h-th battlefield. Similar to the single-resource allocation game, for each player i and each battlefield h, there is a battlefield reward function

$$r_{i,h} : ([0, n_{i,1}]) \times \cdots \times ([0, n_{i,B}]) \times A_{-i} \mapsto \mathbb{R}\,,$$

where we recall A_{-i} is the set of strategy tuples from the players other than i. Let $S \in A_{-i}$ be the strategies used by the other players, the total reward for the i-th player on strategy

$X = \{X_{i,b,h}\}$ is given by summing over the rewards on each individual battlefield:

$$R_i(X, S) = \sum_{h=1}^{k_i} r_{i,h}\left(X_{i,1,h}, \cdots, X_{i,b,h}, S\right).$$

We now prove a variant of Theorem 8.4.1 for the multi-resource setting.

Theorem 8.4.13 (RWM in Multi-Resource Allocation Game)**.** *Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player multi-resource allocation game where $A_i = P_{k_i}(n_{i,1}) \times \cdots \times P_{k_i}(n_{i,B})$ for $n_1, \cdots n_B, k \in \mathbb{Z}^+$. Suppose the reward of the first player is bounded by $L_{\max}$, $\max_{i\in[m],b\in[B]} n_{i,b} = n$, $\max_{i\in[m]} k_i = k$. Then it is possible to implement (exact) RMW_β^T in time*

$$O\left(TkB(n+1)^{2B} + mTk^2B \cdot (n+1)^B\right).$$

assuming query access to $r_{i,h} : ([0, n_{i,1}]) \times \cdots \times ([0, n_{i,B}]) \times A_{-i} \mapsto \mathbb{R}$.

Proof. We proceed to analyze the partition function of the RWM distribution. As usual, we will conduct the analysis from the first player's perspective and drop the subscript used to index the player. Suppose the game is played for T rounds and the observed actions from the other players' are $S^{(1)}, \cdots, S^{(T)}$. Then, similar to Equation (8.6), for an assignment $x \in A_i$, we will define the cumulative reward

$$\ell_h^{(T+1)}(X_{1,h}, \cdots, X_{b,h}) = \sum_{t=1}^{T} r_h\left(X_{1,h}, \cdots, X_{B,h}, S^{(t)}\right).$$

For simplicity, we will abbreviate $(X_{1,h}, \cdots, X_{b,h})$ as $X_{*,h}$. Then, we have the weight for the action x is simply

$$w_T(X) = \prod_{h=1}^{k} \beta^{\ell_h^{(T+1)}(X_{*,h})}.$$

After dropping the superscript marking the rounds, accordingly, the partition function is

now $f_h : ([0,n])^B \mapsto \mathbb{R}^+$ for $h \in [k]$ defined as

$$f_h(y) = \sum_{X \in \mathcal{X}_y} \prod_{h=1}^{k} \beta^{\ell_h(X_{*,h})},$$

where $\mathcal{X}_y = \{X \in (\mathbb{Z}^+)^{B \times k} | \forall b \in [B], \sum_{h=1}^{k} X_{b,h} = y_b\}$. We still have the recursion

$$f_h(y) = \sum_{z \in \mathcal{Z}_y} f_h(y-z) \cdot \beta^{\ell_h(z)}, \tag{8.14}$$

where $\mathcal{Z}_y = \{z \in ([0,n])^B | \forall b, z_b \leq y_b\}$.

Compared to the single-resource allocation game, there are $O\left(k \cdot (n+1)^B\right)$ partition function values we need to compute. Using dynamic programming and the recursion stated in Equation (8.14), each of them now takes time at most $O\left(B(n+1)^B\right)$. Hence, filling the entire DP table takes time $O\left(kB \cdot (n+1)^{2B}\right)$.

After that, we likewise sample the assignment for each battlefield sequentially. We will write $\mathbf{n} = (n_1, \cdots, n_B)$. For the first battlefield, we sample

$$\mathbb{P}[X_{*,1} = y] \propto \beta^{\ell_1(y)} \cdot f_{k-1}(\mathbf{n} - y).$$

To sample from the $(h+1)$-st battlefield, one sample according to the distribution

$$\mathbb{P}\left[X_{*,h+1} = y | X_{*,1\cdots h}\right] \propto \beta^{\ell_{h+1}(y)} \cdot f_{k-h-1}\left(\mathbf{n} - \left(\sum_{j=1}^{h} X_{*,j}\right) - y\right).$$

The domain size of the distributions we sample from is $O\left((n+1)^B\right)$. To compute the probabilities of each element takes $O(B)$ times. Hence, the sampling time for one battlefield is $O\left(B \cdot (n+1)^B\right)$. Hence, the runtime of the sampling process is dominated by that of computing the partition functions.

Finally, we discuss how we maintain the function $\ell_h^{(t)}$ at round t. To do that,

we gather the strategy $S^{(t)}$ observed at the t-th round from other players and then query $r_h(z, S^{(t)})$ for each $z \in ([0, n])^B$, and add that to $\ell_h^{(t-1)}(z)$. Each query takes time $O(mBk)$ (to write down the input). In total, maintaining $\ell_h^{(t)}$ in T rounds takes time $O(mTk^2B(n+1)^B)$. Adding this together with the time for computing the partition function then gives our final runtime. □

As an immediate application, we get no-regret learning and equilibrium computation for the multi-resource allocation games such as the multi-resource Colonel Blotto problem.

Corollary 8.4.14 (Multi-resource without Regret)**.** *Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player multi-resource allocation game where $A_i = P_{k_i}(n_{i,1}) \times \cdots \times P_{k_i}(n_{i,B})$ for $n_1, \cdots n_B, k \in \mathbb{Z}^+$. Suppose the reward of the first player is bounded by $L_{\max}$, $\max_{i\in[m],b\in[B]} n_{i,b} = n$, $\max_{i\in[m]} k_i = k$. Then there is a no-regret learning algorithm for $\mathcal{I}$ with regret:*

$$\mathrm{Reg}_T \leq O\left(L_{\max}\sqrt{T} \cdot \left(\sqrt{\sum_{b=1}^{B} \log(n_B)} + \sqrt{\log(1/\eta)}\right)\right)$$

with probability at least $1 - \eta$ that runs in time

$$O\left(TkB(n+1)^{2B} + mTk^2B \cdot (n+1)^B\right)$$

assuming query access to $r_{i,h} : ([0, n_{i,1}]) \times \cdots \times ([0, n_{i,B}]) \times A_{-i} \mapsto \mathbb{R}$.

Corollary 8.4.15 (Equilibrium Computation for Multi-Resource Allocation Games)**.** *Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player multi-resource allocation game where $A_i = P_{k_i}(n_{i,1}) \times \cdots \times P_{k_i}(n_{i,B})$ for $n_1, \cdots n_B, k \in \mathbb{Z}^+$. Suppose the reward of the first player is bounded by $L_{\max}$, $\max_{i\in[m],b\in[B]} n_{i,b} = n$, $\max_{i\in[m]} k_i = k$. There exists an algorithm to compute an*

ϵ-approximate CCE (Nash if two-player zero-sum) with probability at least $1-\eta$ in time

$$O\left(m(n+1)^{2B}k^2B^2L_{\max}^2\epsilon^{-2}\log(mn/\eta) + m^2(n+1)^Bk^3B^2L_{\max}^2\epsilon^{-2}\cdot\log(mn/\eta)\right)$$

assuming query access to $r_{i,h}:([0,n_{i,1}])\times\cdots\times([0,n_{i,B}])\times A_{-i}\mapsto\mathbb{R}$.

Acknowledgements

We thank Thuy Duong Vuong, Nima Anari, and Kuikui Liu for helpful discussions on MCMC-based approaches for Dueling games and Colonel Blotto, and Saeed Seddighin for discussions on algorithms for Colonel Blotto and related problems. We also thank Maxwell Fishelson for enlightening discussion of optimistic hedge and its variants.

8.5 No-regret Learning and Equilibrium Computation

Lemma 8.5.1 (Lemma 4.1 from [88] Rephrased)**.** *Let $\mathcal{I}=\{(A_i,R_i)_{i=1}^m\}$ be an m-player game played repeatedly for T rounds. Denote $\mathbf{s}_i^{(t)}$ as the mixed strategy chosen by the i-th player, and $a_i^{(t)}\sim\mathbf{s}_i^{(t)}$ as the action sampled. Assume the i-th player follows an algorithm which computes $\mathbf{s}_i^{(t)}$ solely based on $a_j^{(t')}$ for $t'<t$ and $j\in[m]\backslash\{i\}$. Furthermore, suppose the following is true*

$$\sup_{b_j^{(t)}\in A_j \text{ for } t\in[T],j\neq i}\mathbb{E}\left[\max_{e\in A_i}\sum_{t=1}^T R_i(e,b_{-i}^{(t)})-\sum_{t=1}^T R_i(a_i^{(t)},b_{-i}^{(t)})\right]\leq B.$$

Then, for all $\delta\in(0,1)$, with probability at least $1-\delta$, it holds

$$\max_{e\in A_i}\sum_{t=1}^T R_i(e,a_{-i}^{(t)})-\sum_{t=1}^T R_i(a_i^{(t)},a_{-i}^{(t)})\leq B+L_{\max}\sqrt{T/2\log(1/\delta)}.$$

at the end of the repeated play for the i-th player.

Proof of Lemma 8.2.10. If one follows exactly from the Randomized Weighted Majority

algorithm, one has the guarantee that

$$\sup_{b_j^{(t)} \in A_j \text{ for } t\in[T], j\neq i} \mathop{\mathbb{E}}_{a_i^{(t)} \sim RM_i^{(t)}(\beta)} \left[\max_{e\in A_i} \sum_{t=1}^{T} R_i(e, b_{-i}^{(t)}) - \sum_{t=1}^{T} R_i(a_i^{(t)}, b_{-i}^{(t)}) \right] \leq L_{\max}\sqrt{T\log N}.$$

In reality, since we are performing κ-approximate sampling, we have

$$\mathrm{TV}\left(RM_i^{(t)}(\beta), \mathbf{s}_i^{(t)}\right) \leq \kappa.$$

Since the reward of the game is bounded by $L_{\max}$, we have

$$\begin{aligned}
&\sup_{b_j^{(t)} \in A_j \text{ for } t\in[T], j\neq i} \mathop{\mathbb{E}}_{a_i^{(t)} \sim \mathbf{s}_i^{(t)}} \left[\max_{e\in A_i} \sum_{t=1}^{T} R_i(e, b_{-i}^{(t)}) - \sum_{t=1}^{T} R_i(a_i^{(t)}, b_{-i}^{(t)}) \right] \\
\leq & L_{\max}\sqrt{T\log N} + \kappa \cdot L_{\max} \cdot T.
\end{aligned}$$

Then, by Lemma 8.5.1, it then holds

$$\begin{aligned}
&\frac{1}{T}\max_{e\in A_i} \sum_{t=1}^{T} R_i(e, a_{-i}^{(t)}) - \frac{1}{T}\sum_{t=1}^{T} R_i(a_i^{(t)}, a_{-i}^{(t)}) \\
\leq & L_{\max}\sqrt{\log N/T} + \kappa \cdot L_{\max} + L_{\max}\sqrt{\log(1/\delta)/(2T)}.
\end{aligned}$$

with probability at least $1-\delta$. Setting $T = L_{\max}^2\epsilon^{-2}\log(N/\delta)$ and $\kappa = \min(1/2,)$ then gives the average regret is bounded by $O(\epsilon)$. □

Proof of Corollary 8.2.11. By Lemma 8.2.10, if we set $T = C\cdot(L_{\max}^2\epsilon^{-2}\log(Nm/\eta))$, $\delta = \epsilon/(CL_{\max})$ for a sufficient large constant and simulate the repeated game playing for T rounds where each player makes her decision based on δ-RWM, the regret of the i-th player is bounded by ϵ with probability at least $1-\delta/m$. By union bound, this holds for all players simultaneously. Ths results then follows from Theorem 8.2.8. □

8.6 Bit-complexity and Stability of Numeric Operations

To avoid un-necessary technical details of the bit-complexities of numbers and time complexities of algebraic operations, the algorithmic results in the main body of this work are stated in the *Algebraic Computation Model*. In particular, we assume additions, subtractions, multiplications, divisions, exponentiation and comparisons can be carried between real numbers in constant time, and that the computing device has query access to the digits of the real numbers. We remark that, for all the games we study, if the rewards are rational numbers with bounded bit complexities, our algorithms can all be implemented exactly in the RAM computation model with their runtime increased by at most polynomial factors. Unsurprisingly, if one is more careful with the numeric precision needed and maintains only multiplicative approximations of each algebraic operations, our algorithms can be implemented in the RAM model losing only poly-logarithmic factors. In this section, we discuss some standard techniques to this end for the reader interested in any implementation of our algorithms.

We first discuss how we represent and perform algebraic operations on numbers whose absolute values are exponentially large or small. While writing these numbers down exactly is costly, for the purpose of δ-approximate sampling, it is actually sufficient t keep $\text{poly}(\delta)$ multiplicative approximations of these numbers. Fortunately, these approximations can indeed be represented much more succinctly using the scientific notations. For convenience, for a number $a \in \mathbb{R}^+$, we will call $\tilde{a}$ a δ-approximation of a if we have $(1-\delta)a \leq \tilde{a} \leq (1+\delta)a$ and a one-sided δ-approximation if we have $(1-\delta)a \leq \tilde{a} \leq a$.

Fact 8.6.1. *Given $a \in \mathbb{R}^+$ satisfying $\exp(-q) \leq a \leq \exp(q)$ for $q \in \mathbb{Z}^+$, let $\tilde{a}$ be a written in scientific form keeping $\Theta(\log(1/\delta))$ many significant figures. Then, $\tilde{a}$ is a one-sided δ-approximation of a and can be represented using $\Theta(\log(\delta^{-1}) + \log q)$ many bits.*

Instead of performing exact arithmetic computations, we can perform 'approximate'

arithmetic operations on real numbers in all our algorithms.

Claim 8.6.2. Let $a, b \in \mathbb{R}^+$ be two numbers in scientific notations with s significant figures.

- one-sided δ-approximation of additions and multiplications can be performed in time $O(s)$ and $O(s \log s)$ respectively.
- δ-approximation of division can be performed in time $O(s + \log(1/\delta))$.
- Given $1 \leq \alpha \leq 2$ that has s significant figures and $i \in \mathbb{Z}^+ \cup \{0\}$, a one-side δ-approximation of α^i can be computed in time $O(s \cdot \log^2 i \cdot \log(\delta^{-1}))$.

Proof. The first two claims follow from the definition of (one-sided) δ-approximation. We proceed to show that one can perform approximate exponentiation efficiently. In particular, we argue that α^i can be computed fairly accurately via fast exponentiation while keeping $C \cdot \log(i/\delta)$ significant figures throughout the computation for some large enough constant C. By doing so, we can make sure the approximation to α is a one-sided ξ-approximation where $\xi = \delta/i^c$ for some large enough constant c. Consequently, the approximation of β^j for any $j \in [i]$ that is a power of 2 is within $(1 \pm 4^{\log_2(i)} \cdot \xi)$. It then follows α^i can be approximated within $(1 \pm \xi \cdot O(\log i) \cdot 4^{\log_2(i)}) = (1 \pm \delta)$ when c is sufficiently large. □

Unsurprisingly, the output of applying a series of arithmetic operations will be within multiplicative factors of the result obtained by replacing each operation with its approximate counterpart.

Fact 8.6.3. *Given a variable y that is the result of V arithmetic operations including Addition, Multiplication and Division on the inputs $x_1, \cdots, x_n \in \mathbb{R}^+$ in scientific notations with s significant figures, let $\hat{y}$ be the variable obtained by replacing all the arithmetic operations with their $\delta/(10V)$-approximate counterparts for small enough δ. Then, $\hat{y}$ will be a δ-approximation of y. Moreover, if only additions and multiplications are used, the approximation is one-sided i.e. $\hat{y} \leq y$.*

Careful readers may find that subtraction is excluded when we discuss approximate algebraic operations in Claim 8.6.2. For two numbers $a, b \in \mathbb{R}^+$ and $\hat{a}, \hat{b}$ be their δ approximations counterparts, $\hat{a} - \hat{b}$ may be wildly different from $a - b$ when a is substantially larger than b. Yet, subtraction between real numbers is indeed used in two different places. Firstly, subtractions occur in Discrete Fourier Transform, which is used in Proposition 8.4.5 to compute the convolution between functions. The numeric stability of DFT varies among different implementations and depends on a number of subtle factors (See [336]). If numeric stability indeed becomes an issue in the actual implementation, one can fallback to evaluate the convolution in the brute-force manner, which increases the complexity from $O(nk \log n)$ to $O(n^2 k)$.

Another place where subtractions are used is in Algorithm 4 to compute range sum of piece-wise function efficiently. As such, we need a numerically more stable technique for performing range sum query in place of the prefix sum technique. In particular, given the succinct description of a q-piecewise constant function $f : \{0\} \cup [n] \mapsto \mathbb{R}^+$ and $\delta \in (0, 1)$ beforehand, we want to perform some preprocessing in time $q \cdot \text{polylog}(n, q, \delta^{-1})$ and then answer a series of queries of the form $\sum_{i=a}^{b} f(i)$ within $(1 \pm \delta)$ multiplicative factors in time $\text{polylog}(n, q, \delta^{-1})$.

Claim 8.6.4. Given the succinct description D_f of a q-piecewise constant function $f : \{0\} \cup [n] \mapsto \mathbb{R}$ where the function values contain at most s significant figures, there exists an algorithm **Range-Sum-Query** which performs some preprocess in time $O(qs) + \text{polylog}(q, \delta^{-1}, n)$, and can compute one-sided δ-approximation to query of the form $\sum_{i=a}^{b} f(i)$ in time $\text{polylog}(n, q, \delta^{-1})$.

Proof. Let $D_f = \{(a_1, b_1, y_1), \cdots (a_q, b_q, y_q)\}$. In the preprocessing step, we first compute the range sum of all intervals $[a_i, b_i]$ (approximately). Denote the results as an array $[s_1, \cdots, s_q]$. Then, we build a segment tree with the array, where nodes store the approximate range sum of intervals of lengths that are powers of 2. This takes $O(q \log(q))$

arithmetic operations. The data structure then allows us to answer $\sum_{i=a}^{b} s_i$ with $O(\log(q))$ arithmetic operations with enough accuracy. Then, when we receives a query $\sum_{i=a}^{b} f(i)$. We first binary search for the intervals of f that a, b fall into respectively. Next, one uses the pre-built segment tree to answer the range sum of any intervals that are strictly contained in $[a, b]$ and then adds the sum of remaining elements. It is then not hard to see that the pre-processing step takes time at most $O(qs) + \text{polylog}(q, \delta^{-1}, n)$, and answering each query takes time $\text{polylog}(q, \delta^{-1}, n)$. □

Lastly, we discuss the building block of sampling: sampling from multinomial distributions. Typically, our algorithm computes a vector $w_1, \cdots, w_n$ and then samples from the multinomial distribution X where $\mathbb{P}[X = i] \propto w_i$. This is simple in the Algebraic computation model as one can easily reduce this to sampling from uniform distributions over real intervals (which can be done at assumed unit cost). In particular, one first computes the prefix sum $W_1, \cdots W_n$. Then, one samples z from the uniform distribution over the interval $[0, W_n]$ and returns j for $W_{j-1} < z < W_j$. This clearly takes at most time $O(n)$. In the bit-complexity model, we can nonetheless achieve approximate sampling from arbitrary multinomial distributions with similar runtime.

Claim 8.6.5. Given a weight vector $(w_1, \cdots, w_n)$ in scientific notations with s significant figures, δ-approximate sampling from the multinomial distribution X such that $\mathbb{P}[X = i] \propto w_i$ can be done in time $\widetilde{O}(n \cdot (\log(\delta^{-1}) + s))$.

Proof. After reading the input, one first truncates to make sure each w_i has at most $O(\log(n/\delta))$ many significant figures as that is already enough for the specified 'sampling accuracy'. After that, all arithmetic operations will be carried out with their $c \cdot \delta/n$ approximations for some sufficiently small constant c. In the next step, one normalizes the weight vector and rounds each w_i to their nearest multiple of $c \cdot \delta/n$. Doing so changes the distribution by at most $c \cdot \delta$ in total variation distance. One can then multiply all w_i by a factor of $n/(c \cdot \delta)$ to make everything an integer. Finally, one can do the same

thing as sampling in the Algebraic computation model: computing the prefix sums and reducing the problem to sampling from uniform distributions, now over integer intervals. The integers in the interval can be at most $n^2/(c \cdot \delta)$ so the runtime is dominated by the preliminary computations performed. □

We note that in many cases we actually require a slightly more complicated sampling procedure where we wish to sample from a q-piecewise support-n multinomial with $q \ll n$. This can be done similarly in time $\widetilde{O}((q + \log(n))(\log(\delta^{-1}) + s))$ by first sampling one of the q piecewise intervals by the above technique, then sampling uniformly within the interval.

8.7 Implementing Glauber Dynamics

This section is devoted to proving Theorem 8.3.8, which we repeat here for convenience.

Theorem 8.7.1 (RWM on Matroids)**.** *Let $\mathcal{I} = \{(A_i, R_i)_{i=1}^m\}$ be an m-player game on a size-n ground set Ω. If A_i consists of the bases of a rank-k matroid, is linear, and is collision-sensitive with support q, then it is possible to implement δ-RWM$_\beta^T$ in time*

$$O\left(k_i T(CO + q\log(n) + mk_{-i}T\log(n))\log\left(\frac{k_i\log(n) + L_{\max}T\log(\beta^{-1})}{\delta}\right)\right),$$

assuming access to a q-piecewise succinct description of NC_i *encoded under an ordering of Ω and a contraction oracle matching the same ordering.*

Proof. Since A_i is given by the bases of a rank-k matroid, the single-step Glauber Dynamics on Ω^w mix in time

$$T(GD(\Omega^w), \delta) \leq O\left(k\log\left(\frac{\log(|A_i|/w_*)}{\delta}\right)\right)$$

for any external field $w \in \mathbb{R}_+^\Omega$. In our setting, we have that $|A_i| \leq n^k$, and $w_* \geq \beta^{L_{\max}T}$, and the process needs to be repeated once per round bringing the complexity to:

$$O\left(kT \log\left(\frac{k\log(n) + L_{\max}T\log(\beta^{-1})}{\delta}\right)\right)$$

times the implementation time of a single step of Glauber dynamics.

It is left to bound this cost. To implement a step of GD in the t-th round of optimistic hedge, we first remove a uniformly random element from our current basis, then re-sample from the conditional distribution. The first step can easily be implemented in $O(\log(k))$ time. The latter step requires more care. Let $\hat{e}$ denote the $(k-1)$-size set resulting from the down-step of the walk. Query the contraction oracle on $\hat{e}$ and call the resulting set $S_e \subset E$. Notice that by definition, the conditional measure of any $x \in S_e$ is proportional to $\beta^{-r_t(x)}$ where

$$r_t(x) = \sum_{j=1}^{t} R_i^\Omega(x, s^{(j)}).$$

Thus to perform the conditional sampling efficiently, it is sufficient to compute the external field for each element in S_e and sample from the corresponding multinomial distribution.

While implementing this naively would require time at least $|S_e|$ to check the weight of each element in the conditional distribution, this can be circumvented via our assumption that our game is collision-sensitive with bounded support. In particular, assume for the moment we have access to a succinct description for the vertex-wise total rewards $r_t(v)$ that is $(q + tmk_{-i})$-piecewise, and that the output of the contraction oracle respects the order of the description (we will argue this can be constructed efficiently shortly). As a result, the total rewards in S_e are $(q + tmk_{-i})$-piecewise as well. This means that using query access to CO,[12] we can build a succinct description for total rewards on the elements

[12]Formally we are also assuming here one has query access to the size of the output of the contraction oracle. Note this can be easily implemented in polylog(n) time even if one does not assume such access.

in $|S_e|$ (labeled by their index in CO). Sampling from the corresponding multinomial distribution in the algebraic computation model then takes $O((q + tmk_{-i}))$ time, and one can then feed the sampled index into CO to receive the correct vertex. Altogether, a single step of GD can therefore be implemented in $O(CO + (q + tmk_{-i}) \log(n))$ time assuming access to the appropriate description of total rewards.

Finally, we argue we can construct and maintain the succinct descriptions of the vertex-wise reward functions over T rounds efficiently. Recall we start with an q-piecewise succinct description for the no-collision vertex-wise reward values. In each round, at most mk_{-i} new elements of Ω are introduced into the history, and since the game is collision-sensitive the resulting succinct description of rewards is at most $(q + tmk_{-i})$-piecewise in the t-th round as desired. The computational cost stems from noting that it is actually sufficient just to update the rewards for vertices which have appeared in the opponent history (and the number of rounds in which it has appeared). During look-up, computing the total reward for any vertex v that has appeared t times can be computed in $O(1)$ time by simply adding the stored value $(T - t)R_i(v, s)$ for any $s \notin v$. The cost of building the succinct description is therefore asymptotically dominated by the sampling procedure above, which gives the final complexity. □

This chapter, in full, is based on the material as it appears in the Symposium on Discrete Algorithms 2023. Beaglehole, Daniel; Hopkins, Max; Kane, Daniel; Liu, Sihan; Lovett, Shachar. "Sampling Equilibria: Fast No-Regret Learning in Structured Games". The dissertation author was a primary investigator and author of this material.

Bibliography

[1] Heiner Ackermann, Heiko Röglin, and Berthold Vöcking. On the impact of combinatorial structure on congestion games. *Journal of the ACM (JACM)*, 55(6):1–22, 2008.

[2] Naman Agarwal, Guy Kindler, Alexandra Kolla, and Luca Trevisan. Unique games on the hypercube. *Chicago J. Theor. Comput. Sci.*, 2015, 2015.

[3] Rudolf Ahlswede and Peter Gács. Spreading of sets in product spaces and hypercontraction of the markov operator. *The annals of probability*, pages 925–939, 1976.

[4] Amirmahdi Ahmadinejad, Sina Dehghani, Mohammadtaghi Hajiaghayi, Brendan Lucier, Hamid Mahini, and Saeed Seddighin. From duels to battlefields: Computing equilibria of blotto and other games. *Mathematics of Operations Research*, 44(4):1304–1325, 2019.

[5] Shigeki Aida and Daniel Stroock. Moment estimates derived from poincaré and logarithmic sobolev inequalities. *Mathematical Research Letters*, 1(1):75–86, 1994.

[6] Miklós Ajtai, János Komlós, and Endre Szemerédi. Deterministic simulation in logspace. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 132–140, 1987.

[7] Mikhail Alekhnovich, Sanjeev Arora, and Iannis Tourlakis. Towards strong non-approximability results in the lovász-schrijver hierarchy. In *Proceedings of the thirty-seventh annual ACM symposium on theory of computing*, pages 294–303, 2005.

[8] Vedat Alev. *Higher Order Random Walks, Local Spectral Expansion, and Applications.* PhD thesis, University of Waterloo, 2020.

[9] Vedat Levi Alev, Fernando Granha Jeronimo, and Madhur Tulsiani. Approximating constraint satisfaction problems on high-dimensional expanders. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 180–201. IEEE, 2019.

[10] Vedat Levi Alev, Fernando Jeronimo Jeronimo, Dylan Quintana, Shashank Srivastava, and Madhur Tulsiani. List decoding of direct sum codes. In *Proceedings of the*

Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 1412–1425. SIAM, 2020.

[11] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. *arXiv preprint arXiv:2001.02827*, 2020.

[12] Vedat Levi Alev and Ori Parzanchevski. Sequential sweeps and high dimensional expansion. *arXiv preprint arXiv:2312.02089*, 2023.

[13] Yeganeh Alimohammadi, Nima Anari, Kirankumar Shiragur, and Thuy-Duong Vuong. Fractionally log-concave and sector-stable polynomials: counting planar matchings and more. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 433–446, 2021.

[14] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[15] Noga Alon. Explicit expanders of every degree and size. *Combinatorica*, pages 1–17, 2021.

[16] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on information theory*, 38(2):509–516, 1992.

[17] Noga Alon and Michael Capalbo. Explicit unique-neighbor expanders. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pages 73–79. IEEE, 2002.

[18] Noga Alon and Fan RK Chung. Explicit construction of linear sized tolerant networks. *Discrete Mathematics*, 72(1-3):15–19, 1988.

[19] Noga Alon and Michael Luby. A linear time erasure-resilient code with nearly optimal recovery. *IEEE Transactions on Information Theory*, 42(6):1732–1736, 1996.

[20] Noga Alon and Vitali D Milman. $\lambda 1$, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.

[21] Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Noah Golowich, and Tuomas Sandholm. Near-optimal no-regret learning for correlated equilibria in multi-player general-sum games. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 736–749, 2022.

[22] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: optimal mixing of down-up random walks. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1418–1430, 2022.

[23] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Universality of spectral independence with applications to fast mixing in spin glasses. *arXiv preprint arXiv:2307.10466*, 2023.

[24] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. *SIAM Journal on Computing*, (0):FOCS20–1, 2021.

[25] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials ii: high-dimensional walks and an fpras for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 1–12, 2019.

[26] Alexandr Andoni and Daniel Beaglehole. Learning to hash robustly, guaranteed. *arXiv preprint arXiv:2108.05433*, 2021.

[27] Anurag Anshu, Nikolas P Breuckmann, and Chinmay Nirkhe. Nlts hamiltonians from good quantum codes. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1090–1096, 2023.

[28] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. *Journal of the ACM (JACM)*, 62(5):1–25, 2015.

[29] Sanjeev Arora, Subhash A Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, and Nisheeth K Vishnoi. Unique games on expanding constraint graphs are easy. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 21–28, 2008.

[30] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM (JACM)*, 45(3):501–555, 1998.

[31] Sanjeev Arora and Madhu Sudan. Improved low degree testing and its applications. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 485–495, El Paso, Texas, 1997.

[32] W Brian Arthur. Inductive reasoning and bounded rationality. *The American economic review*, 84(2):406–411, 1994.

[33] Srinivasan Arunachalam, Uma Girish, and Noam Lifshitz. One clean qubit suffices for quantum communication advantage. *arXiv preprint arXiv:2310.02406*, 2023.

[34] Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in online combinatorial optimization. *Mathematics of Operations Research*, 39(1):31–45, 2014.

[35] Robert J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1):67–96, 1974.

[36] Yakov Babichenko and Aviad Rubinstein. Settling the complexity of nash equilibrium in congestion games. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1426–1437, 2021.

[37] Mitali Bafna, Boaz Barak, Pravesh K. Kothari, Tselil Schramm, and David Steurer. Playing unique games on certified small-set expanders. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1629–1642. ACM, 2021.

[38] Mitali Bafna, Max Hopkins, Tali Kaufman, and Shachar Lovett. High dimensional expanders: Eigenstripping, pseudorandomness, and unique games. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1069–1128. SIAM, 2022.

[39] Mitali Bafna, Max Hopkins, Tali Kaufman, and Shachar Lovett. Hypercontractivity on high dimensional expanders. pages 185–194, 2022.

[40] Mitali Bafna, Noam Lifshitz, and Dor Minzer. Constant degree direct product testers with small soundness, 2024.

[41] Mitali Bafna and Dor Minzer. Characterizing direct product testing via coboundary expansion. *arXiv preprint arXiv:2308.09668*, 2023.

[42] Mitali Bafna and Dor Minzer. Solving unique games over globally hypercontractive graphs. *arXiv preprint arXiv:2304.07284*, 2023.

[43] Mourad Baïou and Francisco Barahona. Faster algorithms for security games on matroids. *Algorithmica*, 81(3):1232–1246, 2019.

[44] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326, 2012.

[45] Boaz Barak, Siu On Chan, and Pravesh K Kothari. Sum of squares lower bounds from pairwise independence. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 97–106, 2015.

[46] Boaz Barak, Parikshit Gopalan, Johan Håstad, Raghu Meka, Prasad Raghavendra, and David Steurer. Making the long code shorter. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 370–379. IEEE, 2012.

[47] Boaz Barak, Pravesh K Kothari, and David Steurer. Small-set expansion in shortcode graph and the 2-to-2 conjecture. *arXiv preprint arXiv:1804.08662*, 2018.

[48] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *2011 ieee 52nd annual symposium on foundations of computer science*, pages 472–481. IEEE, 2011.

[49] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *arXiv preprint arXiv:1404.5236*, 2014.

[50] Imre Bárány. A generalization of carathéodory's theorem. *Discrete Mathematics*, 40(2-3):141–152, 1982.

[51] Ya M Barzdin. On the realization of networks in three-dimensional space. *Selected Works of AN Kolmogorov: Volume III: Information Theory and the Theory of Algorithms*, pages 194–202, 1993.

[52] Daniel Beaglehole, Max Hopkins, Daniel Kane, Sihan Liu, and Shachar Lovett. Sampling equilibria: Fast no-regret learning in structured games. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3817–3855. SIAM, 2023.

[53] Paul Beame, Russell Impagliazzo, Jan Krajíček, Toniann Pitassi, and Pavel Pudlák. Lower bounds on hilbert's nullstellensatz and propositional proofs. *Proceedings of the London Mathematical Society*, 3(1):1–26, 1996.

[54] William Beckner. Inequalities in fourier analysis. *Annals of Mathematics*, 102(1):159–182, 1975.

[55] Soheil Behnezhad, Avrim Blum, Mahsa Derakhshan, MohammadTaghi Hajiaghayi, Christos H. Papadimitriou, and Saeed Seddighin. Optimal strategies of blotto games: Beyond convexity, 2019.

[56] Soheil Behnezhad, Sina Dehghani, Mahsa Derakhshan, MohammadTaghi HajiAghayi, and Saeed Seddighin. Faster and simpler algorithm for optimal strategies of blotto game. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, page 369–375, Feb 2017.

[57] Mihir Bellare, Shafi Goldwasser, Carsten Lund, and Alexander Russell. Efficient probabilistically checkable proofs and applications to approximations. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 294–304, 1993.

[58] Omri Ben-Eliezer, Max Hopkins, Chutong Yang, and Hantao Yu. Active learning polynomial threshold functions. *Advances in Neural Information Processing Systems*, 35:24199–24212, 2022.

[59] Michael Ben-Or and Nathan Linial. Collective coin flipping, robust voting schemes and minima of banzhaf values. In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, pages 408–416. IEEE, 1985.

[60] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 517–526, 1999.

[61] Siavosh Benabbas. *On hierarchies of strong SDP relaxations for combinatorial optimization problems*. University of Toronto (Canada), 2012.

[62] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. Sdp gaps from pairwise independence. *Theory of Computing*, 8(1):269–289, 2012.

[63] Ivona Bezáková, Daniel Štefankovic, Vijay V Vazirani, and Eric Vigoda. Approximating the permanent in o (n7) time. 2004.

[64] Robi Bhattacharjee, Max Hopkins, Akash Kumar, Hantao Yu, and Kamalika Chaudhuri. Robust empirical risk minimization with tolerance. In *International Conference on Algorithmic Learning Theory*, pages 182–203. PMLR, 2023.

[65] Guy Blanc and Dean Doron. New near-linear time decodable codes closer to the gv bound. In *37th Computational Complexity Conference (CCC 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[66] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of markov chains: Coupling, spectral independence, and entropy factorization. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3670–3692. SIAM, 2022.

[67] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proceedings of the twenty-second annual ACM symposium on Theory of computing*, pages 73–83, 1990.

[68] Sergey Bobkov and Michel Ledoux. Poincaré's inequalities and talagrand's concentration phenomenon for the exponential distribution. *Probability Theory and Related Fields*, 107:383–400, 1997.

[69] Andrej Bogdanov. Gap amplification fails below 1/2. *Comment on ECCC TR05-046, can be found at http://eccc. uni-trier. de/eccc-reports/2005/TR05-046/commt01. pdf*, 2005.

[70] Enric Boix-Adserà, Benjamin L. Edelman, and Siddhartha Jayanti. The multiplayer colonel blotto game. *Proceedings of the 21st ACM Conference on Economics and Computation*, 2020.

[71] Aline Bonami. Étude des coefficients de fourier des fonctions de $L^p(G)$. In *Annales de l'institut Fourier*, volume 20, pages 335–402, 1970.

[72] Emile Borel. The theory of play and integral equations with skew symmetric kernels. *Econometrica*, 21(1):97, 1953.

[73] Christer Borell. Positivity improving operators and hypercontractivity. *Mathematische Zeitschrift*, 180:225–234, 1982.

[74] Endre Boros and Zoltán Füredi. The number of triangles covering the center of an n-set. *Geometriae Dedicata*, 17:69–77, 1984.

[75] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities using the entropy method. *The Annals of Probability*, 31(3):1583–1614, 2003.

[76] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013.

[77] J Bourgain. Walsh subspaces of lp-product spaces. *Séminaire d'analyse fonctionnelle de l'école Polytechnique. Palaiseau*, 1979.

[78] Jean Bourgain, Jeff Kahn, Gil Kalai, Yitzhak Katznelson, and Nathan Linial. The influence of variables in product spaces. *Israel Journal of Mathematics*, 77(1):55–64, 1992.

[79] Petter Brändén, James Haglund, Mirkó Visontai, and David G Wagner. Proof of the monotone column permanent conjecture. In *Notions of Positivity and the Geometry of Polynomials*, pages 63–78. Springer, 2011.

[80] Nikolas P Breuckmann and Jens Niklas Eberhardt. Quantum low-density parity-check codes. *PRX Quantum*, 2(4):040101, 2021.

[81] E Oran Brigham. *The fast Fourier transform and its applications*. Prentice-Hall, Inc., 1988.

[82] Andries E Brouwer and Willem H Haemers. Distance-regular graphs. In *Spectra of graphs*, pages 177–185. Springer, 2012.

[83] Thomas Browning, Max Hopkins, and Zander Kelley. Doppelgangers: the ur-operation and posets of bounded height. *arXiv preprint arXiv:1710.10407*, 2017.

[84] Mark Bun, Marco Gaboardi, Max Hopkins, Russell Impagliazzo, Rex Lei, Toniann Pitassi, Satchit Sivakumar, and Jessica Sorrell. Stability is stable: Connections between replicability, privacy, and adaptive generalization. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 520–527, 2023.

[85] Sam Buss, Dima Grigoriev, Russell Impagliazzo, and Toniann Pitassi. Linear gaps between degrees for the polynomial calculus modulo distinct primes. *Journal of Computer and System Sciences*, 62(2):267–289, 2001.

[86] A Robert Calderbank and Peter W Shor. Good quantum error-correcting codes exist. *Physical Review A*, 54(2):1098, 1996.

[87] Ran Canetti, Guy Even, and Oded Goldreich. Lower bounds for sampling algorithms for estimating the average. *Information Processing Letters*, 53(1):17–25, 1995.

[88] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[89] Nicolo Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

[90] Siu On Chan. Approximation resistance from pairwise-independent subgroups. *Journal of the ACM (JACM)*, 63(3):1–32, 2016.

[91] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for sherali-adams relaxations. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 283–292, 2009.

[92] Zachary Chase, Bogdan Chornomaz, Shay Moran, and Amir Yehudayoff. Local borsuk-ulam, stability, and replicability. *arXiv preprint arXiv:2311.01599*, 2023.

[93] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 110–122. IEEE, 2022.

[94] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1548–1557. SIAM, 2021.

[95] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of glauber dynamics up to uniqueness via contraction. *arXiv preprint arXiv:2004.09083*, 2020.

[96] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of glauber dynamics: Entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1537–1550, 2021.

[97] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.

[98] Radha Chitta, Rong Jin, Timothy C Havens, and Anil K Jain. Approximate kernel k-means: Solution to large scale kernel clustering. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 895–903, 2011.

[99] Matthew Clegg, Jeffery Edmonds, and Russell Impagliazzo. Using the groebner basis algorithm to find proofs of unsatisfiability. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 174–183, 1996.

[100] Alon Cohen, Tamir Hazan, and Tomer Koren. Tight bounds for bandit combinatorial optimization. In *Conference on Learning Theory*, pages 629–642. PMLR, 2017.

[101] Charles J Colbourn, J Scott Provan, and Dirk Vertigan. The complexity of computing the tutte polynomial on transversal matroids. *Combinatorica*, 15(1):1–10, 1995.

[102] Mary Cryan, Heng Guo, and Giorgos Mousa. Modified log-sobolev inequalities for strongly log-concave distributions. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1358–1370. IEEE, 2019.

[103] Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. *Advances in Neural Information Processing Systems*, 34:27604–27616, 2021.

[104] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.

[105] Bart De Schuymer, Hans De Meyer, and Bernard De Baets. Optimal strategies for equal-sum dice games. *Discrete applied mathematics*, 154(18):2565–2576, 2006.

[106] Philippe Delsarte. Association schemes and t-designs in regular semilattices. *Journal of Combinatorial Theory, Series A*, 20(2):230–243, 1976.

[107] Persi Diaconis and Laurent Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996.

[108] Yotam Dikstein. New high dimensional expanders from covers. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 826–838, 2023.

[109] Yotam Dikstein and Irit Dinur. Agreement testing theorems on layered set systems. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1495–1524. IEEE, 2019.

[110] Yotam Dikstein and Irit Dinur. Agreement theorems for high dimensional expanders in the small soundness regime: the role of covers. *CoRR*, abs/2308.09582, 2023.

[111] Yotam Dikstein, Irit Dinur, Yuval Filmus, and Prahladh Harsha. Boolean function analysis on high-dimensional expanders. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[112] Yotam Dikstein, Irit Dinur, Prahladh Harsha, and Noga Ron-Zewi. Locally testable codes via high-dimensional expanders. *arXiv preprint arXiv:2005.01045*, 2020.

[113] Yotam Dikstein, Irit Dinur, and Alexander Lubotzky. Low acceptance agreement tests via bounded-degree symplectic hdxs. *arXiv preprint arXiv:2402.01078*, 2024.

[114] Yotam Dikstein and Max Hopkins. Chernoff bounds and reverse hypercontractivity on hdx. *arXiv preprint arXiv:2404.10961*, 2024.

[115] I. Dinur and D. Steurer. Direct product testing. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 188–196, 6 2014.

[116] Irit Dinur. The pcp theorem by gap amplification. *Journal of the ACM (JACM)*, 54(3):12–es, 2007.

[117] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 357–374. ACM, 2022.

[118] Irit Dinur, Yuval Filmus, and Prahladh Harsha. Analyzing boolean functions on the biased hypercube via higher-dimensional agreement tests: [extended abstract]. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2124–2133, 2019.

[119] Irit Dinur, Yuval Filmus, Prahladh Harsha, and Madhur Tulsiani. Explicit sos lower bounds from high-dimensional expanders. *arXiv preprint arXiv:2009.05218*, 2020.

[120] Irit Dinur and Elazar Goldenberg. Locally testing direct product in the low error range. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 613–622. IEEE, 2008.

[121] Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon Ta-Shma. List-decoding with double samplers. *SIAM Journal on Computing*, 50(2):301–349, 2021.

[122] Irit Dinur, Prahladh Harsha, Tali Kaufman, and Noga Ron-Zewi. From local to robust testing via agreement testing. *Theory of Computing*, 18(1):1–25, 2022.

[123] Irit Dinur, Min-Hsiu Hsieh, Ting-Chun Lin, and Thomas Vidick. Good quantum ldpc codes with linear time decoders. In *Proceedings of the 55th annual ACM symposium on theory of computing*, pages 905–918, 2023.

[124] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 974–985. IEEE, 2017.

[125] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. On non-optimally expanding sets in grassmann graphs. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 940–951, 2018.

[126] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. Towards a proof of the 2-to-1 games conjecture? In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 376–389, 2018.

[127] Irit Dinur, Siqi Liu, and Rachel Yun Zhang. New codes on high dimensional expanders. *arXiv preprint arXiv:2308.15563*, 2023.

[128] Irit Dinur and Inbal Livni Navon. Exponentially small soundness for the direct product z-test. In *32nd Computational Complexity Conference (CCC 2017)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2017.

[129] Irit Dinur and Or Meir. Derandomized parallel repetition via structured pcps. *computational complexity*, 20:207–327, 2011.

[130] Irit Dinur and Omer Reingold. Assignment testers: Towards a combinatorial proof of the pcp theorem. *SIAM Journal on Computing*, 36(4):975–1024, 2006.

[131] Irit Dinur, Madhu Sudan, and Avi Wigderson. Robust local testability of tensor products of ldpc codes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 304–315. Springer, 2006.

[132] Dean Doron and Mary Wootters. High-probability list-recovery, and applications to heavy hitters. In *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[133] Devdatt P Dubhashi and Desh Ranjan. Balls and bins: A study in negative dependence. *BRICS Report Series*, 3(25), 1996.

[134] Martin Dyer, Mark Jerrum, and Haiko Müller. On the switch markov chain for perfect matchings. *Journal of the ACM (JACM)*, 64(2):1–33, 2017.

[135] Roger Eckhardt. Stan ulam, john von neumann, and the monte carlo method. *Los Alamos Science*, 15(131-136):30, 1987.

[136] Lior Eldar and Aram W Harrow. Local hamiltonians whose ground states are hard to approximate. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 427–438. IEEE, 2017.

[137] David Ellis, Guy Kindler, and Noam Lifshitz. Forbidden intersection problems for families of linear maps. *arXiv preprint arXiv:2208.04674*, 2022.

[138] David Ellis, Guy Kindler, and Noam Lifshitz. An analogue of bonami's lemma for functions on spaces of linear maps, and 2-2 games. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 656–660, 2023.

[139] David Ellis, Guy Kindler, Noam Lifshitz, and Dor Minzer. Product mixing in compact lie groups. *arXiv preprint arXiv:2401.15456*, 2024.

[140] Shai Evra. Finite quotients of bruhat–tits buildings as geometric expanders. *Journal of Topology and Analysis*, 9(01):51–66, 2017.

[141] Shai Evra and Tali Kaufman. Bounded degree cosystolic expanders of every dimension. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 36–48, 2016.

[142] Shai Evra, Tali Kaufman, and Gilles Zémor. Decodable quantum ldpc codes beyond the square root distance barrier using high dimensional expanders. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 218–227. IEEE, 2020.

[143] Shai Evra, Guy Kindler, and Noam Lifshitz. Polynomial bogolyubov for special linear groups via tensor rank. *arXiv preprint arXiv:2404.00641*, 2024.

[144] Gabriele Farina, Chung-Wei Lee, Haipeng Luo, and Christian Kroer. Kernelized multiplicative weights for 0/1-polyhedral games: Bridging the gap between learning in extensive-form and normal-form games. *arXiv preprint arXiv:2202.00237*, 2022.

[145] Uriel Feige, Guy Kindler, and Ryan O'Donnell. Understanding parallel repetition requires understanding foams. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 179–192. IEEE, 2007.

[146] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the boolean domain. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1558–1577. SIAM, 2021.

[147] Yuval Filmus, Guy Kindler, Noam Lifshitz, and Dor Minzer. Hypercontractivity on the symmetric group. *arXiv preprint arXiv:2009.05503*, 2020.

[148] Yuval Filmus, Ryan O'Donnell, and Xinyu Wu. A log-sobolev inequality for the multislice, with applications. *arXiv preprint arXiv:1809.03546*, 2018.

[149] Ernst Fischer. Über quadratische formen mit reellen koeffizienten. *Monatshefte für Mathematik und Physik*, 16(1):234–249, 1905.

[150] Noah Fleming, Pravesh Kothari, and Toniann Pitassi. *Semialgebraic proofs and efficient algorithm design*. Foundations and Trends in Theoretical Computer Science, 2019.

[151] Noah Fleming, Pravesh Kothari, and Toniann Pitassi. Semialgebraic proofs and efficient algorithm design. *Foundations and Trends® in Theoretical Computer Science*, 14(1-2):1–221, 2019.

[152] G David Forney. Concatenated codes. 1965.

[153] Jacob Fox, Mikhail Gromov, Vincent Lafforgue, Assaf Naor, and János Pach. Overlap properties of geometric expanders. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 2012(671):49–83, 2012.

[154] Peter Frankl and Vojtěch Rödl. Forbidden intersections. *Transactions of the American Mathematical Society*, 300(1):259–286, 1987.

[155] Yoav Freund and Robert E. Schapire. Game theory, on-line prediction and boosting. *Proceedings of the ninth annual conference on Computational learning theory - COLT '96*, 1996.

[156] Yoav Freund and Robert E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1-2):79–103, 1999.

[157] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[158] Ehud Friedgut and Jean Bourgain. Sharp thresholds of graph properties, and the k-sat problem. *Journal of the American mathematical Society*, 12(4):1017–1054, 1999.

[159] Ehud Friedgut and Yonatan Iluz. Hyper-regular graphs and high dimensional expanders, 2020.

[160] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American mathematical Society*, 124(10):2993–3002, 1996.

[161] Ehud Friedgut, Gil Kalai, and Assaf Naor. Boolean functions whose fourier transform is concentrated on the first two levels. *Advances in Applied Mathematics*, 29(3):427–437, 2002.

[162] Jason Gaitonde, Max Hopkins, Tali Kaufman, Shachar Lovett, and Ruizhe Zhang. Eigenstripping, spectral decay, and edge-expansion on posets. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[163] Howard Garland. p-adic curvature and the cohomology of discrete subgroups of p-adic groups. *Annals of Mathematics*, pages 375–423, 1973.

[164] Konstantinos Georgiou, Avner Magen, Toniann Pitassi, and Iannis Tourlakis. Integrality gaps of 2-o(1) for vertex cover sdps in the lovász–schrijver hierarchy. *SIAM Journal on Computing*, 39(8):3553–3570, 2010.

[165] David Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

[166] Elazar Goldenberg and Karthik CS. Toward a general direct product testing theorem. *ACM Transactions on Computation Theory (TOCT)*, 12(1):1–18, 2019.

[167] Oded Goldreich. A sample of samplers: A computational perspective on sampling. *def*, 1:2n, 1997.

[168] Oded Goldreich and Shmuel Safra. A combinatorial consistency lemma with application to proving the PCP theorem. In *RANDOM: International Workshop on Randomization and Approximation Techniques in Computer Science*. LNCS, 1997.

[169] Russell Golman and Scott E. Page. General blotto: Games of allocative strategic mismatch. *Public Choice*, 138(3/4):279–299, 2009.

[170] Louis Golowich. Improved product-based high-dimensional expanders. *arXiv preprint arXiv:2105.09358*, 2021.

[171] Louis Golowich. From grassmannian to simplicial high-dimensional expanders. *arXiv preprint arXiv:2305.02512*, 2023.

[172] Louis Golowich. Notes on generalizing product-based high-dimensional expanders. Personal communication, May 2024.

[173] Louis Golowich and Tali Kaufman. Nlts hamiltonians and strongly-explicit sos lower bounds from low-rate quantum ldpc codes. *arXiv preprint arXiv:2311.09503*, 2023.

[174] Parikshit Gopalan, Salil Vadhan, and Yuan Zhou. Locally testable codes and cayley graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 81–92, 2014.

[175] Roy Gotlib and Tali Kaufman. Beyond the worst case: Structured convergence of high dimensional random walks. *CoRR*, 2022.

[176] Roy Gotlib and Tali Kaufman. Fine grained analysis of high dimensional random walks. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2023)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2023.

[177] Roy Gotlib and Tali Kaufman. List agreement expansion from coboundary expansion. In Yael Tauman Kalai, editor, *14th Innovations in Theoretical Computer Science Conference, ITCS 2023*, Leibniz International Proceedings in Informatics, LIPIcs. Schloss Dagstuhl- Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing, January 2023.

[178] Dima Grigoriev. Tseitin's tautologies and lower bounds for nullstellensatz proofs. In *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No. 98CB36280)*, pages 648–652. IEEE, 1998.

[179] Dima Grigoriev. Complexity of positivstellensatz proofs for the knapsack. *computational complexity*, 10(2):139–154, 2001.

[180] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1-2):613–622, 2001.

[181] Mikhael Gromov. Filling riemannian manifolds. *Journal of Differential Geometry*, 18(1):1–147, 1983.

[182] Mikhail Gromov. Singularities, expanders and topology of maps. part 2: From combinatorics to topology via algebraic isoperimetry. *Geometric and Functional Analysis*, 20(2):416–526, 2010.

[183] Leonard Gross. Logarithmic sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

[184] Oliver Gross. *The Symmetric Blotto Game [1]*. Rand Corporation, 1950.

[185] Oliver Gross and Robert Wagner. A continuous colonel blotto game. Technical report, Rand Project Air Force Santa Monica Ca, 1950.

[186] Heng Guo and Giorgos Mousa. Local-to-global contraction in simplicial complexes. *arXiv preprint arXiv:2012.14317*, 2020.

[187] Tom Gur, Noam Lifshitz, and Siqi Liu. Hypercontractivity on high dimensional expanders. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 176–184, 2022.

[188] Venkatesan Guruswami, Rajsekar Manokaran, and Prasad Raghavendra. Beating the random ordering is hard: Inapproximability of maximum acyclic subgraph. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 573–582. IEEE, 2008.

[189] Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with psd objectives. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 482–491. IEEE, 2011.

[190] Prahladh Harsha and Ramprasad Saptharishi. A note on the elementary construction of high-dimensional expanders of kaufman and oppenheim. *arXiv preprint arXiv:1912.11225*, 2019.

[191] Sergiu Hart. Discrete colonel blotto and general lotto games. *International Journal of Game Theory*, 36(3):441–460, 2008.

[192] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM (JACM)*, 48(4):798–859, 2001.

[193] Matthew B Hastings, Jeongwan Haah, and Ryan O'Donnell. Fiber bundle codes: breaking the n 1/2 polylog (n) barrier for quantum ldpc codes. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1276–1288, 2021.

[194] Hamed Hatami. A structure theorem for boolean functions with small total influences. *Annals of Mathematics*, pages 509–533, 2012.

[195] Alexander D Healy. Randomness-efficient sampling within nc. *Computational Complexity*, 17:3–37, 2008.

[196] David P Helmbold and Manfred K Warmuth. Learning permutations with exponential weights. *Journal of Machine Learning Research*, 10(7), 2009.

[197] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.

[198] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[199] Max Hopkins. Lecture 13: Hypercontractivity and Small-Set Expansion. https://www.wisdom.weizmann.ac.il/~dinuri/courses/22-HDX/L13.pdf, 2023.

[200] Max Hopkins. Hypercontractivity on hdx ii: Symmetrization and q-norms. 2024.

[201] Max Hopkins, Russell Impagliazzo, Daniel Kane, Sihan Liu, and Christopher Ye. Replicability in high dimensional statistics. *arXiv preprint arXiv:2406.02628*, 2024.

[202] Max Hopkins, Daniel Kane, and Shachar Lovett. The power of comparisons for actively learning linear classifiers. *Advances in Neural Information Processing Systems*, 33:6342–6353, 2020.

[203] Max Hopkins, Daniel Kane, Shachar Lovett, and Gaurav Mahajan. Noise-tolerant, reliable active classification with comparison queries. In *Conference on Learning Theory*, pages 1957–2006. PMLR, 2020.

[204] Max Hopkins, Daniel Kane, Shachar Lovett, and Gaurav Mahajan. Point location and active learning: Learning halfspaces almost optimally. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1034–1044. IEEE, 2020.

[205] Max Hopkins, Daniel Kane, Shachar Lovett, and Michal Moshkovitz. Bounded memory active learning through enriched queries. In *Conference on Learning Theory*, pages 2358–2387. PMLR, 2021.

[206] Max Hopkins, Daniel M Kane, Shachar Lovett, and Gaurav Mahajan. Realizable learning is all you need. In *Conference on Learning Theory*, pages 3015–3069. PMLR, 2022.

[207] Max Hopkins, Daniel M Kane, Shachar Lovett, and Gaurav Mahajan. Do pac-learners learn the marginal distribution? *arXiv preprint arXiv:2302.06285*, 2023.

[208] Max Hopkins and Ting-Chun Lin. Explicit lower bounds against ω (n)-rounds of sum-of-squares. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 662–673. IEEE, 2022.

[209] Samuel B Hopkins, Tselil Schramm, and Luca Trevisan. Subexponential lps approximate max-cut. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 943–953. IEEE, 2020.

[210] Roger A Horn, Noah H Rhee, and So Wasin. Eigenvalue inequalities and equalities. *Linear Algebra and its Applications*, 270(1-3):29–44, 1998.

[211] Nicole Immorlica, Adam Tauman Kalai, Brendan Lucier, Ankur Moitra, Andrew Postlewaite, and Moshe Tennenholtz. Dueling algorithms. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 215–224, 2011.

[212] Russell Impagliazzo, Ragesh Jaiswal, and Valentine Kabanets. Approximate list-decoding of direct product codes and uniform hardness amplification. *SIAM Journal on Computing*, 39(2):564–605, 2009.

[213] Russell Impagliazzo, Ragesh Jaiswal, Valentine Kabanets, and Avi Wigderson. Uniform direct product theorems: simplified, optimized, and derandomized. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 579–588, 2008.

[214] Russell Impagliazzo and Valentine Kabanets. Constructive proofs of concentration bounds. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 617–631. Springer, 2010.

[215] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New direct-product testers and 2-query pcps. *SIAM Journal on Computing*, 41(6):1722–1768, 2012.

[216] Insider. scientists say they can predict two-party outcomes after solving a 95-year-old game theory problem., 2016.

[217] Vishesh Jain, Marcus Michelen, Huy Tuan Pham, and Thuy-Duong Vuong. Optimal mixing of the down-up walk on independent sets of a given size. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1665–1681. IEEE, 2023.

[218] Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. Spectral independence, coupling with the stationary distribution, and the spectral gap of the glauber dynamics. *arXiv preprint arXiv:2105.01201*, 2021.

[219] Fernando Granha Jeronimo, Tushant Mittal, Ryan O'Donnell, Pedro Paredes, and Madhur Tulsiani. Explicit abelian lifts and quantum ldpc codes. *arXiv preprint arXiv:2112.01647*, 2021.

[220] Fernando Granha Jeronimo, Dylan Quintana, Shashank Srivastava, and Madhur Tulsiani. Unique decoding of explicit epsilon-balanced codes near the gilbert-varshamov bound. *arXiv preprint arXiv:2011.05500*, 2020.

[221] Fernando Granha Jeronimo, Shashank Srivastava, and Madhur Tulsiani. Near-linear time decoding of ta-shma's codes via splittable regularity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1527–1536, 2021.

[222] Fernando Granha Jeronimo, Shashank Srivastava, and Madhur Tulsiani. List decoding of tanner and expander amplified codes from distance certificates. *arXiv preprint arXiv:2311.02263*, 2023.

[223] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM (JACM)*, 51(4):671–697, 2004.

[224] Mark R Jerrum, Leslie G Valiant, and Vijay V Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical computer science*, 43:169–188, 1986.

[225] Jean-Pierre Kahane. *Some random series of functions*, volume 5. Cambridge University Press, 1985.

[226] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 68–80, Los Alamitos, CA, USA, oct 1988. IEEE Computer Society.

[227] Daniel M Kane and Raghu Meka. A prg for lipschitz functions of polynomials with applications to sparsest cut. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 1–10, 2013.

[228] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *Journal of the ACM (JACM)*, 51(3):497–515, 2004.

[229] Manuel Kauers, Ryan O'Donnell, Li-Yang Tan, and Yuan Zhou. Hypercontractive inequalities via sos, and the frankl–rödl graph. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 1644–1658. SIAM, 2014.

[230] Tali Kaufman, David Kazhdan, and Alexander Lubotzky. Ramanujan complexes and bounded degree topological expanders. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 484–493. IEEE, 2014.

[231] Tali Kaufman, David Kazhdan, and Alexander Lubotzky. Isoperimetric inequalities for ramanujan complexes and topological expanders. *Geometric and Functional Analysis*, 26(1):250–287, 2016.

[232] Tali Kaufman, Rasmus Kyng, and Federico Soldá. Scalar and matrix chernoff bounds from ℓ_∞-independence. *arXiv preprint arXiv:2111.02177*, 2021.

[233] Tali Kaufman and Alexander Lubotzky. High dimensional expanders and property testing. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 501–506, 2014.

[234] Tali Kaufman and David Mass. High dimensional combinatorial random walks and colorful expansion. *arXiv preprint arXiv:1604.02947*, 2016.

[235] Tali Kaufman and David Mass. Cosystolic expanders over any abelian group. In *Electron. Colloquium Comput. Complex.*, volume 25, page 134, 2018.

[236] Tali Kaufman and David Mass. Local-to-global agreement expansion via the variance method. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

[237] Tali Kaufman and David Mass. Unique-neighbor-like expansion and group-independent cosystolic expansion. In *32nd International Symposium on Algorithms and Computation (ISAAC 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[238] Tali Kaufman and Izhar Oppenheim. Construction of new local spectral high dimensional expanders. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 773–786, 2018.

[239] Tali Kaufman and Izhar Oppenheim. High order random walks: Beyond spectral gap. *Combinatorica*, pages 1–37, 2020.

[240] Tali Kaufman and Izhar Oppenheim. Coboundary and cosystolic expansion from strong symmetry. In *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, volume 198 of *LIPIcs*, pages 84:1–84:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[241] Tali Kaufman and Izhar Oppenheim. High dimensional expansion implies amplified local testability. *CoRR*, 2021.

[242] Tali Kaufman and Ella Sharakanski. Chernoff bound for high-dimensional expanders. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2020.

[243] Tali Kaufman and Ran J Tessler. Local to global high dimensional expansion and garland's method for general posets. *arXiv preprint arXiv:2101.12621*, 2021.

[244] Tali Kaufman and Ran J. Tessler. New cosystolic expanders from tensors imply explicit quantum LDPC codes with $\Omega(\sqrt{n}\log^k n)$ distance. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1317–1329, 2021.

[245] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Hypercontractivity for global functions and sharp thresholds. *arXiv preprint arXiv:1906.05568*, 2019.

[246] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Forbidden intersections for codes. *arXiv preprint arXiv:2103.05050*, 2021.

[247] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Global hypercontractivity and its applications. *arXiv preprint arXiv:2103.04604*, 2021.

[248] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Forbidden intersections for codes. *Journal of the London Mathematical Society*, 108(5):2037–2083, 2023.

[249] Nathan Keller. A tight quantitative version of arrow's impossibility theorem. *Journal of the European Mathematical Society*, 14(5):1331–1355, 2012.

[250] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775, 2002.

[251] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM Journal on Computing*, 37(1):319–357, 2007.

[252] Subhash Khot, Dor Minzer, Dana Moshkovitz, and Muli Safra. Small set expansion in the johnson graph. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 25, page 78, 2018.

[253] Subhash Khot, Dor Minzer, and Muli Safra. On independent sets, 2-to-2 games, and grassmann graphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 576–589, 2017.

[254] Subhash Khot, Dor Minzer, and Muli Safra. On independent sets, 2-to-2 games, and grassmann graphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 576–589, 2017.

[255] Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 592–601. IEEE, 2018.

[256] Subhash Khot and Dana Moshkovitz. Candidate hard unique game. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 63–76, 2016.

[257] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within 2- ε. *Journal of Computer and System Sciences*, 74(3):335–349, 2008.

[258] Subhash Khot, Madhur Tulsiani, and Pratik Worah. A characterization of strong approximation resistance. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 634–643, 2014.

[259] Subhash Khot and Nisheeth K Vishnoi. On the unique games conjecture. In *FOCS*, volume 5, page 3. Citeseer, 2005.

[260] Subhash A Khot and Nisheeth K Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative-type metrics into l1. *Journal of the ACM (JACM)*, 62(1):1–39, 2015.

[261] Philip Klein and Neal Young. On the number of iterations for dantzig-wolfe optimization and packing-covering approximation algorithms. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 320–327. Springer, 1999.

[262] Robert Kleinberg, Georgios Piliouras, and Éva Tardos. Multiplicative updates outperform generic no-regret learning in congestion games. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 533–542, 2009.

[263] Robert David Kleinberg. *Online decision problems with large strategy sets*. PhD thesis, Massachusetts Institute of Technology, 2005.

[264] Adam Klivans, Pravesh K Kothari, and Raghu Meka. Efficient algorithms for outlier-robust regression. In *Conference On Learning Theory*, pages 1420–1430. PMLR, 2018.

[265] Alexandra Kolla. Spectral algorithms for unique games. *computational complexity*, 20(2):177–206, 2011.

[266] Alexandra Kolla, Konstantin Makarychev, and Yury Makarychev. How to play unique games against a semi-random adversary: Study of semi-random models of unique games. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 443–452, 2011.

[267] Wouter M Koolen, Manfred K Warmuth, Jyrki Kivinen, et al. Hedging structured concepts. In *COLT*, pages 93–105. Citeseer, 2010.

[268] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally correctable and locally testable codes with sub-polynomial query complexity. *Journal of the ACM (JACM)*, 64(2):1–42, 2017.

[269] Pravesh K Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any csp. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 132–145, 2017.

[270] Dan Kovenock and Brian Roberson. Coalitional colonel blotto games with application to the economics of alliances. *Journal of Public Economic Theory*, 14(4):653–676, 2012.

[271] Dan Kovenock and Brian Roberson. Generalizations of the general lotto and colonel blotto games. *Economic Theory*, 71(3):997–1032, 2021.

[272] Dmitriy Kvasov. Contests with limited resources. *Journal of Economic Theory*, 136(1):738–748, 2007.

[273] Rasmus Kyng and Zhao Song. A matrix chernoff bound for strongly rayleigh distributions and spectral sparsifiers from a few random spanning trees. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 373–384. IEEE, 2018.

[274] Jean-François Laslier. How two-party competition treats minorities. *Review of Economic Design*, 7(3):297–307, 2002.

[275] Jean-François Laslier and Nathalie Picard. Distributive politics and electoral competition. *Journal of Economic Theory*, 103(1):106–130, 2002.

[276] Holden Lee. Parallelising glauber dynamics. *arXiv preprint arXiv:2307.07131*, 2023.

[277] Anthony Leverrier and Gilles Zémor. Quantum tanner codes. *arXiv preprint arXiv:2202.13641*, 2022.

[278] Anthony Leverrier and Gilles Zémor. Decoding quantum tanner codes. *IEEE Transactions on Information Theory*, 2023.

[279] Anthony Leverrier and Gilles Zémor. Efficient decoding up to a constant fraction of the code length for asymptotically good quantum codes. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1216–1244. SIAM, 2023.

[280] Noam Lifshitz and Dor Minzer. Noise sensitivity on the p-biased hypercube. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1205–1226. IEEE, 2019.

[281] Ting-Chun Lin and Min-Hsiu Hsieh. c3-local testable codes from lossless expanders. *arXiv preprint arXiv:2201.11369*, 2022.

[282] Ting-Chun Lin and Min-Hsiu Hsieh. Good quantum ldpc codes with linear time decoder from lossless expanders. *arXiv preprint arXiv:2203.03581*, 2022.

[283] Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. *Combinatorica*, 26(4):475–487, 2006.

[284] Richard J Lipton and Neal E Young. Simple strategies for large zero-sum games with applications to complexity theory. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 734–740, 1994.

[285] Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and computation*, 108(2):212–261, 1994.

[286] Kuikui Liu. From coupling to spectral independence and blackbox comparison with the down-up walk. *arXiv preprint arXiv:2103.11609*, 2021.

[287] Siqi Liu, Sidhanth Mohanty, and Elizabeth Yang. High-dimensional expanders from expanders. *arXiv preprint arXiv:1907.10771*, 2019.

[288] Alexander Lubotzky. High dimensional expanders. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 705–730. World Scientific, 2018.

[289] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[290] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. Explicit constructions of ramanujan complexes of type ad. *European Journal of Combinatorics*, 26(6):965–993, 2005.

[291] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. Ramanujan complexes of type $\tilde{A}_d$. *Israel J. Math.*, 149(1):267–299, 2005.

[292] Scott T Macdonell and Nick Mastronardi. Waging simple wars: a complete characterization of two-battlefield blotto equilibria. *Economic Theory*, 58(1):183–216, 2015.

[293] Konstantin Makarychev and Yury Makarychev. How to play unique games on expanders. In *International Workshop on Approximation and Online Algorithms*, pages 190–200. Springer, 2010.

[294] Claire Mathieu and Alistair Sinclair. Sherali-adams relaxations of the matching polytope. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 293–302, 2009.

[295] Robert McEliece, Eugene Rodemich, Howard Rumsey, and Lloyd Welch. New upper bounds on the rate of a code via the delsarte-macwilliams inequalities. *IEEE transactions on Information Theory*, 23(2):157–166, 1977.

[296] Mewright. Umd-led team first to solve well-known game theory scenario, Feb 2016.

[297] Dor Minzer and Kai Zhe Zheng. Near optimal alphabet-soundness tradeoff pcps. *arXiv preprint arXiv:2404.07441*, 2024.

[298] Moshe Morgenstern. Existence and explicit constructions of q+ 1 regular ramanujan graphs for every prime power q. *Journal of Combinatorial Theory, Series B*, 62(1):44–62, 1994.

[299] Dana Moshkovitz. Low-degree test with polynomially small error. *computational complexity*, 26:531–582, 2017.

[300] Dana Moshkovitz and Ran Raz. Two-query pcp with subconstant error. *Journal of the ACM (JACM)*, 57(5):1–29, 2008.

[301] Elchanan Mossel. A quantitative arrow theorem. *Probability Theory and Related Fields*, 154(1-2):49–88, 2012.

[302] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 21–30. IEEE, 2005.

[303] Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey E Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006.

[304] Elchanan Mossel, Krzysztof Oleszkiewicz, and Arnab Sen. On reverse hypercontractivity. *Geometric and Functional Analysis*, 23(3):1062–1097, 2013.

[305] Elchanan Mossel and Miklós Z Rácz. A quantitative gibbard-satterthwaite theorem without neutrality. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 1041–1060, 2012.

[306] John Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.

[307] Alon Nilli. On the second eigenvalue of a graph. *Discrete Mathematics*, 91(2):207–210, 1991.

[308] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[309] Izhar Oppenheim. Local spectral expansion approach to high dimensional expanders part i: Descent of spectral gaps. *Discrete & Computational Geometry*, 59(2):293–330, 2018.

[310] Georg Ostrovski and Sebastian van Strien. Payoff performance of fictitious play. *arXiv preprint arXiv:1308.4049*, 2013.

[311] Shayan Oveis Gharan and Luca Trevisan. A new regularity lemma and faster approximation algorithms for low threshold rank graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 303–316. Springer, 2013.

[312] János Pach. A tverberg-type result on multicolored simplices. *Computational Geometry*, 10(2):71–76, 1998.

[313] Alessandro Panconesi and Aravind Srinivasan. Randomized distributed edge coloring via an extension of the chernoff–hoeffding bounds. *SIAM Journal on Computing*, 26(2):350–368, 1997.

[314] Pavel Panteleev and Gleb Kalachev. Quantum ldpc codes with almost linear minimum distance. *IEEE Transactions on Information Theory*, 68(1):213–229, 2021.

[315] Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. pages 375–388, 2022.

[316] Ori Parzanchevski, Ron Rosenthal, and Ran J Tessler. Isoperimetric inequalities in simplicial complexes. *Combinatorica*, 36(2):195–227, 2016.

[317] Daniel Paulin. The convex distance inequality for dependent random variables, with applications to the stochastic travelling salesman and other problems. *Electronic Journal of Probability*, 19:1–34, 2014.

[318] Sriram V Pemmaraju. Equitable coloring extends chernoff-hoeffding bounds. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 285–296. Springer, 2001.

[319] Vianney Perchet, Philippe Rigollet, and Thibaut Le Gouic. An algorithmic solution to the blotto game using multi-marginal couplings, 2022.

[320] Mark S Pinsker. On the complexity of a concentrator. In *7th International Telegraffic Conference*, volume 4, pages 1–318. Citeseer, 1973.

[321] Ali Pourmiri and Thomas Sauerwald. Cutoff phenomenon for random walks on kneser graphs. *Discrete Applied Mathematics*, 176:100–106, 2014.

[322] Kevin Pratt. Personal communication, March 2022.

[323] Prasad Raghavendra. Optimal algorithms and inapproximability results for every csp? In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 245–254, 2008.

[324] Prasad Raghavendra, Tselil Schramm, and David Steurer. *High-dimensional estimation via sum-of-squares proofs*, pages 3389–3423. World Scientific, 2018.

[325] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 755–764, 2010.

[326] Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. *Advances in Neural Information Processing Systems*, 26, 2013.

[327] Ran Raz. A parallel repetition theorem. In *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, pages 447–456, 1995.

[328] Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability pcp characterization of np. In *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*, STOC '97, pages 475–484, New York, NY, USA, 1997. ACM.

[329] Omer Reingold. Undirected connectivity in log-space. *Journal of the ACM (JACM)*, 55(4):1–24, 2008.

[330] Marcel Riesz. Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires. *Acta mathematica*, 49(3):465–497, 1927.

[331] Brian Roberson. The colonel blotto game. *Economic Theory*, 29(1):1–24, 2006.

[332] Robert W Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.

[333] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.

[334] Aviad Rubinstein. Settling the complexity of computing approximate two-player nash equilibria. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 258–265. IEEE, 2016.

[335] Justin Salez. A sharp log-sobolev inequality for the multislice. *arXiv preprint arXiv:2004.05833*, 2020.

[336] James C Schatzman. Accuracy of the discrete fourier transform and the fast fourier transform. *SIAM Journal on Scientific Computing*, 17(5):1150–1166, 1996.

[337] Jeanette P Schmidt and Alan Siegel. On aspects of university and performance for closed hashing. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 355–366, 1989.

[338] Jeanette P Schmidt and Alan Siegel. The analysis of closed hashing under limited randomness. In *Proceedings of the twenty-second annual ACM symposium on Theory of computing*, pages 224–234, 1990.

[339] Jeanette P Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff–hoeffding bounds for applications with limited independence. *SIAM Journal on Discrete Mathematics*, 8(2):223–250, 1995.

[340] Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 593–602. IEEE, 2008.

[341] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for lovász-schrijver lp relaxations of vertex cover and max cut. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 302–310, 2007.

[342] Galina Schwartz, Patrick Loiseau, and Shankar S Sastry. The heterogeneous colonel blotto game. In *2014 7th international conference on NETwork Games, COntrol and OPtimization (NetGCoop)*, pages 232–238. IEEE, 2014.

[343] Ronen Shaltiel. Recent developments in explicit constructions of extractors. *Current Trends in Theoretical Computer Science: The Challenge of the New Century Vol 1: Algorithms and Complexity Vol 2: Formal Models and Semantics*, pages 189–228, 2004.

[344] Jonah Sherman. Breaking the multicommodity flow barrier for $o(\sqrt{logn})$-approximations to sparsest cut. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 363–372. IEEE, 2009.

[345] Michael Sipser and Daniel A Spielman. Expander codes. *IEEE transactions on Information Theory*, 42(6):1710–1722, 1996.

[346] Richard P Stanley. Differential posets. *Journal of the American Mathematical Society*, 1(4):919–961, 1988.

[347] Andrew M Steane. Error correcting codes in quantum theory. *Physical Review Letters*, 77(5):793, 1996.

[348] Dávid Szeszlér. Security games on matroids. *Mathematical Programming*, 161(1):347–364, 2017.

[349] Michel Talagrand. On russo's approximate zero-one law. *The Annals of Probability*, pages 1576–1587, 1994.

[350] Milind Tambe. *Security and game theory: algorithms, deployed systems, lessons learned*. Cambridge university press, 2011.

[351] R Tanner. A recursive approach to low complexity codes. *IEEE Transactions on information theory*, 27(5):533–547, 1981.

[352] Caroline Thomas. N-dimensional blotto game with heterogeneous battlefield values. *Economic Theory*, 65(3):509–544, 2018.

[353] GO Thorin. An extension of a convexity theorem due to m. riesz. *Comm. sem. math. univ. Lund*, 4:1–5, 1939.

[354] Madhur Tulsiani. Csp gaps and reductions in the lasserre hierarchy. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 303–312, 2009.

[355] Madhur Tulsiani. On sos algorithms and lower bounds for constraint satisfaction problems over high-dimensional expanders. SoS+TCS Reading Group, 2021.

[356] Berthold Vöcking and R Aachen. Congestion games: Optimization in competition. In *ACiD*, pages 9–20, 2006.

[357] Dong Quan Vu and Patrick Loiseau. Colonel blotto games with favoritism: Competitions with pre-allocations and asymmetric effectiveness, 2021.

[358] Dong Quan Vu, Patrick Loiseau, and Alonso Silva. Efficient computation of approximate equilibria in discrete colonel blotto games. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 2018.

[359] Dong Quan Vu, Patrick Loiseau, and Alonso Silva. Combinatorial bandits for sequential learning in colonel blotto games. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 867–872. IEEE, 2019.

[360] Sebastian Wagner-Carena, Max Hopkins, Ana Diaz Rivero, and Cora Dvorkin. A novel cmb component separation method: hierarchical generalized morphological component analysis. *Monthly Notices of the Royal Astronomical Society*, 494(1):1507–1529, 2020.

[361] Manfred K Warmuth and Dima Kuzmin. Randomized online pca algorithms with regret bounds that are logarithmic in the dimension. *Journal of Machine Learning Research*, 9(Oct):2287–2320, 2008.

[362] Avi Wigderson. Randomness extractors–applications and constructions. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2009.

[363] Ruizhe Zhang. Personal communication, December 2020.

[364] Yu Zhao. *Generalizations and Applications of Hypercontractivity and Small-Set Expansion*. PhD thesis, Columbia University, 2021.

[365] Qing Zou. The q-binomial inverse formula and a recurrence relation for the q-catalan–qi numbers. *J. Math. Anal*, 8(1):176–182, 2017.