Limits of Efficient Algorithms in the Worst and Average Cases

By

Ansh Nagda

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Computer Science
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor Prasad Raghavendra, Chair
Professor Nikhil Srivastava
Professor Venkatesan Guruswami
Professor Avishay Tal

Spring 2026

Abstract

Limits of Efficient Algorithms in the Worst and Average Cases

by
Ansh Nagda

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Prasad Raghavendra, Chair

This dissertation studies the computational complexity of various problems as a function of a real-valued parameter controlling the hardness. The motivating question behind all these problems is to precisely understand the value of this parameter where the problem transitions from computationally easy to computationally hard. Pursuing this goal often necessitates developing a deep understanding of both algorithmic and hardness techniques. We study such problems in two different contexts: approximation algorithms and average-case algorithms.

In the first part of this dissertation, we study approximation algorithms for two problems: the permanent of positive semidefinite matrices and the EPR quantum Hamiltonian problem. Both these problems fall outside the scope of the standard hardness of approximation literature that targets NP problems. For both problems, we design improved approximation algorithms based on convex relaxations. For the permanent problem, we also prove the first exponential-factor hardness of approximation, showing that the easy-to-hard transition referenced above must occur at an exponentially large scale.

In the second part, we study two types of problems with random inputs: hypothesis testing and certification. For hypothesis testing where the null hypothesis is a simple product distribution, we design a new recipe to argue that the low-degree likelihood ratio obtains optimal advantage among all polynomial-time tests. We instantiate this recipe for the planted clique problem, proving optimal hardness results that are tight up to arbitrarily small inverse polynomial errors. For certification, we study upper certificates for MAX-CUT and MAX-IND-SET in random d-regular graphs. We specifically focus on the less understood setting of small d (say 3 or 4), and we use computational techniques to prove new lower and upper bounds that are numerically nearly tight.

Acknowledgments

First and foremost I want to acknowledge my advisor Prasad Raghavendra for his support and encouragement, and instilling in me a sense that I can work on anything that interests me. During the course of my PhD, he proposed a diversity of ambitious research directions for us to work (with the final one finding success; this has led to some work that I am perhaps most proud of). I would also like to thank Shirshendu Ganguly for serving on my Qualifying Exam committee, Nikhil Srivastava for serving on my Dissertation committee, and Venkatesan Guruswami and Avishay Tal for serving on both.

I had a large number of collaborators including Shayan Oveis Gharan, Farzam Ebrahimnejad, Nathan Ju, Shivam Nadimpalli, Joseph Slote, Abhradeep Thakurta, Prabhakar Raghavan, Alex Wein, Ilias Zadik, and Kostas Tsirkas. Working with these people not only taught me about many areas of theoretical computer science, but also taught me how to do research.

The theory group at Berkeley has been a wonderful place to grow, socialize, and make close friendships. Soda hall had a vibrant atmosphere where I would inevitable end up learning something new every day. Among others, I would like to thank Ishaq Aden-Ali, Omar Alrabiah, Louis Golowich, Lucas Gretta, Meghal Gupta, Christian Ikeokwu, Malvika Joshi, Nathan Ju, Siqi Liu, Xin Lyu, Sidhanth Mohanty, Angelos Pelecanos, Mihir Singhal, Jack Spilecki, Sriram Sridhar, and David Wu for creating this culture in Soda Hall.

I also spent a significant amount of time working at the Simons institute, and found it a great place to meet new researchers working on familiar and (then) unfamiliar topics. I am grateful to Anthony Chen, Fermi Ma, Shivam Nadimpalli, Joseph Slote, Ewin Tang for their friendships.

I'd like to thank Rajiv Gandhi for introducing me to theoretical computer science early in my life, and also for his continued support and advice throughout my academic journey.

Finally, I would like to thank my family for their support, and always encouraging me to follow my passion.

Contents

Part I

Introduction

Chapter 1

Overview

This dissertation studies limits of efficient algorithms through quantitative thresholds. Suppose we parametrize a computational problem using a real number t representing some way to scale the computational complexity of the problem. For example, t could represent an approximation factor, a runtime budget, or a bound on the probability of failure of an algorithm. In computer science we would like to classify the range of possible values of t as algorithmically easy and algorithmically hard.

Approximation Algorithms. The first part of the dissertation studies approximation algorithms for problems beyond the usual constraint satisfaction template. The first chapter in this part concerns the permanent of a positive semidefinite matrix [ENOG24]. If x is a complex Gaussian vector with covariance $A \succeq 0$, then Wick's formula gives

$$\mathbb{E}\left[\prod_i |x_i|^2\right] = \mathrm{per}(A).$$

Thus the problem is best viewed as *integration* rather than as counting over permutations. The technical narrative of that chapter is the relationship between integration and optimization. Wick's formula relates the permanent of A to an average of products of linear forms over the sphere, while the corresponding maximum is the $q \to 0$ case of a matrix $2 \to q$ norm. In low-rank regimes, the average and maximum are close up to subexponential loss, which allows hardness for product-of-linear-forms optimization to transfer to permanents. In the algorithmic direction, the chapter uses the optimal solution to the SDP relaxation underlying earlier approximation algorithms [AGGS17, YP22] to define a smooth proxy for rank, and then exploits this proxy to improve the known approximation factor.

The second approximation algorithms chapter studies the EPR Hamiltonian, a simplified quantum local Hamiltonian problem [JN25]. Here the objective is not to find a classical assignment, but an efficiently describable quantum state with provably large energy. The chapter gives an improved approximation algorithm by using a fractional matching relaxation and rounding its structure into entangled states. Our improved algorithm comes out of a combination of two ideas: first, a coherent interpolation between product states and specific

entangled states, and second, a way to circumvent the integrality gap of the aforementioned matching relaxation by directly rounding the optimal half-integral points to entangled states.

Average-case algorithms. The second part of the dissertation studies average-case algorithms. In a *certification* problem, the algorithm receives a typical random instance but must output a statement that is sound for every instance on which it accepts. In a *distinguishing* problem, the algorithm receives one sample from one of two known distributions and is measured by its advantage in identifying the source. Both notions are quantitative: the question is how strong a certificate, or how large an advantage, is available to polynomial-time algorithms.

The first average-case chapter concerns certification in sparse random regular graphs [NRT25]. For a graph G, let $\mathrm{MC}(G)$ be the fraction of edges in a maximum cut and let $\mathrm{IS}(G)$ be the fraction of vertices in a maximum independent set. The task is to certify upper bounds on these two quantities when G is drawn from the random d-regular graph model. The lower-bound side follows the *quiet-planting* framework of Kunisky and Yu: finite Ramanujan graphs with large cuts or independent sets serve as random-looking obstructions to stronger certification [KY24]. The chapter improves these constructions using AlphaEvolve to search for explicit, checkable Ramanujan witnesses [NVE+25]. The upper-bound side develops local linear programs over finite neighborhoods, using longer walk correlations to improve on the basic Hoffman-type certificate [Fri08, Hof03, Hae21]. For the small degrees studied here, these lower and upper bounds nearly meet.

Finally, the dissertation studies optimal distinguishing advantage for hypothesis testing. We ask the question "what is the optimal distinguishing advantage that can be achieved by any efficient test"? In the case that the null distribution is a simple product distribution, we hypothesize that for many problems, the optimal efficient test is obtained by thresholding the likelihood ratio to a large constant degree. We provide evidence for this hypothesis by studying the planted clique problem [NR25], where the null distribution is $G(n, 1/2)$, and the alternative distribution is obtained by adding a clique of size k. The *square-root threshold* is the scale $k = \Theta(\sqrt{n})$: known polynomial-time algorithms succeed above this scale, while for $k = n^{1/2-\alpha}$, $0 < \alpha < 1/2$, the problem is statistically distinguishable but conjectured to be computationally hard [Jer92, Kuc95, AKS98, KV02]. The Planted Clique Conjecture formalizes this belief, asserting that within this regime, there is no efficient distinguisher achieving advantage $1 - o(1)$. We prove that the Planted Clique Conjecture imposes a sharper version of the same statement: efficient distinguishers cannot beat the advantage achieved by the truncated likelihood ratio by a nontrivial amount. To demonstrate the robustness of the approach, we also construct planted distributions that are much harder than the standard one by perturbing the Planted Clique distribution to match low-degree moments with the null distribution.

Part II

Approximation Algorithms

Chapter 2

Overview of Approximation Algorithms

The classical theory of approximation algorithms is largely organized around NP-style optimization problems. In a typical example, such as a constraint satisfaction problem, an instance specifies finitely many variables together with local rewards, and the task is to output an assignment whose value is close to the optimum. The algorithmic side of the theory relaxes the space of assignments to a convex object, such as a linear or semidefinite program, and then rounds the relaxed solution back to an assignment. The hardness side builds reductions that preserve this local optimization structure, often through PCPs, dictatorship tests, and gadgets [ALM+98, Hås01, TSSW00, Kho02, KKMO07, Rag08].

The two chapters in this part sit somewhat outside that template. The problems considered here do not ask for a near-optimal Boolean or finite-alphabet assignment. The first is an integration problem: approximate the permanent of a positive semidefinite matrix, equivalently a structured high-dimensional Gaussian moment. The second is a quantum optimization problem: approximate the maximum energy of the EPR Hamiltonian, where the optimization variable is a quantum state in an exponentially large Hilbert space. In both settings, approximation still means producing a polynomial-time certificate whose value is provably close to the optimum quantity. The guarantees, however, are driven by mathematical structure outside the usual local predicate framework. For positive semidefinite permanents, the central analytic issue is to determine when a high-dimensional Gaussian moment can be compared to the maximum of a product of linear forms, and how rank and spectrum govern that comparison. For the EPR Hamiltonian, the central issue is entanglement: which quantum states can be efficiently constructed from a relaxation while retaining enough local energy? Chapter 3 studies the first problem, with supporting details in Chapter A; Chapter 4 studies the second, with the details completed in Chapter B.

Throughout this part, approximation guarantees are stated in a one-sided multiplicative form. For a nonnegative function F on instances, a ρ-approximation, with $\rho \geq 1$, returns an estimate $\widehat{F}(I)$ satisfying

$$\rho^{-1} F(I) \leq \widehat{F}(I) \leq F(I).$$

For a maximization problem, this is the same as producing a feasible solution whose value is

at least a $1/\rho$ fraction of the optimum.

2.1 Gaussian moments

The first setting is the permanent of a positive semidefinite matrix. This is a PSD analogue of the nonnegative-matrix permanent, for which Jerrum, Sinclair, and Vigoda gave an FPRAS [JSV04], but its approximation landscape is quite different [AGGS17, YP21, YP22, Mei23]. For a complex Gaussian vector x with covariance matrix $A \succeq 0$, one has

$$\mathbb{E}\left[\prod_i |x_i|^2\right] = \operatorname{per}(A).$$

Thus the problem is a question about a specific Gaussian moment, not about searching over permutations. If $A = VV^\dagger$, with rows $v_1, \dots, v_n \in \mathbb{C}^d$, then Wick's formula rewrites the permanent as a spherical average

$$\operatorname{per}(VV^\dagger) = c^{\mathbb{C}}_{n,d}\, \mathbb{E}_{\|x\|_2=1} \prod_i |\langle v_i, x\rangle|^2.$$

This representation is the bridge from integration to optimization: one can try to replace the average of the product of linear forms by its maximum

$$r(V) = \max_{\|x\|_2=1} \prod_i |\langle v_i, x\rangle|^2.$$

The replacement is meaningful only when the average and the maximum are controlled relative to one another.

This comparison is governed by rank. In low dimension, the spherical average of a degree-$2n$ product cannot be much smaller than its maximum: the loss is bounded by a factor depending on $\binom{n+d-1}{d}$. Thus in regimes where $\log\binom{n+d-1}{d} = o(n)$, the integration problem and the maximization problem are equivalent up to subexponential factors. In high rank, this equivalence can fail exponentially. The maximizer may sit in a small region of the sphere, while the integral is determined by the distribution of the product over typical directions. This is the sense in which positive semidefinite permanents differ from a direct optimization problem: one must understand when an extremal certificate for $r(V)$ is representative of the Gaussian moment, and when the averaging operation sees different structure.

Chapter 3 improves the known algorithmic approximation for this problem; the main guarantee is Theorem 3.1.1, and the proof is developed in Sections 3.1.1 and 3.3. The algorithm still begins from the semidefinite relaxation for $r(V)$, but its analysis does not treat the relaxation as a black-box substitute for the Gaussian integral. After solving the SDP, the instance is normalized so that $0 \preceq A \preceq I$. The trace of A, used as a smooth proxy for rank, then determines how much of the old integration-to-maximization comparison can be trusted. Low trace gives a stronger upper bound on the permanent itself, while high trace allows a more faithful rounding analysis through an interpolation between Gaussian rounding

and a Rademacher-type construction. The proof is therefore organized around the possible divergence between the integral and the associated maximization problem.

Chapter 3 also proves exponential hardness of approximation. The reduction proceeds through the problem of maximizing a product of linear forms,

$$r(V) = \max_{\|x\|_2=1} \prod_i |\langle v_i, x\rangle|^2,$$

which is the $2 \to 0$ case of a matrix $2 \to q$ "norm" problem. The chapter proves tight hardness for $2 \to q$ quantities in the range $-1 < q < 2$, including the nonconvex regime $q < 1$, and then transfers this hardness to positive semidefinite permanents using Wick's formula and a rank-deficient construction. The analytic issue is precisely the low-rank equivalence above: for the constructed matrices, the Gaussian moment and the maximum product of linear forms agree up to subexponential loss, so hardness for the maximization problem becomes hardness for the integration problem. Together, Theorems 3.1.1 and 3.1.2 make this comparison two-sided: the algorithmic result exploits where the integral deviates from the optimization surrogate, while the hardness result identifies a low-rank regime where the surrogate faithfully transfers hardness back to the permanent. The reduction and its analytic ingredients are given in Sections 3.1.1 and 3.4 and Chapter A.

2.2 Quantum Hamiltonians

The second setting is the EPR Hamiltonian, a simplified quantum local Hamiltonian model motivated by the approximation theory of Quantum Max Cut [GP19, AGM20, LP24, Kin23]. A local Hamiltonian is specified as a sum of local terms, but the feasible region is the set of density matrices on n qubits. Exact optimization is therefore an eigenvalue problem over a space of dimension 2^n, and a polynomial-time approximation algorithm cannot enumerate or write down an arbitrary feasible solution. It must produce an efficiently describable state whose energy is provably large. This is the quantum analogue of rounding, but the rounded object is a state preparation rather than a classical assignment.

The EPR Hamiltonian removes one obstruction present in Quantum Max Cut while leaving the entanglement question intact. Its local terms reward agreement with the EPR state

$$|\mathrm{EPR}\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}.$$

Unlike Quantum Max Cut, the terms are ferromagnetic; in particular, the optimal product state has a simple form. The remaining difficulty is to use entanglement without making the rounded state intractable or locally inconsistent. The algorithmic question is therefore not only which relaxation upper bounds the optimum, but also which entangled states can be assembled from the relaxed solution without losing the global approximation guarantee.

Chapter 4 gives an approximation algorithm with ratio

$$\frac{1+\sqrt{5}}{4} \approx 0.809.$$

The main guarantee is Theorem 4.1.1, and the proof is organized in Section 4.2. The relaxation is a fractional matching linear program, whose value upper bounds the optimum energy. The rounding algorithm interprets the structure of an optimal fractional matching as instructions for constructing a state. In bipartite instances, developed in Section 4.2.3, the improvement is to replace a classical random choice between a product state and an EPR pair by a coherent interpolation,

$$\sqrt{\theta}\,|00\rangle + \sqrt{1-\theta}\,|11\rangle\,,$$

so that the rounding keeps interference terms that a probabilistic mixture would discard. For general instances, Section 4.2.4 uses the half-integral structure of the matching polytope: the remaining obstruction is a collection of odd cycles, and these are rounded using compatible cycle states. The cycle-state construction needed for this last step is proved in Chapter B.

2.3 Conclusion

This part studies approximation problems that do not fit the usual model of rounding a relaxation to a finite assignment. In Chapter 3, the function to be approximated is a Gaussian moment, and the central question is when this integral can be compared to the maximum of a product of linear forms. In Chapter 4, the algorithm must produce an efficiently describable quantum state, and the central question is how a fractional matching relaxation can be rounded into useful entanglement.

Chapter 3

On Approximability of the Permanent of PSD Matrices

This chapter is based on "On Approximability of the Permanent of PSD Matrices" [ENOG24], which is joint work with Farzam Ebrahimnejad and Shayan Oveis Gharan.

3.1 Introduction

Given a matrix $A \in \mathbb{C}^{n\times n}$, the permanent of A is defined as

$$\mathrm{per}(A) = \sum_{\sigma\in\mathbb{S}_n} \prod_{i=1}^{n} A_{i,\sigma_i},$$

where the sum is over all permutations over n elements. It is well-known that the permanent of a matrix with non-negative entries can be approximated up to a $1+\epsilon$-multiplicative factor using the MCMC method [JSV04]. Recently there has been significant interest in studying permanent of Hermitian PSD matrices because of close connections to quantum optics and Boson sampling. A folklore algorithm is to simply take the product of the entries of the main diagonal to get an $n!$-approximation.

A few years ago, [AGGS17] obtained the first (deterministic) simply exponential approximation algorithms with approximation factor $e^{(\gamma+1)n}$. The algorithm proposed in [AGGS17] uses a basic SDP relaxation for the problem; many experts expected that perhaps by using higher-level SDP relaxations one can improve the approximation factor. Later on, several groups attempted to improve the approximation factor (see e.g., [Bar20]), but for the general case, only subexponential improvements to the approximation ratio were found [YP21, YP22]. Very recently, Meiburg showed that contrarily to the permanent of non-negative matrices, it is NP-Hard to approximate the permanent of a PSD matrix within a factor of $e^{-n^{1-\epsilon}}$ for any $\epsilon > 0$ [Mei23]. So, the MCMC method falls short of providing a $1+\epsilon$-approximation for PSD permanents.

It remained an open problem if, perhaps by using randomness or higher level SDP relaxations, one can obtain an $e^{-\epsilon n}$ approximation factor for ϵ arbitrarily small, or at the

very least whether the $\gamma + 1$ factor in the exponent can be improved to a smaller constant. We answer both these questions in our work.

Our first result is an exponential improvement on the $e^{-(\gamma+1)n}$ approximation algorithm mentioned above.

Theorem 3.1.1 (Main Algorithmic Result)**.** *There is a deterministic polynomial time $e^{-(\gamma+0.9999)n}$-approximation algorithm for the permanent of a Hermitian PSD matrix $A \in \mathbb{C}^{n\times n}$.*

Our second result is the first exponential hardness of approximation for this problem. As a corollary of a general hardness of approximation result we prove (see Theorem 3.1.5 below), we show the following:

Theorem 3.1.2 (Main Hardness Result)**.** *For all $\epsilon > 0$, it is* NP*-hard to approximate the permanent of a Hermitian PSD matrix $A \in \mathbb{C}^{n\times n}$ within a factor $e^{-(\gamma-\epsilon)n}$.*

In particular, the above theorem shows that assuming NP $\neq$ RP *even using randomness* the approximation factor of [AGGS17, YP21] cannot be improved by more than a factor of e^n.

Maximizing Product of Linear Forms Our hardness techniques also apply to an optimization problem that happens to be related to the permanent of PSD matrices, called the "maximizing product of linear forms" problem, studied by Yuan and Parrilo [YP22, YP21]: Given a matrix $V \in \mathbb{C}^{n\times d}$ with rows $v_1, \ldots, v_n \in \mathbb{C}^d$, define

$$r(V) := \max_{x\in\mathbb{C}^n:\|x\|_2=1} \prod_{i=1}^{n} |\langle x, v_i\rangle|^2. \tag{3.1}$$

They design a polynomial time $O(e^{-\gamma n\cdot(1-o(1))})$-approximation algorithm for $r(V)$ using semidefinite programming, where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. They also prove APX-hardness for this problem, and raise an open problem of finding the true approximability of this problem. Recently, Meiberg studied an equivalent problem under the name Approximate Quantum Maximum Likelihood Estimation, and showed NP-hardness of approximating it to within any constant factor [Mei23]. Our main technical hardness result, Theorem 3.1.5, immediately implies that the maximizing product of linear forms problem (and the Approximate Quantum Maximum Likelihood Estimation problem) does not admit a $e^{-\gamma n(1+\epsilon)}$-approximation for any constant $\epsilon > 0$, answering the question of Yuan and Parrilo up to sub-exponential factors in n.

3.1.1 Technical Contributions

Algorithmic results

In this part, we show the main ideas behind Theorem 3.1.1. We start by presenting algorithms used by previous work. Let $A = VV^\dagger$ be an $n \times n$ PSD matrix where $v_1, \ldots, v_n \in \mathbb{C}^n$ are the

rows of V. Previous work ([AGGS17, YP22]) showed that the value of the following SDP gives a $e^{-(\gamma+1)n}$ approximation to $\mathrm{per}(A)$:

$$\mathrm{SDP}(V) := \max_{X \succeq 0, \mathrm{tr}(X)=n} \prod_{i \in [n]} v_i^\dagger X v_i.$$

$\mathrm{SDP}(V)$ might seem completely unrelated to the definition of $\mathrm{per}(A)$, but we remark that their relationship is a lot more clear when $\mathrm{per}(A)$ is rewritten using Wick's formula (Lemma 3.2.16). Notice that the objective function of $\mathrm{SDP}(V)$ is log-concave, so it can be optimized in polynomial time. It turns out that upon solving $\mathrm{SDP}(V)$, we can reduce to the case that the maximizer X^* of $\mathrm{SDP}(V)$ satisfies $v_i^\dagger X^* v_i = 1$ for all $i \in [n]$ (see Eq. (3.6)). This property simplifies matters enough that we will assume it for the rest of this section.

The above property implies that $A \preceq I$ (see Claim 3.3.1), so we immediately get $\mathrm{per}(A) \leq 1$. Conversely, [AGGS17, YP22] prove that

$$\mathrm{per}(A) \geq \frac{n!}{n^n} \cdot r(V) \geq e^{-n} \cdot r(V). \tag{3.2}$$

Noticing that $\mathrm{SDP}(V)$ is a semidefinite relaxation of $r(V)$, a simple Gaussian rounding argument ([YP22, Lemma 4.3], Lemma A.3.1) can be used to show

$$r(V) \geq e^{-\gamma n} \mathrm{SDP}(V) = e^{-\gamma n}. \tag{3.3}$$

Putting these together, one gets

$$e^{-(\gamma+1)n} \leq \mathrm{per}(A) \leq 1, \tag{3.4}$$

giving a $e^{-(\gamma+1)n}$ approximation factor.

We remark that both sides of the above inequality can be tight, in particular the upper bound is tight for the identity matrix and the lower bound is tight for a certain family of low rank projection matrices (see [AGGS17]). So one may expect that no improvement is possible along this line.

In our approach we exploit the fact these inequalities are tight for matrices of very different rank – the known tight examples for the upper/lower bounds have very high/low rank respectively. In order to make this intuition concrete, we will use $\mathrm{tr}(A)$ as a smooth analogue of rank. Our main technical results are improvements to both sides of the preceding comparison.

Lemma 3.1.3 (Improved Upper Bound)**.** *Let $\epsilon \in [0,1]$. Any matrix $0 \preceq A \preceq I$ with $\mathrm{tr}(A) \leq (1-\epsilon)n$ satisfies*

$$\mathrm{per}(A) \leq \left(1 - \frac{\epsilon^2}{20}\right)^n.$$

Lemma 3.1.4 (Improved Lower Bound)**.** *Let $VV^\dagger = A \preceq I$, and assume that the maximizer X^* of $\mathrm{SDP}(V)$ satisfies $v_i^\dagger X^* v_i = 1$ for all i. For any $0 \leq \beta \leq 1$,*

$$\mathrm{per}(A) \geq e^{-n} \cdot r(V) \geq e^{-(\gamma+1)n} \cdot \exp\left(n \cdot \left(\ln(1-\beta) + \frac{\beta}{1-\beta} \cdot \frac{\mathrm{tr}(A)}{n} - \frac{0.273\beta^2}{(1-\beta)^2} \cdot \frac{n}{\mathrm{tr}(A)}\right)\right).$$

Our proof of Lemma 3.1.3 is inspired by an identity for $\mathrm{per}(A)$ appearing in [Bar20]. Our proof of Lemma 3.1.4 is based on an improved rounding procedure and is more technical, so we provide a proof overview in Section 3.1.2. We can now state our algorithm.

Algorithm: Given a PSD matrix $A = VV^\dagger$ where $v_1, \ldots, v_n$ are rows of V, first reduce to the case that the maximizer X^* of $\mathrm{SDP}(V)$ satisfies $v_i^\dagger X^* v_i = 1$ for all i (as described in Eq. (3.6)). Output $\left(1 - \frac{\epsilon^2}{20}\right)^n$, where ϵ is defined by $\mathrm{tr}(A) = (1-\epsilon)n$.

We will use this algorithm in our proof of Theorem 3.1.1, which is straightforward to analyze when equipped with Lemmas 3.1.3 and 3.1.4.

Hardness results

In this part we highlight the main technical contributions behind Theorem 3.1.2. Our proof broadly consists of two steps:

1. Show that $r(V)$ does not admit a $e^{-\gamma n(1+\epsilon)}$-approximation algorithm.
2. Give an approximation-preserving reduction from $r(V)$ to PSD permanents.

We start by elaborating on Item 1. In order to draw analogies to the existing hardness of approximation literature, we will first rephrase and generalize the optimization problem of $r(V)$. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ be a field. For a vector $x \in \mathbb{F}^n$ and $p \in \mathbb{R} - \{0\}$, define

$$\|x\|_p = \left(\mathbb{E}_i |x_i|^p\right)^{1/p}.$$

We will be particularly interested in the case that $p = 0$, for which we define $\|x\|_0 = \lim_{p \to 0} \|x\|_p$. It is not too hard to see that for any vector x, $\|x\|_0$ equals $\prod_{i \in [n]} |x_i|^{1/n}$, the geometric mean of the magnitude of the entries of x. Note that in the case that $p < 1$, $\|\cdot\|_p$ is not a norm and not convex, but we will nevertheless refer to it as the p-norm, following the conventional notation for these matrix-norm approximation problems.

Given a matrix $A \in \mathbb{F}^{m \times n}$, the $p \to q$ "norm" of A is defined as

$$\|A\|_{p \to q} = \max_{x \in \mathbb{F}^n : \|x\|_p = 1} \|Ax\|_q.$$

The connection of $\|A\|_{p \to q}$ to $r(V)$ is apparent: for any matrix $V \in \mathbb{C}^{n \times d}$, we have

$$r(V) = \|V\|_{2 \to 0}^{2n}.$$

Over the last decade there has been significant interest in designing approximation algorithms or proving hardness of approximation for matrix $p \to q$ norms for $p, q \geq 1$ [BBH$^+$12, BRS15, BGG$^+$23]. Most notably, the $2 \to 4$ norm has been shown to be closely related to the Unique games and the small set expansion conjectures [BBH$^+$12]. To the best of our knowledge, the problem is not well-studied when $q < 1$. We prove tight hardness of approximation (assuming P $\neq$ NP) for the $2 \to q$ norm when $-1 < q < 2$.

For $q > -1$ let

$$\gamma_{\mathbb{F},q} = \mathbb{E}_{g \sim \mathbb{FN}(0,1)} [|g|^q]^{1/q}$$

be the q-norm of a standard (real/complex) normal random variable. Bhattiprolu, Ghosh, Guruswami, Lee, Tulsiani [BGG+23] showed that for any $1 \leq q < 2$, and any $\epsilon > 0$ it is NP-hard to approximate the $2 \to q$ norm of a real $m \times n$ matrix better than $\gamma_{\mathbb{R},p} + \epsilon$, matching known semidefinite relaxation-based approximation algorithms [Ste05]. In our main theorem, we build on their techniques and we extend their result to all $-1 < q < 2$.

Theorem 3.1.5 (Main Technical Hardness Theorem)**.** *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. For all $-1 < q < 2$ and $\epsilon > 0$, it is* NP*-hard to approximate $\|A\|_{2\to q}$ given a matrix $A \in \mathbb{F}^{m\times n}$ within a factor of $\gamma_{\mathbb{F},q} + \epsilon$.*

For the sake of completeness, in Section A.3 we write down a semidefinite relaxation of $\|A\|_{2\to q}$ for all $-1 < q < 2$ and prove that it gives a $\gamma_{\mathbb{F},q}$-approximation to $\|A\|_{2\to q}$, matching the above hardness result. As $r(V) = \|V\|_{2\to 0}^{2n}$, we also get that it is NP-hard to approximate $r(V)$ within a factor of $(\gamma_{\mathbb{C},0} + \epsilon)^{2n} = e^{-\gamma n(1+\epsilon)}$.

Next, we elaborate on Item 2 – an approximation preserving reduction from $r(V)$ to $\mathrm{per}(A)$. Our main observation is that the permanent of a highly rank-deficient $n \times n$ PSD matrix $A = VV^\dagger$ is essentially (up to subexponential error) the same as $r(V)$. This is a consequence of Wick's formula (Lemma 3.2.16), which allows us to view the permanent of a PSD matrix as a squared absolute moment of a complex multivariate Gaussian. As a result, we are able to use Theorem 3.1.5 to prove Theorem 3.1.2, which we do in Section 3.4.2.

3.1.2 Overview of the proof of Lemma 3.1.4

Let us start by explaining the proof of Eq. (3.3), which Lemma 3.1.4 improves upon. For any distribution $\mathcal{D}$ over $\mathbb{C}^n$ with $\mathbb{E}[\|x\|_2^2] = 1$, we have the bound

$$r(V) = \max_{\|x\|_2=1} \prod_{i\in[n]} |\langle v_i, x\rangle|^2 \geq \left(\frac{\mathbb{E}_{x\sim\mathcal{D}} \prod_{i\in[n]} |\langle v_i, x\rangle|^{2/n}}{\mathbb{E}_{x\sim\mathcal{D}} \|x\|_2^2}\right)^n = \mathbb{E}_{x\sim\mathcal{D}} \left[\prod_{i\in[n]} |\langle v_i, x\rangle|^{2/n}\right]^n.$$

Using Jensen's inequality on the RHS, we get

$$r(V) \geq \exp\left(\sum_{i\in[n]} \mathbb{E}_{x\sim\mathcal{D}} \ln |\langle v_i, x\rangle|^2\right). \tag{3.5}$$

The basic Gaussian rounding scheme picks $x \sim \mathcal{D} = \mathbb{CN}(0, X^*)$ (see Definition 3.2.9 for a definition of the complex Gaussian distribution). Notice that for each i, $\langle v_i, x\rangle \sim \mathbb{CN}(0, v_i^\dagger X^* v_i) = \mathbb{CN}(0, 1)$ by assumption of Lemma 3.1.4. Since $\mathbb{E}_{y\sim\mathbb{CN}(0,1)} \ln |y|^2 = -\gamma$, we immediately get $r(V) \geq \exp(-n\gamma)$.

One can see that the analysis (in particular, the application of Jensen's inequality) is tight if $A = V = X^* = I$ for example. So to improve on Eq. (3.5), we must use a different rounding algorithm. Our first observation is that in the special case that $A = V = I$ where $v_i = e_i$, we can get the optimal lower bound by sampling independent Rademachers $s_1, \ldots, s_n \sim \{\pm 1\}$, and setting $x = \sum_{i\in[n]} s_i v_i$. With this choice, $\mathbb{E}_x \ln |\langle v_i, x\rangle|^2 = \mathbb{E}_x \ln 1 = 0$, implying $r(V) \geq 1$.

One could try to use a similar rounding scheme in the more general case of Lemma 3.1.4, i.e., when A is close to I in the sense that $\mathrm{tr}(A) \geq (1-\epsilon)n$. Unfortunately this strategy ends up failing, as when the v_i's are not exactly orthogonal, there could be a nonzero probability that $\langle v_i, x\rangle = 0$, which would imply $\mathbb{E}_x \ln |\langle v_i, x\rangle|^2 = -\infty$. Note that if A is close to I, this singularity is a very small probability event for most of the vectors v_i, so it is natural to try to avoid it by adding some noise to x. We do this by interpolating between the two rounding schemes. We pick a parameter $0 < \beta < 1$, and set $x = \sqrt{1-\beta}g + \sqrt{\beta}\sum_{i\in[n]} s_i v_i$, up to some normalization, where $g \sim \mathbb{CN}(0, X^*)$.

On the technical side, this interpolation helps us analyze $\mathbb{E}_x \ln |\langle v_i, x\rangle|^2$ in terms of tractable quantities. We use a sharp bound on the expected log of the magnitude squared of a noncentral complex Gaussian (see Lemma 3.2.14): for any $c \in \mathbb{C}$,

$$\mathbb{E}_{g\sim\mathbb{CN}(0,1)} \ln |g+c|^2 \geq -\gamma + |c|^2 - \frac{|c|^4}{4}.$$

As a result of this inequality, when β is bounded away from 1, we can effectively bound $\mathbb{E}_x \ln |\langle v_i, x\rangle|^2$ using only the second and fourth moments of the random variable $\sum_{i\in[n]} s_i v_i$, which are tractable.

3.1.3 Overview of the proof of Theorem 3.1.5

As alluded to in Section 3.1.1, prior to our work, optimal hardness results for the $2 \to q$ norm are already established for $q \geq 1$. Our first observation is that these results [GRSW16, BRS15, BGG+23] can be extended to all $-1 < q < 2$ (see Theorem 3.4.1), or even more generally, to 2-concave f-means (see Definition 3.2.2).

In particular, one can deduce the following theorem.

Theorem 3.1.6 (Informal version of Theorem 3.4.1)**.** *Let $q < 2$. Assume that there is a family $\{E_k\}$ of $k \times d_k$ gadget matrices such that $\|E_k\|_{2\to q} = 1$, but for all "smooth" unit vectors x ($\|x\|_\infty \ll 1$), $\|E_k x\|_q \leq \gamma$. Then for all $\epsilon > 0$, it is NP-Hard to distinguish between the following two cases given a matrix $A : \mathbb{C}^m \to \mathbb{C}^n$ with $\|A\|_{2\to 2} \leq 1$.*

1. *Completeness: $\|A\|_{2\to q} = 1$, or*
2. *Soundness: $\|A\|_{2\to q} \leq \gamma + \epsilon$.*

It remains to construct an appropriate family of gadget matrices $\{E_k\}$. We will use the following family, which was suggested in [BRS15]. For $k \geq 1$, define $E_k^{(\mathbb{C})} \in \mathbb{C}^{4^k \times k}$ as the matrix whose rows consist of the members of $\frac{1}{\sqrt{k}} \cdot \{-1, +1, -i, +i\}^k$ ordered arbitrarily.

It remains to show that the matrices $E_k^{(\mathbb{C})}$ satisfy the requirements of Theorem 3.1.6 with $\gamma \approx \gamma_{\mathbb{C},q}$. By construction, $\left\|E_k^{(\mathbb{C})}\right\|_{2\to q} = 1$.

To prove this, in Lemma A.2.1 we prove a Berry-Esseen type result for test functions of the form $|x|^q$ for $q \neq 0$ and $\log|x|$ otherwise, applied to a sum of independent random variables. In particular, the special case of interest to us (for Theorem 3.1.2) is $q = 0$. In that case, the test function is $\log|x|$ which has a singularity at $x = 0$, but we are nevertheless one can bound the right hand side of the lemma below by an arbitrarily small quantity as $\delta \to 0$.

Lemma 3.1.7 (Informal version of Lemma 3.4.3)**.** *Let $-1 < q < 2$ and $0 < \delta < 1$. For all "smooth" unit vectors x with $\|x\|_\infty \le \delta$,*

$$f(\|E_k x\|_q) - \gamma_{\mathbb{C},q} \lesssim -\int_0^\delta \min(0, f(u)) + \delta \cdot \left(\max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1)\right),$$

where $f(x) = |x|^q$ for $q \neq 0$, and $f(x) = \log|x|$ for $q = 0$.

3.1.4 Future Directions

The most exciting open problem is to determine the correct approximability for PSD permanents. Improving the hardness result seems to be out of reach of current techniques, but our ideas provide a clear path to improving the algorithmic result. Any significant improvement to the algorithm along the lines of our ideas would require significantly better versions of Lemmas 3.1.3 and 3.1.4. In particular, Lemma 3.1.3 is currently the bottleneck to a better approximation ratio, specifically the $O(\epsilon^2)$ dependence. We conjecture that it can be improved to $O(\epsilon)$, which would yield better approximation ratios as a corollary.

Although we don't have concrete new applications of hardness of approximation of $\|A\|_{2\to q}$ for $q \neq 0$, we expect to find further applications of the machinery developed here in addressing counting and optimization of linear algebraic problems, e.g., in estimating mixed discriminant, sub-determinant maximization, Nash-welfare maximization, etc.

Chapter organization. In Section 3.2, we present preliminary definitions and results that we will use. In Section 3.3, we prove Theorem 3.1.1. In Section 3.4, we prove Theorems 3.1.2 and 3.1.5 (with some components of the proof appearing in Sections A.1 and A.2).

3.2 Preliminaries

3.2.1 Generalized Means and Norms

Although we mostly use p-norms that are defined using an expectation, it will be convenient to also define the counting version of 2-norm, which we denote as ℓ_2.

Definition 3.2.1 (ℓ_2-norm, Frobenius norm)**.** Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. For a vector $x \in \mathbb{F}^n$, define $\|x\|_{\ell_2} = \sqrt{\sum_{i\in[n]} |x_i|^2}$. For a matrix $A \in \mathbb{F}^{m\times n}$, define $\|A\|_F = \sqrt{\sum_{i,j\in[n]} |A_{i,j}|^2}$.

We work with a generalization of $\|\cdot\|_p$ using the framework of "f-means".

Definition 3.2.2. Let $f : \mathbb{R}_{\ge 0} \to \mathbb{R}$ be continuous and injective. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ be a field. For a vector $x \in \mathbb{F}^n$, define

$$[x]_f^f := \mathbb{E}_{i\sim[n]} f(|x_i|),$$

$$[x]_f := f^{-1}([x]_f^f) = f^{-1}(\mathbb{E}_{i\sim[n]} f(|x_i|)).$$

We will refer to $[x]_f$ as the *f-mean* of x. More generally, for a random variable X over $\mathbb{F}$ define

$$[X]_f^f = \mathbb{E}f(|X|), \quad [X]_f = f^{-1}(\mathbb{E}f(|X|)).$$

For a matrix $A \in \mathbb{F}^{n\times d}$, define

$$[A]_{2\to f} := \max_{x\in\mathbb{F}^n} \frac{[Ax]_f}{\|x\|_2}.$$

f-means provide a convenient and unified way to talk about $\|\cdot\|_p$, even for the case of $p = 0$.

Observation 3.2.3 (Power Means)**.** *For all $p \in \mathbb{R}$, define*

$$f_p(x) = \begin{cases} x^p & \text{if } p > 0, \\ \log x & \text{if } p = 0, \\ -x^p & \text{if } p < 0. \end{cases}$$

Then, we have $[x]_{f_p} = \|x\|_p$ for any $p \in \mathbb{R}$.

We note that the observation would still hold if we used x^p instead of $-x^p$ in the third case, but it will be convenient for f_p to always be an increasing function.

We will mostly be concerned with f-means that are dominated by the 2-norm. This happens exactly when f is 2-concave:

Definition 3.2.4. A function $f : \mathbb{R}_{>0} \to \mathbb{R}$ is 2-concave if $x \to f(\sqrt{x})$ is concave.

Example 3.2.5. For any $p \le 2$, f_p is 2-concave.

We make some useful observations about 2-concave functions.

Claim 3.2.6. *Let $f : \mathbb{R}_{>0} \to \mathbb{R}$ be a 2-concave increasing function. Then,*

1. *$[x]_f \le \|x\|_2$ for any vector x. Equivalently, $[x]_f^f \le f(\|x\|_2)$.*
2. *$f'(x) \le \frac{f'(y)}{y} \cdot x$ for $0 < y \le x$.*
3. *$f(x) - f(y) \le \frac{f'(y)}{2y} \cdot x^2$ for $0 < y \le x$.*

Proof. 1. Using Jensen's inequality on the concave function $x \to f(\sqrt{x})$,

$$[x]_f^f = \mathbb{E}f(|x_i|) = \mathbb{E}f\left(\sqrt{|x_i|^2}\right) \le f(\sqrt{\mathbb{E}|x_i|^2}) = f(\|x\|_2).$$

2. Follows from the fact that $f'(\sqrt{x}) = \frac{f(\sqrt{x})}{2\sqrt{x}}$ is increasing in x.

3. Integrating the above from y to x,

$$f(x) - f(y) \le \frac{f'(y)}{y} \cdot \frac{(x^2 - y^2)}{2} \le \frac{f'(y)}{2y} \cdot x^2.$$

□

One could ask when $[x]_f$ is a homogeneous function of x. It turns out that this is exactly when $[x]_f$ is a p-norm [HLP52].

Lemma 3.2.7. *$[\cdot]_f$ is 1-homogeneous (that is, $[ax]_f = a[x]_f$ for all scalars a and vectors x) if and only if it equals $[\cdot]_{f_p} = \|\cdot\|_p$ for some $p \in \mathbb{R}$.*

We will require another simple fact about f-means.

Fact 3.2.8. *Let V be a matrix $V \in \mathbb{F}^{n\times d}$ and integer $k > 0$. Then,*

$$[V^{(k)}]_{2\to f} \coloneqq \left[\!\left[\begin{bmatrix} V \\ \vdots \\ V \end{bmatrix}\right]\!\right]_{2\to f} = [V]_{2\to f},$$

where V is copied k times in the right hand side.

Proof. For any vector $x \in \mathbb{F}^d$, consider the two vectors $z = Vx$ and $z^{(k)} = V^{(k)}x$. Note that a uniformly random entry of z has the same distribution as a random entry of $z^{(k)}$, so $[z]_f = [z^{(k)}]_f$. The claim follows from the definition of $[\cdot]_{2\to f}$. □

3.2.2 Gaussians

We will consider both real and complex Gaussians.

Definition 3.2.9. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. For the $n \times n$ identity matrix I_n, $\mathbb{FN}(0, I_n)$ is defined to be the distribution over vectors $x \in \mathbb{F}^n$ given by the density function

$$p_{\mathbb{F}}(x) = \begin{cases} (2\pi)^{-n/2} \cdot \exp\left(-\|x\|_{\ell_2}^2/2\right) & \text{if } \mathbb{F} = \mathbb{R}, \\ \pi^{-n} \exp(-\|x\|_{\ell_2}^2) & \text{if } \mathbb{F} = \mathbb{C}. \end{cases}$$

More generally, given a positive semidefinite covariance matrix $\Sigma = AA^\dagger$ for $A \in \mathbb{F}^{n\times d}$, define $\mathbb{FN}(0, \Sigma)$ to be distributed as Ax, where $x \sim \mathbb{FN}(0, I_d)$. We will sometimes use $\mathbb{N}$ to denote $\mathbb{RN}$.

More concretely, a complex Gaussian $g \sim \mathbb{CN}(0, 1)$ can be sampled by sampling its real and imaginary parts independently from $\mathbb{N}(0, 1/2)$. There are formulas for the moments of univariate real and complex Gaussians in terms of the Gamma function.

Definition 3.2.10. For any $p \in \mathbb{R}$ and $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, define $\gamma_{\mathbb{F},p} = [g]_{f_p}$, where g is a random variable distributed as $\mathbb{FN}(0, 1)$.

Fact 3.2.11. *For any $p \in (-1, \infty) - \{0\}$,*

$$\gamma_{\mathbb{R},p}^p = \mathbb{E}_{g\sim\mathbb{N}(0,1)}[|g|^p] = \frac{2^{p/2} \cdot \Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}},$$

and for any $p \in (-2, \infty) - \{0\}$,

$$\gamma_{\mathbb{C},p}^{p} = \mathbb{E}_{g \sim CN(0,1)}[|g|^p] = \Gamma\left(\frac{p}{2} + 1\right).$$

In particular, this implies

$$\gamma_{\mathbb{R},1} = \sqrt{\frac{2}{\pi}}, \quad \gamma_{\mathbb{C},1} = \sqrt{\frac{\pi}{2}}, \quad \gamma_{\mathbb{R},0} = \lim_{p \to 0} \gamma_{\mathbb{R},p} = \sqrt{\frac{e^{-\gamma}}{2}}, \quad \gamma_{\mathbb{C},0} = \lim_{p \to 0} \gamma_{\mathbb{C},p} = \sqrt{e^{-\gamma}}.$$

Fact 3.2.12 (Moment Generating Function of $|g|^2$)**.** *Let* $g \sim \mathbb{CN}(0,1)$*. For any* $t < 1$*,* $\mathbb{E}[e^{t|g|^2}] = (1-t)^{-1}$*.*

We will need sharp bounds on the expected value of $\ln|g+c|^2$ for Gaussian g and fixed c. First, we prove an estimate on the exponential integral function.

Fact 3.2.13. *For* $x \geq 0$ *it holds that*

$$\mathrm{Ei}(-x) = \int_{-\infty}^{-x} \frac{e^t}{t} dt = \gamma + \ln(x) + \sum_{k=1}^{\infty} \frac{(-x)^k}{k \cdot k!} \leq \gamma + \ln(x) - x + \frac{x^2}{4}.$$

Proof. The identity is due to Equation 5.1.11 in [AS48]. For the inequality, we must show that the function

$$f(x) = \sum_{k=3}^{\infty} \frac{(-x)^k}{k \cdot k!}$$

is nonpositive for $x \geq 0$. To do this, observe first that $f(0) = 0$, and

$$\begin{aligned} f'(x) &= \sum_{k=3}^{\infty} \frac{(-1)^k \cdot x^{k-1}}{k!} \\ &= \frac{1}{x} \sum_{k=3}^{\infty} \frac{(-x)^k}{k!} \\ &= \frac{e^{-x} - \left(1 - x + \frac{x^2}{2}\right)}{x} \\ &\leq 0. \qquad (e^{-x} \leq 1 - x + x^2/2 \text{ for } x \geq 0) \end{aligned}$$

Therefore, $f(x) \leq 0$ for $x \geq 0$. □

Lemma 3.2.14. *Let* $c \in \mathbb{C}$*. Then,* $\mathbb{E}_{g \sim \mathbb{CN}(0,1)}[\ln|g+c|^2] \geq -\gamma + |c|^2 - |c|^4/4$*.*

Proof. Define $x = |c|^2$. By [Mos20, Eqn. 35, Thm. 1] we have the identity

$$\mathbb{E}_{g \sim \mathbb{CN}(0,1)}[\ln|g+c|^2] = -\ln(x) - \mathrm{Ei}(x).$$

By Fact 3.2.13, this is at least $-\gamma + x - x^2/4$, as desired. □

3.2.3 Permanent

For a matrix $A \in \mathbb{C}^{n\times n}$, its permanent is defined as

$$\operatorname{per}(A) := \sum_{\sigma\in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)}.$$

On the domain of positive semidefinite matrices, the permanent has some nice properties. For example, it is monotone w.r.t. the Loewner order [AGGS17].

Lemma 3.2.15. *If $A, B \in \mathbb{C}^{n\times n}$ are hermitian and $A \succeq B \succeq 0$, then*

$$\operatorname{per}(A) \geq \operatorname{per}(B).$$

Proof Sketch. The statement of the lemma follows, because $A \succeq B \succeq 0$ implies that $A^{\otimes n} \succeq B^{\otimes n} \succeq 0$. So, if 1_{S_n} is the indicator vector of all permutations in $\mathbb{R}^{n\otimes n}$,

$$\operatorname{per}(A) = \frac{1}{n!} 1_{S_n}^\dagger A^{\otimes n} 1_{S_n} \geq \frac{1}{n!} 1_{S_n}^\dagger B^{\otimes n} 1_{S_n} = \operatorname{per}(B)$$

as desired. □

Lemma 3.2.16. *For any PSD matrix $VV^\dagger$ with $V \in \mathbb{R}^{n\times d}$, we have*

$$\mathbb{E}_{x\sim\mathbb{N}(0,I)} \left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right] = c^{\mathbb{R}}_{n,d} \cdot \mathbb{E}_{x\in\mathbb{R}^d, \|x\|_2=1} \left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right].$$

For any PSD matrix $VV^\dagger$ with $V \in \mathbb{C}^{n\times d}$, we have

$$\operatorname{per}(VV^\dagger) = \mathbb{E}_{x\sim\mathbb{CN}(0,I)} \left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right] = c^{\mathbb{C}}_{n,d} \cdot \mathbb{E}_{x\in\mathbb{C}^d, \|x\|_2=1} \left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right].$$

Here, $v_1, \ldots, v_n$ are the rows of V. The proportionality constants above are defined by

$$c^{\mathbb{R}}_{n,d} = \frac{\Gamma(n + d/2)}{\Gamma(d/2)\cdot(d/2)^n}, \qquad c^{\mathbb{C}}_{n,d} = \frac{(d+n-1)!}{(d-1)!\cdot d^n}.$$

Proof. The first equality in the second conclusion follows from Isserlis' theorem/Wick's formula (see 3.1.4 in [Bar16]).

We prove the other equality in the real case, but the complex case can be proved similarly. Observe that

$$\begin{aligned}
\mathbb{E}_{x\sim\mathbb{N}(0,I)} \left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right] &= \mathbb{E}_{x\sim\mathbb{N}(0,I)} \|x\|_2^{2n} \cdot \left[\prod_{i\in[n]} \left|\left\langle v_i, \frac{x}{\|x\|_2}\right\rangle\right|^2\right] \\
&= \mathbb{E}_{x\sim\mathbb{N}(0,I)} \|x\|_2^{2n} \cdot \mathbb{E}_{x\in\mathbb{R}^n, \|x\|_2=1} \left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right] \\
&= d^{-n} \mathbb{E}_{x\sim\mathbb{N}(0,I)} \|x\|_{\ell_2}^{2n} \cdot \mathbb{E}_{x\in\mathbb{R}^n, \|x\|_2=1} \left[\prod_{i\in[n]} |\langle v_i, x\rangle|^2\right]
\end{aligned}$$

The first identity uses that for $x \sim \mathbb{N}(0, I)$, $\|x\|_2$ is independent from $x/\|x\|_2$. The second identity uses that fact that if $x \sim \mathbb{N}(0, I)$, then $x/\|x\|_2$ is distributed uniformly on a sphere of radius $\|x\|_2$. To conclude the proof, notice that $\mathbb{E}_{x\sim\mathbb{N}(0,I)} \|x\|_{\ell_2}^{2n}$ is the n^{th} moment of a chi-squared random variable with d-degrees of freedom, which is $2^n \frac{\Gamma(n+d/2)}{\Gamma(d/2)}$. □

We will also require the following formula for the permanent of the sum of two matrices [Per12, Page 2].

Lemma 3.2.17. *For any two matrices $A, B \in \mathbb{C}^{n\times n}$,*

$$\operatorname{per}(A+B) = \sum_{S,T\subseteq[n],|S|=|T|} \operatorname{per}(A_{S,T}) \cdot \operatorname{per}(B_{\bar{S},\bar{T}}).$$

When $A = I$, this simplifies to

$$\operatorname{per}(I+B) = \sum_{S\subseteq[n]} \operatorname{per}(B_{S,S}).$$

Above, $A_{S,T}$ is the $|S| \times |T|$ submatrix of A containing rows only in S and columns only in T.

3.3 Algorithm

We start by expanding on the basic setup of the algorithms of [AGGS17, YP22], which we briefly introduced in Section 3.1.1. After this, we will show how Lemmas 3.1.3 and 3.1.4 imply Theorem 3.1.1. Later on, in Sections 3.3.1 and 3.3.2 respectively, we prove Lemmas 3.1.3 and 3.1.4.

Let $A = VV^\dagger$ be the PSD matrix whose permanent we wish to compute. Let $v_1, \ldots, v_n \in \mathbb{C}^n$ be the rows of V, so by Lemma 3.2.16,

$$\operatorname{per}(A) = \mathbb{E}_{x\sim\mathbb{CN}(0,I)} \left[\prod_{i\in[n]} |\langle x, v_i\rangle|^2 \right].$$

Recall the log-concave maximization problem $\mathrm{SDP}(V)$ we associated with this problem:

$$\mathrm{SDP}(V) = \max_{X:X\succeq 0,\operatorname{tr}(X)=n} \prod_{i\in[n]} v_i^\dagger X v_i.$$

Let X^* be the optimal solution to $\mathrm{SDP}(V)$. Note that X^* can be found efficiently. It will be convenient to make a simplification to our problem. We will replace the matrix A by $\tilde{A} = D^{-1/2} A D^{-1/2}$, where D is a positive semidefinite diagonal matrix defined as $D_{i,i} = v_i^\dagger X^* v_i$. Since D is diagonal,

$$\operatorname{per}(A) = \operatorname{per}(\tilde{A}) \cdot \operatorname{per}(D) = \operatorname{per}(\tilde{A}) \cdot \mathrm{SDP}(V),$$

so it suffices to approximate $\operatorname{per}(\tilde{A})$ instead of $\operatorname{per}(A)$. Writing $\tilde{A} = \tilde{V}\tilde{V}^\dagger$ for $\tilde{V} = D^{-1/2}V$, we can see that the objective functions of $\mathrm{SDP}(V)$ and $\mathrm{SDP}(\tilde{V})$ are positive scalar multiples of

each other, so $\mathrm{SDP}(\tilde{V})$ is also maximized by X^*. Note that $\tilde{A}$ enjoys the additional property $\tilde{v}_i^\dagger X^* \tilde{v}_i = 1$ for all $i \in [n]$, where $\tilde{v}_i = D^{-1/2} v_i$. Replacing A by $\tilde{A}$, we will henceforth assume that the maximizer X^* of $\mathrm{SDP}(V)$ satisfies

$$v_i^\dagger X^* v_i = 1 \text{ for all } i \in [n]. \tag{3.6}$$

In particular, this implies $\mathrm{SDP}(V) = 1$. Under this assumption, A satisfies an important property.

Claim 3.3.1. *We have $A \preceq I$.*

Proof. Let $f(X) = \prod_{i\in[n]} v_i^\dagger X v_i$ be the objective function of $\mathrm{SDP}(V)$. We can compute

$$\nabla(\ln f)(X) = \sum_{i\in[n]} \frac{v_i v_i^\dagger}{v_i^\dagger X v_i}.$$

In particular, by Eq. (3.6), $\nabla(\ln f)(X^*) = \sum_{i\in[n]} v_i v_i^\dagger = V^\dagger V$.

The optimality conditions for X^* imply that for all symmetric matrices M with $\mathrm{tr}(M) = 0$ and $W_- \subseteq \mathrm{Range}(X^*)$ it holds that

$$\langle V^\dagger V, M\rangle = \langle \nabla(\ln f)(X^*), M\rangle \le 0.$$

Here, W_- denotes the vector space spanned by the negative eigenvectors of M. Now, let $Q \succeq 0$ be an arbitrary PSD matrix, and set $M = Q - \frac{\mathrm{tr}(Q)}{n} X^*$. M satisfies both the conditions above, and therefore we have

$$\begin{aligned}
0 &\ge \langle V^\dagger V, M\rangle \\
&= \langle V^\dagger V, Q\rangle - \frac{\mathrm{tr}(Q)}{n}\langle V^\dagger V, X^*\rangle \\
&= \langle V^\dagger V, Q\rangle - \frac{\mathrm{tr}(Q)}{n}\sum_{i\in n} v_i^\dagger X^* v_i \\
&= \langle V^\dagger V, Q\rangle - \mathrm{tr}(Q). && \text{(by Eq. (3.6))}
\end{aligned}$$

In other words, $\langle V^\dagger V, Q\rangle \le \mathrm{tr}(Q)$ for all $Q \succeq 0$, implying $V^\dagger V \preceq I$. Therefore $A = VV^\dagger \preceq I$.

□

Claim 3.3.1 immediately implies $\mathrm{per}(A) \le 1$. In [AGGS17, YP22], the authors prove the complimentary inequalities

$$\mathrm{per}(A) \ge \frac{n!}{n^n}\cdot r(V) \ge \exp(-\gamma n)\cdot \frac{n!}{n^n}\cdot \mathrm{SDP}(V) = \exp(-\gamma n)\cdot \frac{n!}{n^n} \gtrsim \exp(-(\gamma+1)n), \tag{3.7}$$

and together, the two inequalities above provide a $e^{-(1+\gamma)n}$ approximation for $\mathrm{per}(A)$.

Recall Lemmas 3.1.3 and 3.1.4, which (under Claim 3.3.1) improve the above inequalities to

$$e^{-(\gamma+1)n} \cdot \exp\left(n \cdot \ell\left(\frac{\mathrm{tr}(A)}{n}\right)\right) \leq \mathrm{per}(A) \leq \exp\left(n \cdot r\left(\frac{\mathrm{tr}(A)}{n}\right)\right). \tag{3.8}$$

Here, $\ell(x) = \max_{0\leq\beta\leq 1} \ln(1-\beta) + \frac{\beta x}{(1-\beta)} - \frac{0.273\beta^2}{(1-\beta)^2 x}$, and $r(x) = \ln\left(1 - \frac{(1-x)^2}{20}\right)$. We are now ready to prove Theorem 3.1.1.

Proof of Theorem 3.1.1. Let $A = VV^\dagger \succeq 0$, where V has rows $v_1, \ldots, v_n$. Our algorithm will first solve SDP(V) and use Eq. (3.6) and Claim 3.3.1 to reduce to the case that $0 \preceq A \preceq I$ and $v_i^\dagger X^* v_i = 1$ for all i, where X^* is the optimal solution to SDP(V). We will then output $\exp\left(n \cdot r\left(\frac{\mathrm{tr}(A)}{n}\right)\right) = \left(1 - \frac{\epsilon^2}{20}\right)^n$.

Eq. (3.8) implies that the approximation factor of this algorithm is at least $e^{-(\gamma+1-\alpha)n}$, where α is the minimum value of $r(x) - \ell(x)$ over all $x \in [0,1]$. Write $\ell(x) \geq \ell'(x) := \max(0, \ln(1-\beta^*) + \frac{\beta^* x}{(1-\beta^*)} - \frac{0.273(\beta^*)^2}{(1-\beta^*)^2 x})$ for $\beta^* = 0.34$. One can numerically determine that $\alpha \geq \min_{0\leq x\leq 1} r(x) - \ell'(x) \geq 10^{-4}$. □

3.3.1 Proof of Lemma 3.1.3

We first prove an inequality that we will require. The proof of this inequality is inspired by an identity of Barvinok [Bar20].

Lemma 3.3.2. *For any matrix $0 \preceq B \prec I$, $\mathrm{per}(I+B) \leq \det\left((I-B)^{-1}\right)$.*

Proof. Write $B = VV^\dagger$ for $V \in \mathbb{C}^{\times n}$. Let $v_1, \ldots, v_n$ be the rows of V.

$$\begin{aligned}
\mathrm{per}(I+B) &= \sum_{S\subseteq[n]} \mathrm{per}(B_{S,S}) && \text{(Lemma 3.2.17)} \\
&= \sum_{S\subseteq[n]} \mathbb{E}_{g\sim\mathbb{CN}(0,I)} \left[\prod_{i\in S} |\langle v_i, g\rangle|^2\right] && \text{(Lemma 3.2.16)} \\
&= \mathbb{E}_{g\sim\mathbb{CN}(0,I)} \left[\prod_{i\in[n]} \left(1 + |\langle v_i, g\rangle|^2\right)\right] \\
&\leq \mathbb{E}_{g\sim\mathbb{CN}(0,I)} \left[\prod_{i\in[n]} e^{|\langle v_i, g\rangle|^2}\right] && (1+x \leq e^x \text{ for all } x) \\
&= \mathbb{E}_{g\sim\mathbb{CN}(0,I)} \left[\exp\left(g^\dagger V^\dagger V g\right)\right].
\end{aligned}$$

Let $\sigma_1, \ldots, \sigma_n$ be the eigenvalues of $V^\dagger V$. Since g is invariant under unitary transformations,

we can rotate g into the eigenbasis of $V^\dagger V$ to get that

$$\begin{aligned}
\operatorname{per}(I+B) &\le \mathbb{E}_{g\sim\mathbb{CN}(0,I)}\left[\exp\left(\sum_{i\in[n]}\sigma_i|g_i|^2\right)\right] \\
&= \prod_{i\in[n]}\mathbb{E}_{g\sim\mathbb{CN}(0,1)}\left[e^{\sigma_i|g|^2}\right] && \text{(Independence of } g_i\text{)} \\
&= \prod_{i\in[n]}\frac{1}{1-\sigma_i}. && \text{(Fact 3.2.12)}
\end{aligned}$$

Noting that the eigenvalues of $V^\dagger V$ match those of $B = VV^\dagger$, this is equal to $\det\left((I-B)^{-1}\right)$. □

Now, we are ready to prove Lemma 3.1.3. Let $0 \preceq A \preceq I$ be a matrix with $\operatorname{tr}(A) \le (1-\epsilon)n$. Let $0 \le \lambda_1 \le \ldots \le \lambda_n \le 1$ be the eigenvalues of A, and let $v_1, \ldots, v_n$ be the corresponding eigenvectors. Let $t \in (1/2, 1]$ be a parameter we will set later, and let i_t be the smallest index i such that $\lambda_i > t$. For any parameter $t \in (1/2, 1]$, we can write

$$A \preceq tI + \sum_{i\ge i_t}(\lambda_i - t)v_i v_i^\dagger = t\cdot\left(I + \sum_{i\ge i_t}\frac{\lambda_i - t}{t}v_i v_i^\dagger\right).$$

Write $B = \sum_{i\ge i_t}\frac{\lambda_i - t}{t}v_i v_i^\dagger$. Since $t > 1/2$, $B \prec I$, so it satisfies the conditions of Lemma 3.3.2.

$$\begin{aligned}
\operatorname{per}(A) &\le t^n \cdot \operatorname{per}(I+B) && \text{(Lemma 3.2.15)} \\
&\le \frac{t^n}{\det(I-B)}. && \text{(Lemma 3.3.2)}
\end{aligned}$$

We now pick $t = 1 - \epsilon/5$. For this choice of t, we must have $i_t \ge \frac{\epsilon n}{2}$, since otherwise, $\operatorname{tr}(A) \ge t\cdot(n - i_t) \ge (1-\epsilon/5)\cdot(1-\epsilon/2)n > (1-\epsilon)n$ contradicts the fact that $\operatorname{tr}(A) \le (1-\epsilon)n$. We compute

$$\det(I-B) = \prod_{i\ge i_t}\left(1 - \frac{\lambda_i - t}{t}\right) = \prod_{i\ge i_t}\left(2 - \frac{\lambda_i}{t}\right) \ge \left(2 - \frac{1}{t}\right)^{n-i_t}.$$

Plugging in the definition of t and our lower bound on i_t, this is at least

$$\left(2 - \frac{1}{1-\epsilon/5}\right)^{(1-\epsilon/2)n} = \left(\frac{1-2\epsilon/5}{1-\epsilon/5}\right)^{(1-\epsilon/2)n}.$$

We now have our upper bound on $\operatorname{per}(A)$:

$$\operatorname{per}(A) \le (1-\epsilon/5)^n \cdot \left(\frac{1-\epsilon/5}{1-2\epsilon/5}\right)^{(1-\epsilon/2)n} = \left(\frac{(1-\epsilon/5)\cdot(1-\epsilon/5)^{1-\epsilon/2}}{(1-2\epsilon/5)^{1-\epsilon/2}}\right)^n.$$

To complete the proof, we use that $\frac{(1-\epsilon/5)\cdot(1-\epsilon/5)^{1-\epsilon/2}}{(1-2\epsilon/5)^{1-\epsilon/2}} \le 1 - \frac{\epsilon^2}{20}$ for all $\epsilon \in [0,1]$.

3.3.2 Proof of Lemma 3.1.4

By Eq. (3.7), it suffices to prove a lower bound on $r(V)$. Let X^* be the optimal solution to SDP(V). Consider the following randomized rounding scheme to a solution of $r(V)$: sample $g \sim \mathbb{CN}(0, X^*)$ and $s_i \sim \{z \in \mathbb{C} : |z| = 1\}$ independently for all $i \in [n]$. Let $x = \sqrt{1-\beta}g + \sqrt{\frac{\beta n}{\mathrm{tr}(A)}} \sum_{i\in[n]} s_i v_i$. We will use the bound

$$r(V)^{1/n} = \max_{\|x\|_2^2=1} \prod_{i\in[n]} |\langle v_i, x\rangle|^{2/n} \geq \frac{\mathbb{E}_x[\prod_{i\in n} |\langle v_i, x\rangle|^{2/n}]}{\mathbb{E}_x[\|x\|_2^2]}.$$

First we compute the denominator.

$$\begin{aligned}
n \cdot \mathbb{E}[\|x\|_2^2] &= n \cdot \mathbb{E}[\|x\|_2^2] \\
&= \mathbb{E}\left[\left\|\sqrt{1-\beta}g + \sqrt{\frac{\beta n}{\mathrm{tr}(A)}} \sum_{i\in[n]} s_i v_i\right\|_{\ell_2}^2\right] \\
&= (1-\beta)\mathbb{E}[\|g\|_{\ell_2}^2] + \frac{\beta n}{\mathrm{tr}(A)}\mathbb{E}\left[\sum_{i,j} s_i s_j \langle v_i, v_j\rangle\right] + \sqrt{\frac{\beta n}{\mathrm{tr}(A)}(1-\beta)}\mathbb{E}\left[\left\langle g, \sum_i s_i v_i\right\rangle\right] \\
&= (1-\beta)\mathbb{E}[\|g\|_{\ell_2}^2] + \frac{\beta n}{\mathrm{tr}(A)} \sum_i \|v_i\|_{\ell_2}^2 && \text{(Independence)} \\
&= (1-\beta)\cdot \mathrm{tr}(X^*) + \frac{\beta n}{\mathrm{tr}(A)} \cdot \sum_{i\in[n]} \|v_i\|_{\ell_2}^2 && (g \sim \mathbb{CN}(0, X^*), \text{ definition of } \|\cdot\|_{\ell_2}) \\
&= n. && (\mathrm{tr}(X^*) = n,\ \textstyle\sum_{i\in[n]} \|v_i\|_{\ell_2}^2 = \mathrm{tr}(VV^\dagger) = \mathrm{tr}(A))
\end{aligned}$$

So, $\mathbb{E}[\|x\|_2^2] = 1$. It remains to lower bound the numerator. We start by applying Jensen's inequality to get

$$\mathbb{E}_x\left[\prod_{i\in n} |\langle v_i, x\rangle|^{2/n}\right] \geq \exp\left(\frac{1}{n}\sum_{i\in[n]} \mathbb{E}_x[\ln |\langle v_i, x\rangle|^2]\right). \tag{3.9}$$

We will bound each of the terms inside the sum. Fix some $i \in [n]$, and let $y_i = \sqrt{\frac{n}{\mathrm{tr}(A)}} \sum_{j\in[n]} s_j \langle v_i, v_j\rangle$ and $z_i = \langle g, v_i\rangle$, so $\langle v_i, x\rangle = \sqrt{1-\beta} z_i + \sqrt{\beta} y_i$. Notice that $z_i \sim \mathbb{CN}(0, v_i^\dagger X^* v_i) = \mathbb{CN}(0,1)$ by assumption. Let us bound

$$\begin{aligned}
\mathbb{E}[\ln |\langle v_i, x\rangle|^2] &= \mathbb{E}[\ln |\sqrt{1-\beta} z_i + \sqrt{\beta} y_i|^2] \\
&= \ln(1-\beta) + \mathbb{E}\left[\ln \left|z_i + \sqrt{\frac{\beta}{1-\beta}} y_i\right|^2\right] \\
&\geq -\gamma + \ln(1-\beta) + \frac{\beta}{1-\beta}\mathbb{E}[|y_i|^2] - \frac{\beta^2}{4(1-\beta)^2}\mathbb{E}[|y_i|^4]. \\
& && \text{(Lemma 3.2.14, } z_i \sim \mathbb{CN}(0,1) \text{ and is independent of } y_i)
\end{aligned}$$

We bound the second and fourth moments of y_i using the below claim, whose proof we defer to Section 3.3.2.

Claim 3.3.3. *For all $i \in [n]$,*

$$\mathbb{E}[|y_i|^2] = \frac{n}{\operatorname{tr}(A)} v_i^\dagger V^\dagger V v_i,$$

$$\mathbb{E}[|y_i|^4] \le \frac{1.09n^2}{\operatorname{tr}(A)^2} \|v_i\|_{\ell_2}^2.$$

Plugging in the bounds from Claim 3.3.3 and summing over all i, we get

$$\begin{aligned}
\frac{1}{n}\sum_{i\in[n]} \mathbb{E}_x[\ln |\langle v_i, x\rangle|^2] &\ge -\gamma + \ln(1-\beta) + \frac{\beta}{(1-\beta)\operatorname{tr}(A)} \sum_{i\in[n]} v_i^\dagger V^\dagger V v_i \\
&\quad - \frac{0.273\beta^2 n}{(1-\beta)^2 \operatorname{tr}(A)^2} \sum_{i\in[n]} \|v_i\|_{\ell_2}^2 \\
&= -\gamma + \ln(1-\beta) + \frac{\beta}{1-\beta} \cdot \frac{\|A\|_F^2}{\operatorname{tr}(A)} - \frac{0.273\beta^2}{(1-\beta)^2} \cdot \frac{n}{\operatorname{tr}(A)} \\
&\ge -\gamma + \ln(1-\beta) + \frac{\beta}{1-\beta} \cdot \frac{\operatorname{tr}(A)}{n} - \frac{0.273\beta^2}{(1-\beta)^2} \cdot \frac{n}{\operatorname{tr}(A)}.
\end{aligned}$$

The last inequality uses $\|A\|_F^2 \ge \operatorname{tr}(A)^2/n$, by Jensen's inequality. This completes the proof of Lemma 3.1.4.

Proof of Claim 3.3.3

Proof. We can directly compute

$$\begin{aligned}
\frac{\operatorname{tr}(A)}{n}\mathbb{E}[|y_i|^2] &= \sum_{j,k\in[n]} \mathbb{E}[s_j \overline{s_k}] \cdot \langle v_i, v_j\rangle \overline{\langle v_i, v_k\rangle} \\
&= \sum_{j\in[n]} |\langle v_i, v_j\rangle|^2 = v_i^\dagger V^\dagger V v_i \qquad (s_j \text{ is independent from } s_k \text{ for } j \ne k)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{\operatorname{tr}(A)^2}{n^2}\mathbb{E}[|y_i|^4] &= \sum_{j,k,l,m\in[n]} \mathbb{E}[s_j s_k \overline{s_l s_m}]\langle v_i, v_j\rangle\langle v_i, v_k\rangle\overline{\langle v_i, v_l\rangle\langle v_i, v_m\rangle} \\
&= \sum_{j\in[n]} |\langle v_i, v_j\rangle|^4 + 2\sum_{j\neq k} |\langle v_i, v_j\rangle\langle v_i, v_k\rangle|^2 \\
&= 2\left(\sum_{j\in[n]} |\langle v_i, v_j\rangle|\right)^2 - \sum_{j\in[n]} |\langle v_i, v_j\rangle|^4 \\
&= 2(v_i^\dagger V^\dagger V v_i)^2 - \sum_{j\in[n]} |\langle v_i, v_j\rangle|^4 \\
&\leq 2\|v_i\|_{\ell_2}^4 - \|v_i\|_{\ell_2}^8 && (V^\dagger V \preceq I, \text{ since } A = VV^\dagger \preceq I) \\
&\leq 1.09 \cdot \|v_i\|_{\ell_2}^2 && (x^2 - x^4 \leq 1.09x \text{ for } x \geq 0)
\end{aligned}$$

The second equality is because $\mathbb{E}[s_j s_k \overline{s_l s_m}] = 0$ unless each index appears an equal number of times in $\{j, k\}$ and $\{l, m\}$. □

3.4 Hardness of Approximation

As mentioned in Section 3.1.1, we will first prove Theorem 3.1.5. Later on, in Section 3.4.2, we will use Theorem 3.1.5 to prove Theorem 3.1.2 using an approximation-preserving reduction to the permanent problem.

Our first result is a general inapproximability result for the f-mean version of $\|A\|_{2\to q}$ that is dependent on an appropriate family of gadgets $\{E_k\}$ as defined below.

Theorem 3.4.1. *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and let $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a continuous increasing 2-concave function such that $\lim_{x\to\infty} \frac{f(x)}{x^2} = 0$. Let $\delta, \gamma > 0$. Assume that for all k, there is a matrix $E_k : \mathbb{F}^k \to \mathbb{F}^{d_k}$ satisfying the following:*

1. *$\|E_k\|_{2\to 2} = 1$.*
2. *The entries of E_k have magnitude equal to $\frac{1}{\sqrt{k}}$.*
3. *for all vectors $x \in \mathbb{F}^k$ with $\|x\|_\infty \leq \delta \cdot \|x\|_{\ell_2}$, $[E_k x]_f \leq \gamma \cdot \|x\|_2$.*

Then for all $\epsilon > 0$, it is NP-Hard to distinguish between the following two cases given a matrix $A : \mathbb{F}^m \to \mathbb{F}^n$ with $\|A\|_{2\to 2} \leq 1$.

1. *Completeness: $[A]_{2\to f} = 1$, or*
2. *Soundness: $[A]_{2\to f} \leq \gamma + \epsilon$.*

The proof of the above theorem is in Section 3.4.1 and closely follows the arguments used in [BRS15, BGG+23]. In order to instantiate it, we will have to construct a family of gadgets $\{E_k\}$ that have $\|E_k\|_{2\to 2} = 1$, but at the same time have small $2 \to f$-norm when restricted to "smooth" vectors.

Definition 3.4.2. For $k \geq 1$, let us define $E_k^{(\mathbb{R})} \in \mathbb{R}^{2^k \times k}$ as the matrix whose rows consist of the members of $\frac{1}{\sqrt{k}} \cdot \{-1, +1\}^k$ ordered arbitrarily. Similarly, we define $E_k^{(\mathbb{C})} \in \mathbb{C}^{4^k \times k}$ as the matrix whose rows consist of the members of $\frac{1}{\sqrt{k}} \cdot \{-1, +1, -i, +i\}^k$ ordered arbitrarily.

Observe that these matrices are normalized so that $\|E_k^{(\mathbb{F})}\|_{2\to 2} = 1$. The following Lemma shows that condition 3 of Theorem 3.4.1 is satisfied with $\gamma \approx \gamma_{\mathbb{F},p}$.

Lemma 3.4.3. *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let f be an absolutely continuous 2-concave increasing function. Let $x \in \mathbb{F}^k$, and $E = E_k^{(\mathbb{F})}$. For all $0 < \delta < 1$, if $\|x\|_\infty \leq \delta \|x\|_{\ell_2}$ then*

$$\left[\frac{Ex}{\|x\|_2}\right]_f^f \leq [g]_f^f + C \cdot \left(-\int_0^{C\delta} \min(0, f(u)) + \delta \cdot \left(\max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1)\right)\right),$$

where $g \sim \mathbb{FN}(0,1)$ and $C > 0$ is a universal constant. In particular if $f = f_p$ for some $-1 < p < 2$, we have

$$[Ex]_{f_p} \leq \|x\|_2 \cdot (\gamma_{\mathbb{F},p} + \epsilon_\delta),$$

where $\epsilon_\delta \to 0$ as $\delta \to 0$.

We prove Lemma 3.4.3 in Section A.1. The proof requires a Berry-Esseen type result for test functions of the form $f(\|.\|_2)$ applied to a sum of independent random vectors, which we prove in Section A.2.

With these results in hand, we can now prove Theorem 3.1.5.

Proof of Theorem 3.1.5. We pick δ to be such that Lemma 3.4.3 implies $\|Ex\|_q \leq \|x\|_2 \cdot (\gamma_{\mathbb{F},q} + \epsilon/2)$ for all x satisfying $\|x\|_\infty \leq \delta \|x\|_{\ell_2}$.

We apply Theorem 3.4.1 to the increasing 2-concave function $f = f_p$, gadget family $\{E_k^{(\mathbb{F})}\}$, and parameters δ, $\gamma = \gamma_{\mathbb{F},q} + \epsilon/2$, and $\epsilon/2$. By Lemma 3.4.3, the three conditions are satisfied, implying that it is NP-Hard to distinguish the case that $\|A\|_{2\to q} = 1$ and $\|A\|_{2\to q} \leq \gamma_{\mathbb{F},q} + \epsilon$. □

3.4.1 Proof of Theorem 3.4.1

We will closely follow the arguments used in [BRS15, BGG+23]. The starting point of our reduction will be the following result implicit in [BRS15], which informally says that it is NP-hard to find a sparse vector in a subspace, according to a certain block-wise notion of sparsity.

Theorem 3.4.4. *For all $\epsilon, \delta, \alpha > 0$ and $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, there is a $k = \mathrm{poly}(1/\epsilon, 1/\delta, 1/\alpha)$ such that given a subspace $W \subseteq \mathbb{F}^{n\times k}$ in the form of a projection matrix $P \in \mathbb{F}^{(n\times k)\times(n\times k)}$, it is* NP*-Hard to distinguish between the following:*

- *There is a vector $x \in W$ such that for all $i \in [n]$, the vector $x_i \in \mathbb{F}^k$ is in $\{e_1, \ldots, e_k\}$.*

- *For all vectors $x \in W$ with $\|x\|_{\ell_2}^2 = n$, the set*

$$S = \{i \in [n] : \|x_i\|_{\ell_2} \leq 1/\alpha, \|x_i\|_\infty \geq \delta\}$$

has size at most ϵn.

Let $0 < \epsilon', \alpha \leq 1$ be constant parameters depending on δ, γ, and ϵ that we will specify later. We prove hardness of the $[A]_{2\to f}$ problem by a reduction from the NP-Hard problem described in Theorem 3.4.4 with parameters ϵ', δ, and α. Theorem 3.4.4 implies that there is some $k = \text{poly}(1/\epsilon', 1/\delta, 1/\alpha)$ such that given a projection matrix $P \in \mathbb{F}^{(n\times k)\times(n\times k)}$ for a subspace $W \subseteq \mathbb{F}^{n\times k}$, it is NP-Hard to distinguish between the following:

- There is a vector $x \in W$ such that for all $i \in [n]$, the vector $x_i \in \mathbb{F}^k$ is in $\{e_1, \ldots, e_k\}$.
- For all vectors $x \in W$ with $\|x\|_{\ell_2}^2 = n$, the set

$$S = \{i \in [n] : \|x_i\|_{\ell_2} \leq 1/\alpha, \|x_i\|_\infty \geq \delta\}$$

has size at most ϵn.

Our reduction will map the projection matrix $P \in \mathbb{F}^{(n\times k)\times(n\times k)}$ to the matrix $A = (I_n \otimes E_k) \cdot P$. Note that $A \in \mathbb{F}^{(n\times d_k)\times(n\times k)}$. To analyze the reduction, we must prove completeness and soundness.

Completeness

If there is a vector $x \in W$ such that for all $i \in [n]$, $x_i \in \{e_1, \ldots, e_k\}$, we need to show $[A]_{2\to f} = 1$.

Indeed, we can consider the vector $z := Ax = (E_k \otimes I_n) \cdot Px = (E_k \otimes I_n)x$. We have for all $i \in [n]$, $z_i = E_k x_i$. Since x_i is a standard basis vector, z_i must be equal to some column of E_k. So by Assumption 2, all entries of z have magnitude $1/\sqrt{k}$, implying $[z]_f = f^{-1}(f(\frac{1}{\sqrt{k}})) = \frac{1}{\sqrt{k}}$. Therefore $[A]_{2\to f} \geq \frac{[z]_f}{\|x\|_2} = 1$.

On the other hand, $[A]_{2\to f} \leq \|A\|_{2\to 2} \leq \|P\|_{2\to 2} \cdot \|E_k\|_{2\to 2} \leq 1$ by Claim 3.2.6.

Soundness

Assuming for all $x \in W$ with $\|x\|_{\ell_2}^2 \leq n$, the set

$$S = \{i \in [n] : \|x_i\|_{\ell_2} \leq 1/\alpha, \|x_i\|_\infty \geq \delta\}$$

has size at most $\epsilon' n$, we need to show $[A]_{2\to f} \leq \gamma + \epsilon$.

Let $y \in \mathbb{F}^{n\times k}$ be an arbitrary vector with $\|y\|_2 = 1$, and set $x = \frac{1}{k} \cdot Py$ and $z = Ay = k \cdot (E_k \otimes I_n)x$. Note that because P is a projection matrix, $\|x\|_{\ell_2}^2 = nk \cdot \|x\|_2^2 \leq n\|y\|_2^2 = n$, and $\|z\|_2 \leq \|y\|_2 = 1$. By virtue of the normalization on x, we have $\|x_i\|_{\ell_2} = \|z_i\|_2$ for each block $i \in [n]$.

We must show $[z]_f \leq \gamma + \epsilon$. We will upper bound the contribution of different indices $i \in [n]$ to $[z]_f$ separately. To do this, define the following partition of $[n]$:

$$V_0 := S = \{i \in [n] : \|x_i\|_{\ell_2} \le 1/\alpha, \|x_i\|_\infty \ge \delta\},$$
$$V_1 := \{i \in [n] : \|x_i\|_{\ell_2} \le \alpha, \|x_i\|_\infty \le \delta\alpha\},$$
$$V_2 := \{i \in [n] : \|x_i\|_{\ell_2} \ge \alpha, \|x_i\|_\infty \le \delta\alpha\},$$
$$V_3 := \{i \in [n] : \|x_i\|_{\ell_2} > 1/\alpha\}.$$

For all $u \in \{0,1,2,3\}$, define $z^{(u)} \in \mathbb{F}^{V_u \times d_k}$ as the collection of $z_i \in \mathbb{F}^{d_k}$ for all $i \in V_u$. Note that

$$[z]_f^f = \sum_{u\in\{0,1,2,3\}} \frac{|V_u|}{|V|} \cdot [z^{(u)}]_f^f. \tag{3.10}$$

We will prove bounds on $[z^{(u)}]_f^f$ for $u \in \{0,1,2,3\}$.

For $u = 0$, Claim 3.2.6 applied to z_i implies

$$[z^{(0)}]_f^f = \mathbb{E}_{i\sim V_0}[z_i]_f^f \le \mathbb{E}_{i\sim V_0} f(\|z_i\|_2) = \mathbb{E}_{i\sim V_0} f(\|x_i\|_{\ell_2}) \le f(1/\alpha). \tag{3.11}$$

For $u = 1$, a similar application of Claim 3.2.6 implies

$$[z^{(1)}]_f^f = \mathbb{E}_{i\sim V_1}[z_i]_f^f \le \mathbb{E}_{i\sim V_1} f(\|z_i\|_2) = \mathbb{E}_{i\sim V_1} f(\|x_i\|_{\ell_2}) \le f(\alpha). \tag{3.12}$$

For $u = 2$, we have

$$[z^{(2)}]_f^f = \mathbb{E}_{i\sim V_2}[z_i]_f^f \underset{\text{Assumption 3}}{\le} \mathbb{E}_{i\sim V_2}\left[f(\gamma \cdot \|z_i\|_2)\right] \underset{Claim\ 3.2.6}{\le} f(\gamma\cdot\|z^{(2)}\|_2) \underset{\|z\|_2\le 1}{\le} f\left(\gamma \cdot \sqrt{\frac{|V|}{|V_2|}}\right). \tag{3.13}$$

Finally, for $u = 3$,

$$\begin{aligned}
[z^{(3)}]_f^f &= \mathbb{E}_{i\sim V_3}[z_i]_f^f \\
&\underset{Claim\ 3.2.6}{\le} \mathbb{E}_{i\sim V_3} f(\|z_i\|_2) \\
&= \mathbb{E}_{i\sim V_3}\left[\|z_i\|_2^2 \cdot \frac{f(\|z_i\|_2)}{\|z_i\|_2^2}\right] \\
&\le \sup_{w\ge 1/\alpha} \frac{f(w)}{w^2} \cdot \mathbb{E}_{i\sim V_3}\|z_i\|_2^2 \\
&= \sup_{w\ge 1/\alpha} \frac{f(w)}{w^2} \cdot \|z^{(3)}\|_2^2 \\
&\underset{\|z\|_2^2=1}{\le} \sup_{w\ge 1/\alpha} \frac{f(w)}{w^2} \cdot \frac{|V|}{|V_3|}.
\end{aligned} \tag{3.14}$$

Now we are equipped to bound Eq. (3.10).

$$
\begin{aligned}
[z]_f^f &\le \sum_{u\in\{0,1,2\}} \frac{|V_u|}{|V\setminus V_3|}[z^{(u)}]_f^f + \sup_{w\ge 1/\alpha}\frac{f(w)}{w^2} && \text{(Eqs. (3.10) and (3.14))}\\
&\le f\left(\sqrt{\sum_{u\in\{0,1,2\}} \frac{|V_u|}{|V\setminus V_3|}[z^{(u)}]_f^2}\right) + \sup_{w\ge 1/\alpha}\frac{f(w)}{w^2} && \text{(Jensen's inequality for } x\to f(\sqrt{x}))\\
&\le f\left(\sqrt{\sum_{u\in\{0,1,2\}} \frac{|V_u|}{(1-\alpha^2)|V|}[z^{(u)}]_f^2}\right) + \sup_{w\ge 1/\alpha}\frac{f(w)}{w^2} && (|V_3|\le \alpha^2\cdot|V_0| \text{ by def of } V_3)\\
&\le f\left(\sqrt{\frac{\epsilon'/\alpha^2+\alpha^2+\gamma^2}{(1-\alpha^2)}}\right) + \sup_{w\ge 1/\alpha}\frac{f(w)}{w^2} && \text{(Eqs. (3.11) to (3.13) and } V_0\le\epsilon'|V|)\\
&\le f\left(\frac{\gamma+\sqrt{\epsilon'}/\alpha+\alpha}{\sqrt{1-\alpha^2}}\right) + \sup_{w\ge 1/\alpha}\frac{f(w)}{w^2} && (\sqrt{a+b}\le\sqrt{a}+\sqrt{b},\ f \text{ is monotone})\\
&= f\left(\frac{\gamma+2\alpha}{\sqrt{1-\alpha^2}}\right) + \sup_{w\ge 1/\alpha}\frac{f(w)}{w^2}. && \text{(Setting } \epsilon'=\alpha^4)
\end{aligned}
$$

Using the assumption that f is continuous and $\lim_{x\to\infty} f(x)/x^2 = 0$, we get that the limit of the right hand side as $\alpha \to 0$ is exactly $f(\gamma)$. Therefore, there exists some $\alpha > 0$ independent of n such that $[z]_f \le \gamma + \epsilon$. We choose α in our invocation of Theorem 3.4.4 accordingly, completing the proof that $[A]_{2\to f} \le \gamma + \epsilon$.

3.4.2 Proof of Theorem 3.1.2

In this section we prove Theorem 3.1.2. We use the following lemma which proves that the approximability of the permanent of highly rank-deficient $n \times n$ PSD matrices is essentially the same as the approximability of of the $2 \to 0$ norm.

Lemma 3.4.5. *Let $V \in \mathbb{C}^{n\times d}$. Then,*

$$
c^{\mathbb{C}}_{n,d}\cdot\binom{n+d-1}{d}^{-1}\cdot\|V\|_{2\to 0}^{2n} \le \mathrm{per}(VV^\dagger) \le c^{\mathbb{C}}_{n,d}\cdot\|V\|_{2\to 0}^{2n},
$$

where $c^{\mathbb{C}}_{n,d}$ is defined in Lemma 3.2.16.

Proof. Let $v_1,\dots,v_n$ be the rows of V. For the upper bound, we can write

$$
\begin{aligned}
\mathrm{per}(VV^\dagger) &= \mathbb{E}_{x\sim\mathbb{C}N(0,I)}\prod_{i\in[n]}|\langle x, v_i\rangle|^2\\
&= c_{n,d}\cdot\mathbb{E}_{\|x\|_2=1}\prod_{i\in[n]}|\langle x, v_i\rangle|^2 && \text{(Lemma 3.2.16)}\\
&\le c_{n,d}\cdot\max_{\|x\|_2=1}\prod_{i\in[n]}|\langle x, v_i\rangle|^2\\
&= c_{n,d}\cdot\|V\|_{2\to 0}^{2n}.
\end{aligned}
$$

Next we prove the lower bound. Let $z \in \mathbb{C}^d$ be a vector with $\|z\|_2 = 1$ vector maximizing $\|Vz\|_0 = \prod_{i\in[n]} |\langle z, v_i\rangle|^{1/n}$. We have $zz^\dagger \preceq \|z\|_{\ell_2}^2 \cdot I = d \cdot I$, so $Vzz^\dagger V^\dagger \preceq d \cdot VV^\dagger$. By Lemma 3.2.15, this implies $\mathrm{per}(Vzz^\dagger V^\dagger) \leq d^n \cdot \mathrm{per}(VV^\dagger)$. Since $Vzz^\dagger V^\dagger$ is rank 1, we can compute its permanent as $\mathrm{per}(Vzz^\dagger V^\dagger) = n! \cdot \prod_{i\in[n]} |\langle z, v_i\rangle|^2$.

$$\begin{aligned}
\|Vz\|_0^{2n} &= \max_{\|x\|_2=1} \prod_{i\in[n]} |\langle x, v_i\rangle|^2 \\
&= \frac{1}{n!} \cdot \mathrm{per}(Vzz^\dagger V^\dagger) \\
&\leq \frac{d^n}{n!} \cdot \mathrm{per}(VV^\dagger) \\
&= \binom{n+d-1}{d} \cdot c_{n,d}^{-1} \cdot \mathrm{per}(VV^\dagger).
\end{aligned}$$

□

Proof of Theorem 3.1.2. We start from Theorem 3.1.5 for the case $\mathbb{F} = \mathbb{C}$ and $q = 0$, to get that it is NP-hard to approximate $\|A\|_{2\to 0}$ within a factor of $e^{-\gamma/2+\epsilon/4}$ (recall that $\gamma_{\mathbb{C},0} = e^{-\gamma/2}$ by Fact 3.2.11).

We reduce the problem of approximating $\|A\|_{2\to 0}$ for $A \in \mathbb{C}^{n\times d}$ to approximating the permanent of the positive semidefinite matrix $B = A^{(k)}(A^{(k)})^\dagger$, where $A^{(k)} \in \mathbb{C}^{nk\times d}$ is as in Fact 3.2.8. By Lemma 3.4.5, $\mathrm{per}(B)$ is proportional to $\|A^{(k)}\|_{2\to 0}^{2nk}$ up to a multiplicative error of

$$\binom{nk+d-1}{d} \leq e^{\epsilon kn/2},$$

which holds for $k = O(\frac{d}{n\epsilon^2})$ and $\epsilon > 0$ small enough. Note that the reduction is efficient because k is polynomial in the size of A.

By Fact 3.2.8, we have $\|A^{(k)}\|_{2\to 0}^{2nk} = \|A\|_{2\to 0}^{2nk}$ which is hard to approximate within a factor of $e^{-kn(\gamma-\epsilon/2)}$. Therefore, it is NP-hard to approximate $\mathrm{per}(B)$ within a factor of $e^{-kn(\gamma-\epsilon)}$, where $B \in \mathbb{C}^{kn\times kn}$. □

Chapter 4

Improved Approximation Algorithms for the EPR Hamiltonian

This chapter is based on "Improved Approximation Algorithms for the EPR Hamiltonian" [JN25], which is joint work with Nathan Ju.

4.1 Introduction

For a quantum Hamiltonian H and a (mixed) state ρ, we say that the *energy* of ρ is $\mathrm{tr}(\rho \cdot H)$. In this work, we refer to a state achieving maximum energy as the *ground state*[1]. Note that the energy of the ground state, or the ground energy, is equal to the maximum eigenvalue of H

$$\lambda_{\max}(H) = \max_{\rho \geq 0, \mathrm{tr}(\rho)=1} \mathrm{tr}(\rho \cdot H).$$

While exactly computing the ground energy of local Hamiltonians is computationally infeasible (QMA-hard), *approximating* the ground energy is possibly tractable and is the focus of this work. Let $\mathcal{F}$ be a family of local Hamiltonians. For $\alpha \in [0,1]$, an *α-approximation algorithm* for $\mathcal{F}$ is one that, upon input $H \in \mathcal{F}$, outputs a description of a (mixed) state ρ achieving energy

$$\mathrm{tr}(\rho \cdot H) \geq \alpha \cdot \lambda_{\max}(H).$$

We are interested in the question "what is the supremum of all $\alpha \in [0,1]$ such that there is a polynomial time α-approximation algorithm[2] for $\mathcal{F}$?". We will use $\alpha^*_{\mathcal{F}}$ to denote the answer to this question.

Quantum Max Cut. In this context, the most commonly studied problem is Quantum Max Cut, where the family of Hamiltonians is $\mathcal{F} = \mathrm{QMC}$. Hamiltonians $H \in \mathrm{QMC}$ are

[1] This is a slight deviation from the usual notion of the ground state being the *minimum* energy state. We remark that both the notions are equivalent up to replacing H by $-H$; we use the maximization notion because it is notationally convenient.

[2] For the purposes of this chapter we only consider classical algorithms, but one can also consider quantum algorithms.

parametrized by a symmetric $n \times n$ interaction matrix w with nonnegative entries, and defined as

$$H = \sum_{i<j \in [n]} w_{ij} H_{ij},$$

where H_{ij} is a projection onto the singlet state $\frac{|01\rangle - |10\rangle}{\sqrt{2}}$ on qubits i and j tensored with identity on the rest. This model is equivalent to the well-known quantum Heisenberg model and is a simple QMA-hard example of a 2-local Hamiltonian.

A line of work initiated by Gharibian and Parekh [GP19] has studied approximation algorithms for this problem [PT21, AGM20, PT22, Lee22, Kin23, HTPG24, WCE+24, TRZ+23]. The best known results in the literature are an algorithm showing $\alpha^*_{\text{QMC}} \geq 0.599$ [LP24, JKK+24], and a (conditional) hardness result showing $\alpha^*_{\text{QMC}} \leq 0.956$ [HNP+23], leaving open a large range of possible values of α^*_{QMC}.

Our current understanding suggests that Quantum Max Cut presents two key conceptual barriers in closing this gap.

- First, the Hamiltonian terms are antiferromagnetic, encouraging pairs of qubits to magnetize in opposite directions on the Bloch sphere. This is akin to the difficulty already found in classical problems like Max Cut.
- Second, there is a purely quantum question of finding the correct ground state (or approximate ground state) entanglement structure.

Existing algorithmic and hardness techniques seem unlikely to fully capture the interplay between these two obstacles. In order to isolate the second challenge, King [Kin23] proposed a new simpler Hamiltonian, known as the EPR Hamiltonian.

EPR Hamiltonian. The EPR Hamiltonian family $\mathcal{F} = \text{EPR}$ is again parametrized by a symmetric $n \times n$ interaction matrix w with nonnegative entries, and is defined as

$$H = \sum_{i<j \in [n]} w_{ij} E_{ij},$$

where E_{ij} is a projection onto the EPR state $|\text{EPR}\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$ on qubits i and j tensored with identity on the rest. Notice that the Hamiltonian terms E_{ij} are ferromagnetic; for example, the product state achieving optimal energy for any EPR Hamiltonian is the trivial state $\rho = |0^n\rangle \langle 0^n|$. In contrast, for QMC Hamiltonians, computing even a 0.956-approximation to the optimal product state is computationally hard [HNP+23].

King [Kin23] gave an algorithm showing $\alpha^*_{\text{EPR}} \geq 1/\sqrt{2} \approx 0.707$, which was later slightly improved to 0.72 [JKK+24]. However, no hardness result are currently known and whether α^*_{EPR} is strictly less than 1 is currently open.

Our main result is an improved approximation algorithm for EPR:

Theorem 4.1.1. *$\alpha^*_{EPR} \geq \frac{1+\sqrt{5}}{4} \approx 0.809$, i.e., there is a polynomial time $\frac{1+\sqrt{5}}{4}$-approximation algorithm for the EPR problem.*

Bipartite QMC. We call an instance of QMC or EPR *bipartite* if its interaction matrix w corresponds to the adjacency matrix of a bipartite graph. In other words, the n qubits can be partitioned into two disjoint sets $[n] = A \cup B$ such that $w_{ij} = 0$ for all pairs $i, j \in A$ and pairs $i, j \in B$, meaning interactions occur only between qubits in different sets. Bipartite instances of QMC and EPR are equivalent up to a Pauli Y applied to every qubit of the subset A (or B). As a result, our algorithm also applies to bipartite QMC instances, improving upon the previously best known 3/4-approximation that one can infer from the work of Lee and Parekh [LP24].

Corollary 4.1.2. *There is a polynomial time $\frac{1+\sqrt{5}}{4}$-approximation algorithm for bipartite instances of the Quantum Max Cut problem.*

4.2 Algorithm

4.2.1 An upper bound on $\lambda_{\max}(H)$

We begin by describing an efficiently computable upper bound on $\lambda_{\max}(H)$. Let us denote by $\mathrm{LP}(w)$ the optimal value of the following linear program, known as the *fractional matching LP*:

$$\begin{aligned} \max & \sum_{i<j} w_{ij} \cdot x_{ij} \\ \text{s.t.} & \sum_{j \neq i} x_{ij} \leq 1 \quad \text{for all } i \in [n] \\ & x_{ij} \geq 0 \quad \text{for all } i, j \in [n] \end{aligned}$$

Let $x = (x_{ij})_{i<j}$ be an optimal solution to the above LP. We say that the *LP energy* on edge $\{i, j\}$ is $\tilde{E}_{ij} = \frac{1+x_{ij}}{2}$.

The *star bound*, proved by Anshu-Gosset-Morenz [AGM20], gives a way to bound the ground state energy $\lambda_{\max}(H)$ in terms of the above linear program. We will need a version of this bound that applies to the EPR Hamiltonian (appearing in [Kin23, Lemma 7] for example).

Lemma 4.2.1. *For any n-qubit state ρ and $i \in [n]$,*

$$\sum_{j \neq i} \max(0, 2\,\mathrm{tr}(\rho \cdot E_{ij}) - 1) \leq 1.$$

As a consequence, $\{\max(0, 2\,\mathrm{tr}(\rho \cdot E_{ij}) - 1)\}_{i<j}$ always gives a feasible solution to the LP, implying that $\sum_{i<j} w_{ij} \max(0, 2\,\mathrm{tr}(\rho \cdot E_{ij}) - 1) \leq \mathrm{LP}(w)$. This implies the upper bound

$$\lambda_{\max}(H) = \max_{\rho \succeq 0, \mathrm{tr}(\rho)=1} \mathrm{tr}(\rho \cdot H) \leq \frac{\sum_{i<j} w_{ij} + \mathrm{LP}(w)}{2} = \sum_{i<j} w_{ij} \tilde{E}_{ij}. \tag{4.1}$$

All the algorithms we present in this chapter involve computing an optimal solution x to $\mathrm{LP}(w)$, and applying a *rounding algorithm* to x to compute the description of a state ρ achieving energy

$$\mathrm{tr}(\rho \cdot E_{ij}) \geq \alpha \cdot \tilde{E}_{ij} \tag{4.2}$$

on every edge $\{i,j\} \in \binom{[n]}{2}$ for some $\alpha > 0$. By Eq. (4.1), we have

$$\mathrm{tr}(\rho \cdot H) = \sum_{i<j} w_{ij} \cdot \mathrm{tr}(\rho \cdot E_{ij}) \geq \alpha \cdot \sum_{i<j} w_{ij} \tilde{E}_{ij} \geq \alpha \cdot \lambda_{\max}(H),$$

so this results in an α-approximation algorithm.

Let us describe the rough plan for our presentation of this rounding algorithm. First, in Section 4.2.2, we describe how to use a rounding strategy used in the work of Lee and Parekh [LP24] to achieve $\alpha = 3/4$ in the special case that the interaction matrix w is bipartite. Next, in Section 4.2.3 we give an improved rounding algorithm achieving $\alpha = \frac{1+\sqrt{5}}{4}$ for bipartite instances. Finally, in Section 4.2.4 we show how to achieve the same approximation ratio $\alpha = \frac{1+\sqrt{5}}{4}$ for general non-bipartite instances, proving Theorem 4.1.1.

4.2.2 A 3/4-approximation in the bipartite case

In the bipartite case, it is well known that the fractional matching LP is integral, i.e., its vertices are indicator vectors of matchings [Edm65].

Lemma 4.2.2. *For any weighted bipartite graph with adjacency matrix w, $\mathrm{LP}(w) = w(M)$, where M is the maximum weight matching of the graph. That is, the optimal solution to $\mathrm{LP}(w)$ is achieved by an integer vector $x_{ij} = \mathbf{1}_{\{\{i,j\} \in M\}}$.*

We can find such an optimal solution using an algorithm for maximum weighted matching. Let $U \subseteq [n]$ be the set of vertices that do not appear in the matching M. Consider the two states

$$\rho_{\mathrm{prod}} = |0^n\rangle\langle 0^n|, \quad \rho_{\mathrm{match}} = \left(\bigotimes_{\{i,j\} \in M} (|\mathrm{EPR}\rangle\langle \mathrm{EPR}|)_{i,j} \right) \otimes \left(\bigotimes_{i \in U} \frac{I_i}{2} \right).$$

The rounding algorithm will output a random choice between ρ_{prod} and ρ_{match}. In particular, we output the state $\rho = (1-p) \cdot \rho_{\mathrm{prod}} + p \cdot \rho_{\mathrm{match}}$ for some choice of $p \in [0,1]$.

Claim 4.2.3. *The energy attained on edge $\{i,j\}$ is*

$$\mathrm{tr}(\rho \cdot E_{ij}) = \begin{cases} \frac{1}{2} + \frac{p}{2} & \text{if } \{i,j\} \in M, \\ \frac{1}{2} - \frac{p}{4} & \text{if } \{i,j\} \notin M. \end{cases}$$

On the other hand, the LP energy on edge $\{i,j\}$ is

$$\tilde{E}_{ij} = \frac{1 + x_{ij}}{2} = \begin{cases} 1 & \text{if } \{i,j\} \in M, \\ \frac{1}{2} & \text{if } \{i,j\} \notin M. \end{cases}$$

If we set $p = 1/2$, then one can check that for each edge $\{i, j\}$,

$$\mathrm{tr}(\rho \cdot E_{ij}) = \frac{3}{4}\tilde{E}_{ij},$$

proving Eq. (4.2) for $\alpha = \frac{3}{4}$.

4.2.3 An improvement in the bipartite case

Our improvement to this algorithm is simple to state – we interpolate between ρ_{prod} and ρ_{match} using quantum superposition rather than in probability. Concretely, let M be the maximum matching and U be the unmatched vertices as in Section 4.2.2. Define the tilted EPR state $|\mathrm{EPR}_\theta\rangle := \sqrt{\theta}\,|00\rangle + \sqrt{1-\theta}\,|11\rangle$. We will output the state

$$\rho = \left(\bigotimes_{\{i,j\}\in M} (|\mathrm{EPR}_\theta\rangle\langle\mathrm{EPR}_\theta|)_{i,j}\right) \otimes \left(\bigotimes_{i\in U} (\theta\,|0\rangle\langle 0| + (1-\theta)\,|1\rangle\langle 1|)_i\right). \tag{4.3}$$

Claim 4.2.4. *The energy attained on edge $\{i, j\}$ is*

$$\mathrm{tr}(\rho \cdot E_{ij}) = \begin{cases} \frac{1}{2} + \sqrt{\theta(1-\theta)} = \frac{1}{2} + \gamma & \text{if } \{i,j\} \in M, \\ \frac{1}{2} - \theta(1-\theta) = \frac{1}{2} - \gamma^2 & \text{if } \{i,j\} \notin M, \end{cases}$$

where $\gamma = \sqrt{\theta(1-\theta)}$.

Comparing Claim 4.2.4 to Claim 4.2.3, one can readily see that this algorithm performs strictly better; for instance, set $\gamma = p/2$. Our new rounding scheme achieves strictly better energy on edges outside M and identical energy on edges inside M.

It remains to prove Claim 4.2.4, and then to analyze the approximation ratio of this rounding algorithm. We will repeatedly use the following simple properties of $|\mathrm{EPR}_\theta\rangle$.

Claim 4.2.5. *Let $\gamma = \sqrt{\theta(1-\theta)}$. The following are true.*

1. $|\langle\mathrm{EPR}\,|\,\mathrm{EPR}_\theta\rangle|^2 = \frac{1}{2} + \gamma$.
2. $\mathrm{tr}_{[1]}\,|\mathrm{EPR}_\theta\rangle\langle\mathrm{EPR}_\theta| = \mathrm{tr}_{[2]}\,|\mathrm{EPR}_\theta\rangle\langle\mathrm{EPR}_\theta| = \theta\,|0\rangle\langle 0| + (1-\theta)\,|1\rangle\langle 1|$.
3. $\langle\mathrm{EPR}|\,(\theta\,|0\rangle\langle 0| + (1-\theta)\,|1\rangle\langle 1|)^{\otimes 2}\,|\mathrm{EPR}\rangle = \frac{1}{2} - \gamma^2$.

Now, one can compute all the 2-qubit marginals:

$$\mathrm{tr}_{[n]\setminus\{i,j\}}(\rho) = \begin{cases} |\mathrm{EPR}_\theta\rangle\langle\mathrm{EPR}_\theta| & \text{if } \{i,j\} \in M, \\ (\theta\,|0\rangle\langle 0| + (1-\theta)\,|1\rangle\langle 1|)^{\otimes 2} & \text{if } \{i,j\} \notin M, \end{cases}$$

where for the second case we used that the 1-qubit marginals are all equal to $\mathrm{tr}_{[n]\setminus\{1\}}[|\mathrm{EPR}_\theta\rangle\langle\mathrm{EPR}_\theta|] = \theta\,|0\rangle\langle 0| + (1-\theta)\,|1\rangle\langle 1|$. By Claim 4.2.5, the energy achieved by each term $\{i, j\}$ is determined by

$$\langle\mathrm{EPR}|\,\mathrm{tr}_{[n]\setminus\{i,j\}}(\rho)\,|\mathrm{EPR}\rangle = \begin{cases} \frac{1}{2} + \sqrt{\theta(1-\theta)} = \frac{1}{2} + \gamma & \text{if } \{i,j\} \in M, \\ \frac{1}{2} - \theta(1-\theta) = \frac{1}{2} - \gamma^2 & \text{if } \{i,j\} \notin M, \end{cases}$$

and Claim 4.2.4 follows immediately. Now let us analyze the approximation ratio. Similarly to Section 4.2.2, we have the LP energy

$$\tilde{E}_{ij} = \begin{cases} 1 & \text{if } \{i,j\} \in M, \\ \frac{1}{2} & \text{if } \{i,j\} \notin M. \end{cases} \tag{4.4}$$

Setting $\gamma = \frac{\sqrt{5}-1}{4}$, one can check using Eq. (4.4) and Claim 4.2.4 that

$$\operatorname{tr}(\rho \cdot E_{ij}) = \frac{1+\sqrt{5}}{4} \cdot \tilde{E}_{ij}$$

for each edge $\{i,j\} \in \binom{[n]}{2}$. Thus, we have shown Eq. (4.2) with $\alpha = \frac{1+\sqrt{5}}{4}$ as desired.

4.2.4 General Case

In the general non-bipartite case, we cannot apply Lemma 4.2.2; the fractional matching LP is not necessarily integral. Previous works proceeded by bounding the integrality gap of this LP, which is quite lossy. Our main idea is to take advantage of the fact that the fractional matching LP is *half-integral* [LP09, Theorem 7.5.1].

Lemma 4.2.6. *For any weighted n-vertex graph with adjacency matrix w, there is a vertex-disjoint collection of edges $M \subset \binom{[n]}{2}$ and cycles $\mathcal{C} \subset 2^{\binom{[n]}{2}}$ such that $\mathrm{LP}(w) = w(M) + \frac{1}{2}\sum_{C\in\mathcal{C}} w(C)$. That is, the optimal solution to $\mathrm{LP}(w)$ is achieved by the half-integer vector $x_{ij} = \mathbf{1}_{\{\{i,j\}\in M\}} + \frac{1}{2}\sum_{C\in\mathcal{C}} \mathbf{1}_{\{\{i,j\}\in C\}}$.*

In particular, (e.g. using the algorithm of Anstee [Ans87]), one can efficiently find such a solution x, i.e. a vertex-disjoint collection of edges $M \subset \binom{[n]}{2}$ and odd length cycles $\mathcal{C}$ such that the LP energies can be written as

$$\tilde{E}_{ij} = \frac{1+x_{ij}}{2} = \begin{cases} 1 & \text{if } \{i,j\} \in M, \\ \frac{3}{4} & \text{if } \{i,j\} \in C \text{ for some } C \in \mathcal{C}, \\ \frac{1}{2} & \text{otherwise .} \end{cases} \tag{4.5}$$

Let U be the set of vertices absent from M and $\mathcal{C}$. The algorithm in the bipartite case (Eq. (4.3)) already suggests what quantum state we will output on the qubits involved in M and U. On the qubits in C for some odd cycle $C \in \mathcal{C}$ of length k, we will output a high energy state for the EPR Hamiltonian on the length k cycle, i.e. $H_{\text{cycle}} = \sum_{i\in[k]} E_{i,i+1(\text{mod } k)}$, subject to the constraint that all 1-qubit marginals are equal to $\theta\,|0\rangle\langle 0| + (1-\theta)\,|1\rangle\langle 1|$. The following lemma shows that there is a way to achieve sufficiently large energy on the cycle edges.

Lemma 4.2.7. *For any integer k, there is an efficient algorithm to compute the description of a k-qubit state ρ_k satisfying the following, where $\gamma = \frac{\sqrt{5}-1}{4}$ and $\theta \geq 1/2$ is the solution to $\sqrt{\theta(1-\theta)} = \gamma$.*

1. *For all* $i \in [k]$, $\mathrm{tr}(\rho_k \cdot E_{i,i+1(\mathrm{mod\ k})}) > \frac{3}{4} \cdot \frac{1+\sqrt{5}}{4}$.
2. *For all* $i \in [k]$, *the* 1*-qubit marginal* $\mathrm{tr}_{[k]-i}[\rho_k]$ *equals* $\theta \left|0\right\rangle \left\langle 0\right| + (1-\theta) \left|1\right\rangle \left\langle 1\right|$.
3. *For all* $i, j \in [k]$, $\mathrm{tr}(\rho_k \cdot E_{ij}) > \frac{1}{2} \cdot \frac{1+\sqrt{5}}{4}$.

We will provide a computer-assisted proof of Lemma 4.2.7 later on in Section B.1. We remark that bounds Items 1 and 3 above are loose; for ease of exposition we have chosen to use the weakest bounds that suffice to attain the required approximation ratio.

Our algorithm will output the state

$$\rho = \left(\bigotimes_{C \in \mathcal{C}} (\rho_{|C|})_{V(C)} \right) \otimes \left(\bigotimes_{\{i,j\} \in M} (\left|\mathrm{EPR}_\theta\right\rangle \left\langle \mathrm{EPR}_\theta\right|)_{i,j} \right) \otimes \left(\bigotimes_{i \in U} (\theta \left|0\right\rangle \left\langle 0\right| + (1-\theta) \left|1\right\rangle \left\langle 1\right|)_i \right), \tag{4.6}$$

where $V(C) \subseteq [n]$ is the set of qubits in the cycle C and $\rho_{|C|}$ is the state guaranteed by Lemma 4.2.7. Similarly to Section 4.2.3, we set $\gamma = \frac{\sqrt{5}-1}{4}$ and $\theta \geq 1/2$ such that $\sqrt{\theta(1-\theta)} = \gamma$. Let us analyze the energy achieved by this algorithm. We have

$$\mathrm{tr}(\rho \cdot E_{ij}) \geq \begin{cases} \frac{1+\sqrt{5}}{4} & \text{if } \{i,j\} \in M, \\ \frac{3}{4} \cdot \frac{1+\sqrt{5}}{4} & \text{if } \{i,j\} \in C \text{ for some } C \in \mathcal{C}, \\ \frac{1}{2} \cdot \frac{1+\sqrt{5}}{4} & \text{otherwise.} \end{cases}$$

The three cases use Claim 4.2.4, Lemma 4.2.7 and Item 1, and the 1-qubit marginal condition together with Lemma 4.2.7 and Item 3, respectively. To elaborate on the last case, we note that it must be that either qubits i and j are nonadjacent vertices in the *same* cycle $C \in \mathcal{C}$, or they are in tensor product. If the former is true, we use the guarantee of Lemma 4.2.7, Item 3. Otherwise, we use the fact that all the 1-qubit marginals are equal to $\theta \left|0\right\rangle \left\langle 0\right| + (1-\theta) \left|1\right\rangle \left\langle 1\right|$ along with the calculations from Section 4.2.3 to conclude that the energy is $\frac{1}{2} - \gamma^2 = \frac{1+\sqrt{5}}{8}$.

Comparing with Eq. (4.5), we immediately get that for every edge $\{i,j\} \in \binom{[n]}{2}$,

$$\mathrm{tr}(\rho \cdot E_{ij}) \geq \frac{1+\sqrt{5}}{4} \tilde{E}_{ij},$$

proving Eq. (4.2) with $\alpha = \frac{1+\sqrt{5}}{4}$. As a result, we obtain an approximation algorithm for the EPR Hamiltonian with approximation ratio $\frac{1+\sqrt{5}}{4}$, proving Theorem 4.1.1.

Part III

Average-Case Algorithms

Chapter 5

Overview of Average-Case Algorithms

The second part of the dissertation studies average-case computation, where instances are sampled from specified probability models rather than chosen adversarially. The chapters use two different notions of algorithmic success. In a certification problem, the algorithm is evaluated on typical instances from a random model, but any certificate it outputs must be sound for the particular instance it was given. In a distinguishing problem, the output is itself a statistical decision: the algorithm receives a sample from one of two distributions, often a null model and a planted model, and its performance is measured by its advantage over random guessing.

5.1 Sparse-graph certification

Chapter 6 studies certification for random d-regular graphs. For a graph G, let $\mathrm{MC}(G)$ denote the fraction of edges in a maximum cut, and let $\mathrm{IS}(G)$ denote the fraction of vertices in a maximum independent set. The certification problem asks for the smallest numbers σ_d^{MC} and σ_d^{IS} such that, with high probability over $G \sim \mathcal{G}(n,d)$, a polynomial-time algorithm can certify

$$\mathrm{MC}(G) \le \sigma_d^{\mathrm{MC}}, \qquad \mathrm{IS}(G) \le \sigma_d^{\mathrm{IS}}.$$

Lower bounds for this problem can be obtained by the quiet-planting framework of Kunisky and Yu [KY24], which conjectures that spectral information is the only computationally accessible information about the input random graph. More precisely, such a graph passes the same basic spectral tests as a typical random regular graph, and the Kunisky–Yu conjecture asserts that random large lifts of these Ramanujan graphs cannot be efficiently distinguished from a random d-regular graph. As a consequence, any finite d-regular Ramanujan graph with an unusually large cut or independent set provides a lower bound for this question.

The chapter strengthens this obstruction by searching for better finite Ramanujan witnesses. The search uses AlphaEvolve [NVE+25], but the mathematical certificate is not the behavior of the search procedure. Each accepted object is an explicit pair (G,S), where G is checked to be Ramanujan and S is a displayed cut or independent set. Thus the computation produces rigorous lower bounds on the extremal Ramanujan quantities γ_d^{MC} and γ_d^{IS}. The

resulting bounds are stated in Theorem 6.1.7; the comparison with previous lower bounds appears in Table 6.1, and one of the discovered witnesses is shown in Figure 6.1.

The upper-bound side asks how much can be certified from efficiently checkable global information. Friedman's theorem [Fri08] implies that a random d-regular graph satisfies

$$\lambda^*(G) \le 2\sqrt{d-1} + o(1)$$

with high probability, where $\lambda^*(G)$ is the largest nontrivial adjacency eigenvalue in absolute value. It is therefore enough to prove a deterministic statement: every d-regular graph satisfying this spectral condition has bounded $\mathrm{MC}(G)$ and $\mathrm{IS}(G)$. Hoffman's classical argument [Hof03, Hae21] uses the quadratic form $\mathbf{x}^\top \mathbf{A}\mathbf{x}$. The chapter refines this by also using longer-walk correlations $\mathbf{x}^\top \mathbf{A}^L \mathbf{x}$.

The resulting certificate is local. A cut or independent set induces a distribution on labelings of the depth-L d-regular tree, obtained by looking at the labels seen from a random vertex and following nonbacktracking walks. Spectral expansion imposes linear constraints on this local distribution, and the cut value or independent-set density is also a linear function of it. This gives a finite linear program whose optimum upper-bounds all graphs satisfying the spectral condition. With symmetry reductions and local optimality reductions, these LPs become small enough to solve in the degrees studied here. The resulting certificates are stated in Theorem 6.1.8 and developed in Section 6.2. For several of the small-degree cases, the upper and lower bounds are separated by only a few thousandths.

5.2 Planted clique distinguishers

Chapter 7 studies distinguishing rather than certification. The null distribution is $G(n, 1/2)$, and the planted distribution is obtained by planting a clique of size $k = n^{1/2-\alpha}$. This is the standard planted clique problem [Jer92, Kuc95, AKS98, KV02]. In this regime, the two distributions are statistically distinguishable but are conjectured to be hard to distinguish in polynomial time. The chapter asks for a quantitative version of this conjectural hardness: assuming the planted clique hypothesis, what is the largest advantage that any efficient test can achieve?

There is always a small advantage coming from the edge count: the planted distribution has slightly more edges in expectation than the null distribution. Low-degree polynomial tests predict that this advantage is optimal [Hop18, KWB19], with sharp leading constant $k^2/(\sqrt{\pi}n)$ for the standard planted distribution. The main result, Theorem 7.1.4, turns this prediction into a computational hardness statement under the planted clique hypothesis. The explicit sharp bound is given in Corollary 7.1.5: no polynomial-time test achieves asymptotically larger advantage, while the edge-counting test matches it.

The proof is a hardness self-amplification argument. Suppose there were an efficient test for the standard planted distribution whose advantage exceeded the low-degree benchmark. A vertex-resampling Markov chain maps a planted clique instance with a larger clique to one with a smaller clique while preserving the null distribution. The associated noise operator has an explicit Fourier spectrum. After decomposing the alleged test into its low-degree part and

its orthogonal complement, any advantage beyond the low-degree benchmark must live in the orthogonal part. The noise operator strongly contracts this part under the null distribution, while its planted expectation survives enough to be thresholded. An anticoncentration argument then converts the resulting real-valued statistic into a constant-advantage test for an easier planted clique instance, contradicting the planted clique hypothesis. The proof overview appears in Section 7.3, the general reduction in Section 7.4, and the anticoncentration step in Section 7.6.

The chapter also constructs planted distributions that are much harder than the standard one. The starting point is a low-degree hard-core construction. By choosing acceptance probabilities for samples from the usual planted distribution, one obtains a small perturbation whose low-degree moments match the null distribution up to an arbitrary prescribed error. This is formulated as a linear program over acceptance probabilities, analyzed through its convex dual, and made algorithmic using stochastic gradient descent. The resulting moment-matching theorem is Theorem 7.1.8, with the general hard-core statement proved in Section 7.5. Combining this construction with the perturbation version of the self-amplification theorem gives Theorem 7.1.9: under the planted clique hypothesis, for any fixed D, there is an efficiently samplable planted distribution supported on graphs containing a clique of size $n^{1/2-\alpha}$ that every polynomial-time test distinguishes from $G(n, 1/2)$ with advantage $o(n^{-D})$.

Chapter 6

Average-Case Algorithms for Sparse-Graph Certification

This chapter is based on the sparse-random-graph certification results from "Reinforced Generation of Combinatorial Structures: Hardness of Approximation" [NRT25], which is joint work with Prabhakar Raghavan and Abhradeep Thakurta.

AlphaEvolve. The lower-bound computations in this chapter use AlphaEvolve [NVE+25], an evolutionary coding agent for discovery problems in which candidate solutions can be generated by programs and evaluated automatically. AlphaEvolve maintains a population of programs, executes those programs to produce candidate objects, evaluates the candidates using a user-specified scoring function, and then uses large language models to propose modifications to the most promising programs. The iteration is therefore code-based rather than object-based: the model proposes code, the code is executed, and the resulting object is accepted, rejected, or ranked by a verifier. This structure is well suited to mathematical search problems where the quality of a proposed solution can be checked mechanically.

In this chapter, AlphaEvolve is used as an optimizer for objectives that can be checked automatically. The user specifies the search space and the score: which objects count as feasible, how infeasible objects are penalized, and how feasible objects are compared. AlphaEvolve then repeatedly generates new search programs, preserving useful partial structure from previous attempts while introducing larger code-level changes than a conventional local search over the objects themselves. This is a good fit for the searches here because candidate graphs are hard to find by hand, but once a graph and witness are proposed, their relevant properties can be checked directly.

The Max-Cut lower-bound search illustrates this distinction. The desired lower bound requires a Ramanujan graph G with a large maximum cut, but computing $\mathrm{MC}(G)$ from scratch for every candidate graph would require solving an NP-hard optimization problem at each evaluation step. Instead, we searched over pairs (G, S), where G is a d-regular graph and $S \subseteq V(G)$ is a proposed cut. The verifier checked the structural conditions on G, including regularity and the Ramanujan eigenvalue bound, and then scored the candidate by counting

the fraction of edges crossing the displayed cut:

$$\text{score}(G, S) = \frac{|\delta_G(S, \bar{S})|}{|E(G)|}$$

for valid Ramanujan graphs, with invalid candidates assigned a low score. Thus each accepted candidate was accompanied by an explicit witness S. The score may underestimate $\text{MC}(G)$ if S is not a maximum cut, but it never overestimates it; in particular, it is already a rigorous lower bound on the maximum-cut fraction of G. The same principle applies to the independent-set searches, where the candidate includes both the graph and a proposed independent set. This witness-based scoring was the main computational device: it allowed AlphaEvolve to optimize a quantity relevant to the theorem without repeatedly recomputing the true optimum from scratch.

6.1 Introduction

A central problem in average case complexity is that of *certification*, where our goal is to certify a property of typical samples from a distribution. We focus on the setting of *sparse random graphs*, where the distribution in question is $\mathcal{G}(n, d)$, the uniform distribution over all n-vertex d-regular graphs, where $d \in \mathbb{N}$ is small. Let Ω be the set of n-vertex graphs, and consider a *property* $P : \Omega \to \{0, 1\}$ such that $\Pr_{G \sim \mathcal{G}(n,d)}[P(G) = 1] = 1 - o_n(1)$. We are given as input a graph $G \sim \mathcal{G}(n, d)$, and our algorithmic task is to efficiently certify that G satisfies P with high probability.

Definition 6.1.1. An efficient certificate of P is any property $P' : \Omega \to \{0, 1\}$ that is computable in polynomial time such that the following conditions hold.

1. $P'(G) = 1$ implies $P(G) = 1$.
2. $\Pr_{G \sim \mathcal{G}(n,d)}[P'(G) = 1] = 1 - o_n(1)$.

Here, $\mathcal{G}(n, d)$ is the uniform distribution over all n-vertex d-regular graphs for some fixed constant $d \in \mathbb{N}$.

We focus specifically on the setting of *refutation*, where P is a co-NP property (recall that a problem is in co-NP if the `NO` instance of a decision problem has an efficiently verifiable certificate). This setting has seen considerable activity, leading to innovation in various algorithmic techniques such as spectral and sum-of-squares algorithms, along with the development of a variety of methods to establish hardness (for instance, [BHK$^+$19a, RRS17, BBK$^+$21, KY24, KVWX23]).

In this work, we consider the problem of certifying upper bounds on the Max-Cut (or Max-Independent-Set) of a graph drawn from $\mathcal{G}(n, d)$. For example, for Max-Cut, we study the following question: for what values of σ can we efficiently certify with high probability that a random graph drawn from $\mathcal{G}(n, d)$ has Max-Cut at most $\sigma \times \#(\texttt{of edges})$?

Definition 6.1.2 (Max-Cut fraction, Max-Independent Set fraction)**.** For an undirected graph G, $\text{MC}(G)$ is defined as the fraction of edges in a maximum cut of G. Similarly, $\text{IS}(G)$ is defined as the fraction of vertices in a maximum independent set of G.

Definition 6.1.3. σ_d^{MC} is the infimum over all σ such that the property $\mathbf{1}\{\mathrm{MC}(G) \leq \sigma\}$ has an efficient certificate. Similarly, σ_d^{IS} is the infimum over all σ such that $\mathbf{1}\{\mathrm{IS}(G) \leq \sigma\}$ has an efficient certificate.

There is evidence that this is a nontrivial problem exhibiting a *certification gap*, in the sense that $\mathrm{MC}(G)$ concentrates around a constant value strictly smaller than σ_d^{MC} for $G \sim \mathcal{G}(n,d)$ [BBK+21, KY24]. These represent the structurally most elementary refutation-style properties (involving only binary variables and pairwise constraints) that exhibit interesting complexity-theoretic phenomena. In this work, we focus on accurately characterizing the value of σ_d^{MC} and σ_d^{IS}. We remark that the asymptotics of these quantities as $d \to \infty$ are already understood (thanks to [BBK+21, KY24]). In service of a precise understanding of this problem in all regimes, we focus on the least understood case where d is small; specifically we study $d \in \{3,4\}$.

Towards this goal, [KY24] introduced a conjecture[1] that relates σ_d^{MC} and σ_d^{IS} to a fundamental question in spectral graph theory, specifically about Ramanujan graphs (Definition 6.1.4 below), which are (in a spectral sense) the most "random" looking regular graphs constructed via a deterministic process.

The intuition behind this framework relies on the strategy of "quiet planting". Specifically, if one can exhibit a Ramanujan graph failing to satisfy a certain property P (e.g., a small Max-Cut value), then under the hardness assumption of [KY24, Conjecture 1.6], it is impossible for a polynomial-time algorithm to certify P on the random graph distribution $\mathcal{G}(n,d)$. In the context of Max-Cut, if we construct a Ramanujan graph with a maximum cut of at least $\gamma_d^{\mathrm{MC}} \times \#(\texttt{edges})$, then the certification bound σ_d^{MC} must essentially be at least γ_d^{MC}, where γ_d^{MC} is defined in Definition 6.1.5 below. Therefore, the objective in establishing stronger lower bounds for certification is to maximize this value γ_d^{MC} over the set of Ramanujan graphs. We employ AlphaEvolve to construct such extremal Ramanujan graphs for both Max-Cut and Max-Independent-Set. The precise conjecture we operate with is stated in Conjecture 6.1.6.

Definition 6.1.4 (Ramanujan graphs)**.** Let G be a d-regular multigraph on n vertices with adjacency matrix A. Suppose the eigenvalues of A are $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. We say that G is Ramanujan if $\max_{i>1} |\lambda_i| \leq 2\sqrt{d-1}$.

Definition 6.1.5. Define γ_d^{MC} (resp. γ_d^{IS}) as the supremum of $\mathrm{MC}(G)$ (resp. $\mathrm{IS}(G)$) over all d-regular Ramanujan graphs G.

γ_d^{MC} (resp. γ_d^{IS}) can be thought of as the performance of the optimal "spectral" certificate for bounds on Max-Cut (resp. Max-Independent-Set). Kunisky and Yu formulate the following conjecture [KY24].

Conjecture 6.1.6. $\sigma_d^{\mathrm{MC}} \geq \gamma_d^{\mathrm{MC}}$ and $\sigma_d^{\mathrm{IS}} \geq \gamma_d^{\mathrm{IS}}$ for all $d \geq 3$.

[KY24] exhibited finite d-regular Ramanujan graphs for $d \in \{3,4\}$, yielding concrete lower bounds on σ_d^{MC} and σ_d^{IS} that are strong enough to prove a certification gap. On the

[1] We refer the reader to [KY24] for supporting evidence and intuition.

other hand, there was a large gap between the best known upper bound [Hof03, Hae21] and their lower bounds (see Table 6.1 for a comparison). Our results make progress towards a more precise understanding of this question. Firstly, we use AlphaEvolve to construct better d-regular Ramanujan graphs, exhibiting stronger lower bounds.

Theorem 6.1.7 (Lower bound)**.** *We have the lower bounds* $\gamma_4^{\mathrm{MC}} \geq 113/124$*,* $\gamma_3^{\mathrm{IS}} \geq 17/36$*, and* $\gamma_4^{\mathrm{IS}} \geq 74/163$*. Under Conjecture 6.1.6, this implies the following computational hardness results:* $\sigma_4^{\mathrm{MC}} \geq 113/124 > 0.911$*,* $\sigma_3^{\mathrm{IS}} \geq 17/36 > 0.472$*, and* $\sigma_4^{\mathrm{IS}} \geq 74/163 > 0.453$*.*

This Theorem improves upon the results of [KY24], who prove $\gamma_4^{\mathrm{MC}} \geq 7/8$, $\gamma_3^{\mathrm{IS}} \geq 11/24$, and $\gamma_4^{\mathrm{IS}} \geq 3/7$. An explicit description of the constructions found by AlphaEvolve can be found in Appendix C.1.1, and a figure of the construction achieving the γ_4^{MC} bound is presented in Figure 6.1. We complement these hardness results with an algorithmic improvement over Hoffman's classic bound [Hof03, Hae21].

Theorem 6.1.8 (Upper bound)**.** *We have* $\sigma_3^{\mathrm{MC}} \leq 0.953$*,* $\sigma_4^{\mathrm{MC}} \leq 0.916$*,* $\sigma_3^{\mathrm{IS}} \leq 0.476$*, and* $\sigma_4^{\mathrm{IS}} \leq 0.457$*.*

Proof ideas for the Max-Cut upper bound. Similar to existing bounds, we will use the spectral bound $\max_{i>1} |\lambda_i| \lesssim 2\sqrt{d-1}$, that is satisfied with high probability when $G \sim \mathcal{G}(n,d)$. Here $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the eigenvalues of the adjacency matrix $\mathbf{A}$ of G.

Consider a cut $(S, \bar{S})$ in G such that α fraction of edges cross $(S, \bar{S})$. Hoffman's bound proceeds by constructing the vector $\mathbf{x} = (2 \cdot \mathbf{1}\{i \in S\} - 1, i \in [n])$. The spectral bound above implies an upper bound on $|\mathbf{x}^\top \mathbf{A}\mathbf{x}|$, which in turn constrains α. We generalize this approach by considering inequalities involving new higher-order quadratic forms of the form $\mathbf{x}^\top \mathbf{A}^L \mathbf{x}$, where $L \in \mathbb{N}$. It is much less clear how bounds on $\mathbf{x}^\top \mathbf{A}^L \mathbf{x}$ imply better bounds on α; we show that the strongest bound one can derive on α can be captured by a linear program over probability distributions on $\{\pm 1\}$-labelings of a d-regular tree of depth L. Such a distribution is meant to model the local view of $(S, \bar{S})$ experienced by a random vertex. We solve this LP for small fixed values of $L > 1$, finding that one can improve on Hoffman's bound (corresponding to $L = 1$). □

We prove this result in Appendix C.1.2. For reference, we have collected our results and how they compare to previous results in Table 6.1. Notably, we obtain upper and lower bounds on σ_4^{MC}, σ_3^{IS}, and σ_4^{IS} that match up to a small absolute error of 0.005, so both Theorems 6.1.7 and 6.1.8 are nearly optimal for these cases. In particular, this raises the exciting possibility that one might be able to obtain clues toward complete characterizations of σ_d^{MC} and σ_d^{IS} by studying the proofs of Theorems 6.1.7 and 6.1.8.

Comments on finding near-optimal lower bounds with AlphaEvolve. We conclude this section with remarks on our methodology in Theorem 6.1.7, and our results. The constructions given in [KY24] for $d \in \{3, 4\}$ are graphs on up to 12 vertices that were generated by computer-assisted experimentation. We were able to replicate their results by randomly sampling a large number of regular graphs, albeit with the post-hoc knowledge of the number of self-loops in their construction.

	LB ([KY24])	LB (Theorem 6.1.7)	UB (Theorem 6.1.8)	UB ([Hof03, Hae21])
Max-Cut, $d = 3$	0.944	0.944	**0.953**	0.971
Max-Cut, $d = 4$	0.875	**0.911**	**0.916**	0.933
Max-Independent-Set, $d = 3$	0.458	**0.472**	**0.476**	0.485
Max-Independent-Set, $d = 4$	0.428	**0.454**	**0.457**	0.464

Table 6.1: Comparison of hardness (LB) and algorithms (UB) for certifying upper bounds on Max-Cut and Max-Independent-Set for $\mathcal{G}(n, d)$, in terms of lower bounds and upper bounds on σ_d^{MC} or σ_d^{IS} (see Definition 6.1.3). The LB results are conditional on Conjecture 6.1.6. Improvements are shown in boldface.

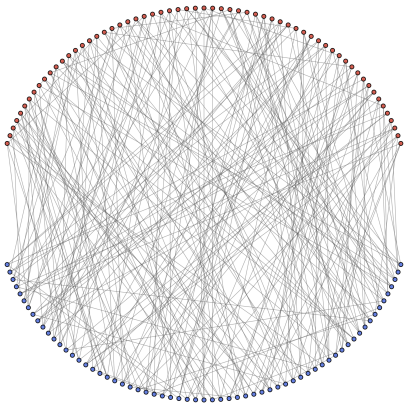


Figure 6.1: 4-regular Ramanujan graph found by AlphaEvolve for the lower bound on γ_4^{MC} in Theorem 6.1.7.

Improving their lower bounds necessitates constructions on many more vertices, as the granularity of $\mathrm{MC}(G)$ or $\mathrm{IS}(G)$ is limited by the size of G. At our target scale, the approach of randomly sampling constructions does not work for two reasons: (1) the space of d-regular n-vertex graphs blows up combinatorially, meaning random sampling will not find interesting "extremal" graphs, and (2) the complexity of computing $\mathrm{MC}(G)$ or $\mathrm{IS}(G)$ is exponential in n, so it is hard to even compute these bounds for a construction.

As stated before, we used AlphaEvolve as a tool to search for finite Ramanujan graphs witnessing the bounds in Theorem 6.1.7. We circumvent the above limitation by searching over the space $\mathcal{S} = \bigcup_{n=1}^{\infty} \Omega_n \times 2^{[n]}$, where Ω_n is the set of n-vertex d-regular graphs. That is, we search directly for a pair (G, S), where G is a graph and S is a cut (or independent set) in G, scoring the pair as $\mathrm{score}(G, S) = -\infty$ if G is not Ramanujan, and $\mathrm{score}(G, S) = |\delta_G(S, \bar{S})|/|E(G)|$ otherwise. Here, $\delta_G(S, \bar{S})$ is the set of edges crossing the cut $(S, \bar{S})$ in G. This gives an efficiently computable score function that lower bounds γ_d^{MC} (or γ_d^{IS}). Consequently, AlphaEvolve isn't limited to the space of small graphs. Indeed, it outputs constructions on as many as 163 vertices (for γ_4^{IS}).

For $d = 3$ for Max-Cut, AlphaEvolve recovered the existing bound $\gamma_3^{\mathrm{MC}} \geq 17/18$ of [KY24], but wasn't able to improve on it. We note that for the sake of conciseness we restricted ourselves to the $d \in \{3, 4\}$ cases. We expect our techniques to achieve similar

improvements over known bounds on γ_d^{MC} and γ_d^{IS} ([BBK+21, Hof03]) for all reasonably small values of $d > 4$ except when $d-1$ is an odd perfect square, when [BBK+21] already gives the optimal lower bound.

6.2 Local Spectral Certificates for Upper Bounds

This section expands on the upper-bound side of Theorem 6.1.8; the detailed proof appears in Section C.1.2. The certificate is spectral in the same sense as Hoffman's classical bound, but it extracts more information from the same spectral hypothesis. For an n-vertex d-regular graph G, let

$$\lambda^*(G) = \max_{i>1} |\lambda_i|,$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the eigenvalues of the adjacency matrix $\mathbf{A}$. Friedman's theorem [Fri08] implies that, for every fixed $\epsilon > 0$, a random graph $G \sim \mathcal{G}(n,d)$ satisfies

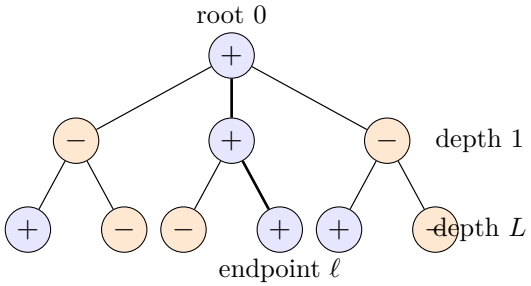
$$\lambda^*(G) \leq 2\sqrt{d-1} + \epsilon$$

with probability $1 - o_n(1)$. It therefore suffices to prove a deterministic statement: every d-regular graph with $\lambda^*(G) \leq \lambda$, where $\lambda = 2\sqrt{d-1} + \epsilon$, has bounded $\mathrm{MC}(G)$ and $\mathrm{IS}(G)$. The algorithmic certificate checks this spectral condition and then invokes the deterministic bound.

Hoffman's argument uses the quadratic form associated with the adjacency matrix. If $S \subseteq V(G)$ is a cut and $\mathbf{x} \in \{\pm 1\}^n$ is its cut vector, then $\mathbf{x}^\top \mathbf{A}\mathbf{x}$ determines the number of crossing edges. The spectral assumption constrains this quadratic form, and hence constrains the cut value. Our improvement follows the same spectral strategy while incorporating longer-walk information. Instead of only using the one-step correlation $\mathbf{x}^\top \mathbf{A}\mathbf{x}$, we also use constraints coming from $\mathbf{x}^\top \mathbf{A}^L \mathbf{x}$ for small values of L. The resulting bound is represented by a finite linear program over local neighborhoods.

Let $T_{d,L}$ be the rooted d-ary tree of depth L: the root has d children, and every non-root vertex at distance less than L from the root has $d-1$ children. Let $\Lambda_{d,L}$ denote the set of probability distributions over $\{\pm 1\}$-labelings of $T_{d,L}$. Given a graph G and a cut S, we associate to them a distribution $\mu_{G,S} \in \Lambda_{d,L}$ as follows. First choose a random ordering of the incident edges at every vertex of G. Then choose a random root vertex $i \in V(G)$. The labels on $T_{d,L}$ are obtained by following the corresponding nonbacktracking walks of length at most L starting from i, and assigning to each tree vertex the value of the cut vector at the endpoint of that walk. This construction is illustrated in Figure 6.2.

This local distribution is a coarse representation of the pair (G,S), but it retains the quantities needed by the certificate. First, it satisfies local consistency constraints. If r_0 is the root of $T_{d,L}$ and r_1 is a child of r_0, then the marginal view obtained by rooting the local neighborhood at r_0 is the same, up to the natural tree isomorphism, as the view obtained by rooting at r_1. This follows from the reversibility of a random directed edge (i,j) in a regular graph.



Variables	Root density	Walk correlation
$\mu(y)$, $y : T_{d,L} \to \{\pm 1\}$	$\mathbb{E}_{\mu}[y_0] \approx 2\alpha - 1$	$\mathbb{E}_{\mu,\ell}[y_0 y_\ell] = \mathbf{x}^\top (\mathbf{A}/d)^L \mathbf{x}/n$

Figure 6.2: The local distribution induced by a cut. After choosing a random root and random incident-edge orderings, the cut vector on G pulls back to a random $\{\pm 1\}$-labeling of the depth-L tree $T_{d,L}$. The linear program optimizes over such local distributions rather than over global cuts.

Second, $\mu_{G,S}$ records the size of the cut side. If α approximates $|S|/n$ up to an additive discretization error δ, then

$$2\alpha - 1 - 2\delta \le \mathbb{E}_{y\sim\mu_{G,S}}[y_0] \le 2\alpha - 1 + 2\delta,$$

where 0 denotes the root of the tree. Third, the spectral condition constrains the correlation between the root and the endpoint of a random L-step walk on the tree model. Indeed,

$$\mathbb{E}_{y\sim\mu_{G,S},\ell}[y_0 y_\ell] = \frac{\mathbf{x}^\top (\mathbf{A}/d)^L \mathbf{x}}{n},$$

where ℓ is the endpoint of a random L-step walk on $T_{d,L}$ starting from the root. Writing

$$\mathbf{A} = d\frac{\mathbf{1}\mathbf{1}^\top}{n} + \widetilde{\mathbf{A}}$$

and using $\|\widetilde{\mathbf{A}}\|_{\mathsf{op}} \le \lambda$, this expectation is constrained to lie within an interval determined by α, δ, and $(\lambda/d)^L$. Thus the spectral expansion assumption becomes a linear constraint on the probability mass function of $\mu_{G,S}$.

Finally, the objective can also be expressed locally. The fraction of edges crossing the cut is

$$\frac{|\delta_G(S, \bar{S})|}{|E(G)|} = \frac{1 - \mathbb{E}_{y\sim\mu_{G,S},j\sim[d]}[y_0 y_j]}{2},$$

where j ranges over the children of the root. Therefore, after fixing d, L, and a discretized value of α, the maximum cut value compatible with the spectral hypothesis is upper-bounded by a linear program whose variables are the probabilities assigned to labelings of $T_{d,L}$. Solving this LP for all $\alpha \in \{0, \delta, 2\delta, \dots, 1\}$ gives a uniform upper bound on every cut in every G satisfying $\lambda^*(G) \le \lambda$. When $L = 1$, this recovers Hoffman's bound; taking $L > 1$ adds local

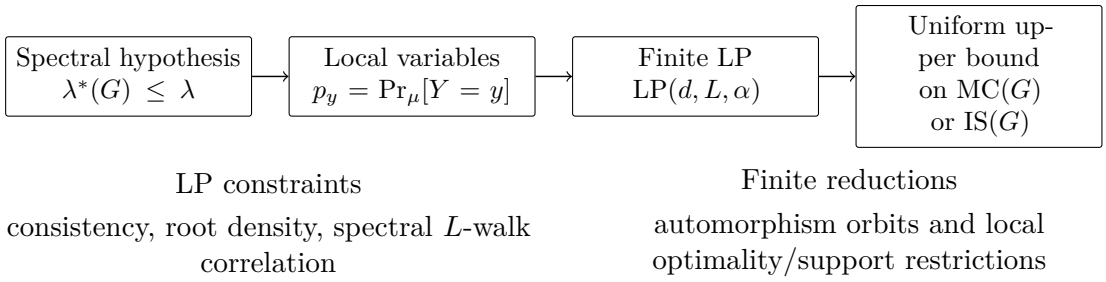


Figure 6.3: Structure of the local spectral certificate. The spectral event supplies globally valid inequalities, which become linear constraints on the local distribution μ. After symmetry and support reductions, the resulting finite LP gives a deterministic upper bound for every graph satisfying the spectral hypothesis.

constraints from longer walks and improves the numerical certificate. The structure of the resulting certificate is summarized in Figure 6.3.

The unreduced LP is computationally intractable at the relevant parameters. For example, when $d = 3$ and $L = 4$, the tree has 46 vertices, so there are 2^{46} possible labelings. We make the computation tractable using two reductions. First, labelings related by automorphisms of $T_{d,L}$ can share a single variable. Second, for Max-Cut, we restrict to locally optimal labelings: after replacing S by a locally improved cut, the label pattern on every depth-L neighborhood must be optimal subject to its boundary labels, so suboptimal local patterns can be discarded.

With $\lambda = 2\sqrt{d-1} + 10^{-5}$ and $\delta = 0.0005$, these reductions lead to tractable LPs. The Max-Cut computation uses $d = 3, L = 4$, with 4396 variables, and $d = 4, L = 2$, with 88 variables. The resulting deterministic bounds are

$$\mathrm{MC}(G) \le 0.953 \quad \text{for } G \in \Omega_{3,\lambda}, \qquad \mathrm{MC}(G) \le 0.916 \quad \text{for } G \in \Omega_{4,\lambda}.$$

Together with Friedman's theorem, these become efficient average-case certificates for random 3- and 4-regular graphs.

The Max-Independent-Set upper bounds use the same local-distribution framework. The main change is that the support of μ is restricted to labelings of $T_{d,L}$ that are themselves independent-set labelings, meaning no edge of the tree has both endpoints labeled 1. There is then no cut objective to maximize; instead, one asks for the largest feasible value of the root density parameter α. The tractable instances are $d = 3, L = 4$, with 1771 variables, and $d = 4, L = 3$, with 8855 variables, giving

$$\sigma_3^{\mathrm{IS}} \le 0.476, \qquad \sigma_4^{\mathrm{IS}} \le 0.457.$$

In this way the upper-bound side is also local and finite: spectral expansion supplies globally valid inequalities, and the LP optimizes over locally consistent neighborhood distributions satisfying those inequalities.

Chapter 7

On Optimal Distinguishers for Planted Clique

This chapter is based on "On Optimal Distinguishers for Planted Clique" [NR25], which is joint work with Prasad Raghavendra.

7.1 Introduction

In a *distinguishing problem*, one receives a sample drawn from one of two distributions, which we will refer to as the *"planted"* and *"null"* distributions. The task of the algorithm is to determine the source distribution.

The archetypal example of a *distinguishing problem* is the Planted Clique problem. Here the null distribution $\mathcal{N} = G(n, 1/2)$ is the Erdős–Rényi model over n-vertex graphs, where each edge is included independently with probability $\frac{1}{2}$. The planted distribution $\mathcal{P} = G(n, 1/2, k)$ is parametrized by a number $0 \leq k \leq n$, and is sampled as follows.[1]

- Sample a random subset $S \subseteq [n]$ of vertices by including each vertex independently with probability $\frac{k}{n}$.
- Generate an Erdős–Rényi random graph $G \sim G(n, 1/2)$.
- Plant a clique amongst the vertices in S. That is, set $G_{ij} = 1$ for all $i, j \in S$.

We will focus on the computationally hard regime where $k = n^{1/2-\alpha}$ for some $0 < \alpha < 1/2$. In this regime, the Planted Clique problem is information-theoretically easy, but has long been hypothesized to be intractable for polynomial-time algorithms [Jer92, Kuc95, AKS98, KV02]. The scale $k = \Theta(\sqrt{n})$ is the usual square-root algorithmic threshold: spectral methods and related polynomial-time algorithms succeed above it, while no polynomial-time algorithm is known below it for $k = n^{1/2-\alpha}$.

The intractability of distinguishing problems is quantified in terms of the *distinguishing advantage*. Let A be a *test*, that is, a randomized mapping A from the set of n-vertex graphs

[1] This planted distribution has been referred to as the *binomial k* model [HWX15a, BJ23]. See [HS24] for a comparison with other models.

to $\{\pm 1\}$. We will define the *advantage* achieved by A as

$$\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(A) := |\mathbb{E}_{\mathcal{P}}[A] - \mathbb{E}_{\mathcal{N}}[A]| = 2 \cdot |\Pr_{x\sim\mathcal{P}}[A(x) = 1] - \Pr_{x\sim\mathcal{N}}[A(x) = 1]| \in [0, 2]^2.$$

Above, the expectation is taken over both x and the internal randomness of A. In this terminology, the Planted Clique Hypothesis can be stated as follows.

Conjecture 7.1.1. *(Planted Clique Hypothesis) Let $\alpha > 0$ and set $k = n^{1/2-\alpha}$. Let $\mathcal{N} = G(n, 1/2)$ and $\mathcal{P} = G(n, 1/2, k)$. For every randomized polynomial time test A and all sufficiently large n,*

$$\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(A) \leq 1/4.$$

In light of the seeming intractability of the Planted Clique problem, it is natural to ask the following question.

What is the optimal (i.e., smallest) advantage achievable by any efficient algorithm?

In an elegant sequence of works, Hirahara and Shimizu [HS23, HS24] demonstrated that random self-reductions can be used to amplify the hardness of the Planted Clique problem. In particular, these works showed that Conjecture 7.1.1 implies the following stronger bound on the advantage.

Theorem 7.1.2. *(Theorem 1.3, (7) in [HS24]) Assume the Planted Clique Hypothesis. Let $\alpha, \gamma > 0$ and set $k = n^{1/2-\alpha}$. Let $\mathcal{N} = G(n, 1/2)$ and $\mathcal{P} = G(n, 1/2, k)$. For every randomized polynomial time test A,*

$$\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(A) = n^{\gamma} \cdot O\left(\frac{k^2}{n}\right).$$

Notice that on average, graphs sampled from the planted distribution $\mathcal{P}$ have $\binom{k}{2}/2$ more edges than an Erdős–Rényi graph sampled from $\mathcal{N}$, owing to the clique planted among k vertices. This slight mismatch in the expected number of edges can be used to efficiently distinguish the planted and null distributions with an advantage of $\Omega\left(\frac{k^2}{n}\right)$. Thus the above bound on the distinguishing advantage is correct up to an arbitrary polynomial factor in n.

This leaves the question of whether we can characterize the optimal advantage for Planted Clique? Subsequently, one may ask which class of algorithms are optimal distinguishers for the problem? Conversely, can we construct alternate planted distributions where the optimal advantage is significantly smaller than $\frac{k^2}{n}$, say $o(n^{-100})$? This work make progress towards answering all these questions. In order to explain our results, we will first give an introduction to the *low-degree method*.

[2]Our choice of normalization is unconventional – advantage is usually defined as *half* of this quantity. Nevertheless, this will be more notationally convenient.

Low-degree polynomials. Starting with the works of Hopkins and Steurer [Hop18, HS17], the low-degree method has emerged as an effective heuristic to determine whether a distinguishing problem is computationally tractable or not. Roughly speaking, the low-degree method is based on the hypothesis that if all degree D polynomials fail on a distinguishing task, then all algorithms running in time 2^D fail too.

To elucidate further, we will first set up some terminology on real-valued test functions. Fix a distinguishing problem between a planted distribution $\mathcal{P}$ and a null distribution $\mathcal{N}$, both over a set Ω. For a real-valued test function $f : \Omega \to \mathbb{R}$, let us define $\mathrm{R}^{(\mathcal{P},\mathcal{N})}(f)$ as the ratio

$$\mathrm{R}^{(\mathcal{P},\mathcal{N})}(f) := \frac{\mathbb{E}_{\mathcal{P}}[f] - \mathbb{E}_{\mathcal{N}}[f]}{\|f\|_{2,\mathcal{N}}} = \frac{\mathbb{E}_{\mathcal{P}}[f] - \mathbb{E}_{\mathcal{N}}[f]}{\sqrt{\mathbb{E}_{\mathcal{N}}[f^2]}} \in \mathbb{R}.$$

Informally, one can think of $\mathrm{R}^{(\mathcal{P},\mathcal{N})}(f)$ measuring the "performance" of f at distinguishing between $\mathcal{P}$ and $\mathcal{N}$ – note that if the output of f is in $\{\pm 1\}$, then one recovers the advantage

$$|\,\mathrm{R}^{(\mathcal{P},\mathcal{N})}(f)| = \mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(f).$$

$\mathrm{R}^{(\mathcal{P},\mathcal{N})}(f)$ is not as readily interpreted if the output of f is *not* restricted to $\{\pm 1\}$, but we will nevertheless treat it as a sort of distinguishing advantage. For a family of test functions $\mathcal{F} \subseteq L^2(\mathcal{N})$, let $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}]$ denote the optimal advantage achieved by a function in $\mathcal{F}$, i.e.,

$$\mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}] := \sup_{f \in \mathcal{F}} \mathrm{R}^{(\mathcal{P},\mathcal{N})}(f).$$

For simplicity, let us restrict our attention to the case where $\Omega = \{\pm 1\}^N$ is the N-dimensional hypercube for some $N \in \mathbb{N}$. In our running example of Planted Clique, one should think of $N = \binom{n}{2}$; a graph G will be represented by its flattened $\{\pm 1\}$-valued adjacency matrix in $\{\pm 1\}^{\binom{n}{2}}$. Let $\mathcal{F}_{\leq d}$ denote the subspace of polynomials of degree at most d over Ω. The quantity $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}_{\leq d}]$ denotes the "advantage" achieved by the best polynomial of degree at most d, and we will refer to it as the *low-degree advantage*. A particularly nice feature of the low-degree advantage is that it can often be explicitly computed for many distinguishing problems using analytic techniques. The central thesis behind the low-degree method is the following conjecture:

Conjecture 7.1.3 (Low-degree conjecture, informal)**.** *For many "natural" distinguishing tasks between $\mathcal{P}$ and $\mathcal{N}$, $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}_{\leq O(\log N)}] = o(1)$ implies that for every polynomial time test A, $\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(A) = o(1)$.*

We refer the reader to [Hop18, KWB19] for a more precise statement.

Computational hardness based on low-degree advantage. In light of the low-degree conjecture, it may be natural to hypothesize that low-degree polynomials yield the optimal advantage among all efficient algorithms. For Planted Clique, our results support this hypothesis by proving that under Conjecture 7.1.1, no efficient algorithm can achieve advantage better than the low-degree advantage. Specifically, we show the following.

Theorem 7.1.4. *Assume the Planted Clique Hypothesis (Conjecture 7.1.1). Let $\alpha > 0$ and set $k = n^{1/2-\alpha}$. Let $d \in \mathbb{N}$ be a sufficiently large constant. Let $\mathcal{N} = G(n, 1/2)$ and $\mathcal{P} = G(n, 1/2, k)$. For every randomized polynomial time test A,*

$$\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(A) \le (1 + o_n(1)) \cdot \mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}_{\le d}].$$

As stated before, the low-degree advantage $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}_{\le d}]$ can be computed explicitly, giving a concrete upper bound of about $k^2/(\sqrt{\pi}n)$ on the advantage. In fact, our proof can be refined slightly to yield the following nearly optimal bound.

Corollary 7.1.5. *Assume the Planted Clique Hypothesis (Conjecture 7.1.1). Let $\alpha > 0$ and set $k = n^{1/2-\alpha}$. Let $\mathcal{N} = G(n, 1/2)$ and $\mathcal{P} = G(n, 1/2, k)$. For every randomized polynomial time test A,*

$$\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(A) \le (1 + o_n(1)) \cdot \frac{k^2}{\sqrt{\pi}n}.$$

Remark 7.1.6 (Optimality). Conversely, one can show that a simple algorithm based on a low-degree polynomial achieves this bound – testing whether the input graph has more than $\binom{n}{2}/2$ edges achieves advantage $\approx \frac{k^2}{\sqrt{\pi}n}$, so the above hardness result is tight up to $1 + o_n(1)$ factors.

The above results fall out as a special case of a much more general result that can be applied beyond the Planted Clique problem. In particular, we prove the above result for any distinguishing problems where the underlying planted measure satisfies a few natural properties such as hypercontractivity of low-degree polynomials; we expect that these properties hold in many settings where the planted distribution is obtained by adding a sparse signal onto the null distribution. We defer the full statement of the general theorem to Theorem 7.4.3.

Another immediate consequence of our general result (Theorem 7.4.3) is that the optimality of low-degree polynomials holds even for perturbations of the planted distribution. Suppose $\mathcal{P}'$ is a small perturbation of the planted distribution $\mathcal{P} = G(n, 1/2, k)$ in that the relative density $\frac{d\mathcal{P}'}{d\mathcal{P}}$ is bounded pointwise by $\mathrm{poly}(n)$. Let $\mathcal{P}^*$ be a "noisy" version of $\mathcal{P}'$ in the sense that it is obtained by adding a small amount of additional randomness to $\mathcal{P}'$.

Theorem 7.1.7 (Informal version of Theorem 7.7.1)**.** *Assume the Planted Clique hypothesis (Conjecture 7.1.1). Let $\mathcal{P}'$ and $\mathcal{P}^*$ be as above, and let d be a sufficiently large constant. For any efficiently computable test A,*

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) \le \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[\mathcal{F}_{\le d}] + n^{-\Omega(\sqrt{d})} \cdot \left\| \frac{d\mathcal{P}'}{d\mathcal{P}} \right\|_\infty$$

For the sake of readability, here we do not elaborate on the details of the aforementioned noise-addition process, other than to remark that it is done in a way that approximately preserves clique size. That is, $\mathcal{P}^*$ is still essentially supported on graphs containing large cliques.

A Hard-Core Distribution. As noted earlier, the natural planted distribution $\mathcal{P}$ can be distinguished from the null distribution with an advantage of $\Omega(\frac{k^2}{n}) \gg \frac{1}{n}$ simply by counting the number of edges. Motivated by recent applications of Planted Clique-like problems to cryptography [ABI+23, BJRZ24], it is natural to ask whether there exist other planted distributions over graphs with k-cliques that make the distinguishing problem much harder. For instance, does there exist a planted distribution $\mathcal{P}^*$ such that

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) = o(1/n^{100})$$

for efficient tests A?

This is analogous to Impagliazzo's hard-core lemma [Imp95] in complexity theory, wherein starting with the assumption that a function f is hard to compute with probability 0.99 for circuits of size s, one can construct a distribution $\mathcal{D}$ over inputs under which the function cannot be computed with probability $\frac{1}{2} + \delta$ for slightly smaller circuits.

Our general theorem on perturbations of the planted distribution (see Theorem 7.1.7) paves the way towards constructing a hard distribution $\mathcal{P}^*$ by giving a way to turn low-degree hardness into computational hardness. We begin by constructing a small perturbation $\mathcal{P}'$ of $\mathcal{P}$ so that the low-degree advantage $\mathrm{R}^{(\mathcal{P}',\mathcal{N})}[\mathcal{F}_{\le d}]$ is at most a desired bound δ. Specifically, we will show the following.

Theorem 7.1.8. *For any constants $\alpha > 0$ and $d \in \mathbb{N}$, and any $\delta = \delta(n) > 0$, there is a $\mathrm{poly}(n^d, 1/\delta)$ time algorithm to sample from a distribution $\mathcal{P}'$ satisfying the following:*

1. *The low-degree advantage for the $(\mathcal{P}',\mathcal{N})$ distinguishing problem is at most δ, i.e., $\mathrm{R}^{(\mathcal{P}',\mathcal{N})}[\mathcal{F}_{\le d}] \le \delta$.*
2. *$\mathcal{P}'$ is a small perturbation of $\mathcal{P}$, i.e., $\left\|\frac{d\mathcal{P}'}{d\mathcal{P}}\right\|_\infty \le 1 + o(1)$.*

The above result can be viewed as a constructive hard-core lemma against low-degree polynomials. Let $\mathcal{P}^*$ be a noisy version of $\mathcal{P}'$ as in the setup for Theorem 7.1.7. Appealing to Theorem 7.1.7, we conclude the following computational hardness result, affirmatively answering the above question.

Theorem 7.1.9. *Assume the Planted Clique Hypothesis (Conjecture 7.1.1). For any $0 < \alpha < 1/2$ and $D \in \mathbb{N}$, there is a $\mathrm{poly}(n)$ time algorithm to sample from a distribution $\mathcal{P}^*$ supported on graphs containing a clique of size at least $k = n^{1/2-\alpha}$ such that for any polynomial time test A,*

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) = o(n^{-D}).$$

Finally, we note that our proof of Theorem 7.1.8 generalizes to a much broader class of distinguishing problems beyond Planted Clique, and may be of independent interest. Let us state an informal version of this generalization.

Theorem 7.1.10 (Informal version of Theorem 7.5.1)**.** *Let $\mathcal{P}$ and $\mathcal{N}$ be two distributions over Ω, and let V be an arbitrary subspace of $L^2(\mathcal{N})$ such that V is computationally "tractable", and V satisfies a hypercontractive inequality. If $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[V] = o(1)$, then for any $d \in \mathbb{N}$ one can efficiently find a small perturbation $\mathcal{P}'$ of $\mathcal{P}$ in the sense that $\|\frac{d\mathcal{P}'}{d\mathcal{P}}\|_\infty = 1 + o(1)$, such that $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[V] = 1/n^d$.*

We formally state and prove this result in Section 7.5.

7.1.1 Related Work

Planted Clique. The planted clique problem was originally proposed by Jerrum [Jer92] and Kučera [Kuc95] as a search version wherein the task was to identify a k-clique hidden within an n-vertex Erdős–Rényi random graph. The decision version of the problem was suggested by Saks [AKS98, KV02] has since resisted all algorithmic attacks. For $k = o(\sqrt{n})$, there is a growing body of lower bounds in restricted models of computation which suggest that the problem is intractable. We refer the reader to [HS24] for a survey on the Planted Clique problem.

Recognized as one of the most prominent problems in average-case complexity, its presumed computational difficulty has led to wide-ranging applications in areas including average-case complexity [ERSY22], cryptography [JP00, ABW10, ABI+23], hardness of approximation [MRS20], game theory [HK11] and property testing [AAK+07]. More recently, Planted Clique and its variants have been used to develop a web of reductions amongst problems in high-dimensional statistics [BR13a, BBH18, BB20, BR13b, MW15, Che15, HWX15b, WBS16, GMZ17, CLR+17, WBP16, ZX18, CW18, BB19].

Low Degree Method. The idea of using low-degree polynomials as benchmark algorithms for hypothesis testing was first explored in the work of Barak et al. [BHK+19b] on sum-of-squares lower bounds for Planted Clique. This was then further developed by Hopkins and Steurer [HS17], and subsequently by others (see, e.g. [HKP+17, Hop18], and also [KWB19] for a survey). This technique has since become known as the "low-degree hardness" paradigm, and has been applied to a variety of statistical problems [BHK+19b, HS17, HKP+17, Hop18, BKW19, KWB19, DKWB23, BB19, MRX19, CHK+20].

Perhaps the work that is most related to ours is the the work of Moitra and Wein [MW23] that studies the precise error rates for the Spiked Wigner model. Towards exactly characterizing the asymptotic error achievable by efficient algorithms, they propose the following strengthening of the low-degree conjecture, and investigate its consequences for the *spiked Wigner model*.

Conjecture 7.1.11 (Moitra-Wein strengthening of the Low-degree conjecture, informal)**.** *Let $\mathcal{F}_{\mathrm{alg}}^{\mathbb{R}}$ be the class of real-valued functions $f \in L^2(\mathcal{N})$ computable in polynomial time. For many "natural" distinguishing tasks between $\mathcal{P}$ and $\mathcal{N}$,*

$$\mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}_{\mathrm{alg}}^{\mathbb{R}}] \leq (1 + o(1)) \cdot \mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}_{O(\log n)}].$$

More recently, Ding, Hua, Slot, and Steurer [DHSS25] also leveraged a similar strengthening of the Low-degree conjecture for the stochastic block model. Unfortunately due to technical issues, the above conjecture turns out to be false as stated even for the sparse Spiked Wigner and Planted Clique models, and needs to be amended. One possible amendment states that the above holds for f belonging to a certain "nice" set of efficiently computable real-valued functions that seem to circumvent these technical issues. We believe that Theorems 7.1.4 and 7.1.7 nicely complement this hypothesis that under suitable restrictions, the behaviour of low-degree polynomials characterizes the advantage of optimal distinguishers.

Hard-Core Lemma and Hardness amplification. Fix a class $\mathcal{C}$ of circuits and a boolean function $f : \{0,1\}^n \to \{0,1\}$ which is somewhat hard to compute for the class $\mathcal{C}$, in that no circuit from $\mathcal{C}$ succeeds with probability $> 1 - \delta$. The (nonuniform) hard-core lemma asserts that there exists a large subset of inputs on which the function f is essentially unpredictable for a slightly smaller class of circuits. Impagliazzo [Imp95] gave the first proof of such a theorem and used it to derive an alternative proof of Yao's XOR lemma [Yao82].

Constructive versions of the hard-core lemma can be obtained by a boosting algorithm [KS99, BHK09].

A technical point is that in general, one cannot prove exactly analogous statements against the class of uniform polynomial-time algorithms. Holenstein [Hol05] proved a useful statement that can be used to port many of the key consequences of the nonuniform hard-core lemma to the uniform setting. In particular, assuming f is somewhat hard to compute for polynomial-time algorithms, for any polynomial-time oracle algorithm $A^{(\cdot)}$ that can make input-independent membership queries to a set $S \subseteq \{0,1\}^n$, there is a hard-core set S such that $A^{(S)}$ cannot compute f on a nontrivial fraction of inputs from S. This statement does not guarantee the existence of a *universal* hard-core set that works against all polynomial-time algorithms A, and this seems to be a generic limitation that cannot be worked around without nonuniform hardness assumptions.

In Theorem 7.1.9, we prove a constructive, uniform hard-core lemma for the Planted Clique Problem, avoiding the aforementioned generic limitation. We begin by first obtaining an assumption-free hard-core planted distribution against the class of low-degree polynomials. Then, with Theorem 7.1.7, we show how to lift this to a computational hardness statement by proving that a slightly noisy version of this distribution is hard for the class of efficient algorithms under the mild assumption Conjecture 7.1.1.

Chapter organization. Section 7.2 contains preliminaries. Section 7.3 contains an overview of our proof ideas. In Section 7.4 we state and prove a generalized version of the reduction. In Section 7.5 we state and prove a generalized version of our technique to find hard-core distributions. In Section 7.7 we describe how to instantiate our general theorems to recover our main results on Planted Clique (Theorems 7.1.4 and 7.1.9).

7.2 Preliminaries

Probability and Analysis. For a finite set Ω, $L^2(\Omega)$ denotes the vector space of all real-valued functions $f : \Omega \to \mathbb{R}$. Let $\mathcal{N}$ be a distribution over Ω. The L^p norms and inner product induced by $\mathcal{N}$ are defined as follows:

$$\|f\|_{p,\mathcal{N}} = \mathbb{E}_{\mathcal{N}}[f^p]^{1/p}, \qquad \langle f, g\rangle_{\mathcal{N}} = \mathbb{E}_{\mathcal{N}}[f \cdot g],$$

where $f \cdot g : \Omega \to \mathbb{R}$ is the pointwise product $f \cdot g(x) = f(x) \cdot g(x)$. Since we will be simultaneously working with multiple distributions, we will always write the distribution's label in the subscript to avoid ambiguity. We say that f and g are *orthogonal with respect to*

$\mathcal{N}$ if $\langle f, g\rangle_{\mathcal{N}} = 0$. As syntactic sugar, often we will write $L^2(\mathcal{N})$ to denote $L^2(\Omega)$ with the suggestion that inputs to functions in $L^2(\mathcal{N})$ are typically drawn from $\mathcal{N}$.

For any $1 \le p \le q$, Jensen's inequality applied to the convex function $x \to x^{q/p}$ implies that $\|f\|_{p,\mathcal{N}} \le \|f\|_{q,\mathcal{N}}$. The following is a consequence of Hölder's inequality.

Lemma 7.2.1 (Log-convexity of L^p norms)**.** *For $f \in L^2(\mathcal{N})$,*

$$\|f\|_{1,\mathcal{N}} \ge \frac{\|f\|_{2,\mathcal{N}}^3}{\|f\|_{4,\mathcal{N}}^2} \quad \text{and} \quad \frac{\|f\|_{4,\mathcal{N}}}{\|f\|_{2,\mathcal{N}}} \le \frac{\|f\|_{8,\mathcal{N}}^2}{\|f\|_{4,\mathcal{N}}^2}.$$

Definition 7.2.2 ((a,b)-anticoncentrated)**.** A real-valued random variable x is (a,b)-anticoncentrated if $\Pr[|x| \ge b \cdot \mathbb{E}[x^2]^{1/2}] \ge a$.

Lemma 7.2.3 (Paley-Zygmund inequality)**.** *A real-valued random variable x is $\left((1-\sqrt{\theta})^2 \cdot \frac{\mathbb{E}[x^2]^2}{\mathbb{E}[x^4]}, \theta\right)$-anticoncentrated for any $\theta \in [0,1]$.*

Definition 7.2.4 (Pointwise product)**.** For two subspaces $V, V' \subset L^2(\Omega)$, the pointwise product $V \odot V'$ is the space consisting of fg for $f \in V$ and $g \in V'$. Define $V^{\odot k} = \underbrace{V \odot \ldots \odot V}_{k \text{ times}}$.

The Bernoulli distribution $\mathrm{Ber}(p)$ is the distribution over $\{0,1\}$ that outputs 1 with probability p and 0 with probability $1-p$. The Rademacher distribution Rad is the uniform distribution over $\{\pm 1\}$.

Distinguishing Problems. For two distributions $\mathcal{P}$ and $\mathcal{N}$ over Ω, let $\frac{d\mathcal{P}}{d\mathcal{N}} \in L^2(\mathcal{N})$ be the relative density $\frac{d\mathcal{P}}{d\mathcal{N}}(x) = \frac{\Pr_{\mathcal{P}}[x]}{\Pr_{\mathcal{N}}[x]}$. For a function $f \in L^2(\mathcal{N})$ define

$$\mathrm{R}^{(\mathcal{P},\mathcal{N})}(f) = \frac{\mathbb{E}_{\mathcal{P}}[f] - \mathbb{E}_{\mathcal{N}}[f]}{\|f\|_{2,\mathcal{N}}}.$$

Let $\mathcal{F} \subseteq L^2(\mathcal{N})$. Define

$$\mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}] = \sup_{f \in \mathcal{F}} \mathrm{R}^{(\mathcal{P},\mathcal{N})}(f).$$

The following claim gives a way to compute $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[V]$ for subspaces V in terms of an orthonormal eigenbasis of V.

Claim 7.2.5. *Let $V \subseteq L^2(\mathcal{N})$ be a subspace containing $\mathbf{1}$, and let $\mathbf{1}, f_1, \ldots, f_\ell$ be a basis that is orthonormal with respect to $\langle \cdot, \cdot \rangle_{\mathcal{N}}$. Then, $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[V] = \sqrt{\sum_{i \in [\ell]} \mathbb{E}_{\mathcal{P}}[f_i]^2}$.*

Proof. Let $f \in V$. Expand $f = c_0 \mathbf{1} + \sum_{i \in [\ell]} c_i f_i$. We have

$$\begin{aligned}
\mathrm{R}^{(\mathcal{P},\mathcal{N})}(f) &= \frac{\sum_{i \in [\ell]} c_i \cdot \mathbb{E}_{\mathcal{P}}[f_i]}{\sqrt{c_0^2 + \sum_{i \in [\ell]} c_i^2}} && (\mathbb{E}_{\mathcal{N}}[f_i] = 0 \text{ for all } i \in [\ell] \text{ by orthogonality}) \\
&\le \frac{\sum_{i \in [\ell]} c_i \cdot \mathbb{E}_{\mathcal{P}}[f_i]}{\sqrt{\sum_{i \in [\ell]} c_i^2}} && (c_0^2 \ge 0) \\
&\le \sqrt{\sum_{i \in [\ell]} \mathbb{E}_{\mathcal{P}}[f_i]^2}. && (\text{Cauchy-Schwarz})
\end{aligned}$$

This proves $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[V] \leq \sqrt{\sum_{i\in[\ell]} \mathbb{E}_{\mathcal{P}}[f_i]^2}$. For the other direction, notice that one could have picked $c_i = \mathbb{E}_{\mathcal{P}}[f_i]$ and $c_0 = 0$, in which case all inequalities would be tight, proving $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[V] = \sqrt{\sum_{i\in[\ell]} \mathbb{E}_{\mathcal{P}}[f_i]^2}$. □

Markov Chains and Noise Operators. A Markov chain M over Ω is defined by a transition rule which maps an element $x \in \Omega$ to a probability distribution over Ω. We will denote a sample from this distribution by $y \sim M(x)$. For a distribution $\mathcal{N}$, $M(\mathcal{N})$ will denote the distribution of y in the following process: (1) sample $x \sim \mathcal{N}$, (2) sample $y \sim M(x)$. If $M(\mathcal{N}) = \mathcal{N}$, we say $\mathcal{N}$ is a stationary distribution of M.

We associate a *noise operator* $T : L^2(\mathcal{N}) \to L^2(\mathcal{N})$ to any Markov chain M that is defined by $Tf(x) = \mathbb{E}_{y\sim M(x)}[f(y)]$. Note that T always has the constant function $\mathbf{1}(x) = 1$ as an eigenvector with eigenvalue 1.

Fact 7.2.6. *Let T be the noise operator associated with a reversible Markov chain with stationary distribution $\mathcal{N}$. T has a complete basis of eigenvectors $\mathbf{1} = f_1, f_2, \ldots, f_{|\Omega|}$ that are orthogonal with respect to $\mathcal{N}$ with real eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{|\Omega|}$.*

We will denote by $\lambda_{\geq c}(T)$ (resp. $\lambda_{<c}(T)$) the span of all eigenvectors with eigenvalue $\geq c$ (resp. $< c$).

Boolean functions. Often, Ω will be equal to the boolean hypercube $\{\pm 1\}^{[m]}$ with $\mathcal{N} = \mathrm{Unif}(\Omega)$. For $x \in \{\pm 1\}^{[m]}$ and a set $S \subseteq [m]$, write $\chi_S(x) = x_S = \prod_{i\in S} x_i$. The *Fourier basis* $\{\chi_S\}_{S\subseteq[m]}$ forms an orthonormal basis of $L^2(\mathcal{N})$. Let $\mathcal{F}_{\leq d} = \mathrm{Span}\{\chi_S : |S| \leq d\}$ be the set of all functions having degree at most d.

The following lemma is a consequence of [Bon70].

Lemma 7.2.7 (Bonami's lemma)**.** *Every degree $\leq d$ function f over the boolean hypercube satisfies $\|f\|_{4,\mathcal{N}} \leq \sqrt{3}^d \cdot \|f\|_{2,\mathcal{N}}$.*

We will also consider the setting of the p-biased hypercube, where $\Omega = \{0,1\}^{[m]}$ and $\mathcal{N} = \mathrm{Ber}(p)^{\otimes m}$. For $x \in \{0,1\}^{[m]}$ and $S \subseteq [m]$, define the p-biased Fourier character $\chi_S^p(x) = \prod_{i\in S} \frac{x_i - p}{\sqrt{p(1-p)}}$. The Fourier basis $\{\chi_S^p\}_{S\subseteq[m]}$ is orthonormal with respect to $\mathcal{N}$. Again, we say that $f \in L^2(\mathcal{N})$ has degree at most d if $f \in \mathrm{Span}\{\chi_S^p : |S| \leq d\}$. Say that f is S_m-symmetric if $f(x) = f(\pi(x))$ for any permutation π, where π acts on x by permuting its coordinates. The following lemma is a consequence of [KLLM21].

Lemma 7.2.8. *Suppose $d \leq mp$. Every S_m-symmetric degree $\leq d$ function f over the m-dimensional p-biased hypercube satisfies $\|f\|_{4,\mathcal{N}} \leq 8^d \cdot \|f\|_{2,\mathcal{N}}$.*

Proof. Assume without loss of generality that $p \leq 1/2$. Write $f = \sum_{|T|\leq d} \hat{f}_T \chi_T^p$, and let $I_S(f) = (p(1-p))^{-|S|} \sum_{T\supseteq S} \hat{f}_T^2$. We will apply Lemma II.1.6 from [KLLM21], which states that if $I_S(f) \leq b \cdot \|f\|_{2,\mathcal{N}}^2$ for all $|S| \leq d$, then $\|f\|_{4,\mathcal{N}} \leq 4^d \cdot b^{1/4} \cdot \|f\|_{2,\mathcal{N}}$. In order to apply this, the key property about S_m-symmetric functions we will use is that $\hat{f}_T = \hat{f}_{T'}$ if $|T| = |T'|$.

$$(p(1-p))^{|S|} I_S(f) = \sum_{T\supseteq S} \hat{f}_T^2 = \mathbb{E}_{S'\sim\binom{[m]}{|S|}} \left[\sum_{T\supseteq S'} \hat{f}_T^2\right] = \binom{m}{|S|}^{-1} \sum_{|T|\leq d} \binom{|T|}{|S|} \cdot \hat{f}_T^2,$$

where the second inequality used S_m-symmetry and the last equality used the fact that $T \supseteq S'$ with probability equal to $\binom{|T|}{|S|}/\binom{m}{|S|}$.

Since $|T|$ and $|S|$ are always at most d, we can bound $\binom{|T|}{|S|} \leq 2^d$. Therefore

$$I_S(f) \leq (p(1-p))^{-|S|} \cdot \binom{m}{|S|}^{-1} \cdot 2^d \cdot \|f\|_{2,\mathcal{N}}^2 \leq 4^d \cdot \left(\frac{|S|}{mp}\right)^{|S|} \cdot \|f\|_{2,\mathcal{N}}^2 \leq 4^d \cdot \|f\|_{2,\mathcal{N}}^2,$$

where we used the fact that $p \leq 1/2$ and $|S| \leq d \leq mp$. In conclusion, we have proved that $I_S(f) \leq 4^d \cdot \|f\|_{2,\mathcal{N}}^2$, implying $\|f\|_{4,\mathcal{N}} \leq 8^d \cdot \|f\|_{2,\mathcal{N}}$. □

Graphs. In this work, we will always represent a graph by a vector $G \in \{\pm 1\}^{\binom{[n]}{2}}$. The entries of G should be interpreted as follows: the edge $\{i,j\}$ is present if and only if $G_{ij} = 1$.

Definition 7.2.9 (Random Graph Models)**.** We will consider two distributions over n-vertex graphs. The first is the Erdős–Rényi model $G(n,1/2) = \mathrm{Unif}(\{\pm 1\}^{\binom{[n]}{2}})$. The second is the binomial-k Planted Clique model $G(n,1/2,k)$ that is sampled as follows:

- Sample a graph $G \sim G(n,1/2)$.
- Sample a random vector $x \sim \mathrm{Ber}(k/n)^{\otimes n}$. For all i,j such that $x_i = x_j = 1$, set $G_{ij} = 1$.
- Output G.

7.3 Technical Overview for Planted Clique

In this section we describe the main proof ideas behind our results for planted clique. We will discuss Theorems 7.1.4 and 7.1.8 in Sections 7.3.1 and 7.3.2 respectively.

7.3.1 Hardness Self-Amplification

Let us begin by discussing Theorem 7.1.4 which states that the Planted Clique Hypothesis (Conjecture 7.1.1) implies a much stronger hardness statement. Let $\beta > 0$ be a constant, let $\Omega = \{\pm 1\}^{\binom{[n]}{2}}$, and define the distributions

$$\mathcal{P}^* = G(n,1/2,n^{1/2-\beta}), \quad \mathcal{N} = G(n,1/2).$$

Our goal is to show that $\mathcal{P}^*$ and $\mathcal{N}$ cannot be distinguished with advantage significantly larger than $\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[\mathcal{F}_{\leq d}]$ for some constant d. For the sake of contradiction, assume that there is an efficient test A mapping Ω to $\{\pm 1\}$ achieving advantage

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) > (1+\Omega(1)) \cdot \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[\mathcal{F}_{\leq d}]. \tag{7.1}$$

Let us pick another constant α such that $0 < \alpha < \beta$, and define $\mathcal{P} = G(n,1/2,n^{1/2-\alpha})$. We will use A to construct another efficient test B such that

$$\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(B) \geq \Omega(1),$$

which contradicts the Planted Clique Hypothesis by Theorem 7.1.2. Our construction of B proceeds in two steps. First, in the *amplification* step, we construct an efficiently computable *real-valued* function $g : \Omega \to \mathbb{R}$ achieving $\mathrm{R}^{(\mathcal{P},\mathcal{N})}(g) = \omega(1)$. Next, we perform a *rounding* step to convert g into an efficiently computable $\{\pm 1\}$-valued test B achieving $\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(B) = \Omega(1)$.

Assumption 7.3.1. *We assume that A is deterministic, so it can be interpreted as a function $A : \Omega \to \{\pm 1\}$. We also assume that $\mathbb{E}_{\mathcal{N}}[A] = 0$. All functions f we will construct from A will also share this property, allowing us to write the simplified expression*

$$\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}(f) = \frac{\mathbb{E}_{\mathcal{P}^*}[f]}{\|f\|_{2,\mathcal{N}}}.$$

These assumptions are completely avoidable, and solely serve the purpose of simplifying the presentation.

Step 1: Amplification. In this part, our goal is, given A satisfying Eq. (7.1), to construct a function $g \in L^2(\mathcal{N})$ that achieves $\mathrm{R}^{(\mathcal{P},\mathcal{N})}(g) = \omega(1)$. As in any average case reduction, we will need an efficiently computable randomized map M from Ω to Ω that maps $\mathcal{N}$ to $\mathcal{N}$ and $\mathcal{P}$ to $\mathcal{P}^*$. We will use the *vertex-resampling Markov Chain* M. We define M by specifying how to sample $H \sim M(G)$ for a fixed $G \in \Omega$. Set $p := n^{\alpha-\beta}$. H is sampled as follows.

- Sample $z \sim \mathrm{Ber}(p)^{\otimes n}$.
- For each $\{i,j\} \in \binom{[n]}{2}$, set $H_{\{i,j\}} = G_{\{i,j\}}$ if $z_i = z_j = 1$, and independently sample $H_{\{i,j\}} \sim \mathrm{Rad}$ otherwise.

It is not too difficult to see that $M(\mathcal{N}) = \mathcal{N}$. As for $\mathcal{P}$, one can check that M essentially independently removes vertices from the clique with probability $1 - p$, and hence $M(\mathcal{P}) = \mathcal{P}^*$ is another Planted Clique distribution with slightly smaller clique size.

Remark 7.3.2. We remark that similar maps have appeared in prior work. In particular M is the composition of the *embedding* and *shrinking* reductions used by Hirahara and Shimizu [HS23, HS24]. This map has also been used in a different context in the work of Brennan et al. [BBH$^+$20].

Let $T : L^2(\mathcal{N}) \to L^2(\mathcal{N})$ be the noise operator associated with M, that is, $Tf(G) = \mathbb{E}_{H\sim M(G)}[f(H)]$. The main observation we will need is that the Fourier basis gives an explicit orthonormal basis of eigenvectors of T with respect to the inner product $\langle f, g\rangle_{\mathcal{N}} = \mathbb{E}_{\mathcal{N}}[f \cdot g]$.

Fact 7.3.3. *For all $S \subseteq \binom{[n]}{2}$, $\chi_S(G) = G_S = \prod_{\{i,j\}\in S} G_{\{i,j\}}$ is an eigenvector of T with eigenvalue $\lambda_S = p^{|V(S)|}$, where $V(S) \subseteq [n]$ is the set of all vertices belonging to some edge in S.*

In particular we see that M has a large spectral gap; its largest nontrivial eigenvalue is $\lambda_2 := \max_{S\neq\emptyset} |\lambda_S| = p^2 = n^{2(\alpha-\beta)} = o(1)$. Note that since $\mathbb{E}_{\mathcal{N}}[A] = 0$, A is orthogonal to the constant function $\mathbf{1}$, and lies completely in the nontrivial eigenspace of T. This implies the following contractive inequality:

$$\|TA\|_{2,\mathcal{N}} \leq p^2 \cdot \|A\|_{2,\mathcal{N}} = p^2.$$

This suggests that we might want to set $g = TA$. Let us compute

$$\mathrm{R}^{(\mathcal{P},\mathcal{N})}(TA) = \frac{\mathbb{E}_{\mathcal{P}}[TA]}{\|TA\|_{2,\mathcal{N}}} = \frac{\mathbb{E}_{\mathcal{P}^*}[A]}{\|TA\|_{2,\mathcal{N}}} \geq \frac{\mathbb{E}_{\mathcal{P}^*}[A]}{p^2} = \frac{\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A)}{p^2} \gtrsim \frac{\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}(\mathcal{F}_{\leq d})}{p^2}. \quad (7.2)$$

The second equality above uses the definition of T and the fact that $M(\mathcal{P}) = \mathcal{P}^*$, and the final equality uses our simplifying assumption that $\mathbb{E}_{\mathcal{N}}[A] = 0$. A simple calculation (for instance, the one in Section D.3) shows that the numerator is approximately $n^{-2\beta}$ for any constant d. On the other hand, the denominator is $p^2 = n^{2(\alpha-\beta)}$. Consequently, we have shown

$$\mathrm{R}^{(\mathcal{P},\mathcal{N})}(TA) \gtrsim \Omega(n^{-2\alpha}).$$

Unfortunately this is not strong enough; recall that we wanted to compute g achieving $\mathrm{R}^{(\mathcal{P},\mathcal{N})}(g) = \omega(1)$. So we must look beyond the spectral gap to achieve the desired amplification. One might imagine that if A was orthogonal not only to $\mathbf{1}$, but also to *all* eigenvectors corresponding to eigenvalues that are too large, TA would obtain much larger advantage.

Our main idea is to achieve this by explicitly projecting A onto the orthogonal complement of $\mathcal{F}_{\leq d} = \mathrm{Span}\{\chi_S : |S| \leq d\}$. Let us write $A = A_+ + A_-$ for some low degree function $A_+ \in \mathcal{F}_{\leq d}$ and some high degree function $A_- \perp \mathcal{F}_{\leq d}$, and imagine replacing A with A_- in the prior discussion. Observe that for every set S of edges with $|S| > d$, we have $|V(S)| \geq \sqrt{|S|} > \sqrt{d}$, so $\lambda_S = p^{2|V(S)|} < p^{2\sqrt{d}}$. As a consequence, T acts as a $p^{2\sqrt{d}}$-contraction on A_-, an improvement from the p^2 before.

Next, we will lower bound $\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}(A_-)$. Since A_- and A_+ are orthogonal, $\max(\|A_-\|_{2,\mathcal{N}}, \|A_+\|_{2,\mathcal{N}}) \leq \|A\|_{2,\mathcal{N}} = 1$, so one can show the following kind of "triangle inequality":

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) = \mathbb{E}_{\mathcal{P}^*}[A] = \mathbb{E}_{\mathcal{P}^*}[A_+] + \mathbb{E}_{\mathcal{P}^*}[A_-] \leq \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}(A_+) + \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}(A_-).$$

Recall that $\mathbb{E}_{\mathcal{P}^*}[A] \geq (1 + \Omega(1)) \cdot \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[\mathcal{F}_{\leq d}]$ by Eq. (7.1), and note that $\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}(A_+) \leq \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[\mathcal{F}_{\leq d}]$ because $A_+ \in \mathcal{F}_{\leq d}$. Together, these imply that A_- achieves a nontrivial "advantage" of

$$\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}(A_-) \geq \Omega\left(\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[\mathcal{F}_{\leq d}]\right).$$

In a similar fashion to Eq. (7.2), we can conclude that

$$\mathrm{R}^{(\mathcal{P},\mathcal{N})}(TA_-) \geq \frac{\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}(A_-)}{p^{2\sqrt{d}}} = \Omega\left(\frac{\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[\mathcal{F}_{\leq d}]}{p^{2\sqrt{d}}}\right) \approx \Omega\left(n^{-2\beta} \cdot n^{2\sqrt{d}(\beta-\alpha)}\right).$$

Picking d to be a constant greater than $\left(\frac{\beta}{\beta-\alpha}\right)^2$, we get $\mathrm{R}^{(\mathcal{P},\mathcal{N})}(TA_-) = \omega(1)$, as desired. Therefore we will use $g := TA_-$.

Remark 7.3.4. $TA_-(G)$ is not directly computable, but we can approximate it sufficiently well by making a polynomial number of queries to A.

Step 2: Rounding. In this step, we show how to round g to our final distinguisher B for $\mathcal{P}$ and $\mathcal{N}$ achieving advantage $\Omega(1)$. We will construct B by thresholding the value of $g = TA_-$. Concretely, we will compute B in the following way.

- On input $G \in \Omega$, compute $g(G)$.
- If $|g(G)| \geq \|g\|_{2,\mathcal{P}}/10$, return 1. Otherwise, return -1.

We will interpret an output of 1 (resp. -1) as the algorithm guessing that the input was sampled from $\mathcal{P}$ (resp. $\mathcal{N}$). It is quite straightforward to bound the success probability under $\mathcal{N}$; by Markov's inequality, one can show

$$\Pr[|g(G)| \geq \|g\|_{2,\mathcal{P}}/10] \leq 100 \cdot \frac{\|g\|_{2,\mathcal{N}}}{\|g\|_{2,\mathcal{P}}} \leq 100 \cdot \frac{\|g\|_{2,\mathcal{N}}}{|\mathbb{E}_{\mathcal{P}}[g]|} = \frac{100}{|\,\mathrm{R}^{(\mathcal{P},\mathcal{N})}(g)|} = o(1).$$

The main challenge is to bound the probability of success under $\mathcal{P}$. We will aim prove that g is, say, $(\Omega(1), 0.1)$-anticoncentrated (Definition 7.2.2) under $\mathcal{P}$, that is, $|g(G)| \geq \|g\|_{2,\mathcal{P}}/10$ with probability at least $\Omega(1)$ for $G \sim \mathcal{P}$. If this is true, then B achieves the requisite advantage

$$\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}[B] = 2 \cdot (\Pr_{\mathcal{P}}[B(G) = 1] - \Pr_{\mathcal{N}}[B(G) = 1]) = \Omega(1) - o(1) = \Omega(1).$$

So it suffices to prove that $g(\mathcal{P}) = TA_-(\mathcal{P})$ is anticoncentrated. Our key insight here is that A_- is a *noise-resistant* function under $\mathcal{P}$ with respect to the noise operator T. Concretely, one can prove

$$\frac{\|TA_-\|_{2,\mathcal{P}}}{\|A_-\|_{2,\mathcal{P}^*}} \gtrsim n^{-O(d)}.$$

Since $\mathcal{P}$ is a relatively simple distribution, one should expect such noise resistant functions to behave like low-degree functions do over the boolean hypercube, in particular, it should be anticoncentrated. We prove the following anticoncentration statement for Tf such noise-resistant functions f.

Lemma 7.3.5 (Lemma 7.6.2, simplified version)**.** *Suppose $f \in L^2(\mathcal{P}^*)$ satisfies $\|Tf\|_{2,\mathcal{P}} \geq n^{-(\beta-\alpha)D} \cdot \|f\|_{2,\mathcal{P}^*}$ for some $D = o(\log n)$. Suppose that f is invariant under permutating the vertices $[n]$. Then, $Tf(\mathcal{P})$ is $(c^{D^2}, 0.1)$-anticoncentrated for some universal constant c.*

The proof of of the above lemma is quite technical so we will avoid explaining it here, pointing the reader to Section 7.6 for the proof. We will invoke the lemma with $f = A_-$. Note that we can achieve permutation-invariance by simply randomly permuting the inputs to A_-. As a result, we obtain that TA_- is indeed $(\Omega(1), 0.1)$-anticoncentrated, completing the argument that B achieves advantage $\Omega(1)$.

Remark 7.3.6. In this overview we have not shown how to obtain Theorem 7.1.7, our result for perturbations of $\mathcal{P}$. Our argument above generalizes straightforwardly to this setting, and we directly prove a version of the reduction for perturbations of $\mathcal{P}$ in Section 7.4.

7.3.2 Computing a Hardcore Distribution

The goal of this part is to discuss the main ideas behind Theorem 7.1.8, which states that one can slightly perturb the standard Planted Clique distribution to make it arbitrarily hard to distinguish from an Erdős–Rényi graph with respect to the set of low-degree polynomials $\mathcal{F}_{\leq d}$. Concretely, let $\alpha > 0$, and define

$$\mathcal{P} = G(n, 1/2, n^{1/2-\alpha}), \quad \mathcal{N} = G(n, 1/2), \quad \Omega = \{\pm 1\}^{\binom{[n]}{2}}.$$

We will show that

$$\mathrm{R}^{(\mathcal{P}',\mathcal{N})}[\mathcal{F}_{\leq d}] = 0 \tag{7.3}$$

for some distribution $\mathcal{P}'$ satisfying $\|\frac{d\mathcal{P}'}{d\mathcal{P}}\|_\infty \leq 1 + o(1)$.

We will construct $\mathcal{P}'$ by computing *acceptance probabilities* $p(G) \in [0,1]$ for each graph G. Let $G \sim \mathcal{P}$, and sample $a \sim \mathrm{Ber}(p(G))$. We set $\mathcal{P}'$ to be the distribution on G conditioned on $a = 1$. That is, the density $d\mathcal{P}'$ will be given by

$$d\mathcal{P}'(G) = \frac{p(G) \cdot d\mathcal{P}(G)}{\mathbb{E}_{\mathcal{P}}[p]}.$$

Notice that for any function $f : \Omega \to \mathbb{R}$,

$$\mathbb{E}_{\mathcal{P}'}[f] = \frac{\langle f, p\rangle_{\mathcal{P}}}{\mathbb{E}_{\mathcal{P}}[p]}.$$

Recall (Claim 7.2.5) that one can compute the low-degree advantage $\mathrm{R}^{(\mathcal{P}',\mathcal{N})}[\mathcal{F}_{\leq d}]$ by

$$\mathrm{R}^{(\mathcal{P}',\mathcal{N})}[\mathcal{F}_{\leq d}] = \sqrt{\sum_{0<|S|\leq d} \mathbb{E}_{\mathcal{P}'}[\chi_S]^2} = \frac{1}{\mathbb{E}_{\mathcal{P}}[p]} \sqrt{\sum_{0<|S|\leq d} \langle \chi_S, p\rangle_{\mathcal{P}}^2}. \tag{7.4}$$

We would like p to satisfy exactly the following two conditions:

1. The acceptance probability over $\mathcal{P}$ is close to 1. In particular, $\mathbb{E}_{\mathcal{P}}[p] \geq 1 - o(1)$. This would imply that $\|\frac{d\mathcal{P}'}{d\mathcal{P}}\|_\infty = \max_G \frac{p(G)}{\mathbb{E}_{\mathcal{P}}[p]} \leq \frac{1}{\mathbb{E}_{\mathcal{P}}[p]} = 1 + o(1)$.
2. The low degree moments of $\mathcal{P}'$ are equal to those of $\mathcal{N}$, in the sense that $\langle \chi_S, p\rangle_{\mathcal{P}} = 0$ for all $0 < |S| \leq d$. This would imply that $\mathrm{R}^{(\mathcal{P}',\mathcal{N})}[\mathcal{F}_{\leq d}] = 0$ by Eq. (7.4).

Notice that both of these conditions are linear constraints on the vector $p : \Omega \to [0,1]$. Therefore, we can write our objective as the following linear program:

$$\mathrm{LP} \equiv \max \ \mathbb{E}_{\mathcal{P}}[p] \quad \text{s.t.} \ \begin{cases} \langle p, \chi_S\rangle_{\mathcal{P}} = 0 & \text{for all } S \subseteq \Omega, 0 < |V(S)| \leq d, \\ p : \Omega \to [0,1]. \end{cases}$$

Our goal is to show that the optimal value of this LP is close to 1. By computing the dual with respect to the first set of constraints, one can reduce this to proving an inequality for low degree functions over $\mathcal{P}$. In particular, one can show

$$\mathrm{LP} = \inf_{\alpha} \mathbb{E}_{G\sim\mathcal{P}}[\max(0, 1 + \alpha(G))], \tag{7.5}$$

where the infimum is taken over all low degree functions α with zero constant coefficient, that is, $\alpha = \sum_{0<|S|\le d} \hat{\alpha}_S \cdot \chi_S$ for some Fourier coefficients $\{\hat{\alpha}_S\}_{0<|S|\le d}$. Fix an arbitrary function α of this form. If $\max(0, 1+\alpha)$ was a low-degree polynomial, one would easily be able to use the fact that $\mathcal{P}$ and $\mathcal{N}$ are mildly indistinguishable with low-degree polynomials to bound

$$\mathbb{E}_{\mathcal{P}}[\max(0, 1+\alpha)] \ge \mathbb{E}_{\mathcal{N}}[\max(0, 1+\alpha)] - o(1) \ge \mathbb{E}_{\mathcal{N}}[1+\alpha] - o(1) = 1 - o(1).$$

Unfortunately $\max(0, 1+\alpha)$ is clearly not a low-degree polynomial. Still, it is a "simple" enough function of the low-degree polynomial $1+\alpha$ that we are able to prove exactly the above bound.

With this, we have shown that $\mathrm{LP} \ge 1 - o(1)$, implying the existence of a p satisfying Conditions 1 and 2. We have not yet commented on how to efficiently compute p. As written, the linear program has exponentially many variables, but nevertheless it is possible to find an approximately optimal solution that achieves

$$\mathrm{R}^{(\mathcal{P}',\mathcal{N})}[\mathcal{F}_{\le d}] \le \delta$$

in time $\mathrm{poly}(n/\delta)$. The main idea is to find a smooth convex relaxation to the dual convex program (Eq. (7.5)), and optimize over α using stochastic gradient descent. Once we have found an approximate maximizer, we can push it back through the duality argument to find an approximate minimizer p of the LP. We leave the details of this construction to Section 7.5.

7.4 Hardness Amplification from Subspace Hardness

The purpose of this section is to abstract the key ideas of our reduction and present a proof of a general hardness amplification result. Consider the problem of efficiently distinguishing two sequences of distributions $\mathcal{P} = \mathcal{P}_n$ and $\mathcal{N} = \mathcal{N}_n$ over $\Omega = \Omega_n$. Let $M = M_n$ be a reversible Markov chain with stationary distribution $\mathcal{N}$, and let $T : L^2(\mathcal{N}) \to L^2(\mathcal{N})$ be its associated noise operator. Let $\mathcal{P}'$ be a distribution that is close to $\mathcal{P}$ in the sense that $\|\frac{d\mathcal{P}'}{d\mathcal{P}}\|_\infty$ is bounded, and set $\mathcal{P}^* = M(\mathcal{P}')$.

We will provide conditions on $\mathcal{P}$, $\mathcal{N}$, and T under which weak hardness for distinguishing $\mathcal{P}$ and $\mathcal{N}$ implies strong hardness for distinguishing $\mathcal{P}^*$ and $\mathcal{N}$. The performance of our reduction depends primarily on a parameter $\epsilon > 0$. A central object in our statement is a subspace $V \subseteq L^2(\mathcal{N})$ containing the top $\ge \epsilon$-eigenspace of T, that is, $\lambda_{>\epsilon}(T) \subseteq V$. V is often interpretable as a space consisting of the "simplest" functions in $L^2(\mathcal{N})$.

Example 7.4.1 (Low-degree functions)**.** When $T = T_\rho$ is the standard noise operator on the boolean hypercube, $V = \mathcal{F}_{\le d}$ can be taken as the set of degree $\le d$ functions for $d = \log_\rho \epsilon$.

When T is the vertex-resampling noise operator with noise rate $1-p$, $V = \mathcal{F}_{\le d}$ can be taken as the set of degree $\le d$ functions for $d = \binom{\log_p \epsilon}{2}$ (by Fact 7.3.3).

We will require that ϵ is large enough so one can find such a V that is *computationally tractable*. Concretely, we impose the following set of conditions.

Definition 7.4.2 (Tractability of function spaces)**.** Let $\mathcal{N} = \mathcal{N}_n$ be a sequence of distributions over $\Omega = \Omega_n$. A sequence of subspaces $V = V_n \subseteq L^2(\mathcal{N})$ containing the constant function $\mathbf{1}$ is $q(n)$-tractable if it has an orthonormal basis $\{\mathbf{1}, f_1, \ldots, f_\ell\}$ such that

1. $\ell \leq q(n)$.
2. For all i, one can evaluate f_i in time $q(n)$.
3. Functions f in V are $q(n)$-smooth, that is, $\|f\|_\infty \leq q(n) \cdot \|f\|_{2,\mathcal{N}}$.

We note that both examples in Example 7.4.1 are $n^{O(d)}$-tractable. We will also require a technical condition stating that functions $f \in L^2(\mathcal{P})$ that are noise-resistant should be well-behaved as random variables over $\mathcal{P}$, in the sense that they are anticoncentrated (Definition 7.2.2). Finally, we assume that the problem of distinguishing $\mathcal{P}$ from $\mathcal{N}$ is mildly hard. Informally, the conclusion of our theorem states that if the above conditions hold, then for any efficiently computable test A,

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) < \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[V] + \delta \tag{7.6}$$

for some small δ.

Theorem 7.4.3. *Let $\mathcal{N} = \mathcal{N}_n$ and $\mathcal{P} = \mathcal{P}_n$ be two sequences of distributions over a finite sample space $\Omega = \Omega_n$. Let $M = M_n$ be a reversible Markov chain with stationary measure $\mathcal{N}$, and let $T : L^2(\mathcal{N}) \to L^2(\mathcal{N})$ be its associated noise operator. Let $q = q(n) \geq 1, \epsilon = \epsilon(n) \in [0,1]$, and $\gamma = \gamma(n) \in [0,1]$ be parameters, and let $V = V_n$ be a subspace of $L^2(\mathcal{N})$ containing the top eigenspace $\lambda_{>\epsilon}(T)$. Assume the following "niceness" conditions.*

1. *In time $q(n)$, one can sample $y \sim \mathcal{N}$ and $y \sim M(x)$ for any $x \in \Omega$.* *(sampleability)*
2. *V is q-tractable.* *(tractability of top eigenspace)*
3. *If $\|Tf\|_{2,\mathcal{P}} \geq \epsilon/q \cdot \|f\|_{2,M(\mathcal{P})}$, then $Tf(\mathcal{P})$ is $(\gamma, 0.1)$-anticoncentrated.* *(anticoncentration)*

Let $\delta = 400\gamma^{-1}\epsilon$. Let $\mathcal{P}' = \mathcal{P}'_n$ be another sequence of distributions such that $\|\frac{d\mathcal{P}'}{d\mathcal{P}}\|_\infty < \infty$, and set $\mathcal{P}^* = M(\mathcal{P}')$. Suppose one can compute a randomized mapping A from Ω to $\{\pm 1\}$ in time $t(n)$ satisfying

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) \geq \mathrm{Var}_{\mathcal{N}}(\Pi_V f)^{1/2} \cdot \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[V] + \delta \cdot \left\| \frac{d\mathcal{P}'}{d\mathcal{P}} \right\|_\infty \tag{7.7}$$

for infinitely many n, where $f(G) = \mathbb{E}_A[A(G)]$ and $\mathrm{Var}_{\mathcal{N}}(\Pi_V f) = \|\Pi_V f - \mathbb{E}_{\mathcal{N}}[\Pi_V f]\|_{2,\mathcal{N}}^2$. Then, there is a randomized mapping B from Ω to $\{\pm 1\}$ with runtime $t(n) \cdot \mathrm{poly}(q, \epsilon^{-1}, \gamma^{-1})$ such that for infinitely many n,

$$\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(B) \geq \gamma.$$

7.4.1 Proof of Theorem 7.4.3

Assume without loss of generality that $\mathbb{E}_{\mathcal{P}^*}[A] \geq \mathbb{E}_{\mathcal{N}}[A]$, so $\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) = \mathbb{E}_{\mathcal{P}^*}[A] - \mathbb{E}_{\mathcal{N}}[A]$. We will define $f(x) = \mathbb{E}_A[A(x)]$ to be the average output of the algorithm under input x.

Let us recall the rough plan of our reduction, as explained in Section 7.3.1. We will decompose $f = f_+ + f_-$ for some $f_+ \in V$ and $f_- \perp V$. We will obtain the test B by thresholding $|Tf_-|$.

Of course, we aren't necessarily able to compute Tf_- efficiently – we are only guaranteed that f is efficiently computable. We will instead compute an approximation that relies on the fact that V is tractable.

Claim 7.4.4. *There is a randomized mapping C from Ω to $\mathbb{R}$ computable in time $t(n) \cdot \mathrm{poly}(q, \epsilon^{-1}, \gamma^{-1})$ such that for any $x \in \Omega$,*

$$\mathbb{E}_C[|C(x) - Tf_-(x)|] \leq \epsilon. \tag{7.8}$$

We prove Claim 7.4.4 in Section D.1. Let us continue by defining our test B.

Definition of our test B for $\mathcal{P}$ and $\mathcal{N}$.

1. On input x, compute $C(x)$.
2. If $|C(x)| \geq \delta/20$, output $+1$. Otherwise output -1.

In the rest of the proof, we will analyze the probability that this algorithm succeeds under $\mathcal{N}$ and $\mathcal{P}$ and conclude that it achieves distinguishing advantage $\Omega(\gamma)$.

Probability of success under $\mathcal{N}$. We will show that $|C(x)|$ is small with high probability by bounding its average and applying Markov's inequality.

$$\begin{aligned}
\mathbb{E}_{x\sim\mathcal{N},C}[|C(x)|] &\leq \|Tf_-\|_{1,\mathcal{N}} + \epsilon && \text{(triangle inequality, Eq. (7.8))}\\
&\leq \|Tf_-\|_{2,\mathcal{N}} + \epsilon && \text{(Jensen's inequality on } t \to t^2)\\
&\leq \epsilon \cdot \|f_-\|_{2,\mathcal{N}} + \epsilon && (V^\perp \text{ is contained in the } < \epsilon\text{-eigenspace of } T)\\
&\leq 2\epsilon. && (\|f_-\|_{2,\mathcal{N}} \leq \|f\|_{2,\mathcal{N}} \leq \|f\|_\infty \leq 1)
\end{aligned}$$

For the last inequality, we used that the outputs of f are bounded in $[-1, 1]$ because A has range $\{\pm 1\}$. By Markov's inequality, the probability that $|C(x)| \geq \delta/20$ is at most $400\epsilon/\delta$. Plugging in the definition of δ, this is at most $\gamma/10$. Therefore, B outputs $+1$ with probability at most $\gamma/10$.

Probability of success under $\mathcal{P}$. We start by showing that the anticoncentration condition (condition 3) can be applied to Tf_-, i.e. $\|Tf_-\|_{2,\mathcal{P}} \geq \epsilon/q \cdot \|f_-\|_{2,M(\mathcal{P})}$. Let us bound

$$\begin{aligned}
\|Tf_-\|_{2,\mathcal{P}} &\geq \|Tf_-\|_{1,\mathcal{P}} && \text{(Jensen's inequality on } t \to t^2)\\
&\geq \left\|\frac{d\mathcal{P}'}{d\mathcal{P}}\right\|_\infty^{-1} \cdot \|Tf_-\|_{1,\mathcal{P}'} && (\|\tfrac{dP}{dQ}\|_\infty = \sup_f \tfrac{\|f\|_{1,P}}{\|f\|_{1,Q}})\\
&\geq \left\|\frac{d\mathcal{P}'}{d\mathcal{P}}\right\|_\infty^{-1} \cdot \mathbb{E}_{\mathcal{P}'}[Tf_-]\\
&= \left\|\frac{d\mathcal{P}'}{d\mathcal{P}}\right\|_\infty^{-1} \cdot \mathbb{E}_{\mathcal{P}^*}[f_-].
\end{aligned}$$

Eq. (7.7) will allow us to further lower bound this quantity. Write $\delta' = \delta \cdot \|\frac{d\mathcal{P}'}{d\mathcal{P}}\|_\infty$. For infinitely many n, we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{P}^*}[f_-] &= \mathbb{E}_{\mathcal{P}^*}[f] - \mathbb{E}_{\mathcal{P}^*}[f_+] \\
&= \mathbb{E}_{\mathcal{P}^*}[A] - \mathbb{E}_{\mathcal{P}^*}[f_+] && \text{(definition of } f\text{)} \\
&\geq \mathbb{E}_{\mathcal{N}}[A] + \mathrm{Var}_{\mathcal{N}}(f_+)^{1/2} \cdot \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[V] + \delta' - \mathbb{E}_{\mathcal{P}^*}[f_+] && \text{(Eq. (7.7))} \\
&\geq \mathbb{E}_{\mathcal{N}}[f] + \mathrm{Var}_{\mathcal{N}}(f_+)^{1/2} \cdot \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[V] + \delta' - \mathbb{E}_{\mathcal{N}}[f_+] - \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[V] \cdot \mathrm{Var}_{\mathcal{N}}(f_+)^{1/2} \\
& && (f_+ - \mathbb{E}_{\mathcal{N}}[f_+] \in V) \\
&= \mathbb{E}_{\mathcal{N}}[f] + \delta' - \mathbb{E}_{\mathcal{N}}[f_+] \\
&= \mathbb{E}_{\mathcal{N}}[f_-] + \delta' = \delta'. && (f_- \perp V \text{ and } \mathbf{1} \in V)
\end{aligned}$$

To elaborate on the fourth step, we applied the definition of $\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[V]$ on the function $g = f_+ - \mathbb{E}_{\mathcal{N}}[f_+]$, which also belongs to V because V contains the constant function $\mathbf{1}$. The inequality then follows by the fact $\|g\|_{2,\mathcal{N}}^2 = \|f_+ - \mathbb{E}_{\mathcal{N}}[f_+]\|_{2,\mathcal{N}}^2 = \mathrm{Var}_{\mathcal{N}}(f_+)$.

Combining the two inequalities proved above, we conclude that $\|Tf_-\|_{2,\mathcal{P}} \geq \delta \geq 2\epsilon$. On the other hand,

$$\begin{aligned}
\|f_-\|_{2,M(\mathcal{P})} &\leq \|f\|_{2,M(\mathcal{P})} + \|f_+\|_{2,M(\mathcal{P})} && \text{(triangle inequality)} \\
&\leq \|f\|_\infty + \|f_+\|_\infty \\
&\leq q\|f\|_{2,\mathcal{N}} + q\|f_+\|_{2,\mathcal{N}} && \text{(smoothness of } V \text{ (Definition 7.4.2))} \\
&\leq 2q. && (\|f_+\|_{2,\mathcal{N}} \leq \|f\|_{2,\mathcal{N}} \leq 1)
\end{aligned}$$

Therefore we have $\|Tf_-\|_{2,\mathcal{P}} \geq \epsilon/q \cdot \|f_-\|_{2,M(\mathcal{P})}$, and we can conclude that $Tf_-(\mathcal{P})$ is $(\gamma, 0.1)$-anticoncentrated. Finally, we can bound the probability that B outputs $+1$. By a union bound, we write

$$\Pr_{\mathcal{P},C}[|C(x)| \geq \delta/20] \geq \Pr_{x\sim\mathcal{P}}[|Tf_-(x)| \geq \delta/10] - \Pr_{x\sim\mathcal{P},C}[|C(x) - Tf_-(x)| \geq \delta/20].$$

By $(\gamma, 0.1)$-anticoncentration, the first term is at least γ. Using Markov's inequality along with Eq. (7.8), the second term is at most $20\epsilon/\delta \leq \gamma/10$, so the probability above can be bounded by $\gamma - \gamma/10 = 9\gamma/10$.

Therefore, B outputs $+1$ with probability at least $9\gamma/10$ under $\mathcal{P}$. Combining our bounds on the success probability in both cases, we get that B achieves the required advantage

$$\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(B) = 2 \cdot (\Pr_{x\sim\mathcal{P}}[B(x) = 1] - \Pr_{x\sim\mathcal{N}}[B(x) = -1]) \geq 2 \cdot (9\gamma/10 - \gamma/10) \geq \gamma$$

for infinitely many n.

7.5 Hardcore Distributions against a Subspace of Distinguishers

In this section, we will state and prove a sort of hard-core Lemma for distinguishing two distributions $\mathcal{P}$ and $\mathcal{N}$ over a sample space Ω using functions belonging to a computationally

tractable subspace $V \subseteq L^2(\mathcal{N})$. Assume that V contains the constant function $\mathbf{1}$. Recall our measure of performance of a V at the task:

$$\mathrm{R}^{(\mathcal{P},\mathcal{N})}[V] = \max_{f \in V} \mathrm{R}^{(\mathcal{P},\mathcal{N})}(f) = \max_{f \in V} \frac{\mathbb{E}_{\mathcal{P}}[f] - \mathbb{E}_{\mathcal{N}}[f]}{\|f\|_{2,\mathcal{N}}}.$$

Our result states that if V is nice enough and $\mathcal{P}$ satisfies a mild hardness condition against $V^{\odot 4}$, then one can "perturb" $\mathcal{P}$ very slightly to find another distribution $\mathcal{P}^*$ that satisfies arbitrarily strong hardness against V.

Theorem 7.5.1. *Let $\mathcal{P} = \mathcal{P}_n$ and $\mathcal{N} = \mathcal{N}_n$ be distributions over a finite set $\Omega = \Omega_n$, and let $V = V_n \subseteq L^2(\mathcal{N})$ be a subspace of functions containing the constant function $\mathbf{1}$. Let $q = q(n) \geq 1$ and $c = c(n) \geq 1$ be parameters. Assume the following conditions.*

1. *One can sample from $\mathcal{P}$ in time $q(n)$.* *(efficient sampleability)*
2. *V is q-tractable.* *(tractability of V)*
3. *Any $f \in V$ satisfies $\|f\|_{8,\mathcal{N}} \leq c^{1/4} \cdot \|f\|_{4,\mathcal{N}}$* *($V$ is $(4,8)$-hypercontractive)*
4. *$\mathrm{R}^{(\mathcal{P},\mathcal{N})}[V^{\odot 4}] \leq 1/(8c)$.* *(mild hardness assumption)*

For any $\delta = \delta(n) > 0$, one can using $\mathrm{poly}(q, \delta^{-1})$ preprocessing time prepare the description of a distribution $\mathcal{P}^*$ such that with high probability,

1. *$\left\|\frac{d\mathcal{P}^*}{d\mathcal{P}}\right\|_\infty \leq 1 + O\left(c \cdot \mathrm{R}^{(\mathcal{P},\mathcal{N})}[V]\right)$* *(closeness of $\mathcal{P}^*$ and $\mathcal{P}$)*
2. *One can sample from $\mathcal{P}^*$ in expected time $\mathrm{poly}(q, \log(\delta^{-1}))$.* *(efficient sampleability)*
3. *$\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[V] \leq \delta$.* *(hardness amplified to δ)*

As a concrete example, one can consider the setting of the boolean hypercube, where $\mathcal{N} = \mathrm{Unif}(\{\pm 1\}^n)$ and $V = \mathcal{F}_{\leq d}$ is the space of degree $\leq d$ functions. In this case, the relevant parameters can be shown to be $q(n) = n^d$ and $c(n) = 2^{O(d)}$.

We make a few remarks about the theorem statement.

- The bound on $\left\|\frac{d\mathcal{P}^*}{d\mathcal{P}}\right\|_\infty$ in the conclusion is much stronger than a TV distance bound – in particular, $\mathcal{P}^*$ is contiguous with respect to $\mathcal{P}$, that is, unlikely events in $\mathcal{P}$ remain unlikely in $\mathcal{P}^*$. This fact is false for two distributions that are δ-close in TV distance.
- Under the conditions of the theorem, one can show using a compactness argument that there exists a distribution $\mathcal{P}^*$ satisfying $\left\|\frac{d\mathcal{P}^*}{d\mathcal{P}}\right\|_\infty \leq 1 + O\left(c \cdot \mathrm{R}^{(\mathcal{P},\mathcal{N})}[V]\right)$ and $\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[V] = 0$. That is, functions in V achieve advantage exactly 0.

7.5.1 Proof of Theorem 7.5.1

Let us define $\epsilon = \mathrm{R}^{(\mathcal{P},\mathcal{N})}[V^{\odot 4}]$. We will assume $\delta \leq \epsilon$, since otherwise $\mathcal{P}^* = \mathcal{P}$ already satisfies the requirement that $\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[V] \leq \delta$. As explained in Section 7.3.2, we will construct $\mathcal{P}^*$

by computing *acceptance probabilities* $p : \Omega \to [0,1]$. We set $\mathcal{P}^*$ to be $\mathcal{P}$ conditioned on the event that $\mathrm{Ber}(p(x)) = 1$. In other words,

$$d\mathcal{P}^*(x) = \frac{d\mathcal{P}(x) \cdot p(x)}{\mathbb{E}_{\mathcal{P}}[p]}. \tag{7.9}$$

We will sample from $\mathcal{P}^*$ by performing rejection sampling with $\mathcal{P}$.

Recall that in Section 7.3.2, we defined p to be the optimal solution to a particular LP. As solving the LP directly takes exponential time, we will follow the outline from the end of the technical overview and attempt to solve a *smoothing* of the dual formulation. Recall the dual formulation:

$$\min_{g=\sum_{i\in[\ell]} g_i f_i} \mathbb{E}_{x\sim\mathcal{P}}[\max(0, 1+g(x))].$$

Let us define a smooth approximation to the univariate function $t \to \max(0,t)$; let $\sigma(t) = \int_{-\infty}^{t} \frac{1}{1+2^{-x/\delta}} dx$. Observe that σ is convex and continuously differentiable, and satisfies the following properties for all $t \in \mathbb{R}$.

1. $\sigma'(t) \in [0,1]$.
2. $\sigma(t) \geq \max(0,t)$.
3. $\sigma(0) \leq \delta$.
4. $\sigma''(t) \leq O(1/\delta)$.

Let $\{\mathbf{1}, f_1, \ldots, f_\ell\}$ be a computationally tractable orthonormal basis of V, where $\mathbf{1}$ is the constant function. Our algorithm will consist of finding the optimal solution to the following convex program:

$$\min_{g=\sum_{i\in[\ell]} g_i f_i} \mathbb{E}_{x\sim\mathcal{P}}[\sigma(1+g(x))].$$

Say we find a solution g^* to the convex program. We will set the acceptance probabilities to

$$p(x) := \sigma'(1+g^*(x)). \tag{7.10}$$

One cannot expect to find an exact optimizer in general, since even exactly evaluating the objective function is intractable. We will leverage the fact that the objective function is written as an average over efficiently computable functions, and use stochastic gradient descent to find an approximate optimizer. We will only require approximate first-order optimality in the sense that the gradient of the objective function at g^* is small. Let us calculate the first derivatives.

$$\left.\frac{\partial \mathbb{E}_{\mathcal{P}}[\sigma(1+g)]}{\partial g_i}\right|_{g=g^*} = \mathbb{E}_{x\sim\mathcal{P}}\left[\left.\frac{\partial}{\partial g_i}\sigma\left(1+\sum_{i\in[\ell]} g_i f_i\right)\right|_{g=g^*}\right] = \mathbb{E}_{x\sim\mathcal{P}}\left[f_i \cdot \sigma'(1+g^*)\right] = \langle f_i, p\rangle_{\mathcal{P}}.$$

Claim 7.5.2. *With* $\mathrm{poly}(q, \delta^{-1})$ *preprocessing time, one can prepare the description of a function* $g^* = \sum_{i\in[\ell]} g_i^* f_i$ *that can be queried in time* $\mathrm{poly}(q, \log \delta^{-1})$ *such that with high probability over the preprocessing randomness,*

$$\left\|\nabla_{\{g_1,\dots,g_\ell\}} \mathbb{E}_{\mathcal{P}}[\sigma(1+g)]\right\|_2\Big|_{g=g^*} = \sqrt{\sum_{i\in[\ell]} \langle f_i, p\rangle_{\mathcal{P}}^2} \le \delta/3,$$

where $\left\|\nabla_{\{g_1,\dots,g_\ell\}} \mathbb{E}_{x\sim\mathcal{P}}[\sigma(1+g(x))]\right\|_2\Big|_{g=g^*}$ *is the gradient of* $\mathbb{E}_{x\sim\mathcal{P}}[\sigma(1+g(x))]$ *viewed as a function of the coefficients* $g_1, \dots, g_n$ *of* $g = \sum_{i\in\ell} g_i f_i$*, evaluated at* g^**.*

The proof of Claim 7.5.2 involves invoking some off-the-shelf convergence guarantees for stochastic gradient descent, and we will defer it to Section D.2. We will require another claim that bounds the optimal value of the convex program above.

Claim 7.5.3. *For any* $g = \sum_{i\in[\ell]} g_i f_i$*, we have* $\mathbb{E}_{\mathcal{P}}[\sigma(1+g)] \ge 1 - 4c\epsilon + \epsilon\|g\|_{2,\mathcal{N}}$.

The proof of Claim 7.5.3 involves using the fact that $\mathcal{P}$ and $\mathcal{N}$ are mildly indistinguishable with respect to polynomials over $V^{\odot 4}$, which allows us to switch the expectation with respect to $\mathcal{P}$ to an expectation with respect to $\mathcal{N}$, which is then easily bounded. Let us argue that the three conditions on $\mathcal{P}^*$ are satisfied.

Closeness of $\mathcal{P}^*$ and $\mathcal{P}$. We will bound

$$\left\|\frac{d\mathcal{P}^*}{d\mathcal{P}}\right\|_\infty = \max_{x\in\Omega} \frac{p(x)}{\mathbb{E}_{\mathcal{P}}[p]} \le \frac{1}{\mathbb{E}_{\mathcal{P}}[p]}.$$

Let us lower bound the denominator.

$$\begin{aligned}
\mathbb{E}_{\mathcal{P}}[p] &= \mathbb{E}_{\mathcal{P}}[\sigma'(1+g^*)] && \text{(Eq. (7.10))}\\
&= \mathbb{E}_{\mathcal{P}}[(1+g^*)\cdot\sigma'(1+g^*)] - \mathbb{E}_{\mathcal{P}}[g^*\cdot\sigma'(1+g^*)] \\
&&& \text{(adding/subtracting } \mathbb{E}_{\mathcal{P}}[g^*\cdot\sigma'(1+g^*)]\text{)}\\
&\ge \mathbb{E}_{\mathcal{P}}[\sigma(1+g^*)] - \sigma(0) - \mathbb{E}_{\mathcal{P}}[g^*\cdot\sigma'(1+g^*)] && (\sigma \text{ is convex, so } \sigma(t)-\sigma(0)\le t\cdot\sigma'(t))\\
&\ge 1 - 5c\epsilon + \epsilon\|g^*\|_{2,\mathcal{N}} - \mathbb{E}_{\mathcal{P}}[g^*\cdot\sigma'(1+g^*)] && (\text{Claim 7.5.3, } \sigma(0)\le\delta\le\epsilon\le c\epsilon)\\
&\ge 1 - 5c\epsilon + \epsilon\|g^*\|_{2,\mathcal{N}} - \sum_{i\in[\ell]} g_i^*\cdot\mathbb{E}_{\mathcal{P}}[f_i\cdot p] && (\text{expanded } g^*, \text{ Eq. (7.10)})\\
&\ge 1 - 5c\epsilon + \epsilon\|g^*\|_{2,\mathcal{N}} - \|g^*\|_{2,\mathcal{N}}\cdot\sqrt{\sum_{i\in[\ell]}\langle f_i,p\rangle_{\mathcal{P}}^2} && \text{(Cauchy-Schwarz)}\\
&\ge 1 - 5c\epsilon + \epsilon\|g^*\|_{2,\mathcal{N}} - \delta/3\cdot\|g^*\|_{2,\mathcal{N}} \ge 1 - 5c\epsilon. && (\text{Claim 7.5.2, } \delta\le\epsilon)
\end{aligned}$$

By the mild hardness assumption on $V^{\odot 4}$, $c\epsilon \le 1/8$, implying that $\|\frac{d\mathcal{P}^*}{d\mathcal{N}}\|_\infty = 1 + O(c\cdot\epsilon)$.

Efficient Sampleability We will sample from $\mathcal{P}^*$ by rejection sampling with $\mathcal{P}$. Concretely, we sample $x\sim\mathcal{P}$, and a Bernoulli $b\sim\mathrm{Ber}(p(x))$. If $b=1$, we output x, otherwise we repeat this process. The output is distributed as $\mathcal{P}^*$ by definition, and since $\mathbb{E}[p(x)] \ge 1 - 5c\epsilon > \Omega(1)$, the expected number of iterations before terminating is $O(1)$. Since g^* can be computed in time $\mathrm{poly}(q, \log\delta^{-1})$, the expected runtime is $\mathrm{poly}(q, \log\delta^{-1})$.

Hardness amplified to δ. Finally, we bound the quantity $\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[V]$ using the expression given by Claim 7.2.5.

$$\begin{aligned}
\mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[V] &= \sqrt{\sum_{i\in[\ell]} \mathbb{E}_{\mathcal{P}^*}[f_i]^2} && \text{(Claim 7.2.5)}\\
&= \frac{1}{\mathbb{E}_{\mathcal{P}}[p]}\sqrt{\sum_{i\in[\ell]} \langle p, f_i\rangle_{\mathcal{P}}^2} && \text{(Eq. (7.9))}\\
&\le \frac{\delta}{3\cdot(1-5c\epsilon)} \le \delta. && \text{(Claim 7.5.2, bound on } \mathbb{E}_{\mathcal{P}}[p],\ c\epsilon \le 1/8)
\end{aligned}$$

This completes the proof of Theorem 7.5.1 modulo Claim 7.5.2 which we prove in in Section D.2, and Claim 7.5.3 which we prove below.

Proof of Claim 7.5.3. Recall the hypercontractivity property we assumed, which states that for $f \in V$, $\|f\|_{8,\mathcal{N}} \le c^{1/4}\cdot\|f\|_{4,\mathcal{N}}$. We will use this in several ways throughout this proof, so it will be helpful to collect all the consequences we will need before getting into the proof. Firstly, for any function $f \in V$, $(4,8)$-hypercontractivity implies $(2,4)$-hypercontractivity by Lemma 7.2.1:

$$\|f\|_{4,\mathcal{N}} \le c^{1/2}\cdot\|f\|_{2,\mathcal{N}}. \tag{7.11}$$

As a result, one can apply the definition of $\epsilon = \mathrm{R}^{(\mathcal{P},\mathcal{N})}[V^{\odot 4}]$ on the functions f^2 and f^4 to bound $|\|f\|_{2,\mathcal{P}}^2 - \|f\|_{2,\mathcal{N}}^2| \le \epsilon\cdot\|f\|_{4,\mathcal{N}}^2$, and $|\|f\|_{4,\mathcal{P}}^4 - \|f\|_{4,\mathcal{N}}^4| \le \epsilon\cdot\|f\|_{8,\mathcal{N}}^2$. Applying $(2,4)$-hypercontractivity and $(4,8)$-hypercontractivity respectively, one has the following inequalities.

$$\|f\|_{2,\mathcal{P}} \ge \sqrt{1-c\epsilon}\cdot\|f\|_{2,\mathcal{N}} \ge 2^{-1/5}\cdot\|f\|_{2,\mathcal{N}}, \quad \|f\|_{4,\mathcal{P}} \le (1+c^{1/2}\epsilon)^{1/4}\cdot\|f\|_{4,\mathcal{N}} \le 2^{1/5}\cdot\|f\|_{2,\mathcal{N}}, \tag{7.12}$$

where we used the bound $c\epsilon \le 1/8$. Let us finally begin with the proof. Recall that we want to prove a lower bound on $\mathbb{E}_{\mathcal{P}}[\sigma(1+g)]$. Let $0 < t < 1/2$ be some parameter to be picked later. We will bound

$$\begin{aligned}
\mathbb{E}_{\mathcal{P}}[\sigma(1+g)] &\ge \mathbb{E}_{\mathcal{P}}[\max(0,1+g)] && (\sigma(x) \ge \max(0,x))\\
&= \frac{1}{2}\mathbb{E}_{\mathcal{P}}[1+g] + \frac{1}{2}\mathbb{E}_{\mathcal{P}}[|1+g|] && (\max(0,x) = x/2 + |x|/2)\\
&\ge (1-t)\cdot\mathbb{E}_{\mathcal{P}}[1+g] + t\cdot\mathbb{E}_{\mathcal{P}}[|1+g|] && (\mathbb{E}_{\mathcal{P}}[|1+g|] \ge \mathbb{E}_{\mathcal{P}}[1+g])\\
&= 1 - t + (1-t)\cdot\mathbb{E}_{\mathcal{P}}[g] + t\cdot\|1+g\|_{1,\mathcal{P}}.
\end{aligned}$$

We will bound the second and third terms separately. For the second term, recall that $\epsilon = \mathrm{R}^{(\mathcal{P},\mathcal{N})}(V^{\odot 4})$ and $\mathbb{E}_{\mathcal{N}}[g] = 0$ because g has no $\mathbf{1}$ component. So,

$$\mathbb{E}_{\mathcal{P}}[g] \ge \mathbb{E}_{\mathcal{N}}[g] - \epsilon\cdot\|g\|_{2,\mathcal{N}} = -\epsilon\cdot\|g\|_{2,\mathcal{N}}.$$

Next we bound the third term.

$$\begin{aligned}
\|1+g\|_{1,\mathcal{P}} &\geq \frac{\|1+g\|_{2,\mathcal{P}}^3}{\|1+g\|_{4,\mathcal{P}}^2} && \text{(Lemma 7.2.1)}\\
&\geq \frac{1}{2}\cdot\frac{\|1+g\|_{2,\mathcal{N}}^3}{\|1+g\|_{4,\mathcal{N}}^2} && \text{(Eq. (7.12) on } f=1+g)\\
&\geq (2c)^{-1}\cdot\|1+g\|_{2,\mathcal{N}} \geq (2c)^{-1}\cdot\|g\|_{2,\mathcal{N}}. && \text{(Eq. (7.11) on } f=1+g)
\end{aligned}$$

Plugging these back into our original bound, we get

$$\begin{aligned}
\mathbb{E}_{\mathcal{P}}[\sigma(1+g)] &\geq 1-t+\|g\|_{2,\mathcal{N}}\cdot\left(\frac{t}{2c}-\epsilon\right)\\
&= 1-4\cdot c\epsilon+\epsilon\|g\|_{2,\mathcal{N}}, && \text{(picking } t=4\epsilon c\leq 1/2.)
\end{aligned}$$

completing the proof. □

7.6 Anticoncentration for Planted Clique

In this section, we will prove the main technical statement that is required to invoke our general results in the setting of Planted Clique. We will show that Condition 3 in the statement of Theorem 7.4.3 holds, that is, functions in $L^2(\mathcal{P})$ that survive the noise operator are anticoncentrated as real-valued random variables.

We will diverge slightly from the technical overview in Section 7.3 and consider a slight modification of the vertex resampling Markov Chain introduced in Section 7.3.1.

Definition 7.6.1 (Permuted Vertex Resampling Markov Chain)**.** Let $\Omega = \{\pm 1\}^{\binom{[n]}{2}}$. The Permuted Vertex Resampling Markov Chain with noise rate $1-p$ is a Markov chain M^{sym} with the following transition rule: for any $G \in \Omega$, $M^{\mathrm{sym}}(G)$ is distributed as H in the following.

- Sample a random permutation $\pi \sim S_n$. Replace G by $\pi(G)$.
- Sample $z \sim \mathrm{Ber}(p)^{\otimes n}$.
- Sample the entries of H as follows:

$$H_{\{i,j\}} \sim \begin{cases} G_{\{i,j\}} & \text{if } x_i = x_j = 1,\\ \mathrm{Rad} & \text{otherwise.}\end{cases}$$

The required anticoncentration statement is captured in the following Lemma.

Lemma 7.6.2. *Let $0 < p < 1$ and $k \in \mathbb{N}$. Let $\mathcal{P} = G(n, 1/2, k)$, $\mathcal{P}' = G(n, 1/2, pk)$. Let M^{sym} be the permuted vertex resampling Markov chain over $\{\pm 1\}^{\binom{[n]}{2}}$ with noise rate $1-p$, so $M(\mathcal{P}) = \mathcal{P}'$. Let T be the noise operator associated with M^{sym}. Suppose $f \in L^2(\mathcal{P}')$ satisfies $\|Tf\|_{2,\mathcal{P}} \geq p^d \cdot \|f\|_{2,\mathcal{P}'}$ for some $d = o(\min(k, \log p^{-1}))$. Then, $Tf(\mathcal{P})$ is $(c^{d^2}, 0.1)$-anticoncentrated for some universal constant c.*

The goal of the rest of this section is to prove the above Lemma. Let us begin by describing a high-level plan. Observe that since $\mathcal{P}$ is not a stationary distribution of M^{sym}, we are not guaranteed a basis of eigenvectors of T that is orthonormal with respect to $\mathcal{P}$. To address this issue, consider the following equivalent way to sample a graph $\tilde{G} \sim \mathcal{P}$. Define $q := k/n$.

- Sample $x \sim \mathrm{Ber}(q)^{\otimes n}$. Define $\mathrm{clique}(x) \in \{0,1\}^{\binom{[n]}{2}}$ by $\mathrm{clique}(x)_{\{i,j\}} = x_i x_j$.
- Sample $G \sim G(n, 1/2)$.
- Output $\tilde{G} = G \wedge \mathrm{clique}(x)$. That is, output the graph obtained from G by planting a clique on the vertices i such that $x_i = 1$.

The main idea is to treat f not as a function of $\tilde{G}$ under $\tilde{G} \sim \mathcal{P}$, but as a function of (x, G) where $x \sim \mathrm{Ber}(q)^{\otimes n}$ and $G \sim G(n, 1/2)$ are independent. Since the input is now a simple product distribution, we will be able to better understand the effect of the noise operator T on it. We will begin by defining Ψ and Ψ', the "lifted" versions of $\mathcal{P}$ and $\mathcal{P}'$, by

$$\Psi = \mathrm{Ber}(p)^{\otimes n} \times G(n, 1/2), \qquad \Psi' = \mathrm{Ber}(pq)^{\otimes n} \times G(n, 1/2).$$

Note that Ψ and Ψ' are both distributions over $\{0,1\}^{[n]} \times \{\pm 1\}^{\binom{[n]}{2}}$. For clarity, we will always denote a sample from Ψ as (x, G) and a sample from Ψ' as (y, H) for $x, y \in \{0,1\}^{[n]}$ and $G, H \in \{\pm 1\}^{\binom{[n]}{2}}$. We define a lifted Markov chain M^* by the following sampling proceduer for $M^*(x, G)$ for any (x, G):

- Randomly permute the vertices by sampling $\pi \sim S_n$, and replacing x with $\pi(x)$ and G with $\pi(G)$.
- Sample $z \sim \mathrm{Ber}(p)^{\otimes n}$. Set $y = x \wedge z$.
- Set $H_{ij} = \begin{cases} G_{ij} & \text{if } z_i = z_j = 1, \\ \mathrm{Rad} & \text{otherwise.} \end{cases}$
- Output (y, H).

Let us denote by T^* the noise operator associated with M^*. Note that $T^*(\Psi) = \Psi'$, and for any (x, G), the distribution of $H \wedge \mathrm{clique}(y)$ for $(y, H) \sim M^*(x, G)$ is identical to that of $M^{\mathrm{sym}}(G \wedge \mathrm{clique}(y))$. As a consequence, we have the following.

Observation 7.6.3. *For any function $f \in L^2(\mathcal{P}')$, consider $g(y, H) = f(H \wedge \mathrm{clique}(y))$. Then, $f(\mathcal{P}')$ and $g(\Phi')$ are identical as real-valued random variables, and so are $Tf(\mathcal{P})$ and $T^* g(\Phi)$.*

We will be able to coarsely understand the action of T^* on $L^2(\Psi')$ in terms of the Fourier basis. In particular, we will "merge" all monomials belonging to a certain equivalence class, resulting in a partitioning on the space $L^2(\Psi)$ into a collection of orthogonal subspaces. We say that a pair (A, B) for $A \subseteq \binom{[n]}{2}$ and $B \subseteq [n]$ is "valid" if $V(A) \cap B = 0$. For a valid pair (A, B), define the subspace $W_{A,B} \subseteq L^2(\Psi)$ to be all functions of the form

$$(x, G) \to G_A \cdot \chi_B^q(x) \cdot r(x_{V(A)}),$$

where r is an arbitrary function in $L^2(\{0,1\}^{V(A)})$. Here, $G_A = \prod_{\{ij\}\in A} G_{ij}$ is the boolean Fourier character and $\chi_B^q(x) = \prod_{i\in B} \frac{x_i - q}{\sqrt{q(1-q)}}$ is the q-biased character.Similarly, define $W'_{A,B} \subseteq L^2(\Psi')$ as the space of all functions of the form

$$(y, H) \to H_A \cdot \chi_B^{pq}(y) \cdot r(y_{V(A)}).$$

The following is a direct result of the fact that the Fourier characters form an orthogonal basis, and the fact that Ψ and Ψ' are product distributions.

Fact 7.6.4. *The collection of subspaces $W_{A,B}$ (resp. $W'_{A,B}$) form an orthogonal decomposition of $L^2(\Psi)$ (resp. $L^2(\Psi')$), that is, they are an orthogonal collection of subspaces that span the entire space.*

We will require two results about these subspaces.

Claim 7.6.5. *For any valid pair (A, B), T^* is a $p^{\frac{|V(A)|+|B|}{2}}$-contractive mapping from $W'_{A,B}$ to $W_{A,B}$. That is, for any function $w \in W'_{A,B}$, $T^*w \in W_{A,B}$, and $\|T^*w\|_{2,\Psi} \le p^{\frac{|V(A)|+|B|}{2}} \cdot \|w\|_{2,\Psi'}$.*

Claim 7.6.6. *Any S_n-symmetric function in $\mathrm{Span}\{W_{A,B} : |V(A)| + |B| \le 4d, (A, B)\}$ is $((2c)^{d^2}, 1/2)$-anticoncentrated for some absolute constant c.*

We prove both of the above results in Section 7.6.1. For now we will use them to obtain Lemma 7.6.2. Let $f \in L^2(\mathcal{P}')$ be the given function that satisfies $\|Tf\|_{2,\mathcal{P}} \ge p^d \cdot \|f\|_{2,\mathcal{P}'}$. We will define $g \in L^2(\Psi')$ by $g(y, H) = f(H \wedge \mathrm{clique}(y))$. By Observation 7.6.3, we have $\|T^*g\|_{2,\Psi} \ge p^d \cdot \|g\|_{2,\Psi'}$. We will use Fact 7.6.4 to write $g = g_+ + g_-$, where

$$g_+ \in \mathrm{Span}\{W_{A,B} : |V(A)| + |B| \le 4d\}, \quad g_- \in \mathrm{Span}\{W_{A,B} : |V(A)| + |B| > 4d\}.$$

By Claim 7.6.5, we get

$$\|T^*g_-\|_{2,\Psi} \le p^{2d} \cdot \|g_-\|_{2,\Psi'} \le p^{2d} \cdot \|g\|_{2,\Psi} \le p^d \cdot \|T^*g\|_{2,\Psi}. \tag{7.13}$$

Let us calculate the probability that T^*g is at least $\|T^*g\|_{2,\Psi}/10$. We will lower bound it by the probability that both (a) $|T^*g_+| \ge \|T^*g\|_{2,\Psi}/4$ and (b) $|T^*g_-| < \|T^*g\|_{2,\Psi}/10$.

(a) Eq. (7.13) implies $\|T^*g_+\|_{2,\Phi}^2 = \|T^*g\|_{2,\Phi}^2 - \|T^*g_-\|_{2,\Phi}^2 \ge (1-p^{2d}) \cdot \|T^*g\|_{2,\Phi}^2 \ge \|T^*g\|_{2,\Phi}^2/2$. We conclude by Claim 7.6.6 that $|T^*g_+(x, G)| \ge \|T^*g(\Psi)\|_{2,\Psi}/4$ with probability at least c^{d^2}.

(b) By Markov's inequality,

$$\begin{aligned}
\Pr_{\Psi}[|Tg_-| > \|T^*g\|_{2,\Psi}/10] &= \Pr_{\Psi}[|Tg_-|^2 > \|T^*g\|_{2,\Psi}^2/100] \\
&\le \frac{100\|T^*g_-\|_{2,\Psi}^2}{\|T^*g\|_{2,\Psi}^2} \\
&\le 100p^{2d}. && \text{(Eq. (7.13))}
\end{aligned}$$

As a result, T^*g is at least $\|T^*g\|_{2,\Psi}/10$ with probability $(2c)^{-d^2} - 100p^{2d} \ge c^{-d^2}$, where we used that $d = o(\log p^{-1})$. That is, T^*g is $(c^{-d^2}, 0.1)$-anticoncentrated. By Observation 7.6.3, this is equivalent to the statement that Tf is $(c^{-d^2}, 0.1)$-anticoncentrated, completing the proof.

7.6.1 Proof of Auxiliary Claims

It remains to prove Claims 7.6.5 and 7.6.6.

Proof of Claim 7.6.5. Let $w \in W'_{A,B}$, that is, $w(y, H) = H_A \cdot \chi_B^{pq}(y) \cdot r(y_{V(A)})$ for an arbitrary function r. We need to prove that $T^*w \in W_{A,B}$, and that $\|T^*w\|_{2,\Psi} \le p^{\frac{|V(A)|+|B|}{2}} \cdot \|w\|_{2,\Psi'}$. Let us write an expression for T^*w. Let $z \sim \mathrm{Ber}(p)^{\otimes n}$. Since B is disjoint from $V(A)$, we can write

$$\begin{aligned} T^*w(x, G) &= \mathbb{E}_z[\mathbb{E}_{H|G,z}[H_A] \cdot \chi_B^{pq}(x \wedge z) \cdot r((x \wedge z)_{V(A)})] \\ &= \left(\prod_{i \in B} \mathbb{E}_z[\chi_{\{i\}}^{pq}(x \wedge z)]\right) \cdot \mathbb{E}_z[\mathbb{E}_{H|G,z}[H_A] \cdot r((x \wedge z)_{V(A)})], \end{aligned}$$

We will consider each term separately. First, let $i \in B$. We have

$$\begin{aligned} \mathbb{E}_z[\chi_{\{i\}}^{pq}(x \wedge z)] &= \mathbb{E}_{z_i \sim \mathrm{Ber}(p)}\left[\frac{x_i z_i - pq}{\sqrt{pq(1-pq)}}\right] \\ &= \frac{p \cdot (x_i - p)}{\sqrt{pq(1-pq)}} = \sqrt{\frac{p \cdot (1-p)}{1-pq}} \cdot \chi_{\{i\}}^{q}(x). \end{aligned}$$

Next, we will simplify the second term in our expression for $T^*w(x, G)$. Recall that conditioned on values for G and z, $H_{\{i,j\}}$ is set to $G_{\{i,j\}}$ if $z_i = z_j = 1$, and an independent Rademacher otherwise. As a result we have $\mathbb{E}_{H|G,z}[H_A] = \mathbf{1}_{\{z_{V(A)}=1\}} \cdot G_A$, so

$$\mathbb{E}_z[\mathbb{E}_{H|G,z}[H_A] \cdot r((x \wedge z)_{V(A)})] = p^{|V(A)|} \cdot G_A \cdot r(x_{V(A)}).$$

Combining these two expressions, we can write

$$T^*w(x, G) = p^{|V(A)|+|B|/2} \cdot \sqrt{\frac{1-p}{1-pq}}^{|B|} \cdot G_A \cdot \chi_B^q(x) \cdot r(x_{V(A)}).$$

This proves that $T^*w \in W_{A,B}$. It remains to show that T^* contracts the norm of w.

$$\begin{aligned} \|T^*w\|_2^2 &\le p^{2|V(A)|+|B|} \cdot \mathbb{E}_{(x,G)\sim\Psi}\left[\left(G_A \cdot \chi_B^q(x) \cdot r(x_{V(A)})\right)^2\right] && (\sqrt{\tfrac{1-p}{1-pq}} \le 1) \\ &= p^{2|V(A)|+|B|} \cdot \mathbb{E}_{x\sim\mathrm{Ber}(q)^{\otimes n}}\left[\chi_B^q(x)^2 \cdot r(x_{V(A)})^2\right] && (G_A \in \{\pm 1\}) \\ &= p^{2|V(A)|+|B|} \cdot \mathbb{E}_{x\sim\mathrm{Ber}(q)^{\otimes n}}\left[r(x_{V(A)})^2\right]. && (B \cap V(A) = \emptyset,\ \mathbb{E}[\chi_B^q(x)^2] = 1) \end{aligned}$$

We will bound this further by noting that $\mathbb{E}_{x\sim\mathrm{Ber}(q)^{\otimes n}}\left[r(x_{V(A)})^2\right] = \mathbb{E}_{X\sim\mathrm{Ber}(q)^{\otimes|V(A)|}}[R(X)]$ for some nonnegative function R. We will bound

$$\mathbb{E}_{X\sim\mathrm{Ber}(q)^{\otimes|V(A)|}}[R(X)] \le \left\|\frac{d\,\mathrm{Ber}(q)^{\otimes|V(A)|}}{d\,\mathrm{Ber}(pq)^{\otimes|V(A)|}}\right\|_\infty \mathbb{E}_{X\sim\mathrm{Ber}(pq)^{\otimes|V(A)|}}[R(X)].$$

Calculating $\left\| \frac{d\,\mathrm{Ber}(q)^{\otimes |V(A)|}}{d\,\mathrm{Ber}(pq)^{\otimes |V(A)|}} \right\|_\infty = \left\| \frac{d\,\mathrm{Ber}(q)}{d\,\mathrm{Ber}(pq)} \right\|_\infty^{|V(A)|} = p^{-|V(A)|}$, we get the bound

$$\|T^* w\|_2^2 \le p^{|V(A)|+|B|} \cdot \mathbb{E}_{y \sim \mathrm{Ber}(pq)^{\otimes n}}[r(y_{V(A)})^2].$$

On the other hand, we can compute

$$\|w\|_2^2 = \mathbb{E}_{(y,H)\sim\Psi'}[(H_A \cdot \chi_B^{pq}(y) \cdot r(y_{V(A)}))^2] = \mathbb{E}_{y \sim \mathrm{Ber}(pq)^{\otimes n}}[r(y_{V(A)})^2]\,.$$

We conclude that $\|T^* w\|_2 \le p^{\frac{|V(A)|+|B|}{2}} \cdot \|w\|_2$, completing the proof. □

Proof of Claim 7.6.6. We must show that any S_n-symmetric function $h \in \mathrm{Span}\{W_{A,B} : |V(A)| + |B| \le 4d\}$ is $((2c)^{d^2}, 1/2)$-anticoncentrated for some absolute constant c. We will prove an upper bound on $\frac{\|h\|_{8,\Psi}}{\|h\|_{4,\Psi}}$ and then use the Paley-Zygmund inequality (Lemma 7.2.3) to obtain anticoncentration. Let us write

$$\|h\|_{8,\Psi}^8 = \mathbb{E}_\Psi[h^8] = \mathbb{E}_x[\mathbb{E}_G[h^8(x,G)]].$$

Observation 7.6.7. *For any fixed x, the map $G \to h(x,G)$ is a degree $\le \binom{|V(A)|}{2} \le 8d^2$ function over the boolean hypercube $\{\pm 1\}^{\binom{[n]}{2}}$. As a consequence, $G \to h^2(x,G)$ is a degree $\le 16d^2$ function.*

Consequently for any x we can use Bonami's lemma (Lemma 7.2.7) twice to bound the inner expectation by

$$\mathbb{E}_{G\sim G(n,1/2)}[h^8(x,G)] \le 9^{8d^2}\mathbb{E}_{G\sim G(n,1/2)}[h^4(x,G)]^2 \le 9^{16d^2}\mathbb{E}_{G\sim G(n,1/2)}[h^2(x,G)]^4.$$

Now consider the function $s(x) = \mathbb{E}_G[h^2(x,G)]$ under the q-biased hypercube, that is, under $x \sim \mathrm{Ber}(q)^{\otimes n}$. Since h is symmetric under permutations $\pi \in S_n$, s is also S_n-symmetric and has degree $\le 8d = o(np)$, so we can apply Lemma 7.2.8 to obtain $\mathbb{E}[s^4(x)] \le 8^{32d} \cdot \mathbb{E}[s^2(x)]^2$. Combining these two bounds,

$$\begin{aligned}
\|h\|_{8,\Psi}^8 &\le 9^{16d^2} \cdot \mathbb{E}_x[s^4(x)] \\
&\le 9^{16d^2} 8^{32d} \cdot \mathbb{E}_x[s^2(x)]^2 \\
&= 9^{16d^2} 8^{32d} \cdot \mathbb{E}_x[\mathbb{E}_G[h^2(x,G)]^2]^2 \\
&\le 9^{16d^2} 8^{32d} \cdot \mathbb{E}_x[\mathbb{E}_G[h^4(x,G)]]^2 = 9^{16d^2} 8^{32d} \cdot \|h\|_{4,\Psi}^8. && \text{(Jensen's inequality)}
\end{aligned}$$

We will apply the Paley-Zygmund inequality (Lemma 7.2.3) on h^2 to obtain anticoncentration of h.

$$\begin{aligned}
\Pr_\Psi[|h| \ge \|h\|_{2,\Psi}/2] &\ge \Pr_\Psi[|h| \ge \|h\|_{4,\Psi}/2] && \text{(Jensen's inequality)} \\
&= \Pr_\Psi[h^2 \ge \|h^2\|_{2,\Psi}/4] \\
&\ge \frac{\mathbb{E}_\Psi[h^4]^2}{4\mathbb{E}_\Psi[h^8]} && \text{(Lemma 7.2.3)} \\
&\ge \frac{1}{4 \cdot 9^{16d^2} 8^{32d}} \ge (2c)^{d^2}. && (\text{for } 2c = (4 \cdot 9^{16} \cdot 8^{32})^{-1})
\end{aligned}$$

□

7.7 Proof of Main Theorems for Planted Clique

In this section we will prove our results on Planted Clique. Having already proved all the pieces in previous sections, this section will not introduce any new ideas; it will mostly involve invoking these pieces with an appropriate choice of parameters.

7.7.1 Central Theorem

We will begin by stating and proving a formal version of Theorem 7.1.7, which will imply all our other results.

Theorem 7.7.1. *Assume the Planted Clique Hypothesis (Conjecture 7.1.1). Let $0 < \alpha < \beta$, and let $d \in \mathbb{N}$. Consider $\mathcal{P} = G(n, 1/2, n^{1/2-\alpha})$ and $\mathcal{N} = G(n, 1/2)$. Let $\mathcal{P}'$ be another distribution over Ω. Define $\mathcal{P}^* = M^{\mathrm{sym}}(\mathcal{P}')$, where M is the permuted vertex-resampling Markov chain with noise rate $1 - p$ for $p = n^{\alpha-\beta}$. For every randomized polynomial time test A,*

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) \le \mathrm{Var}_{\mathcal{N}}(\Pi_{\mathcal{F}_{\le d}} f)^{1/2} \cdot \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[\mathcal{F}_{\le d}] + O(n^{-(\beta-\alpha)d'}) \cdot \left\| \frac{d\mathcal{P}'}{d\mathcal{P}} \right\|_\infty,$$

where $f(G) = \mathbb{E}_A[A(G)]$, $\mathrm{Var}_{\mathcal{N}}(\Pi_V f) = \|\Pi_V f - \mathbb{E}_{\mathcal{N}}[\Pi_V f]\|_{2,\mathcal{N}}^2$, and d' is the minimum number of vertices in any simple graph with $d+1$ edges. In particular, we have $\mathrm{Var}_{\mathcal{N}}(\Pi_{\mathcal{F}_{\le d}} f)^{1/2} \le \|f\|_{2,\mathcal{N}} \le 1$ and $d' \ge \sqrt{d}$, implying the simplified statement

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) \le \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[\mathcal{F}_{\le d}] + O(n^{-(\beta-\alpha)\sqrt{d}}) \cdot \left\| \frac{d\mathcal{P}'}{d\mathcal{P}} \right\|_\infty.$$

Proof of Theorem 7.7.1. We will argue the contrapositive. Suppose there is a polynomial time computable test A such that for infinitely many values of n,

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) \ge \|\Pi_{\mathcal{F}_{\le d}} f\|_{2,\mathcal{N}} \cdot \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[\mathcal{F}_{\le d}] + \omega(n^{-(\beta-\alpha)d'}) \cdot \left\| \frac{d\mathcal{P}'}{d\mathcal{P}} \right\|_\infty \tag{7.14}$$

In light of Theorem 7.1.2, we would like to conclude that there is a polynomial time computable test B such that for infinitely many values of n,

$$\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(B) = \Omega(1),$$

which would imply that the Planted Clique Hypothesis is false. To do this, we will apply Theorem 7.4.3 with an appropriate choice of parameters. We will set

$$\epsilon = p^{d'} = n^{(\alpha-\beta)d'}, \quad q = n^{2d}, \quad \gamma = c^{(4d/(\beta-\alpha))^2}$$

for c as in Lemma 7.6.2. We will set $V = \mathcal{F}_{\le d}$ to be the space of degree $\le d$ polynomials. Let T be the noise operator associated with M^{sym}.

Fact 7.7.2. *For any $0 < p < 1$ and d, $\mathcal{F}_{\le d}$ contains the top eigenspace $\lambda_{>\epsilon}(T)$ for $\epsilon = p^{d'}$, where d' is the minimum number of vertices an any simple graph with $d+1$ edges.*

Let us verify all requirements of Theorem 7.4.3.

- One can sample from $\mathcal{N}$, and sample $H \sim M^{\mathrm{sym}}(G)$ in time $O(n^2) = O(q)$.
- $V = \mathcal{F}_{\leq d}$ is $\binom{n}{2}^d \leq q$-tractable.
- By Lemma 7.6.2, if $\|Tf\|_{2,\mathcal{P}} \geq \epsilon/q \cdot \|f\|_{2,M^{\mathrm{sym}}(\mathcal{P})} \geq n^{-4d} \cdot \|f\|_{2,M^{\mathrm{sym}}(\mathcal{P})}$, we have that $Tf(\mathcal{P})$ is $(\gamma, 0.1)$-anticoncentrated by definition of γ.

By Eq. (7.14), we have $\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) \geq \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}(\mathcal{F}_{\leq d}) + \omega(\delta) \cdot \left\| \frac{d\mathcal{P}'}{d\mathcal{P}} \right\|_\infty$, where $\delta = 1600\gamma^{-1}\epsilon = O(p^{d'}) = O(n^{-(\beta-\alpha)d'})$. Therefore we can apply Theorem 7.4.3 to deduce the existence of an efficiently computable test B achieving $\mathrm{Adv}^{(\mathcal{P},\mathcal{N})}(B) \geq \gamma = \Omega(1)$ for infinitely many n, as required. □

7.7.2 Implications for Vanilla Planted Clique

Theorem 7.1.4 follows directly from Theorem 7.7.1 – it is exactly equivalent to the simplified statement in the special case that $\mathcal{P}' = \mathcal{P}$. Next we will deduce Corollary 7.1.5.

Proof of Corollary 7.1.5. Let us set $d = 1$. We have that d', the minimum number of vertices in any simple graph with $d+1$ edges, equals 3. The following claim is a simple calculation which we defer to Section D.3.

Claim 7.7.3. *Let $\mathcal{P} = G(n, 1/2, k)$ and $\mathcal{N} = G(n, 1/2)$ for $k = n^{1/2-\beta}$ for some $\beta > 0$. We have $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}_{\leq d}] = (1 + o_n(1)) \cdot k^2/(\sqrt{2}n)$ for any constant $d \geq 1$.*

Set $\alpha = \beta/10$ for some arbitrary $0 < \beta < 1$. We will apply the strong bound in Theorem 7.7.1 for $\mathcal{P}' = \mathcal{P}$. We have $\mathcal{P}^* = G(n, 1/2, k)$ for $k = n^{1/2-\beta}$. Let A be a randomized polynomial time test. Since $\mathcal{P}^*$ and $\mathcal{N}$ are both invariant under permuting the n vertices of the graph, we can assume at no cost to A's advantage that A is also invariant under permuting the vertices of the input graph. Applying Theorem 7.7.1, we conclude that for any randomized polynomial time test A,

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) \leq \mathrm{Var}_{\mathcal{N}}(\Pi_{\mathcal{F}_{\leq 1}} f)^{1/2} \cdot \frac{k^2}{\sqrt{2}n} + O(n^{-0.9 \cdot \beta \cdot 3}) = \mathrm{Var}_{\mathcal{N}}(\Pi_{\mathcal{F}_{\leq 1}} f)^{1/2} \cdot \frac{k^2}{\sqrt{2}n} + o\left(\frac{k^2}{\sqrt{n}}\right).$$

To complete the proof, we will argue that $\mathrm{Var}_{\mathcal{N}}(\Pi_{\mathcal{F}_{\leq 1}} f)^{1/2} \leq \sqrt{\frac{2}{\pi}} + o_n(1)$.

Claim 7.7.4. *Let $p : \{\pm 1\}^T \to [-1, 1]$ be a bounded function over the m-dimensional boolean hypercube with Fourier decomposition $p = \sum_{S \subseteq T} \hat{p}_S \cdot \chi_S$. If its degree-1 part $p_{=1} = \sum_{i \in T} \hat{p}_{\{i\}} \chi_{\{i\}}$ is symmetric under permuting the $|T|$ coordinates, then $\|p_{=1}\|_{2,\mathcal{N}} \leq \sqrt{2/\pi} + o_{|T|}(1)$ where $\mathcal{N} = \mathrm{Unif}(\{\pm 1\}^T)$.*

Proof. By symmetry, we have $\hat{p}_{\{i\}} = \hat{p}_{\{j\}}$ for all $i, j \in T$. So,

$$\begin{aligned}
|\hat{p}_{\{j\}}| &= |T|^{-1} \cdot \left| \sum_{i \in T} \hat{p}_{\{i\}} \right| \\
&= |T|^{-1} \cdot \left| \mathbb{E}_{x \sim \{\pm 1\}^T}[p(x) \cdot \sum_{i \in T} x_i] \right| \\
&\leq |T|^{-1} \cdot \mathbb{E}_{x \sim \{\pm 1\}^T} \left| \sum_{i \in T} x_i \right| && (p \text{ has range } [-1, 1])
\end{aligned}$$

By the Central Limit Theorem, we have that the expected value of $|\sum_{i \in T} x_i|$ is at most $(1 + o_{|T|}(1)) \cdot \sqrt{|T|} \cdot \mathbb{E}_{g \sim N(0,1)}[|g|] = (1 + o_{|T|}(1))\sqrt{|T|} \cdot \sqrt{\frac{2}{\pi}}$. Therefore we have

$$\|p_{=1}\|_{2,\mathcal{N}} = \sqrt{\sum_{i \in T} \hat{p}_{\{i\}}^2} \leq (1 + o_{|T|}(1)) \cdot \sqrt{\frac{2}{\pi}}.$$

□

We will apply Claim 7.7.4 to $p = f - \mathbb{E}_{\mathcal{N}}[f]$ for $T = \binom{[n]}{2}$. Since f is symmetric under permuting the vertices of the input graph, its degree-1 Fourier coefficients are indeed symmetric under permuting the $\binom{n}{2}$ potential edges – for any two edges $e, e' \in \binom{[n]}{2}$, there is a permutation on vertices that sends e to e'. Therefore, we get $\mathrm{Var}_{\mathcal{N}}(\Pi_{\mathcal{F}_{\leq 1}} f)^{1/2} = \|p_{=1}\|_{2,\mathcal{N}} \leq \sqrt{2/\pi} + o_n(1)$. We have now proved the desired bound

$$\mathrm{Adv}^{(\mathcal{P}^*, \mathcal{N})}(A) \leq (1 + o_n(1)) \cdot \frac{k^2}{\sqrt{\pi} n}.$$

□

7.7.3 Hard-core Lemma style results

Finally, we prove our hard-core lemma style results, beginning with Theorem 7.1.8, which states that one can find a perturbation $\mathcal{P}'$ of $\mathcal{P}$ that is arbitrarily hard for low-degree polynomials. We will do this by invoking Theorem 7.5.1.

Proof of Theorem 7.1.8. We will simply apply Theorem 7.5.1 to $\mathcal{P} = G(n, 1/2, k)$ for $k = n^{1/2-\alpha}$ to $V = \mathcal{F}_{\leq d}$. Set $q(n) = n^{O(d)}$ and $c = \sqrt{3}^d = O(1)$. Let us check that the conditions of the Theorem 7.5.1 hold.

- One can sample from $\mathcal{P}$ in time $O(n^2) \leq q(n)$.
- V is indeed $\binom{n}{2}^d \leq q$-tractable.
- For hypercontractivity, we will use Bonami's lemma (Lemma 7.2.7) on $f^2 \in \mathcal{F}_{\leq 2d}$, which states that $\|f^2\|_{4,\mathcal{N}} \leq 3^d \cdot \|f^2\|_{2,\mathcal{N}}$ for $f \in \mathcal{F}_{\leq d}$. Let us bound

  $$\mathbb{E}_{\mathcal{N}}[f^8] = \|f^2\|_{4,\mathcal{N}}^4 \leq 3^{4d} \cdot \mathbb{E}_{\mathcal{N}}[f^4]^2.$$

  So, $\|f\|_{8,\mathcal{N}} \leq \sqrt{3}^d \cdot \|f\|_{4,\mathcal{N}}$.

- We have by Claim 7.7.3 that $\mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}_{\leq d}] = o_n(1) = o(c)$.

The statement of Theorem 7.5.1 then directly implies Theorem 7.1.8. □

Finally, we show how to combine Theorems 7.1.8 and 7.7.1 to get Theorem 7.1.9.

Proof of Theorem 7.1.9. Let $\alpha > 0$, and set $\beta = 2\alpha$. We will set $d = 10D^2/\alpha$ and $\delta = n^{-2D}$. Let us apply Theorem 7.1.8 to obtain $\mathcal{P}'$ satisfying $\left\|\frac{d\mathcal{P}'}{d\mathcal{P}}\right\|_\infty = 1 + o_n(1)$ and $\mathbb{R}^{(\mathcal{P}',\mathcal{N})}(A) \leq \delta$. Applying Theorem 7.7.1 to $\mathcal{P}'$, we obtain $\mathcal{P}^*$ so that for any efficient test A,

$$\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A) \leq \mathrm{R}^{(\mathcal{P}^*,\mathcal{N})}[\mathcal{F}_{\leq d}] + O(n^{-2D}) \leq \mathrm{R}^{(\mathcal{P}',\mathcal{N})}[\mathcal{F}_{\leq d}] + O(n^{-2D}) \leq \delta + O(n^{-2D}).$$

By definition of δ, this is $O(n^{-2D}) = o(n^{-D})$ as required. Note that $\mathcal{P}^*$ does not contain a large clique with probability 1. Instead, we can show that $T(\mathcal{P})$ contains a clique of size, say $n^{1/2-10\beta}$ except with exponentially small probability $\exp(-O(n))$. Since $\|\frac{d\mathcal{P}^*}{M^{\mathrm{sym}}(\mathcal{P})}\|_\infty \leq \|\frac{d\mathcal{P}'}{\mathcal{P}}\|_\infty = 1+o_n(1)$, this implies that $\mathcal{P}^*$ also contains a clique of size $n^{1/2-10\beta}$ with probability $1 - \exp(-O(n))$. So the sampler can keep track of the sampled clique and use rejection sampling to ensure that the resulting graph has a large clique with probability 1.

This results in a $\exp(-O(n))$ perturbation to $\mathcal{P}^*$ in TV distance, which can only change the value of $\mathrm{Adv}^{(\mathcal{P}^*,\mathcal{N})}(A)$ by $\exp(-O(n))$, and the theorem follows. □

Chapter 8

Conclusion

The projects in this dissertation all study sharp transitions between efficient and inefficient computation. The first technical part studies approximation algorithms: in the permanent chapter, the transition is an approximation factor, and in the EPR chapter, it is the approximation ratio achieved by an efficient quantum-state rounding algorithm. The second technical part studies average-case algorithms: in the sparse-graph certification chapter, the transition is a certifiable value for maximum cut or maximum independent set, and in the planted clique chapter, it is the distinguishing advantage achievable by a polynomial-time test.

The reason to focus on sharp thresholds is that they force the algorithmic and hardness results to be stated on the same numerical scale. They also expose the structure that each method is really using: an SDP normalization and trace parameter for positive semidefinite permanents, fractional matchings and tilted EPR states for quantum Hamiltonian approximation, Kunisky-Yu quiet planting through extremal Ramanujan graphs for sparse-graph certification, local linear programs for matching upper bounds, and low-degree polynomials and noise operators for planted clique.

8.1 Future Directions

Several directions remain open. For positive semidefinite permanents, the algorithmic and hardness constants still do not match; closing this gap would likely require a sharper understanding of the geometry of Gaussian moments. For the EPR Hamiltonian, the improved approximation algorithm leaves open whether the optimal ratio is strictly below 1, and whether the tilted-state rounding principle can be extended to richer quantum Hamiltonian families. For sparse random graphs, the near-matching certification thresholds in small degrees suggest that exact constants may be within reach, perhaps by combining larger searches for extremal Ramanujan graphs with stronger local or semidefinite certificates. For planted clique, the hard-core distribution construction suggests a broader program of turning low-degree hardness into precise computational hardness for other average-case problems.

Across the projects, the first step is to formulate the numerical target precisely enough that simple algorithms can be tested against it. The next step is to find the obstruction to improving

them. Sometimes that obstruction is analytic; sometimes it is a finite combinatorial structure; and in the sparse-graph chapter, some of the relevant structures were found computationally and then verified by conventional arguments.

Bibliography

[AAK+07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence, 2007.

[ABI+23] Damiano Abram, Amos Beimel, Yuval Ishai, Eyal Kushilevitz, and Varun Narayanan. Cryptography from planted graphs: security with logarithmic-size messages, 2023.

[ABW10] Benny Applebaum, Boaz Barak, and Avi Wigderson. Public-key cryptography from different assumptions, 2010.

[AGGS17] Nima Anari, Leonid Gurvits, Shayan Oveis Gharan, and Amin Saberi. Simply exponential approximation of the permanent of positive semidefinite matrices, 2017.

[AGM20] Anurag Anshu, David Gosset, and Karen Morenz. Beyond product state approximations for a quantum analogue of max cut. *arXiv preprint arXiv:2003.14394*, 2020.

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph, 1998.

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

[Ans87] Richard P. Anstee. A polynomial algorithm for b-matchings: An alternative approach. *Information Processing Letters*, 24(3):153–157, 1987.

[AS48] Milton Abramowitz and Irene A Stegun. Handbook of mathematical functions with formulas, graphs, and mathematical tables, 1948.

[Bar16] Alexander Barvinok. Combinatorics and complexity of partition functions, 2016.

[Bar20] Alexander Barvinok. A remark on approximating permanents of positive definite matrices, 2020.

[BB19] Matthew Brennan and Guy Bresler. Average-case lower bounds for learning sparse mixtures, robust estimation and semirandom adversaries, 2019.

[BB20] Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage, 2020.

[BBH+12] Boaz Barak, Fernando G.S.L. Brandao, Aram W. Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications, 2012.

[BBH18] Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure, 2018.

[BBH+20] Matthew Brennan, Guy Bresler, Samuel B Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low-degree tests are almost equivalent, 2020.

[BBK+21] Afonso S Bandeira, Jess Banks, Dmitriy Kunisky, Christopher Moore, and Alex Wein. Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs, 2021.

[Ben05] Vidmantas Bentkus. A lyapunov-type bound in rd. *Theory of Probability and Its Applications*, 49(2):311–323, 2005.

[BGG+23] Vijay Bhattiprolu, Mrinal Kanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Inapproximability of matrix $p \to q$ norms, 2023.

[BHK09] Boaz Barak, Moritz Hardt, and Satyen Kale. The uniform hardcore lemma via approximate bregman projections, 2009.

[BHK+19a] Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem, 2019.

[BHK+19b] Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem, 2019.

[BJ23] Guy Bresler and Tianze Jiang. Detection-recovery and detection-refutation gaps via reductions from planted clique, 2023.

[BJRZ24] Andrej Bogdanov, Chris Jones, Alon Rosen, and Ilias Zadik. Low-degree security of the planted random subgraph problem, 2024.

[BKW19] Afonso S Bandeira, Dmitriy Kunisky, and Alexander S Wein. Computational hardness of certifying bounds on constrained pca problems, 2019.

[Bon70] Aline Bonami. Etude des coefficients de fourier des fonctions de lp (g), 1970.

[BR13a] Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection, 2013.

[BR13b] Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection, 2013.

[BRS15] Jop Briët, Oded Regev, and Rishi Saket. Tight hardness of the non-commutative grothendieck problem, 2015.

[Che15] Yudong Chen. Incoherence-optimal matrix completion, 2015.

[CHK+20] Yeshwanth Cherapanamjeri, Samuel B Hopkins, Tarun Kathuria, Prasad Raghavendra, and Nilesh Tripuraneni. Algorithms for heavy-tailed statistics: Regression, covariance estimation, and beyond, 2020.

[CLR+17] T Tony Cai, Tengyuan Liang, Alexander Rakhlin, et al. Computational and statistical boundaries for submatrix localization in a large noisy matrix, 2017.

[CW18] T. Tony Cai and Yihong Wu. Statistical and computational limits for sparse matrix detection, 2018.

[DG03] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures and Algorithms*, 22(1):60–65, 2003.

[DHSS25] Jingqiu Ding, Yiding Hua, Lucas Slot, and David Steurer. Low degree conjecture implies sharp computational thresholds in stochastic block model, 2025.

[DKWB23] Yunzi Ding, Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Subexponential-time algorithms for sparse pca, 2023.

[Edm65] Jack Edmonds. Maximum matching and a polyhedron with 0, 1-vertices. *Journal of Research of the National Bureau of Standards B*, 69:55–56, 1965.

[ENOG24] Farzam Ebrahimnejad, Ansh Nagda, and Shayan Oveis Gharan. On approximability of the permanent of psd matrices, 2024.

[ERSY22] Reyad Abed Elrazik, Robert Robere, Assaf Schuster, and Gal Yehuda. Pseudorandom self-reductions for np-complete problems, 2022.

[Fri08] Joel Friedman. *A Proof of Alon's Second Eigenvalue Conjecture and Related Problems*. American Mathematical Society, 2008.

[GG23] Guillaume Garrigos and Robert M. Gower. Handbook of convergence theorems for stochastic gradient methods. *arXiv preprint arXiv:2301.11235*, 2023.

[GMZ17] Chao Gao, Zongming Ma, and Harrison H. Zhou. Sparse cca: Adaptive estimation and computational barriers, 2017.

[GP19] Sevag Gharibian and Ojas Parekh. Almost optimal classical approximation algorithms for a quantum generalization of max-cut. *arXiv preprint arXiv:1909.08846*, 2019.

[GRSW16] Venkatesan Guruswami, Prasad Raghavendra, Rishi Saket, and Yi Wu. Bypassing ugc from some optimal geometric inapproximability results, 2016.

[Hae21] Willem H Haemers. Hoffman's ratio bound, 2021.

[Hås01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[HK11] Elad Hazan and Robert Krauthgamer. How hard is it to approximate the best nash equilibrium?, 2011.

[HKP+17] Samuel B. Hopkins, Pravesh K. Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures, 2017.

[HLP52] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. Inequalities, 1952.

[HNP+23] Yeongwoo Hwang, Joe Neeman, Ojas Parekh, Kevin Thompson, and John Wright. Unique games hardness of quantum max-cut, and a conjectured vector-valued borell's inequality. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1319–1384. SIAM, 2023.

[Hof03] Alan J Hoffman. On eigenvalues and colorings of graphs, 2003.

[Hol05] Thomas Holenstein. Key agreement from weak bit agreement, 2005.

[Hop18] Samuel Hopkins. Statistical inference and the sum of squares method., 2018.

[HS17] Samuel B Hopkins and David Steurer. Efficient bayesian estimation from few samples: community detection and related problems, 2017.

[HS23] Shuichi Hirahara and Nobutaka Shimizu. Hardness self-amplification: Simplified, optimized, and unified, 2023.

[HS24] Shuichi Hirahara and Nobutaka Shimizu. Planted clique conjectures are equivalent, 2024.

[HTPG24] Felix Huber, Kevin Thompson, Ojas Parekh, and Sevag Gharibian. Second order cone relaxations for quantum max cut, 2024.

[HWX15a] Bruce Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs, 2015.

[HWX15b] Bruce Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs, 2015.

[Imp95] R. Impagliazzo. Hard-core distributions for somewhat hard problems, 1995.

[Jer92] Mark Jerrum. Large cliques elude the metropolis process, 1992.

[JKK+24] Zackary Jorquera, Alexandra Kolla, Steven Kordonowy, Juspreet Singh Sandhu, and Stuart Wayland. Monogamy of entanglement bounds and improved approximation algorithms for qudit hamiltonians. *arXiv preprint arXiv:2410.15544*, 2024.

[JN25] Nathan Ju and Ansh Nagda. Improved approximation algorithms for the epr hamiltonian, 2025.

[JP00] Ari Juels and Marcus Peinado. Hiding cliques for cryptographic security, 2000.

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries, 2004.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 767–775, 2002.

[Kin23] Robbie King. An improved approximation algorithm for quantum max-cut on triangle-free graphs. *Quantum*, 7:1180, 2023.

[KKMO07] 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.

[KLLM21] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Global hypercontractivity and its applications, 2021.

[KS99] Adam R Klivans and Rocco A Servedio. Boosting and hard-core sets, 1999.

[Kuc95] Ludek Kucera. Expected complexity of graph partitioning problems, 1995.

[KV02] Michael Krivelevich and Van H Vu. Approximating the independence number and the chromatic number in expected polynomial time, 2002.

[KVWX23] Pravesh Kothari, Santosh S Vempala, Alexander S Wein, and Jeff Xu. Is planted coloring easier than planted clique?, 2023.

[KWB19] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio, 2019.

[KY24] Dmitriy Kunisky and Xifan Yu. Computational hardness of detecting graph lifts and certifying lift-monotone properties of random regular graphs, 2024.

[Lee22] Eunou Lee. Optimizing quantum circuit parameters via sdp, 2022.

[LP09] Laszlo Lovasz and Michael D. Plummer. *Matching Theory*, volume 367. American Mathematical Society, 2009.

[LP24] Eunou Lee and Ojas Parekh. An improved quantum max cut approximation via maximum matching. In *51st International Colloquium on Automata, Languages, and Programming*, volume 297 of *Leibniz International Proceedings in Informatics*, pages 105:1–105:11. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2024.

[McK25] Brendan McKay. Graph formats, 2025. Accessed 2025-09-08.

[Mei23] A Meiburg. Inapproximability of positive semidefinite permanents and quantum state tomography, 2023.

[Mos20] Stefan M Moser. Expected logarithm and negative integer moments of a noncentral 2-distributed random variable, 2020.

[MRS20] Pasin Manurangsi, Aviad Rubinstein, and Tselil Schramm. The strongish planted clique hypothesis and its consequences, 2020.

[MRX19] Sidhanth Mohanty, Prasad Raghavendra, and Jeff Xu. Lifting sum-of-squares lower bounds: Degree-2 to degree-4, 2019.

[MW15] Zongming Ma and Yihong Wu. Computational barriers in minimax submatrix detection, 2015.

[MW23] Ankur Moitra and Alexander S. Wein. Precise error rates for computationally efficient testing, 2023.

[NR25] Ansh Nagda and Prasad Raghavendra. On optimal distinguishers for planted clique, 2025.

[NRT25] Ansh Nagda, Prabhakar Raghavan, and Abhradeep Thakurta. Reinforced generation of combinatorial structures: Hardness of approximation, 2025.

[NVE+25] Alexander Novikov, Ngân Vũ, Marvin Eisenberger, Emilien Dupont, Po-Sen Huang, Adam Zsolt Wagner, Sergey Shirobokov, Borislav Kozlovskii, Francisco J. R. Ruiz, Abbas Mehrabian, M. Pawan Kumar, Abigail See, Swarat Chaudhuri, George Holland, Alex Davies, Sebastian Nowozin, Pushmeet Kohli, and Matej Balog. AlphaEvolve: A coding agent for scientific and algorithmic discovery, 2025. arXiv:2506.13131.

[Per12] Jerome K Percus. Combinatorial methods, 2012.

[Pin94] Iosif Pinelis. An approach to inequalities for the distributions of infinite-dimensional martingales. *Progress in Probability*, 30:128–134, 1994.

[PT21] Ojas Parekh and Kevin Thompson. Application of the level-2 quantum lasserre hierarchy in quantum approximation algorithms. In *International Conference on Approximation Algorithms for Combinatorial Optimization Problems*, 2021.

[PT22] Ojas Parekh and Kevin Thompson. An optimal product-state approximation for 2-local quantum hamiltonians with positive terms. *arXiv preprint arXiv:2206.08342*, 2022.

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 245–254, 2008.

[RRS17] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random csps below the spectral threshold, 2017.

[Ste05] Daureen Steinberg. Computation of matrix norms with applications to robust optimization, 2005.

[TRZ+23] Jun Takahashi, Chaithanya Rayudu, Cunlu Zhou, Robbie King, Kevin Thompson, and Ojas Parekh. An su(2)-symmetric semidefinite programming hierarchy for quantum max cut, 2023.

[TSSW00] Luca Trevisan, Gregory B. Sorkin, Madhu Sudan, and David P. Williamson. Gadgets, approximation, and linear programming. *SIAM Journal on Computing*, 29(6):2074–2097, 2000.

[WBP16] Tengyao Wang, Quentin Berthet, and Yaniv Plan. Average-case hardness of rip certification, 2016.

[WBS16] Tengyao Wang, Quentin Berthet, and Richard J. Samworth. Statistical and computational trade-offs in estimation of sparse principal components, 2016.

[WCE+24] Adam Bene Watts, Anirban Chowdhury, Aidan Epperly, J. William Helton, and Igor Klep. Relaxations and exact solutions to quantum max cut via the algebraic structure of swap operators. *Quantum*, 8:1352, 2024.

[Yao82] Andrew C Yao. Theory and application of trapdoor functions, 1982.

[YP21] Chenyang Yuan and Pablo A. Parrilo. Semidefinite relaxations of products of nonnegative forms on the sphere, 2021.

[YP22] Chenyang Yuan and Pablo A. Parrilo. Maximizing products of linear forms, and the permanent of positive semidefinite matrices, 2022.

[ZX18] Anru Zhang and Dong Xia. Tensor svd: Statistical and computational limits, 2018.

Appendix A

Appendix for Chapter 3

A.1 Proof of Lemma 3.4.3

In this section we prove Lemma 3.4.3. We use the following corollary which we will prove in Section A.2.

Corollary A.1.1. *Let $f : \mathbb{R}_{>0} \to \mathbb{R}$ be an absolutely continuous 2-concave increasing function with $0 \in \overline{\text{Range}(f)}$. Let $\mathcal{F} \in \{\mathbb{R}, \mathbb{C}\}$, and let $z \in \mathcal{F}^n$ be a vector with $\|z\|_{\ell_2} = 1$ and $\|z\|_\infty \leq \delta$. For each i, let σ_i be an independently and uniformly sampled member of $\{-1, +1\}$ if $\mathcal{F} = \mathbb{R}$ and be an independently and uniformly sampled member of $\{-1, +1, -i, +i\}$ otherwise. Then there exists a universal $C > 0$ such that $Z = \sum_{i\in[n]} \sigma_i z_i$ satisfies*

$$[Z]_f^f \leq [g]_f^f + C \cdot \left(- \int_0^{C\delta} \min(0, f(u)) + \delta \cdot \left(f(2\sqrt{\log(1/\delta)}) + f'(1) \right) \right),$$

where $g \sim \mathcal{F}\mathbb{N}(0, 1)$.

Now we are ready to prove the main result of this section.

Proof of Lemma 3.4.3. Let $z = x/\|x\|_2$ and $y = Ez = Ex/\|x\|_2$. Let m be the number of rows of E. Note that for $i \in [m]$, we have

$$y_i = \sum_{j\in[n]} E_{i,j} \cdot z_j.$$

Let $Z_j = E_{i,j} \cdot z_j$ be the random variable where i chosen uniformly at random from $[m]$. We invoke Corollary A.1.1 on $z/\sqrt{k}$. We verify its conditions: First, $\left\| \frac{z}{\sqrt{k}} \right\|_{\ell_2} = \|z\|_2 = 1$. Second, by definition of $E_k^{(\mathcal{F})}$ we can write $Z_j = \sigma_j \cdot \frac{z_j}{\sqrt{k}}$, where the random variables $\sigma_j \in \{+1, -1\}$ when $\mathcal{F} = \mathbb{R}$ and $\sigma_j \in \{+1, -1, +i, -i\}$ chosen independently and uniformly at random Lastly,

$$\frac{|z_j|}{\sqrt{k}} = \frac{|x_j|}{\sqrt{k}\,\|x\|_2} = \frac{|x_j|}{\|x\|_{\ell_2}} \leq \frac{\|x\|_\infty}{\|x\|_{\ell_2}} \leq \delta.$$

Now by invoking Lemma A.2.1 on the random variables $Z_1, \dots, Z_k$, we get

$$\begin{aligned}\left[\frac{Ex}{\|x\|_2}\right]_f^f &= [y]_f^f = \mathbb{E}_{i\sim[m]} f(y_i) = \mathbb{E}\left[f\left(\sum_{j\in[n]} Z_j\right)\right] \\ &\leq [g]_f^f + C\cdot\left(-\int_0^{C\delta}\min(0, f(u)) + \delta\cdot\left(f(2\sqrt{\log(1/\delta)}) + f'(1)\right)\right),\end{aligned}$$

as desired.

Now assume $f = f_p$ for some $-1 < p < 2$. By Lemma 3.2.7, $\|\cdot\|_p$ is homogeneous, and we get

$$\begin{aligned}[Ex]_f &= \|x\|_2 \cdot \left[\frac{Ex}{\|x\|_2}\right]_f \\ &\leq \|x\|_2 \cdot f^{-1}\left(\gamma_{\mathcal{F},p}^p + C\cdot\left(-\int_0^{C\delta}\min(0, f(u)) + \delta\cdot\left(f(2\sqrt{\log(1/\delta)}) + f'(1)\right)\right)\right)\end{aligned}$$

The second term is 0 if $p > 0$, otherwise it is equal to $\frac{\delta^{p+1}}{p+1}$. The third term is 2δ if $p < 0$, otherwise it is bounded by $4\delta(\log(1/\delta)+1)$ for δ small enough. Therefore, the limit of the right hand side as $\delta \to 0$ is $\|x\|_2 \cdot \gamma_{F,p}$. □

A.2 Proof of Corollary A.1.1

Corollary A.1.1 follows directly as a special case of the following Lemma, for $k = 1$ if $\mathcal{F} = \mathbb{R}$ and for $k = 2$ if $\mathcal{F} = \mathbb{C}$.

Lemma A.2.1. *Let $f : \mathbb{R}_{>0} \to \mathbb{R}$ be an absolutely continuous 2-concave increasing function with $0 \in \overline{\text{Range}(f)}$. Let $k \geq 1$ and let $X_1, \dots, X_n$ are bounded independent random variables in $\mathbb{R}^k$ with $\mathbb{E}[X_i] = 0$, $Cov(\sum_i X_i) = I_k/k$, and $\|X_i\|_{\ell_2} \leq \delta_i$ such that $\sum_i \delta_i^2 \leq 1$ and $\delta_i \leq \delta$ for some $0 < \delta < 1$, then there exists $\eta_k > 0$ such that*

$$\begin{aligned}\left[\left\|\sum_{i\in[n]} X_i\right\|_{\ell_2}\right]_f^f \leq [\|g\|_{\ell_2}]_f^f - e\int_0^{C_k\delta/e}\min(0, f(u))\,du \\ + \delta\cdot\left(C_k\cdot\max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1)\right),\end{aligned}$$

where $g \sim N(0, I_k/k)$, and C_k is the constant in Theorem A.2.2.

Before proving Lemma A.2.1, we state some probabilistic tools we will require in the proof. The multivariate Berry-Esseen theorem below is due to [Ben05].

Theorem A.2.2 (Multivariate Berry-Esseen)**.** *Let $k \geq 1$ and let $X_1, \dots, X_n$ be independent random variables in $\mathbb{R}^k$ satisfying $\mathbb{E}[X_i] = 0$ for $1 \leq i \leq n$. Define $X = X_1 + \cdots + X_n$ and*

suppose $\mathbb{E}[XX^T] = I_k$. *Further let* $g \sim \mathbb{N}(0, I_k)$. *Then there exists* $C_k > 0$ *such that for all convex sets* $U \subseteq \mathbb{R}^k$ *it holds that*

$$|\Pr[g \in U] - \Pr[X \in U]| \leq C_k \cdot \mathbb{E}_{i\in[n]}[\|X_i\|_{\ell_2}^3].$$

In particular, for all $u \geq 0$ *it holds that*

$$|\Pr[\|g\|_{\ell_2} \leq u] - \Pr[\|X\|_{\ell_2} \leq u]| \leq C_k \cdot \mathbb{E}_{i\in[n]}[\|X_i\|_{\ell_2}^3].$$

The next form of multivariate Hoeffding is due to [Pin94, Theorem 3].

Theorem A.2.3 (Multivariate Hoeffding)**.** *Let* $k \geq 1$ *and let* $X_1, \ldots, X_n$ *be independent random variables in* $\mathbb{R}^k$ *satisfying* $\mathbb{E}[X_i] = 0$ *for* $1 \leq i \leq n$. *Further suppose* $\|X_i\|_{\ell_2} \leq \delta_i$ *almost surely for* $\delta_1, \ldots, \delta_n > 0$. *Define* $X = X_1 + \cdots + X_n$.

$$|\Pr[\|X\|_{\ell_2} \geq u]| \leq 2 \cdot e^{-u^2/2\sum_{i\in[n]}\delta_i^2}.$$

Fact A.2.4. *Let* $f : \mathbb{R}_{>0} \to \mathbb{R}$ *be an absolutely continuous function. If* $\int_0^b f(u)$ *is bounded then* $\lim_{u\to 0} uf(u) = 0$.

Claim A.2.5. *Let* X *be a random variable over* $\mathbb{R}_{>0}$, *and let* $f : \mathbb{R}_{>0} \to \mathbb{R}$ *be an absolutely continuous function with* $f(a) = 0$ *for some* $a \in \mathbb{R}_{>0}$. *Then, if* $\mathbb{E}[f(X)]$ *exists,*

$$\begin{aligned}\mathbb{E}[f(X)] = &-\int_0^a f'(u)\Pr[X \leq u]\,du + \int_a^\infty f'(u)\Pr[X \geq u]\,du \\ &+ \lim_{u\to 0} f(u)\Pr[X \leq u] - \lim_{u\to\infty} f(u)\Pr[X \geq u].\end{aligned}$$

Proof. Let $\mu : \mathbb{R}_{>0} \to \mathbb{R}$ denote the PDF of the random variable X.

$$\begin{aligned}\mathbb{E}[f(X)] &= \int_0^\infty f(u)d\mu(u) \\ &= \int_0^a f(u)d\mu(u) + \int_a^\infty f(u)d\mu(u) \\ &= f(u)\Pr[X \leq u]\Big|_0^a - \int_{I_1} f'(u)\Pr[X \leq u] - f(u)\Pr[X \geq u]\Big|_a^\infty + \int_{I_2} f'(u)\Pr[X \geq u] \\ &= \lim_{u\to 0} f(u)\Pr[X \leq u] - \lim_{u\to\infty} f(u)\Pr[X \geq u] \\ &\quad - \int_{I_1} f'(u)\Pr[X \leq u] + \int_{I_2} f'(u)\Pr[X \geq u],\end{aligned}$$

where the last equality uses $f(a) = 0$. □

Fact A.2.6. *Let* $k \geq 1$ *and let* $g \sim N(0, I_k/k)$. *Then for all* $0 \leq u \leq 1$, $\Pr[\|g\|_{\ell_2} \leq u] \leq eu$.

Proof. $k \cdot \|g\|_{\ell_2}^2$ is distributed as a Chi-squared random variable with k degrees of freedom. In [DG03, Lemma 2.2], the authors show that

$$\Pr[\|g\|_{\ell_2} \leq u] \leq (u^2 e^{1-u^2})^{k/2} \leq e \cdot u,$$

as desired. □

With these facts in hand, we are ready to prove Lemma A.2.1.

Proof of Lemma A.2.1. Define the random variable $X = \|\sum_i X_i\|_{\ell_2}$. We split the domain of f into $I_1 = f^{-1}([-\infty, 0])$ and $I_2 = f^{-1}([0, \infty])$, where $I_1 = [0, a]$ and $I_2 = [a, \infty]$. Since $0 \in \overline{\text{Range}(f)}$, we have $f(a) = 0$. We use Claim A.2.5 to write the left hand side as

$$\begin{aligned}\mathbb{E}[f(X)] &= -\int_0^a f'(u)\Pr[X \le u] + \int_a^\infty f'(u)\Pr[X \ge u] \\ &\quad + \lim_{u\to 0} f(u)\Pr[X \le u] - \lim_{u\to\infty} f(u)\Pr[X \ge u] \\ &\le -\int_0^a f'(u)\Pr[X \le u] + \int_a^\infty f'(u)\Pr[X \ge u].\end{aligned}$$

The final inequality uses $f(0) \le f(a) = 0$ and $f(\infty) \ge f(a) = 0$.

We bound each of the above integrals separately.

Claim A.2.7. *We have*

$$\int_0^a f'(u)\Pr[X \le u] \ge \int_0^a f'(u)\Pr[\|g\|_{\ell_2} \le u] + e\int_0^{C_k\delta/e} \min(0, f(u)).$$

Claim A.2.8. *We have*

$$\int_a^\infty f'(u)\Pr[X \ge u] \le \int_a^\infty f'(u)\Pr[|g| \ge u] + \delta\cdot\left(C_k\cdot\max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1)\right).$$

We will complete the proof of Lemma A.2.1 using these two claims, and prove the claims later.

$$\begin{aligned}\mathbb{E}[f(X)] &\le -\int_0^a f'(u)\Pr[X \le u] + \int_a^\infty f'(u)\Pr[X \ge u] \\ &\le -\int_0^a f'(u)\Pr[\|g\|_{\ell_2} \le u] - e\int_0^{C_k\delta/e} \min(0, f(u)) \\ &\quad + \int_a^\infty f'(u)\Pr[|g| \ge u] + \delta\cdot\left(C_k\cdot f(2\sqrt{\log(1/\delta)}) + 2f'(1)\right) \\ &= \mathbb{E}[f(\|g\|_{\ell_2})] - \int_0^{C_k\delta} \min(0, f(u)) + \delta\cdot\left(C_k\cdot\max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1)\right).\end{aligned}$$

Here, the last equality is by applying Claim A.2.5 on the random variable $\|g\|_{\ell_2}$:

$$\begin{aligned}\mathbb{E}[\|g\|_{\ell_2}] &= -\int_0^a f'(u)\Pr[\|g\|_{\ell_2} \le u] + \int_a^\infty f'(u)\Pr[|g| \ge u] \\ &\quad + \lim_{u\to 0} f(u)\Pr[\|g\|_{\ell_2} \le u] - \lim_{u\to\infty} f(u)\Pr[\|g\|_{\ell_2} \ge u]\end{aligned}$$

Observe that by Fact A.2.4, the third term above is zero, and by Claim 3.2.6,

$$\lim_{u\to\infty} f(u)\Pr[\|g\|_{\ell_2} \ge u] \le \lim_{u\to\infty} (f'(1)\cdot u^2 + f(1))\cdot\Pr[\|g\|_{\ell_2} \ge u] = 0.$$

This completes the justification of the last inequality. It remains to prove Claim A.2.7 and Claim A.2.8. We begin by writing an inequality which will be used in both proofs. By Theorem A.2.2, for any $u \geq 0$,

$$\begin{aligned} |\Pr[X \leq u] - \Pr[\|g\|_{\ell_2} \leq u]| &\leq C_k \cdot \mathbb{E}_{i\in[n]}[\|X_i\|_{\ell_2}^3] \\ &\leq C_k \cdot \sum_{i\in[n]} [\|X_i\|_{\ell_2}^2] \cdot \max_{i\in[n]} \|X_i\|_{\ell_2} \\ &\leq C_k \cdot \delta. \end{aligned} \tag{A.1}$$

Proof of Claim A.2.7. For a parameter b with $0 \leq b \leq a$, we write

$$\begin{aligned} \int_{I_1} f'(u) \Pr[X \leq u] &\geq \int_b^a f'(u) \Pr[X \leq u] && (f'(u) \geq 0) \\ &\geq \int_b^a f'(u)(\Pr[\|g\|_{\ell_2} \leq u] - C_k\delta) && (\text{Eq. (A.1)}) \\ &= \int_0^a f'(u) \Pr[\|g\|_{\ell_2} \leq u] - \int_0^b f'(u) \Pr[\|g\|_{\ell_2} \leq u] - C_k\delta(f(a) - f(b)) \\ &\geq \int_0^a f'(u) \Pr[\|g\|_{\ell_2} \leq u] - e\int_0^b f'(u)u + C_k\delta f(b) \\ & && (\text{using } f(a) = 0, \text{ and Fact A.2.6}) \\ &= \int_0^a f'(u) \Pr[\|g\|_{\ell_2} \leq u] - euf(u)\Big|_0^b + e\int_0^b f(u) + C_k\delta f(b) \\ &= \int_0^a f'(u) \Pr[\|g\|_{\ell_2} \leq u] - ebf(b) + e\int_0^b f(u) + C_k\delta f(b) && (\text{Fact A.2.4}) \\ &= \int_0^a f'(u) \Pr[\|g\|_{\ell_2} \leq u] + e\int_0^{\min(a, C_k\delta/e)} f(u) \\ & && (\text{Setting } b = \min(a, C_k\delta/e)) \\ &= \int_0^a f'(u) \Pr[\|g\|_{\ell_2} \leq u] + e\int_0^{C_k\delta/e} \min(f(u), 0). \end{aligned}$$

The penultimate equality is because if $\min(a, C_k\delta/e) = a$, then $f(b) = 0$, and if $\min(a, C_k\delta/e) = C_k\delta/e$, then $bf(b) = C_k\delta f(b)/e$.

□

Proof of Claim A.2.8. Applying Theorem A.2.3 on X,

$$\Pr[X \geq u] \leq 2 \cdot e^{-u^2/2\sum\delta_i^2} \leq 2 \cdot e^{-u^2/2}. \tag{A.2}$$

For $\max(a,1) < c < \infty$ we write

$$\begin{aligned}
\int_{I_2} f'(u)\Pr[X \geq u] &= \int_a^c f'(u)\Pr[X \geq u] + \int_c^\infty f'(u)\Pr[X \geq u] \\
&\leq \int_a^c f'(u)(\Pr[|g| \geq u] + C_k\delta) + \int_c^\infty 2f'(u)e^{-u^2/2} \\
&\qquad \text{(Eq. (A.1), Eq. (A.2))} \\
&\leq \int_a^c f'(u)\Pr[|g| \geq u] + C_k\delta \cdot (f(c) - f(a)) + 2f'(1) \cdot \int_c^\infty ue^{-u^2/2} \\
&\qquad \text{(using } f \text{ is 2-concave, Claim 3.2.6, and } c \geq 1) \\
&\leq \int_a^\infty f'(u)\Pr[|g| \geq u] + C_k\delta \cdot f(c) + 2f'(1) \cdot \int_c^\infty ue^{-u^2/2} \\
&\qquad (f \text{ is increasing, } f(a) = 0) \\
&= \int_a^\infty f'(u)\Pr[|g| \geq u] + C_k\delta \cdot f(c) + 2f'(1) \cdot e^{-c^2/2} \qquad \text{(A.3)}
\end{aligned}$$

Setting $c = \max(a, 2\sqrt{\log(1/\delta)})$, we get

$$\int_{I_2} f'(u)\Pr[X \geq u] \leq \int_a^\infty f'(u)\Pr[|g| \geq u] + \delta \cdot \left(C_k \cdot \max(0, f(2\sqrt{\log(1/\delta)})) + 2f'(1)\right). \tag{A.4}$$

The inequality above is by bounding $e^{-c^2/2} \leq \delta$, and if $\max(a, 2\sqrt{\log(1/\delta)}) = a$, then $f(c) = 0$. □

□

A.3 Algorithm for approximating $2 \to q$ norm

Let $q < 2$. Given a matrix $A \in \mathcal{F}^{n\times d}$ with rows $\{a_i\}_{i\in[n]}$, we write an expression for $\|A\|_{2\to q}$:

$$\|A\|_{2\to q} = f_q^{-1}\left(\max_{x\in\mathcal{F}^d, \|x\|_2=1}\left[\sum_{i\in[n]} f_q\left(|\langle a_i, x\rangle|\right)\right]\right) = f_q^{-1}\left(\max_{x\in\mathcal{F}^d, \|x\|_2=1}\left[\sum_{i\in[n]} f_q\left(\sqrt{a_i^\dagger x x^\dagger a_i}\right)\right]\right).$$

Notice that $xx^\dagger$ is a rank-1 PSD matrix with $\mathrm{tr}(xx^\dagger) = x^\dagger x = \|x\|_{\ell_2} = d \cdot \|x\|_2 = d$. We can relax the rank 1 constraint to obtain a relaxation of $\|A\|_{2\to q}$ which we denote by $\mathrm{SDP}_{2\to q}(A)$:

$$\mathrm{SDP}_{2\to q}(A) := f_q^{-1}\left(\max_{X\succeq 0, \mathrm{tr}(X)=d}\left[\sum_{i\in[n]} f_q\left(\sqrt{a_i^\dagger X a_i}\right)\right]\right),$$

where the maximum is taken over all matrices in $\mathcal{F}^{d\times d}$. Note that the objective function being maximized is concave for all $q \leq 2$ because the function $X \to a_i^\dagger X a_i$ is linear and f_q is 2-concave.

Lemma A.3.1. *For any matrix $A \in \mathcal{F}^{n\times d}$ and any $-1 < q \le 2$,*

$$\|A\|_{2\to q} \le \mathrm{SDP}_{2\to q}(A) \le \gamma_{\mathcal{F},q}^{-1} \cdot \|A\|_{2\to q}.$$

Proof. The first inequality is because $\mathrm{SDP}_{2\to q}(A)$ is a relaxation of $\|A\|_{2\to q}$. For the second inequality, we describe a rounding procedure for the relaxation.

Let $X^* \in \mathcal{F}^{d\times d}$ be the optimal solution to $\mathrm{SDP}_{2\to q}(A)$. We sample a vector $x \sim \mathcal{F}\mathbb{N}(0, X^*)$. Clearly,

$$\|A\|_{2\to q}^2 = \max_{x\in\mathcal{F}^n, x\neq 0} \frac{\|Ax\|_q^2}{\|x\|_2^2} \ge \frac{\mathbb{E}\|Ax\|_q^2}{\mathbb{E}\|x\|_2^2}. \tag{A.5}$$

First we calculate the denominator of the right hand side: $\mathbb{E}\|x\|_2^2 = \frac{1}{d}\cdot \mathrm{tr}(X) = 1$. For the numerator, we bound

$$\begin{aligned}
\mathbb{E}\|Ax\|_q^2 &= \mathbb{E}\left(f_q^{-1}\left(\sum_{i\in[n]} f_q(|\langle a_i, x\rangle|)\right)^2\right) \\
&\ge f_q^{-1}\left(\sum_{i\in[n]} \mathbb{E} f_q(|\langle a_i, x\rangle|)\right)^2 \qquad \text{(Jensen's inequality, and } x\to f_q^{-1}(x)^2 \text{ is convex)} \\
&= f_q^{-1}\left(\sum_{i\in[n]} \mathbb{E}_{g\sim\mathcal{F}\mathbb{N}(0,1)} f_q\left(\sqrt{a_i^\dagger X^* a_i}\cdot |g|\right)\right)^2 \\
&\qquad\qquad (\langle a_i, x\rangle \text{ is a Gaussian with variance } a_i^\dagger X^* a_i) \\
&= f_q^{-1}\left(\mathbb{E}_{g\sim\mathcal{F}\mathbb{N}(0,1)} f_q(|g|)\right)^2 \cdot f_q^{-1}\left(\sum_{i\in[n]} f_q\left(\sqrt{a_i^\dagger X^* a_i}\right)\right)^2 \qquad \text{(Homogeneity)} \\
&= \gamma_{\mathcal{F},q}^2 \cdot \mathrm{SDP}_{2\to q}(A)^2.
\end{aligned}$$

Together with Eq. (A.5), this implies $\|A\|_{2\to q} \ge \gamma_{\mathcal{F},q}\cdot \mathrm{SDP}_{2\to q}(A)$, completing the proof. □

Appendix B

Appendix for Chapter 4

B.1 Proof of Lemma 4.2.7

For the reader's convenience we restate Lemma 4.2.7 below. The code for verifying the numerical claims made in the below proof is available at `https://github.com/anshnagda/EPR-algorithm`.

Lemma B.1.1 (Restatement of Lemma 4.2.7)**.** *For any integer k, there is an efficient algorithm to compute the description of a k-qubit state ρ_k satisfying the following, where $\gamma = \frac{\sqrt{5}-1}{4}$ and $\theta \geq 1/2$ is the solution to $\sqrt{\theta(1-\theta)} = \gamma$.*

1. *For all $i \in [k]$, $\operatorname{tr}(\rho_k \cdot E_{i,i+1(\text{mod k})}) > \frac{3}{4} \cdot \frac{1+\sqrt{5}}{4}$.*
2. *For all $i \in [k]$, the 1-qubit marginal $\operatorname{tr}_{[k]-i}[\rho_k]$ equals $\theta \left|0\right\rangle \left\langle 0\right| + (1-\theta) \left|1\right\rangle \left\langle 1\right|$.*
3. *For all $i, j \in [k]$, $\operatorname{tr}(\rho_k \cdot E_{ij}) > \frac{1}{2} \cdot \frac{1+\sqrt{5}}{4}$.*

Proof. For $k \leq 5$, we numerically verify the claim by explicitly solving an SDP.

Similarly, one can numerically verify that there exists a 5-qubit state ψ satisfying the following:

1. $\frac{1}{4} \cdot \sum_{i \in [4]} \operatorname{tr}(E_{i,i+1} \cdot \psi) \geq 0.668$.
2. For all $i, j \in [5]$, $\operatorname{tr}(E_{ij} \cdot \psi) > \frac{1}{2} \cdot \frac{1+\sqrt{5}}{4}$.
3. For all $i \in [5]$, the 1-qubit marginal $\operatorname{tr}_{[5]-i}[\psi]$ equals $\theta \left|0\right\rangle \left\langle 0\right| + (1-\theta) \left|1\right\rangle \left\langle 1\right|$.

Now assume $k \geq 7$. Define the state

$$\rho'_k = \left|\mathrm{EPR}_\theta\right\rangle \left\langle \mathrm{EPR}_\theta\right|^{\otimes \frac{k-5}{2}} \otimes \psi.$$

Let Shift_i be the unitary that shifts qubits in the k-cycle i spots. We will set ρ_k to be the shift-invariant state

$$\rho_k = \mathbb{E}_{i \sim [k]} \left[\mathsf{Shift}_i \cdot \rho'_k \cdot \mathsf{Shift}_i^\dagger \right].$$

We will verify that ρ_k satisfies the three requirements.

1. Let $i \in [k]$. We can write the energy $\mathrm{tr}(\rho_k \cdot E_{i,i+1(\mathrm{mod\ k})})$ as the average energy of a random edge $\{j, j+1(\mathrm{mod}\ k)\}$ under ρ'_k. By definition of ρ'_k, one can compute

$$\mathrm{tr}(\rho'_k \cdot E_{j,j+1(\mathrm{mod\ k})}) \geq \begin{cases} \frac{1+\sqrt{5}}{4} & j \leq k-5 \text{ odd}, \\ \frac{1+\sqrt{5}}{8} & j \leq k-5 \text{ even or } j = k, \\ 0.668 & k-5 < j < k. \end{cases}$$

   Therefore

$$\mathrm{tr}(\rho_k \cdot E_{i,i+1(\mathrm{mod\ k})}) = \frac{\frac{k-5}{2} \cdot \frac{1+\sqrt{5}}{4} + \frac{k-5}{2} \cdot \frac{1+\sqrt{5}}{8} + 4 \cdot 0.668 + \frac{1+\sqrt{5}}{8}}{k}.$$

   One can verify that $4 \cdot 0.668 + \frac{1+\sqrt{5}}{8} > 5 \cdot \frac{3}{4} \cdot \frac{1+\sqrt{5}}{4}$, implying

$$\mathrm{tr}(\rho \cdot E_{i,i+1(\mathrm{mod\ k})}) > \frac{(k-5) \cdot \frac{3}{4} \cdot \frac{1+\sqrt{5}}{4} + 5 \cdot \frac{3}{4} \cdot \frac{1+\sqrt{5}}{4}}{k} = \frac{3}{4} \cdot \frac{1+\sqrt{5}}{4}.$$

2. It suffices to prove that the 1-qubit marginals of ρ'_k are equal to $\theta \left|0\right\rangle \left\langle 0\right| + (1-\theta) \left|1\right\rangle \left\langle 1\right|$. For qubits $i \leq k-5$, this follows from Claim 4.2.5, and for $i > k-4$, this follows by definition of ψ.

3. Let $i \neq j$. It suffices to prove $\mathrm{tr}(\rho'_k \cdot E_{ij}) > \frac{1}{2} \cdot \frac{1+\sqrt{5}}{4}$ and the same for ρ'_k. We consider three cases.

   - If $j = i+1$ for some odd $i \leq k-5$, this is implied by Item 1.
   - If $k-5 < i, j \leq k$, this is implied by the definition of ψ.
   - Otherwise, the 2-qubit marginal $\mathrm{tr}_{[n]\setminus\{i,j\}}(\rho'_k)$ equals $(\theta \left|0\right\rangle \left\langle 0\right| + (1-\theta) \left|1\right\rangle \left\langle 1\right|)^{\otimes 2}$, and this is implied by Claim 4.2.5.

□

Appendix C

Appendix for Chapter 6

C.1 Proofs of Average Case Theorems

C.1.1 Proof of Theorem 6.1.7

Below we specify three Ramanujan graphs G_4^{MC}, G_3^{IS}, and G_4^{IS} in the sparse6 format [McK25]. We also specify a cut S_4^{MC} (in the form of a subset of vertices), and independent sets S_3^{IS} and S_4^{IS} that witness lower bounds on the actual values of $\mathrm{MC}(G_4^{\mathrm{MC}})$, $\mathrm{IS}(G_3^{\mathrm{IS}})$, and $\mathrm{IS}(G_4^{\mathrm{IS}})$. The vertices of all graphs are 0-indexed, as dictated by the sparse6 format. We start by defining G_4^{MC} and S_4^{MC}:

```
>>sparse6<<:~?@{‘oQ?bOCHBGIPBHECdOmReom\\A?aECwaAEGKCEiKIGi[NfG?PCwka_AEijomS_qIadPetIJ[
     DFaOlgqyXKbYAIqolbpi?@@qVJJmx‘ryfhGSQDhSfKhonOhOwOYSudgKQ_OoYaPS^e@tGgH]PHhAWMC}ECCaIJS]
     eIsm@EbQ@@q@OdAgodctRbAwsfQaZNGWnOtQNGryNESzLRxwfJTmA@qasNSeDBcz@QWgQJB}
     oODDYlUDciRLCTwwVHbAXNeQWJrDfcASuWxOcNR}ELT@Q_rCrVx[eLRYZJc\\]aOh?UH{cOSadKcTMmsX\\
     YgsdQdiZNstTfBPOVwSHJcIOFspXbcHZWn

 0 2 6 7 13 15 16 17 19 20 21 24 25 27 29 31 33 35 36 37 40 42 47 49 50 54 55 56 58 60 61 62 63 65 67 69 72
       73 74 77 78 79 83 84 85 86 89 96 99 101 102 103 104 105 110 111 112 113 114 117 120 121 123
```

G_3^{IS} and S_3^{IS} are defined by the following.

```
>>sparse6<<:cb?gGMGE_OGo]FBDoggiAGabYCETCESa\\YGHFTGC}aTLObPmcVLqQ|KrKNraaAIXO~\n

0 15 29 6 4 14 20 28 13 7 12 35 16 30 31 22 32
```

Finally, G_4^{IS} and S_4^{IS} are defined by the following.

```
>>sparse6<<:~?Ab_O?_g@a??@gMa_K_WJ‘WNacX?cD?sMbwA‘?P_O[cWF@_Q_CF@Kf?Wa‘o\\CK[C{]Ck_dgWBS]C_f‘OLD?jeG^
     cS@egdc_ea[x@?u_c{?sVBKXf?}bSDAwhgW@ECQF[B@WZE{LCOk‘OWEstFs?B{QgSHE?}gCsHcDB?khkkEG|fKjaGlHSH@XBbWw__o
     ‘@S_wkHSRID[@?NEcReXX‘OjJkvF‘\\JsKAXLIS]DKta_sHHTbGqH{CGPQa?ybW|K{PBgeJcICPCH[‘‘GwFHMbhBI[
     eHKNBcSEGr_wzJs~J@ne@JkkvK[{J[Ef‘Magz‘_wJ\\IIXvePWapCgH\\L[VEX?g@AM‘tms[BwzIS@BotLs_E‘RNtYMPuagbM‘
     vc_eI|gMXwdpGNlbMXtb@LKxs‘xVKQGgPPdPallxPINfPE_OUItZLYLQKBNAFaPJNi@dhxP{yNYIQK\\
     MAUeHOMxyaHOKqMdPFJQ@ip{OIQ‘h^NaLfxmN|GLyDRCMKYMP|ZLXpQCEEQSRv\n

 0 1 3 4 7 8 9 10 11 12 13 15 18 24 26 27 28 31 34 40 42 43 44 46 48 53 54 55 56 59 60 61 62 65 75 77 78 82
       84 90 91 93 94 98 100 101 103 105 107 109 110 112 114 123 124 125 127 128 129 132 134 135 136 138
      141 144 145 146 150 153 154 159 160 162
```

The following properties can easily be verified and immediately imply Theorem 6.1.7.

1. G_3^{IS} is a 3-regular Ramanujan graph on 36 vertices. G_4^{MC} and G_4^{IS} are 4-regular Ramanujan graphs on 124 and 163 vertices respectively.

2. S_3^{IS} and S_4^{IS} are independent sets in G_3^{IS} and G_4^{IS} respectively with $|S_3^{\mathrm{IS}}| = 17$ and $|S_4^{\mathrm{IS}}| = 74$. Consequently, $\mathrm{IS}(G_3^{\mathrm{IS}}) \geq 17/36$, and $\mathrm{IS}(G_4^{\mathrm{IS}}) \geq 74/163$.

3. The number of edges in G_4^{MC} with exactly one endpoint in S_4^{MC} is equal to 226. Consequently, $\mathrm{MC}(G_4^{\mathrm{MC}}) \geq 113/124$.

It can be easily verified that all the graphs are Ramanujan proving the claimed lower bounds on γ_4^{MC}, γ_3^{IS}, and γ_4^{IS}.

C.1.2 Proof of Theorem 6.1.8

The goal of this section is to prove improved upper bounds on σ_d^{MC} and σ_d^{IS}. The same upper bounds apply to γ_d^{MC} and γ_d^{IS}, as will be clear in the proof. We will give a detailed argument for the Max-Cut case, and sketch the (minor) changes required for Max-Independent-Set at the end.

As with previous refutation algorithms [Hof03, Hae21], our efficient certificate will be a bound on the *spectral expansion* of a random graph $G \sim \mathcal{G}(n,d)$, which is defined as $\lambda^*(G) = \max_{i>1} |\lambda_i|$ where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the eigenvalues of the adjacency matrix $\mathbf{A}$ of the graph G. Friedman's theorem [Fri08] proves that λ^* concentrates around $2\sqrt{d-1}$.

Theorem C.1.1. *For all $d \in \mathbb{N}$ and $\epsilon > 0$, with probability $1 - o_n(1)$ over $G \sim \mathcal{G}(n,d)$, we have $\lambda^*(G) \leq 2\sqrt{d-1} + \epsilon$.*

Definition C.1.2 (d-ary tree)**.** The d-ary tree of depth L, $T_{d,L}$ is a rooted tree, where the root has d children, and all non-root vertices at distance $< L$ from the root have $d-1$ children.

Our improvement on Hoffman's classical bounds [Hof03, Hae21] consists of concluding stronger bounds on $\mathrm{MC}(G)$ and $\mathrm{IS}(G)$ given that $\lambda^*(G) \leq \lambda$ for $\lambda = 2\sqrt{d-1} + \epsilon$. Let us denote by $\Omega_{d,\lambda}$ the set of d-regular graphs G such that $\lambda^*(G) \leq \lambda$. We will require the notion of a *labeled* d-ary tree, which we define below.

Definition C.1.3 (Labeled d-ary tree)**.** A $\{\pm 1\}$-labeling of $T_{d,L}$ is a mapping y from the vertices of $T_{d,L}$ to $\{\pm 1\}$.

Let $\Lambda_{d,L}$ be the collection of distributions over $\{\pm 1\}$-labelings of $T_{d,L}$, where the vertices are labeled by elements of $\{\pm 1\}$. For most of the remainder of this section, we will show how we obtain our upper bounds on $\mathrm{MC}(G)$ given $G \in \Omega_{d,\lambda}$. Our upper bound will be parametrized by a nonnegative integer $L \in \mathbb{N}$. Interestingly, we will exactly recover Hoffman's bound [Hof03] when $L = 1$. The idea is essentially to write a *finite* LP relaxation of the problem of finding the supremum of the cut fraction $|\delta_G(S, \bar{S})|/|E(G)|$ over all $G \in \Omega_{d,\lambda}$, by projecting all the information about G and S onto "local" neighborhoods at distance L.

Let $G \in \Omega_{d,\lambda}$ be an n-vertex graph, and let $S \subseteq V(G)$ be a cut in G. Let $\alpha \in [0, 1/2]$ be a value that approximates $|S|/n$, in the sense that $|\alpha - |S|/n| \leq \delta$. Define the cut vector $\mathbf{x} \in \{-1,1\}^n$ by $x_i = (-1)^{\mathbf{1}\{i \in S\}}$. We will associate the pair (G, S) with a distribution $\mu_{G,S} \in \Lambda_{d,L}$, which can be sampled from as follows:

- Sample a random ordering of the neighboring edges of each vertex in G.
- Sample a random vertex $i \sim [n]$.
- We will label $T_{d,L}$ by the values of x at all length $\leq L$ nonbacktracking walks from i (that is, walks that don't use the same edge twice in a row). Formally, a vertex v of $T_{d,L}$ can be written as a sequence $(q_1, \ldots, q_k)$ for $0 \leq k \leq L$, $1 \leq q_1 \leq d$, and $1 \leq q_l \leq d-1$ for $l > 1$. We will label v by the x value of the final vertex in the corresponding path starting from i in G. In particular, let $i_0 = i$, and for $1 \leq l \leq k$ let i_l be the q_l^{th} neighbor of i_{l-1} in the sampled ordering, excluding i_{l-2} if $l > 1$. We will label v by $y_v := x_{i_k}$.
- Output the resulting labeling y of $T_{d,L}$.

We will deduce some linear constraints on $\mu_{G,S}$, in terms of its probability distribution function.

Lemma C.1.4 (Local Consistency Constraints)**.** *Let r_0 be the root of $T_{d,L}$, and let r_1 be a child of r_0. For $i \in \{0,1\}$, let T_i be the subtree of $T_{d,L}$ rooted at r_i containing all vertices at distance at most L from r_{1-i}. Then, T_0 is isomorphic to T_1, and the marginal distribution of $\mu_{G,S}$ on T_0 is identical (up to the isomorphism) of that on T_1.*

Proof. The marginal distribution on T_0 can be viewed as the following: sample a random vertex i in G and a random neighbor $j \sim i$. Output the labels x at all endpoints of a nonbacktracking walk from i of length at most L if (i,j) is the first edge in the walk, and length at most $L-1$ otherwise.

The marginal distribution on T_1 is identical, with the roles of i and j switched. Since the distribution of (i,j) is identical to that of (j,i), both marginal distributions are identical. □

Lemma C.1.5 (Average Value of root)**.** *We have $2\alpha - 1 - 2\delta \leq \mathbb{E}_{y \sim \mu_{G,S}}[y_0] \leq 2\alpha - 1 + 2\delta$.*

Proof. We can directly compute $\mathbb{E}_{y \sim \mu_{G,S}}[y_0] = \mathbb{E}_{i \sim [n]}[x_i] = 2|S|/n - 1$. The sandwiching inequalities follow from the fact that $|\alpha - |S|/n| \leq \delta$. □

Lemma C.1.6 (Spectral Constraints)**.** *The bound $\lambda^*(G) \leq \lambda$ implies the bounds*

$$(2\alpha - 1 - 2\delta)^2 \leq \mathbb{E}_{y \sim \mu_{G,S}, \ell}[y_0 \cdot y_\ell] \leq (2\alpha - 1 + 2\delta) + (\lambda/d)^L$$

if L is even, and

$$(2\alpha - 1 - 2\delta)^2 - (\lambda/d)^L \leq \mathbb{E}_{y \sim \mu_{G,S}, \ell}[y_0 \cdot y_\ell] \leq (2\alpha - 1 + 2\delta)^2 + (\lambda/d)^L$$

if L is odd. Here 0 denotes the root, and ℓ denotes the (random) endpoint of an L-step random walk in $T_{d,L}$ starting from the root.

Proof. We begin by computing

$$\mathbb{E}_{y \sim \mu_{G,S}, \ell}[y_0 \cdot y_\ell] = \frac{\mathbf{x}^\top (\mathbf{A}/d)^L \mathbf{x}}{n} = \frac{\mathbf{x}^\top \mathbf{A}^L \mathbf{x}}{d^L \cdot n}.$$

Let us write $\mathbf{A} = d\frac{\mathbf{1}\mathbf{1}^\top}{n} + \tilde{\mathbf{A}}$, where $d = \lambda_1$ is the trivial eigenvalue of $\mathbf{A}$ and $\tilde{\mathbf{A}}$ is the portion of $\mathbf{A}$ orthogonal to the trivial eigenvector $\mathbf{1}$. That is, $\tilde{A} \cdot \mathbf{1} = \mathbf{0}$. Note that $\|\tilde{\mathbf{A}}\|_{\mathsf{op}} \leq \lambda$. Let us proceed with our computation.

$$\begin{aligned}\mathbb{E}_{y\sim\mu_{G,S},\ell}[y_0 \cdot y_\ell] &= \frac{d^L \cdot \mathbf{x}^\top \mathbf{1}\mathbf{1}^T \mathbf{x}}{d^L \cdot n^2} + \frac{\mathbf{x}^\top \tilde{\mathbf{A}}^L \mathbf{x}}{d^L \cdot n} \\ &= \frac{(2|S| - n)^2}{n^2} + \frac{\mathbf{x}^\top \tilde{\mathbf{A}}^L \mathbf{x}}{d^L \cdot n}.\end{aligned}$$

The Lemma follows from (1) the fact that $|\alpha - |S|/n| \leq \delta$, and (2) the bounds $|\mathbf{x}^\top \tilde{\mathbf{A}}^L \mathbf{x}| \leq \|\tilde{\mathbf{A}}\|_{\mathsf{op}} \cdot \|\mathbf{x}\|_2^2 \leq \lambda n$, along with the bound $\mathbf{x}^\top \tilde{A}^L \mathbf{x} \geq 0$ if L is even. □

Lemma C.1.7 (Objective Value)**.** *The cut fraction* $|\delta_G(S, \bar{S})|/|E(G)|$ *equals* $(1 - \mathbb{E}_{y\sim\mu_{G,S},i\sim[d]}[y_0 \cdot y_i])/2$*, where* 0 *denotes the root, and* i *denotes the* i^{th} *neighbor of the root of the* d*-regular tree of depth* L*.*

Proof. This is again a direct computation. We can write

$$\mathbb{E}_{y\sim\mu_{G,S},i\sim[d]}[y_0 \cdot y_i] = \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{nd} = 1 - 2\frac{|\delta_G(S, \bar{S})|}{|E(G)|},$$

and rearranging gives the identity. □

Lemmas C.1.4 to C.1.6 are all linear constraints in the pdf of the distribution $\mu = \mu_{G,S}$, and Lemma C.1.7 is a linear objective function. We may solve this LP in the variables μ for small values of L and d to obtain valid upper bounds on objective function, the cut value $|\delta_G(S, \bar{S})|/|E(G)|$. If we solve this LP for a collection of discrete values $\alpha \in \{0, \delta, 2\delta, \ldots, 1\}$, we obtain an upper bound on the cut value of *any* cut of a graph $G \in \Omega_{d,\lambda}$.

Computational Efficiency. The above LP appears intractable – it has 2^{46} variables even for $d = 3$, $L = 4$. We apply two optimizations to make it more tractable in practice.

1. Use the same variable for isomorphic labelings of $T_{d,L}$. In particular, if there is a graph isomorphism of $T_{d,L}$ that maps a labeling y to another labeling y', then Lemma C.1.4 implies that they should receive the same probability in μ.
2. Only consider "locally optimal" labelings of $T_{d,L}$. To do this, we apply a preprocessing step, where we ensure that the cut S is locally optimal in the sense that for any vertex $i \in S$, the labeling of $T_{d,L}$ generated by restricting $\mathbf{x} \in \{\pm 1\}^n$ to the L-step neighborhood of i corresponds to a maximum cut of $T_{d,L}$, subject to the "boundary conditions" of labels of the leaves of $T_{d,L}$. Therefore, we can also discard any variables corresponding to such suboptimal labelings of $T_{d,L}$.

Recall that $\lambda = 2\sqrt{d-1} + \epsilon$ for arbitrarily small ϵ. We will now pick $\epsilon = 10^{-5}$. Using the above optimizations, we are able to solve the $1/\delta$ LPs for $d = 3, L = 4$ (with 4396 variables), and $d = 4, L = 2$ (with 88 variables) for $\delta = 0.0005$, resulting in the universal upper bounds $\mathrm{MC}(G) \leq 0.953$ for $G \in \Omega_{3,\lambda}$ and $\mathrm{MC}(G) \leq 0.916$ for $G \in \Omega_{4,\lambda}$.

An interesting point is that the $d = 4, L = 3$ case has 11880 variables and is technically within reach of current LP solvers, but resulted in the *same* bound on σ_4^{MC} as the $d = 4, L = 2$ case.

Upper bounds on $\mathrm{IS}(G)$**.** We conclude with some comments on the modifications needed to obtain upper bounds on σ_d^{IS}. The main modification is that we need to restrict ourselves to distributions in $\Lambda_{d,L}$ that are supported only on $\{\pm 1\}$-labelings that correspond to actual Independent Sets of $T_{d,L}$. That is, we only consider labelings y such that for any edge $\{i,j\}$ in $T_{d,L}$, y_i and y_j are not both equal to 1. The objective function disappears, and we now only need to find the largest possible value α such that the LP is still feasible. The $1/\delta$ LPs are tractable for $d = 3, L = 4$ (with 1771 variables) and $d = 4, L = 3$ (with 8855 variables), and they result in the bounds $\sigma_3^{\mathrm{IS}} \leq 0.476$ and $\sigma_4^{\mathrm{IS}} \leq 0.457$.

Appendix D

Appendix for Chapter 7

D.1 Proof of Claim 7.4.4

In this section we present the proof of Claim 7.4.4. Recall that our goal is to approximate $Tf_-(x)$, where $f(x) = \mathbb{E}_A[A(x)]$ is the average output of A. We will do this by computing A on a polynomial number of inputs.

Recall that the function $f_+ = f - f_-$ is in the q-tractable (Definition 7.4.2) subspace V, i.e., it can be written as $f_+ = \sum_{0\le i\le \ell} c_i \cdot f_i$ for real-valued coefficients $c_i = \mathbb{E}_{\mathcal{N}}[f_i \cdot f]$, where $\{f_1, \ldots, f_\ell\}$ is an orthonormal basis of V that can be efficiently evaluated.

We begin by producing a large collection of independent samples $\mathbf{z} = \{z_1, \ldots, z_k\}$ from $\mathcal{N}$. Define the function $g^{\mathbf{z}}(y) = A(y) - \sum_{0\le i\le \ell} f_i(y) \cdot \mathbb{E}_{j\sim[k]}[f_i(z_j) \cdot A(z_j)]$. We will argue that $g^z(y)$ is close to f_-. In the interest of conciseness, we will use very crude bounds.

$$\begin{aligned}
\mathbb{E}_{\mathbf{z}}[|g^{\mathbf{z}}(y) - f_-(y)|] &= \mathbb{E}_{\mathbf{z}}\left[\left|\sum_{0\le i\le \ell} f_i(y) \cdot (\mathbb{E}_{j\sim[k]}[f_i(z_j)\cdot A(z_j)] - \mathbb{E}_{\mathcal{N}}[f_j \cdot f])\right|\right] \\
&\le \sum_{0\le i\le \ell} |f_i(y)| \cdot \mathbb{E}_{\mathbf{z}}\left[\left|\mathbb{E}_{j\sim[k]}[f_i(z_j)\cdot A(z_j)] - \mathbb{E}_{\mathcal{N}}[f_i \cdot f]\right|\right] \\
&\le \sum_{0\le i\le \ell} |f_i(y)| \cdot \mathbb{E}_{\mathbf{z}}\left[\left(\mathbb{E}_{j\sim[k]}[f_i(z_j)\cdot A(z_j)] - \mathbb{E}_{\mathcal{N}}[f_i \cdot f]\right)^2\right]^{1/2} \\
&\qquad\qquad \text{(Jensen's inequality)} \\
&= \sum_{0\le i\le \ell} |f_i(y)| \cdot \mathrm{Var}_{\mathbf{z}}\left[\mathbb{E}_{j\sim[k]}[f_i(z_j)\cdot A(z_j)]\right]^{1/2} \\
&\le k^{-1/2} \sum_{0\le i\le \ell} |f_i(y)| \cdot \mathrm{Var}_{z_1\sim\mathcal{N}}\left[f_i(z_1)\cdot A(z_1)\right]^{1/2} \\
&\le k^{-1/2} \cdot (\ell+1) \cdot \max_i \|f_i\|_\infty^2 \le 2k^{-1/2}q^3. \\
&\qquad\qquad (\ell \le q,\ \|f_i\|_\infty \le q \cdot \|f_i\|_{2,\mathcal{N}} = q \text{ by } q\text{-tractability})
\end{aligned}$$

As long as $k \gg q^6/\epsilon^2$, this is at most $\epsilon/2$. Finally, we will approximate $Tf_-(x)$ by computing the empirical average of $g^{\mathbf{z}}(y)$ across a large number of independent samples

$y \sim M(x)$. Concretely, sample another collection $\mathbf{y} = \{y_1, \ldots, y_k\}$ where $y_i \sim M(x)$ define $C(x) = C^{\mathbf{z},\mathbf{y}}(x) = \mathbb{E}_{j\sim[k]}[g^{\mathbf{z}}(y_j)]$. For any $x \in \Omega$ and $\mathbf{z}$ we will bound

$$\begin{aligned}\mathbb{E}_{\mathbf{y}}[|C(x) - Tg^{\mathbf{z}}(x)|] &\leq \mathbb{E}_{\mathbf{y}}[(C(x) - Tg^{\mathbf{z}}(x))^2]^{1/2} \\ &= \mathrm{Var}_{\mathbf{y}}\left[\mathbb{E}_{j\sim[k]}[g^{\mathbf{z}}(y_j)]\right]^{1/2} \\ &= k^{-1/2}\,\mathrm{Var}_{y_1\sim M(x)}\left[\mathbb{E}_{j\sim[k]}[g^{\mathbf{z}}(y_1)]\right]^{1/2} \leq k^{-1/2}\|g^{\mathbf{z}}\|_\infty.\end{aligned}$$

One can bound $\|g^{\mathbf{z}}\|_\infty \leq 2q^3$ for any z, so again this is at most $\epsilon/2$ for $k \gg q^6/\epsilon^2$. By the triangle inequality, we can conclude

$$\begin{aligned}\mathbb{E}_C[|C(x) - Tf_{<\epsilon}(x)|] &\leq \mathbb{E}_{\mathbf{y},\mathbf{z}}[|C(x) - Tg^{\mathbf{z}}(x)|] + \mathbb{E}_{\mathbf{z}}[|Tf_-(x) - Tg^{\mathbf{z}}(x)|] \\ &\leq \mathbb{E}_{\mathbf{y},\mathbf{z}}[|C(x) - Tg^{\mathbf{z}}(x)|] + \mathbb{E}_{y\sim M(x),\mathbf{z}}[|f_-(y) - g^{\mathbf{z}}(y)|] \\ &\leq 2 \cdot \epsilon/2 = \epsilon.\end{aligned}$$

D.2 Details of Stochastic Gradient Descent Convergence

In this section we present the proof of Claim 7.5.2, which completes the proof of Theorem 7.5.1. Let us begin by recalling the setup. We are trying to solve the following problem:

$$\min_{g=\sum_{i\in[\ell]} g_i f_i} \mathbb{E}_{x\sim\mathcal{P}}[\sigma(1 + g(x))].$$

Recall that the objective function is a convex function of the coefficients $g_1, \ldots, g_\ell$ of g. Throughout this section, we will use the notation $\|v\|_2 = \sqrt{\sum_{i\in[n]} v_i^2}$ exclusively for n-dimensional vectors. The goal is to find a point $g^* = \sum_i g_i^* f_i$ such that the 2-norm of the gradient is bounded as follows.

$$\|\nabla_{\{g_1,\ldots,g_n\}} \mathbb{E}_{x\sim\mathcal{P}}[\sigma(1 + g(x))]\|_2 \leq \delta/3. \tag{D.1}$$

We will use stochastic gradient descent to find an approximate optimizer. The bulk of this proof will be spent in verifying that the conditions for convergence of stochastic gradient descent are met. We will apply the following result [GG23, Corollary 5.6].

Theorem D.2.1. *Let μ be a distribution over differentiable convex functions $F : \mathbb{R}^n \to \mathbb{R}$ such that each function F in the support of μ is R-smooth, i.e. for all $x, y \in \mathbb{R}^n$,*

$$\|\nabla F(x) - \nabla F(y)\|_2 \leq R \cdot \|x - y\|_2.$$

Say $\|\nabla F(x)\|_2 \leq K$ for all f in the support of μ and for all $x \in \mathbb{R}^n$. Suppose $\mathbb{E}_\mu[F]$ is minimized at $\hat{x}$. Assume that one can sample $F \sim \mu$, evaluate $F(x)$ and evaluate $\nabla F(x)$ in time T. Then there is a stochastic gradient descent based algorithm running in time $\mathrm{poly}(n, K, R, 1/\delta', T)$ that finds a point x satisfying the following with high probability:

$$\mathbb{E}_{F\sim\mu}[F(x)] \leq \mathbb{E}_{F\sim\mu}[F(\hat{x})] + \delta'.$$

Let us describe what parameters we will instantiate Theorem D.2.1 with and verify that its conditions are met. We set $n = \ell$, so our variables are the coefficients $g_1, \dots, g_\ell$ of $g = \sum_{i \in [\ell]} g_i f_i$. We will set μ to be the following distribution over functions: sample a point $x \sim \mathcal{P}$, and output the function F that maps $(g_1, \dots, g_\ell)$ to $\sigma(1 + g(x))$.

Note that we can write $\nabla F(g) = \sigma'(1 + g(x)) \cdot \mathbf{f}(x)$, where $\mathbf{f}(x)$ is the ℓ-dimensional vector $[f_1(x), \dots, f_\ell(x)]$. We can indeed sample from μ and evaluate F and ∇F in time $T := \mathrm{poly}(q)$. Using the fact that $\sigma'(\cdot) \in [0, 1]$ and $\|f_i\|_\infty \le q$, we will bound

$$\|\nabla F(g)\|_2^2 = \sigma'(1 + g(x))^2 \cdot \|\mathbf{f}(x)\|^2 \le \sum_{i \in [\ell]} f_i(x)^2 \le \ell q^2 \le q^3.$$

So K can be set to q^3. Similarly, we can bound

$$\begin{aligned}
\|\nabla F(g) - \nabla F(h)\|_2 &= \|(\sigma'(1 + g(x)) - \sigma'(1 + h(x))) \cdot \mathbf{f}(x)\|_2 \\
&\le \ell q^3 \cdot |\sigma'(1 + h(x)) - \sigma'(1 + g(x))| \\
&\le O(\delta^{-1} q^3) \cdot |h(x) - g(x)| \\
&\qquad \text{(Lipschitz constant of } \sigma' \text{ is } \max_t |\sigma''(t)| = O(\delta^{-1})) \\
&\le O(\delta^{-1} q^3) \cdot \sum_{i \in \ell} f_i(x) \cdot |h_i - g_i| \\
&\le O(\delta^{-1} q^3) \cdot \|\mathbf{f}(x)\|_2 \cdot \|h - g\|_2. \qquad \text{(Cauchy-Schwarz)} \\
&\le O(\delta^{-1} q^{4.5}) \cdot \|h - g\|_2.
\end{aligned}$$

As a result, we can set $R = O(\delta^{-1} q^{4.5})$. Applying Theorem D.2.1, we get an algorithm running in time $\mathrm{poly}(q, 1/\delta, 1/\delta')$ that finds a point g^* such that with high probability, $\mathbb{E}_{\mathcal{N}}[\sigma(1 + g^*)] \le \mathbb{E}_{\mathcal{N}}[\sigma(1 + \hat{g})] + \delta'$.

Recall that we wanted an upper bound on the gradient (Eq. (D.1)). We will use another off-the-shelf result to bound this [GG23, Lemma 2.28].

Lemma D.2.2. *If $F : \mathbb{R}^d \to \mathbb{R}$ is convex and R-smooth, then for all $x \in \mathbb{R}^d$, $\frac{1}{2R}\|\nabla F(x)\|_2^2 \le F(x) - F(\hat{x})$.*

We immediately get $\|\nabla F(g^*)\|_2 \le O(\delta^{-1} q^{4.5}) \cdot \delta'$. Choosing δ' to be small enough, this is at most $\delta/4$. To complete the proof, note that rounding the coefficients of g^* to $O(\log(q/\delta))$ bits of precision can change $\|\nabla F(g^*)\|_2$ by at most $\mathrm{poly}(\delta/q)$. Therefore we can evaluate g^* in time $\mathrm{poly}(q, \log 1/\delta)$ while still guaranteeing that $\|\nabla F(g^*)\|_2 \le \delta/2$, completing the proof.

D.3 Calculation of Low-Degree Advantage for Planted Clique

Let $\mathcal{N} = G(n, 1/2)$ and $\mathcal{P} = G(n, 1/2, k)$ for $k = o(\sqrt{n})$. The goal of this section is to calculate $\mathrm{R}^{(\mathcal{P}, \mathcal{N})}[\mathcal{F}_{\le d}]^2$ up to $1 + o(1)$ factors. We begin with the well-known expression

$\mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}_{\leq d}]^2 = \sum_{|S|\leq d}(k/n)^{2\cdot|V(S)|}$ (e.g. [Hop18], Lemma 2.4.1). Let us bound

$$\begin{aligned}\mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}_{\leq d}]^2 &= \sum_{|S|\leq d}(k/n)^{2\cdot|V(S)|} \\ &= \binom{n}{2}\cdot\frac{k^4}{n^4} + \sum_{1<|S|\leq d}(k/n)^{2\cdot|V(S)|} \\ &\leq \frac{k^4}{2n^2} + \sum_{\mathcal{S}} n^{|V(\mathcal{S})|}(k/n)^{2\cdot|V(\mathcal{S})|},\end{aligned}$$

where $\mathcal{S}$ ranges over all $\leq d$-vertex graphs on at least 3 vertices. Since the number of d-vertex graphs is bounded by $2^{\binom{d}{2}} = O(1)$ and $k = o(\sqrt{n})$, we have

$$\frac{k^4}{2n^2} \leq \mathrm{R}^{(\mathcal{P},\mathcal{N})}[\mathcal{F}_{\leq d}]^2 \leq \frac{k^4}{2n^2} + O\left(\frac{k^6}{n^3}\right) = (1+o(1))\cdot\frac{k^4}{2n^2}.$$