

UCLA

UCLA Electronic Theses and Dissertations

Title

Dimensional Asymptotics for Norms of Maximal Averaging Operators on Cartesian Powers of Finite Graphs

Permalink

<https://escholarship.org/uc/item/16k3f6p1>

Author

Greenblatt, Jordan Schor

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

**Dimensional Asymptotics for Norms of Maximal
Averaging Operators on Cartesian Powers of
Finite Graphs**

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

by

Jordan Schor Greenblatt

2016

© Copyright by
Jordan Schor Greenblatt
2016

ABSTRACT OF THE DISSERTATION

Dimensional Asymptotics for Norms of Maximal Averaging Operators on Cartesian Powers of Finite Graphs

by

Jordan Schor Greenblatt

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2016

Professor Terence Chi-Shen Tao, Chair

We examine the L^p norm dimensional asymptotics of spherical and ball maximal averaging operators on Cartesian powers of finite graphs. We extend the techniques [8], [3], and [7], which together prove that the spherical maximal operators on Cartesian powers of finite cliques satisfy L^p bounds independent of dimension for all $p \in (1, \infty]$. We use these tools to prove computable sufficient conditions for dimension-independent L^2 bounds on the ball and spherical maximal operators on powers of some distance regular graphs. For instance, our techniques prove that the ball maximal operators on powers of the 5-cycle graph and the spherical maximal operators on powers of the Petersen graph each satisfy dimension-independent L^2 bounds.

We go on to discuss some less well-behaved asymptotics. We prove that for a substantial family of graphs, including all trees with more than 2 vertices and even some regular graphs, the spherical maximal averaging operators have exponentially growing L^p norms with respect to dimension for all $p \in [1, \infty)$. We prove that for another family, the ball maximal averaging operators have exponentially growing L^p norms for some $p > 1$. Finally, we identify some graphs whose behavior lies between these two extremes and discuss a few related open questions.

The dissertation of Jordan Schor Greenblatt is approved.

Amit Sahai

Mario Bonk

Terence Chi-Shen Tao, Committee Chair

University of California, Los Angeles

2016

For Grandpa

TABLE OF CONTENTS

1	Introduction	1
1.1	Standard Definitions	1
1.2	Background and Past Results	3
1.3	Important Families of Graphs	4
1.4	New Results	6
1.5	Notation and Conventions	9
2	The Dimension-Independent Bound	11
2.1	Maximal Spherical Samples	11
2.1.1	Definition of a Radial Space	12
2.1.2	Spherical Samples and Maximal Samples	13
2.1.3	Distance Regular Graphs and Radial Spaces	15
2.2	Structure of Spheres	23
2.3	Mass Concentration	36
2.4	The Smooth Maximal Operator	63
2.5	Noise Semigroups	77
2.6	Noise Coefficient Formulas	88
2.7	Fourier Analysis	98
2.8	Difference Bound for Neighboring Spheres	105
2.9	Spectral Smoothness and Decay	121
3	Dimensionally Unbounded Maximal Operators	153
3.1	Global Antipode Graphs	153

3.1.1	The Exponential Lower Bound	153
3.1.2	Pathological Examples	157
3.2	Non-Mean Distance Regular Graphs	166
3.3	L^p Asymptotics for Other Graphs	174
4	Future Questions	178
4.1	Small Questions	178
4.2	Medium Questions	179
4.3	Main Question	181
	References	184

ACKNOWLEDGMENTS

First and foremost, I would like to thank Rachael Nisenkier for her love and, equally important, her patience. I would like to thank my mom for teaching me how to turn thoughts into words and put words onto paper. I would like to thank my dad for teaching me how to think with precision, rigor, and a critical eye. I owe both of them extra gratitude for their immeasurable support and for providing towering role models throughout my life.

I would like to thank Terry Tao not only for all the time and useful suggestions he gave me, but also for teaching me how to approach a big problem in little pieces and for reminding me that puzzles should be at least as fun as they are difficult. I would like to thank Phil Gressman for first exposing me to the beauty of Fourier analysis and, through his engaging lectures, providing me with a strong foundation in the subject. I would like to thank Christoph Thiele for offering me so much attention and opportunity from my very first quarter at UCLA. I would like to thank Monica Visan for pushing me when that was what I needed.

I would like to thank Ben Krause for his mentorship and friendship. When I was floundering for a research direction, he gave me encouragement and brought me onto the project that led to the chapters below. I would like to thank Alexandra Kolla and Leonard Schulman for welcoming me onto the project. I would also like to thank Leonard for his tutelage, intellectual exchanges, and financial support in the summer of 2015. I would like to thank Nick Cook for helping me to enter the fields of probability and graph theory by sharing his expertise with me.

Finally, I would like to thank my graduate school comrades, particularly Andreas Aaserud, Adam Azzam, Zach Norwood, Andy Soffer, and Joe Woodworth, for many engaging discussions, mathematical and otherwise. Graduate school experiences vary widely, and the time I shared with them made mine a good one.

VITA

2011	Intensive B.S. (Mathematics), Yale University, <i>Distinction in the Major</i>
2011-2016	Teaching Assistant, University of California, Los Angeles
2013	M.A. (Mathematics), University of California, Los Angeles
2014	Candidate in Philosophy (Mathematics), University of California, Los Angeles
2015	Research Assistant, California Institute of Technology

CHAPTER 1

Introduction

1.1 Standard Definitions

Unless otherwise stated, all graphs will be simple, connected, undirected, unweighted, and finite, all functions will be complex valued, and all L^p norms on graphs will be taken with respect to the counting measure. For a fixed graph G and $x, y \in V(G)$, we use the notation $x \sim_G y$ to signify that the pair $\{x, y\}$ is in $E(G)$. We also denote by d_G , or simply d when clear from context, the “shortest distance” metric on G . On any finite, discrete space, the notation $\mathbb{P}$ refers to the uniform probability measure and $\mathbb{E}$ refers to the expectation over $\mathbb{P}$.

Definition 1.1.1 (Cartesian Graph Product). *For graphs G and H , the Cartesian product of G and H will be denoted by $G \square H$. The vertex set $V(G \square H)$ is the Cartesian set product $V(G) \times V(H)$. Two vertices $(g, h), (g', h') \in V(G \square H)$ are adjacent in $G \square H$ if and only if either*

1. $g = g'$ and $h \sim_H h'$ or

2. $h = h'$ and $g \sim_G g'$.

Definition 1.1.2 (Cartesian Graph Power). *The N^{th} Cartesian power of a graph G will be denoted by G^N . For a positive integer N we inductively define*

$$G^N := \begin{cases} G & \text{if } N = 1 \\ G^{N-1} \square G & \text{if } N > 1. \end{cases}$$

More concretely, $V(G^N)$ is the Cartesian set power $V(G)^N$ and for two vertices $x, y \in V(G^N)$, we say $x \sim_{G^N} y$ if and only if there exists a unique index $1 \leq n \leq N$ such that

1. for all $m \neq n$, $x_m = y_m$ and

2. $x_n \sim_G y_n$.

Remark 1.1.3. By direct computation one observes that for any $(g, h), (g', h') \in V(G \square H)$,

$$d_{G \square H}((g, h), (g', h')) = d_G(g, g') + d_H(h, h').$$

Inductively applying this computation, one observes that for any $x, y \in G^N$,

$$d_{G^N}(x, y) = \sum_{n=1}^N d_G(x_n, y_n).$$

We view a graph G as a natural metric measure space under the (integer valued) metric d_G and the counting measure. We define

$$\mathbb{S}_k(x) := \{y \in V(G) : d_G(x, y) = k\}$$

to be the radius k sphere in G centered at vertex $x \in V(G)$ and

$$\mathbb{B}_k(x) := \{y \in V(G) : d_G(x, y) \leq k\}$$

to be the radius k ball in G centered at vertex x .

Definition 1.1.4 (Spherical Averaging Operator). *The radius k spherical averaging operator is defined for any $f : G \rightarrow \mathbb{C}$ by*

$$S_k^G f(x) := \frac{1}{|\mathbb{S}_k(x)|} \sum_{y \in \mathbb{S}_k(x)} f(y) = \mathbb{E}_{\mathbb{S}_k(x)} f.$$

Definition 1.1.5 (Spherical Maximal Operator). *The spherical maximal operator M^G is defined for any $f : G \rightarrow \mathbb{C}$ by*

$$M^G f(x) := \sup_{\mathbb{S}_k(x) \neq \emptyset} |S_k^G f(x)|.$$

Definition 1.1.6 (Ball Averaging Operator). *The radius k spherical averaging operator is defined for any $f : G \rightarrow \mathbb{C}$ by*

$$B_k^G f(x) := \frac{1}{|\mathbb{B}_k(x)|} \sum_{y \in \mathbb{B}_k(x)} f(y) = \mathbb{E}_{\mathbb{B}_k(x)} f.$$

Definition 1.1.7 (Spherical Maximal Operator). *The spherical maximal operator M^G is defined for any $f : G \rightarrow \mathbb{C}$ by*

$$M_B^G f(x) := \sup_{\mathbb{B}_k(x) \neq \emptyset} |B_k^G f(x)|.$$

Throughout the paper, when it is clear from context, we will suppress the dependence of metrics and operators on G in order to minimize clutter.

1.2 Background and Past Results

M^G satisfies an immediate operator bound of 1 on $L^\infty(G)$. Therefore, for any $1 \leq p < \infty$,

$$\sum_{x \in G} |M^G f(x)|^p \leq \sum_{x \in G} \|f\|_\infty^p \leq \|f\|_\infty^p |G| \leq \|f\|_p^p |G|$$

so M^G also satisfies an L^p bound of $|G|^{1/p}$. Although this simple observation quickly shows that M^G is bounded in L^p , the quantitative bound $|G|^{1/p}$ is not in general sharp. The optimal bound depends heavily on the structure of G . In [8], [3], and [7] the authors explored the dimensional asymptotics of the L^p norms of spherical maximal averaging operators on Cartesian powers of finite cliques $G = K_m$ for all $m \geq 2$.

If $p < \infty$, the a priori L^p bound for the spherical maximal operator on K_m^N is exponential in the dimension N , namely $m^{N/p}$. By viewing M more carefully as the maximum of $N+1$ contractions on $L^1(K_m^N)$ and $L^\infty(K_m^N)$, namely the spherical averaging operators, we can bound M in L^1 by $N+1$, simply dominating Mf by the sum of all the spherical averages for any $f \geq 0$. Linearizing the maximal

operator as in [9] (top of p. 151) and interpolating between the L^1 and L^∞ endpoints provides the improved bound $(N + 1)^{1/p}$.

Testing M on a delta mass in K_m^N shows that the L^1 bound of $N + 1$ is sharp. However, the three papers collectively prove the surprising result that if $1 < p < \infty$, M satisfies an L^p bound depending *only* on m and p . In particular, the bound is *independent of the dimension* N . We state the main results of the papers for the reader's convenience:

Theorem 1.2.1 ([8], Theorem 1). *The spherical maximal operator on the N dimensional hypercube, i.e. K_2^N , satisfies the bound*

$$\left\| M^{K_2^N} f \right\|_2 \lesssim \|f\|_2$$

with a constant independent of $f \in L^2$ and the dimension N .

Theorem 1.2.2 ([3], Theorem 2.5). *For all $p \in (1, \infty]$ the spherical maximal operator on the N dimensional hypercube K_2^N satisfies the bound*

$$\left\| M^{K_2^N} f \right\|_p \lesssim_p \|f\|_p$$

with a constant depending on p but independent of $f \in L^p$ and the dimension N .

Theorem 1.2.3 ([7], Theorem 1.3). *For all $p \in (1, \infty]$ and integers $m > 2$, the spherical maximal operator on the N^{th} Cartesian power of the clique on m vertices, i.e. K_m^N , satisfies the bound*

$$\left\| M^{K_m^N} f \right\|_p \lesssim_{m,p} \|f\|_p$$

with a constant depending on p and m but independent of $f \in L^p$ and the dimension N .

1.3 Important Families of Graphs

In this section, we consolidate the definitions of standard graph families for the convenience of reference. We will use G for a generic graph of diameter d .

Definition 1.3.1 (Regular Graph). G is regular if every vertex has the same degree, i.e.

$$|\mathbb{S}_1(x)| = |\mathbb{S}_1(y)| \text{ for all } x, y \in G.$$

Definition 1.3.2 (Distance Degree Regular Graph (DDRG)). G is distance degree regular if the size of a sphere depends only on its radius, i.e.

$$|\mathbb{S}_k(x)| = |\mathbb{S}_k(y)| \text{ for all } x, y \in G, k \in \{0, \dots, d\}.$$

Note that any vertex transitive, and therefore any Cayley, graph is a DDRG.

Definition 1.3.3 (Distance Regular Graph (DRG)). G is distance regular if there exist finite sequences $\{b_0, \dots, b_d\}$ and $\{c_0, \dots, c_d\}$ such that for all $x \in G, k \in \{0, \dots, d\}, y \in \mathbb{S}_k(x)$,

$$|\mathbb{S}_1(y) \cap \mathbb{S}_{k+1}(x)| = b_k \text{ and } |\mathbb{S}_1(y) \cap \mathbb{S}_{k-1}(x)| = c_k. \quad (1.1)$$

Note that applying (1.1) to $k = 0$ immediately implies that G is regular of degree b_0 . Also, because G is connected, it follows that

$$b_j > 0 \text{ for all } j \in \{0, \dots, d-1\} \text{ and} \quad (1.2)$$

$$c_j > 0 \text{ for all } j \in \{1, \dots, d\}. \quad (1.3)$$

Also it follows from the Definition 1.3.3 that $c_0 = b_d = 0$

Definition 1.3.4 (Distance Regular Graph, alternate). A graph G is distance regular if the size of the intersection of two spheres depends only on their radii and the distance between their centers, i.e., if $x, y, x', y' \in G$ and $d(x, y) = d(x', y')$,

$$|\mathbb{S}_a^G(x) \cap \mathbb{S}_b^G(y)| = |\mathbb{S}_a^G(x') \cap \mathbb{S}_b^G(y')| \text{ for all } a, b \in \{0, \dots, d\}. \quad (1.4)$$

Applying (1.4) to $x = y, x' = y'$, and $a = b$ implies that G is distance degree regular. Note that the equivalence of Definitions 1.3.3 and 1.3.4 is straightforward and classical (see [2], p. 1).

Example 1 (Distance Regular Graphs). *The following are some examples of common distance regular graphs:*

1. *Cycles*
2. *Cliques*
3. *Petersen graph*
4. *Platonic Solids*
5. *Cartesian powers of cliques*

Notice that, although Cartesian powers of cliques are distance regular, in general distance regularity is not preserved under the operation of Cartesian powers. For instance, C_5^N is distance regular if and only if $N = 1$.

Although its significance is a priori obscure, distance regularity is the most important graph property for our present discussion. In §2.1.3 we will provide some practical intuition for this definition which should cast light on its importance in the context of spherical averages.

Definition 1.3.5 (Tree). *G is a tree if*

- (1) *it contains at least 3 vertices and*
- (2) *there is a unique simple path between any pair of vertices.*

We deliberately include the non-standard condition (1) because K_2 , which technically satisfies (2), behaves like a clique rather than a tree in our present discussion.

1.4 New Results

The results of [8], [3], and [7], which together prove dimension-independent L^p bounds for the spherical maximal operators on Cartesian powers of arbitrary finite

cliques, inspired the more general project of identifying analogous dimensional asymptotics for the norms of maximal averaging operators on Cartesian powers of more general graphs. In particular, they raised the central question, “Which graphs’ Cartesian powers have spherical maximal operators that satisfy L^p bounds independent of dimension?”

Chapter 2 constitutes the bulk of the new results. Although we will explicitly state the main result of that chapter below, we emphasize that it is not intended only as an end unto itself. It proves a calculable condition for distance regular graphs to have dimension-independent L^p bounds on their spherical and/or ball maximal operators for all $p \geq 2$. Indeed, using our condition, it is straightforward (if a bit laborious) to verify that, for instance, Cartesian powers of the 5-cycle have dimension-independent bounds on their ball maximal operators and Cartesian powers of the Petersen graph have dimension-independent bounds on their spherical maximal operators.

However, the techniques of Chapter 2 provide a richer understanding of the dimension-independence phenomenon first observed by Harrow, Kolla, and Schulman in [8] for the hypercube. Even after [3] and [7] extended the result, it remained in the context of a base graph with the enormous benefits of maximal symmetry and diameter 1 among other useful properties, i.e. a clique. Although Theorem 1.4.1 does not provide a comprehensive answer to our question of which graphs’ Cartesian powers have spherical maximal operators that satisfy L^p bounds independent of dimension, its proof reveals many of the important properties that were previously cloaked by the idyllic behavior of cliques.

The following theorem is an immediate consequence of Theorem 2.1.10 and Corollary 2.1.11:

Theorem 1.4.1. *Let G be a non-bipartite distance regular graph, $[w_{j,n}^{G_{in}}]_{\substack{0 \leq j \leq d \\ 0 \leq n}}$ be the two parameter array defined by Formula 2.1.6, and $[a_{j,n}^{G_{in}}]_{\substack{1 \leq j \leq d \\ 0 \leq n}}$ be the two*

parameter array defined by Formulas 2.6.1 and 2.6.4 with respect to $[w_{j,n}^{G_{in}}]_{\substack{0 \leq j \leq d \\ 0 \leq n}}$.
For each $j \in \{1, \dots, d\}$, define

$$a_j^{G_{in}} := \begin{cases} 0 & \text{if } a_{j,n} = 0 \text{ for all } n \geq 0 \text{ and} \\ a_{j,m} & \text{where } m \text{ is the minimal index such that } a_{j,m} \neq 0 \text{ otherwise.} \end{cases}$$

If $a_j^{G_{in}} \geq 0$ for all $j \in \{1, \dots, d\}$, then for all $p \geq 2$, we have the uniform bound

$$\left\| M_B^{G^N} f \right\|_p \lesssim_G \|f\|_p$$

with a constant depending on G but not on the dimension N or $f \in L^p(G^N)$.

Moreover, if P_{d-1} has a spectral gap, let $[w_{j,n}^{G_{out}}]_{\substack{0 \leq j \leq d \\ 0 \leq n}}$ be the two parameter array defined by Formula 2.1.6 and $[a_{j,n}^{G_{out}}]_{\substack{1 \leq j \leq d \\ 0 \leq n}}$ be the two parameter array defined by Formulas 2.6.1 and 2.6.4 with respect to $[w_{j,n}^{G_{out}}]_{\substack{0 \leq j \leq d \\ 0 \leq n}}$.

$$a_j^{G_{out}} := \begin{cases} 0 & \text{if } a_{j,n} = 0 \text{ for all } n \geq 0 \text{ and} \\ a_{j,m} & \text{where } m \text{ is the minimal index such that } a_{j,m} \neq 0 \text{ otherwise.} \end{cases}$$

has the property that $a_j^{G_{out}} \geq 0$ for all $j \in \{1, \dots, d\}$, then for all $p \geq 2$, we have the uniform bound

$$\left\| M^{G^N} f \right\|_p \lesssim_G \|f\|_p$$

with a constant depending on G but not on the dimension N or $f \in L^p(G^N)$.

Corollary 1.4.2. *There exist graphs with diameter greater than 1 whose Cartesian powers' ball (e.g. the 5-cycle) and spherical (e.g. Petersen) maximal averaging operators satisfy dimension-independent L^p bounds for all $p \geq 2$.*

Proof. The proof consists simply of applying Theorem 1.4.1 to the relevant arrays for the 5-cycle and the Petersen graphs. While time-consuming, the computation of these arrays is straightforward. $\square$

To provide context for dimensional asymptotics for L^p norms of maximal operators, we include the following negative results:

Theorem 1.4.3. *There exists a non-empty family of graphs $\mathcal{G}$ such that for any $G \in \mathcal{G}$, there exists a constant $C > 1$ such that for any $p \in [1, \infty)$,*

$$\left\| M^{G^N} \right\|_p^p \gtrsim C^N.$$

Moreover, $\mathcal{G}$ contains all trees and some regular graphs.

Theorem 1.4.4. *There exists a non-empty family of graphs $\mathcal{H}$ such that for any $H \in \mathcal{H}$, there exists a constant $C > 1$ and an exponent $p > 1$ such that*

$$\left\| M_B^{H^N} \right\|_p^p \gtrsim C^N.$$

Moreover, $\mathcal{H}$ contains all trees and some regular graphs.

The machinery and explicit constructions required for Theorems 1.4.3 and 1.4.4 are contained in §3.1 and §3.2 respectively. Theorem 1.4.3 follows immediately from Theorem 3.1.6, Theorem 3.1.7, and Proposition 3.1.14 while Theorem 1.4.4 follows immediately from Theorem 3.2.4, Corollary 3.2.7, and Proposition 3.2.8.

1.5 Notation and Conventions

For any $\mathbf{a} \in \mathbb{N}^d$, we use the notation

$$\binom{N}{\mathbf{a}} := \binom{N}{a_1, \dots, a_d}$$

for a multinomial coefficient in d variables. For any vector $\mathbf{b} \in \mathbb{R}^d$ we use the notation

$$\mathbf{b}^{\mathbf{a}} = \prod_{j=1}^d b_j^{a_j}$$

for a multi-index exponential. In general if $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{R}^d$ and f is a scalar valued function, the notation $f(\mathbf{a})$ refers to the vector $[f(a_1), \dots, f(a_d)]$. Similarly, if $\mathbf{a} = (a_1, \dots, a_d)$ and $\mathbf{b} = (b_1, \dots, b_d)$ are vectors of the same dimension and f is a scalar valued function with two arguments, the notation $f(\mathbf{a}, \mathbf{b})$ refers to the vector $[f(a_1, b_1), \dots, f(a_d, b_d)]$.

Our asymptotic notation, i.e. O , o , Ω , ω , and Θ generally refers to behavior as N approaches ∞ . In Chapter 2, it may refer to behavior as k approaches ∞ (which will imply that N also approaches ∞), which should be discernible from context. Subscripts will be used to indicate dependence of implicit constants, however all constants are free to depend on the graph (or radial space, see §2.1.1 for details) G . In particular, ω , Ω , and Θ notation imply that for sufficiently large parameters, the quantities in question share a sign, which is not a requirement for O and o .

Often we will use $\lesssim$, $\gtrsim$, and $\sim$ bounds up to constant factor, which are subject to the same conventions (including sign) as ω , Ω , and Θ . Whenever we say “sufficiently large” without further specification, it means “sufficiently large by comparison to a constant” which is free to depend on G .

The notation $f(t) = O_{an}(t^n)$ is equivalent to the statement “there exists a function $g(t)$ which is real analytic in a neighborhood of $t = 0$ such that $f(t) = t^n g(t)$.” When we refer to functions being analytic on a closed interval, we mean it is analytic on an open neighborhood of the interval. For a vector $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{R}^d$, $\Sigma \mathbf{a} := \sum_{j=1}^d a_j$. If the indices of $\mathbf{a}$ begin at 0 rather than 1, i.e. $\mathbf{a} = (a_0, \dots, a_d)$, then $\Sigma \mathbf{a} := \sum_{j=0}^d a_j$. For a given event S in a probability space $\mathbb{P}_S$ refers to probability conditioned on S and $\mathbb{E}_S$ refers to expectation conditioned on S .

CHAPTER 2

The Dimension-Independent Bound

2.1 Maximal Spherical Samples

We begin by introducing a probabilistic framework for our exploration of maximal operators on powers of distance regular graphs. After we have established our definitions and terminology, we will state our probabilistic maximal theorem. Then we will prove that this probabilistic maximal theorem quickly implies dimension-independent L^p bounds for the ball averaging operators on sufficiently well behaved graphs and dimension-independent L^p bounds for the spherical averaging operators of even better behaved graphs.

Throughout this chapter, we will frequently refer to a base measure space G equipped with a number of parameters. We refer to such a space as a *radial space* and we have collected all the properties of radial spaces in §2.1.1 for ease of reference. It is a glut of information, but there is no need to digest it all at once as its facets appear in the proof gradually. Although radial spaces have total mass 1, we avoid referring to them as “probability spaces” to emphasize that we think of them domains for L^p functions rather than probabilistic sample spaces. The normalization is merely a convenience to simplify the L^2 theory.

We emphasize that the concept of a radial space is essentially a convenient packaging of the relevant properties of distance regular graphs. In §2.1.3 we will flesh this notion out and make it concrete with Definitions 2.1.7 and 2.1.8.

2.1.1 Definition of a Radial Space

Definition 2.1.1 (Radial Space). *Let (G, μ) be a measure space with total mass*

1. *All $L^p(G)$ spaces are constructed with respect to μ and likewise all $L^p(G^N)$ spaces are with respect to the Cartesian power measure μ^N . For some $d \in \mathbb{N}$, G is equipped with a family of $d + 2$ linear operators $\{P_0, \dots, P_d, P\} \in \mathcal{B}(L^p(G))$ for all $1 \leq p \leq \infty$, a two parameter column-stochastic array $\{w_{j,n}\}_{\substack{0 \leq j \leq d, \\ n \geq 0}}$, an orthonormal basis $\widehat{G}$ of $L^2(G)$, a spectral bound $\lambda \in (0, 1)$, and a stochastic vector $\mathbf{p} = (p_0, \dots, p_d) \in (0, 1)^{d+1}$ (note all components are strictly positive).*

For all $A \in \{P_0, \dots, P_d, P\}$, $j \in \{1, \dots, d\}$, $f \in L^p(G)$ (any $1 \leq p \leq \infty$), the radial space G satisfies the following axioms:

1. $\|A\|_1 \leq 1$,
2. $\|A\|_\infty \leq 1$,
3. $A^* = A$,
4. $Af \geq 0$ if $f \geq 0$,
5. $A(\mathbf{1}) = \mathbf{1}$,
6. $P^n P_0 = \sum_{j=0}^d w_{j,n} P_j$,
7. $n < j \implies w_{j,n} = 0$,
8. $n = j \implies w_{j,n} > 0$,
9. $\{P_0, \dots, P_d\}$ is linearly independent,
10. $\widehat{G}$ diagonalizes A ,
11. $\widehat{G}$ contains $\mathbf{1}$,
12. for all $\mathbf{1} \neq \xi \in \widehat{G}$, $P_1(\xi) = \lambda_\xi \xi$ where $|\lambda_\xi| \leq \lambda$.

2.1.2 Spherical Samples and Maximal Samples

Throughout this chapter we will see several sequences $\{X_n\}_{n=1}^N$ of i.i.d. random variables taking finitely many integer values. Because the same characteristics come up repeatedly, we adopt the following shorthand (adapted to the appropriate lettering, parameterizations, etc.):

Notation 1 (Discrete I.I.D. Sequences). *Let $\{X_n\}_{n=1}^N$ be a sequence of i.i.d. real-valued random variables. We will define*

$$\begin{aligned} X &:= \sum_{n=1}^N X_n, \\ X_n^j &:= \mathbb{1}_{\{X_n=j\}}, \text{ and} \\ X^j &:= \sum_{n=1}^N X_n^j. \end{aligned}$$

Fixing $N \in \mathbb{N}$, we define a sequence $\{D_1, \dots, D_N\}$ of i.i.d. $\{0, \dots, d\}$ -valued random variables with the distribution

$$\mathbb{P}(D_n = j) = p_j \text{ for } j \in \{0, \dots, d\}.$$

Definition 2.1.2 (Sphere). *For any integer $k \in \{0, \dots, dN\}$, the radius k sphere, denoted by $\mathbb{S}_k$ is the event $\{D = k\}$.*

Note that, because all entries in $\mathbf{p}$ were positive, all spheres have positive probability.

Definition 2.1.3 (Spherical Sampling Operator). *We define the k spherical sampling operator on $L^1(G^N)$ pointwise by*

$$S_k^N f(x) = \mathbb{E}_{\mathbb{S}_k} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right).$$

Observe that for all $1 \leq p \leq \infty, j \in \{0, \dots, d\}$, Reisz-Thorin interpolation implies the operator bound $\|P_j\|_p \leq 1$. Moreover, for any $1 \leq p \leq \infty, \{j_n\}_{n=1}^N \in$

$\{0, \dots, d\}^N$, because L^p norms multiple through tensor products, $\|\otimes_{n=1}^N P_{j_n}\|_p \leq 1$.

Finally, by the Minkowski integral inequality, for all $1 \leq p \leq \infty$,

$$\begin{aligned} \|S_k^N f(x)\|_p &= \left\| \mathbb{E}_{\mathbb{S}_k} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right) \right\|_p \\ &\leq \mathbb{E}_{\mathbb{S}_k} \left(\left\| \left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right\|_p \right) \\ &\leq \mathbb{E}_{\mathbb{S}_k} \left(\|f(x)\|_p \right) \\ &= \|f\|_p \end{aligned} \tag{2.1}$$

Directly from the properties of $\{P_0, \dots, P_d\}$ and interpolation, we can now see that $\{S_k^N\}_{k=0}^{dN}$ is a family of commuting, positivity preserving L^p -contractions for all $1 \leq p \leq \infty$.

Now we can define the maximal operator motivating our present discussion.

Definition 2.1.4 (Spherical Maximal Operator). *Fix $N \in \mathbb{N}$. For any $f \in L^1(G)$, we denote and define the spherical maximal operator by*

$$Mf(x) = \max_{0 \leq k \leq \mathbb{E}(D)} |S_k^N f(x)|.$$

The cutoff at radius $\mathbb{E}(D)$ rather than dN will be explained in §2.1.3. The remaining sections of this chapter are directed at proving the following theorem:

Theorem 2.1.5. *Let $[a_{j,n}]_{\substack{1 \leq j \leq d \\ 0 \leq n}}$ be the two parameter array defined by Formulas 2.6.1 and 2.6.4 and, for each $j \in \{1, \dots, d\}$, define*

$$a_j := \begin{cases} 0 & \text{if } a_{j,n} = 0 \text{ for all } n \geq 0 \text{ and} \\ a_{j,m} & \text{where } m \text{ is the minimal index such that } a_{j,m} \neq 0 \text{ otherwise.} \end{cases}$$

If $a_j \geq 0$ for all $j \in \{1, \dots, d\}$, then for all $p \geq 2$, we have the uniform bound

$$\|Mf\|_p \lesssim_G \|f\|_p$$

with a constant depending on the parameters of G as laid out in §2.1.1 but not on the dimension N .

2.1.3 Distance Regular Graphs and Radial Spaces

Theorem 2.1.5 and our discussion of radial spaces in general is vacuous without context. Every non-bipartite distance regular graph is naturally endowed with a radial space, which we call its inner radial space. Moreover, if P_{d-1} has a spectral gap, it is naturally endowed with a second radial space, which we call its outer radial space. In this section, for a given DRG G with diameter d , we will construct the inner and outer radial spaces explicitly, show that they satisfy the axioms of Definition 2.1.1, and use Theorem 2.1.5 (if applicable) to attain dimension-independent L^2 bounds for ball and/or spherical maximal operators on powers of G .

Before any of this rigorous discussion, we lay out some heuristics for the significance of DRGs. Simply put, a DRG is a graph for which random walks are radial. That is, if a mass is placed at an arbitrary vertex x and takes a fixed number of random steps, the probability of finding it at any vertex y depends only on the distance between x and y . Notice that this property of random walks is familiar from Euclidean space with Brownian motion as a substitute for random walks.

However, distance regularity is far from ubiquitous in graphs. Even graphs with relatively large symmetry groups and good spectral qualities are not guaranteed to satisfy it. For instance the Cayley graph of $\mathbb{Z}_8$ with generator set $\{1, 2, 6, 7\}$ is vertex transitive and its spherical averaging operators are simultaneously diagonalizable with a spectral gap, the importance of which will arise implicitly in §2.7. Nonetheless, a random walk does not weight all vertices of a given distance from a starting point equally.

From a practical standpoint, the radial space framework of §2.1.1 allows us to divorce the study of spheres of small radius and spheres of large radius in a sense to be made more concrete shortly. This is useful because, although we use the same tools for both, our tools cannot handle them simultaneously. From a pedagogical

standpoint, the radial space framework distills the attributes of distance regular graphs that make their radial maximal operators particularly amenable to bounds.

We will define the inner space in full detail in Definition 2.1.7 below, but first we provide some motivation. For any $j \in \{0, \dots, d\}$, we denote by P_j the averaging operator over the sphere of radius j . We also define P to be P_1 , the stochastically normalized adjacency operator. For a DRG, the radially of a random step (the transition matrix of which is P) implies that P sends $\text{span}(\{P_0, \dots, P_d\})$ to itself. Moreover, simply by the triangle inequality, if a mass sits at a vertex y and takes a random step, for any vertex x it can only land at least $d(x, y) - 1$ and at most $d(x, y) + 1$ from x .

In the context of a radial space, we think of the set $\{P_0, \dots, P_d\}$ as a sort of ladder of distances and P as a random up, down, or neutral step. The adaptation of a DRG to its inner space suggests that we think of its spherical averages as the primary object of study rather than the actual graph. That is, instead thinking of the boolean-valued binary relation of adjacency as the fundamental property defining the graph, we view the $\{0, \dots, d\}$ -valued distance function (which does not single out distance 1 as all-important) to be the fundamental property. These pairs are conveniently encapsulated by the “distance ladder” $\{P_0, \dots, P_d\}$.

Although the spherical averaging operators can be defined for any metric measure space that gives finite positive measure to spheres, in particular any graph, the operators are not necessarily well-adapted to the geometry of the space. However, in a distance regular graph, a random step on the underlying space corresponds perfectly to a random “distance ladder” step. That is, if the DRG is regular of degree D (recall that regularity follows directly from the definition of a DRG), looking back at the standard Definition 1.3.3, it is straightforward to verify that the probability of a step “up” the ladder from P_j to P_{j+1} (for $j < d$) is $\frac{b_j}{D}$, the probability of a step “down” to P_{j-1} (for $j > 0$) is $\frac{c_j}{D}$, and the probability of remaining stationary at P_j (i.e. the “neutral step”) is $\frac{D-b_j-c_j}{D}$. We will adopt the

standard notation

$$a_j := D - b_j - c_j$$

for convenience.

It is straightforward to verify that for any family of averaging operators over pairwise disjoint, positive measure sets (in this case, non-empty spheres of distinct radii) are linearly independent for any metric measure space. We can therefore use $\{P_0, \dots, P_d\}$ as an (ordered) basis of its span. Moreover, if we view P is an operator on $\text{span}(\{P_0, \dots, P_d\})$ rather than an operator acting on the functions on G , we have,

$$PP_j = \begin{cases} \frac{a_0 P_j + b_0 P_{j+1}}{D} = P_1 & \text{for } j = 0 \\ \frac{c_j P_{j-1} + a_j P_j + b_j P_{j+1}}{D} & \text{for } j \in \{1, \dots, d-1\} \\ \frac{a_d P_{j-1} + c_d P_j}{D} & \text{for } j = d. \end{cases} \quad (2.2)$$

Representing P in the basis $\{P_0, \dots, P_d\}$ provides a practical formula for converting a DRG into its inner radial space.

Formula 2.1.6. Define the square matrix $A := [a_{j,n}]_{0 \leq j,n \leq d}$ with

$$a_{j,n} = \begin{cases} b_n & \text{for } j = n + 1 \\ a_j = a_n & \text{for } j = n \\ c_j & \text{for } j = n - 1 \\ 0 & \text{otherwise.} \end{cases} \quad (2.3)$$

or, presented more visually,

$$A = \frac{1}{D} \begin{bmatrix} a_0 & c_1 & 0 & \dots & 0 & 0 \\ b_0 & a_1 & c_2 & \dots & 0 & 0 \\ 0 & b_1 & a_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & c_{d-1} & 0 \\ 0 & 0 & 0 & \dots & a_{d-1} & c_{d-1} \\ 0 & 0 & 0 & \dots & b_{d-1} & a_d \end{bmatrix} = \frac{1}{D} \begin{bmatrix} 0 & c_1 & 0 & \dots & 0 & 0 \\ D & a_1 & c_2 & \dots & 0 & 0 \\ 0 & b_1 & a_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & c_{d-1} & 0 \\ 0 & 0 & 0 & \dots & a_{d-1} & c_{d-1} \\ 0 & 0 & 0 & \dots & b_{d-1} & a_d \end{bmatrix}. \quad (2.4)$$

Define the array $[w_{j,n}^{G_{in}}]_{\substack{0 \leq j \leq d \\ n \geq 0}}$ by

$$w_{j,n}^{G_{in}} = A^n|_{j,0} \quad (2.5)$$

so the n^{th} column of $[w_{j,n}^{G_{in}}]_{\substack{0 \leq j \leq d \\ n \geq 0}}$ matches the 0^{th} column of A^n .

Then, simply chasing definitions, we have the identity

$$P^n P_0 = \sum_{j=0}^d w_{j,n}^{G_{in}} P_j. \quad (2.6)$$

for all $n \geq 0$. Note also that for $j > n$, $w_{j,n}^{G_{in}} = 0$ which can be seen directly from the expression (2.4). In concrete terms, this corresponds to the fact that a mass that has taken only n steps can move no further than distance n . Also, recall from (1.2) that $b_j > 0$ for all $j \in \{0, \dots, d-1\}$. Examining (2.4) reveals that for any $j \in \{0, \dots, d\}$,

$$w_{j,j}^{G_{in}} = \frac{\prod_{i=1}^j b_i}{D^j} > 0.$$

This corresponds to the fact that if a mass takes j steps, as long as j is less than the diameter, there is a positive probability that each step moves away from the starting point, landing the mass a distance j away.

To avoid confusion, in the definitions of the inner and outer radial spaces, we will use the superscripts G_{in} , G_{out} , and G when there is potential ambiguity.

Also, at this stage we restrict G to be non-bipartite and let $\widehat{G}$ be an arbitrary orthonormal eigenbasis of $L^2(G)$ for the operator P_1 containing $\mathbf{1}$. This exists because P_1 is real symmetric (because G is regular) and $P_1(\mathbf{1}) = \mathbf{1}$ because P_1 is stochastic. Notice also that degree distance regularity follows directly from the definition of distance regularity so for any $x, y \in G$, $|\mathbb{S}_j^G(x)| = |\mathbb{S}_j^G(y)|$

Definition 2.1.7 (Inner Radial Space). *The inner radial space G , denoted G_{in} , is equipped with*

- $\mu = \mathbb{P}$,
- $d^{G_{in}} = d^G$,
- $P_j^{G_{in}} = P_j^G$ for $j \in \{0, \dots, d\}$,
- $P^{G_{in}} = P_1^G$,
- $[w_{j,n}^{G_{in}}]_{\substack{0 \leq j \leq d \\ n \geq 0}}$ defined by (2.5),
- $\widehat{G_{in}} = \widehat{G}$,
- $\lambda = \max_{\mathbf{1} \neq \xi \in \widehat{G}} \|P_1(\xi)\|_2$, and
- $p_j = \frac{|\mathbb{S}_j^G(x)|}{|G|}$ for arbitrary $x \in G$ for all $j \in \{0, \dots, d\}$.

The only axioms of Definition 2.1.1 that are not standard facts and that we have not already shown are axiom (12) which holds because G is non-bipartite by assumption and (10) which holds because $\widehat{G}$ diagonalizes $P = P_1$ by construction and $P_0 = I$ trivially and these two operators generate $\{P_0, \dots, P_d\}$ as an algebra by (2.6).

The outer radial space is similar, but instead of mimicking the behavior of spherical averages from radius 0 up, it mimics their behavior from radius d down:

Definition 2.1.8 (Outer Radial Space). *The outer radial space G , denoted G_{out} , is equipped with*

- $\mu = \mathbb{P}$,
- $d^{G_{out}} = d^G$,
- $P_j^{G_{out}} = P_{d-j}^G$ for $j \in \{0, \dots, d\}$,
- $P^{G_{out}} = P_1^G$,
- $[w_{j,n}^{G_{out}}]_{\substack{0 \leq j \leq d \\ n \geq 0}}$ defined by (2.9) below,
- $\widehat{G_{out}} = \widehat{G}$,
- $\lambda = \max_{\mathbf{1} \neq \xi \in \widehat{G}} \|P_1(\xi)\|_2$, and
- $p_j = \frac{|\mathbb{S}_{d-j}^G(x)|}{|G|}$ for arbitrary $x \in G$ for all $j \in \{0, \dots, d\}$.

To derive $w_{j,n}^{G_{out}}$ we follow the logic of we used to find $w_{j,n}^{G_{in}}$. Using (2.2) (adapting the numbering appropriately) we derive an analog of Formula 2.1.6.

Formula 2.1.9. Define the square matrix $B := [b_{j,n}]_{0 \leq j, n \leq d}$ by

$$b_{j,n} = \begin{cases} c_{d-n} & \text{for } j = n + 1 \\ a_{d-j} = a_n & \text{for } j = n \\ b_{d-j} & \text{for } j = n - 1 \\ 0 & \text{otherwise.} \end{cases} \quad (2.7)$$

or, presented more visually,

$$B = \frac{1}{D} \begin{bmatrix} a_d & b_{d-1} & 0 & \dots & 0 & 0 \\ c_d & a_{d-1} & b_{d-2} & \dots & 0 & 0 \\ 0 & c_{d-1} & a_{d-2} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & b_1 & 0 \\ 0 & 0 & 0 & \dots & a_1 & b_0 \\ 0 & 0 & 0 & \dots & c_1 & a_0 \end{bmatrix} = \frac{1}{D} \begin{bmatrix} a_d & b_{d-1} & 0 & \dots & 0 & 0 \\ c_d & a_{d-1} & b_{d-2} & \dots & 0 & 0 \\ 0 & c_{d-1} & a_{d-2} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & b_1 & 0 \\ 0 & 0 & 0 & \dots & a_1 & D \\ 0 & 0 & 0 & \dots & c_1 & 0 \end{bmatrix}. \quad (2.8)$$

Define the array $[w_{j,n}^{G_{out}}]_{\substack{0 \leq j \leq d \\ n \geq 0}}$ by

$$w_{j,n} = B^n|_{j,0} \quad (2.9)$$

so the n^{th} column of $[w_{j,n}^{G_{out}}]_{\substack{0 \leq j \leq d \\ n \geq 0}}$ matches the 0^{th} column of B^n .

As before, we have the identity

$$P^n P_0 = \sum_{j=0}^d w_{j,n}^{G_{out}} P_j.$$

and the axioms of Definition 2.1.1 follow.

The purpose of defining these spaces is to frame the following theorems:

Theorem 2.1.10. *Let G be a non-bipartite distance regular graph and $[a_{j,n}]_{\substack{1 \leq j \leq d \\ 0 \leq n}}$ be the two parameter array defined by Formulas 2.6.1 and 2.6.4 and let $R := N \sum_{j=0}^d j p_j^{G_{in}}$. For each $j \in \{1, \dots, d\}$, define*

$$a_j := \begin{cases} 0 & \text{if } a_{j,n} = 0 \text{ for all } n \geq 0 \text{ and} \\ a_{j,m} & \text{where } m \text{ is the minimal index such that } a_{j,m} \neq 0 \text{ otherwise.} \end{cases}$$

If $a_j \geq 0$ for all $j \in \{1, \dots, d\}$, then for all $p \geq 2$, we have the uniform bound

$$\left\| \max_{k \leq R} S_k^{G^N} f \right\|_p \lesssim_G \|f\|_p$$

with a constant depending on G but not on the dimension N or $f \in L^p(G^N)$.

Moreover, if P_{d-1} has a spectral gap, and

$$b_j := \begin{cases} 0 & \text{if } a_{j,n} = 0 \text{ for all } n \geq 0 \text{ and} \\ a_{j,m} & \text{where } m \text{ is the minimal index such that } a_{j,m} \neq 0 \text{ otherwise.} \end{cases}$$

has the property that $b_j \geq 0$ for all $j \in \{1, \dots, d\}$, then for all $p \geq 2$, we have the uniform bound

$$\left\| \max_{k \geq R} S_k^{G^N} f \right\|_p \lesssim_G \|f\|_p$$

with a constant depending on G but not on the dimension N or $f \in L^p(G^N)$.

This theorem constitutes little more than definition chasing. However, it has an important consequence:

Corollary 2.1.11. *If the inner radial space of G satisfies the hypotheses of (2.1.10), then for all $p \geq 2$, we have the uniform bound*

$$\left\| M_B^{G^N} f \right\|_p \lesssim_G \|f\|_p$$

with a constant depending on G but not on the dimension N or $f \in L^p(G^N)$.

If the inner and the outer radial spaces of G both satisfy the hypotheses of (2.1.10), then for all $p \geq 2$, we have the uniform bound

$$\left\| M^{G^N} f \right\|_p \lesssim_G \|f\|_p$$

with a constant depending on G but not on the dimension N or $f \in L^p(G^N)$.

Proof. The second claim is a trivial consequence of Theorem 2.1.10. The first claim follows from two observations. The first is that, because expectation over a ball is nothing but the expectation over a (weighted) random choice of spheres, we immediately have the inequality

$$\max_{0 \leq k \leq R} |\mathbb{E}_{\mathbb{B}_k(x)} f| \leq \max_{0 \leq k \leq R} |S_k f(x)|. \quad (2.10)$$

The second observation is that for any fixed $x \in G^N$,

$$\begin{aligned} R &= \mathbb{E}_{Y \in G^N} d_{G^N}(x, Y) \\ &= \mathbb{E}_{Y \in G^N} \sum_{n=1}^N d_{\Gamma}(x_n, Y_n) \\ &= \sum_{n=1}^N \mathbb{E}_{Y_n \in \Gamma} d_{\Gamma}(x_n, Y_n). \end{aligned}$$

In particular, R is the expectation of a sum of N i.i.d. random variables so the central limit theorem guarantees that

$$\lim_{N \rightarrow \infty} \mathbb{P}_{G^N}(\mathbb{B}_R(x)) = \lim_{N \rightarrow \infty} \mathbb{P}_{G^N}(\{d_{G^N}(x, Y) \leq R\}) = \frac{1}{2}$$

so we have uniform bound over all $N, k > R$

$$\mathbb{P}_{\mathbb{B}_k(x)} \gtrsim 1.$$

Therefore

$$\begin{aligned} \max_{k \geq R} |\mathbb{E}_{\mathbb{B}_k(x)} f| &= \frac{1}{|\mathbb{B}_k(x)|} \sum_{y \in \mathbb{B}_k(x)} |f(y)| \\ &\lesssim \frac{1}{|G^N|} \sum_{y \in \mathbb{B}_k(x)} |f(y)| \\ &\leq \frac{1}{|G^N|} \sum_{y \in G^N} |f(y)| \\ &= \|f\|_1 \end{aligned}$$

so the operator $\max_{k \geq R} |\mathbb{E}_{\mathbb{B}_k(x)} f|$ satisfies a universal L^1 bound. Because it is an averaging operator, it trivially satisfies a universal L^∞ bound so, by Reisz-Thorin interpolation, it satisfies universal L^p bounds for all $1 \leq p \leq \infty$. In conjunction with (2.10), this proves the claim. $\square$

2.2 Structure of Spheres

We now use our notation D^j from Notation 1 to help break down and analyze spheres in G^N .

Definition 2.2.1 (Type $\mathbf{a}$). *For any $\mathbf{a} = (a_0, \dots, a_d) \in \mathbb{N}^{d+1}$ such that $\Sigma \mathbf{a} = N$, Type $\mathbf{a}$ is the event*

$$\mathbb{T}_{\mathbf{a}} := \{ (D^0, \dots, D^d) = \mathbf{a} \}.$$

It is straightforward to verify combinatorially that for any type $\mathbb{T}_{\mathbf{a}}$,

$$\mathbb{P}(\mathbb{T}_{\mathbf{a}}) = \binom{N}{\mathbf{a}} \mathbf{p}^{\mathbf{a}}. \quad (2.11)$$

Notation 2 (Vector $\mathbf{d}$). *We reserve the notation*

$$\mathbf{d} := (0, 1, \dots, d)$$

due to the frequency with which this vector will appear.

The array of types $\{\mathbb{T}_{\mathbf{a}} : \mathbf{a} \in \mathbb{N}^{d+1}, \Sigma \mathbf{a} = N\}$ partitions the spheres into the (disjoint) union

$$\mathbb{S}_k = \bigcup_{\substack{\Sigma \mathbf{a} = N \\ \mathbf{a} \cdot \mathbf{d} = k}} \mathbb{T}_{\mathbf{a}}.$$

and the spheres in turn partition the sample space into the (disjoint) union

$$\bigcup_{k=0}^{dN} \mathbb{S}_k.$$

Notice that, while the spheres' sizes in K_m^N (in our present language, this corresponds to $d = 1$) are binomially distributed (see [7] for more detail), the mass of a sphere is more generally given by a sum over families of parallel affine codimension 2 spaces in an ambient $d + 1$ parameter multinomial distribution. As a result, spheres in spaces with $d > 1$ are far less amenable to simple recurrence relations and are more difficult to describe probabilistically than those in clique powers.

We use the standard gamma function to build an analytic function that describes type masses.

Definition 2.2.2 (Analytic Type Mass Function and Domain). *We denote the analytic type mass function by*

$$T(\mathbf{a}) := \mathbf{p}^{\mathbf{a}} \frac{\Gamma(N + 1)}{\prod_{j=0}^d \Gamma(a_j + 1)}$$

defined on the domain

$$\mathcal{D} := \{\mathbf{a} = (a_0, \dots, a_d) \in (-1, N + d)^{d+1} : \Sigma \mathbf{a} = N\}.$$

Notice that if $\mathbf{a} \in \mathbb{N}^d$ then $T(\mathbf{a})$ matches (2.11). Therefore we have $T(\mathbf{a}) = \mathbb{P}(\mathbb{T}_{\mathbf{a}})$, hence the label “type mass function.” To facilitate quantitative estimates on the distribution of type masses $\mathbb{P}(\mathbb{T}_{\mathbf{a}})$ with respect to the parameter $\mathbf{a}$, we will take advantage of the logarithmic concavity of the type sizes by using that of the analytic type size function (to be proved shortly). Accordingly, we denote the log gamma function as follows:

Notation 3 (Log Gamma Function). *The Log Gamma function, defined on $(0, \infty)$, is denoted by*

$$\Lambda(x) := (\ln \circ \Gamma)(x)$$

We identify the common properties of the derivatives of Λ that will be relevant to our discussion, which are either directly from [1] or slight manipulations of those formulas. Although some are easily equivalent, we list everything we need to facilitate later references.

Claim 2.2.3. *Λ satisfies the following conditions on its specified domain $(0, \infty)$:*

1. $\Lambda''(x) = \sum_{n=0}^{\infty} \frac{1}{(x+n)^2}$ (formula 6.4.10),
2. $\Lambda''(x)$ is decreasing (direct consequence of formula 6.4.10),
3. $\Lambda''(x) > 0$ (direct consequence of formula 6.4.10),
4. the only positive root of $\Lambda'(x)$ rounded to 3 decimal places is $x \approx 1.462$ (table 6.3.19),
5. $\Lambda(x)$ is bounded from below (it is convex by (3) and has a critical point by (4)), and
6. $\lim_{x \rightarrow 0^+} \Lambda(x) = +\infty$ (Γ has a simple pole with positive residue at $z = 0$ by formula 6.1.3 so it, and therefore its logarithm, go to $+\infty$ as x approaches 0).

Claim 2.2.4. *Λ satisfies the following asymptotic formulas on intervals bounded away from 0, i.e. on any interval (δ, ∞) with constants dependent on δ :*

1. $\Lambda(x) = (x - \frac{1}{2}) \ln(x) + O_{\delta}(1)$ (direct consequence of formula 6.4.41),
2. $\Lambda'(x) = \ln(x) + O_{\delta}(\frac{1}{x})$ (direct consequence of formula 6.3.18),
3. $\Lambda'(x) = \ln(x) - \frac{1}{2x} + O_{\delta}(\frac{1}{x^2})$ (direct consequence of formula 6.3.18), and

4. $\Lambda''(x) = \frac{1}{x} + O_\delta\left(\frac{1}{x^2}\right)$ (direct consequence of formula 6.4.12).

Defining Λ facilitates the definition of the (analytic) log type mass function on $\mathcal{D}$ as follows:

Definition 2.2.5 (Log Type Mass Function).

$$\begin{aligned} L(\mathbf{a}) &:= \ln(T(\mathbf{a})) \\ &= \mathbf{a} \cdot \ln(\mathbf{p}) + \Lambda(N+1) - \sum_{j=0}^d \Lambda(a_j + 1). \end{aligned} \quad (2.12)$$

Directly from (2.12), we compute the first and second derivatives of $L(\mathbf{a})$.

Lemma 2.2.6. *For any $i, j \in \{0, \dots, d\}$, the first and second derivatives of $L(\mathbf{a})$ are given (respectively) by*

$$L_i(\mathbf{a}) = \ln(p_i) - \Lambda'(a_i + 1)$$

$$L_{ij}(\mathbf{a}) = -\Lambda''(a_i + 1) \delta_{ij}.$$

From the simple Lemma 2.2.6 we derive a few crucial lemmas concerning the distribution of types in this section and §2.3.

Lemma 2.2.7. *$L(\mathbf{a})$ is concave on $\mathcal{D}$, i.e. the Hessian $\mathcal{H}L(\mathbf{a})$ is negative definite for all $\mathbf{a} \in \mathcal{D}$.*

Proof. We convert the second derivative computation in Lemma 2.2.6 into a matrix format to find

$$\begin{bmatrix} -\Lambda''(a_0 + 1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & -\Lambda''(a_d + 1) \end{bmatrix} \quad (2.13)$$

Recall that $\Lambda'' > 0$ (Claim 2.2.3, item (3)) and thus (2.13) is negative definite for all $\mathbf{a} \in \mathcal{D}$. □

The main result of this section will help us locate where the types in a given sphere peak in mass:

Proposition 2.2.8. *For any real number $k \in (0, \mathbb{E}(D)]$, let*

$$\mathcal{D}^k := \{\mathbf{a} \in \mathcal{D} : \mathbf{a} \cdot \mathbf{d} = k\}.$$

Then $L(\mathbf{a})$ has a unique maximum in $\mathcal{D}^k$, denoted here by $\mathbf{b}$. Moreover,

$$b_0 \sim N$$

and, for any $j \in \{1, \dots, d\}$,

$$b_j = \begin{cases} O(1) & \text{if } k = O\left(N^{\frac{j-1}{j}}\right) \\ \Theta\left(\frac{k^j}{N^{j-1}}\right) & \text{otherwise.} \end{cases}$$

Proof. Because $L(\mathbf{a})$ is strictly concave on $\mathcal{D}$ by Lemma 2.2.7, it is also strictly concave when restricted to $\mathcal{D}^k$. Therefore, $L(\mathbf{a})$ has at most one critical point $\mathbf{b}$ in the domain $\mathcal{D}^k$ and, if existent, $L(\mathbf{b})$ must be a maximum. Our goal is to show that L has a maximum in the interior of $\mathcal{D}^k$ where $\mathcal{D}^k$ is viewed as a subset of codimension 2 affine space $\{\mathbf{a} \in \mathbb{R}^{d+1} : \Sigma \mathbf{a} = N, \mathbf{a} \cdot \mathbf{d} = k\}$. To this end, we note that for any $\mathbf{a} \in \mathcal{D}^k$ lying within a distance δ from the boundary of $\mathcal{D}^k$ (in the context of $\{\mathbf{a} \in \mathbb{R}^{d+1} : \Sigma \mathbf{a} = N, \mathbf{a} \cdot \mathbf{d} = k\}$) has at least one component above $N + d - \delta$ or below $\delta - 1$. Moreover, if there is a component over $N + d - \delta$, the sum of the remaining components is $\delta - d$ so at least one must be below $\delta - 1$. Thus there is guaranteed to be at least one component before $\delta - 1$.

Because $\lim_{x \rightarrow 0^+} \Lambda(x) = +\infty$ and $\Lambda(x)$ is bounded from below, (Claim (2.2.3), items (6) and (5)), as x approaches the boundary, $L(\mathbf{a})$ goes to $-\infty$ uniformly (with respect to the distance between x and the boundary). Therefore $L(\mathbf{a})$ must indeed have a maximum in the interior of $\mathcal{D}^k$.

With this in mind, we may use the method of Lagrange multipliers to find the maximum of L by searching for a vector $\mathbf{b} \in \mathcal{D}^k$ such that $\nabla L(\mathbf{b})$ is normal to

the surface $\mathcal{D}^k$. Equivalently we seek $\mathbf{b} \in \mathcal{D}^k$ such that $\nabla L(\mathbf{b})$ lies in the span of $\mathbf{d}$ and the constant vector $\mathbf{1} := (1, \dots, 1)$.

If $\nabla L(\mathbf{b}) = \mu_1 \mathbf{1} + \mu_d \mathbf{d}$, examining the 0^{th} component shows

$$\mu_1 = L_0(\mathbf{b}) = \ln(p_0) - \Lambda'(b_0 + 1).$$

Thus we seek $\mathbf{b}$ such that

$$\nabla L(\mathbf{b}) - [\ln(p_0) - \Lambda'(b_0 + 1)] \mathbf{1} = \mu_d \mathbf{d},$$

i.e. for all $j \in \{1, \dots, d\}$,

$$L_j(\mathbf{b}) - [\ln(p_0) - \Lambda'(b_0 + 1)] = jL_1(\mathbf{b}) - j[\ln(p_0) - \Lambda'(b_0 + 1)]. \quad (2.14)$$

We can also use the equivalent statement that for all $j \in \{1, \dots, d\}$,

$$dL_j(\mathbf{b}) - d[\ln(p_0) - \Lambda'(b_0 + 1)] = jL_j(\mathbf{b}) - j[\ln(p_0) - \Lambda'(b_0 + 1)]. \quad (2.15)$$

We proceed by using (2.15) to prove the condition

$$b_0 \sim N \quad (2.16)$$

in the lemma statement. Applying Lemma 2.2.6, for all $j \in \{1, \dots, d\}$ we derive from (2.15),

$$\begin{aligned} d \ln(p_j) - d\Lambda'(b_j + 1) + d\Lambda'(b_0 + 1) - d \ln(p_0) = \\ j \ln(p_d) - j\Lambda'(b_d + 1) + j\Lambda'(b_0 + 1) - j \ln(p_0) \end{aligned}$$

and compute

$$\begin{aligned} \Lambda'(b_j + 1) &= \frac{1}{d} \ln \left(\frac{p_j^d}{p_d^j p_0^{d-j}} \right) + \frac{d-j}{d} \Lambda'(b_0 + 1) + \frac{j}{d} \Lambda'(b_d + 1) \\ &= O(1) + \frac{d-j}{d} \Lambda'(b_0 + 1) + \frac{j}{d} \Lambda'(b_d + 1). \end{aligned} \quad (2.17)$$

We can quickly adapt Claim 2.2.3 to suit our needs with the bound

$$\Lambda'(x + 1) \leq \begin{cases} 0 & \text{for } 0 < x \leq .4 \text{ and} \\ \ln(x) + O\left(\frac{1}{x}\right) & \text{for } x > .4 \end{cases} \quad (2.18)$$

as follows:

Rounding to 3 decimal places, the only positive root of Λ' is approximately 1.462 (Claim 2.2.3, item (4)) and Λ' is increasing on $\mathbb{R}_+$ (Claim 2.2.3, item (3)). Thus, for $x + 1$ below the first root (we can restrict this regime to $x \leq .4$ to be concrete) $\Lambda'(x) < 0$. For $x > .4$, we use Claim 2.2.4, (2), which directly implies that for $x + 1 \in \mathbb{R}_+$ bounded away from 0, $\Lambda'(x + 1) = \ln(x + 1) + O\left(\frac{1}{x+1}\right)$. Applying the Taylor remainder theorem to approximate $\ln(x + 1) = \ln(x) + O\left(\frac{1}{x}\right)$ for positive x away from 0 (in this case $x > .4$) we have now established (2.18).

To prove (2.16), we will show that for any $\mathbf{b}$ satisfying (2.15), if b_0 is below a sufficiently small multiple of N , then $k > \mathbb{E}(D)$. Suppose $j \in \{1, \dots, d - 1\}$ and $b_0 < cN$, where $c = \Omega(1)$ is a small constant less than 1 to be determined later. As long as c does not depend on N (this will require justification), we can pick c first and then pick N so large that $cN > .4$ simply because we may neglect boundedly many cases of N . Then, recalling that Λ' is an increasing function, we use (2.17) and (2.18) to see that

$$\begin{aligned} \Lambda'(b_j + 1) &\leq O(1) + \frac{d-j}{d}\Lambda'(cN + 1) + \frac{j}{d}\Lambda'(N + 1) \\ &\leq O\left(1 + \frac{1}{cN}\right) + \frac{d-j}{d}\ln(cN) + \frac{j}{d}\ln(N) \\ &\leq O\left(1 + \frac{1}{cN}\right) + \frac{d-j}{d}\ln(c) + \ln(N) \end{aligned} \quad (2.19)$$

Because $\frac{1}{cN} \leq (.4)^{-1} = O(1)$, (2.19) becomes

$$\Lambda'(b_j + 1) \leq O(1) + \frac{d-j}{d}\ln(c) + \ln(N). \quad (2.20)$$

Because c will be less than 1, $\ln(c) < 0$ so we may decrease its coefficient to majorize the expression, i.e.

$$\Lambda'(b_j + 1) \leq O(1) + \frac{1}{d}\ln(c) + \ln(N). \quad (2.21)$$

Exponentiating both sides of (2.21) shows that, for some large positive $C = O(1)$ independent of c and N ,

$$\exp[\Lambda'(b_j + 1)] \leq Cc^{1/d}N. \quad (2.22)$$

We use the approximation $\Lambda'(x) = \ln(x) - \frac{1}{2x} + O\left(\frac{1}{x^2}\right)$ (Claim 2.2.4, item (3)) to estimate

$$\exp[\Lambda'(x)] = x \exp\left[-\frac{1}{2x} + O\left(\frac{1}{x^2}\right)\right] = x - \frac{1}{2} + O\left(\frac{1}{x}\right) \quad (2.23)$$

for $x > 0$ bounded away from 0. We increase C if necessary to ensure that whenever $x > C$, we have $\exp[\Lambda'(x)] \geq x - 1$. Now, if $b_j > C$, (2.22) can be rewritten as

$$b_j \leq \exp[\Lambda'(b_j + 1)] \leq Cc^{1/d}N. \quad (2.24)$$

Again, we allow ourselves to pick N large enough (relative to c) that $c^{1/d}N \geq 1$ and thereby ensure that (2.24) holds when $b_j \leq C$ as well, thus justifying (2.24) for all possible values of b_j .

We apply (2.24) to see that

$$b_d = N - b_0 - \sum_{j=1}^{d-1} b_j \geq N(1 - c) - N(d-1)Cc^{1/d}.$$

Then, simply by the definition of k , we have

$$\begin{aligned} k &\geq db_d + O(1) \\ &\geq dN[1 - c - (d-1)Cc^{1/d}] + O(1). \end{aligned}$$

Notice that by construction $\mathbb{E}(D_1) < d$ and $N\mathbb{E}(D_1) = \mathbb{E}(D)$. Thus, as long as we pick c small enough (relative to the constants C and $d - \mathbb{E}(D_1)$, in particular independent of N), we can ensure that $k > \mathbb{E}(D)$. Thus we have shown that $b_0 \gtrsim N$ uniformly over $k \leq \mathbb{E}(D)$.

Now we proceed to the claim that for any $j \in \{1, \dots, d\}$,

$$b_j = \begin{cases} O(1) & \text{if } k = O\left(N^{\frac{j-1}{j}}\right) \\ \Theta\left(\frac{k^j}{N^{j-1}}\right) & \text{otherwise.} \end{cases}$$

Applying Lemma 2.2.6, we derive from the Lagrange multiplier condition (2.14),

$$\ln(p_j) - \Lambda'(b_j + 1) + \Lambda'(b_0 + 1) - \ln(p_0) = j \ln(p_1) - j\Lambda'(b_1 + 1) + j\Lambda'(b_0 + 1) - j \ln(p_0)$$

$$\begin{aligned}\Lambda'(b_j + 1) &= \ln \left(\frac{p_j p_0^{j-1}}{p_1^j} \right) - (j-1)\Lambda'(b_0 + 1) + j\Lambda'(b_1 + 1) \\ \exp [\Lambda'(b_j + 1)] &= \frac{p_j p_0^{j-1}}{p_1^j} \frac{\left(\exp [\Lambda'(b_1 + 1)] \right)^j}{\left(\exp [\Lambda'(b_0 + 1)] \right)^{j-1}}\end{aligned}\tag{2.25}$$

We use the fact that $b_0 \gtrsim N$ and our earlier estimate $\Lambda'(x) = \ln(x) - \frac{1}{2x} + O\left(\frac{1}{x^2}\right)$ (and simply absorb constants) to derive from (2.25),

$$\exp [\Lambda'(b_j + 1)] \sim \frac{\left(\exp [\Lambda'(b_1 + 1)] \right)^j}{N^{j-1}}.\tag{2.26}$$

We will use (2.26) twice, first to show that b_1 is bounded from below by a multiple of k (we will show this for k sufficiently large, however it is true in general), and then to complete the proof of the lemma. Noting that $b_1 \leq k \leq N\mathbb{E}(D_1)$ (and recalling the Λ' is increasing), we apply (2.18) to (2.26) to achieve the generally suboptimal bounds

$$\begin{aligned}\exp [\Lambda'(b_1 + 1)] &\lesssim N \\ \implies \exp [\Lambda'(b_j + 1)] &\lesssim \exp [\Lambda'(b_1 + 1)].\end{aligned}\tag{2.27}$$

Note that the claim

$$k = O\left(N^{\frac{j-1}{j}}\right) \implies b_j = O(1)$$

is vacuously true if $b_j < 0$ (as b_j is bounded below by -1 by construction) and it is thus safe to assume $b_j + 1$ is bounded away from 0. Therefore we are free to apply (2.23) to the left side of (2.27) (and (2.18) to the right) to further derive

$$b_j + O(1) \begin{cases} \leq 0 & \text{for } 0 \leq b_1 \leq .4 \text{ and} \\ \lesssim b_1 & \text{for } b_1 > .4. \end{cases}\tag{2.28}$$

Thus, if $b_1 \leq .4$, then

$$k = \sum_{j=1}^d j b_j = O(1)\tag{2.29}$$

and, if $b_1 > .4$, then

$$k = \sum_{j=1}^d j b_j \lesssim \sum_{j=1}^d j b_1 \lesssim b_1. \quad (2.30)$$

Neglecting the finitely many cases of k that could a priori fall under the (2.29) regime, we have (2.30) for general k and thus we have proven that $b_1 \sim k$.

Returning to (2.26), we apply (2.30) and (2.23) to get

$$b_j + O(1) = \exp [\Lambda'(b_j + 1)] \sim \frac{\left(\exp [\Lambda'(b_1 + 1)] \right)^j}{N^{j-1}} \sim \frac{[k + O(1)]^j}{N^{j-1}}.$$

Therefore, if $k = O\left(N^{\frac{j-1}{j}}\right)$, we have

$$b_j = O(1),$$

while if k is above a sufficiently large (constant) multiple of $N^{\frac{j-1}{j}}$, we have

$$b_j = \Theta\left(\frac{k^j}{N^{j-1}}\right),$$

proving the lemma. □

Definition 2.2.9 (Peak of $\mathbb{S}_k$). *We denote the (unique) point at which L achieves its maximum in $\mathcal{D}^k$, whose existence and properties are established in Proposition 2.2.8, by $\mathbf{b}^k$.*

Lemma 2.2.10. *The $\mathbb{R}^{d+1}$ -valued map*

$$g(k) = \begin{bmatrix} g_0(k) \\ g_1(k) \\ \vdots \\ g_d(k) \end{bmatrix} = \begin{bmatrix} b_0^k \\ b_1^k \\ \vdots \\ b_d^k \end{bmatrix}$$

from is analytic on $[0, \mathbb{E}(D)]$ and satisfies the smoothness bound

$$|g'(k)| = O(1).$$

Proof. For fixed $k \in [0, \mathbb{E}(D)]$, we define the map $F_k : \mathbb{R}^{d+1} \rightarrow \mathbb{R}^{d+1}$ by

$$\begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_d \end{bmatrix} \mapsto \begin{bmatrix} \sum_{i=0}^N b_i - N \\ \sum_{i=1} b_i - k \\ \Lambda'(b_0 + 1) - 2j\Lambda'(b_1 + 1) + \Lambda'(b_2 + 1) - \ln(p_0 p_1^{-2} p_2) \\ \vdots \\ (j-1)\Lambda'(b_0 + 1) - j\Lambda'(b_1 + 1) + \Lambda'(b_j + 1) - \ln(p_0^{j-1} p_1^{-j} p_j) \\ \vdots \\ (d-1)\Lambda'(b_0 + 1) - d\Lambda'(b_1 + 1) + \Lambda'(b_d + 1) - \ln(p_0^{d-1} p_1^{-d} p_d) \end{bmatrix}.$$

Note that the solutions to the equation $F_k(\mathbf{b}) = \mathbf{0}$ are precisely the critical points of $L(\mathbf{a})$ restricted to $\mathcal{D}^k$ which we know from the proof of Proposition 2.2.8 to be exactly $\{\mathbf{b}^k\}$. We compute the Jacobian $\mathcal{J}F_k(\mathbf{b}^k)$ to be the matrix $[f_{i,j}^k]_{\substack{0 \leq i \leq d \\ 0 \leq j \leq d}}$ given by

$$f_{i,j}^k = \begin{cases} 1 & \text{for } i = 0 \\ j & \text{for } i = 1 \\ (j-1)\Lambda''(b_0^k + 1) & \text{for } j = 0, i > 1 \\ -j\Lambda''(b_1^k + 1) & \text{for } j = 1, i > 1 \\ \Lambda''(b_j^k + 1) & \text{for } j = i > 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.31)$$

or, laid out more visually,

$$\mathcal{J}F_k(\mathbf{b}^k) = \left[\begin{array}{cc|cccc} 1 & 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 2 & 3 & \dots & d \\ \hline \Lambda''(b_0^k + 1) & -2\Lambda''(b_1^k + 1) & \Lambda''(b_2^k + 1) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (d-1)\Lambda''(b_0^k + 1) & -d\Lambda''(b_1^k + 1) & 0 & 0 & \dots & \Lambda''(b_d^k + 1) \end{array} \right].$$

where the partitions are merely for readability. Recall from Claim 2.2.3, item (3) that Λ'' is strictly positive and note that therefore $\mathcal{J}F_k(\mathbf{b}^k)$ can row reduce

to identity and is therefore invertible. Because F is analytic and $[F \circ g](k)$ is identically 0 for all $k \in [0, \mathbb{E}(D)]$, by the analytic implicit function theorem, g is also analytic.

Moreover, we can explicitly compute the $\mathbb{R}^{d+1}$ -valued derivative $g'(k)$ using implicit differentiation. Here the partitions separate the k parameter from the $b_0, \dots, b_d$ variables.

$$\begin{aligned}
\mathbf{0} = \partial_k [F_k \circ g](k) &= \left[\begin{array}{c|c} \begin{matrix} 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{matrix} & \mathcal{J}F_k(\mathbf{b}^{\mathbf{k}}) \end{array} \right] \left[\begin{array}{c} 1 \\ \hline g'(k) \end{array} \right] \\
&= \left[\begin{array}{c} 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{array} \right] + \left[\mathcal{J}F_k(\mathbf{b}^{\mathbf{k}}) \right] g'(k) \\
\Rightarrow g'(k) &= \left[\mathcal{J}F_k(\mathbf{b}^{\mathbf{k}}) \right]^{-1} \left[\begin{array}{c} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{array} \right] \tag{2.32}
\end{aligned}$$

so $g'(k)$ is simply the second column of $\left[\mathcal{J}F_k(\mathbf{b}^{\mathbf{k}}) \right]$, which we directly compute

to be

$$g'(k) = \begin{bmatrix} -1 \\ 1 \\ \frac{\Lambda''(b_0^k+1)+2\Lambda''(b_1^k+1)}{\Lambda''(b_2^k+1)} \\ \vdots \\ \frac{(j-1)\Lambda''(b_0^k+1)+j\Lambda''(b_1^k+1)}{\Lambda''(b_j^k+1)} \end{bmatrix}. \quad (2.33)$$

Proposition 2.2.8 provides the bounds $b_0^k \sim N$ and $b_1^k \sim k$ which, in conjunction with Claim 2.2.4, item (4), proves $\Lambda''(b_0^k+1) = O(\frac{1}{N})$ and $\Lambda''(b_1^k+1) = O(\frac{1}{k})$. Proposition 2.2.8 also states that for all $j \in \{2, \dots, d\}$, $b_j^k = O\left(1 + \frac{k^j}{N^{j-1}}\right) = O(k)$ so, because Λ'' is decreasing (Claim 2.2.3, item (2)), $\Lambda''(b_j^k) = \Omega(\frac{1}{k})$. Together with (2.33), this proves

$$g'(k) = \begin{bmatrix} O(1) \\ O(1) \\ \frac{O(1/N)+O(1/k)}{\Omega(1/k)} \\ \vdots \\ \frac{O(1/N)+O(1/k)}{\Omega(1/k)} \end{bmatrix} = \begin{bmatrix} O(1) \\ O(1) \\ O(1) \\ \vdots \\ O(1) \end{bmatrix}$$

so $|g'(k)| = O(1)$. □

Corollary 2.2.11. *For any $k \in \{1, \dots, \mathbb{E}(D)\}$,*

$$|\mathbf{b}^k - \mathbf{b}^{k-1}| = O(1)$$

Proof. Using the language and result of Lemma 2.2.10 as well as the mean value theorem, we bound

$$|\mathbf{b}^k - \mathbf{b}^{k-1}| \leq \sup_{\ell \in [k-1, k]} |g'(\ell)| = O(1).$$

□

Heuristically, $\mathbf{b}^k$ is the epicenter of mass concentration for types $\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k$ with respect to the type parameter $\mathbf{a}$. However, in general there is no reason for the

components of $\mathbf{b}^k$ to be integers and therefore it does not have to represent a nonempty type in $\mathbb{S}_k$. In §2.3 we will deal with many of the technical hurdles caused by the fact that $\mathbf{b}^k$ cannot be used directly to justify properties of the sphere.

2.3 Mass Concentration

This section is composed of technical estimates entirely toward the goal of proving Proposition 2.3.2 below and a few important consequences. The lemmas preceding the proof of the Proposition have no direct bearing outside of this section.

We will now quantify the Gaussian-like mass concentration of the log concave distribution of types within $\mathbb{S}_k$. Modeling the Gaussian case, we will formalize the notion of standard deviation from the peak in order to achieve exponential decay with respect to this deviation.

Definition 2.3.1 (Deviation of $\mathbf{a}$ and Deviant Index). *For any $\mathbf{a} \in \mathcal{D}^k$, define the deviation of $\mathbf{a}$ by*

$$\delta_{\mathbf{a}} := \max_{j \in \{1, \dots, d\}} \frac{|a_j - b_j^k|}{\max(1, |b_j^k|^{1/2})}. \quad (2.34)$$

We define the deviant index of $\mathbf{a}$ to be the minimal J that attains the outer maximum in (2.34).

Proposition 2.3.2. *We have the decay condition*

$$\mathbb{P}(\mathbb{T}_{\mathbf{a}}) = T(\mathbf{a}) \lesssim e^{-\Omega(\delta_{\mathbf{a}})} \left(\max_{\mathbf{b} \in \mathbb{N}^d} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}}) \right) \quad (2.35)$$

uniformly in $k \in [0, \mathbb{E}(D)]$, $\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k$.

Although intuitive for our log concave probability distribution among types, Proposition 2.3.2 requires extra attention in cases where $k = o(N)$. Lemmas 2.3.3 and 2.3.4 will handle the case $k = \Omega(N)$.

Lemma 2.3.3. *As long as $\delta_{\mathbf{a}} \geq 2$, L satisfies the decay condition*

$$L(\mathbf{a}) - L(\mathbf{b}^k) = -\Omega(\delta_{\mathbf{a}}). \quad (2.36)$$

Proof. Let $\mathbf{a} \in \mathcal{D}^k$ be such that $\delta_{\mathbf{a}} \geq 2$ and J be the deviant index of $\mathbf{a}$. Note that, because $\mathbf{b}^k$ is a critical point of L in $\mathcal{D}^k$,

$$\nabla L(\mathbf{b}^k) \cdot (\mathbf{a} - \mathbf{b}^k) = 0. \quad (2.37)$$

We parameterize the path from $\mathbf{b}^k$ to $\mathbf{a}$ by

$$\gamma(t) := \mathbf{b}^k + t(\mathbf{a} - \mathbf{b}^k)$$

for t going from 0 to 1 and use the chain rule and (2.37) to observe

$$\begin{aligned} \partial_t [L \circ \gamma](t) &= \nabla L[\mathbf{b}^k + t(\mathbf{a} - \mathbf{b}^k)] \cdot (\mathbf{a} - \mathbf{b}^k) - \nabla L(\mathbf{b}^k) \cdot (\mathbf{a} - \mathbf{b}^k) \\ &= \int_0^t \partial_s \left[\nabla L[\mathbf{b}^k + s(\mathbf{a} - \mathbf{b}^k)] \cdot (\mathbf{a} - \mathbf{b}^k) \right] ds \\ &= \int_0^t \left[(\mathbf{a} - \mathbf{b}^k) \cdot \mathcal{H}L(\mathbf{b}^k + s(\mathbf{a} - \mathbf{b}^k)) (\mathbf{a} - \mathbf{b}^k) \right] ds \\ &\leq \int_0^t \left[(a_J - b_J^k) \left(-\Lambda''(b_J^k + s(a_J - b_J^k) + 1) (a_J - b_J^k) \right) \right] ds \\ &= -(a_J - b_J^k)^2 \int_0^t \Lambda''(b_J^k + s(a_J - b_J^k + 1)) ds. \end{aligned} \quad (2.38)$$

Because $\Lambda'' > 0$ (Claim 2.2.3, item (3)), (2.38) implies for any $t \geq \frac{1}{\delta_{\mathbf{a}}}$ that

$$\partial_t [L \circ \gamma](t) \leq -(a_J - b_J^k)^2 \int_0^{\frac{1}{\delta_{\mathbf{a}}}} \Lambda''(b_J^k + s(a_J - b_J^k)) ds \quad (2.39)$$

First we handle the case where $b_J^k \leq 1$ so that $\delta_{\mathbf{a}} = |a_J - b_J^k|$. In this regime, the argument of Λ'' in the integrand of (2.39) is bounded by

$$b_J^k + \frac{(a_J - b_J^k)}{\delta_{\mathbf{a}}} = b_J^k + 1 \leq 2.$$

Thus, because $\Lambda''(2) > 0$ and Λ'' is decreasing (Claim 2.2.3, items (3) and (2)), we have

$$\Lambda''(b_J^k + s(a_J - b_J^k)) \geq \Lambda''(2) = \Omega(1)$$

for all $s \in \left[0, \frac{1}{\delta_{\mathbf{a}}}\right]$. From here (2.39) becomes

$$\partial_t [L \circ \gamma](t) \leq -\delta_{\mathbf{a}}^2 \int_0^{\frac{1}{\delta_{\mathbf{a}}}} \Omega(1) \, ds = -\Omega(\delta_{\mathbf{a}}) \quad (2.40)$$

If $b_J^k > 1$, we note that

$$\frac{1}{\delta_{\mathbf{a}}} = \frac{|b_J^k|^{1/2}}{|a_J - b_J^k|}$$

so, on the region of integration in (2.39), we have an upper bound on the argument of Λ'' , namely

$$\begin{aligned} b_J^k + s(a_J - b_J^k) &\leq b_J^k + \frac{|a_J - b_J^k|}{\delta_{\mathbf{a}}} \\ &= b_J^k + |b_J^k|^{1/2} \\ &\leq 2b_J^k. \end{aligned}$$

By Claim 2.2.4, item (4), because $b_J^k > 1$, i.e. bounded away from 0, $\Lambda''(2b_J^k) \sim \frac{1}{b_J^k}$. Moreover, by Claim 2.2.3, item (2), Λ'' decreases so on the entire region of integration we have the lower bound

$$\Lambda''(b_J^k + s(a_J - b_J^k)) = \Omega\left(\frac{1}{|b_J^k|}\right) \quad (2.41)$$

Returning to (2.39), the bound (2.41) implies

$$\begin{aligned} \partial_t [L \circ \gamma](t) &\leq -(a_J - b_J^k)^2 \int_0^{\frac{1}{\delta_{\mathbf{a}}}} \Omega\left(\frac{1}{|b_J^k|}\right) \, ds \\ &\leq -\int_0^{\frac{1}{\delta_{\mathbf{a}}}} \Omega\left(\frac{(a_J - b_J^k)^2}{|b_J^k|}\right) \, ds \\ &\leq -\int_0^{\frac{1}{\delta_{\mathbf{a}}}} \Omega(\delta_{\mathbf{a}}^2) \, ds \\ &= -\Omega(\delta_{\mathbf{a}}) \end{aligned} \quad (2.42)$$

Combined, (2.40) and (2.42) prove the general bound $\partial_t [L \circ \gamma](t) = -\Omega(\delta_{\mathbf{a}})$ for all $t \geq \frac{1}{\delta_{\mathbf{a}}}$. Finally we recall $\frac{1}{\delta_{\mathbf{a}}}$ is bounded away from 1 simply because $\delta_{\mathbf{a}} \geq 2$ and integrate the (negative on its domain) function $\partial_t [L \circ \gamma](t)$ to get

$$L(\mathbf{a}) - L(\mathbf{b}^k) = \gamma(1) - \gamma(0)$$

$$\begin{aligned}
&= \int_0^1 \partial_t [L \circ \gamma](t) dt \\
&\leq \int_{\frac{1}{\delta_{\mathbf{a}}}}^1 \partial_t [L \circ \gamma](t) dt \\
&\leq - \left(1 - \frac{1}{\delta_{\mathbf{a}}}\right) \Omega(\delta_{\mathbf{a}}) \\
&\leq -\Omega(\delta_{\mathbf{a}}),
\end{aligned}$$

thus proving the claim $L(\mathbf{a}) - L(\mathbf{b}^k) = -\Omega(\delta_{\mathbf{a}})$ for any b_j^k . $\square$

Lemma 2.3.4. *Fix a (small) constant $c > 0$. Then for $k \geq cN$ we have the decay condition*

$$\mathbb{P}(\mathbb{T}_{\mathbf{a}}) = T(\mathbf{a}) \lesssim_c e^{-\Omega(\delta_{\mathbf{a}})} \left(\max_{\mathbf{b} \in \mathbb{N}^d} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}}) \right) \quad (2.43)$$

where the implied constant depends only on c , i.e. it is independent of $cN < k \leq \mathbb{E}(D)$, $\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k$.

Proof. Fixing $k \in [cN, \mathbb{E}(D)]$, $\mathbf{a} \in \mathcal{D}^k$, we observe that L and therefore T have maxima at $\mathbf{b}^k$ so $T(\mathbf{a}) \leq T(\mathbf{b}^k)$. Therefore, if $\delta_{\mathbf{a}} < 2$, the condition (2.43) holds vacuously. If $\delta_{\mathbf{a}} \geq 2$, we apply Lemma 2.3.3 and exponentiate the inequality $L(\mathbf{a}) - L(\mathbf{b}^k) = -\Omega(\delta_{\mathbf{a}})$ to get $T(\mathbf{a}) \leq T(\mathbf{b}^k)e^{-\Omega(\delta_{\mathbf{a}})}$. The lemma will follow if we can show there exists a nonempty type $\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}_k$ such that $T(\mathbf{b}) \sim_c T(\mathbf{b}^k)$. To this end, we define an approximate peak type parameter by

$$b_j = \begin{cases} \lceil b_j^k \rceil & \text{for } j \in \{2, \dots, d\}, \\ k - \sum_{j=2}^d j b_j, & \text{for } j = 1, \text{ and} \\ N - \sum_{j=1}^d b_j & \text{for } j = 0. \end{cases}$$

Because all components of $\mathbf{b}^k$ are bounded below by -1 , for each $j \in \{2, \dots, d\}$, $b_j \geq 0$. Moreover by Proposition 2.2.8 $b_1^k \sim k$ so we have

$$b_1 = k - \sum_{j=2}^d j b_j \quad (2.44)$$

$$= k - \sum_{j=2}^d j b_j^k + O(1) \quad (2.45)$$

$$= b_1^k + O(1) = \Theta(k) \quad (2.46)$$

so $b_1 \geq 0$ for k sufficiently large and similarly by Proposition 2.2.8 $b_0^k \sim N$ so

$$b_0 = N - \sum_{j=1}^d b_j \quad (2.47)$$

$$= N - \sum_{j=1}^d b_j^k + O(1) \quad (2.48)$$

$$= b_0^k + O(1) = \Theta(N) \quad (2.49)$$

so $b_0 \geq 0$ for N sufficiently large. Because $\mathbf{b}^k$ is a critical point of $L(\mathbf{a})$ and $\mathbf{b}^k, \mathbf{b} \in \mathcal{D}^k$, we have

$$\nabla L(\mathbf{b}^k) \cdot (\mathbf{b} - \mathbf{b}^k) = 0. \quad (2.50)$$

Defining $\gamma(t) = \mathbf{b}^k + t(\mathbf{b} - \mathbf{b}^k)$, we have

$$\begin{aligned} L(\mathbf{b}) - L(\mathbf{b}^k) &= \int_0^1 \partial_t [L \circ \gamma](t) dt \\ &= \int_0^1 \nabla L(\mathbf{b}^k + t(\mathbf{b} - \mathbf{b}^k)) \cdot (\mathbf{b} - \mathbf{b}^k) dt \\ &= \int_0^1 \int_0^t \partial_s \left[\nabla L(\mathbf{b}^k + s(\mathbf{b} - \mathbf{b}^k)) \cdot (\mathbf{b} - \mathbf{b}^k) \right] ds dt \quad (2.51) \\ &= \int_0^1 \int_0^t \left[(\mathbf{b} - \mathbf{b}^k) \cdot \mathcal{H}L(\mathbf{b}^k + s(\mathbf{b} - \mathbf{b}^k)) (\mathbf{b} - \mathbf{b}^k) \right] ds dt \\ &= \int_0^1 \int_0^t \sum_{j=0}^d \left[-\Lambda''(b_j^k + s(b_j - b_j^k) + 1) (b_j - b_j^k)^2 \right] ds dt \end{aligned} \quad (2.52)$$

where the equality (2.51) used (2.50) with the fundamental theorem of calculus.

First we observe that, by construction, for all $j \in \{0, \dots, d\}$,

$$(b_j - b_j^k)^2 \lesssim 1 \quad (2.53)$$

Also, notice that for all $j \in \{0, \dots, d\}$, because $k \sim_c N$ in this lemma, $b_j^k \gtrsim_c N^{1/j}$. Moreover for all $j \in \{2, \dots, d\}$, $b_j \geq b_j^k$ and, by (2.44) and (2.47), $b_1 \sim k \sim_c N$ and $b_0 \sim N$. Therefore for all $j \in \{0, \dots, d\}$, $s \in [0, 1]$, the expression $b_j^k + s(b_j - b_j^k)$ is a convex combination of two values of size $\omega(1)$, in particular both bounded away from -1 . Therefore the argument $b_j^k + s(b_j - b_j^k) + 1$ is bounded away from 0 so Claim 2.2.4, item (4) implies that

$$\Lambda''(b_j^k + s(b_j - b_j^k) + 1) = O\left(\frac{1}{b_j^k}\right) \quad (2.54)$$

because (2.53) and the bounds away from 0 on b_j and b_j^k imply

$$b_j^k \sim b_j \implies b_j^k + s(b_j - b_j^k) \sim b_j^k$$

for all $s \in [0, 1]$. Then (2.53) and (2.54) imply that the magnitude of each summand in the integrand in (2.52) satisfies the bound

$$\Lambda''(b_j^k + s(b_j - b_j^k) + 1) (b_j - b_j^k)^2 = O\left(\frac{1}{b_j^k}\right) = O(N^{-1/d}). \quad (2.55)$$

In particular, weakening the bound (2.55), the entire integrand in (2.52) is bounded from below by some negative constant. Therefore

$$L(\mathbf{b}) - L(\mathbf{b}^k) = \int_0^1 \int_0^t O(1) ds dt = O(1). \quad (2.56)$$

Exponentiating both sides of (2.56) proves that $T(\mathbf{b}) \sim T(\mathbf{b}^k)$ and so we may apply Lemma 2.3.3 to get

$$T(\mathbf{a}) \lesssim e^{-\Omega(\delta_{\mathbf{a}})} T(\mathbf{b}^k) \sim_c e^{-\Omega(\delta_{\mathbf{a}})} T(\mathbf{b}) \leq e^{-\Omega(\delta_{\mathbf{a}})} \left(\max_{\mathbf{b} \in \mathbb{N}^d} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}}) \right),$$

thus concluding the proof. $\square$

Lemma 2.3.5. *Fix a (large) constant $C > 1$. Then for any $\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k$, if $J \in \{1, \dots, d\}$ is the deviant index of $\mathbf{a}$, then we have the decay condition*

$$\mathbb{P}(\mathbb{T}_{\mathbf{a}}) = T(\mathbf{a}) \lesssim e^{-\Omega(\delta_{\mathbf{a}})} \left(\max_{\mathbf{b} \in \mathbb{N}^d} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}}) \right) \quad (2.57)$$

where the implied constant is independent of $k \leq \mathbb{E}(D)$, $\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k$, $J \in \{1, \dots, d\}$ as long as $b_J^k < C$.

Proof. As in Lemma 2.3.3, let J be the deviant index of $\mathbf{a}$. Throughout this proof, our asymptotic notation assumes $b_J^k \leq C$. By Proposition 2.2.8, if $b_J^k = O_C(1)$ then $k^J \lesssim N^{J-1}$. Moreover, also by Proposition 2.2.8, $b_0^k \sim N$ and $b_1^k \sim k = o(N)$. Heuristically, in this case $\mathbb{S}_k$ is so weighted against distance J that the type masses $\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}})$ will decrease exponentially in a_J for sufficiently large a_J .

However, in order to establish decay with respect to a_J , it is crucial that we examine only types in the sphere $\mathbb{S}_k$. Therefore we compute for any $\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}_k$

$$\begin{aligned}
\frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}+(J-1)\mathbf{e}_0-J\mathbf{e}_1+\mathbf{e}_J})}{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}})} &= \frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}+(J-1)\mathbf{e}_0-J\mathbf{e}_1+\mathbf{e}_J}) \mathbb{P}(\mathbb{S}_k)}{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}}) \mathbb{P}(\mathbb{S}_k)} \\
&= \frac{\mathbb{P}(\mathbb{T}_{\mathbf{b}+(J-1)\mathbf{e}_0-J\mathbf{e}_1+\mathbf{e}_J})}{\mathbb{P}(\mathbb{T}_{\mathbf{b}})} \\
&= \frac{\mathbf{p}^{\mathbf{b}+(J-1)\mathbf{e}_0-J\mathbf{e}_1+\mathbf{e}_J}(\mathbf{b}+(J-1)\mathbf{e}_0-J\mathbf{e}_1+\mathbf{e}_J)^N}{\mathbf{p}^{\mathbf{b}}(\mathbf{b})^N} \\
&= \left(\frac{p_0^{J-1} p_J}{p_1^J} \right) \frac{\prod_{i=0}^{J-1} (b_1 - i)}{(b_J + 1) \prod_{i=1}^{J-1} (b_0 + i)} \quad (2.58)
\end{aligned}$$

Recall that all components of $\mathbf{b}$ are bounded from below by -1 and

$$k = \sum_{j=1}^d j b_j.$$

Thus, for all $j \in \{1, \dots, d\}$,

$$b_j \leq k + \left(\sum_{i \neq j} i \right) = O(k) = o(N) \quad (2.59)$$

and thus

$$b_0 = N - \sum_{j=1}^d b_j \geq \frac{N}{2}$$

$$b_1 \leq k + O(1) \leq 2k$$

as long as k and N are sufficiently large. Thus,

$$\left(\frac{p_0^{J-1} p_J}{p_1^J} \right) \frac{\prod_{i=0}^{J-1} (b_1 - i)}{\prod_{i=1}^{J-1} (b_0 + i)} \leq \left(\frac{p_0^{J-1} p_J}{p_1^J} \right) \frac{(2k)^J}{\left(\frac{N}{2}\right)^{J-1}} \lesssim 1 \quad (2.60)$$

because $k^J \lesssim N^{J-1}$. Therefore there is a constant C such that

$$\left(\frac{p_0^{J-1} p_J}{p_1^J} \right) \frac{\prod_{i=0}^{J-1} (b_1 - i)}{\prod_{i=1}^{J-1} (b_0 + i)} \leq C.$$

Because (2.59) implies that for any $\mathbb{T}_{\mathbf{b}} \in \mathbb{S}_k$, $b_J = o(N)$ (and $b_0 \geq \frac{N}{2}$), we can assume that $b_0 - (J-1)b \geq 0$ for any $b \leq b_J$. Therefore we can apply (2.58) to each of the types

$$\left\{ \mathbb{T}_{\mathbf{a} - a[(J-1)\mathbf{e}_0 - J\mathbf{e}_1 + a_J]} \right\}_{a=0}^{J-1}$$

sequentially and apply (2.60) to find

$$\begin{aligned} & \frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}})}{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a} - a_J[(J-1)\mathbf{e}_0 - J\mathbf{e}_1 + \mathbf{e}_J]})} \\ &= \prod_{m=1}^{a_J} \left[\left(\frac{p_0^{J-1} p_J}{p_1^J} \right) \frac{\prod_{i=0}^{J-1} [(a_1 + mJ) - i]}{(a_J - m + 1) \prod_{i=1}^{J-1} [a_0 - m(J-1) + i]} \right] \\ &\leq \prod_{m=1}^{a_J} \frac{C}{a_J - m + 1} \\ &\leq \frac{C^{a_J}}{a_J!} \\ &\lesssim e^{-a_J}. \end{aligned} \tag{2.61}$$

Condition (2.57) holds vacuously if $\delta_{\mathbf{a}} \lesssim 1$, so we may assume $\delta_{\mathbf{a}} \geq C$. Then, because J is the deviant index of $\mathbf{a}$ and $|b_J^k| \leq C$,

$$\begin{aligned} \delta_{\mathbf{a}} &= \frac{|a_J - b_J^k|}{\max\left(1, |b_J^k|^{1/2}\right)} \\ \implies \frac{|a_J - b_J^k|}{C^{1/2}} &\leq \delta_{\mathbf{a}} \leq |a_J - b_J^k| \\ \implies |a_J - b_J^k| &\sim_C \delta_{\mathbf{a}}. \end{aligned} \tag{2.62}$$

Moreover,

$$3C \leq \delta_{\mathbf{a}} \leq |a_J - b_J^k|.$$

Therefore, because $\mathbf{a} \in \mathbb{N}^{d+1}$, $a_J \geq 0$ and, because $|b_J^k| \leq C$, if a_J and b_J^k are at least $3C$ apart, then $a_J \geq 2C$. Therefore we have

$$a_J \sim_C |a_J - b_J^k|,$$

which, in conjunction with (2.62), yields

$$a_J \sim_C \delta_{\mathbf{a}}. \quad (2.63)$$

Thus, by (2.61),

$$\begin{aligned} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) &\lesssim e^{-a_J} \left(\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a} - a_J[(J-1)\mathbf{e}_0 - J\mathbf{e}_1 + \mathbf{e}_J]} \right) \\ &\leq e^{-\Omega(\delta_{\mathbf{a}})} \left(\max_{\mathbf{b} \in \mathbb{N}^d} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}}) \right), \end{aligned}$$

completing the proof. $\square$

Lemma 2.3.6. *There exist a (large) constant $C > 1$ and a (small) constant $c > 0$ such that for any $\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k$, if $J \in \{1, \dots, d\}$ is the deviant index of $\mathbf{a}$, then we have the decay condition*

$$\mathbb{P}(\mathbb{T}_{\mathbf{a}}) \lesssim e^{-\Omega(\delta_{\mathbf{a}})} \left(\max_{\mathbf{b} \in \mathbb{N}^d} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}}) \right) \quad (2.64)$$

where the implied constant is independent of $k \leq cN$, $\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k$, $J \in \{1, \dots, d\}$ as long as $b_J^k \geq C$.

Proof. The constants c and C will be determined over the course of the proof, but will always be chosen and adjusted independently of N , k , $\mathbf{a}$, and J . Due to the many instances of nested quantifiers, we will often chose clarity over brevity in this proof and use named constants (n for large constant or $1/n$ for small constants) rather than asymptotic notation when it could be confusing.

As in the proofs of Lemmas 2.3.4 and 2.3.5, the condition (2.64) holds vacuously when $\delta_{\mathbf{a}} \leq C$ so we may assume $\delta_{\mathbf{a}} > C$ (as long as C is determined before $\mathbf{a}$). Because $\mathbf{a}, \mathbf{b}^k \in \mathcal{D}^k$,

$$\sum_{j=1}^d j(a_j - b_j^k) = \sum_{j=1}^d j a_j - \sum_{j=1}^d j b_j^k = k - k = 0. \quad (2.65)$$

By Proposition 2.2.8, there exists a constant n such that for all $j \in \{1, \dots, d\}$,

$$b_j^k \leq n \max \left(1, \frac{k^j}{N^{j-1}} \right). \quad (2.66)$$

Moreover, if C is chosen sufficiently large relative to the implicit constants in Proposition 2.2.8, we can ensure the bound

$$b_j^k \geq \frac{k^j}{nN^{j-1}}. \quad (2.67)$$

for all $j \in \{1, \dots, J\}$. We now show that for any $j > J$, $|a_j - b_j^k|$ is substantially smaller than $|a_J - b_J^k|$. Because J is the deviant index of $\mathbf{a}$, for all $j \in \{1, \dots, d\}$,

$$\frac{|a_j - b_j^k|}{\max\left(1, |b_j^k|^{1/2}\right)} \leq \delta_{\mathbf{a}} = \frac{|a_J - b_J^k|}{|b_J^k|^{1/2}} \leq \frac{|a_J - b_J^k|}{C^{1/2}}. \quad (2.68)$$

Thus, if $b_j^k \leq n$, (2.68) implies

$$|a_j - b_j^k| \leq \sqrt{n} \frac{|a_j - b_j^k|}{\max\left(1, |b_j^k|^{1/2}\right)} \leq |a_J - b_J^k| \left(\frac{n}{C}\right)^{1/2}. \quad (2.69)$$

If $j > J$ but $b_j^k > n$, we combine (2.66) and (2.67) to find,

$$\frac{b_j^k}{b_J^k} \leq \frac{(nk^j)/N^{j-1}}{k^J/(nN^{J-1})} = n^2 \frac{k^{j-J}}{N^{j-J}} \leq c^{j-J} n^2 \leq cn^2 \quad (2.70)$$

where the last equality requires only that $c \leq 1$. In this case we use (2.70) to observe

$$\begin{aligned} \frac{|a_j - b_j^k|}{|b_j^k|^{1/2}} &\leq \frac{|a_J - b_J^k|}{|b_J^k|^{1/2}} \\ \implies |a_j - b_j^k| &\leq |a_J - b_J^k| \left(\frac{|b_j^k|}{|b_J^k|}\right)^{1/2} \\ &\leq |a_J - b_J^k| n\sqrt{c}. \end{aligned} \quad (2.71)$$

Combining (2.69) and (2.71), we can pick C sufficiently large and c sufficiently small relative to d and n , we can ensure that for all $j > J$,

$$|a_j - b_j^k| \leq \frac{|a_J - b_J^k|}{d^2}. \quad (2.72)$$

Combining (2.72) and (2.65) then yields the bound

$$\left| \sum_{j=1}^J j(a_j - b_j^k) \right| \leq \left| \sum_{j=1}^d j(a_j - b_j^k) \right| + \sum_{j=J+1}^d j|a_j - b_j^k|$$

$$\begin{aligned}
&= \sum_{j=J+1}^d j|a_j - b_j^k| \\
&\leq \sum_{j=J+1}^d j \frac{|a_J - b_J^k|}{d^2} \\
&\leq |a_J - b_J^k|. \tag{2.73}
\end{aligned}$$

Rearranging (2.73) then proves $\sum_{j=1}^{J-1} j(a_j - b_j^k)$ lies between $-(J-1)(a_J - b_J^k)$ and $-(J+1)(a_J - b_J^k)$ so there must exist some $I \in \{1, \dots, J-1\}$ such that

$$\begin{aligned}
|a_I - b_I^k| &\geq \frac{|a_J - b_J^k|}{I} \text{ and} \\
(a_I - b_I^k) &\text{ has the opposite sign as } (a_J - b_J^k).
\end{aligned}$$

Moreover, as long as $C \geq n$, (2.66) implies

$$b_J^k \leq n \frac{k^J}{N^{J-1}}$$

and thus

$$\frac{b_I^k}{b_J^k} \geq \frac{k^I / (nN^{I-1})}{(nk^J / N^{J-1})} = \frac{1}{n^2} \left(\frac{N}{k} \right)^{J-I} \geq \frac{1}{cn^2}. \tag{2.74}$$

Therefore, if c is sufficiently small relative to n , (2.74) implies

$$b_I^k \geq b_J^k \geq C.$$

We have now completed the first portion of the proof. To summarize, we have found an $I \in \{1, \dots, J-1\}$ such that $a_I - b_I^k$ negates and is at least comparable to $a_J - b_J^k$ and b_I^k is at least as large as b_k^I or, more precisely,

$$|a_I - b_I^k| \geq \frac{1}{I} |a_J - b_J^k| \geq \frac{1}{d} |a_J - b_J^k|, \tag{2.75}$$

$$b_I^k \geq b_J^k \geq C, \text{ and} \tag{2.76}$$

$$(a_I - b_I^k) \text{ has the opposite sign as } (a_J - b_J^k). \tag{2.77}$$

If there is more than one such index, let I be the minimal index that achieves the maximum of the set

$$\{|a_i - b_i^k| : 1 \leq i < J \text{ and } (a_i - b_i^k) \text{ has the opposite sign as } (a_J - b_J^k)\}. \tag{2.78}$$

Heuristically, a_I deviates substantially from the concentration of types within $\mathbb{S}_k$ in order to compensate for the deviation of a_J in the opposite direction. Our strategy for the remainder of the proof is to show that if we examine types in $\mathbb{S}_k$ whose I^{th} component is closer than a_I to b_I^k and whose J^{th} component is closer than a_J to b_J^k , we will find that they have substantially higher probability than $\mathbb{T}_{\mathbf{a}}$. Indeed, the increase in probability as we move our attention from $\mathbb{T}_{\mathbf{a}}$ to types closer to $\mathbf{b}^k$ (in the I^{th} and J^{th} components) will suffice to prove our desired bound (2.64).

To avoid handling the cases $a_J < b_J^k$ and $a_J > b_J^k$ separately, we denote the sign of $a_J - b_J^k$ by

$$\sigma := \text{sgn}(a_J - b_J^k)$$

and define the function $F : \mathbb{R}^3 \rightarrow \mathbb{R}$ by

$$F(b_0, b_I, b_J) = \ln \left[p_0^{\sigma(I-J)} p_I^{\sigma J} p_J^{-\sigma I} \right] - \ln \left[b_0^{\sigma(I-J)} b_I^{\sigma J} b_J^{-\sigma I} \right]$$

to approximate the log of the ratio

$$R(\mathbf{b}) := \frac{\mathbb{P}(\mathbb{T}_{\mathbf{b} + \sigma(I-J)\mathbf{e}_0 + \sigma J\mathbf{e}_I - \sigma I\mathbf{e}_J})}{\mathbb{P}(\mathbb{T}_{\mathbf{b}})}$$

for a generic type $\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}_k$. The parameter

$$\mathbf{b} + (I - J)\mathbf{e}_0 + J\mathbf{e}_I - I\mathbf{e}_J$$

heuristically represents a “neighboring” type to $\mathbb{T}_{\mathbf{b}}$ that is slightly closer than $\mathbf{b}$ to $\mathbf{b}^k$ and therefore more probable. To estimate $F(b_0, b_I, b_J)$, we apply the Lagrange multiplier condition (2.14) to b_I^k and b_J^k and derive

$$\begin{aligned} IL_J(\mathbf{b}^k) - I \ln(p_0) + I\Lambda'(b_0^k + 1) &= JL_I(\mathbf{b}^k) - J \ln(p_0) + J\Lambda'(b_0^k + 1) \\ \implies \ln[p_J^I p_0^{J-I} p_I^{-J}] &= I\Lambda'(b_J^k + 1) + (J - I)\Lambda'(b_0^k + 1) - J\Lambda'(b_I^k + 1) \\ \implies \ln[p_J^I p_0^{J-I} p_I^{-J}] &= \ln \left[(b_J^k)^I (b_0^k)^{J-I} (b_I^k)^{-J} \right] + O\left(\frac{1}{b_0^k} + \frac{1}{b_I^k} + \frac{1}{b_J^k}\right) \end{aligned} \quad (2.79)$$

where the last implication in (2.79) used Claim 2.2.4, item (2) and the fact that b_0^k , b_I^k , and b_J^k are bounded away from -1 simply because $b_I^k, b_J^k \geq C$ by (2.76) and

$b_0^k \sim N$ by Proposition 2.2.8. We may simplify this bound a bit by recalling, again by (2.76), that $b_I^k \geq b_J^k$ and, as in (2.59), b_I^k and b_J^k are bounded by a constant multiple of k . Because b_0^k is bounded below by a constant multiple of N , as long as c is sufficiently small (i.e. k and therefore b_I^k and b_J^k are sufficiently small relative to N), we can ensure that

$$b_J^k \leq b_I^k \leq b_0^k. \quad (2.80)$$

Applying (2.80) to (2.79), we have

$$\begin{aligned} \ln [p_J^I p_0^{J-I} p_I^{-J}] &= \ln \left[(b_J^k)^I (b_0^k)^{J-I} (b_I^k)^{-J} \right] + O\left(\frac{1}{b_J^k}\right) \\ \implies \ln [p_J^I p_0^{J-I} p_I^{-J}] - \ln \left[(b_J^k)^I (b_0^k)^{J-I} (b_I^k)^{-J} \right] &= O\left(\frac{1}{b_J^k}\right) \\ \implies F(b_0^k, b_I^k, b_J^k) &= O\left(\frac{1}{b_J^k}\right) \end{aligned} \quad (2.81)$$

because by construction

$$\sigma F(b_0^k, b_I^k, b_J^k) = \ln [p_J^I p_0^{J-I} p_I^{-J}] - \ln \left[(b_J^k)^I (b_0^k)^{J-I} (b_I^k)^{-J} \right].$$

We now seek a parameter $\mathbf{b}$ for a nonempty type that lies roughly in the middle of $\mathbf{a}$ and $\mathbf{b}^k$, in a sense to be made more concrete. We define

$$\begin{aligned} b &:= \left\lfloor \min \left(\frac{|a_J - b_J^k|}{2I}, \frac{|a_I - b_I^k|}{2J} \right) \right\rfloor \\ \mathbf{b} &:= \mathbf{a} + b\sigma[(I - J)\mathbf{e}_0 + J\mathbf{e}_I - I\mathbf{e}_J] \end{aligned}$$

Notice that for $j \in \{I, J\}$, b_j lies strictly between a_j and b_j^k . The first implication of this fact is that

$$\operatorname{sgn}(b_J - b_J^k) = \sigma = -\operatorname{sgn}(b_I - b_I^k). \quad (2.82)$$

For $j \in \{I, J\}$, if $a_j \geq b_j^k/2$, then by $b_j = \Omega(b_j^k)$ because b_j lies between a_j and b_j^k . If $a_j < b_j^k/2$ and therefore $|a_J - b_J^k| \gtrsim b_J^k$, then by the construction of $\mathbf{b}$,

$$b_j = a_j + \Omega(b) = a_j + \Omega(|a_J - b_J^k|) = \Omega(b_J^k)$$

where the bound $b = \Omega(|a_J - b_J^k|)$ came from (2.75). Thus we have that $b_0, b_I, b_J > 0$ and

$$b_j = \begin{cases} = \Omega(b_j^k) & \text{for } j = I, J \\ = \Omega(N) & \text{for } j = 0. \end{cases} \quad (2.83)$$

In order to estimate $F(b_0, b_I, b_J)$, we parameterize the segment from (b_0^k, b_I^k, b_J^k) to (b_0, b_I, b_J) by $\gamma : [0, 1] \rightarrow \mathbb{R}^3$ given by

$$\gamma(t) := (b_0^k, b_I^k, b_J^k) + t(b_0 - b_0^k, b_I - b_I^k, b_J - b_J^k).$$

and compute

$$\partial_t[F \circ \gamma](t) = \frac{\sigma(J - I)(b_0 - b_0^k)}{b_0^k + t(b_0 - b_0^k)} - \frac{\sigma J(b_I - b_I^k)}{b_I^k + t(b_I - b_I^k)} + \frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)}. \quad (2.84)$$

By (2.83) all on the denominators in (2.84) are positive. Furthermore, by (2.82),

$$-\frac{\sigma J(b_I - b_I^k)}{b_I^k + t(b_I - b_I^k)}, \frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)} > 0. \quad (2.85)$$

If $\text{sgn}(b_0 - b_0^k) = \sigma$, then the term $\frac{\sigma(J-I)(b_0 - b_0^k)}{b_0^k + t(b_0 - b_0^k)} > 0$ so (2.84) implies that

$$\partial_t[F \circ \gamma](t) \geq \frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)} \quad (2.86)$$

for all $t \in [0, 1]$. This corresponds geometrically to the situation in which the direction $\sigma[(I - J)\mathbf{e}_0 + J\mathbf{e}_I - I\mathbf{e}_J]$ not only brings $\mathbf{b}$ toward the concentration of types in $\mathbb{S}_k$ with respect to the I^{th} and J^{th} components, but also in the 0^{th} component.

This is not necessarily the case so we must show that even if $\text{sgn}(b_0 - b_0^k) = -\sigma$, i.e. the $\sigma[(I - J)\mathbf{e}_0 + J\mathbf{e}_I - I\mathbf{e}_J]$ moves the 0^{th} component of $\mathbf{b}$ away from the concentration of types in $\mathbb{S}_k$, the term $\frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)}$ in (2.84) will counteract the effect of the term $\frac{\sigma(J-I)(b_0 - b_0^k)}{b_0^k + t(b_0 - b_0^k)}$ and the bound (2.86) will hold (up to a constant) despite the loss of probability due to the 0^{th} component.

To bound $\frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)}$ from below, we first observe that, if $b_J > 2b_J^k$, then

$$\frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)} \geq \frac{I(b_J/2)}{b_J} \geq \frac{1}{2}. \quad (2.87)$$

If $b_J \leq 2b_J^k$, we immediately have the upper bound

$$b_J^k + t(b_J - b_J^k) \leq 2b_J^k \quad (2.88)$$

on the denominator. To get a lower bound on the numerator we observe

$$|b_J - b_J^k| \geq |a_J - b_J^k| - |a_J - b_J| = |a_J - b_J^k| - Ib \geq \frac{|a_J - b_J^k|}{2} \quad (2.89)$$

by the construction of $\mathbf{b}$. Combining (2.88) and (2.89) provides the bound

$$\frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)} \geq \left(\frac{I}{4}\right) \frac{\delta_{\mathbf{a}}}{|b_J^k|^{1/2}}. \quad (2.90)$$

In order to achieve the lower bound (2.86) in the current case, namely when $\text{sgn}(b_0 - b_0^k) = -\sigma$, we must compare $b_0 - b_0^k$ to the bounds (2.87) and (2.90).

First we note that, because $\mathbf{b}, \mathbf{b}^k \in \mathcal{D}^k$, we have

$$\begin{aligned} \sum_{j=1}^d |b_j| &\leq \sum_{j=1}^d j|b_j| \leq k + \sum_{j=1}^d j \leq 2k \text{ and} \\ \sum_{j=1}^d |b_j^k| &\leq \sum_{j=1}^d j|b_j^k| \leq k + \sum_{j=1}^d j \leq 2k \end{aligned} \quad (2.91)$$

for k sufficiently large relative to d . Then for all $j \in \{1, \dots, d\}$, (2.91) implies the blunt bound

$$|b_j - b_j^k| \leq 4k. \quad (2.92)$$

Because $b_0 = N - \sum_{j=1}^d b_j$ and $b_0^k = N - \sum_{j=1}^d b_j^k$, (2.91) implies

$$b_0, b_0^k \geq N - 2dk \quad (2.93)$$

and (2.92) implies

$$|b_0 - b_0^k| \leq \sum_{j=1}^d |b_j - b_j^k| \leq 4dk. \quad (2.94)$$

Putting (2.92) and (2.94) together then proves

$$\left| \frac{(J-I)(b_0 - b_0^k)}{b_0^k + t(b_0 - b_0^k)} \right| \leq \left| \frac{(J-I)(4dk)}{N - 2dk} \right|$$

so, as long as c is sufficiently small relative to d ,

$$\left| \frac{(J-I)(b_0 - b_0^k)}{b_0^k + t(b_0 - b_0^k)} \right| \leq \frac{1}{4}. \quad (2.95)$$

If $b_J > 2b_J^k$ and therefore (2.87) holds, $\frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)} \geq \frac{1}{2}$ so (2.95) implies

$$\left| \frac{(J-I)(b_0 - b_0^k)}{b_0^k + t(b_0 - b_0^k)} \right| \leq \left(\frac{1}{2} \right) \frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)}. \quad (2.96)$$

By (2.91), $b_J^k \leq 2k$ so (2.90) implies

$$\frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)} \geq \left(\frac{I}{4} \right) \frac{\delta_{\mathbf{a}}}{(2k)^{1/2}} \geq \left(\frac{I}{4\sqrt{2}} \right) \frac{\delta_{\mathbf{a}}}{(cN)^{1/2}}. \quad (2.97)$$

Notice by the construction of $\mathbf{b}$ that for all $j \notin \{0, I, J\}$, $b_j = a_j$. Moreover, because of our earlier observation that b_I lies between a_I and b_I^k while b_J lies between a_J and b_J^k , we also have the inequalities $|b_j - b_j^k| \leq |a_j - b_j^k|$ for $j = I, J$. Therefore we have for all $j \in \{1, \dots, d\}$ that

$$|b_j - b_j^k| \leq |a_j - b_j^k|$$

and we may refine (2.94) to obtain

$$\begin{aligned} |b_0 - b_0^k| &\leq \sum_{j=1}^d |b_j - b_j^k| \leq \sum_{j=1}^d |a_j - b_j^k| \\ &\leq \sum_{j=1}^d |b_j^k|^{1/2} \delta_{\mathbf{a}} \leq d(2k)^{1/2} \delta_{\mathbf{a}} \leq \sqrt{2}d(cN)^{1/2} \delta_{\mathbf{a}} \end{aligned} \quad (2.98)$$

If we pick c sufficiently small that $b_0, b_0^k > \frac{N}{2}$, then (2.98) implies

$$\begin{aligned} \left| \frac{(J-I)(b_0 - b_0^k)}{b_0^k + t(b_0 - b_0^k)} \right| &\leq \frac{(J-I)\sqrt{2}d(cN)^{1/2}\delta_{\mathbf{a}}}{N/2} \\ &= 2\sqrt{2}(J-I)d\left(\frac{c}{N}\right)^{1/2}\delta_{\mathbf{a}}. \end{aligned} \quad (2.99)$$

Combining (2.99) with (2.97) yields

$$\left| \frac{(J-I)(b_0 - b_0^k)}{b_0^k + t(b_0 - b_0^k)} \right| \left| \frac{I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)} \right| \leq \frac{16(J-I)d}{I} c \leq 16d^2 c. \quad (2.100)$$

We have already established (2.96) in the $b_J > 2b_J^k$ case so, if we choose c sufficiently small relative to d , (2.100) completes the justification of (2.96) for all $t \in [0, 1]$, $b_J \geq 0$.

Finally we return to (2.84) with (2.96) and recall the positivity of the term $-\frac{\sigma J(b_I - b_I^k)}{b_I^k + t(b_I - b_I^k)}$ from (2.85) to compute for all $t \in [0, 1]$

$$\begin{aligned} \partial_t [F \circ \gamma](t) &= \frac{\sigma(J-I)(b_0 - b_0^k)}{b_0^k + t(b_0 - b_0^k)} - \frac{\sigma J(b_I - b_I^k)}{b_I^k + t(b_I - b_I^k)} + \frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)} \\ &\geq \frac{\sigma(J-I)(b_0 - b_0^k)}{b_0^k + t(b_0 - b_0^k)} + \frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)} \\ &\geq \frac{1}{2} \frac{\sigma I(b_J - b_J^k)}{b_J^k + t(b_J - b_J^k)} \\ &= \Omega \left(\left| \frac{b_J - b_J^k}{b_J^k + t(b_J - b_J^k)} \right| \right) \end{aligned} \quad (2.101)$$

We integrate (2.101) to bound the difference between $F(b_0, b_I, b_J)$ and $F(b_0^k, b_I^k, b_J^k)$:

$$\begin{aligned} F(b_0, b_I, b_J) - F(b_0^k, b_I^k, b_J^k) &\geq \Omega \left(\int_0^1 \left| \frac{b_J - b_J^k}{b_J^k + t(b_J - b_J^k)} \right| dt \right) \\ &= |\ln(b_J) - \ln(b_J^k)| \end{aligned} \quad (2.102)$$

We make use of (2.102) by once again splitting into the regimes around the cutoff criterion $b_J > 2b_J^k$. If $b_J \leq 2b_J^k$, we use the mean value theorem and if $b_J > 2b_J^k$ we use elementary log properties to obtain

$$F(b_0, b_I, b_J) - F(b_0^k, b_I^k, b_J^k) = \begin{cases} \Omega \left(\frac{|b_J - b_J^k|}{b_J^k} \right) & \text{if } b_J \leq 2b_J^k \\ \Omega(\ln[b_J/b_J^k]) = \Omega(1) & \text{if } b_J > 2b_J^k \end{cases} \quad (2.103)$$

We recall the bound $F(b_0^k, b_I^k, b_J^k) = O(1/b_J^k)$ from (2.81) and note that if C , and therefore $|b_J - b_J^k|$, is sufficiently large relative to the implicit constants in

(2.103), then $F(b_0^k, b_I^k, b_J^k)$ can be bounded by an arbitrarily small multiple of $F(b_0, b_I, b_J) - F(b_0^k, b_I^k, b_J^k)$. Thus, (2.103) implies

$$F(b_0, b_I, b_J) = \begin{cases} \Omega\left(\frac{|b_J - b_J^k|}{b_J^k}\right) & \text{if } b_J \leq 2b_J^k \\ \Omega(1) & \text{if } b_J > 2b_J^k \end{cases} \quad (2.104)$$

Recalling that F was built to approximate the log of

$$R(\mathbf{b}) = \frac{\mathbb{P}(\mathbb{T}_{\mathbf{b} + \sigma(I-J)\mathbf{e}_0 + \sigma J\mathbf{e}_I - \sigma I\mathbf{e}_J})}{\mathbb{P}(\mathbb{T}_{\mathbf{b}})},$$

the ratio of masses of neighboring types with respect to the

$$-\sigma(J - I)\mathbf{e}_0 + \sigma J\mathbf{e}_I - \sigma I\mathbf{e}_J$$

direction, we have now made substantial progress toward our goal of showing that from $\mathbf{a}$ to $\mathbf{b}$, the mass of types increases substantially in this direction. What remains to show before our final bound is that

1. $F(b_0, b_I, b_J)$ is a sufficiently good approximation of $(\ln \circ R)(\mathbf{b})$ and
2. R remains large between $\mathbf{b}$ and $\mathbf{a}$.

To prove (1), we note by (2.83) that $\left(\frac{1}{b_0} + \frac{1}{b_I} + \frac{1}{b_J}\right) = O\left(\frac{1}{b_J^k}\right)$ and directly compute

$$\begin{aligned} (\ln \circ R)(\mathbf{b}) &= \ln \left(\frac{\mathbb{P}(\mathbb{T}_{\mathbf{b} + \sigma(I-J)\mathbf{e}_0 + \sigma J\mathbf{e}_I - \sigma I\mathbf{e}_J})}{\mathbb{P}(\mathbb{T}_{\mathbf{b}})} \right) \\ &= \ln \left(p_0^{\sigma(I-J)} p_I^{\sigma J} p_J^{-\sigma I} \right) + \ln \left(\frac{\binom{N}{\mathbf{b} + \sigma(I-J)\mathbf{e}_0 + \sigma J\mathbf{e}_I - \sigma I\mathbf{e}_J}}{\binom{N}{\mathbf{b}}} \right) \\ &= \ln \left(p_0^{\sigma(I-J)} p_I^{\sigma J} p_J^{-\sigma I} \right) - \ln \left(b_0^{\sigma(I-J)} b_I^{\sigma J} b_J^{-\sigma I} \left[1 + O\left(\frac{1}{b_0} + \frac{1}{b_I} + \frac{1}{b_J}\right) \right] \right) \\ &= \ln \left(p_0^{\sigma(I-J)} p_I^{\sigma J} p_J^{-\sigma I} \right) - \ln \left(b_0^{\sigma(I-J)} b_I^{\sigma J} b_J^{-\sigma I} \left[1 + O\left(\frac{1}{b_J^k}\right) \right] \right) \\ &= F(b_0, b_I, b_J) + O\left(\frac{1}{b_J^k}\right) \end{aligned} \quad (2.105)$$

where the final line resulted from an application of the Taylor remainder theorem.

Together with (2.104), we have the useful lower bound

$$(\ln \circ R)(\mathbf{b}) \gtrsim \min \left(\frac{|b_J - b_J^k|}{b_J^k}, 1 \right) \quad (2.106)$$

and use the bound (2.89), i.e. $|b_J - b_J^k| \gtrsim |a_J - b_J^k|$, and the definition of $\delta_{\mathbf{a}}$ to improve (2.106) to

$$(\ln \circ R)(\mathbf{b}) \gtrsim \min \left(\frac{\delta_{\mathbf{a}}}{\sqrt{b_J^k}}, 1 \right). \quad (2.107)$$

Finally we return to (2), namely that R remains large between $\mathbf{b}$ and $\mathbf{a}$. Let $a \in \{0, \dots, b-1\}$ so the components of the parameter

$$\tilde{\mathbf{a}} := \mathbf{a} + a\sigma[(I-J)e_0 + Je_I - Ie_J]$$

lie between the respective components of $\mathbf{a}$ and $\mathbf{b}$. We perform to combinatorial computation

$$\begin{aligned} (\ln \circ R)(\tilde{\mathbf{a}}) - \ln \left(p_0^{\sigma(I-J)} p_I^{\sigma J} p_J^{-\sigma I} \right) = \\ \begin{cases} \sum_{j=1}^{J-I} \ln(\tilde{a}_0 - j) - \sum_{j=0}^{J-1} \ln(\tilde{a}_I + j) + \sum_{j=1}^I \ln(\tilde{a}_J - j) & \text{if } \sigma = 1 \\ - \sum_{j=0}^{J-I-1} \ln(\tilde{a}_0 + j) + \sum_{j=1}^J \ln(\tilde{a}_I - j) - \sum_{j=0}^{I-1} \ln(\tilde{a}_J + j) & \text{if } \sigma = -1. \end{cases} \end{aligned} \quad (2.108)$$

This expression conveniently extends to a smooth function on the path $\mathbf{a}(t) = \mathbf{a} + t(\mathbf{b} - \mathbf{a})$. We compute the derivative

$$\begin{aligned} \partial_t [(\ln \circ R)(\tilde{\mathbf{a}}(t))] &= [\nabla(\ln \circ R)](\tilde{\mathbf{a}}(t)) \cdot (\mathbf{b} - \mathbf{a}) \\ &= [\partial_0(\ln \circ R)](\tilde{\mathbf{a}}(t))(b_0 - a_0) + [\partial_I(\ln \circ R)](\tilde{\mathbf{a}}(t))(b_I - a_I) + \\ &\quad [\partial_J(\ln \circ R)](\tilde{\mathbf{a}}(t))(b_J - a_J) \end{aligned} \quad (2.109)$$

If $\sigma = 1$, then $\tilde{a}_I \leq b_I$ and by (2.83), $b_I \gtrsim b_J^k$ so

$$\begin{aligned} [\partial_I(\ln \circ R)](\tilde{\mathbf{a}}(t)) &= - \sum_{j=0}^{J-1} \frac{1}{\tilde{a}_I + j} \\ &\leq - \frac{J}{b_I + J} \leq -\Omega \left(\frac{1}{b_J^k} \right). \end{aligned} \quad (2.110)$$

Also $\sigma = 1$ implies $\tilde{a}_J \geq b_J^k \geq I$ so

$$[\partial_J(\ln \circ R)](\tilde{\mathbf{a}}(t)) = \sum_{j=1}^I \frac{1}{\tilde{a}_J - j} > 0 \quad (2.111)$$

as long as C is large enough.

In close analogy, if $\sigma = -1$ then $\tilde{a}_J \leq b_J$ and by (2.83), $b_J \gtrsim b_J^k$ so

$$\begin{aligned} [\partial_J(\ln \circ R)](\tilde{\mathbf{a}}(t)) &= -\sum_{j=0}^{I-1} \frac{1}{\tilde{a}_I - j} \\ &\leq -\frac{I}{b_J + I} \leq -\Omega\left(\frac{1}{b_J^k}\right). \end{aligned} \quad (2.112)$$

Also, $\tilde{a}_I \geq b_I \geq J$ so

$$[\partial_I(\ln \circ R)](\tilde{\mathbf{a}}(t)) = \sum_{j=0}^{J-1} \frac{1}{\tilde{a}_I - j} > 0 \quad (2.113)$$

as long as C is large enough. In particular these bounds imply that, regardless of σ ,

$$[\partial_I(\ln \circ R)](\tilde{\mathbf{a}}(t))(b_I - a_I), [\partial_J(\ln \circ R)](\tilde{\mathbf{a}}(t))(b_J - a_J) < 0 \quad (2.114)$$

Because $\tilde{a}_0(t)$ lies between $\mathbf{a}$ and $\mathbf{b}$ for all $t \in [0, 1]$, we have $\tilde{a}_0(t) + O(1) = \Omega(N)$ so, regardless of σ ,

$$[\partial_0(\ln \circ R)](\tilde{\mathbf{a}}(t)) = O\left(\frac{1}{N}\right). \quad (2.115)$$

By the construction of $\mathbf{b}$ and (2.75), we have

$$|a_I - b_I| \gtrsim \min\left(|a_I - b_I^k|, |a_J - b_J^k|\right) \gtrsim |a_J - b_J^k|$$

The bounds (2.110) and (2.111) if $\sigma = 1$ or (2.112) and (2.113) if $\sigma = -1$ in conjunction with (2.114) then imply

$$\begin{aligned} &[\partial_I(\ln \circ R)](\tilde{\mathbf{a}}(t))(b_I - a_I) + [\partial_J(\ln \circ R)](\tilde{\mathbf{a}}(t))(b_J - a_J) \\ &= -\Omega\left(\frac{a_J - b_J^k}{b_J^k}\right) = -\Omega\left(\frac{\delta_{\mathbf{a}}}{\sqrt{b_J^k}}\right) \end{aligned} \quad (2.116)$$

and, combining (2.115) with the bounds (2.116), (2.109) provides the estimate

$$\partial_t[(\ln \circ R)(\tilde{\mathbf{a}}(t))] = O\left(\frac{|b_0 - a_0|}{N}\right) - \Omega\left(\frac{\delta_{\mathbf{a}}}{\sqrt{b_J^k}}\right) \quad (2.117)$$

Following the same logic as (2.98), we have

$$\begin{aligned}
|b_0 - a_0| &\leq \sum_{j=1}^d |b_j - a_j| \leq \sum_{j=1}^d |a_j - b_j^k| \\
&\leq \sum_{j=1}^d |b_j^k|^{1/2} \delta_{\mathbf{a}} \lesssim (cN)^{1/2} \delta_{\mathbf{a}}
\end{aligned} \tag{2.118}$$

Recalling that $b_j^k \lesssim k \leq cN$, (2.118) and (2.117) allow us to conclude

$$\partial_t [(\ln \circ R)(\tilde{\mathbf{a}}(t))] = \delta_{\mathbf{a}} \left[O\left(\frac{c}{N^{1/2}}\right) - \Omega\left(\frac{1}{cN^{1/2}}\right) \right]. \tag{2.119}$$

Therefore, as long as c is chosen sufficiently small,

$$\partial_t [(\ln \circ R)(\tilde{\mathbf{a}}(t))] < 0 \tag{2.120}$$

for all $t \in [0, 1]$.

Then, (2.120) proves that

$$(\ln \circ R)(\tilde{\mathbf{a}}(t)) \geq (\ln \circ R)(\tilde{\mathbf{a}}(1)) \geq (\ln \circ R)(\mathbf{b}) \tag{2.121}$$

for all $t \in [0, 1]$. In light of (2.107), i.e. $(\ln \circ R)(\mathbf{b}) \gtrsim \min\left(\frac{\delta_{\mathbf{a}}}{(b_j^k)^{1/2}}, 1\right)$, (2.121) implies for all $a \in \{0, \dots, b-1\}$, $\tilde{\mathbf{a}} := \mathbf{a} + a\sigma[(I-J)e_0 + Je_I - Ie_J]$,

$$(\ln \circ R)(\tilde{\mathbf{a}}) \gtrsim \min\left(\frac{\delta_{\mathbf{a}}}{(b_j^k)^{1/2}}, 1\right). \tag{2.122}$$

Finally we have the necessary ingredients for our overall goal. We compute

$$\begin{aligned}
L(\mathbf{b}) - L(\mathbf{a}) &= \sum_{a=0}^{b-1} \left[\ln \left(\mathbf{a} + (a+1)\sigma[(I-J)e_0 + Je_I - Ie_J] \right) - \right. \\
&\quad \left. \ln \left(\mathbf{a} + a\sigma[(I-J)e_0 + Je_I - Ie_J] \right) \right] \\
&= \sum_{a=0}^{b-1} \left(\ln \circ R \right) \left(\mathbf{a} + a\sigma[(I-J)e_0 + Je_I - Ie_J] \right) \\
&\gtrsim b \min\left(\frac{\delta_{\mathbf{a}}}{(b_j^k)^{1/2}}, 1\right) \gtrsim \min(\delta_{\mathbf{a}}^2, |a_J - b_j^k|) \geq \delta_{\mathbf{a}}.
\end{aligned} \tag{2.123}$$

Exponentiating (2.123) and rearranging then yields

$$L(\mathbf{a}) \lesssim L(\mathbf{b}) e^{-\Omega(\delta_{\mathbf{a}})} \leq e^{-\Omega(\delta_{\mathbf{a}})} \left(\max_{\mathbf{b} \in \mathbb{N}^d} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}}) \right),$$

and proves the lemma. $\square$

Proposition 2.3.2 follows directly from the last few lemmas.

Proof of Proposition 2.3.2. Let C and c be the constants from Lemma (2.3.6). If $\mathbf{a}$ does not satisfy the hypotheses of (2.3.6), then it must satisfy those of (2.3.4) with the constant c or (2.3.5) with the constant C . $\square$

Corollary 2.3.7. *If $\mathbb{T}_{\mathbf{a}}$ is a maximal probability type in $\mathbb{S}_k$, then $\delta_{\mathbf{a}} \lesssim 1$.*

Proof. By Proposition 2.3.2,

$$\begin{aligned} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) &\lesssim e^{-\Omega(\delta_{\mathbf{a}})} \left(\max_{\mathbf{b} \in \mathbb{N}^d} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}}) \right) = e^{-\Omega(\delta_{\mathbf{a}})} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \\ \implies e^{-\Omega(\delta_{\mathbf{a}})} &\gtrsim 1 \\ \implies \delta_{\mathbf{a}} &\lesssim 1. \end{aligned}$$

$\square$

Definition 2.3.8 (n^{th} Deviation in $\mathbb{S}_k$). *For any $n \in \mathbb{N}$, we denote and define the n^{th} deviation in $\mathbb{S}_k$ to be the event*

$$\text{Dev}_k(n) = \bigcup_{\delta_{\mathbf{a}} \in [n, n+1)} \mathbb{T}_{\mathbf{a}}$$

and reserve the notation

$$\text{Dev}_k(\leq n) = \bigcup_{m \leq n} \text{Dev}_k(m)$$

Lemma 2.3.9. *For any $k \in [0, \mathbb{E}(D)]$, $n \in \mathbb{N}$,*

$$|\{\mathbf{a} \in \mathbb{N}^{d+1} \cap \mathcal{D}^k : \delta_{\mathbf{a}} \in [n, n+1)\}| \lesssim n^{1-d} \left(\prod_{j=2}^d \max(1, |b_j^k|^{1/2}) \right)$$

Proof. To avoid dealing with combinatorial hassles caused by the boundary of $\mathcal{D}^k$ and the deviation restriction on the index 1 component, we will overestimate the size of a given set in the lemma statement by allowing $\mathbf{a}$ to have arbitrary integer components. More concretely, $\{\mathbf{a} \in \mathbb{N}^{d+1} \cap \mathcal{D}^k : \delta_{\mathbf{a}} \in [n, n+1)\}$ is a subset of

$$\begin{aligned} \widetilde{\text{Dev}}_k(n) &:= \\ \left\{ \mathbf{a} \in \mathbb{Z}^{d+1} : \Sigma \mathbf{a} = N, \mathbf{a} \cdot \mathbf{d} = k, \frac{|a_j - b_j^k|}{\max(1, |b_j^k|^{1/2})} \in [n, n+1) \text{ for all } j \in \{2, \dots, d\} \right\} \end{aligned}$$

so the lemma will follow from the bound

$$\left| \widetilde{\text{Dev}}_k(n) \right| \lesssim n^{1-d} \left(\prod_{j=2}^d \max \left(1, |b_j^k|^{1/2} \right) \right) \quad (2.124)$$

Using the notation

$$\bigcup_{m \leq n} \widetilde{\text{Dev}}_k(m),$$

We view $\text{Dev}_k(n)$ as the difference of $\widetilde{\text{Dev}}_k(\leq n)$ and $\widetilde{\text{Dev}} \leq k(\leq n-1)$, i.e.

$$\begin{aligned} \widetilde{\text{Dev}}_k(n) &= \widetilde{\text{Dev}}_k(\leq n) \setminus \widetilde{\text{Dev}} \leq k(\leq n-1) \\ \implies \left| \widetilde{\text{Dev}}_k(n) \right| &= \left| \widetilde{\text{Dev}}_k(\leq n) \right| - \left| \widetilde{\text{Dev}} \leq k(\leq n-1) \right| \end{aligned} \quad (2.125)$$

We can parameterize the set $\widetilde{\text{Dev}}_k(\leq n)$ by

$$(a_2, \dots, a_d) \mapsto \left(N - k + \sum_{j=2}^d (j-1)a_j, k - \sum_{j=2}^d ja_j, a_2, \dots, a_d \right)$$

where $\frac{|a_j - b_j^k|}{\max(1, |b_j^k|^{1/2})} \in (-n-1, n+1)$ for all $j \in \{2, \dots, d\}$. This provides the identity

$$\left| \widetilde{\text{Dev}}_k(\leq n) \right| = \prod_{j=2}^d \left| \mathbb{Z} \cap \left[\left(-n-1, n+1 \right) \cup \left(-(n+1)|b_j^k|^{1/2}, (n+1)|b_j^k|^{1/2} \right) \right] \right| \quad (2.126)$$

We note that for $r \in \mathbb{R}_+$,

$$\left| \mathbb{Z} \cap (-r-1, r+1) \right| = 2r + O(1)$$

and use (2.126) to compute for any $n \geq 1$

$$\left| \widetilde{\text{Dev}}_k(\leq n) \right| = \left(\prod_{\substack{2 \leq j \leq d \\ |b_j^k| > 1}} \left[2(n+1)|b_j^k|^{1/2} + O(1) \right] \right) \left(\prod_{\substack{2 \leq j \leq d \\ |b_j^k| \leq 1}} \left[2(n+1) + O(1) \right] \right) \quad (2.127)$$

Thus we have

$$\begin{aligned}
|\widetilde{\text{Dev}}_k(n)| &= |\widetilde{\text{Dev}}_k(\leq n)| - |\widetilde{\text{Dev}} \leq k(\leq n-1)| \\
&= \left(\prod_{\substack{2 \leq j \leq d \\ |b_j^k| > 1}} \left[2(n+1) |b_j^k|^{1/2} + O(1) \right] \right) \left(\prod_{\substack{2 \leq j \leq d \\ |b_j^k| \leq 1}} \left[2(n+1) + O(1) \right] \right) - \\
&\quad \left(\prod_{\substack{2 \leq j \leq d \\ |b_j^k| > 1}} \left[2n |b_j^k|^{1/2} + O(1) \right] \right) \left(\prod_{\substack{2 \leq j \leq d \\ |b_j^k| \leq 1}} \left[2n + O(1) \right] \right) \\
&= O \left(n^{1-d} \left[\prod_{j=2}^d \max(1, |b_j^k|^{1/2}) \right] \right) \tag{2.128}
\end{aligned}$$

by Taylor expansion. $\square$

Lemma 2.3.10. *If $\mathbb{T}_{\mathbf{a}}$ is a maximal probability type in $\mathbb{S}_k$, then*

$$\mathbb{P}(\mathbb{T}_{\mathbf{a}}) \sim \left(\prod_{j=2}^d \max(1, |b_j^k|^{1/2}) \right)^{-1}.$$

Proof. By Proposition 2.2.8 there exists a large $n > 1$ such that for any $j \in \{1, \dots, d\}$,

$$k^j \geq nN^{j-1} \implies b_j^k \geq \frac{k^j}{nN^{j-1}}. \tag{2.129}$$

Let $m > 1$ be a large constant to be determined later and $J \in \{1, \dots, d\}$ be the maximal index such that $k^J \geq nN^{J-1}$ and $b_J^k \leq m$. Note that if $k \leq cN$ for a sufficiently small constant $c > 0$, we can ensure that for all $j \leq J$, $b_j^k \geq b_J^k$ as in the proof of (2.3.6). If $k > cN$, then for all $j \in \{1, \dots, J\}$, $b_j^k \geq \frac{k^j}{nN^{j-1}} \geq \frac{c^j}{n} N$. In either case we are free to assume that for all $j \leq J$, $b_j^k \geq m$ if N is sufficiently large.

Notice then that for all $j > J$, $b_j^k \lesssim 1$ so the lemma is equivalent to

$$\mathbb{P}(\mathbb{T}_{\mathbf{a}}) \sim \prod_{j=2}^J |b_j^k|^{-1/2}. \tag{2.130}$$

We reserve the notation

$$\mathbf{f}_j = (j-1)\mathbf{e}_0 - j\mathbf{e}_1 + \mathbf{e}_j$$

for all $j \in \{2, \dots, J\}$. Our goal is to show that for any such j , we can find $\prod_{j=2}^J |b_j^k|^{1/2}$ distinct types of equivalent in probability to $\mathbb{P}_{\mathbf{s}_k}(\mathbb{T}_{\mathbf{a}})$ up to a constant factor, from which (2.130) and therefore the lemma as a whole would follow. Note that $\mathbf{b}^k$ could have components close to -1 for $j \in \{J+1, \dots, d\}$ so we must be cautious in our handling of j in this range because our asymptotics on Λ are no longer reliable.

Let $\mathbf{b} \in \mathcal{D}^k$ be such that for all $j \in \{2, \dots, J\}$, $|b_j - b_j^k| \leq 2\delta_{\mathbf{a}} |b_j^k|^{1/2}$ and for all $j > J$, $b_j \leq m$. Because $(\mathbf{b} - \mathbf{b}^k) \cdot \mathbf{d} = 0$, this immediately implies $|b_1 - b_1^k| \lesssim \max_{j \in \{2, \dots, d\}} |b_j^k|^{1/2} \lesssim k^{1/2}$ and, because $\Sigma(\mathbf{b} - \mathbf{b}^k) = 0$, we have the additional bound $|b_0 - b_0^k| \lesssim k^{1/2}$. In particular, $b_1 \sim b_1^k$, $b_0 \sim b_0^k$, and, if m is chosen sufficiently large relative to $\delta_{\mathbf{a}} \lesssim 1$, $b_J \sim b_J^k$.

For any $j \in \{2, \dots, J\}$ we use (2.2.6) and the estimate Claim 2.2.4, item (4) to compute

$$\begin{aligned} & \mathbf{f}_j \cdot [\nabla L(\mathbf{b}) - \nabla L(\mathbf{b}^k)] \\ &= \int_0^1 \mathbf{f}_j \cdot \mathcal{H}L(\mathbf{b}^k + t(\mathbf{b} - \mathbf{b}^k))(\mathbf{b} - \mathbf{b}^k) dt \\ &= \int_0^1 \sum_{i=0}^d -\Lambda''[b_i^k + t(b_i - b_i^k) + 1] f_i^j(b_i - b_i^k) dt \\ &= -(j-1)(b_0 - b_0^k) \int_0^1 \Lambda''[b_0^k + t(b_0 - b_0^k) + 1] dt + \\ & \quad j(b_1 - b_1^k) \int_0^1 \Lambda''[b_1^k + t(b_1 - b_1^k) + 1] dt + \\ & \quad - (b_j - b_j^k) \int_0^1 \Lambda''[b_j^k + t(b_j - b_j^k) + 1] dt \\ &= O\left(\left|\frac{b_0 - b_0^k}{\min(b_0, b_0^k)}\right| + \left|\frac{b_1 - b_1^k}{\min(b_1, b_1^k)}\right| + \left|\frac{b_j - b_j^k}{\min(b_j, b_j^k)}\right|\right) \\ &= O\left(\frac{\sqrt{k}}{N} + \frac{\sqrt{k}}{k} + \frac{|b_j^k|^{1/2}}{|b_j^k|}\right) \end{aligned}$$

$$= O\left(|b_j^k|^{-1/2}\right). \quad (2.131)$$

Notice that because $L(\mathbf{b}^k)$ is a critical point of $L(\mathbf{a})$ on $\mathcal{D}^k$, $\nabla L(\mathbf{b}^k) \cdot \mathbf{f}_j = 0$. Then (2.131) implies

$$\nabla L(\mathbf{b}) = O\left(|b_j^k|^{-1/2}\right) \quad (2.132)$$

For all $j \in \{1, \dots, J\}$, $|a_j - b_j^k| \leq \delta_{\mathbf{a}} |b_j^k|^{1/2}$ for all $j \in \{1, \dots, d\}$ and for all $j > J$, $a_j < m$ as long as m is sufficiently large relative to $\delta_{\mathbf{a}}$. Now let $\mathbf{b} \in \mathcal{D}^k$ be of the form

$$\mathbf{b} = \mathbf{b}(c_2, \dots, c_J) = \mathbf{a} + \sum_{j=2}^J c_j \mathbf{f}_j \quad (2.133)$$

for some $c_j \in \mathbb{Z}$ such that $|c_j| \leq \delta_{\mathbf{a}} |b_j^k|^{1/2}$ and notice that for any $t \in [0, 1]$, $\gamma(t) := \mathbf{a} + t(\mathbf{b} - \mathbf{a})$ satisfies the hypotheses of (2.131) (and therefore (2.132)).

Then we apply (2.132) to get the bound

$$\begin{aligned} L(\mathbf{b}) - L(\mathbf{a}) &= \int_0^1 \partial_t [L \circ \gamma](t) dt \\ &= \int_0^1 (\nabla L)(\gamma(t)) \cdot (\mathbf{b} - \mathbf{a}) dt \\ &= \sum_{j=2}^J \left(\int_0^1 c_j (\nabla L)(\gamma(t)) \cdot \mathbf{f}_j dt \right) \\ &= \sum_{j=2}^J O\left(c_j |b_j^k|^{-1/2}\right) \\ &= O(1). \end{aligned} \quad (2.134)$$

We denote by B the corresponding Cartesian product, that is,

$$B := \prod_{j=2}^J \left(\mathbb{Z} \cap \left[-\delta_{\mathbf{a}} |b_j^k|^{1/2}, \delta_{\mathbf{a}} |b_j^k|^{1/2} \right] \right)$$

and observe that

$$|B| = \prod_{j=2}^J \left(\delta_{\mathbf{a}} |b_j^k|^{1/2} \right) \sim \prod_{j=2}^J \left(|b_j^k|^{1/2} \right)$$

Applying the trivial bound to the probability of the union of all such types and calling on (2.134) then yields

$$\begin{aligned}
1 &= \mathbb{P}_{\mathbb{S}_k} \left(\left\{ \mathbb{T}_{\mathbf{b}(c_2, \dots, c_d)} : c_j \in \mathbb{Z} \cap \left[-\delta_{\mathbf{a}} |b_j^k|^{1/2}, \delta_{\mathbf{a}} |b_j^k|^{1/2} \right] \text{ for } j \in \{2, \dots, J\} \right\} \right) \\
&= \sum_{\mathbf{c} \in B} \mathbb{P}_{\mathbb{S}_k} (\mathbb{T}_{\mathbf{b}(c_2, \dots, c_d)}) \\
&= \sum_{\mathbf{c} \in B} \Theta (\mathbb{P}_{\mathbb{S}_k} (\mathbb{T}_{\mathbf{a}})) \\
&= \Theta(|B| \mathbb{P}_{\mathbb{S}_k} (\mathbb{T}_{\mathbf{a}})) \\
\Rightarrow \mathbb{P}_{\mathbb{S}_k} (\mathbb{T}_{\mathbf{a}}) &\lesssim \prod_{j=2}^J \left(|b_j^k|^{-1/2} \right) \tag{2.135}
\end{aligned}$$

□

Corollary 2.3.11. *For any $k \in [0, \mathbb{E}(D)]$*

$$\mathbb{P}_{\mathbb{S}_k} (\text{Dev}_k(n)) \lesssim e^{-\Omega(n)}.$$

Proof. We simply use Lemmas 2.3.10 and 2.3.9 with Proposition 2.3.2 to obtain

$$\begin{aligned}
\mathbb{P}_{\mathbb{S}_k} (\text{Dev}_k(n)) &= \bigcup_{\delta_{\mathbf{a}} \in [n, n+1)} \mathbb{P}_{\mathbb{S}_k} (\mathbb{T}_{\mathbf{a}}) \\
&= \sum_{\delta_{\mathbf{a}} \in [n, n+1)} \mathbb{P}_{\mathbb{S}_k} (\mathbb{T}_{\mathbf{a}}) \\
&\lesssim \sum_{\delta_{\mathbf{a}} \in [n, n+1)} e^{-\Omega(\delta_{\mathbf{a}})} \left(\max_{\mathbf{b} \in \mathbb{N}^d} \mathbb{P}_{\mathbb{S}_k} (\mathbb{T}_{\mathbf{b}}) \right) \\
&\leq \left| \{ \mathbf{a} \in \mathbb{N}^{d+1} \cap \mathcal{D}^k : \delta_{\mathbf{a}} \in [n, n+1) \} \right| e^{-\Omega(\delta_{\mathbf{a}})} \left(\max_{\mathbf{b} \in \mathbb{N}^d} \mathbb{P}_{\mathbb{S}_k} (\mathbb{T}_{\mathbf{b}}) \right) \\
&\lesssim n^{1-d} \left(\prod_{j=2}^d \max \left(1, |b_j^k|^{1/2} \right) \right) e^{-\Omega(\delta_{\mathbf{a}})} \left(\max_{\mathbf{b} \in \mathbb{N}^d} \mathbb{P}_{\mathbb{S}_k} (\mathbb{T}_{\mathbf{b}}) \right) \\
&\lesssim n^{1-d} \left(\prod_{j=2}^d \max \left(1, |b_j^k|^{1/2} \right) \right) e^{-\Omega(n)} \left(\prod_{j=2}^d \max \left(1, |b_j^k|^{1/2} \right) \right)^{-1} \\
&= n^{d-1} e^{-\Omega(n)} \\
&= e^{-\Omega(n)}
\end{aligned}$$

completing the proof. □

Corollary 2.3.12. *For any $k \in [0, \mathbb{E}(D)]$, there exists some constant n such that*

$$\mathbb{P}_{\mathbb{S}_k}(\text{Dev}_k(\leq n)) > \frac{1}{2}.$$

Proof. Let $C > 1$ be a (large) constant such that $\mathbb{P}_{\mathbb{S}_k}(\text{Dev}_k(n)) \leq Ce^{-n/C}$, whose existence follows from Corollary 2.3.11. If we pick n so large that $\sum_{m>n} Ce^{-m/C} < \frac{1}{2}$. The result follows immediately. $\square$

2.4 The Smooth Maximal Operator

We begin by introducing a smoother variant of our spherical maximal operator.

Definition 2.4.1 (Smooth Maximal Operator). *In G^N , the smooth maximal operator is defined by*

$$M_S f(x) = \max_{0 \leq K \leq \mathbb{E}(D)} \frac{1}{K+1} \left| \sum_{k=0}^K S_k f(x) \right|$$

The goal of this section and §2.5 is to bound M_S in L^2 independent of dimension, however we will prove it for all $p > 1$ because it is no more difficult. The remainder of this chapter is dedicated to bootstrapping a bound on M from the bound on M_S . As in [8] and [7], the strategy for bounding M_S is to dominate the family of Cesàro means $\left(\frac{1}{K+1} \sum_{k=0}^K S_k \right)$ by appropriate means of noise averages. This allows us to take advantage of the powerful operator semigroup tools of [10]. However, the complexity of spheres in G^N relative to K_m^N forces us to examine multi-parameter noise operators.

Definition 2.4.2 (Noise Operator). *Let $\mathbf{q} \in (0, 1)^{d+1}$ be a stochastic vector (as with $\mathbf{p}$, all components are strictly positive). The noise operator with parameter $\mathbf{q}$ is the operator on $L^1(G^N)$ pointwise by*

$$\mathcal{N}_{\mathbf{q}}^N f(x) := \mathbb{E} \left(\left[\bigotimes_{n=1}^N P_{Q_n} \right] f(x) \right).$$

where $\{Q_1, \dots, Q_N\}$ is a sequence of i.i.d. random variables with the distribution

$$\mathbb{P}(Q_n = j) = q_j \text{ for } j \in \{0, \dots, d\}.$$

In keeping with our prior conventions, we generally suppress the dimension and write $\mathcal{N}_{\mathbf{q}}$ in place of $\mathcal{N}_{\mathbf{q}}^N$. As with the distance variable D , we define the noise distance variable

$$Q := \sum_{n=1}^N Q_n.$$

Lemma 2.4.3. *The operator $\mathcal{N}_{\mathbf{q}}$ is a radial operator (linear combination of spherical samples) and, in particular, decomposes into*

$$\mathcal{N}_{\mathbf{q}} = \sum_{k=0}^{dN} \mathbb{P}(Q = k) S_k$$

if and only if $\mathbf{q}$ satisfies the following algebraic relations for each $j \in \{2, \dots, d\}$:

$$\frac{q_j q_0^{j-1}}{q_1^j} = \frac{p_j p_0^{j-1}}{p_1^j} \quad (2.136)$$

Proof. We decompose $\mathcal{N}_{\mathbf{q}}$ radially, i.e.

$$\begin{aligned} \mathcal{N}_{\mathbf{q}} f(x) &= \mathbb{E} \left(\left[\otimes_{n=1}^N P_{Q_n} \right] f(x) \right) \\ &= \sum_{k=0}^{dN} \mathbb{P}(Q = k) \left[\mathbb{E} \left(\left[\otimes_{n=1}^N P_{(Q_n|Q=k)} \right] f(x) \right) \right] \end{aligned}$$

Thus, the lemma will follow directly if we prove that the $\{0, \dots, d\}^N$ valued random variable $(Q_1, \dots, Q_N | Q = k)$ is identically distributed to $(D_1, \dots, D_N | D = k)$. Note that, for any $\mathbf{a} = (a_0, \dots, a_d) \in \mathbb{N}^{d+1}$ such that $\Sigma \mathbf{a} = N$, by the componentwise symmetry of the noise operator, the event $\{(Q^0, \dots, Q^d) = \mathbf{a}\}$ gives equal weight to each $\{0, \dots, d\}^N$ -valued outcome for $(Q_1, \dots, Q_N)$ that it contains. Similarly, by the componentwise symmetry of the spherical samples, the event $\{(D^0, \dots, D^d) = \mathbf{a}\}$ (that is, $\mathbb{T}_{\mathbf{a}}$) gives equal weight to each $\{0, \dots, d\}^N$ -valued outcome for $(D_1, \dots, D_N)$ that it contains. Thus the lemma will follow if the identity

$$\mathbb{P}_{\{Q=\mathbf{a}, \mathbf{d}\}} \left(\{(Q^0, \dots, Q^d) = \mathbf{a}\} \right) = \mathbb{P}_{\mathbb{S}_{\mathbf{a}, \mathbf{d}}}(\mathbb{T}_{\mathbf{a}}), \quad (2.137)$$

is satisfied for all types $\mathbb{T}_{\mathbf{a}}$ if and only if $\mathbf{q}$ satisfies (2.136). Moreover, direct computation reveals as in (2.11)

$$\mathbb{P}\left(\{(Q^0, \dots, Q^d) = \mathbf{a}\}\right) = \binom{N}{\mathbf{a}} \mathbf{q}^{\mathbf{a}}. \quad (2.138)$$

We then apply Bayes' law to (2.138) and (2.11) in (2.137) and reduce the lemma to showing that for all $k \leq dN$, $\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k$, the identity

$$\begin{aligned} \frac{\binom{N}{\mathbf{a}} \mathbf{q}^{\mathbf{a}}}{\mathbb{P}(\{Q = k\})} &= \frac{\binom{N}{\mathbf{a}} \mathbf{p}^{\mathbf{a}}}{\mathbb{P}(\{D = k\})} \\ \iff \left(\frac{\mathbf{q}}{\mathbf{p}}\right)^{\mathbf{a}} &= \frac{\mathbb{P}(\{Q = k\})}{\mathbb{P}(\{D = k\})} \end{aligned} \quad (2.139)$$

is equivalent to (2.136). Note that, because

$$\begin{aligned} \mathbb{P}(\{Q = k\}) &= \sum_{\mathbf{a} \in \mathbb{N}^{d+1} \cap \mathcal{D}^k} \binom{N}{\mathbf{a}} \mathbf{q}^{\mathbf{a}} \text{ and} \\ \mathbb{P}(\{D = k\}) &= \sum_{\mathbf{a} \in \mathbb{N}^{d+1} \cap \mathcal{D}^k} \binom{N}{\mathbf{a}} \mathbf{p}^{\mathbf{a}}, \end{aligned}$$

if the ratio $(\mathbf{q}/\mathbf{p})^{\mathbf{a}}$ is constant on $\mathbf{a} \in \mathbb{N}^{d+1} \cap \mathcal{D}^k$, it must equal $\frac{\mathbb{P}(\{Q=k\})}{\mathbb{P}(\{D=k\})}$.

It is straightforward to verify that

$$\begin{aligned} &\{\mathbf{a} \in \mathbb{N}^{d+1} : \mathbf{a} \in \mathcal{D}^k\} \\ &= \left\{ \mathbf{a} = (N-k)\mathbf{e}_0 + k\mathbf{e}_1 + \sum_{j=2}^d b_j [\mathbf{e}_j + (j-1)\mathbf{e}_0 - j\mathbf{e}_1] \in \mathbb{N}^{d+1} : (b_2, \dots, b_d) \in \mathbb{N}^{d-1} \right\} \end{aligned}$$

so the ratio $(\mathbf{q}/\mathbf{p})^{\mathbf{a}}$ is constant, and thus equal to $\frac{\mathbb{P}(\{Q=k\})}{\mathbb{P}(\{D=k\})}$, if and only if for all $\mathbf{a} \in \mathbb{N}^{d+1} \cap \mathcal{D}^k, j \in \{2, \dots, d\}$,

$$\begin{aligned} \left(\frac{\mathbf{q}}{\mathbf{p}}\right)^{\mathbf{a}} &= \left(\frac{\mathbf{q}}{\mathbf{p}}\right)^{(\mathbf{a} + \mathbf{e}_j + (j-1)\mathbf{e}_0 - j\mathbf{e}_1)} \\ \iff 1 &= \left(\frac{\mathbf{q}}{\mathbf{p}}\right)^{(\mathbf{e}_j + (j-1)\mathbf{e}_0 - j\mathbf{e}_1)} \\ \iff \frac{q_j q_0^{j-1}}{q_1^j} &= \frac{p_j p_0^{j-1}}{p_1^j}. \end{aligned}$$

as claimed. □

We will now use the implicit function theorem to parameterize the curve induced by the equations (2.136) in terms of p_1 maintaining the constraint $\Sigma \mathbf{p} = 1$ in a neighborhood of $\mathbf{p} = \mathbf{e}_0$. Moreover, we will establish the properties of the curve γ that are necessary to bound M_S .

Lemma 2.4.4. *There exists some $\epsilon \in (0, 1)$ and a curve $\gamma : [0, \epsilon] \rightarrow [0, 1]^{d+1}$ given by*

$$\gamma(t) = \begin{bmatrix} q_0(t) \\ q_1(t) \\ \vdots \\ q_d(t) \end{bmatrix}$$

such that the following conditions hold:

- (1) *For all $t \in (0, \epsilon]$, $j \in \{2, \dots, d\}$, $\mathbf{q} = \gamma(t)$ satisfies the equations (2.136),*
- (2) *$\mathbf{q} \cdot \mathbf{d} = t$,*
- (3) *$\Sigma \mathbf{q} = 1$,*
- (4) *$\gamma(t)$ is real analytic on (an open interval containing) its domain,*
- (5) *for all $0 \leq i < j \in \{1, \dots, d\}$, $p_j^{(i)}(0) = 0$, and*
- (6) *$q_j^{(j)}(0) = \frac{p_j p_0^{j-1}}{p_1^j} j!$*

Proof. Temporarily allowing $\{q_0, \dots, q_d\}$ to take arbitrary real values, we fix $t \in \mathbb{R}$

to define the map $F_t(\cdot) : \mathbb{R}^{d+1} \rightarrow \mathbb{R}^{d+1}$ by

$$\begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ \vdots \\ q_j \\ \vdots \\ q_d \end{bmatrix} \mapsto \begin{bmatrix} \sum_{j=0}^d q_j \\ \sum_{j=1}^d j q_j - t \\ q_2 q_0 - \left(\frac{p_2 p_0}{p_1^2} \right) q_1^2 \\ \vdots \\ q_j q_0^{j-1} - \left(\frac{p_j p_0^{j-1}}{p_1^j} \right) q_1^j \\ \vdots \\ q_d q_0^{d-1} - \left(\frac{p_d p_0^{d-1}}{p_1^d} \right) q_1^d \end{bmatrix}$$

and observe that $F_0(\mathbf{e}_0) = (\mathbf{e}_0)$. To directly compute the first derivatives of $F_0(\cdot)$ at $(1, 0, \dots, 0)$ to find the Jacobian,

$$\mathcal{J}F_0(\mathbf{e}_0) = \begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 2 & 3 & \dots & d \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

We note that F is analytic in all of $q_0, q_1, \dots, q_d$, and t and that F_0 has an invertible Jacobian at $(1, 0, \dots, 0)$. Thus, by the analytic implicit function theorem there exists some $\epsilon > 0$ and analytic functions $\{q_0(t), q_1(t), \dots, q_d(t)\}$ defined on the domain $(-\epsilon, 2\epsilon)$ such that for all $t \in (-\epsilon, 2\epsilon)$, $F_t(q_0(t), q_1(t), \dots, q_d(t)) = (1, 0, \dots, 0)$. Throughout this chapter, we will reduce ϵ as needed to match our claims but we will ensure it remains positive and we will reduce it only finitely many times. Defining $\gamma(t)$ as in the lemma statement on the domain $[0, \epsilon]$ immediately justifies conditions (1), (2), (3), and (4). Because each component of the vector-valued function $F_t \circ \gamma(t)$ is constant, differentiating each component in t yields the system

$$0 = \sum_{j=0}^d q'_j \tag{2.140}$$

$$0 = \sum_{j=1}^d j q'_j - 1 \quad (2.141)$$

$$0 = q'_j q_0^{j-1} + (j-1) q_j q_0^{j-2} q'_0 - j \left(\frac{p_j p_0^{j-1}}{p_1^j} \right) q_1^{j-1} q'_1 \text{ for all } j \in \{2, \dots, d\}. \quad (2.142)$$

Because $q_j(0) = 0$ for all $j \in \{1, \dots, d\}$ and $q_j(0) = 1$, evaluating (2.142) at 0 yields

$$\begin{aligned} 0 &= q'_j q_0^{j-1} + (j-1) q_j q_0^{j-2} q'_0 - j \left(\frac{p_j p_0^{j-1}}{p_1^j} \right) q_1^{j-1} q'_1 \\ \implies 0 &= q'_j(0). \end{aligned}$$

for all $j \in \{2, \dots, d\}$. Then, solving (2.141) and (2.140) proves

$$\gamma'(t) = \begin{bmatrix} -1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (2.143)$$

Because $q_0(0) = 1$, by continuity we can assume $q_0(t) \neq 0$ on $[0, \epsilon]$. In particular, for all $j \in \{2, \dots, d\}$, $q_0(t)^{1-j}$ is analytic with $q_0(0)^{1-j} = 1$. Calling on the fact $q_1(0) = 0$ and $q'_1(0) = 1$ from (2.143), we now have

$$q_0(t)^{1-j} = 1 + O_{an}(t) \text{ and}$$

$$q_1(t) = t + O_{an}(t^2)$$

Then we can rewrite the equation

$$q_j q_0^{j-1} = \left(\frac{p_j p_0^{j-1}}{p_1^j} \right) q_1^j,$$

from the implicit function theorem to obtain

$$\begin{aligned} q_j(t) &= \left(\frac{p_j p_0^{j-1}}{p_1^j} \right) q_1(t)^j q_0(t)^{1-j} \\ &= \left(\frac{p_j p_0^{j-1}}{p_1^j} \right) [t + O_{an}(t^2)]^j [1 + O_{an}(t)]^{1-j} \\ &= \left(\frac{p_j p_0^{j-1}}{p_1^j} \right) t^j + O_{an}(t^{j+1}). \end{aligned} \quad (2.144)$$

Conditions (5) and (6), which have already been proven for $j = 1$, immediately follow for $j \in \{2, \dots, d\}$ by differentiating (2.144) i times for any $i \in \{0, \dots, j\}$. The only technicality that remains is to show that γ maps $[0, \epsilon]$ into the unit cube $[0, 1]^{d+1}$. Notice that for all $j \in \{1, \dots, d\}$, the lowest order term in (2.144) is positive and therefore $q_j(t)$ increases on $[0, \epsilon]$. Thus we have the bound $q_j(t) \geq 0$. Also, because $q_j(0) = 0$, simply by continuity we can enforce the bound $q_j(t) \leq 1$. Similarly, by (2.143), $q'_0(0) = -1$ and thus q_0 decreases on $[0, \epsilon]$, thus providing the bound $q_0(t) \leq 1$. Again, by continuity we can ensure $q_0(t) \geq 0$. Then, for any $t \in [0, \epsilon]$, $\gamma(t) \in [0, 1]^{d+1}$, thus concluding the lemma. $\square$

Definition 2.4.5 (Bi-Analytic Noise Mass Function). *For $\mathbf{q} \in \mathbb{R}_+^{d+1}$ with $\Sigma \mathbf{q} = 1$, $\mathbf{a} \in \mathcal{D}$, we define the Bi-Analytic Noise Mass Function by*

$$P(\mathbf{q}, \mathbf{a}) := \mathbf{p}^{-\mathbf{a}} \mathbf{q}^{\mathbf{a}} T(\mathbf{a}). \quad (2.145)$$

Recall from (2.11) that for $\mathbf{a} \in \mathbb{N}^{d+1}$, $T(\mathbf{a}) = \mathbf{p}^{\mathbf{a}} \binom{N}{\mathbf{a}}$ and therefore $P(\mathbf{q}, \mathbf{a}) = \mathbf{q}^{\mathbf{a}} \binom{N}{\mathbf{a}}$ so (2.145) equals $\mathbb{P}(Q^j = a_j : j = 0, \dots, d)$. Our goal at this stage is to show that, when t is sufficiently close to $\frac{k}{N}$, the corresponding noise distance Q is relatively likely to be k , i.e. $\mathbb{P}_{\{\mathbf{q}=\gamma(t)\}}(Q = k)$ is large in an appropriate sense.

Lemma 2.4.6. *For a fixed $k \leq \epsilon N$, we have*

$$\mathbb{P}_{\{\mathbf{q}=\gamma(k/N)\}}(Q = k) \sim \frac{1}{\sqrt{k}}.$$

Proof. Suppose for a given $t_0 \in [0, \epsilon]$ that there exists a type $\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k$ such that

$$\mathbb{P}_{\{\mathbf{q}=\gamma(t_0)\}}(Q^j = a_j : j = 0, \dots, d) \sim \frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}})}{\sqrt{k}}. \quad (2.146)$$

We observe that for any $\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k$,

$$\frac{P(\gamma(t_0), \mathbf{a})}{T(\mathbf{a})} = \left(\frac{\gamma(t_0)}{\mathbf{p}} \right)^{\mathbf{a}}. \quad (2.147)$$

Thus, in close analogy with Lemma 2.4.3, (2.147) is independent of $\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1}$ because $\gamma(t)$ satisfies the algebraic relations (2.136). Therefore, assuming (2.146),

we would have

$$\begin{aligned}\mathbb{P}_{\{\mathbf{q}=\gamma(t_0)\}}(Q=k) &= \sum_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k} \mathbb{P}_{\{\mathbf{q}=\gamma(t_0)\}}(Q^j = a_j : j = 0, \dots, d) \\ &\sim \sum_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k} \frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}})}{\sqrt{k}} = \frac{1}{\sqrt{k}}\end{aligned}\tag{2.148}$$

We build such a type parameter $\mathbf{a}$ for $t_0 = \frac{k}{N}$ as follows:

$$a_j = \begin{cases} \lfloor Nq_j \left(\frac{k}{N}\right) \rfloor & \text{for } j \geq 2 \\ k - \sum_{i=2}^d ia_i & \text{for } j = 1 \\ N - \sum_{i=1}^d a_i & \text{for } j = 0 \end{cases}$$

and note that $\mathbf{a} \in \mathbb{N}^{d+1}$ as because for all $j \in \{2, \dots, d\}$, q_j is non-negative and $q_j \left(\frac{k}{N}\right) = o\left(\frac{k}{N}\right)$ by Lemma 2.4.4. We note that, because $a_j \geq 0$ for all $j \in \{0, \dots, d\}$, $a_j + 1$ is bounded away from 0 and we may use Claim (2.2.4), item (1), i.e. $\Lambda(x) = \left(x - \frac{1}{2}\right) \ln(x) + O(1)$. However, for any j such that $a_j = 0$, we use the identity $\Lambda(a_j + 1) = \ln \circ \Gamma(1) = \ln(1) = 0$ in lieu of the approximation. Thus we have

$$\begin{aligned}\ln[P(\mathbf{q}, \mathbf{a})] &= \left[\sum_{j=0}^d a_j \ln(q_j) \right] + \left[\Lambda(N+1) - \sum_{j=0}^d \Lambda(a_j + 1) \right] \\ &= \left[\sum_{\substack{0 \leq j \leq d \\ a_j \geq 1}} a_j \ln(q_j) \right] + \left[\Lambda(N+1) - \sum_{\substack{0 \leq j \leq d \\ a_j \geq 1}} \Lambda(a_j + 1) \right] \\ &= \left[\sum_{\substack{0 \leq j \leq d \\ a_j \geq 1}} a_j \left(\ln(Nq_j) - \ln(N) \right) \right] + \left[\Lambda(N+1) - \sum_{\substack{0 \leq j \leq d \\ a_j \geq 1}} \Lambda(a_j + 1) \right] \\ &= \left[\sum_{\substack{0 \leq j \leq d \\ a_j \geq 1}} \left(a_j \ln(Nq_j) - \Lambda(a_j + 1) \right) \right] + \left[\Lambda(N+1) - N \ln(N) \right] \\ &= \left[\sum_{\substack{0 \leq j \leq d \\ a_j \geq 1}} \left(a_j \ln(Nq_j) - \left(a_j + \frac{1}{2}\right) \ln(a_j) + O(1) \right) \right] +\end{aligned}$$

$$\begin{aligned}
& \left[\left(N + \frac{1}{2} \right) \ln(N) - N \ln(N) + O(1) \right] \\
&= \left[\sum_{\substack{0 \leq j \leq d \\ a_j \geq 1}} \left(a_j \left[\ln(Nq_j) - \ln(a_j) \right] \right) \right] + \left[\ln \left(N^{1/2} \prod_{\substack{0 \leq j \leq d \\ a_j \geq 1}} a_j^{-1/2} \right) \right] + O(1) \\
&= \left[\sum_{\substack{0 \leq j \leq d \\ a_j \geq 1}} O \left(\frac{a_j |Nq_j - a_j|}{\min(Nq_j, a_j)} \right) \right] + \left[\ln \left(N^{1/2} \prod_{\substack{0 \leq j \leq d \\ a_j \geq 1}} a_j^{-1/2} \right) \right] + O(1) \\
&= \left[\sum_{\substack{0 \leq j \leq d \\ a_j \geq 1}} O \left(\frac{a_j}{\min(Nq_j, a_j)} \right) \right] + \left[\ln \left(N^{1/2} \prod_{\substack{0 \leq j \leq d \\ a_j \geq 1}} a_j^{-1/2} \right) \right] + O(1)
\end{aligned} \tag{2.149}$$

$$= \left[\sum_{\substack{0 \leq j \leq 2 \\ a_j \geq 1}} O \left(\frac{a_j}{\min(Nq_j, a_j)} \right) \right] + \left[\ln \left(N^{1/2} \prod_{\substack{0 \leq j \leq d \\ a_j \geq 1}} a_j^{-1/2} \right) \right] + O(1) \tag{2.150}$$

Note that to (2.149) used the fact that, by the construction of $\mathbf{a}$, for all $j \in \{0, \dots, d\}$, we have $|Nq_j - a_j| = O(1)$. Similarly, (2.150) used the fact that (again by the construction of $\mathbf{a}$) for each $j \in \{2, \dots, d\}$ we have $\min(Nq_j, a_j) = a_j$ so we may fold any of the corresponding summands into the $O(1)$ error, ignoring them in the estimate.

Because $\mathbf{q} = \gamma(k/N)$, by (2.4.4) we have $q_1 = \frac{k}{N} + O\left(\left(\frac{k}{N}\right)^2\right)$ so $q_1 \sim \frac{k}{N}$ and thus $Nq_1 \sim k$. Similarly, $q_0 = 1 + O\left(\frac{k}{N}\right)$ so for small enough ϵ , (2.4.4) implies that $q_0 \sim 1$ and thus $Nq_0 \sim N$. Because

$$a_1 - Nq_1, a_0 - Nq_0 = O(1),$$

as long as k and N are sufficiently large, we have

$$\begin{aligned}
a_1 &\sim k \sim Nq_1 \text{ and} \\
a_0 &\sim N \sim Nq_0.
\end{aligned} \tag{2.151}$$

Therefore, we have the bounds

$$\frac{a_1}{\min(Nq_1, a_1)}, \frac{a_0}{\min(Nq_0, a_0)} = O(1)$$

and (2.150) reduces to

$$\begin{aligned} \ln [P(\mathbf{q}, \mathbf{a})] &= \ln \left(N^{1/2} \prod_{\substack{0 \leq j \leq d \\ a_j \geq 1}} a_j^{-1/2} \right) + O(1) \\ &= \ln \left(\prod_{\substack{2 \leq j \leq d \\ a_j \geq 1}} a_j^{-1/2} \right) + \frac{1}{2} \ln \left(\frac{N}{a_0} \right) + \ln \left(a_1^{-1/2} \right) + O(1) \\ &= \ln \left(\prod_{\substack{2 \leq j \leq d \\ a_j \geq 1}} a_j^{-1/2} \right) + \frac{1}{2} \ln \left(\frac{N}{a_0} \right) + \frac{1}{2} \ln \left(\frac{k}{a_1} \right) + \ln \left(k^{-1/2} \right) + O(1). \end{aligned} \tag{2.152}$$

Another application of (2.151) proves

$$\ln \left(\frac{N}{a_0} \right), \ln \left(\frac{k}{a_1} \right) = O(1)$$

so (2.152) becomes

$$\ln [P(\mathbf{q}, \mathbf{a})] = \ln \left(\prod_{\substack{2 \leq j \leq d \\ a_j \geq 1}} a_j^{-1/2} \right) + \ln \left(k^{-1/2} \right) + O(1). \tag{2.153}$$

We now note that, for any $j \in \{2, \dots, d\}$ with $a_j \geq 1$,

$$a_j \sim Nq_j = N \left(\frac{p_0^{j-1} p_j}{p_1^j} \left(\frac{k}{N} \right)^j + O \left(\left(\frac{k}{N} \right)^{j+1} \right) \right),$$

which immediately implies

$$a_j \sim \frac{k^j}{N^{j-1}}.$$

In conjunction with Proposition 2.2.8, this proves that

$$b_j^k \sim \frac{k^j}{N^{j-1}} \sim a_j \gtrsim 1. \tag{2.154}$$

Therefore (2.154) implies

$$a_j^{1/2} \sim \max\left(1, |b_j^k|^{1/2}\right). \quad (2.155)$$

Conversely, if $a_j < 1$ (that is, $a_j = 0$) then by construction

$$1 > Nq_j = N \left(\frac{p_0^{j-1} p_j}{p_1^j} \left(\frac{k}{N} \right)^j + O \left(\left(\frac{k}{N} \right)^{j+1} \right) \right) \quad (2.156)$$

$$\implies \frac{k^j}{N^{j-1}} = O(1) \quad (2.157)$$

In this case, once again by Proposition 2.2.8, we have $b_j^k = O(1)$ so

$$\max\left(1, |b_j^k|^{1/2}\right) \sim 1. \quad (2.158)$$

Then exponentiating (2.153) and applying (2.155) and (2.158) yields

$$\begin{aligned} P(\mathbf{q}, \mathbf{a}) &\sim \frac{1}{\sqrt{k}} \left(\prod_{\substack{2 \leq j \leq d \\ a_j \geq 1}} \max\left(1, |b_j^k|^{1/2}\right) \right)^{-1} \\ &\sim \frac{1}{\sqrt{k}} \left(\prod_{j=2}^d \max\left(1, |b_j^k|^{1/2}\right) \right)^{-1} \end{aligned} \quad (2.159)$$

Finally, applying Lemma 2.3.10, (2.159) implies

$$\mathbb{P}_{\{\mathbf{q}=\gamma(k/N)\}}(Q = k) = P(\mathbf{q}, \mathbf{a}) \sim \frac{1}{\sqrt{k}} \left(\max_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \right),$$

proving (2.146) for $t_0 = \frac{k}{N}$ and thereby completing the lemma. $\square$

Lemma 2.4.6 is essentially an autonomous portion of the proof of Lemma 2.4.7 below. As described in the beginning of the section, the goal of Lemma 2.4.7 is to show that if t is sufficiently close to $\frac{k}{N}$, the probability mass $\mathbb{P}_{\{\mathbf{q}=\gamma(t)\}}$ concentrates the random variable Q near k . A useful probabilistic heuristic is think of the distribution of Q as mimicking a Poisson distribution with expectation tN .

Lemma 2.4.7. *For a fixed $k \leq \epsilon N$, we have*

$$\mathbb{P}_{\{\mathbf{q}=\gamma(t)\}}(Q = k) \sim \frac{1}{\sqrt{k}}.$$

uniformly in $t \in \left[\frac{k-\sqrt{k}}{N}, \frac{k}{N} \right]$.

Proof. We adopt the same notation as (2.4.6) without reintroduction. We will show that for t_0 sufficiently close to $\frac{k}{N}$, the derivative $\partial_t [\ln \circ P(\gamma(t), \mathbf{a})]$ is small enough between t_0 and $\frac{k}{N}$ to ensure that $P(\gamma(t_0), \mathbf{a}) \sim P(\gamma(k/N), \mathbf{a}) \sim \frac{1}{\sqrt{k}}$. By (2.148), this will complete the lemma. Examining only the t -dependent terms in $\ln \circ P(\gamma(t), \mathbf{a})$, we compute

$$\partial_t [\ln \circ P(\gamma(t), \mathbf{a})] = \sum_{j=0}^d a_j \partial_t [\ln (q_j(t))] = \sum_{j=0}^d \frac{a_j}{q_j(t)} q'_j(t). \quad (2.160)$$

Fixing $t_1 \in \left[\frac{k-\sqrt{k}}{N}, \frac{k}{N} \right]$ and $j \in \{1, \dots, d\}$, the mean value theorem and Lemma 2.4.4 provide the bound

$$\begin{aligned} |q_j(t_1) - q_j\left(\frac{k}{N}\right)| &\leq \left|t_1 - \frac{k}{N}\right| \sup_{t \in [t_1, k/N]} |q'_j(t)| \\ &= \frac{\sqrt{k}}{N} \sup_{t \in [t_1, k/N]} O(t^{j-1}) \\ &= O\left(\frac{k^{j-1/2}}{N^j}\right). \end{aligned} \quad (2.161)$$

Moreover, by Lemma 2.4.4 we have

$$\begin{aligned} q_j(t_1) &\sim t_1^j \sim \left(\frac{k}{N}\right)^j \\ q'_j(t_1) &\sim t_1^{j-1} \sim \left(\frac{k}{N}\right)^{j-1}, \end{aligned}$$

which we use to estimate

$$\begin{aligned} \frac{a_j}{q_j(t_1)} q'_j(t_1) &= \frac{N q_j\left(\frac{k}{N}\right) + O(1)}{q_j(t_1)} q'_j(t_1) \\ &= \frac{N q_j\left(\frac{k}{N}\right) + O(1)}{q_j\left(\frac{k}{N}\right) + O\left(\frac{k^{j-1/2}}{N^j}\right)} q'_j(t_1) \\ &= \frac{N + O\left(\frac{N^j}{k^j}\right)}{1 + O\left(\frac{1}{\sqrt{k}}\right)} q'_j(t_1) \\ &= \left[N + O\left(\frac{N^j}{k^j}\right) + O\left(\frac{N}{\sqrt{k}}\right) \right] q'_j(t_1) \\ &= N q'_j(t_1) + O\left(\frac{N^j}{k^j} \frac{k^{j-1}}{N^{j-1}}\right) + O\left(\frac{N}{\sqrt{k}} \frac{k^{j-1}}{N^{j-1}}\right) \\ &= N q'_j(t_1) + O\left(\frac{N}{\sqrt{k}}\right). \end{aligned} \quad (2.162)$$

The $q_0(t_1)$ term is handled similarly, although it is less delicate. As in (2.161), we compute

$$|q_0(t_1) - q_0\left(\frac{k}{N}\right)| \leq \left|t_1 - \frac{k}{N}\right| \sup_{t \in [t_1, k/N]} |q'_0(t)| \lesssim \frac{\sqrt{k}}{N}.$$

Note also that by (2.4.4), $q_0(0) = 1$ so as long as ϵ is small enough, $q_0(t) \sim 1$ uniformly in $t \in [0, \epsilon]$. Following the lead of (2.162), we compute

$$\begin{aligned} \frac{a_0}{q_0(t_1)} q'_0(t_1) &= \frac{Nq_0\left(\frac{k}{N}\right) + O(1)}{q_0\left(\frac{k}{N}\right) + O\left(\frac{\sqrt{k}}{N}\right)} q'_0(t_1) \\ &= \frac{N + O(1)}{1 + O\left(\frac{\sqrt{k}}{N}\right)} q'_0(t_1) \\ &= \left[N + O\left(\sqrt{k}\right)\right] q'_0(t_1) \\ &= Nq'_0(t_1) + O\left(\sqrt{k}\right) \\ &= Nq'_0(t_1) + O\left(\frac{N}{\sqrt{k}}\right) \end{aligned} \tag{2.163}$$

where the last line used the blunt bound $k \lesssim N \implies \sqrt{k} \lesssim \frac{N}{\sqrt{k}}$. Note that by (2.4.4), $\gamma(t_1)$ satisfies

$$\sum_{j=0}^d q_j = 1 \implies q_0 = 1 - \sum_{j=1}^d q_j,$$

so we immediately have the identity

$$q'_0(t) = - \sum_{j=1}^d q'_j(t)$$

for all $t \in [0, \epsilon]$. Therefore we can rewrite (2.163) as

$$\frac{a_0}{q_0(t_1)} q'_0(t_1) = -N \sum_{j=1}^d q'_j(t_1) + O\left(\frac{N}{\sqrt{k}}\right). \tag{2.164}$$

Setting $t = t_1$ and applying the estimates (2.162) and (2.164) to (2.160) yields

$$\partial_t [\ln \circ P(\gamma(t), \mathbf{a})]_{t=t_1} = \frac{a_0}{q_0(t_1)} q'_0(t_1) + \sum_{j=1}^d \frac{a_j}{q_j(t_1)} q'_j(t_1)$$

$$\begin{aligned}
&= \left(-N \sum_{j=1}^d q'_j(t_1) + O\left(\frac{N}{\sqrt{k}}\right) \right) + \left(\sum_{j=1}^d N q'_j(t_1) + O\left(\frac{N}{\sqrt{k}}\right) \right) \\
&= O\left(\frac{N}{\sqrt{k}}\right).
\end{aligned} \tag{2.165}$$

We use (2.165) to estimate the difference of $\ln \circ P(\gamma(t_0), \mathbf{a})$ and $\ln \circ P(\gamma(k/N), \mathbf{a})$, noting that (2.165) applies to every $t \in [t_0, \frac{k}{N}] \subset \left[\frac{k-\sqrt{k}}{N}, \frac{k}{N}\right]$. With the mean value theorem, we estimate

$$\begin{aligned}
\ln \circ P\left(\gamma\left(\frac{k}{N}\right), \mathbf{a}\right) - \ln \circ P(\gamma(t_0), \mathbf{a}) &= O\left(\left|\frac{k}{N} - t_0\right| \sup_{t \in [t_0, k/N]} \left|\partial_t [\ln \circ P(\gamma(t), \mathbf{a})]\right|\right) \\
&= O\left(\frac{\sqrt{k}}{N} \frac{N}{\sqrt{k}}\right) = O(1).
\end{aligned} \tag{2.166}$$

Exponentiating (2.166) and rearranging then yields

$$P(\gamma(t_0), \mathbf{a}) \sim P(\gamma(k/N), \mathbf{a}) \sim \frac{1}{\sqrt{k}}.$$

As mentioned above, by (2.148) this completes the proof. $\square$

Lemma 2.4.8. *For any $f : G^N \rightarrow \mathbb{R}_+$, $1 \leq K \leq \epsilon N$, we have the pointwise inequality*

$$\frac{1}{K} \sum_{k=1}^K S_k f \lesssim \frac{N}{K} \int_0^{K/N} \mathcal{N}_{\gamma(t)} f \, dt.$$

Proof. For clarity we will explicitly note the t dependence of the noise operator, writing $\{Q_1(t), \dots, Q_N(t)\}$ for the i.i.d. sequence with distribution $\gamma(t)$ in the definition of $\mathcal{N}_{\gamma(t)}$. By Lemmas 2.4.3 and 2.4.4, we can write $\mathcal{N}_{\gamma(t)}$ as

$$\mathcal{N}_{\gamma(t)} = \sum_{k=0}^{dN} \mathbb{P}(Q(t) = k) S_k. \tag{2.167}$$

Because the spherical operators and the noise operators all preserve non-negativity of functions, the lemma will follow from the bound

$$\frac{1}{K+1} \lesssim \frac{N}{K} \int_0^{K/N} \mathbb{P}(Q(t) = k) \, dt$$

$$\Longleftrightarrow \frac{1}{N} \lesssim \int_0^{K/N} \mathbb{P}(Q(t) = k) dt \quad (2.168)$$

for any $k \leq K \leq \epsilon N$. Using (2.4.7), this is now a minor task:

$$\begin{aligned} \int_0^{K/N} \mathbb{P}(Q(t) = k) dt &\geq \int_{k/N - \sqrt{k}/N}^{k/N} \mathbb{P}(Q(t) = k) dt \\ &\gtrsim \int_{k/N - \sqrt{k}/N}^{k/N} \frac{1}{\sqrt{k}} dt \\ &= \frac{\sqrt{k}}{N} \frac{1}{\sqrt{k}} \\ &= \frac{1}{N}. \end{aligned}$$

Thus proving (2.168) and therefore the whole lemma. $\square$

The power of Lemma 2.4.8 lies in the following auxiliary maximal operator:

Definition 2.4.9 (Parameterized Maximal Operator). *In G^N , the parameterized maximal operator is defined by*

$$M^\gamma f(x) := \sup_{0 < T \leq \epsilon} \frac{1}{T} \int_0^T |\mathcal{N}_{\gamma(t)} f(x)| dt.$$

In §2.5 we will describe our condition on G to ensure a dimension-independent L^p bound on M^γ for all $p \in (1, \infty]$ and in §2.6 we will establish a concrete diagnostic test for it. We will conclude §2.5 by showing that if G satisfies the condition (and therefore M^γ is bounded in L^p independent of dimension), we can quickly bound M_S in L^p independent of dimension using Lemma 2.4.8.

2.5 Noise Semigroups

Definition 2.5.1 (Natural Noise Semigroups). *For any $j \in \{1, \dots, d\}$ we denote the j^{th} natural noise semigroup by $\{\tilde{\mathcal{N}}_{x_j}^j\}_{x_j \in \mathbb{R}_+}$ and define it to be the 1-parameter family of operators on $L^p(G)$ for all $p \in [1, \infty]$ given by*

$$\tilde{\mathcal{N}}_{x_j}^j := e^{x_j(P^j - I)}.$$

We denote the total natural noise semigroup (usually referred to simply as the “noise semigroup”) by $\{\tilde{\mathcal{N}}_{\mathbf{x}}\}_{\mathbf{x} \in \mathbb{R}_+^d}$ and define it to be the d -parameter family of operators on $L^p(G)$ for all $p \in [1, \infty]$ given by

$$\tilde{\mathcal{N}}_{\mathbf{x}} := \tilde{\mathcal{N}}_{x_1}^1 \dots \tilde{\mathcal{N}}_{x_d}^d = \exp \left(\sum_{j=1}^d x_j (P^j - I) \right). \quad (2.169)$$

The same notation and terminology will be used for the N^{th} tensor powers of natural noise operators on $L^p(G^N)$ when the distinction is easily discernible from context.

Note that rearrangement in (2.169) is justified because all relevant operators commute.

Claim 2.5.2. *For all $j \in \{1, \dots, d\}$, the semigroup $\{\tilde{\mathcal{N}}_{x_j}^j\}_{x_j \in \mathbb{R}_+}$ on the base space G satisfies the following properties:*

- (1) *for all $x_j \in \mathbb{R}_+$, $\tilde{\mathcal{N}}_{x_j}^j$ is self-adjoint on $L^2(G)$,*
- (2) *for all $x_j \in \mathbb{R}_+$ and $p \in [1, \infty]$, $\tilde{\mathcal{N}}_{x_j}^j$ is a contraction on $L^p(G)$,*
- (3) *for all $x_j \in \mathbb{R}_+$ and $\tilde{\mathcal{N}}_{x_j}^j$ preserves non-negativity,*
- (4) *for all $x_j \in \mathbb{R}_+$, the operator $\tilde{\mathcal{N}}_{x_j}^j P_0$ is a convex combination of the operators $P_0, \dots, P_d$, and*
- (5) *as $x_j \rightarrow 0$, $\tilde{\mathcal{N}}_{x_j}^j$ converges to the identity in the strong operator topology on $\mathcal{B}(L^2(G))$, i.e. the semigroup is strongly continuous at $x_j = 0$ and therefore on all of $\mathbb{R}_+$.*

Proof. Recall from Definition 2.1.1, axioms (1), (2), (3), and (4) that P , and therefore P^j , are self-adjoint, non-negativity preserving L^1 and L^∞ contractions. By Reisz-Thorin interpolation, they are also L^p contractions for all $p \in [1, \infty]$. We will show that the formal expression

$$\left(I + \frac{x_j (P^j - I)}{n} \right)^n \rightarrow e^{x_j (P^j - I)} \quad (2.170)$$

applies in the sense of L^p operator norm convergence. To do so, we compute for any $\mathcal{B}(L^p(G^N))$, $n \in \mathbb{N}$,

$$\begin{aligned}
\left\| e^T - \left(1 + \frac{T}{n}\right)^n \right\|_p &= \left\| \sum_{m=0}^{\infty} \frac{1}{m!} T^m - \sum_{m=0}^n \binom{n}{m} \left(\frac{T}{n}\right)^m \right\|_p \\
&= \left\| \sum_{m=0}^n \frac{1}{m!} \left(1 - \frac{\prod_{i=0}^{m-1} (n-i)}{n^m}\right) T^m + \sum_{m=n+1}^{\infty} \frac{1}{m!} T^m \right\|_p \\
&\leq \sum_{m=0}^n \left| \frac{1}{m!} \left(1 - \frac{\prod_{i=0}^{m-1} (n-i)}{n^m}\right) \right| \|T\|_p^m + \sum_{m=n+1}^{\infty} \left| \frac{1}{m!} \right| \|T\|_p^m \\
&\leq \sum_{m=0}^n \frac{1}{m!} \left(1 - \frac{\prod_{i=0}^{m-1} (n-i)}{n^m}\right) \|T\|_p^m + \sum_{m=n+1}^{\infty} \frac{1}{m!} \|T\|_p^m \\
&= e^{\|T\|_p} - \left(1 + \frac{\|T\|_p}{n}\right)^n
\end{aligned} \tag{2.171}$$

so, in particular, the classical numeric identity $\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x$ implies the L^p -norm convergence (2.170). Note that, as long as $n \geq x_j \geq 0$, the operator

$$I + \frac{x_j(P^j - I)}{n} = \left(1 - \frac{x_j}{n}\right) I + \frac{x_j}{n} P^j$$

is a convex combination of non-negativity preserving L^p contractions for all $p \in [1, \infty]$. Thus $\left(I + \frac{x_j(P^j - I)}{n}\right)$ and therefore $\left(I + \frac{x_j(P^j - I)}{n}\right)^n$ are also non-negativity preserving L^p contractions and, in light of (2.170), proves that the limit $\tilde{\mathcal{N}}_{x_j}^j$ must also be a non-negativity preserving L^p contraction, i.e. conditions (2) and (3) both hold. Moreover, each operator $\left(1 + \frac{T}{n}\right)^n$ is a polynomial of the self-adjoint operator P and is therefore itself self-adjoint. Thus, the limit $\tilde{\mathcal{N}}_{x_j}^j$ is also self-adjoint, proving condition (1).

Using the Taylor expansion

$$\left\| e^{x_j(P^j - I)} - I \right\|_p = \left\| \sum_{n=1}^{\infty} \frac{x_j^n (P^j - I)^n}{n!} \right\|_p \leq \sum_{n=1}^{\infty} \frac{x_j^n \|P^j - I\|_p^n}{n!} = O_{an}(x_j),$$

we have L^p operator norm continuity for all $p \in [1, \infty]$ (in particular strong L^2 continuity) of the map

$$x_j \mapsto e^{x_j(P^j - I)}$$

at $x_j = 0$.

proving (5). Recall from Definition 2.1.1, axiom (5) that the operators P and P_j for all $j \in \{0, \dots, d\}$ fix the function $\mathbf{1}$ so for any convex combinations T and S of these operators, $TS(\mathbf{1}) = T(\mathbf{1}) = \mathbf{1}$. Therefore, TS is also a convex combination of $\{P_0, P_1, \dots, P_d, P\}$. Moreover, because P sends $\text{span}(\{P_0, \dots, P_d\})$ to itself, if S is a convex combination of $\{P_0, \dots, P_d\}$, then TS must be a convex combination of $\{P_0, \dots, P_d\}$. In particular, this means that if $n \geq x_j \geq 0$, the operator

$$\left(I + \frac{x_j(P^j - I)}{n}\right)^n P_0$$

is a convex combination of $(\{P_0, \dots, P_d\})$ and so too therefore is the limit $\tilde{\mathcal{N}}_{x_j}^j$ proving condition (4). $\square$

Corollary 2.5.3. *Properties (1), (2), (3), and (5) from Claim 2.5.2 hold for the semigroup $\left\{\tilde{\mathcal{N}}_{x_j}^j\right\}_{x_j \in \mathbb{R}_+}$ on the Cartesian power G^N .*

Proof. All of these properties are preserved by tensor products. $\square$

Definition 2.5.4 (Natural Noise Maximal Operators). *For all $j \in \{1, \dots, d\}$, the j^{th} natural noise maximal operator for functions on G^N is defined by*

$$\tilde{\mathcal{N}}_*^j f(y) := \sup_{x_j \in \mathbb{R}_+} \left| \tilde{\mathcal{N}}_{x_j}^j f(y) \right|$$

The total natural noise maximal operator (usually referred to simply as the “noise maximal operator”) is defined by

$$\tilde{\mathcal{N}}_* f(y) := \sup_{\mathbf{x} \in \mathbb{R}_+^d} \left| \tilde{\mathcal{N}}_{\mathbf{x}} f(y) \right|.$$

Stein’s continuous Maximal Ergodic Theorems (in [10]) readily yields dimension-independent L^p bounds on $\tilde{\mathcal{N}}_*$ for all $p \in [1, \infty]$.

Proposition 2.5.5. *For all $p > 1$, we have the bound*

$$\left\| \tilde{\mathcal{N}}_* f \right\|_p \lesssim_p \|f\|_p$$

independent of $f \in L^p(G^N)$ and the dimension N .

Proof. By Claim 2.5.3, each semigroup $\{\tilde{\mathcal{N}}_{x_j}^j\}$ satisfies the hypotheses of [10, Corollary 2, p. 1895]. Therefore, for any $p > 1$, there is a universal bound C_p such that for all $f \in L^p(G^N)$, $j \in \{1, \dots, d\}$,

$$\left\| \tilde{\mathcal{N}}_*^j f \right\|_p \leq C_p \|f\|_p. \quad (2.172)$$

Also recall from Claim 2.5.3 that for all $j \in \{1, \dots, d\}$ and $x_j \in \mathbb{R}_+$, $\tilde{\mathcal{N}}_{x_j}^j$ preserves non-negativity so by extension $\tilde{\mathcal{N}}_{\mathbf{x}}$ must also preserve non-negativity. Therefore it will suffice to prove the proposition for $f : G^N \rightarrow \mathbb{R}_+$ in $L^p(G^N)$.

Also, if T is a general non-negativity preserving operator and $\{f_i\}_{i \in I}$ is a family of functions on which T is defined, then for any $i_0 \in I$ we have the pointwise bound

$$T \left[\sup_{i \in I} f_i \right] = T \left[\left(\sup_{i \in I} f_i \right) - f_{i_0} \right] + T f_{i_0} \geq T f_{i_0} \quad (2.173)$$

because $(\sup_{i \in I} f_i) - f_{i_0}$ and therefore $T(\sup_{i \in I} f_i - f_{i_0})$ are non-negative. Taking a supremum over i_0 in (2.173) proves the pointwise bound

$$T \left(\sup_{i \in I} f_i \right) \geq \sup_{i \in I} T f_i. \quad (2.174)$$

For any such $f : G^N \rightarrow \mathbb{R}_+$ in $L^p(G^N)$, we use (2.174) repeatedly, $\binom{d}{2}$ times to be precise, to obtain the pointwise bound

$$\begin{aligned} \tilde{\mathcal{N}}_* f &= \sup_{\mathbf{x} \in \mathbb{R}_+^d} \tilde{\mathcal{N}}_{\mathbf{x}} f \\ &= \sup_{x_1 \in \mathbb{R}_+} \dots \sup_{x_d \in \mathbb{R}_+} \tilde{\mathcal{N}}_{x_1}^1 \dots \tilde{\mathcal{N}}_{x_d}^d f \\ &\leq \sup_{x_1 \in \mathbb{R}_+} \tilde{\mathcal{N}}_{x_1}^1 \sup_{x_2 \in \mathbb{R}_+} \tilde{\mathcal{N}}_{x_2}^2 \dots \sup_{x_d \in \mathbb{R}_+} \tilde{\mathcal{N}}_{x_d}^d f \\ &= \tilde{\mathcal{N}}_*^1 \dots \tilde{\mathcal{N}}_*^d f \end{aligned} \quad (2.175)$$

Applying (2.172) for $j \in \{1, \dots, d\}$ to (2.175) immediately proves

$$\left\| \tilde{\mathcal{N}}_* f \right\|_p \leq C_p^d \|f\|_p,$$

which was our desired bound. $\square$

Corollary 2.5.6. *The maximal operator $\sup_{\mathbf{x} \in \mathbb{R}_+^d} \tilde{\mathcal{N}}_{\mathbf{x}} P_0$ satisfies the bound*

$$\left\| \sup_{\mathbf{x} \in \mathbb{R}_+^d} \tilde{\mathcal{N}}_{\mathbf{x}} (P_0^{\otimes N}) f \right\|_p \lesssim_p \|f\|_p$$

independent of $f \in L^p(G^N)$ and the dimension N .

Proof. Because P_0 and therefore $P_0^{\otimes N}$ is an L^p contraction, Proposition 2.5.5 implies

$$\left\| \sup_{\mathbf{x} \in \mathbb{R}_+^d} \tilde{\mathcal{N}}_{\mathbf{x}} (P_0^{\otimes N}) f \right\|_p \lesssim_p \|P_0^{\otimes N} f\|_p \leq \|f\|_p.$$

□

Recall from Definition 2.1.1, axiom (9) that the set $\{P_0, \dots, P_d\}$ is linearly independent. Thus, for any operator in $\text{span}(\{P_0, \dots, P_d\})$ we have a well defined projection operator π_j onto each operator in $\{P_0, \dots, P_d\}$. That is,

$$T = \sum_{j=0}^d a_j P_j \implies \pi_j(T) = a_j,$$

noting that π_j maps $\mathcal{B}(L^2(G^N)) \rightarrow \mathbb{C}$, outputting a component rather than an operator. To motivate our coming discussion we think of the natural noise operator $\tilde{\mathcal{N}}_{x_j}^j$ as the time evolution of a diffusion process of mass distributed over $\{P_0, \dots, P_d\}$, where the “direction” of diffusion is determined by j . In particular, $\tilde{\mathcal{N}}_{x_j}^j$ is unique in the set $\left\{ \tilde{\mathcal{N}}_{x_j}^j \right\}_{j=1}^d$ for sending mass immediately to P_j but not to any P_i for $i > j$. It will serve as a stand-in for a diffusion process sending mass from P_0 (exclusively) to P_j , because such a process is not necessarily attainable by a diffusion process that begins with full mass at P_0 . Our goal is to show that, under a particular condition on G , for any $t \in (0, \epsilon]$, the noise parameter $\gamma(t)$ necessary to induce a radial operator (see Lemmas 2.4.3 and 2.4.4) can be obtained by a natural noise, (diffusion) process.

Lemma 2.5.7. *For the duration of this lemma and its proof, let $\mathcal{N}_{\mathbf{q}}$, $\tilde{\mathcal{N}}_{x_j}^j$, and $\tilde{\mathcal{N}}_{\mathbf{x}}$ refer to the operators on the base space G . Then there exists an analytic function*

$\eta : [0, \epsilon] \rightarrow \mathbb{R}^d$ such that if we allow $\tilde{\mathcal{N}}_{\mathbf{x}}$ to take parameters in $\mathbb{R}^d$ rather than $\mathbb{R}^{d+1}$, then

$$\mathcal{N}_{\gamma(t)} = \tilde{\mathcal{N}}_{\eta(t)} P_0$$

for all $t \in [0, \epsilon]$.

Proof. Looking back at the proof of Claim 2.5.2, it is clear that even if $x_j < 0$, the operator $e^{x_j(P^j - I)} P_0$ still lies in $\text{span}(\{P_0, \dots, P_d\})$, although in general it may have negative coefficients. Thus we can write the natural noise operators in the form

$$\tilde{\mathcal{N}}_{\mathbf{x}} P_0 = \sum_{j=0}^d \left(\pi_j \left[\tilde{\mathcal{N}}_{\mathbf{x}} P_0 \right] \right) P_j. \quad (2.176)$$

However, Definition 2.1.1, axiom (5) implies that for all $n \in \mathbb{N}, x_j \in \mathbb{R}$,

$$\begin{aligned} \mathbf{1} &= \left[1 - \frac{x^j}{N} \right] \mathbf{1} + \frac{x_j}{n} \mathbf{1} = \left(I \left[1 - \frac{x^j}{N} \right] + \frac{x_j}{n} P^j \right) \mathbf{1} \\ \implies \mathbf{1} &= \left(I \left[1 - \frac{x^j}{N} \right] + \frac{x_j}{n} P^j \right)^n \mathbf{1} = \left(I + \frac{x_j (P^j - I)}{n} \right)^n \mathbf{1} \\ \implies \mathbf{1} &= \tilde{\mathcal{N}}_{x_j}^j \mathbf{1}. \end{aligned} \quad (2.177)$$

We use (2.177) and (2.176) to see

$$\begin{aligned} \mathbf{1} &= \tilde{\mathcal{N}}_{\mathbf{x}} P_0(\mathbf{1}) = \sum_{j=0}^d \left(\pi_j \left[\tilde{\mathcal{N}}_{\mathbf{x}} P_0 \right] \right) P_j(\mathbf{1}) = \sum_{j=0}^d \left(\pi_j \left[\tilde{\mathcal{N}}_{\mathbf{x}} P_0 \right] \right) \mathbf{1} \\ \implies 1 &= \sum_{j=0}^d \left(\pi_j \left[\tilde{\mathcal{N}}_{\mathbf{x}} P_0 \right] \right). \end{aligned} \quad (2.178)$$

Recall that for a random variable $Q_1(t)$ with distribution $\gamma(t)$,

$$\mathcal{N}_{\gamma(t)} f(x) = \mathbb{E} (P_{Q_1} f(x)) = \sum_{j=0}^d q_j(t) P_j f(x),$$

so the parameterized noise operator $\mathcal{N}_{\gamma(t)}$ decomposes analogously with (2.176) as

$$\mathcal{N}_{\gamma(t)} = \sum_{j=0}^d q_j(t) P_j.$$

The lemma then amounts to finding an $\eta(t)$ meeting the description in the statement such that for all $j \in \{0, \dots, d\}$, $q_j(t) = \pi_j \left[\tilde{\mathcal{N}}_{\eta(t)} P_0 \right]$. Noting that, by (2.178) and the construction of $\gamma(t)$,

$$\sum_{j=0}^d \left(\pi_j \left[\tilde{\mathcal{N}}_{\mathbf{x}} P_0 \right] \right) = 1 = \sum_{j=0}^d q_j(t),$$

so it will suffice to show $q_j(t) = \pi_j \left[\tilde{\mathcal{N}}_{\eta(t)} P_0 \right]$ for $j \in \{1, \dots, d\}$.

Because $\tilde{\mathcal{N}}_{\mathbf{x}}$ is by definition an exponential, $\tilde{\mathcal{N}}_{\mathbf{x}} P_0$ maps from $\mathbb{R}^d$ to $\text{span}(\{P_0, \dots, P_d\})$ analytically so we may define an analytic map $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$ given by

$$\begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \mapsto \begin{bmatrix} \pi_1 \left[\tilde{\mathcal{N}}_{\mathbf{x}} P_0 \right] \\ \vdots \\ \pi_d \left[\tilde{\mathcal{N}}_{\mathbf{x}} P_0 \right] \end{bmatrix}. \quad (2.179)$$

We compute the first derivatives of $\tilde{\mathcal{N}}_{\mathbf{x}} P_0$ for all $j \in \{1, \dots, d\}$ and use Definition 2.1.1, axioms (6) and (7) to derive the useful expression

$$\partial_j \left[\tilde{\mathcal{N}}_{\mathbf{x}} P_0 \right]_{\mathbf{x}=\mathbf{0}} = (P^j - I) P_0 = \sum_{i=1}^j w_{i,j} P_i + (w_{0,j} - 1) P_0. \quad (2.180)$$

Then (2.180) allows us to calculate the Jacobian of F at $\mathbf{0}$, namely

$$\mathcal{J}F(\mathbf{0}) = [w_{j,n}]_{j,n \in \{1, \dots, d\}}. \quad (2.181)$$

In particular, $\mathcal{J}F(\mathbf{0})$ is upper triangular with j^{th} diagonal entry $w_{j,j} > 0$ (by Definition 2.1.1, axiom (8)) and is therefore invertible. By the analytic inverse function theorem, F has a local analytic inverse F^{-1} in a neighborhood of $F(\mathbf{0}) = \mathbf{0}$.

Finally we define the analytic function $\eta : \mathbb{R} \rightarrow \mathbb{R}^d$ by

$$\eta(t) := F^{-1} [q_1(t), \dots, q_d(t)],$$

and note that $q_j(t)$ is the j^{th} component of

$$\left(F \circ F^{-1}\right)[q_1(t), \dots, q_d(t)] = F[\eta(t)] = \begin{bmatrix} \pi_1 [\tilde{\mathcal{N}}_{\eta(t)} P_0] \\ \vdots \\ \pi_d [\tilde{\mathcal{N}}_{\eta(t)} P_0] \end{bmatrix}.$$

by the construction of F . □

The analytic path η is the missing link in the relationship between the parameters of G and our dimension-independence condition. Roughly speaking, if G is sufficiently well behaved, $\eta(t)$ describes a subfamily of the noise semigroup that can dominate the Cesàro mean samples $\frac{1}{K+1} \sum_{k=0}^K S_k$ thus extending the noise maximal operator bound from Corollary 2.5.6 to M_S .

Definition 2.5.8 (Principal Noise Coefficients). *Let*

$$\eta(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_d(t) \end{bmatrix}$$

be the analytic function defined in Lemma 2.5.7. For each $j \in \{1, \dots, d\}$, the j^{th} principal noise coefficient of G is defined to be 0 if x_j is identically 0 or the lowest order coefficient in the Taylor series (expanded at $t = 0$) of $x_j(t)$ otherwise.

The principal noise coefficients frame our dimension independence condition for the auxiliary maximal operator M^γ from Definition 2.4.9 and subsequently on M_S .

Lemma 2.5.9. *If all principal noise coefficients of G are non-negative, then for any $p > 1$,*

$$\|M^\gamma f\|_{L^p(G^N)} \lesssim_p \|f\|_{L^p(G^N)}$$

with a bound independent of $f \in L^p(G^N)$ and the dimension N .

Proof. In this proof we will indicate whether noise operators are defined on G or G^N with the superscripts 1 and N respectively to avoid ambiguity. Unraveling the definition of an operator tensor product and using the independence of the sequence $\{Q_1(t), \dots, Q_N(t)\}$ provides the practical expression

$$\begin{aligned}
\mathcal{N}_{\gamma(t)}^N f(x) &= \mathbb{E} \left(\left[\bigotimes_{n=1}^N P_{Q_n(t)} \right] f(x) \right) \\
&= \left(\bigotimes_{n=1}^N \left[\sum_{j=0}^d q_j(t) P_j \right] \right) f(x) \\
&= \left(\sum_{j=0}^d q_j(t) P_j \right)^{\otimes N} f(x) \\
&= (\mathbb{E} P_{Q_1(t)})^{\otimes N} f(x) \\
\implies \mathcal{N}_{\gamma(t)}^N &= (\mathcal{N}_{\gamma(t)}^1)^{\otimes N}
\end{aligned} \tag{2.182}$$

Therefore, by Lemma 2.5.7 and (2.182), we have the identity

$$\tilde{\mathcal{N}}_{\eta(t)}^N (P_0^{\otimes N}) = \left(\tilde{\mathcal{N}}_{\eta(t)}^1 P_0 \right)^{\otimes N} = (\mathcal{N}_{\gamma(t)}^1)^{\otimes N} = \mathcal{N}_{\gamma(t)}^N. \tag{2.183}$$

Because noise operators preserve non-negativity, it suffices to prove the lemma for $f \geq 0$. In this case, we can ignore sign and simply dominate means by suprema. Therefore for any $0 < T \leq \epsilon$ and $x \in G^N$,

$$\begin{aligned}
\frac{1}{T} \int_0^T |\mathcal{N}_{\gamma(t)}^N f(x)| dt &= \frac{1}{T} \int_0^T \mathcal{N}_{\gamma(t)}^N f(x) dt \\
&\leq \sup_{t \in (0, T]} \mathcal{N}_{\gamma(t)}^N f(x) \\
&\leq \sup_{t \in (0, \epsilon]} \mathcal{N}_{\gamma(t)}^N f(x)
\end{aligned} \tag{2.184}$$

Taking a supremum over all such T , (2.184) then yields

$$M^\gamma f(x) = \sup_{0 < T \leq \epsilon} \frac{1}{T} \int_0^T |\mathcal{N}_{\gamma(t)}^N f(x)| dt \leq \sup_{t \in (0, \epsilon]} \mathcal{N}_{\gamma(t)}^N f(x) \tag{2.185}$$

Using the assumption that all principal noise coefficients are non-negative, as long as ϵ is sufficiently small, $x_j(t) \geq 0$ for all $t \in (0, \epsilon]$ (i.e. $\eta(t) \in \mathbb{R}_+^d$) so we have

the pointwise bound

$$\sup_{t \in (0, \epsilon]} \tilde{\mathcal{N}}_{\eta(t)} (P_0^{\otimes N}) f \leq \sup_{\mathbf{x} \in \mathbb{R}_+^d} \tilde{\mathcal{N}}_{\mathbf{x}} (P_0^{\otimes N}) f, \quad (2.186)$$

which, in conjunction with (2.183) and (2.185), implies the pointwise bound

$$M^\gamma f(x) = \sup_{\mathbf{x} \in \mathbb{R}_+^d} \tilde{\mathcal{N}}_{\mathbf{x}} (P_0^{\otimes N}) f(x). \quad (2.187)$$

Applying Corollary 2.5.6 to (2.187) implies that for any $p \in (1, \infty]$ and $f \in L^p(G^N)$,

$$\|M^\gamma f\|_{L^p(G^N)} \leq \left\| \sup_{\mathbf{x} \in \mathbb{R}_+^d} \tilde{\mathcal{N}}_{\mathbf{x}} (P_0^{\otimes N}) f \right\|_{L^p(G^N)} \lesssim_p \|f\|_{L^p(G^N)}$$

as claimed. $\square$

Proposition 2.5.10. *For any $p \in (1, \infty]$, if the principle noise coefficients are non-negative, then*

$$\|M_S f\|_{L^p(G^N)} \lesssim_p \|f\|_{L^p(G^N)}$$

with a bound independent of $f \in L^p(G^N)$ and the dimension N .

Proof. As in Lemma 2.5.9, it will suffice to prove the bound for $f : G^N \rightarrow \mathbb{R}_+$ in $L^p(G^N)$. Note that for any such function, we have the blunt pointwise bound.

$$M_S f(x) \leq \max_{0 \leq K \leq \epsilon N} \frac{1}{K+1} \sum_{k=0}^K S_k f(x) + \frac{1}{dN\epsilon} \sum_{k=0}^{\mathbb{E}(D)} S_k f(x)$$

Taking L^p norms for any $p > 1$ this yields

$$\begin{aligned} \|M_S f\|_p &\leq \left\| \max_{0 \leq K \leq \epsilon N} \frac{1}{K+1} \sum_{k=0}^K S_k f(x) \right\|_p + \left\| \frac{1}{dN\epsilon} \sum_{k=0}^{\mathbb{E}(D)} S_k f(x) \right\|_p \\ &\leq \left\| \max_{0 \leq K \leq \epsilon N} \frac{1}{K+1} \sum_{k=0}^K S_k f(x) \right\|_p + \frac{1}{\epsilon} \left[\frac{1}{dN} \sum_{k=0}^{dN} \|S_k f\|_p \right] \\ &= \left\| \max_{0 \leq K \leq \epsilon N} \frac{1}{K+1} \sum_{k=0}^K S_k f(x) \right\|_p + O(\|f\|_p) \end{aligned}$$

simply because spherical samples are L^p contractions. Thus it will suffice to prove

$$\left\| \max_{0 \leq K \leq \epsilon N} \frac{1}{K+1} \sum_{k=0}^K S_k f(x) \right\|_p \lesssim_p \|f\|_p. \quad (2.188)$$

However, by Lemma 2.4.8, we have the pointwise bound

$$\max_{1 \leq K \leq \epsilon N} \frac{1}{K} \sum_{k=1}^K S_k f \lesssim \max_{1 \leq K \leq \epsilon N} \frac{N}{K} \int_0^{K/N} \mathcal{N}_{\gamma(t)} f(x) dt \leq M^\gamma f, \quad (2.189)$$

which is very close to the result we seek. We now handle the technical annoyance of comparing the Cesàro means $\frac{1}{K+1} \sum_{k=0}^K S_k$ and $\frac{1}{K} \sum_{k=1}^K S_k$. We use the following blunt pointwise bounds and simple manipulations:

$$\max_{0 \leq K \leq \epsilon N} \frac{1}{K+1} \sum_{k=0}^K S_k f \leq S_0 f + \max_{0 \leq K \leq \epsilon N} \frac{1}{K+1} \sum_{k=1}^K S_k f \quad (2.190)$$

$$= S_0 f + \max_{1 \leq K \leq \epsilon N} \frac{1}{K+1} \sum_{k=1}^K S_k f \leq S_0 f + \max_{1 \leq K \leq \epsilon N} \frac{1}{K} \sum_{k=1}^K S_k f \quad (2.191)$$

Combined, (2.189) and (2.190) provide the pointwise bound

$$\max_{0 \leq K \leq \epsilon N} \frac{1}{K+1} \sum_{k=0}^K S_k f \leq S_0 f + \max_{1 \leq K \leq \epsilon N} \frac{1}{K} \sum_{k=1}^K S_k f \leq S_0 f + M^\gamma f. \quad (2.192)$$

Recalling that S_0 is an L^p contraction and taking L^p norms of both sides of (2.192) and applying Lemma 2.5.9 shows

$$\left\| \max_{0 \leq K \leq \epsilon N} \frac{1}{K+1} \sum_{k=0}^K S_k f \right\|_p \leq \|S_0 f\|_p + \|M^\gamma f\|_p \lesssim \|f\|_p,$$

proving (2.188) and thus the whole proposition. $\square$

2.6 Noise Coefficient Formulas

In this section, we present and justify recursively computable formulas for the coefficients of the analytic functions $\{q_0(t), \dots, q_d(t)\}$ and $\{x_1(t), \dots, x_d(t)\}$ from §2.4 and §2.5. In practice they should be implemented using dynamic programming for the sake of efficiency, but this is tangential to our present discussion. Not that all noise operators in this section are defined on the base space G .

This is the applicable component of Lemma 2.5.9 and Proposition 2.5.10. Formulas 2.6.1 and 2.6.4 provide allow one to, given a radial space G , test directly if it meets our condition and satisfies dimension-independent $L^2(G^N)$ bounds for M_S and therefore M (although the latter will require §2.8 and §2.7 to justify).

Formula 2.6.1 (Radial Noise Coefficients). *The radial noise coefficients are a 2-parameter family of real numbers denoted by $\{b_{j,n}\}_{\substack{0 \leq j \leq d \\ 0 \leq n}}$. We compute them recursively according to the ordering below:*

(1) If $j > n$, then $b_{j,n} = 0$.

(2) If $j = n \in \{0, 1\}$, then $b_{j,n} = 1$.

(3) If $j = n \in \{2, \dots, d\}$, then $b_{j,n} = \frac{p_0^{j-1} p_j}{p_1^j}$.

(4) If $n > j \geq 2$, then $b_{j,n}$ is the coefficient of t^n in the polynomial

$$b_{j,j} \left(\sum_{m=1}^{n-1} b_{1,m} t^m \right)^j \left(\sum_{m=0}^{n-1} \left[\sum_{\ell=1}^{n-1} b_{0,\ell} t^\ell \right]^m \right)^{j-1}.$$

(5) If $n > j = 1$, then $b_{j,n} = - \sum_{i=2}^d i b_{i,n}$.

(6) If $n > j = 0$, then $b_{j,n} = - \sum_{i=1}^d b_{i,n}$.

Proposition 2.6.2. *The radial noise coefficients in Formula (2.6.1) are the Taylor coefficients of the components of $\gamma(t)$:*

$$\gamma(t) = \begin{bmatrix} q_0(t) \\ q_1(t) \\ \vdots \\ q_d(t) \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{\infty} b_{0,n} t^n \\ \sum_{n=0}^{\infty} b_{1,n} t^n \\ \vdots \\ \sum_{n=0}^{\infty} b_{d,n} t^n \end{bmatrix}$$

Proof. We proved in Lemma 2.4.4 that $b_{j,n}$ as described in Formula (2.6.1) indeed corresponds to the respective Taylor coefficients whenever $j \leq n$, so cases (2),

(3), and (4) have already been justified. Moreover, Lemma 2.4.4 also states that $\gamma(t) \cdot \mathbf{d} = t$ so for any $n > 1$, the analytic function $\sum_{i=1}^d i q_i(t)$ has t^n coefficient 0 and thus $b_{1,n} = -\sum_{i=2}^d i b_{i,n}$, justifying case (5). In the same vein, Lemma 2.4.4 states that $\Sigma \gamma(t) = 1$ so for any $n > 0$, the analytic function $\sum_{i=0}^d q_i(t)$ has t^n coefficient 0 and thus $b_{0,n} = -\sum_{i=1}^d b_{i,n}$, justifying case (6).

Thus it only remains to justify case (4), i.e. $n > j \geq 2$. We recall the algebraic relation

$$\frac{q_j q_0^{j-1}}{q_1^j} = \frac{p_j p_0^{j-1}}{p_1^j} = b_{j,j} \quad (2.193)$$

which holds once again by Lemma 2.4.4. We use (2.193) and the geometric sum formula to find an expression for q_j in terms of $\{b_{0,m}\}_{m=1}^\infty$ and $\{b_{1,m}\}_{m=2}^\infty$ (recall $b_{0,0} = b_{1,1} = 1$ and $b_{1,0} = 0$) as follows:

$$\begin{aligned} q_j &= \frac{q_1^j}{q_0^{j-1}} \\ &= \left(t + \sum_{m=2}^\infty b_{1,m} t^m \right)^j \left(\frac{1}{1 + \sum_{\ell=1}^\infty b_{0,\ell} t^\ell} \right)^{j-1} \\ &= \left(t + \sum_{m=2}^\infty b_{1,m} t^m \right)^j \left(\sum_{m=0}^\infty \left[\sum_{\ell=1}^\infty b_{0,\ell} t^\ell \right]^m \right)^{j-1} \end{aligned} \quad (2.194)$$

Now we will show that any term in the fully expanded form of (2.194) containing $b_{0,\mu}$ or $b_{1,\mu}$ for $\mu \geq n$ has order strictly larger than n . Indeed, any term containing $b_{1,m}$ for $m \geq n$ must also appear in

$$\begin{aligned} & b_{1,\mu} t^\mu \left(t + \sum_{m=2}^\infty b_{1,m} t^m \right)^{j-1} \left(\sum_{m=0}^\infty \left[\sum_{\ell=1}^\infty b_{0,\ell} t^\ell \right]^m \right)^{j-1} \\ &= O_{an} \left(t^\mu \left[t + \sum_{m=2}^\infty b_{1,m} t^m \right]^{j-1} \right) \\ &= O_{an}(t^{\mu+j-1}) \\ &= O_{an}(t^{n+1}) \end{aligned}$$

because $n \geq \mu$ and $j \geq 2$. Therefore the t^n coefficient of (2.194) and therefore q_j is equal to that of

$$\left(t + \sum_{m=2}^{n-1} b_{1,m} t^m\right)^j \left(\sum_{m=0}^{\infty} \left[\sum_{\ell=1}^{\infty} b_{0,\ell} t^{\ell}\right]^m\right)^{j-1}. \quad (2.195)$$

Similarly, any term in (2.194) containing $b_{0,\mu}$ have a factor of $b_{0,\mu} t^{\mu}$ from the factor

$$\left(\sum_{m=0}^{\infty} \left[\sum_{\ell=1}^{\infty} b_{0,\ell} t^{\ell}\right]^m\right)^{j-1}$$

in (2.194). However

$$\left(t + \sum_{m=2}^{n-1} b_{1,m} t^m\right)^j = O_{an}(t^j)$$

so any term in (2.194) containing $b_{0,\mu}$ must vanish to at least order $\mu + j$. Thus the t^n coefficient q_j is equal to that of

$$\left(t + \sum_{m=2}^{n-1} b_{1,m} t^m\right)^j \left(\sum_{m=0}^{\infty} \left[\sum_{\ell=1}^{n-1} b_{0,\ell} t^{\ell}\right]^m\right)^{j-1}. \quad (2.196)$$

Finally we note

$$\sum_{m=n}^{\infty} \left[\sum_{\ell=1}^{n-1} b_{0,\ell} t^{\ell}\right]^m = O_{an}(t^n)$$

so any term in the fully expanded form of (2.195) containing a factor in this tail will vanish to at least order $n + j$ once again because

$$\left(t + \sum_{m=2}^{\infty} b_{1,m} t^m\right)^j = O_{an}(t^j).$$

Therefore the t^n coefficient q_j is equal to that of

$$\left(t + \sum_{m=2}^{n-1} b_{1,m} t^m\right)^j \left(\sum_{m=0}^{n-1} \left[\sum_{\ell=1}^{n-1} b_{0,\ell} t^{\ell}\right]^m\right)^{j-1}, \quad (2.197)$$

justifying case (4). □

Definition 2.6.3 (Principal Walk Parameter Matrix). *We refer to the two parameter column-stochastic array $[w_{j,n}]_{\substack{0 \leq j \leq d \\ n \geq 0}}$ from Definition 2.1.1 as the walk parameter matrix. We define the principal walk parameter matrix to be the submatrix*

$$W := [w_{j,n}]_{\substack{1 \leq j \leq d \\ 1 \leq n \leq d}}$$

The matrix W arose organically in the proof of Lemma 2.5.7. Indeed Formula 2.6.4 and Proposition 2.6.5 serve as a constructive analog to that lemma.

Formula 2.6.4 (Natural Noise Coefficients). *For any $n \in \mathbb{N}$, let*

$$\begin{bmatrix} c_{1,n} \\ \vdots \\ c_{d,n} \end{bmatrix}$$

be the ($\mathbb{R}^d$ -valued) coefficient of t^n in the polynomial

$$\sum_{\substack{\alpha \in \mathbb{N}^d \\ \alpha \cdot \mathbf{d} \leq n \\ |\alpha| > 1}} \frac{1}{\alpha!} \left(\left(\prod_{j=1}^d \left(\sum_{m=j}^{n-1} a_{j,m} t^m \right)^{\alpha_j} \right) \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_d \end{bmatrix} \left(\prod_{j=1}^d (P^j - I)^{\alpha_j} \right) P_0 \right)$$

and let $\{b_{j,n}\}_{\substack{0 \leq j \leq d \\ 0 \leq n}}$ be the array of radial noise coefficients, which are computed via Formula 2.6.1. Then we can compute $\{a_{1,n}, \dots, a_{d,n}\}$ using the formula

$$\begin{bmatrix} a_{1,n} \\ \vdots \\ a_{d,n} \end{bmatrix} = W^{-1} \begin{bmatrix} b_{1,n} - c_{1,n} \\ \vdots \\ b_{d,n} - c_{d,n} \end{bmatrix}$$

It may not be immediately obvious that Formula 2.6.4 is a valid recursive formula, as it must have a base case and a non-redundant recursive step. To see that it indeed has both of these attributes, note that for all $j \in \{1, \dots, d\}$, $c_{j,0} = 0$ by construction, which provides a base case for the recursive formula that depends only on the walk parameter matrix, which is intrinsic to G , and the radial noise coefficients, which are calculable via Formula 2.6.1. Moreover, for any $n \in \mathbb{N}$, the coefficients $c_{1,n}, \dots, c_{d,n}$ have no dependence on any natural noise coefficient $a_{j,\mu}$ for $\mu \geq n$. Thus Formula 2.6.4 can be calculated recursively.

Proposition 2.6.5. *The natural noise coefficients in Formula (2.6.4) are the*

Taylor coefficients of the components of $\eta(t)$:

$$\eta(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_d(t) \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{\infty} a_{1,n} t^n \\ \vdots \\ \sum_{n=0}^{\infty} a_{d,n} t^n \end{bmatrix}$$

Proof. We recall the analytic function $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$ defined in Lemma 2.5.7 by

$$\begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \mapsto \begin{bmatrix} \pi_1 [\tilde{\mathcal{N}}_{\mathbf{x}} P_0] \\ \vdots \\ \pi_d [\tilde{\mathcal{N}}_{\mathbf{x}} P_0] \end{bmatrix}.$$

For any multi-index $\alpha \in \mathbb{N}^d$, we calculate the corresponding partial derivative of $\tilde{\mathcal{N}}_{\mathbf{x}} P_0$:

$$\begin{aligned} \partial^\alpha [\tilde{\mathcal{N}}_{\mathbf{x}} P_0]_{\mathbf{0}} &= \partial^\alpha \left[\exp \left(\sum_{j=1}^d x_j (P^j - I) \right) \right]_{\mathbf{0}} P_0 \\ &= \left(\prod_{j=1}^d (P^j - I)^{\alpha_j} \right) P_0 \end{aligned} \quad (2.198)$$

so that the derivatives of F can be written

$$\partial^\alpha F(\mathbf{0}) = \begin{bmatrix} \pi_1 [\prod_{j=1}^d (P^j - I)^{\alpha_j} P_0] \\ \vdots \\ \pi_d [\prod_{j=1}^d (P^j - I)^{\alpha_j} P_0] \end{bmatrix}. \quad (2.199)$$

For the remainder of this proof, for notational convenience we will ignore the $q_0(t)$ component of $\gamma(t)$ and adopt the notation

$$\gamma(t) = \begin{bmatrix} q_1(t) \\ q_2(t) \\ \vdots \\ q_d(t) \end{bmatrix} = \begin{bmatrix} t \\ \sum_{n=2}^{\infty} b_{2,n} t^n \\ \vdots \\ \sum_{n=d}^{\infty} b_{d,n} t^n \end{bmatrix}, \quad (2.200)$$

where the final equality was a result of Lemma 2.4.4. Note that we deliberately cite Lemma 2.4.4 rather than Proposition 2.6.2 because we will not use the full power

of the latter in this proof. We view Formulas 2.6.1 and 2.6.4 as black box results to be used for the explicit computation of the radial and natural noise coefficients in order to find the principal noise coefficients and test for dimension-independent $L^2(G^N)$ bounds on the maximal operator M .

Recall that η was defined in terms of F by

$$\eta(t) := F^{-1} \circ \gamma(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_d(t) \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{\infty} a_{1,n} t^n \\ \vdots \\ \sum_{n=0}^{\infty} a_{d,n} t^n \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{\infty} a_{1,n} t^n \\ \vdots \\ \sum_{n=1}^{\infty} a_{d,n} t^n \end{bmatrix} \quad (2.201)$$

noting that the Taylor expansions start at the linear term because

$$\eta(0) = F^{-1} \circ \gamma(0) = F^{-1}(\mathbf{0}) = \mathbf{0}.$$

Returning to (2.199), we observe that the highest order term (as a polynomial in P) of the expression $\left(\prod_{j=1}^d (P^j - I)^{\alpha_j}\right) P_0$ is $P^{\alpha \cdot \mathbf{d}}$. Therefore, for any $j \in \{1, \dots, d\}$, we have

$$\pi_j \partial^\alpha \left[\tilde{\mathcal{N}}_{\mathbf{x}} P_0 \right]_{\mathbf{0}} \begin{cases} = w_{j,j} & \text{if } j = \alpha \cdot \mathbf{d} \text{ and} \\ = 0 & \text{if } j > \alpha \cdot \mathbf{d}. \end{cases} \quad (2.202)$$

We will now show that for all $1 \leq n < j \leq d$, $a_{j,n} = 0$. Our strategy is to prove that if there is a j such that $x_j(t)$ does not vanish to order at least j , then for at least one such j , $q_j(t)$ will be forced to have a non-zero term of the same order thus contradicting Lemma 2.4.4.

Suppose there exists a pair (j, n) with $1 \leq n < j \leq d$ such that $a_{j,n} \neq 0$, and let (j, n) be the minimal such pair with respect to the ordering

$$(j, n) \prec (i, m) \iff n < m \text{ or } n = m, j > i.$$

In other words, (j, n) is the unique pair such that $a_{i,m} = 0$ for all $(i, m) \neq (j, n)$ with $m \leq n$ and $i \geq j$.

Example 2 (Minimal Pair (j, n)).

$$\eta(t) = \begin{bmatrix} t + t^2 + t^5 \\ t^2 + t^4 + t^5 \\ t^2 + t^3 + t^4 + t^5 \\ t^3 + t^4 \\ t^3 + t^4 + t^5 + t^7 \end{bmatrix} \implies (j, n) = (3, 2)$$

Then, by the minimality of (j, n) , we have

$$q_j(t) = \left(\frac{a_{j,n} t^n}{j!} + O_{an}(t^{n+1}) \right) \quad (2.203)$$

Recalling (2.202), we Taylor expand the j^{th} component of F to get

$$\pi_j [\tilde{\mathcal{N}}_{\mathbf{x}} P_0] = \pi_j \left[\sum_{\alpha \in \mathbb{N}^d} \frac{\eta(t)^\alpha}{\alpha!} \partial^\alpha [\tilde{\mathcal{N}}_{\mathbf{x}} P_0]_{\mathbf{0}} \right] \quad (2.204)$$

Notice that if $\alpha = \mathbf{e}_j$, we can apply Definition 2.1.1, axiom (6) and (2.203) to obtain

$$\frac{\eta(t)^\alpha}{\alpha!} \partial^\alpha [\tilde{\mathcal{N}}_{\mathbf{x}} P_0]_{\mathbf{0}} = \left(\frac{a_{j,n} t^n}{j!} + O_{an}(t^{n+1}) \right) \left[\sum_{i=0}^j w_{i,j} P_i \right]. \quad (2.205)$$

Now let $\alpha \cdot \mathbf{d} \geq j$ be any multi-index in the sum (2.204) and examine $\eta(t)^\alpha$. If there exists $i > j$ such that $\alpha_i \neq 0$, we observe that, by the minimality of the pair (j, n) , for all $m \leq n$ we have $a_{i,m} = 0$ so that $\eta(t)^\alpha = O_{an}(x_i(t)^{\alpha_i}) = O_{an}(t^{n+1})$. Notice that for any $i < j$, again by the minimality of j , the lowest term in $x_i(t)$ must be of order at least $\min(i, n)$. Thus, if $\alpha_{j+1} = \dots = \alpha_d = 0$ and $\alpha \cdot \mathbf{d} \geq j$, the lowest term of $\eta(t)^\alpha$ must be of order at least

$$\sum_{i=1}^{n-1} i \alpha_i + \left(\sum_{i=n}^j \alpha_i \right) n \quad (2.206)$$

Unless $\alpha = \mathbf{e}_j$, the condition $\alpha \cdot \mathbf{d} \geq j$ implies that $|\alpha|$ must be at least 2. Therefore, if $\alpha_i > 0$ for any $i \geq n$, (2.206) must exceed n .

Finally, if $\alpha_i = 0$ for all $i \geq n$, then, by (2.202)

$$\sum_{i=1}^{n-1} i\alpha_i = \boldsymbol{\alpha} \cdot \mathbf{d} \geq j.$$

The upshot of this computation is that, unless $\boldsymbol{\alpha} = \mathbf{e}_j$, $\eta(t)^\alpha = O_{an}(t^{n+1})$ for any $\boldsymbol{\alpha}$ that corresponds to a non-vanishing term appearing in (2.204), the Taylor series of the j^{th} component of $F(\mathbf{x})$. Plugging $\eta(t)$ into (2.204) and applying (2.205) then yields the expression

$$\begin{aligned} q_j(t) &= \pi_j \left[\tilde{\mathcal{N}}_{\eta(t)} P_0 \right] = \pi_j \left[\sum_{\boldsymbol{\alpha} \cdot \mathbf{d} \geq j} \frac{\eta(t)^\alpha}{\boldsymbol{\alpha}!} \partial^\alpha \left[\tilde{\mathcal{N}}_{\mathbf{x}} P_0 \right]_{\mathbf{0}} \right] \\ &= \left(\frac{a_{j,n} t^n}{j!} + O_{an}(t^{n+1}) \right) w_{j,j} + \sum_{\substack{\boldsymbol{\alpha} \cdot \mathbf{d} \geq j \\ \boldsymbol{\alpha} \neq \mathbf{e}_j}} O_{an}(t^{n+1}) \pi_j \left(\partial^\alpha \left[\tilde{\mathcal{N}}_{\mathbf{x}} P_0 \right]_{\mathbf{0}} \right) \\ &= \frac{a_{j,n} t^n}{j!} w_{j,j} + O_{an}(t^{n+1}). \end{aligned} \quad (2.207)$$

Because of our assumption that $a_{j,n} \neq 0$, this contradicts the fact that $q_j(t) = O_{an}(t^j) = O_{an}(t^{n+1})$ from Lemma 2.4.4 as claimed, so

$$a_{j,n} = 0 \text{ for all } (j, n) \text{ such that } 0 \leq n < j \leq d. \quad (2.208)$$

We will now show that the coefficient array $[a_{j,n}]_{\substack{1 \leq j \leq d \\ 0 \leq n}}$ satisfies the computable Formula 2.6.4 by solving a sequence of d equation linear systems. These systems are obtained by comparing each component of the t^n coefficients the Taylor series of $\gamma(t)$ and $F \circ \eta(t)$ in increasing order. Fixing n , (2.208) implies that for all $j \in \{1, \dots, d\}$, $x_j(t) = O_{an}(t^j)$ and therefore $\eta(t)^\alpha = O_{an}(t^{\boldsymbol{\alpha} \cdot \mathbf{d}})$. Thus the $\mathbb{R}^d$ valued coefficient of t^n in the series

$$\begin{aligned} \gamma(t) &= [F \circ \eta](t) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d} \frac{1}{\boldsymbol{\alpha}!} [\partial^\alpha F(\mathbf{0})] \eta(t)^\alpha \\ &= \sum_{\boldsymbol{\alpha} \cdot \mathbf{d} \leq n} \frac{1}{\boldsymbol{\alpha}!} [\partial^\alpha F(\mathbf{0})] \eta(t)^\alpha + O_{an}(t^{n+1}) \end{aligned} \quad (2.209)$$

matches that of

$$\begin{aligned} \sum_{\alpha \cdot \mathbf{d} \leq n} \frac{1}{\alpha!} [\partial^\alpha F(\mathbf{0})] \eta(t)^\alpha &= \sum_{\alpha \cdot \mathbf{d} \leq n} \frac{1}{\alpha!} \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_d \end{bmatrix} \left(\prod_{j=1}^d \left(\sum_{m=j}^{\infty} a_{j,m} t^m \right)^{\alpha_j} (P^j - I)^{\alpha_j} \right) P_0 \\ &= \sum_{j=1}^d \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_d \end{bmatrix} (P^j - I) P_0 \left(\sum_{m=j}^{\infty} a_{j,m} t^m \right) + \end{aligned} \quad (2.210)$$

$$\sum_{\substack{\alpha \cdot \mathbf{d} \leq n \\ |\alpha| > 1}} \frac{1}{\alpha!} \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_d \end{bmatrix} \left(\prod_{j=1}^d \left(\sum_{m=j}^{\infty} a_{j,m} t^m \right)^{\alpha_j} (P^j - I)^{\alpha_j} \right) P_0 \quad (2.211)$$

Recall from (2.180) that $(P^j - I) P_0 = \sum_{i=1}^j w_{i,j} P_i + (w_{0,j} - 1) P_0$. Therefore, the t^n term of the series (2.210) is

$$\begin{aligned} &\sum_{j=1}^d \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_d \end{bmatrix} \left[\sum_{i=1}^j w_{i,j} P_i + (w_{0,j} - 1) P_0 \right] (a_{j,n} t^n) \\ &= \sum_{j=1}^d \sum_{i=1}^j w_{i,j} \mathbf{e}_i (a_{j,n} t^n) = \sum_{i=1}^j \left(\sum_{j=1}^d w_{i,j} a_{j,n} \right) \mathbf{e}_i t^n \\ &= W \begin{bmatrix} a_{1,n} \\ \vdots \\ a_{d,n} \end{bmatrix} t^n \end{aligned} \quad (2.212)$$

Examining (2.211) now, we note that any term in the fully expanded form with a factor of $a_{i,\mu}$ for some $\mu \geq n$, also appears in

$$a_{i,\mu} t^\mu \sum_{\substack{\alpha \cdot \mathbf{d} \leq n \\ |\alpha| > 1 \\ \alpha_i \geq 1}} \frac{1}{\alpha!} \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_d \end{bmatrix} \left(\prod_{\substack{1 \leq j \leq d \\ j \neq i}} \left(\sum_{m=j}^{\infty} a_{j,m} t^m \right)^{\alpha_j - \delta_{i,j}} (P^j - I)^{\alpha_j} \right) P_0 = O_{an}(t^{\mu+1})$$

where $\delta_{i,j}$ is the standard Kronecker delta. Therefore, $a_{i,\mu}$ does not appear in the

t^n coefficient of (2.211) so it agrees with the t^n coefficient of

$$\sum_{\substack{\alpha \cdot \mathbf{d} \leq n \\ |\alpha| > 1}} \frac{1}{\alpha!} \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_d \end{bmatrix} \left(\prod_{j=1}^d \left(\sum_{m=j}^{n-1} a_{j,m} t^m \right)^{\alpha_j} (P^j - I)^{\alpha_j} \right) P_0. \quad (2.213)$$

Then, defining

$$\begin{bmatrix} c_{1,n} \\ \vdots \\ c_{d,n} \end{bmatrix}$$

to be the t^n coefficient of (2.213), the identity

$$\begin{aligned} \gamma(t) &= \sum_{j=1}^d \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_d \end{bmatrix} (P^j - I) P_0 \left(\sum_{m=j}^{\infty} a_{j,m} t^m \right) + \\ &\quad \sum_{\substack{\alpha \cdot \mathbf{d} \leq n \\ |\alpha| > 1}} \frac{1}{\alpha!} \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_d \end{bmatrix} \left(\prod_{j=1}^d \left(\sum_{m=j}^{\infty} a_{j,m} t^m \right)^{\alpha_j} (P^j - I)^{\alpha_j} \right) P_0 \end{aligned}$$

from (2.209), (2.210), and (2.211) in conjunction with (2.212) implies

$$\begin{bmatrix} a_{1,n} \\ \vdots \\ a_{d,n} \end{bmatrix} = W^{-1} \begin{bmatrix} b_{1,n} - c_{1,n} \\ \vdots \\ b_{d,n} - c_{d,n} \end{bmatrix}$$

which proves the validity of Formula 2.6.4. $\square$

2.7 Fourier Analysis

To take advantage of the smoothness of the map $k \mapsto S_k$ that we will establish in §2.8, we must establish some spectral machinery for $L^2(G^N)$.

Definition 2.7.1 (Frequency). *Let $\xi = (\xi_1, \dots, \xi_N) \in \widehat{G}^N$. Then we define and denote the frequency of ξ by*

$$\|\xi\|_{freq} := |\{n : \xi_n \neq \mathbf{1}\}|.$$

Definition 2.7.2 (Fourier Transform). *Let $f \in L^2(G)$. We define the Fourier transform of f to be the function on $\widehat{G}$ given by*

$$\widehat{f}(\xi) := \langle f, \xi \rangle_{L^2(G^N)}.$$

Also, for any operator $A \in \mathcal{B}(L^2(G))$ such that $\widehat{G}$ diagonalizes A , we define the Fourier transform of A (denoted by $\widehat{A}$) to be the function mapping ξ to the eigenvalue of ξ with respect to A . We use the same notation for functions in $L^2(G^N)$ and operators in $\mathcal{B}(L^2(G^N))$ that are diagonalized by $\widehat{G}^N$, defining the Fourier transform to be the tensor power of the transform defined above. We also use the symbol $\mathcal{F}$ for both maps so that for $f \in L^2(G^N)$, $A \in \mathcal{B}(L^2(G^N))$ diagonalized by $\widehat{G}^N$,

$$\begin{aligned}\mathcal{F}f(\xi) &= \widehat{f}(\xi) \text{ and} \\ \mathcal{F}A(\xi) &= \widehat{A}(\xi).\end{aligned}$$

We recall from Definition 2.1.1, axiom (10) that $\widehat{G}$ is an orthonormal eigenbasis for each operator in $\{P_0, \dots, P_d, P\}$. It follows directly that $\widehat{G}^N$ is an orthonormal eigenbasis for the spherical sampling operators $\{S_k\}_{k=0}^{2N}$. In particular we have the Parseval identities for all $f, g \in L^2(G^N)$:

$$\begin{aligned}\|f\|_{L^2(G^N)}^2 &= \sum_{\xi \in \widehat{G}^N} |\widehat{f}(\xi)|^2 \\ \langle f, g \rangle_{L^2(G^N)} &= \sum_{\xi \in \widehat{G}^N} \widehat{f}(\xi) \overline{\widehat{g}(\xi)}.\end{aligned}\tag{2.214}$$

In this subsection, will review a square function technique of Elias Stein. Our presentation is very close to that found in §3.1 of [8]. For a richer treatment of this square function theory, see [10] or [9].

Definition 2.7.3 (Littlewood-Paley Square Function). *For any $f \in L^2(G^N)$ the Littlewood-Paley Square Function of f is defined and denoted by*

$$Rf(y)^2 := \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) |S_k f(y) - S_{k-1} f(y)|^2.$$

Rather than work directly with the Littlewood-Paley Square Function, we will focus our attention on averaging operators that are sufficiently close to the spherical averaging operators in an $L^2(G^N)$ sense but more amenable to smoothness estimates. For the present, we will assume the existence of such operators and then prove their existence and properties in §2.8 and §2.9.

Lemma 2.7.4. *Let the families $\{S'_k\}_{0 \leq k \leq \mathbb{E}(D)-1}, \{S''_k\}_{0 \leq k \leq \mathbb{E}(D)-1} \subset \mathcal{B}(L^2(G^N))$ be self-adjoint $L^2(G^N)$ contractions that are simultaneously diagonalizable by $\widehat{G}^N$ and satisfy the spectral bound*

$$\left| \widehat{S''_k}(\xi) - \widehat{S'_k}(\xi) \right| \lesssim \frac{r}{N} e^{-\Omega(rk/N)}$$

uniformly in $0 \leq k \leq \mathbb{E}(D) - 1$ and $\xi \in \widehat{G}^N$ with $\|\xi\|_{freq} = r$. Then for any $f \in L^2(G^N)$, the square function

$$R'f(y)^2 := \sum_{0 \leq k \leq \mathbb{E}(D)-1} k |S''_k f(y) - S'_k f(y)|^2.$$

satisfies the $L^2(G^N)$ bound

$$\|R'f\|_2 \lesssim \|f\|_2$$

uniformly in $f \in L^2(G^N)$.

Proof. We first observe that for any $S \in \{S'_k\}_{k=0}^{2N} \cup \{S''_k\}_{k=0}^{2N}$ and $f \in L^2(G^N)$,

$$\begin{aligned} \mathcal{F}(Sf)(\xi) &= \langle Sf, \xi \rangle \\ &= \langle f, S\xi \rangle \\ &= \widehat{S}(\xi) \langle f, \xi \rangle \\ &= \widehat{S}(\xi) \widehat{f}(\xi) \end{aligned}$$

where we used the self-adjointness of $\widehat{S}$ twice; once directly and once to pull its real (by self-adjointness) eigenvalue out of the inner product without conjugation.

We now compute directly

$$\|R'f\|_2^2 = \sum_{y \in G^N} \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) |S''_k f(y) - S'_k f(y)|^2$$

$$\begin{aligned}
&= \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) \sum_{y \in G^N} |S_k'' f(y) - S_k' f(y)|^2 \\
&= \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) \|S_k'' f - S_k' f\|_2^2 \tag{2.215}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) \sum_{\xi \in \widehat{G}^N} |\mathcal{F}(S_k'' f - S_k' f)(\xi)|^2 \\
&\leq \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) \left(\max_{\xi \in \widehat{G}^N} |\widehat{S}_k''(\xi) - \widehat{S}_k'(\xi)|^2 \right) \left(\sum_{\xi \in \widehat{G}^N} |\widehat{f}(\xi)|^2 \right) \\
&= \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) \left(\max_{\xi \in \widehat{G}^N} |\widehat{S}_k''(\xi) - \widehat{S}_k'(\xi)|^2 \right) \|f\|_2^2. \tag{2.216}
\end{aligned}$$

Let $\xi \in \widehat{G}^N$ attain the maximum in (2.216) and $r = \|\xi\|_{freq}$. Then, combining (2.216) with our spectral hypotheses, we can choose $c > 0$ to be small enough that

$$\begin{aligned}
\|R' f\|_2^2 &= \|f\|_2^2 \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) \left| \widehat{S}_k''(\xi) - \widehat{S}_k'(\xi) \right|^2 \\
&\lesssim \|f\|_2^2 \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) \left(\frac{r}{N} \right)^2 e^{-(2cr/N)k} \\
&\lesssim \|f\|_2^2 \left(\frac{r}{N} \right)^2 \sum_{k=1}^{\infty} k e^{-(2cr/N)k}. \tag{2.217}
\end{aligned}$$

Now, using elementary rearrangements and bounds including the geometric series formula, we compute

$$\begin{aligned}
\sum_{k=1}^{\infty} k e^{-(2cr/N)k} &= \sum_{k=1}^{\infty} \left(\sum_{j=k}^{\infty} e^{-(2cr/N)j} \right) \\
&= \left(\sum_{k=1}^{\infty} e^{-(2cr/N)k} \left[\sum_{j=0}^{\infty} e^{-(2cr/N)j} \right] \right) \\
&= \left(\sum_{k=1}^{\infty} e^{-(2cr/N)k} \right) \left(\frac{1}{1 - e^{-(2cr/N)}} \right) \\
&= \left(\frac{e^{-(2cr/N)}}{1 - e^{-(2cr/N)}} \right) \left(\frac{1}{1 - e^{-(2cr/N)}} \right) \\
&\leq \left(\frac{1}{1 - e^{-(2cr/N)}} \right)^2. \tag{2.218}
\end{aligned}$$

To improve on (2.218), we recall that $r \in \{1, \dots, N\}$ and observe

$$\begin{aligned} 1 - e^{-(2cr/N)} &= \int_{-2cr/N}^0 e^x dx \geq \frac{2cr}{N} e^{-2cr/N} \geq \frac{2cr}{N} e^{-2c} \\ \Rightarrow \left(\frac{1}{1 - e^{-(2cr/N)}} \right)^2 &\leq \left(\frac{N}{2cr} e^{2cr/N} \right)^2 \leq \left(\frac{N}{2cr} e^{2c} \right)^2 \lesssim \left(\frac{N}{r} \right)^2. \end{aligned} \quad (2.219)$$

Finally, plugging (2.219) into (2.217) yields our desired bound, namely

$$\begin{aligned} \|R'f\|_2^2 &\lesssim \|f\|_2^2 \\ \Rightarrow \|R'f\|_2 &\lesssim \|f\|_2. \end{aligned}$$

□

Corollary 2.7.5. *Suppose the families $\{S'_k\}_{0 \leq k \leq \mathbb{E}(D)-1}$, $\{S''_k\}_{0 \leq k \leq \mathbb{E}(D)-1} \subset \mathcal{B}(L^2(G^N))$ satisfy the hypotheses of Lemma 2.7.4 and that for all $0 \leq k \leq \mathbb{E}(D) - 1$,*

$$\|S_k - S'_k\|_2, \|S_{k+1} - S''_k\|_2 \lesssim \frac{1}{N^2}.$$

Then the Littlewood-Paley square function satisfies the $L^2(G^N)$ bound

$$\|Rf\|_2 \lesssim \|f\|_2$$

uniformly in $f \in L^2(G^N)$.

Proof. Defining $R'f(x)$ as in Lemma 2.7.4, the corollary will follow if we can show that $\|Rf\|_2 - \|R'f\|_2 = O(\|f\|_2)$. To this end, we apply (2.215) to both R and R' to compute

$$\begin{aligned} &\|Rf\|_2^2 - \|R'f\|_2^2 \\ &= \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) \|S_{k+1}f - S_kf\|_2^2 - \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) \|S''_kf - S'_kf\|_2^2 \\ &= \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) \left[\|S_{k+1}f - S_kf\|_2^2 - \|S''_kf - S'_kf\|_2^2 \right] \\ &= \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) \left[\left(\|S_{k+1}f\|_2^2 - \|S''_kf\|_2^2 \right) + \left(\langle S_{k+1}f, S_kf \rangle - \langle S''_kf, S'_kf \rangle \right) \right] \end{aligned}$$

$$\begin{aligned}
& \left(\|S_k f\|_2^2 - \|S'_k f\|_2^2 \right) \Big] \\
&= \sum_{0 \leq k \leq \mathbb{E}(D)-1} (k+1) O\left(\frac{1}{N^2} \|f\|_2^2\right) \\
&= O\left(\|f\|_2^2\right).
\end{aligned} \tag{2.220}$$

which concludes the corollary by Lemma 2.7.4. $\square$

At last, we are in a position to prove our main theorem.

Proof of Theorem 2.1.5. Fixing $2 \leq p \leq \infty$, by Minkowski's inequality, for any $f \in L^p, k \in \{0, \dots, dN\}$,

$$|S_k f(y)| = \left| \mathbb{E}_{\mathbb{S}_k} \left[\bigotimes_{n=1}^N P_{D_n} f(y) \right] \right| \tag{2.221}$$

$$\leq \mathbb{E}_{\mathbb{S}_k} \left| \bigotimes_{n=1}^N P_{D_n} f(y) \right| \tag{2.222}$$

$$\leq \mathbb{E}_{\mathbb{S}_k} \bigotimes_{n=1}^N P_{D_n} |f|(y) \tag{2.223}$$

$$\tag{2.224}$$

where the final inequality was simply an application of Definition 2.1.1, axiom (4) in each component. Because f and $|f|$ share L^p norm, it suffices to prove the theorem for only non-negative functions. Therefore, we assume without loss of generality that $f \geq 0$ and we will omit absolute values made superfluous by this assumption.

For such a fixed $f \in L^2(G^N)$, we linearize M by letting defining $k_f(y) : G^N \rightarrow \{0, \dots, \lfloor \mathbb{E}(D) \rfloor\}$ to be a radius achieving the maximum of all spherical samples at y , that is

$$Mf(y) = S_{k_f(y)} f(y)$$

Then, for any $y \in G^N$, we have

$$Mf(y) = \left(S_{k_f(y)} f(y) - \frac{1}{k_f(y) + 1} \sum_{k=0}^{k_f(y)} S_k f(y) \right) + \frac{1}{k_f(y) + 1} \sum_{k=0}^{k_f(y)} S_k f(y)$$

$$\begin{aligned}
&= \left(\frac{1}{k_f(y) + 1} \sum_{k=0}^{k_f(y)-1} (k+1) (S_{k+1}f(y) - S_k f(y)) \right) + \frac{1}{k_f(y) + 1} \sum_{k=0}^{k_f(y)} S_k f(y) \\
&\leq \left(\frac{1}{k_f(y) + 1} \sum_{k=0}^{k_f(y)-1} (k+1) [S_{k+1}f(y) - S_k f(y)] \right) + M_S f(y). \quad (2.225)
\end{aligned}$$

To bound the term

$$\left(\frac{1}{k_f(y) + 1} \sum_{k=0}^{k_f(y)-1} (k+1) [S_{k+1}f(y) - S_k f(y)] \right)$$

we apply the Cauchy-Schwarz inequality with respect to the summand k :

$$\begin{aligned}
&\sum_{k=0}^{k_f(y)-1} (k+1) [S_{k+1}f(y) - S_k f(y)] \\
&\leq \left[\sum_{k=0}^{k_f(y)-1} (\sqrt{k+1})^2 \right]^{1/2} \left[\sum_{k=0}^{k_f(y)-1} ((k+1) [S_{k+1}f(y) - S_k f(y)])^2 \right]^{1/2} \\
&\leq k_f(y) \left[\sum_{k=0}^{k_f(y)-1} ((k+1) [S_{k+1}f(y) - S_k f(y)])^2 \right]^{1/2} \\
&\leq k_f(y) Rf(y) \quad (2.226)
\end{aligned}$$

Plugging (2.226) into (2.225), we have the pointwise bound

$$Mf(y) \leq \left(\frac{k_f(y) Rf(y)}{k_f(y) + 1} \right) + M_S f(y) \leq Rf(y) + M_S f(y) \quad (2.227)$$

and the corresponding L^2 bound

$$\|Mf\|_2 \leq \|Rf\|_2 + \|M_S f\|_2. \quad (2.228)$$

By Lemma 2.8.3 and Proposition 2.9.1 below, the operator families

$$\{S'_k\}_{0 \leq k \leq \mathbb{E}(D)-1}, \{S''_k\}_{0 \leq k \leq \mathbb{E}(D)-1} \subset \mathcal{B}(L^2(G^N))$$

defined in Definition 2.8.2 satisfy the hypotheses of Lemma 2.7.4 and Corollary 2.7.5 so $\|Rf\|_2 \lesssim \|f\|_2$. Also, by Proposition 2.5.5, $\|M_S f\|_2 \lesssim \|f\|_2$ so (2.228) implies

$$\|Mf\|_2 \lesssim \|f\|_2. \quad (2.229)$$

Interpolating (2.229) against the trivial L^∞ bound concludes the proof. $\square$

2.8 Difference Bound for Neighboring Spheres

In [8], [3], and [7], combinatorial relations were used to show that the (operator valued) map $k \mapsto S_k$ satisfied a crucial “smoothness” property. Although the relation is not quite so simple, heuristically, for an eigenvector v with frequency r in a sense to be made formal in §2.7, $\left(\widehat{S_k} - \widehat{S_{k-1}}(v)\right) \approx \frac{r}{N} \widehat{S_k}(v)$. Increasing the diameter of the base graph sacrifices the elegant recursion of spheres in a clique power.

In order to compensate for the loss of an explicit relation to ensure smoothness, we must analyze more closely the relationship between the process of sampling over S_k and the process of sampling over S_{k+1} . Our goal in this section is to show that the probability distribution of the random variable D_1 conditioned on S_k , recalling that S_k is by definition the event $\left\{\sum_{n=1}^N D_n = k\right\}$, varies sufficiently smoothly with respect to k . This will provide a crucial quantitative estimate in §2.7.

We introduce the function

$$R(\mathbf{a}) := \frac{p_1}{p_0} \frac{a_0}{(a_1 + 1)},$$

with the domain

$$\mathcal{D}_R := \{\mathbf{a} \in \mathcal{D} : a_0 > 0, a_1 < N + d - 1\}.$$

Note that when $\mathbf{a} \in \mathcal{D}_R \cap \mathbb{N}^{d+1}$, $R(\mathbf{a})$ matches probability ratio

$$\frac{p_1}{p_0} \frac{a_0}{(a_1 + 1)} = \frac{\mathbf{p}^{\mathbf{a}+\mathbf{f}_1} \binom{N}{\mathbf{a}+\mathbf{f}_1}}{\mathbf{p}^{\mathbf{a}} \binom{N}{\mathbf{a}}} = \frac{\mathbb{P}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1})}{\mathbb{P}(\mathbb{T}_{\mathbf{a}})} \quad (2.230)$$

Our goal is to show that, for any $\mathbf{a}, \mathbf{b}$ sufficiently close to the peak type of S_k (i.e. $\mathbf{b}^k$), $R(\mathbf{a})$ and $R(\mathbf{b})$ are sufficiently close to provide useful bounds. We make this notion concrete in Lemma 2.8.1.

Lemma 2.8.1. *For any sufficiently large $k \leq \mathbb{E}(D)$, there exists some (k -dependent) $c > 0$ and corresponding subsets*

$$\begin{aligned}\mathbb{S}'_k &:= \bigcup_{\delta_{\mathbf{a}} \leq c} \mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k \\ \mathbb{S}''_k &:= \bigcup_{\delta_{\mathbf{a}} \leq c} \mathbb{T}_{\mathbf{a}+\mathbf{f}_1} \subset \mathbb{S}_{k+1}.\end{aligned}$$

such that, if

$$R_k := \frac{\mathbb{P}(\mathbb{S}''_k)}{\mathbb{P}(\mathbb{S}'_k)},$$

we have for any $\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3$ with $\delta_{\mathbf{a}} \leq c$,

- (1) $\mathbb{P}_{\mathbb{S}_k}(\mathbb{S}_k \setminus \mathbb{S}'_k) \lesssim \frac{1}{N^2}$,
- $\mathbb{P}_{\mathbb{S}_{k+1}}(\mathbb{S}_{k+1} \setminus \mathbb{S}''_k) \lesssim \frac{1}{N^2}$,
- (2) $\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) = \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \left[1 + O\left(\frac{1}{N^2}\right)\right]$,
- $\mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) = \mathbb{P}_{\mathbb{S}_{k+1}}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) \left[1 + O\left(\frac{1}{N^2}\right)\right]$, and
- (3) $\frac{|R(\mathbf{a}) - R_k|}{R_k} \lesssim \max\left(\frac{1}{\sqrt{N}}, \frac{1}{k}\right) \delta_{\mathbf{a}}$.

Proof. Before choosing c , we build up heuristics for what it “should” be. Proposition 2.2.8 and Lemma 2.3.3 provide powerful tools for localizing $\mathbb{S}_k$ and $\mathbb{S}_{k+1}$ around their peak types $\mathbf{b}^k$ and $\mathbf{b}^{k+1}$. Note that, by Corollary 2.2.11, $|\mathbf{b}^k - \mathbf{b}^{k+1}| \lesssim 1$. In particular, Lemma 2.3.3 says that there exists some $A > 1$ such that for all positive integers c ,

$$\sum_{\substack{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1} \\ \delta_{\mathbf{a}} > c}} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}), \quad \sum_{\substack{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1} \\ \delta_{\mathbf{a}} > c}} \mathbb{P}_{\mathbb{S}_{k+1}}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) \lesssim A^{-c}$$

Picking $c = 2 \log_A(N)$ would ensure

$$\sum_{\substack{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1} \\ \delta_{\mathbf{a}} > c}} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}), \quad \sum_{\substack{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1} \\ \delta_{\mathbf{a}} > c}} \mathbb{P}_{\mathbb{S}_{k+1}}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) \lesssim \frac{1}{N^2}. \quad (2.231)$$

which we will soon see satisfies conditions (1) and (2) in the lemma statement. However, this fails to preclude the problem of a type in $\mathbb{S}_{k+1}$ near $\mathbf{b}^{k+1}$ that has no analog in $\mathbb{S}_k$ and thus leaving $R(\mathbf{a})$ undefined.

To circumvent this issue, we notice that it can only occur for k very small, in which case the mass of $\mathbb{S}_k$ is overwhelmingly concentrated in types close to $\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1}$ while $\mathbb{S}_{k+1}$ is overwhelmingly concentrated in types close to $\mathbb{T}_{N\mathbf{e}_0+(k+1)\mathbf{f}_1}$. To show this, we compute for any $\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1}$ and $j \in \{2, \dots, d\}$,

$$\begin{aligned}
\frac{\mathbb{P}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_j})}{\mathbb{P}(\mathbb{T}_{\mathbf{a}})} &= \frac{p_0^{j-1} p_j}{p_1^j} \frac{\binom{N}{\mathbf{a}+(j-1)\mathbf{e}_0-j\mathbf{e}_1+\mathbf{e}_j}}{\binom{N}{\mathbf{a}}} \\
&= \frac{p_0^{j-1} p_j}{p_1^j} \frac{\prod_{i=0}^{j-1} (a_1 - i)}{(a_j + 1) \prod_{i=1}^{j-1} (a_0 + i)} \\
&\lesssim \frac{k^j}{N^{j-1}} \\
&\leq \frac{N^{\frac{j}{4}}}{N^{j-1}} \\
&= N^{1-\frac{3j}{4}} \\
&\leq N^{-\frac{1}{2}}.
\end{aligned} \tag{2.232}$$

Summing up the exponential decay bound (2.232) over all types $\mathbb{T}_{\mathbf{a}}$ in $\mathbb{S}_k$ with $\sum_{j=2}^d a_j > 3$ proves that if $k \leq N^{1/4}$,

$$\begin{aligned}
\mathbb{P}_{\mathbb{S}_k} \left(\mathbb{S}_k \setminus \bigcup_{\substack{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1} \\ \sum_{j=2}^d a_j \leq 3}} \mathbb{T}_{\mathbf{a}} \right) &= \sum_{\substack{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1} \\ \sum_{j=2}^d a_j > 3}} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \\
&\lesssim \sum_{m>3} \sum_{\substack{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1} \\ \sum_{j=2}^d a_j = m}} N^{-\frac{m}{2}} \\
&\leq \sum_{m>3} (d-1)^m N^{-\frac{m}{2}} \\
&\lesssim \frac{1}{N^2}.
\end{aligned} \tag{2.233}$$

From Proposition 2.2.8, $b_0^k \sim N$ and $b_1^k \sim k$. It follows in the $k > N^{1/4}$ case, if $\mathbf{a} \in \mathcal{D}^k$ with $\delta_{\mathbf{a}} \leq c$, where $c = 2 \log_A(N) = o(k)$, then $\mathbf{a} \in \mathcal{D}_R$. Therefore, if we define

$$c = \begin{cases} 4 & \text{if } k < N^{1/4} \\ \log_A(N) & \text{if } N^{1/4} \leq k \leq \mathbb{E}(D), \end{cases} \tag{2.234}$$

then, as long as k is sufficiently large, R is defined on

$$\mathcal{D}^{k,c} := \{\mathbf{a} \in \mathcal{D}^k : \delta_{\mathbf{a}} < c\}.$$

If we define $\mathbb{S}'_k$ and $\mathbb{S}''_k$ as in the statement of the lemma relative to our choice of c in (2.234) then (2.233) and (2.231) prove that, for all k sufficiently large,

$$\begin{aligned} \mathbb{P}_{\mathbb{S}_k}(\mathbb{S}_k \setminus \mathbb{S}'_k) &= \sum_{\substack{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1} \\ \delta_{\mathbf{a}} > c}} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \lesssim \frac{1}{N^2} \\ \mathbb{P}_{\mathbb{S}_k}(\mathbb{S}_{k+1} \setminus \mathbb{S}''_k) &= \sum_{\substack{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1} \\ \delta_{\mathbf{a}} > c}} \mathbb{P}_{\mathbb{S}_{k+1}}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) \lesssim \frac{1}{N^2} \end{aligned}$$

so we have already proven condition (1). Furthermore,

$$\mathbb{P}_{\mathbb{S}_k}(\mathbb{S}'_k) = 1 - \mathbb{P}_{\mathbb{S}_k}(\mathbb{S}_k \setminus \mathbb{S}'_k) = 1 - O\left(\frac{1}{N^2}\right)$$

and, in perfect analogy,

$$\mathbb{P}_{\mathbb{S}_{k+1}}(\mathbb{S}''_k) = 1 - O\left(\frac{1}{N^2}\right).$$

Thus, for any event $E \subset \mathbb{S}'_k$,

$$\mathbb{P}_{\mathbb{S}'_k}(E) = \frac{\mathbb{P}_{\mathbb{S}_k}(E \cap \mathbb{S}'_k)}{\mathbb{P}_{\mathbb{S}_k}(\mathbb{S}'_k)} = \mathbb{P}_{\mathbb{S}_k}(E) \left[1 + O\left(\frac{1}{N^2}\right)\right]$$

and again in perfect analogy, for any event $E \subset \mathbb{S}''_k$,

$$\mathbb{P}_{\mathbb{S}''_k}(E) = \frac{\mathbb{P}_{\mathbb{S}_{k+1}}(E \cap \mathbb{S}''_k)}{\mathbb{P}_{\mathbb{S}_{k+1}}(\mathbb{S}''_k)} = \mathbb{P}_{\mathbb{S}_{k+1}}(E) \left[1 + O\left(\frac{1}{N^2}\right)\right]$$

Thus, we have condition (2).

It now remains to prove condition (3), namely that

$$\frac{|R(\mathbf{a}) - R_k|}{R_k} \lesssim \max\left(\frac{1}{\sqrt{N}}, \frac{1}{k}\right) \delta_{\mathbf{a}}.$$

uniformly in $\mathbf{a} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}$.

We decompose R_k as follows:

$$R_k = \frac{\mathbb{P}(\mathbb{S}''_k)}{\mathbb{P}(\mathbb{S}'_k)}$$

$$\begin{aligned}
&= \frac{\mathbb{P}\left(\bigcup_{\mathbf{a} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} \mathbb{T}_{\mathbf{a}+\mathbf{f}_1}\right)}{\mathbb{P}(\mathbb{S}'_k)} \\
&= \sum_{\mathbf{a} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} \frac{\mathbb{P}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1})}{\mathbb{P}(\mathbb{S}'_k)} \\
&= \sum_{\mathbf{a} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} \frac{R(\mathbf{a})\mathbb{P}(\mathbb{T}_{\mathbf{a}})}{\mathbb{P}(\mathbb{S}'_k)} \\
&= \sum_{\mathbf{a} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} R(\mathbf{a}) \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}). \tag{2.235}
\end{aligned}$$

We now compare individual ratios $R(\mathbf{a})$ with R_k . For any $\mathbf{a} \in \mathcal{D}^{k,c}$,

$$\begin{aligned}
|R(\mathbf{a}) - R_k| &= \left| R(\mathbf{a}) - \sum_{\mathbf{b} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} R(\mathbf{b}) \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) \right| \\
&= \left| R(\mathbf{a}) \left(\sum_{\mathbf{b} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) \right) - \sum_{\mathbf{b} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} R(\mathbf{b}) \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) \right| \\
&\leq \sum_{\mathbf{b} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} |R(\mathbf{a}) - R(\mathbf{b})| \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}). \tag{2.236}
\end{aligned}$$

Note that, for $\mathbf{a}, \mathbf{b} \in \mathcal{D}^{k,c} \subset \mathcal{D}^k$, $\mathbf{b} - \mathbf{a} \in \text{span}(\mathbf{f}_2, \dots, \mathbf{f}_d)$. Accordingly, in order to bound $|R(\mathbf{a}) - R(\mathbf{b})|$ in (2.236), we bound $\partial_{\mathbf{f}_j} R$ on $\mathcal{D}^{k,c}$ for all $j \in \{2, \dots, d\}$ as follows:

$$\begin{aligned}
|\partial_{\mathbf{f}_j} R(\mathbf{a})| &= |(j-1)R_0(\mathbf{a}) - jR_1(\mathbf{a}) + R_j(\mathbf{a})| \\
&= \left| \frac{(j-1)p_1}{p_0} \frac{1}{a_1+1} - \frac{jp_1}{p_0} \frac{a_0}{(a_1+1)^2} \right| \\
&\lesssim \frac{1}{a_1+1} + \frac{a_0}{(a_1+1)^2} \\
&\lesssim \frac{1}{k} + \frac{N}{k^2} \\
&\lesssim \frac{N}{k^2} \tag{2.237}
\end{aligned}$$

where the penultimate line above used Proposition 2.2.8 and the fact that $|\mathbf{a} - \mathbf{b}^k| = o(k)$ for $\mathbf{a} \in \mathcal{D}^{k,c}$. Note also that, because $(\mathbf{a} - \mathbf{b}^k) \cdot \mathbf{d} = 0$, for any $\mathbf{a} \in \mathcal{D}^{k,c} \subset \mathcal{D}^k$, $|a_1 - b_1^k| = O\left(\max\left(|a_2 - b_2^k|, \dots, |a_d - b_d^k|\right)\right)$. In particular, because $|b_j^k| \lesssim$

$\max\left(1, \frac{k^j}{N^{j-1}}\right)$ for all $j \in \{1, \dots, d\}$, we have that $|b_1^k| \sim \max(|b_1^k|, \dots, |b_d^k|)$ and

$$\begin{aligned}
\delta_{\mathbf{a}} &= \max_{j \in \{1, \dots, d\}} \frac{|a_j - b_j^k|}{\max\left(1, |b_j^k|^{1/2}\right)} \\
&\sim \max_{j \in \{2, \dots, d\}} \frac{|a_j - b_j^k|}{\max\left(1, |b_j^k|^{1/2}\right)} \\
&\lesssim \max_{j \in \{2, \dots, d\}} \frac{|a_j - b_j^k|}{\left[\frac{k^j}{N^{j-1}}\right]^{1/2}} \\
&\leq |\mathbf{a} - \mathbf{b}^k| \max_{j \in \{2, \dots, d\}} \frac{1}{\left[\frac{k^j}{N^{j-1}}\right]^{1/2}} \\
&\lesssim |\mathbf{a} - \mathbf{b}^k| \left\lceil \frac{k}{\sqrt{N}} \right\rceil
\end{aligned} \tag{2.238}$$

Combined, (2.236), (2.237), and (2.238) prove

$$\begin{aligned}
|R(\mathbf{a}) - R_k| &\lesssim \sum_{\mathbf{b} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} \left[\frac{N}{k^2} |\mathbf{a} - \mathbf{b}| \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) \right] \\
&\leq \sum_{\mathbf{b} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} \left[\frac{N}{k^2} \left(|\mathbf{a} - \mathbf{b}^k| + |\mathbf{b} - \mathbf{b}^k| \right) \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) \right] \\
&\lesssim \sum_{\mathbf{b} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} \left[\frac{N}{k^2} \max\left(|\mathbf{a} - \mathbf{b}^k|, |\mathbf{b} - \mathbf{b}^k|\right) \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) \right] \\
&\lesssim \sum_{\mathbf{b} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} \left[\frac{N}{k^2} \max\left(\delta_{\mathbf{a}} \left\lceil \frac{k}{\sqrt{N}} \right\rceil, \delta_{\mathbf{b}} \left\lceil \frac{k}{\sqrt{N}} \right\rceil\right) \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) \right] \\
&\leq \frac{N}{k^2} \left\lceil \frac{k}{\sqrt{N}} \right\rceil \sum_{\mathbf{b} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} \max(\delta_{\mathbf{a}}, \delta_{\mathbf{b}}) \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) \\
&\leq \frac{N}{k^2} \left\lceil \frac{k}{\sqrt{N}} \right\rceil \sum_{0 \leq n < c} \sum_{\mathbf{b} \in \text{Dev}_k(n)} \max(\delta_{\mathbf{a}}, \delta_{\mathbf{b}}) \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) \\
&= \frac{N}{k^2} \left\lceil \frac{k}{\sqrt{N}} \right\rceil \left(\sum_{\delta_{\mathbf{b}} < \delta_{\mathbf{a}}} \delta_{\mathbf{a}} \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) + \sum_{\delta_{\mathbf{a}} \leq n < c} \delta_{\mathbf{b}} \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) \right) \\
&\leq \frac{N}{k^2} \left\lceil \frac{k}{\sqrt{N}} \right\rceil \left(\delta_{\mathbf{a}} + \sum_{\delta_{\mathbf{a}} \leq n < c} (n+1) \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{b}}) \left[1 + O\left(\frac{1}{N}\right) \right] \right) \\
&\lesssim \frac{N}{k^2} \left\lceil \frac{k}{\sqrt{N}} \right\rceil \left(\delta_{\mathbf{a}} + \sum_{\delta_{\mathbf{a}} \leq n < c} (n+1) e^{-\Omega(n)} \right) \\
&\lesssim \frac{N}{k^2} \left\lceil \frac{k}{\sqrt{N}} \right\rceil \delta_{\mathbf{a}}
\end{aligned} \tag{2.239}$$

Recall that $R(\mathbf{a}) \sim \frac{N}{k}$ uniformly for $|\mathbf{a} - \mathbf{b}^k| = o(k)$, in particular for all $\mathbf{a} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}$. Thus, by (2.235) we may compute

$$\begin{aligned} R_k &= \sum_{\mathbf{a} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} R(\mathbf{a}) \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) \\ &\sim \sum_{\mathbf{a} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}} \frac{N}{k} \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) \\ &= \frac{N}{k}. \end{aligned} \tag{2.240}$$

Together, (2.239) and (2.240) prove that for a given $\mathbf{a} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}$,

$$\begin{aligned} \frac{|R(\mathbf{a}) - R_k|}{R_k} &\lesssim \frac{\frac{N}{k^2} \left\lceil \frac{k}{\sqrt{N}} \right\rceil \delta_{\mathbf{a}}}{\frac{N}{k}} \\ &\sim \frac{1}{k} \max \left(1, \frac{k}{\sqrt{N}} \right) \delta_{\mathbf{a}} \\ &= \max \left(\frac{1}{k}, \frac{1}{\sqrt{N}} \right) \delta_{\mathbf{a}}, \end{aligned}$$

thus proving condition (3) and completing the lemma. $\square$

Definition 2.8.2 (Restricted Spherical Sampling Operators). *For any $f \in L^2(G^N)$ and any (integer) $0 \leq k \leq \mathbb{E}(D)$, define the restricted spherical averaging operators by*

$$\begin{aligned} S'_k f(x) &:= \mathbb{E}_{\mathbb{S}'_k} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right) \text{ and} \\ S''_k f(x) &:= \mathbb{E}_{\mathbb{S}''_k} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right). \end{aligned}$$

Lemma 2.8.3. *The operators $\{S'_0, \dots, S'_{\lfloor \mathbb{E}(D) \rfloor - 1}\}$ and $\{S'_0, \dots, S'_{\lfloor \mathbb{E}(D) \rfloor - 1}\}$ are self-adjoint $L^2(G^N)$ contractions that are simultaneously diagonalizable by $\widehat{G}^N$. Moreover, $\|S'_k - S_k\|_2, \|S''_k - S_{k+1}\|_2 \lesssim \frac{1}{N^2}$ uniformly in $0 \leq k \leq \mathbb{E}(D) - 1$.*

Proof. Because all the operators in question can be viewed as convex combinations of tensor products of $\{P_0, \dots, P_d\}$, it is straightforward to verify they are

- self-adjoint,

- $L^2(G^N)$ contractions,
- and diagonalizable by $\widehat{G}^N$.

It only remains to show that $\|S'_k - S_k\|_2, \|S''_k - S_{k+1}\|_2 \lesssim \frac{1}{N^2}$. Because the arguments for $\|S'_k - S_k\|_2$ and $\|S''_k - S_{k+1}\|_2$ are essentially identical, it will suffice to show $\|S'_k - S_k\|_2 \lesssim \frac{1}{N^2}$. To this end we compute

$$\begin{aligned}
& S_k f(x) - S'_k f(x) \\
&= \mathbb{E}_{\mathbb{S}_k} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right) - \mathbb{E}_{\mathbb{S}'_k} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right) \\
&= \sum_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k} \mathbb{E}_{\mathbb{T}_{\mathbf{a}}} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right) \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) - \sum_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}'_k} \mathbb{E}_{\mathbb{T}_{\mathbf{a}}} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right) \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) \\
&= \sum_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k \setminus \mathbb{S}'_k} \mathbb{E}_{\mathbb{T}_{\mathbf{a}}} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right) \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) + \\
&\quad \sum_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}'_k} \mathbb{E}_{\mathbb{T}_{\mathbf{a}}} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right) \left(\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) - \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) \right). \tag{2.241}
\end{aligned}$$

We will use Minkowski's inequality as in (2.221) in the proof of Theorem 2.1.5 to bound the summands in (2.241) in $L^2(G^N)$. First we compute

$$\begin{aligned}
\left\| \sum_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k \setminus \mathbb{S}'_k} \mathbb{E}_{\mathbb{T}_{\mathbf{a}}} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right) \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \right\|_2 &\leq \sum_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}_k \setminus \mathbb{S}'_k} \|f\|_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \\
&= \|f\|_2 \mathbb{P}_{\mathbb{S}_k} \left(\bigcup_{\mathbb{S}_k \setminus \mathbb{S}'_k} \mathbb{T}_{\mathbf{a}} \right) \\
&= \|f\|_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{S}_k \setminus \mathbb{S}'_k) \\
&= \|f\|_2 \frac{1}{N^2}. \tag{2.242}
\end{aligned}$$

and analogously

$$\begin{aligned}
&\left\| \sum_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}'_k} \mathbb{E}_{\mathbb{T}_{\mathbf{a}}} \left(\left[\bigotimes_{n=1}^N P_{D_n} \right] f(x) \right) \left(\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) - \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) \right) \right\|_2 \\
&\leq \sum_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}'_k} \|f\|_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \left| 1 - \frac{\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}})}{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}})} \right|
\end{aligned}$$

$$\begin{aligned}
&\lesssim \|f\|_2 \sum_{\mathbb{T}_{\mathbf{a}} \subset \mathbb{S}'_k} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \frac{1}{N^2} \\
&\leq \|f\|_2 \frac{1}{N^2}.
\end{aligned} \tag{2.243}$$

Taking L^2 norms and plugging (2.242) and (2.243) back into (2.241) yields

$$\|S_k f(x) - S'_k f(x)\|_2 \lesssim \|f\|_2 \frac{1}{N^2}. \tag{2.244}$$

Because, as alluded to above, the same argument works for S''_k , this completes the proof. $\square$

For the remainder of Chapter 2 we will restrict our attention to the $d = 2$ case. We emphasize that, while this restriction is less than ideal, it handles the bulk of the hurdles encountered in generalizing the result from $d = 1$ as in [8], [3], and [7] to arbitrary $d \in \mathbb{N}$. However, if we allow d to exceed 2, we encounter a number of frustrating (if generally superficial) hassles that result from the $k = o\left(N^{\frac{J-1}{J}}\right)$ for a give $J \in \{2, \dots, d\}$ case which necessitates handling $j \leq J$ and $j > J$ separately, reminiscent of the labor intensive Proposition 2.3.2.

The $d = 2$ case provides a happy medium which illustrates how one would split up $\{1, \dots, d\}$ as described above while avoiding a torrent of largely banal special cases. We will return to this restriction in Chapter 4, however we will just mention that, while this proof has not been extended to $d > 2$, the skeleton for such an argument is here. Moreover, we shall try to give some suggestions concerning this generalization along the way for the curious reader, including leaving the arguments in terms of general d when we do not rely on the fact that $d = 2$. We heartily encourage intrepid readers to attempt the generalization to $d > 2$ independently.

Lemma 2.8.4. *If $d = 2$, for $j \in \{0, 1, 2\}$,*

$$\mathbb{E}_{\mathbb{S}_k} D^j \sim \frac{k^j}{N^{j-1}}.$$

Proof. We decompose

$$\begin{aligned}
\mathbb{E}_{\mathbb{S}_k} D^j &= \sum_{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1}} a_j \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \\
&= \sum_{n \geq 0} \sum_{\mathbf{a} \in \text{Dev}_k(n)} a_j \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \\
&= b_j^k + \sum_{n \geq 0} \sum_{\mathbf{a} \in \text{Dev}_k(n)} (a_j - b_j^k) \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}). \tag{2.245}
\end{aligned}$$

Applying Corollary 2.3.11 to the error term in (2.245) yields

$$\begin{aligned}
\left| \sum_{n \geq 0} \sum_{\mathbf{a} \in \text{Dev}_k(n)} (a_j - b_j^k) \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \right| &= \sum_{n \geq 0} \sum_{\mathbf{a} \in \text{Dev}_k(n)} |a_j - b_j^k| \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \\
&\leq \sum_{n \geq 0} (n+1) \max \left(1, |b_j^k|^{1/2} \right) e^{-\Omega(n)} \\
&\lesssim \max \left(1, |b_j^k|^{1/2} \right). \tag{2.246}
\end{aligned}$$

Thus, if $k = \Omega \left(N^{\frac{j-1}{j}} \right)$, Proposition 2.2.8 implies $b_j^k = \Theta \left(\frac{k^j}{N^{j-1}} \right)$ and $\max \left(1, |b_j^k|^{1/2} \right) = o \left(\frac{k^j}{N^{j-1}} \right)$ so (2.245) implies

$$\mathbb{E}_{\mathbb{S}_k} D^j \sim \frac{k^j}{N^{j-1}}. \tag{2.247}$$

Because the condition $k = \Omega \left(N^{\frac{j-1}{j}} \right)$ is vacuous of $j \in \{0, 1\}$, we need only worry about the case in which $j \geq 2$.

At this stage we invoke the assumption (unused thus far in the proof) that $d = 2$. Our remaining case is now $j = 2, k = o(N^{1/2})$. We begin with the decomposition

$$\begin{aligned}
\mathbb{E}_{\mathbb{S}_k} D^2 &= \sum_{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^3} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \\
&= \sum_{0 \leq a_2 \leq \frac{k}{2}} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + a_2\mathbf{f}_2}) \\
&= \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + \mathbf{f}_2}) + \sum_{2 \leq a_2 \leq \frac{k}{2}} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + a_2\mathbf{f}_2}) \tag{2.248}
\end{aligned}$$

Next we note that, for any $\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^3$ such that $\mathbf{a} + \mathbf{f}_2 \in \mathcal{D}^k \cap \mathbb{N}^3$, we compute

$$\frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_2})}{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}})} = \frac{p_2 p_0}{p_1^2} \frac{a_1(a_1 - 1)}{(a_0 + 1)(a_2 + 1)} \quad (2.249)$$

Because $k = o(N)$ (in particular, k is bounded away from N), we have $a_0 > \frac{N}{2}$ for $\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^{d+1}$. Thus, if k is bounded by a sufficiently small multiple of $\sqrt{N}$, which we are free to assume,

$$\frac{p_2 p_0}{p_1^2} \frac{a_1(a_1 - 1)}{a_0 + 1} \leq \frac{p_2 p_0}{p_1^2} \frac{k^2}{N/2} \leq \frac{1}{2} \quad (2.250)$$

for any $\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^3$.

We first bound the tail

$$\sum_{2 \leq a_2 \leq \frac{k}{2}} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + a_2\mathbf{f}_2}). \quad (2.251)$$

To do so, we apply (2.250) inductively to bound the summands in (2.251) from above by finding that for any $2 \leq a_2 \leq \frac{k}{2}$,

$$\begin{aligned} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + a_2\mathbf{f}_2}) &\leq a_2 \frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + (a_2-1)\mathbf{f}_2})}{2(a_2 + 1)} \\ &\leq \frac{1}{2} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + (a_2-1)\mathbf{f}_2}) \\ &\leq \frac{1}{2} \frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + \mathbf{f}_2})}{2^{a_2-1}} \\ &= \frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + \mathbf{f}_2})}{2^{a_2}}. \end{aligned} \quad (2.252)$$

Plugging (2.252) back into (2.251) proves

$$\sum_{2 \leq a_2 \leq \frac{k}{2}} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + a_2\mathbf{f}_2}) \leq \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + \mathbf{f}_2}). \quad (2.253)$$

Because (2.251) is non-negative, (2.253) implies in conjunction with (2.248) that

$$\mathbb{E}_{\mathbb{S}_k} D^2 \sim \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + \mathbf{f}_2}). \quad (2.254)$$

Using (2.249), we estimate

$$\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + \mathbf{f}_2}) = \frac{p_2 p_0}{p_1^2} \frac{(k)(k-1)}{N-k+1} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1})$$

$$\sim \frac{k^2}{N} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1}). \quad (2.255)$$

We estimate $\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1})$ in turn as follows:

$$\begin{aligned} 1 &= \sum_{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^3} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \\ &= \sum_{0 \leq a_2 \leq \frac{k}{2}} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1+a_2\mathbf{f}_2}) \\ &= \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1}) + \sum_{1 \leq a_2 \leq \frac{k}{2}} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1+a_2\mathbf{f}_2}). \end{aligned} \quad (2.256)$$

We observe that the error term $\sum_{1 \leq a_2 \leq \frac{k}{2}} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1+a_2\mathbf{f}_2})$ is non-negative so we may bound it below by 0. Also, we note that (2.249) and (2.250) immediately imply that for all $\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^3$, $\frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_2})}{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}})} \leq \frac{1}{2}$ and therefore

$$\sum_{1 \leq a_2 \leq \frac{k}{2}} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1+a_2\mathbf{f}_2}) \leq \sum_{1 \leq a_2 \leq \frac{k}{2}} \frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1})}{2^{a_2}} \leq \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1}). \quad (2.257)$$

Combining (2.256) and (2.257), we see that $\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1}) \in [\frac{1}{2}, 1]$ and, in particular,

$$\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1}) \sim 1$$

and therefore

$$\mathbb{E}_{\mathbb{S}_k}(D^2) \sim \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{N\mathbf{e}_0+k\mathbf{f}_1+\mathbf{e}_1}) \sim \frac{k^2}{N}$$

as claimed. $\square$

Remark 2.8.5. *Lemma 2.8.4 generalizes naturally for $j > 2$. Extending the Lemma 2.8.4 is not only a crucial point of entry to the extension of our argument to $d > 2$, it is also a useful pedagogical point of entry as it provides some familiarity with the subtleties of the $d > 2$ case but is more straightforward than the rest of the generalization.*

Now we come to the main result of the section. This will likely prove one of the more involved pieces to generalize to $d > 2$ because the necessary analogous

statement for $d > 2$ is not immediately clear and must be determined by the needs of the lemmas in §2.9.

Proposition 2.8.6. *For any $0 \leq k \leq \mathbb{E}(D) - 1$,*

$$|\mathbb{P}_{\mathbb{S}_{k+1}}(D^j) - \mathbb{P}_{\mathbb{S}_k}(D^j)| \lesssim \begin{cases} 1 & \text{if } j \in \{0, 1\} \text{ and} \\ \frac{k}{N} & \text{if } j = 2 \end{cases}$$

Proof. We begin with some reductions and rearrangements. By definition, $k = \mathbb{E}_{\mathbb{S}_k}(D^1) + 2\mathbb{E}_{\mathbb{S}_k}(D^2)$ and thus

$$|\mathbb{E}_{\mathbb{S}_{k+1}}(D^1) - \mathbb{E}_{\mathbb{S}_k}(D^1)| \leq 1 + 2|\mathbb{E}_{\mathbb{S}_{k+1}}(D^2) - \mathbb{E}_{\mathbb{S}_k}(D^2)|.$$

Therefore, the desired bound $|\mathbb{E}_{\mathbb{S}_{k+1}}(D^1) - \mathbb{E}_{\mathbb{S}_k}(D^1)| \lesssim 1$ follows immediately from the bound $|\mathbb{E}_{\mathbb{S}_{k+1}}(D^2) - \mathbb{E}_{\mathbb{S}_k}(D^2)| \lesssim \frac{k}{N}$. Similarly, $N = \mathbb{E}_{\mathbb{S}_k}(D^0) + \mathbb{E}_{\mathbb{S}_k}(D^1) + \mathbb{E}_{\mathbb{S}_k}(D^2)$ so

$$|\mathbb{E}_{\mathbb{S}_{k+1}}(D^0) - \mathbb{E}_{\mathbb{S}_k}(D^0)| \leq |\mathbb{E}_{\mathbb{S}_{k+1}}(D^1) - \mathbb{E}_{\mathbb{S}_k}(D^1)| + |\mathbb{E}_{\mathbb{S}_{k+1}}(D^2) - \mathbb{E}_{\mathbb{S}_k}(D^2)|.$$

Therefore the desired bound $|\mathbb{E}_{\mathbb{S}_{k+1}}(D^0) - \mathbb{E}_{\mathbb{S}_k}(D^0)| \lesssim 1$ follows immediately from the bound $|\mathbb{E}_{\mathbb{S}_{k+1}}(D^2) - \mathbb{E}_{\mathbb{S}_k}(D^2)| \lesssim \frac{k}{N}$. Thus it only remains to prove the latter bound.

Also, note that $\mathbb{E}_{\mathbb{T}_a}(D_2) = a_2$ so, splitting $\mathbb{S}_k$ and $\mathbb{S}_{k+1}$ into their types, the total probability law then dictates

$$\begin{aligned} \mathbb{E}_{\mathbb{S}_k}(D^2) &= \sum_{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^3} \mathbb{E}_{\mathbb{T}_a}(D_2) \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_a) = \sum_{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^3} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_a) \\ \mathbb{E}_{\mathbb{S}_{k+1}}(D^2) &= \sum_{\mathbf{a} \in \mathcal{D}^{k+1} \cap \mathbb{N}^3} \mathbb{E}_{\mathbb{T}_a}(D_2) \mathbb{P}_{\mathbb{S}_{k+1}}(\mathbb{T}_a) = \sum_{\mathbf{a} \in \mathcal{D}^{k+1} \cap \mathbb{N}^3} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{k-2b,b}) \end{aligned} \tag{2.258}$$

The purpose of defining $R(\mathbf{a})$ at the beginning of §2.8 is to allow us to compare $\mathbb{E}_{\mathbb{S}_k}(D^2)$ and $\mathbb{E}_{\mathbb{S}_{k+1}}(D^2)$ by decomposing as in (2.258) and matching types corresponding to the same ‘ a_2 ’ value. The purpose of defining and establishing the properties of R_k in Lemma 2.8.1 is to focus on only the relevant summands in (2.258). By Lemma 2.8.1 we have

$$\begin{aligned}
\mathbb{E}_{\mathbb{S}_k}(D^2) &= \sum_{\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^3} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \\
&= \sum_{\delta_{\mathbf{a}} < c} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) + \sum_{\delta_{\mathbf{a}} \geq c} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \\
&\leq \sum_{\delta_{\mathbf{a}} < c} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) + k \sum_{\delta_{\mathbf{a}} \geq c} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \\
&= \sum_{\delta_{\mathbf{a}} < c} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) + O(k \mathbb{P}_{\mathbb{S}_k}(\mathbb{S}_k \setminus \mathbb{S}'_k)) \\
&= \sum_{\delta_{\mathbf{a}} < c} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) + O\left(\frac{k}{N^2}\right) \\
&= \sum_{\delta_{\mathbf{a}} < c} a_2 \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) + O\left(\frac{k}{N}\right) \\
&= \sum_{\delta_{\mathbf{a}} < c} a_2 \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) \left(1 + O\left(\frac{1}{N^2}\right)\right) + O\left(\frac{k}{N}\right) \\
&= \sum_{\delta_{\mathbf{a}} < c} a_2 \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) \left(1 + O\left(\frac{1}{N}\right)\right) + O\left(\frac{k}{N}\right) \\
&= \sum_{\delta_{\mathbf{a}} < c} a_2 \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) + O\left(\frac{k}{N}\right)
\end{aligned}$$

Computing analogously in $\mathbb{S}''_k$, we get

$$\begin{aligned}
\mathbb{E}_{\mathbb{S}_k}(D^2) &= \sum_{\delta_{\mathbf{a}} \leq c} a_2 \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) + O\left(\frac{k}{N}\right) \\
\mathbb{E}_{\mathbb{S}_{k+1}}(D^2) &= \sum_{\delta_{\mathbf{a}} \leq c} a_2 \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) + O\left(\frac{k}{N}\right).
\end{aligned} \tag{2.259}$$

We use (2.259) to establish our difference bound:

$$\begin{aligned}
|\mathbb{E}_{\mathbb{S}_{k+1}}(D^2) - \mathbb{E}_{\mathbb{S}_k}(D^2)| &= \left| \sum_{\delta_{\mathbf{a}} < c} a_2 (\mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) - \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}})) \right| + O\left(\frac{k}{N}\right) \\
&= \left| \sum_{\substack{\delta_{\mathbf{a}} < c \\ a_2 \geq 1}} a_2 (\mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) - \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}})) \right| + O\left(\frac{k}{N}\right) \\
&= \left| \sum_{\substack{\delta_{\mathbf{a}} < c \\ a_2 \geq 1}} a_2 (\mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) - \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}})) - b_2^k + b_2^k \right| + O\left(\frac{k}{N}\right)
\end{aligned}$$

$$\begin{aligned}
&= \left| \sum_{\substack{\delta_{\mathbf{a}} < c \\ a_2 \geq 1}} a_2 \left(\mathbb{P}_{\mathbb{S}_k''}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) - \mathbb{P}_{\mathbb{S}_k'}(\mathbb{T}_{\mathbf{a}}) \right) - \right. \\
&\quad \left. b_2^k \sum_{\substack{\delta_{\mathbf{a}} < c \\ a_2 \geq 1}} \mathbb{P}_{\mathbb{S}_k''}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) + b_2^k \sum_{\substack{\delta_{\mathbf{a}} < c \\ a_2 \geq 1}} \mathbb{P}_{\mathbb{S}_k'}(\mathbb{T}_{\mathbf{a}}) \right| + O\left(\frac{k}{N}\right) \\
&\leq \sum_{\substack{\delta_{\mathbf{a}} < c \\ a_2 \geq 1}} |a_2 - b_2^k| |\mathbb{P}_{\mathbb{S}_k''}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}) - \mathbb{P}_{\mathbb{S}_k'}(\mathbb{T}_{\mathbf{a}})| + O\left(\frac{k}{N}\right) \\
&= \sum_{\substack{\delta_{\mathbf{a}} < c \\ a_2 \geq 1}} |a_2 - b_2^k| \mathbb{P}_{\mathbb{S}_k'}(\mathbb{T}_{\mathbf{a}}) \left| \frac{\mathbb{P}_{\mathbb{S}_k''}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1})}{\mathbb{P}_{\mathbb{S}_k'}(\mathbb{T}_{\mathbf{a}})} - 1 \right| + O\left(\frac{k}{N}\right) \\
&= \sum_{\substack{\delta_{\mathbf{a}} < c \\ a_2 \geq 1}} |a_2 - b_2^k| \mathbb{P}_{\mathbb{S}_k'}(\mathbb{T}_{\mathbf{a}}) \left| \frac{\mathbb{P}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1})}{\mathbb{P}(\mathbb{S}_k'')} \frac{\mathbb{P}(\mathbb{S}_k')}{\mathbb{P}(\mathbb{T}_{\mathbf{a}})} - 1 \right| + O\left(\frac{k}{N}\right) \\
&= \sum_{\substack{\delta_{\mathbf{a}} < c \\ a_2 \geq 1}} |a_2 - b_2^k| \mathbb{P}_{\mathbb{S}_k'}(\mathbb{T}_{\mathbf{a}}) \left| \frac{R(\mathbf{a})}{R_k} - 1 \right| + O\left(\frac{k}{N}\right) \\
&= \sum_{\substack{\delta_{\mathbf{a}} < c \\ a_2 \geq 1}} |a_2 - b_2^k| \mathbb{P}_{\mathbb{S}_k'}(\mathbb{T}_{\mathbf{a}}) \frac{|R(\mathbf{a}) - R_k|}{R_k} + O\left(\frac{k}{N}\right) \quad (2.260)
\end{aligned}$$

To bound (2.260), we split into cases for small k and large k . If $k = o(\sqrt{N})$, Lemma 2.2.8 shows that $b_2^k = O(1)$ so $a_2, |a_2 - b_2^k| \lesssim \delta_{\mathbf{a}}$ for all relevant, i.e. $a_2 \geq 1$. Also, $\frac{1}{k} \geq \frac{1}{\sqrt{N}}$ so we use condition (3) in Lemma 2.8.1 to bound $\frac{|R(\mathbf{a}) - R_k|}{R_k}$ by $\frac{1}{k} \delta_{\mathbf{a}} \lesssim \frac{a_2}{k}$. For notational ease we temporarily adopt the notation

$$b := b_2^k + c \max\left(1, |b_2^k|^{1/2}\right)$$

so that as long as k is sufficiently large, for any $\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^3$ we have the equivalence

$$a_2 < b \iff \delta_{\mathbf{a}} < c.$$

Note that (here we use our assumption that $d = 2$) each $\mathbf{a} \in \mathcal{D}^{k,c} \cap \mathbb{N}^{d+1}$ can be written in the form $\mathbf{a} = N\mathbf{e}_0 + k\mathbf{f}_1 + a_2\mathbf{f}_2$ for some $a_2 < b$. Now (2.260) implies

$$|\mathbb{E}_{\mathbb{S}_{k+1}}(D^2) - \mathbb{E}_{\mathbb{S}_k}(D^2)| = \sum_{\substack{\delta_{\mathbf{a}} < c \\ a_2 \geq 1}} |a_2 - b_2^k| \mathbb{P}_{\mathbb{S}_k'}(\mathbb{T}_{\mathbf{a}}) \frac{|R(\mathbf{a}) - R_k|}{R_k} + O\left(\frac{k}{N}\right)$$

$$\begin{aligned}
&\lesssim \sum_{\substack{\delta_{\mathbf{a}} \leq c \\ a_2 \geq 1}} a_2 \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) \frac{1}{k} \delta_{\mathbf{a}} + O\left(\frac{k}{N}\right) \\
&\lesssim \sum_{\substack{\delta_{\mathbf{a}} \leq c \\ a_2 \geq 1}} (a_2)^2 \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) \frac{1}{k} + O\left(\frac{k}{N}\right) \\
&= \frac{1}{k} \sum_{1 \leq a_2 < b} (a_2)^2 \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + a_2\mathbf{f}_2}) \tag{2.261}
\end{aligned}$$

Recall from (2.249) that

$$\frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_2})}{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}})} = \frac{p_2 p_0}{p_1^2} \frac{a_1(a_1 - 1)}{(a_0 + 1)(a_2 + 1)}$$

for any $\mathbf{a} \in \mathcal{D}^k \cap \mathbb{N}^3$ such that $\mathbf{a} + \mathbf{f}_2 \in \mathcal{D}^k \cap \mathbb{N}^3$ and therefore we have

$$\begin{aligned}
\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + \mathbf{f}_2}) &= \frac{p_2 p_0}{p_1^2} \frac{k(k-1)}{N-k+1} \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + \mathbf{f}_2}) \\
&\leq \frac{p_2 p_0}{p_1^2} \frac{k(k-1)}{N-k+1} \\
&\lesssim \frac{k^2}{N}. \tag{2.262}
\end{aligned}$$

Furthermore, if k is bounded by a sufficiently small multiple of $\sqrt{N}$, (2.249) that for all $1 \leq a_2 < b-1$,

$$\frac{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_2})}{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}})} \leq \frac{1}{2}, \tag{2.263}$$

which inductively provides the bound

$$\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{N\mathbf{e}_0 + k\mathbf{f}_1 + a_2\mathbf{f}_2}) \lesssim \frac{k^2}{N} 2^{-a_2} \tag{2.264}$$

in conjunction with (2.262). Plugging (2.264) back into (2.261) yields

$$|\mathbb{E}_{\mathbb{S}_{k+1}}(D^2) - \mathbb{E}_{\mathbb{S}_k}(D^2)| \lesssim \frac{1}{k} \sum_{1 \leq a_2 < b} (a_2)^2 \frac{k^2}{N} 2^{-a_2} \lesssim \frac{k}{N} \tag{2.265}$$

thus completing the bound for k below a sufficiently small multiple of $\sqrt{N}$, i.e. $k = o(\sqrt{N})$.

In the $\Omega(\sqrt{N}) \leq k \leq \mathbb{E}(D)$ case, $\frac{1}{k} \lesssim \frac{1}{\sqrt{N}}$ so we use condition 3 in Lemma 2.8.1 to bound $\frac{|R(\mathbf{a}) - R_k|}{R_k}$ up to a constant by $\frac{1}{\sqrt{N}} \delta_{\mathbf{a}}$ and return to (2.260)

$$\begin{aligned}
|\mathbb{E}_{\mathbb{S}_k}(D^2) - \mathbb{E}_{\mathbb{S}_{k+1}}(D^2)| &\lesssim \sum_{\delta_{\mathbf{a}} < c} |a_2 - b_2^k| \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) \frac{|R(\mathbf{a}) - R_k|}{R_k} + O\left(\frac{k}{N}\right) \\
&\lesssim \frac{1}{\sqrt{N}} \sum_{\delta_{\mathbf{a}} < c} |a_2 - b_2^k| \delta_{\mathbf{a}} \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) + O\left(\frac{k}{N}\right) \\
&\lesssim \frac{1}{\sqrt{N}} \sum_{\delta_{\mathbf{a}} < c} \left(\delta_{\mathbf{a}} \sqrt{b_2^k}\right) \delta_{\mathbf{a}} \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) + O\left(\frac{k}{N}\right) \\
&\lesssim \frac{\sqrt{k^2/N}}{\sqrt{N}} \sum_{\delta_{\mathbf{a}} < c} (\delta_{\mathbf{a}})^2 \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) + O\left(\frac{k}{N}\right) \\
&\leq \frac{k}{N} \sum_{0 \leq n < c} n^2 \sum_{\mathbb{T}_{\mathbf{a}} \subset \text{Dev}_k(n)} \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}) + O\left(\frac{k}{N}\right) \\
&\leq \frac{k}{N} \sum_{0 \leq n < c} n^2 \left[\sum_{\mathbb{T}_{\mathbf{a}} \subset \text{Dev}_k(n)} \left(1 + O\left(\frac{1}{N}\right)\right) \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \right] + O\left(\frac{k}{N}\right) \\
&\lesssim \frac{k}{N} \sum_{0 \leq n < c} n^2 \left[\sum_{\mathbb{T}_{\mathbf{a}} \subset \text{Dev}_k(n)} \mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}) \right] + O\left(\frac{k}{N}\right) \\
&\leq \frac{k}{N} \sum_{0 \leq n < c} n^2 \mathbb{P}_{\mathbb{S}_k}(\text{Dev}_k(n)) + O\left(\frac{k}{N}\right) \\
&\lesssim \frac{k}{N} \sum_{0 \leq n < c} n^2 e^{-\Omega(n)} + O\left(\frac{k}{N}\right) \tag{2.266} \\
&\lesssim \frac{k}{N}.
\end{aligned}$$

were we used Corollary 2.3.11 to pass to (2.266). This completes the proof. $\square$

2.9 Spectral Smoothness and Decay

At a crucial step in §2.7, we used Proposition 2.9.1 below.

Proposition 2.9.1. *For any $\|\xi\|_{\text{freq}} = r, 1 \leq k < \mathbb{E}(D)$, we have the difference bound*

$$|\widehat{S''_k}(\xi) - \widehat{S'_k}(\xi)| \lesssim \frac{r}{N} e^{-\Omega(rk/N)}$$

Heuristically, Proposition 2.9.1 describes the two key spectral properties of

the spherical sampling operators (in some cases using the restricted operators as proxies), namely that their Fourier transforms

1. decay exponentially with respect to radius and frequency, tempered by a factor of dimension, and
2. vary smoothly in radius for lower frequencies.

To understand (1), observe that a uniform sample of k component of an eigenvector with r non-constant components is expected to see $\frac{rk}{N}$ of the non-constant components. If each sampled non-constant component contributes some factor in $(-\lambda, \lambda)$ to the Fourier transform where $\lambda \in (0, 1)$, intuitively the Fourier transform should be bounded by something like $\lambda^{rk/N} = e^{-\Omega(rk/n)}$. If $d = 1$, a uniform sample over $\mathbb{S}_k$ is the same as applying P_1 to a uniformly random sample of k components. Each such P_1 operator then contributes a factor in $(-\lambda, \lambda)$ to the Fourier transform where λ is the spectral bound in the definition of a radial space in §2.1.1 (Definition 2.1.1, axiom (12)). Although allowing $d > 1$ complicates issues, this heuristic from the $d = 1$ case is a good one and in fact the approximation of a spherical sample by an application of P_1 to a uniformly random choice of k components is often practical for $k = o(\sqrt{N})$.

To understand (2), we will again borrow a heuristic from the $d = 1$ case. In that setting, the event $\mathbb{S}_k$ gives probability $\frac{k}{N}$ while the event $\mathbb{S}_{k+1}$ gives probability $\frac{k+1}{N}$ to the event $\{D_n = 1\}$ for a given n . Therefore, the change in these probabilities, i.e. $\frac{1}{N}$, is $\frac{1}{k}$ times the probability given by $\mathbb{S}_k$, i.e. $\frac{k}{N}$. Similarly, the difference in probability from $\mathbb{S}_k$ to $\mathbb{S}_{k+1}$ of the event $\{D_n = 0\}$ is $\frac{1}{N-k} \sim \frac{1}{N}$ because $k \leq \mathbb{E}(D) = \mathbb{E}(D_1)N$ and $\mathbb{E}(D_1) \in (0, 1)$. Moreover, because the expected number of applications of P_1 to the r non-constant components of ξ is $\frac{rk}{N}$ as we saw in our heuristic for (1), the difference in probability that $D_n = 1$ for a given choice of approximately $\frac{rk}{N}$ components n should be roughly $\left(1 + \frac{1}{k}\right)^{rk/N} - 1 \approx \frac{r}{N}$ and the difference in probability that $D_n = 0$ for a given choice of approximately $r \left(1 - \frac{rk}{N}\right)$

components n should be roughly $(1 + \frac{1}{N})^{r-rk/N} - 1 \leq (1 + \frac{1}{N})^r - 1 \approx \frac{r}{N}$. In short, $\mathbb{S}_k$ and $\mathbb{S}_{k+1}$ weight each operator applied to a non-constant component (which are the only relevant operators because $P_j(\mathbf{1}) = \mathbf{1}$ for all j , even when $d > 1$) similarly if $\frac{r}{N}$ is small, differing by a factor of roughly $\frac{r}{N}$. Recall from §2.8 that when $d = 2$ (this should hold for $d > 1$ in general, although we have not proven that) the events $\mathbb{S}_k$ and $\mathbb{S}_{k+1}$ both give probability $\Theta(\frac{k}{N})$ to the event $\{D_n = 1\}$ for a given n and those probabilities change by $O(\frac{1}{N})$. Similarly, $\mathbb{S}_k$ and $\mathbb{S}_{k+1}$ both give probability $\Theta(1)$ to the event $\{D_n = 1\}$ for a given n and those probabilities also change by $O(\frac{1}{N})$. Although dealing with $d > 1$ presents several difficulties (with which we will deal in the current section) a similar calculation to the $d = 1$ heuristic above provides the appropriate smoothness bounds.

Our first lemma and corollary quantify (1) above. Before that, however, we introduce some notation that will prove useful throughout the section.

Notation 4 (Random Parameter). *We adopt the notation $\mathbf{D}$ for the $\{0, 1, 2\}^N$ valued random variable $(D_1, \dots, D_N)$.*

Notation 5 (Truncated Vectors and Random Sums). *We adopt the notation*

$$D_{\leq r}^j := \sum_{n=1}^r D_n^j \text{ and}$$

$$D_{\leq r} := \sum_{n=1}^r D_n.$$

Similarly, for any vector $\mathbf{x} = (x_1, \dots, x_N)$ (including the random parameter vector $\mathbf{D}$), we adopt the notation

$$\mathbf{x}_{\leq r} = (x_1, \dots, x_r).$$

Variants of this notation, such as the subscripts “ $< r$,” “ $\geq r$,” or “ $> r$ ” are to be interpreted analogously.

Notation 6 (Truncated Spheres and Types). *For any $r \leq N$ and any $\mathbf{a} \in \mathbb{N}^d$*

with $\Sigma \mathbf{a} = r$, we adopt the notation

$$\mathbb{S}_k^r = \{D_{\leq r} = k\} \text{ and } \mathbb{T}_{\mathbf{a}}^r = \{(D_{\leq r}^0, \dots, D_{\leq r}^d) = \mathbf{a}\}.$$

Because our spectral decay condition will rely on a large portion of the P_1 components of the random operator $[\otimes_{n=1}^N P_{D_n}]$ coinciding with the non-constant components of a given eigenvector ξ , we first show that a low rate of coincidence is sufficiently unlikely.

Definition 2.9.2 (Log Truncated Type Mass Function and Domains). *For a fixed type $\mathbb{T}_{\mathbf{b}} \subset G^N$ and $r \in \{0, \dots, N\}$, we designate the domain*

$$\mathcal{D}_r^{\mathbf{b}} := \left\{ \mathbf{a} \in \prod_{j=0}^2 (-1, \min(r+2, b_j+1)) : \Sigma \mathbf{a} = r \right\}$$

for the function

$$L_{\mathbf{b}}(\mathbf{a}) := \left[\Lambda(r+1) - \sum_{j=0}^2 \Lambda(a_j+1) \right] + \left[\Lambda(N-r+1) - \sum_{j=0}^2 \Lambda(b_j-a_j+1) \right] - \left[\Lambda(N+1) - \sum_{j=0}^2 \Lambda(b_j+1) \right].$$

Moreover, for a fixed $0 \leq k < \mathbb{E}(D)$ we further define the domain

$$\mathcal{D}_r^k := \bigcup_{\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}_k^r} \mathcal{D}_r^{\mathbf{b}}.$$

Note that if $\mathbf{a} \in \mathbb{N}^3$, then

$$L_{\mathbf{b}}(\mathbf{a}) = \ln \left[\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \right], \tag{2.267}$$

hence the name “log truncated type mass function.” Lastly, before we begin proving the requisite lemmas for Proposition 2.9.1, we define

$$\mathbf{f}_1 := \mathbf{e}_1 - \mathbf{e}_0.$$

Notice that the definition of $\mathbf{f}_1$ doesn’t match our notation

$$\mathbf{f}_j = (j-1)\mathbf{e}_0 - j\mathbf{e}_1 + \mathbf{e}_j$$

for $j > 1$ from §2.3, however it is more useful for $j = 1$.

Lemma 2.9.3. *There exists a (small) constant $c > 0$ such that for any $r \in \{0, \dots, N\}$ and $k \in \{0, \dots, 2N\}$,*

$$\mathbb{P}_{\mathbb{S}'_k} \left(D_{\leq r}^1 \leq c \frac{rk}{N} \right) \lesssim e^{-\Omega(rk/N)} \text{ and}$$

$$\mathbb{P}_{\mathbb{S}''_k} \left(D_{\leq r}^1 \leq c \frac{rk}{N} \right) \lesssim e^{-\Omega(rk/N)}.$$

Proof. We decompose $\mathbb{S}'_k$ and $\mathbb{S}''_k$, computing

$$\begin{aligned} \mathbb{P}_{\mathbb{S}'_k} \left(D_{\leq r}^1 \leq c \frac{rk}{N} \right) &= \sum_{\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}'_k} \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{b}}) \mathbb{P}_{\mathbb{T}_{\mathbf{b}}} \left(D_{\leq r}^1 \leq c \frac{rk}{N} \right) \text{ and} \\ \mathbb{P}_{\mathbb{S}''_k} \left(D_{\leq r}^1 \leq c \frac{rk}{N} \right) &= \sum_{\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}''_k} \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{b}}) \mathbb{P}_{\mathbb{T}_{\mathbf{b}}} \left(D_{\leq r}^1 \leq c \frac{rk}{N} \right) \end{aligned}$$

so it will suffice to prove the uniform bound

$$\mathbb{P}_{\mathbb{T}_{\mathbf{b}}} \left(D_{\leq r}^1 \leq c \frac{rk}{N} \right) \lesssim e^{-\Omega(rk/N)}$$

over all $\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}'_k \cup \mathbb{S}''_k$ for some $c > 0$. For the remainder of the lemma, we fix a generic $\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}'_k \cup \mathbb{S}''_k$ (the structural similarity of $\mathbb{S}'_k$ and $\mathbb{S}''_k$ allows us to avoid splitting cases).

We proceed very much as we did in §2.2, to which we refer the reader for greater detail (particularly Proposition 2.2.8). Following the computations and reasoning of 2.2.6 and 2.2.7, we see

$$\mathcal{H}L_{\mathbf{b}} = \begin{bmatrix} - \begin{pmatrix} \Lambda''(a_0 + 1) \\ +\Lambda''(b_0 - a_0 + 1) \end{pmatrix} & 0 & 0 \\ 0 & - \begin{pmatrix} \Lambda''(a_1 + 1) \\ \Lambda''(b_1 - a_1 + 1) \end{pmatrix} & 0 \\ 0 & 0 & - \begin{pmatrix} \Lambda''(a_2 + 1) \\ \Lambda''(b_2 - a_2 + 1) \end{pmatrix} \end{bmatrix} \quad (2.268)$$

and that $L_{\mathbf{b}}(\mathbf{a})$ is therefore strictly concave. Also, as $\mathbf{a}$ approaches the boundary of $\mathcal{D}_r^{\mathbf{b}}$, $L_{\mathbf{b}}(\mathbf{a})$ uniformly approaches $-\infty$ so $L_{\mathbf{b}}(\mathbf{a})$ must have a maximum (which

will be the only critical point by concavity) in the interior of $\mathcal{D}_r^{\mathbf{b}}$. Using the method of Lagrange multipliers and the concavity of $L_{\mathbf{b}}(\mathbf{a})$, this maximum occurs precisely at the (unique) solution in $\mathcal{D}_r^{\mathbf{b}}$ to the system

$$\begin{aligned}\Lambda'(b_0 - a_0 + 1) - \Lambda'(a_0 + 1) &= \Lambda'(b_1 - a_1 + 1) - \Lambda'(a_1 + 1) \\ &= \Lambda'(b_2 - a_2 + 1) - \Lambda'(a_2 + 1).\end{aligned}\tag{2.269}$$

Note that, because $\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}'_k$ or $\mathbb{S}''_k$, we have the bounds $|b_0 - b_0^k| = o(b_0^k)$ and $|b_1 - b_1^k| = o(b_1^k)$ so, by Proposition (2.2.8),

$$b_0 \sim b_0^k \sim N \tag{2.270}$$

$$b_1 \sim b_1^k \sim k. \tag{2.271}$$

At this point it will be expedient to handle the exceptional cases in which a_0 is close to $b_0 + 1$, a_1 is close to $b_1 + 1$, or a_0 is close to $b_2 + 1$. Note that by Proposition 2.2.8, either $b_2 = O(1)$ or $b_2 \sim \frac{k^2}{N} \lesssim k$ so $b_2 = O(b_1) = O(b_0)$. Therefore, by (2.269), if $b_1 - a_1 + 1$ or $b_0 - a_0 + 1$ is sufficiently small, we can ensure $a_2 > b_2$ so it will suffice to handle the case in which $a_2 > b_2$.

Although we cannot force a_0 to be arbitrarily close to $b_0 + 1$ or a_1 to be arbitrarily close to $b_1 + 1$ (in an additive sense) by insisting that a_2 be sufficiently close (relative to a constant independent of $\mathbf{b}$) to $b_2 + 1$, the condition $a_2 > b_2$ in conjunction with (2.269) does ensure that $a_0 \sim b_0 (\sim N)$ and $a_1 \sim b_1 (\sim k)$. Moreover, this implies $N > r = a_0 + a_1 + a_2 \gtrsim N$, i.e. $r \sim N$ in this regime.

Then we compute for any $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3$

$$\begin{aligned}\frac{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}-\mathbf{f}_1}^r)}{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)} &= \frac{\binom{r}{\mathbf{a}+\mathbf{e}_0-\mathbf{e}_1} \binom{N-r}{\mathbf{b}-\mathbf{a}-\mathbf{e}_0+\mathbf{e}_1}}{\binom{r}{\mathbf{a}} \binom{N-r}{\mathbf{b}-\mathbf{a}}} \\ &= \frac{a_1}{a_0 + 1} \frac{b_0 - a_0}{b_1 - a_1 + 1} \\ &\leq \frac{a_1}{b_1 - a_1 + 1} \frac{N}{a_0 + 1}.\end{aligned}\tag{2.272}$$

We observe that for $c > 0$ sufficiently small, $c \frac{rk}{N} \leq ck \leq \frac{a_1}{2}$. For such a c , let $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3$ have $a_1 \leq 2c \frac{rk}{N} \leq a_1$. Because $a_2 \leq b_2 \leq a_2$ and $\Sigma \mathbf{a} = r = \Sigma \mathbf{a}^{\mathbf{b}}$, we

now have

$$N \gtrsim a_0 = r - a_1 - a_2 > r - a_1 - a_2 = a_0 \sim N. \quad (2.273)$$

Plugging (2.273) into (2.272) provides the bound

$$\frac{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}-\mathbf{f}_1}^r)}{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)} \lesssim \frac{a_1}{b_1 - a_1}. \quad (2.274)$$

Reducing c if necessary, this allows us to enforce the bound

$$\frac{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}-\mathbf{f}_1}^r)}{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)} \leq \frac{1}{2}. \quad (2.275)$$

for all $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3$ with $a_1 \leq 2c \frac{rk}{N}$.

Applying (2.275) inductively for any $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3$ with $a_1 = 2 \lfloor c \frac{rk}{N} \rfloor$ shows that

$$\begin{aligned} & \mathbb{P}_{\mathbb{T}_{\mathbf{b}}} \left([D_{\leq r}^0, D_{\leq r}^1, D_{\leq r}^2] = \mathbf{a} - \left\lfloor c \frac{rk}{N} \right\rfloor \mathbf{f}_1 \right) \\ & \leq \frac{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)}{2^{\lfloor c \frac{rk}{N} \rfloor}} \\ & = e^{-\Omega(rk/N)} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r). \end{aligned} \quad (2.276)$$

We strengthen the bound (2.276) by again applying (2.275) inductively to find

$$\begin{aligned} & \sum_{m=\lfloor c \frac{rk}{N} \rfloor}^{2\lfloor c \frac{rk}{N} \rfloor} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}-m\mathbf{f}_1}^r) \\ & \lesssim \mathbb{P}_{\mathbb{T}_{\mathbf{b}}} \left([D_{\leq r}^0, D_{\leq r}^1, D_{\leq r}^2] = \mathbf{a} - \left\lfloor c \frac{rk}{N} \right\rfloor \mathbf{f}_1 \right) \\ & \leq e^{-\Omega(rk/N)} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r). \end{aligned} \quad (2.277)$$

We finish the $a_2 > b_2$ case by decomposing $\mathbb{T}_{\mathbf{b}}$:

$$\begin{aligned} & \mathbb{P}_{\mathbb{T}_{\mathbf{b}}} (D_{\leq r}^1 \leq c \frac{rk}{N}) \\ & = \sum_{m=0}^{\lfloor c \frac{rk}{N} \rfloor} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}} (D_{\leq r}^1 = m) \end{aligned}$$

$$\begin{aligned}
&= \sum_{m=\lfloor c \frac{rk}{N} \rfloor}^{2\lfloor c \frac{rk}{N} \rfloor} \mathbb{P}_{\mathbb{T}_b} \left(D_{\leq r}^1 = 2 \left\lfloor c \frac{rk}{N} \right\rfloor - m \right) \\
&= \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^b \cap \mathbb{N}^3 \\ a_1 = 2\lfloor c \frac{rk}{N} \rfloor}} \sum_{m=\lfloor c \frac{rk}{N} \rfloor}^{2\lfloor c \frac{rk}{N} \rfloor} \mathbb{P}_{\mathbb{T}_b} (\mathbb{T}_{\mathbf{a}-m\mathbf{f}_1}^r) \\
&\leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^b \cap \mathbb{N}^3 \\ a_1 = 2\lfloor c \frac{rk}{N} \rfloor}} e^{-\Omega(rk/N)} \mathbb{P}_{\mathbb{T}_b} (\mathbb{T}_{\mathbf{a}}^r) \\
&\leq e^{-\Omega(rk/N)}.
\end{aligned}$$

Thus, we are hereafter free to assume

$$(b_0 - a_0 + 1), (b_1 - a_1 + 1), (b_2 - a_2 + 1) = \Omega(1). \quad (2.278)$$

Our next goal is to prove that

$$\frac{r}{N} \sim \frac{a_0}{b_0} \sim \frac{a_1}{b_1}. \quad (2.279)$$

In the process, we will frequently use Claim 2.2.4 to approximate Λ' away from 0. We observe that if $a_0 = O(1)$, then

$$\Lambda'(b_0 - a_0 + 1) - \Lambda'(a_0 + 1) \geq \ln[b_0 + O(1)] + O(1) = \ln(N) + O(1) \quad (2.280)$$

by (2.270). Looking back to (2.269), this would force $a_1, a_2 = O(1)$ and therefore $r = \sum_{j=0}^2 a_j = O(1)$. However, in this case, the conclusion of the lemma is vacuously true. Thus, $a_0 = \omega(1)$ and, in particular, a_0 is bounded away from 0.

On the other hand, if $a_0 \sim b_0 (\sim N)$, then

$$\Lambda'(b_0 - a_0 + 1) - \Lambda'(a_0 + 1) \leq [\ln(N) + O(1)] - \ln(N) \leq O(1) \quad (2.281)$$

and (2.269) would force the condition

$$a_1 \sim b_1 (\sim k). \quad (2.282)$$

and therefore we would have

$$\frac{r}{N} \leq 1 \sim \frac{a_1}{b_1} \sim \frac{a_0}{b_0} \lesssim \frac{r}{N},$$

thus proving (2.279) in the $a_0 \sim b_0$ case.

Now we are left with the situation in which $a_0 = \omega(1), o(N)$. In this case we have

$$\begin{aligned} \Lambda'(b_0 - a_0 + 1) - \Lambda'(a_0 + 1) &= \ln(b_0 - a_0) - \ln(a_0) + O\left(\frac{1}{a_0}\right) \\ &= \ln\left(\frac{b_0}{a_0} + 1\right) + O\left(\frac{1}{a_0}\right) \\ &= \ln\left(\frac{b_0}{a_0}\right) + O\left(\frac{1}{a_0} + \frac{a_0}{b_0}\right) \\ &= \ln\left(\frac{b_0}{a_0}\right) + o(1) \end{aligned} \tag{2.283}$$

By (2.278) we may use the approximations

$$\Lambda'(b_1 - a_1 + 1) = \ln(b_1 - a_1 + 1) + O(1) \text{ and}$$

$$\Lambda'(b_2 - a_2 + 1) = \ln(b_2 - a_2 + 1) + O(1).$$

Moreover, if $a_1 = O(1)$ and $a_0 = o(r)$ then we have

$$r = a_0 + a_1 + a_2 = a_2 + o(r)$$

and therefore

$$\begin{aligned} \ln\left[\omega\left(\frac{N}{r}\right)\right] &= \ln\left(\frac{b_0}{a_0}\right) \\ &= \Lambda'(b_2 - a_2 + 1) - \Lambda'(a_2 + 1) + o(1) \\ &= \ln(b_2 - a_2 + 1) - \ln(r) + O(1) \\ &\leq \ln\left(\frac{b_2 + 1}{r}\right) + O(1) \\ \implies b_2 &= \omega(N) \end{aligned} \tag{2.284}$$

which contradicts the trivial bound $b_2 \leq N$. Thus we have that $a_1 = O(1) \implies a_0 = \Theta(r)$. In this case, we have

$$\begin{aligned}
& \Lambda'(b_0 - a_0 + 1) - \Lambda'(a_0 + 1) = \Lambda'(b_1 - a_1 + 1) - \Lambda'(a_1 + 1) \\
\implies & \ln\left(\frac{b_0}{a_0}\right) = \ln(k) - \Lambda'(a_1 + 1) + O(1) \\
\implies & \ln\left(\frac{N}{r}\right) = \ln(k) - \Lambda'(a_1 + 1) + O(1) \\
\implies & \ln\left(\frac{N}{rk}\right) = -\Lambda'(a_1 + 1) + O(1) \\
\implies & \ln\left(\frac{N}{rk}\right) \geq O(1) \\
\implies & \ln\left(\frac{rk}{N}\right) \leq O(1) \\
\implies & \frac{rk}{N} = O(1).
\end{aligned}$$

However, if $\frac{rk}{N} = O(1)$, the lemma holds vacuously. Therefore, we may assume $a_1 = \omega(1)$ and we use the approximation $\Lambda'(a_1 + 1) = \ln(a_1 + 1) + O(1)$ and (2.269) to observe

$$\Lambda'(b_0 - a_0 + 1) - \Lambda'(a_0 + 1) = \Lambda'(b_1 - a_1 + 1) - \Lambda'(a_1 + 1) \quad (2.285)$$

$$\implies \ln\left(\frac{b_0}{a_0}\right) = \ln\left(\frac{b_1 - a_1 + 1}{a_1}\right) + O(1) \quad (2.286)$$

$$\implies \ln\left(\frac{b_0}{a_0}\right) = \ln\left(\frac{b_1}{a_1}\right) + O(1). \quad (2.287)$$

so the ratios $\frac{a_0}{b_0}$ and $\frac{a_1}{b_1}$ are comparable. Note that, if $a_2 = O(1)$, then

$$\begin{aligned}
r &= a_0 + a_1 + a_2 = \frac{a_0}{b_0}b_0 + \frac{a_1}{b_1}b_1 + O(1) \\
\implies r + O(1) &\sim \frac{a_0}{b_0}(b_0 + b_1) \sim \frac{a_0}{b_0}N \\
\implies \frac{r}{N} &\sim \frac{a_0}{b_0} \left(\sim \frac{a_1}{b_1} \right).
\end{aligned}$$

Similarly if $a_2 = \omega(1)$ (in particular, a_2 is bounded away from $b_2 + 1$) then, in perfect analogy with (2.285), we have

$$\frac{a_2}{b_2} \sim \frac{a_0}{b_0} \left(\sim \frac{a_1}{b_1} \right)$$

and therefore

$$r = a_0 + a_1 + a_2 = \frac{a_0}{b_0}b_0 + \frac{a_1}{b_1}b_1 + \frac{a_2}{b_2}b_2 \sim \frac{a_0}{b_0}(b_0 + b_1 + b_2) = \frac{a_0}{b_0}N.$$

This proves that, regardless of a_2 , we have (2.279). In conjunction with (2.282) and our earlier observation that we may ignore the $a_0 = O(1)$ case, this proves (2.279) in general.

Define $\mathbf{a}^{\mathbf{b}}$ to be the element of $\mathcal{D}_r^{\mathbf{b}}$ such that $L_{\mathbf{b}}(\mathbf{a}^{\mathbf{b}})$ is maximal, whose properties we just established. We now focus on two cases, namely when $a_2^{\mathbf{b}}$ is bounded away from -1 and when $a_2^{\mathbf{b}}$ is “close” to -1 . First we handle the former. Fix a (small) $\delta > 0$ to be determined later and suppose $a_2^{\mathbf{b}} \geq \delta - 1$. For any $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}}$,

$$\begin{aligned} L_{\mathbf{b}}(\mathbf{a}) - L_{\mathbf{b}}(\mathbf{a}^{\mathbf{b}}) &= \int_0^1 \partial_t L_{\mathbf{b}}(\mathbf{a}^{\mathbf{b}} + t(\mathbf{a} - \mathbf{a}^{\mathbf{b}})) dt \\ &= \int_0^1 \int_0^t (\mathbf{a} - \mathbf{a}^{\mathbf{b}}) \cdot \mathcal{H}L_{\mathbf{b}}(\mathbf{a}^{\mathbf{b}} + s(\mathbf{a} - \mathbf{a}^{\mathbf{b}})) ds dt \\ &= - \int_0^1 \left[\sum_{j=0}^2 (a_j - a_j^{\mathbf{b}})^2 \int_0^t \left(\Lambda''(a_j^{\mathbf{b}} + s(a_j - a_j^{\mathbf{b}}) + 1) \right. \right. \\ &\quad \left. \left. + \Lambda''(b_j - a_j^{\mathbf{b}} - s(a_j - a_j^{\mathbf{b}}) + 1) \right) ds \right] dt. \end{aligned} \quad (2.288)$$

From (2.288), we draw two conclusions. The first is that, because we are in the regime $a_2^{\mathbf{b}} > \delta - 1$ and we already established that $a_j^{\mathbf{b}}$ is bounded away from $b_j + 1$ for $j \in \{0, 1, 2\}$ (the analogous conditions for a_j follow directly from the fact that $a_j \in \{0, \dots, b_j\}$), all arguments of Λ'' are bounded away from 0 so the integral (2.288) is $O_\delta(1)$ as long as $|\mathbf{a} - \mathbf{a}^{\mathbf{b}}| = O(1)$. Picking such an arbitrary $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3$, we can exponentiate (2.288) to obtain

$$\begin{aligned} L_{\mathbf{b}}(\mathbf{a}^{\mathbf{b}}) &= L_{\mathbf{b}}(\mathbf{a}) + O_\delta(1) \\ &= \ln \left[\mathbb{P}_{\mathbf{T}_{\mathbf{b}}}([D_{\leq r}^0, D_{\leq r}^1, D_{\leq r}^2] = \mathbf{a}) \right] + O_\delta(1) \\ &\leq \ln(1) + O_\delta(1) = O_\delta(1). \end{aligned} \quad (2.289)$$

Our second conclusion from (2.288) will conclude the $a_2^{\mathbf{b}} > \delta - 1$ regime. By combining (2.289) with (2.288) (and ignoring some non-negative terms in the inner integrand in the latter), we conclude that for any $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}}$,

$$L_{\mathbf{b}}(\mathbf{a}) \leq O_{\delta}(1) - \int_0^1 \left[\sum_{j=1}^2 (a_j - a_j^{\mathbf{b}})^2 \int_0^t \Lambda''(a_j^{\mathbf{b}} + s(a_j - a_j^{\mathbf{b}}) + 1) ds \right] dt. \quad (2.290)$$

Recall that $a_1^{\mathbf{b}} \sim \frac{rk}{N}$ and $a_1 \geq 0$ so the argument $a_1^{\mathbf{b}} + s(a_1 - a_1^{\mathbf{b}}) + 1$ is bounded away from 0 uniformly in $s \in [0, 1]$. In particular, as long as $a_1 \leq a_1^{\mathbf{b}}$ (as in our present $a_1 \leq c\frac{rk}{N}$ case),

$$\Lambda''(a_1^{\mathbf{b}} + s(a_1 - a_1^{\mathbf{b}}) + 1) \gtrsim \frac{1}{a_1^{\mathbf{b}}}$$

so (2.290) implies

$$\begin{aligned} |L_{\mathbf{b}}(\mathbf{a})| &\gtrsim (a_2 - a_2^{\mathbf{b}})^2 \int_0^1 \int_0^t \Lambda''(a_2^{\mathbf{b}} + s(a_2 - a_2^{\mathbf{b}}) + 1) ds dt \\ &\quad + \frac{(a_1 - a_1^{\mathbf{b}})^2}{a_1^{\mathbf{b}}} + O_{\delta}(1). \end{aligned} \quad (2.291)$$

Because $a_1^{\mathbf{b}} \sim \frac{rk}{N}$, if a_1 is bounded from above by a sufficiently small multiple of $\frac{rk}{N}$ (i.e. $a_1 < c\frac{rk}{N}$), $a_1^{\mathbf{b}} \sim (a_1^{\mathbf{b}} - a_1)$. Moreover, $a_2^{\mathbf{b}} > \delta - 1$ so, as long as $a_2 > \delta - 1$, (2.291) implies

$$\begin{aligned} L_{\mathbf{b}}(\mathbf{a}) &\leq -\Omega_{\delta} \left(\frac{(a_2 - a_2^{\mathbf{b}})^2}{\max(a_2^{\mathbf{b}}, a_2) + 1} \right) \\ &\quad - \Theta(a_1^{\mathbf{b}} - a_1) + O_{\delta}(1). \end{aligned} \quad (2.292)$$

Noting that each a_2 in (2.293) below is bounded below by 0 and therefore $1 - \delta$, we exponentiate (2.292) and compute

$$\mathbb{P}_{\mathbb{T}_{\mathbf{b}}} (D_{\leq r}^1 \leq c\frac{rk}{N}) = \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 \leq c\frac{rk}{N}}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}} (\mathbb{T}_{\mathbf{a}}^r) = \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 \leq c\frac{rk}{N}}} \exp[L_{\mathbf{b}}(\mathbf{a})]$$

$$\begin{aligned}
&\leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 \leq c \frac{rk}{N}}} \exp \left[-\Omega_\delta \left(\frac{(a_2 - a_2^{\mathbf{b}})^2}{\max(a_2^{\mathbf{b}}, a_2) + 1} \right) - \Theta(a_1^{\mathbf{b}} - a_1) + O_\delta(1) \right] \\
&\lesssim_\delta \sum_{\substack{0 \leq a_1 \leq c \frac{rk}{N} \\ 0 \leq a_2}} e^{-\Theta(a_1^{\mathbf{b}} - a_1)} \exp \left[-\Omega_\delta \left(\frac{(a_2 - a_2^{\mathbf{b}})^2}{\max(a_2^{\mathbf{b}}, a_2) + 1} \right) \right] \\
&\lesssim e^{-\Theta(a_1^{\mathbf{b}} - c \frac{rk}{N})} \sum_{0 \leq m \leq c \frac{rk}{N}} e^{-\Theta(m)} \sum_{a_2=0}^{\infty} \exp \left[-\Omega_\delta \left(\frac{(a_2 - a_2^{\mathbf{b}})^2}{\max(a_2^{\mathbf{b}}, a_2) + 1} \right) \right] \\
&\lesssim e^{-\Theta(\frac{rk}{N})} \sum_{a_2=0}^{\infty} \exp \left[-\Omega_\delta \left(\frac{(a_2 - a_2^{\mathbf{b}})^2}{\max(a_2^{\mathbf{b}}, a_2) + 1} \right) \right] \tag{2.293}
\end{aligned}$$

We split the remaining sum in the final expression in (2.293) into $a_2 \leq 2 \lceil a_2^{\mathbf{b}} \rceil$ and the remainder. First we use the blunt estimate

$$\sum_{a_2=0}^{2 \lceil a_2^{\mathbf{b}} \rceil} \exp \left[-\Omega_\delta \left(\frac{(a_2 - a_2^{\mathbf{b}})^2}{\max(a_2^{\mathbf{b}}, a_2) + 1} \right) \right] \leq \sum_{0 \leq a_2 \leq 2 \lceil a_2^{\mathbf{b}} \rceil} 1 \lesssim \max(1, a_2^{\mathbf{b}}) \lesssim \frac{rk}{N}. \tag{2.294}$$

The larger values of a_2 benefit from the exponential decay:

$$\begin{aligned}
\sum_{a_2 > 2 \lceil a_2^{\mathbf{b}} \rceil} \exp \left[-\Omega_\delta \left(\frac{(a_2 - a_2^{\mathbf{b}})^2}{\max(a_2^{\mathbf{b}}, a_2) + 1} \right) \right] &= \sum_{a_2 > 2 \lceil a_2^{\mathbf{b}} \rceil} e^{-\Omega_\delta(a_2)} \\
&\leq \sum_{a_2=0}^{\infty} e^{-\Omega_\delta(a_2)} \lesssim_\delta 1. \tag{2.295}
\end{aligned}$$

Plugging (2.294) and (2.295) into (2.293) yields

$$\begin{aligned}
\mathbb{P}_{\mathbb{T}_{\mathbf{b}}} (D_{\leq r}^1 \leq c \frac{rk}{N}) &\lesssim e^{-\Theta(\frac{rk}{N})} \sum_{a_2=0}^{\infty} \exp \left[-\Omega_\delta \left(\frac{(a_2 - a_2^{\mathbf{b}})^2}{\max(a_2^{\mathbf{b}}, a_2) + 1} \right) \right] \\
&\lesssim_\delta e^{-\Omega_\delta(\frac{rk}{N})} \left(\frac{rk}{N} \right) \\
&\lesssim_\delta e^{-\Omega_\delta(\frac{rk}{N})}, \tag{2.296}
\end{aligned}$$

which was our desired bound. Our goal for the remainder of the proof is to find a $\delta > 0$ such that if $a_2^{\mathbf{b}} \leq \delta - 1$,

$$\mathbb{P}_{\mathbb{T}_{\mathbf{b}}} (D_{\leq r}^1 \leq c \frac{rk}{N}) \lesssim_\delta e^{-\Omega_\delta(\frac{rk}{N})}$$

Heuristically, if δ is sufficiently small, the condition $a_2^{\mathbf{b}} \leq \delta - 1$ should imply that a uniformly random element of $\mathbb{T}_{\mathbf{b}}$ is unlikely to have $D_{\leq r}^2 > 0$. Looking back at (2.279) and (2.269), we have that, as long as $b_2 \geq 1$,

$$\begin{aligned} \Lambda'(b_2) - \Lambda'(a_2^{\mathbf{b}} + 1) &\leq \Lambda'(b_2 - a_2^{\mathbf{b}} + 1) - \Lambda'(a_2^{\mathbf{b}} + 1) \\ &= \Lambda'(b_0 - a_0^{\mathbf{b}} + 1) - \Lambda'(a_0^{\mathbf{b}} + 1) \\ &\leq \ln(N) - \ln(r) = \ln\left(\frac{N}{r}\right) \end{aligned} \quad (2.297)$$

Then, because $b_2 \geq 1$, (2.297) implies

$$\begin{aligned} \ln(b_2) &\leq \ln\left(\frac{N}{r}\right) + \Lambda'(a_2^{\mathbf{b}} + 1) + O(1) \\ \implies \frac{b_2 r}{N} &\lesssim e^{\Lambda'(a_2^{\mathbf{b}} + 1)}. \end{aligned} \quad (2.298)$$

Thus, if $r = \Omega(N)$, we can pick δ small enough that $b_2 = 0$ and therefore $\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(D_{\leq r}^2 > 0) = 0$. Now, as we bound $\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)$, we are free to assume that either

$$b_2 = a_2 = 0 \text{ or} \quad (2.299)$$

$$r \text{ is bounded above by an arbitrarily small multiple of } N. \quad (2.300)$$

Moreover, because $b_0 \sim N$ and $a_0 \leq r + 2$, (2.300) implies

$$N \gtrsim b_0 - a_0 \gtrsim N - r \gtrsim N. \quad (2.301)$$

Let $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3$ have $a_2 \geq 1$ so that, by (2.300) and (2.301) we have $b_0 - a_0 \sim N$. Then we compute

$$\begin{aligned} \frac{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}+\mathbf{e}_0-\mathbf{e}_2}^r)}{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)} &= \frac{\binom{r}{\mathbf{a}+\mathbf{e}_0-\mathbf{e}_2} \binom{N-r}{\mathbf{b}-\mathbf{a}-\mathbf{e}_0+\mathbf{e}_2}}{\binom{r}{\mathbf{a}} \binom{N-r}{\mathbf{b}-\mathbf{a}}} \\ &= \frac{a_2}{a_0 + 1} \frac{b_0 - a_0}{b_2 - a_2 + 1} \\ &\gtrsim \frac{1}{r} \frac{b_0 - a_0}{b_2} \end{aligned} \quad (2.302)$$

Because $b_0 - a_0 \gtrsim N$, (2.302) implies

$$\frac{\mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{\mathbf{a}+\mathbf{e}_0-\mathbf{e}_2}^r)}{\mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{\mathbf{a}}^r)} \gtrsim \frac{N}{rb_2} \quad (2.303)$$

Because $\lim_{x \rightarrow 0^+} \Lambda'(x) = -\infty$, (2.298) implies that, if δ is sufficiently small, $\frac{b_2 r}{N}$ can be forced below any positive number. In conjunction with (2.303), this implies that, as long as $\delta > 0$ is sufficiently small,

$$\frac{\mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{\mathbf{a}+\mathbf{e}_0-\mathbf{e}_2}^r)} \leq \frac{1}{2}. \quad (2.304)$$

We decompose

$$\begin{aligned} & \mathbb{P}_{\mathbb{T}_b}(D_{\leq r}^1 \leq c \frac{rk}{N}) \\ &= \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^b \cap \mathbb{N}^3 \\ a_1 \leq c \frac{rk}{N}}} \mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{\mathbf{a}}^r) \\ &\leq \sum_{\substack{a_1 \leq c \frac{rk}{N}}} \sum_{n=0}^{r-a_1} \mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{(r-a_1-n, a_1, n)}^r) \end{aligned} \quad (2.305)$$

and apply (2.304) inductively to achieve the bound

$$\mathbb{P}_{\mathbb{T}_b}(D_{\leq r}^1 \leq c \frac{rk}{N}) \lesssim \sum_{0 \leq m \leq c \frac{rk}{N}} \mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{(r-m, m, 0)}^r) \quad (2.306)$$

Note that we have now proven (2.306) in the case where $b_2 > 0$, however, if $b_2 = 0$ we have the same bound (with equality in fact) simply because of our earlier observation that, in this case, $\mathbb{P}_{\mathbb{T}_b}(D_{\leq r}^2 > 0) = 0$. Computing as in (2.302) shows

$$\frac{\mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{(r-m+1, m-1, 0)}^r)}{\mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{(r-m, m, 0)}^r)} = \frac{m}{r-m+1} \frac{b_0 - r + m}{b_1 - m + 1}. \quad (2.307)$$

Then, as long as c is sufficiently small, (2.307) implies

$$\frac{\mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{(r-m+1, m-1, 0)}^r)}{\mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{(r-m, m, 0)}^r)} \leq \frac{1}{2} \quad (2.308)$$

for all $m \leq 2c\frac{rk}{N}$. Applying (2.308) inductively as m decreases from $\lfloor 2c\frac{rk}{N} \rfloor$ to $\lfloor c\frac{rk}{N} \rfloor$ implies

$$\mathbb{P}_{\mathbb{T}_b} \left([D_{\leq r}^0, D_{\leq r}^1, D_{\leq r}^2] = \left[r - \left\lfloor c\frac{rk}{N} \right\rfloor, \left\lfloor c\frac{rk}{N} \right\rfloor, 0 \right] \right) \leq e^{-\Omega(rk/N)}. \quad (2.309)$$

Applying (2.308) inductively to (2.309) and plugging the bound into (2.306), we calculate the exponential sum and obtain

$$\mathbb{P}_{\mathbb{T}_b} (D_{\leq r}^1 \leq c\frac{rk}{N}) \lesssim e^{-\Omega(rk/N)}. \quad (2.310)$$

This handles the case in which $a_2^b \leq \delta - 1$ and therefore, in conjunction with (2.296) completes the proof. $\square$

Corollary 2.9.4. *The averaging operators S'_k and S''_k satisfy the spectral decay condition*

$$\left| \widehat{S'_k}(\xi) \right|, \left| \widehat{S''_k}(\xi) \right| \lesssim e^{-\Omega(rk/N)}$$

where $r = \|\xi\|_{freq}$.

Proof. By component-permutation symmetry we can assume without loss of generality that

$$\xi = \left(\bigotimes_{n=1}^r \xi_{x_n} \right) \otimes \left(\mathbf{1}^{\otimes (N-r)} \right)$$

where $\left| \widehat{P_1}(\xi_n) \right| \leq \lambda < 1$ for all $n \in \{1, \dots, r\}$. Here λ is the spectral bound in the definition of a radial space in §2.1.1 (Definition 2.1.1, axiom (12)). Using the constant c from Lemma 2.9.3, we can decompose and bound

$$\begin{aligned} \left| \widehat{S'_k}(\xi) \right| &= \left| \sum_{\mathbf{x} \in \{0,1,2\}^r} \left(\prod_{n=1}^r \widehat{P_{x_n}}(\xi_n) \right) \mathbb{P}_{S'_k}(\mathbf{D}_{\leq r} = \mathbf{x}) \right| \\ &\leq \sum_{\mathbf{x} \in \{0,1,2\}^r} \lambda^{|\{m: x_m=1\}|} \mathbb{P}_{S'_k}(\mathbf{D}_{\leq r} = \mathbf{x}) \\ &\leq \sum_{\substack{\mathbf{x} \in \{0,1,2\}^r \\ |\{m: x_m=1\}| \leq c\frac{rk}{N}}} \mathbb{P}_{S'_k}(\mathbf{D}_{\leq r} = \mathbf{x}) + \lambda^{c\frac{rk}{N}} \sum_{\substack{\mathbf{x} \in \{0,1,2\}^r \\ |\{m: x_m=1\}| > c\frac{rk}{N}}} \mathbb{P}_{S'_k}(\mathbf{D}_{\leq r} = \mathbf{x}) \end{aligned}$$

$$\begin{aligned}
&\leq \mathbb{P}_{\mathbb{S}'_k} \left(D_{\leq r}^1 \leq c \frac{rk}{N} \right) + e^{-\Omega\left(\frac{rk}{N}\right)} \\
&\lesssim e^{-\Omega\left(\frac{rk}{N}\right)}
\end{aligned}$$

by Lemma 2.9.3. The same argument proves

$$\left| \widehat{S''_k}(\xi) \right| \lesssim e^{-\Omega(rk/N)}$$

completing the corollary. $\square$

Notice that for high frequency eigenvectors, i.e. when $r \sim N$, the bound $\frac{r}{N}e^{-\Omega(rk/N)}$ is equivalent to $e^{-\Omega(rk/N)}$ so spectral smoothness with respect to k is a moot point. The next lemma and corollary allow us to take advantage of smoothness in the $r = o(N)$ case.

Lemma 2.9.5. *If $r < cN$ for a sufficiently small $c > 0$, then for any $\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3$ such that $a_1 + 2a_2 \leq \frac{k}{2}$,*

$$\frac{\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)} = 1 + O\left(\frac{a_1 + a_2}{k} + \frac{r}{N}\right).$$

Proof. Define $\mathbf{x} \in \{0, 1, 2\}^r$ by

$$x_n = \begin{cases} 2 & \text{for } 0 < n \leq a_2 \\ 1 & \text{for } a_2 < n \leq a_1 + a_2 \\ 0 & \text{for } a_1 + a_2 < n \leq r \end{cases}$$

By the construction of $\mathbb{S}'_k$ and $\mathbb{S}''_k$, for any event E , we have

$$\begin{aligned}
\mathbb{P}_{\mathbb{S}'_k}(E) &= \frac{\mathbb{P}_{\mathbb{S}_k}(E)}{\mathbb{P}_{\mathbb{S}_k}(\mathbb{S}'_k)} = \mathbb{P}_{\mathbb{S}_k}(E) \left[1 + O\left(\frac{1}{N}\right) \right] \quad \text{and} \\
\mathbb{P}_{\mathbb{S}''_k}(E) &= \frac{\mathbb{P}_{\mathbb{S}_{k+1}}(E)}{\mathbb{P}_{\mathbb{S}_{k+1}}(\mathbb{S}''_k)} = \mathbb{P}_{\mathbb{S}_{k+1}}(E) \left[1 + O\left(\frac{1}{N}\right) \right].
\end{aligned}$$

and therefore

$$\frac{\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)} = \frac{\mathbb{P}_{\mathbb{S}_{k+1}}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{\mathbb{S}_k}(\mathbb{T}_{\mathbf{a}}^r)} \left[1 + O\left(\frac{1}{N}\right) \right] \quad (2.311)$$

Because we are examining different spheres in spaces of changing dimension, for the remainder of the argument we will specify dimension of spheres. Recall that

$$\mathbb{S}_\kappa^\nu = \left\{ \sum_{n=1}^\nu D_n = \kappa \right\}.$$

By component-permutation symmetry, we have

$$\begin{aligned} & \frac{\mathbb{P}_{\mathbb{S}_k^N}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{\mathbb{S}_{k+1}^N}(\mathbb{T}_{\mathbf{a}}^r)} \\ &= \frac{\binom{r}{\mathbf{a}} \mathbb{P}_{\mathbb{S}_k^N}(\mathbf{D}_{\leq r} = \mathbf{x})}{\binom{r}{\mathbf{a}} \mathbb{P}_{\mathbb{S}_{k+1}^N}(\mathbf{D}_{\leq r} = \mathbf{x})} \\ &= \frac{\mathbb{P}_{\mathbb{S}_k^N}(\mathbf{D}_{\leq r} = \mathbf{x})}{\mathbb{P}_{\mathbb{S}_{k+1}^N}(\mathbf{D}_{\leq r} = \mathbf{x})} \\ &= \prod_{m=1}^r \frac{\mathbb{P}_{\mathbb{S}_k^N}(D_m = x_m | \mathbf{D}_{< m} = \mathbf{x}_{< m})}{\mathbb{P}_{\mathbb{S}_{k+1}^N}(D_m = x_m | \mathbf{D}_{< m} = \mathbf{x}_{< m})} \\ &= \left(\prod_{m=0}^{a_2-1} \frac{\mathbb{P}_{\mathbb{S}_{k-2m}^N}(D_1 = 2)}{\mathbb{P}_{\mathbb{S}_{k-2m+1}^N}(D_1 = 2)} \right) \left(\prod_{m=a_2+1}^r \frac{\mathbb{P}_{\mathbb{S}_k^N}(D_m = x_m | \mathbf{D}_{< m} = \mathbf{x}_{< m})}{\mathbb{P}_{\mathbb{S}_{k+1}^N}(D_m = x_m | \mathbf{D}_{< m} = \mathbf{x}_{< m})} \right) \\ &= \left(\prod_{m=0}^{a_2-1} \frac{\mathbb{P}_{\mathbb{S}_{k-2m}^N}(D_1 = 2)}{\mathbb{P}_{\mathbb{S}_{k-2m+1}^N}(D_1 = 2)} \right) \left(\prod_{m=0}^{a_1-1} \frac{\mathbb{P}_{\mathbb{S}_{(k-2a_2)-m}^N}(D_1 = 1)}{\mathbb{P}_{\mathbb{S}_{(k-2a_2)-m+1}^N}(D_1 = 1)} \right) \\ &\quad \left(\prod_{m=0}^{a_0-1} \frac{\mathbb{P}_{\mathbb{S}_{(k-2a_2-a_1)-m}^N}(D_1 = 0)}{\mathbb{P}_{\mathbb{S}_{(k-2a_2-a_1)-m+1}^N}(D_1 = 0)} \right) \end{aligned} \tag{2.312}$$

In conjunction with Lemma 2.8.4 and Proposition 2.8.6, (2.312) implies

$$\begin{aligned} & \frac{\mathbb{P}_{\mathbb{S}_k^N}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{\mathbb{S}_{k+1}^N}(\mathbb{T}_{\mathbf{a}}^r)} \\ &= \left(\prod_{m=0}^{a_2-1} \left[1 + O\left(\frac{1}{k-2m}\right) \right] \right) \left(\prod_{m=0}^{a_1-1} \left[1 + O\left(\frac{1}{k-2a_2-m}\right) \right] \right) \\ &\quad \left(\prod_{m=0}^{a_0-1} \left[1 + O\left(\frac{1}{N}\right) \right] \right) \\ &= \left[1 + O\left(\frac{a_2}{k}\right) \right] \left[1 + O\left(\frac{a_1}{k}\right) \right] \left[1 + O\left(\frac{r}{N}\right) \right] \\ &= 1 + O\left(\frac{a_1 + a_2}{k} + \frac{r}{N}\right) \end{aligned} \tag{2.313}$$

Finally, we combine (2.311) and (2.313) to find

$$\begin{aligned} & \frac{\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)} \\ &= \left[1 + O\left(\frac{a_1 + a_2}{k} + \frac{r}{N}\right) \right] \left[1 + O\left(\frac{1}{N}\right) \right] \\ &= 1 + O\left(\frac{a_1 + a_2}{k} + \frac{r}{N}\right), \end{aligned}$$

which is our desired bound. $\square$

Corollary 2.9.6. *If $r < cN$ for a sufficiently small $c > 0$, then for any $\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^d$ such that $a_1, a_2 = O\left(\frac{rk}{N}\right)$,*

$$\frac{\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)} = 1 + O\left(\frac{r}{N}\right).$$

Proof. This is an immediate consequence of Lemma 2.9.5. $\square$

Although Corollary 2.9.6 on its own is very useful for proving Proposition 2.9.1, the case in which $D_{\leq r} = \omega\left(\frac{rk}{N}\right)$ must be handled more subtly. In this regime, we must balance the more general (but relatively weak) smoothness bound from Lemma 2.9.5 with the rapid decay of $\mathbb{P}_{\mathbb{S}'_k}(D_{\leq r} > m)$ and $\mathbb{P}_{\mathbb{S}''_k}(D_{\leq r} > m)$ with respect to m .

Lemma 2.9.7. *If $r < cN$ for $c > 0$ sufficiently small and $\frac{rk}{N} = \Omega(1)$, there exists a (large) constant $C > 1$ such that*

$$\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > C \frac{rk}{N}}} \left| \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r) - \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) \right| \lesssim \frac{r}{N} e^{-\Omega(rk/N)}$$

Proof. Throughout the proof, keep in mind that many probabilistic bounds will be vacuously true simply because the event in question is empty. When in doubt, we advise first checking if this fact clears up the confusion.

Invoking Lemma 2.9.5, we observe that the quantity we wish to bound can be decomposed as

$$\begin{aligned}
& \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > C \frac{rk}{N}}} |\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r) - \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)| \\
& \leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > C \frac{rk}{N} \\ a_1, a_2 \leq \frac{k}{6}}} |\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r) - \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)| + \sum_{j=1}^2 \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > C \frac{rk}{N} \\ a_j > \frac{k}{6}}} |\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r) - \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)| \\
& \leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > C \frac{rk}{N} \\ a_1, a_2 \leq \frac{k}{6}}} \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) \left| \frac{\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| + \sum_{j=1}^2 \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > C \frac{rk}{N} \\ a_j > \frac{k}{6}}} [\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r) + \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)] \\
& \lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > C \frac{rk}{N} \\ a_1, a_2 \leq \frac{k}{6}}} \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) \max\left(\frac{r}{N}, \frac{a_1 + a_2}{k}\right) + \\
& \quad \sum_{j=1}^2 \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_j > \frac{k}{6}}} \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r) + \sum_{j=1}^2 \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_j > \frac{k}{6}}} \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) \\
& \lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > C \frac{rk}{N} \\ a_1, a_2 \leq \frac{k}{6}}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) + \\
& \quad \sum_{j=1}^2 \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_j > \frac{k}{6}}} \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r) + \sum_{j=1}^2 \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_j > \frac{k}{6}}} \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)
\end{aligned} \tag{2.314}$$

In order to prove the lemma, we will bound (2.314) up to a constant by $\frac{r}{N}e^{-\Omega(rk/N)}$.

We first handle the term

$$\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > C \frac{rk}{N} \\ a_1, a_2 \leq \frac{k}{6}}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) \leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > C \frac{rk}{N}}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)$$

$$\begin{aligned}
&\leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N}}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{S}_k''}(\mathbb{T}_{\mathbf{a}}^r) + \\
&\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_2 > \frac{C}{2} \frac{rk}{N}}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{S}_k''}(\mathbb{T}_{\mathbf{a}}^r)
\end{aligned} \tag{2.315}$$

Our immediate goal at this stage is to bound both terms in (2.315) up to constants by $\frac{r}{N}e^{-\Omega(rk/N)}$. The two terms are treated identically, so it will suffice to bound the term

$$\begin{aligned}
\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N}}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{S}_k''}(\mathbb{T}_{\mathbf{a}}^r) &= \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N}}} \frac{a_1 + a_2}{k} \left(\sum_{\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}_k''} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \mathbb{P}_{\mathbb{S}_k''}(\mathbb{T}_{\mathbf{b}}) \right) \\
&= \sum_{\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}_k''} \left(\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N}}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \right) \mathbb{P}_{\mathbb{S}_k''}(\mathbb{T}_{\mathbf{b}}). \tag{2.316}
\end{aligned}$$

Thus, it will further suffice to bound

$$\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N}}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \tag{2.317}$$

uniformly in $\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}_k''$. Fixing such a $\mathbb{T}_{\mathbf{b}}$, we compute $\partial_{\mathbf{e}_j - \mathbf{e}_0}^{\mathbf{a}} L_{\mathbf{b}}(\mathbf{a})$ directly for $j \in \{1, 2\}$:

$$\begin{aligned}
\partial_{\mathbf{e}_j - \mathbf{e}_0}^{\mathbf{a}} L_{\mathbf{b}}(\mathbf{a}) &= -\Lambda'(a_j + 1) + \Lambda'(b_j - a_j + 1) \\
&\quad + \Lambda'(a_0 + 1) - \Lambda'(b_0 - a_0 + 1).
\end{aligned} \tag{2.318}$$

If $a_j \in [0, b_j]$, we are free to use the approximation $\Lambda(x) = \ln(x) + O(1)$ for all terms in (2.318). In this case we have

$$\begin{aligned}
\partial_{\mathbf{e}_j - \mathbf{e}_0}^{\mathbf{a}} L_{\mathbf{b}}(\mathbf{a}) &= \ln\left(\frac{a_0 + 1}{b_0 - a_0 + 1}\right) - \ln\left(\frac{a_j + 1}{b_j - a_j + 1}\right) + O(1) \\
&\leq \ln\left(\frac{r}{N}\right) - \ln\left(\frac{a_j + 1}{k}\right) + O(1).
\end{aligned} \tag{2.319}$$

We note that, if $a_j \geq C \left(\frac{rk}{N} \right)$ for a sufficiently large $C > 1$, $\partial_{\mathbf{e}_j - \mathbf{e}_0}^{\mathbf{a}} L_{\mathbf{b}}(\mathbf{a}) \leq -\ln(2)$. In particular, if $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3$ with $\lceil C \left(\frac{rk}{N} \right) \rceil < a_j < b_j$ and $a_0 > 0$ for a given $\mathbb{T}_{\mathbf{b}}$, we have

$$\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a} + \mathbf{e}_j - \mathbf{e}_0}^r) \leq \frac{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)}{2}. \quad (2.320)$$

Also, because of our hypothesis that $\frac{rk}{N} = \Omega(1)$, we can increase C based on the lower bound for $\frac{rk}{N}$ in order to ensure that whenever $a_j \geq \frac{C}{4} \left(\frac{rk}{N} \right)$, (2.320) holds.

Our next step is dispense with those summands $\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)$ in (2.317) corresponding to $\mathbf{a}$ with $a_2 > \frac{C}{2} \frac{rk}{N}$. To this end, we apply (2.320) inductively (for $j = 2$) to show

$$\begin{aligned} \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1, a_2 > \frac{C}{2} \frac{rk}{N}}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) &= \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N}}} \sum_{a_2 > \frac{C}{2} \frac{rk}{N}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \\ &\leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N} \\ a_2 = \lceil \frac{C}{2} \frac{rk}{N} \rceil}} \sum_{m \geq 0} \frac{a_1 + a_2 + m}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a} + m(\mathbf{e}_2 - \mathbf{e}_0)}^r) \\ &\lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N} \\ a_2 = \lceil \frac{C}{2} \frac{rk}{N} \rceil}} \sum_{m \geq 0} \frac{a_1 + m}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a} + m(\mathbf{e}_2 - \mathbf{e}_0)}^r) \\ &\lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N} \\ a_2 = \lfloor \frac{C}{2} \frac{rk}{N} \rfloor}} \sum_{m > 0} \frac{a_1 + m}{k} \frac{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)}{2^m} \\ &= \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N} \\ a_2 = \lfloor \frac{C}{2} \frac{rk}{N} \rfloor}} \frac{a_1}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \sum_{m > 0} \left(1 + \frac{m}{a_1} \right) 2^{-m} \\ &\lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N} \\ a_2 = \lfloor \frac{C}{2} \frac{rk}{N} \rfloor}} \frac{a_1}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \left(\sum_{m > 0} m 2^{-m} \right) \end{aligned}$$

$$\lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N} \\ a_2 = \lfloor \frac{C}{2} \frac{rk}{N} \rfloor}} \frac{a_1}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \quad (2.321)$$

We use (2.321) to simplify (2.317).

$$\begin{aligned} & \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N}}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \\ &= \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_2 \leq \frac{C}{2} \frac{rk}{N} < a_1}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) + \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1, a_2 > \frac{C}{2} \frac{rk}{N}}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \\ &\lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_2 \leq \frac{C}{2} \frac{rk}{N} < a_1}} \frac{a_1 + a_2}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) + \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N} \\ a_2 = \lfloor \frac{C}{2} \frac{rk}{N} \rfloor}} \frac{a_1}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \\ &\lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_2 \leq \frac{C}{2} \frac{rk}{N} < a_1}} \frac{a_1}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) + \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 > \frac{C}{2} \frac{rk}{N} \\ a_2 = \lfloor \frac{C}{2} \frac{rk}{N} \rfloor}} \frac{a_1}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \\ &\lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_2 \leq \frac{C}{2} \frac{rk}{N} < a_1}} \frac{a_1}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \quad (2.322) \end{aligned}$$

Applying (2.320) inductively implies that for any $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3$ with $a_1 = \lceil \frac{C}{2} \frac{rk}{N} \rceil$, we have

$$\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \leq e^{-\Omega(rk/N)} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}} \left([D_{\leq r}^0, D_{\leq r}^1, D_{\leq r}^2] = \mathbf{a} - \left\lfloor \frac{C}{4} \frac{rk}{N} \right\rfloor (\mathbf{e}_j - \mathbf{e}_0) \right). \quad (2.323)$$

Summing (2.323) over all $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3$ such that $a_1 = \lceil \frac{C}{2} \frac{rk}{N} \rceil$ provides the bound

$$\begin{aligned} & \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 = \lceil \frac{C}{2} \frac{rk}{N} \rceil}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \leq e^{-\Omega(rk/N)} \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 = \lceil \frac{C}{2} \frac{rk}{N} \rceil - \lfloor \frac{C}{4} \frac{rk}{N} \rfloor}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \\ &= e^{-\Omega(rk/N)} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}} \left(\bigcup_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 = \lceil \frac{C}{2} \frac{rk}{N} \rceil - \lfloor \frac{C}{4} \frac{rk}{N} \rfloor}} \mathbb{T}_{\mathbf{a}}^r \right) \\ &\leq e^{-\Omega(rk/N)} \quad (2.324) \end{aligned}$$

Moreover, another inductive application of (2.320) proves that

$$\begin{aligned}
\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_2 \leq \frac{C}{2} \frac{rk}{N} < a_1}} \frac{a_1}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) &= \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_2 \leq \frac{C}{2} \frac{rk}{N}}} \sum_{a_1 > \frac{C}{2} \frac{rk}{N}} \frac{a_1}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \\
&\lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_2 \leq \frac{C}{2} \frac{rk}{N} \\ a_1 = \lceil \frac{C}{2} \frac{rk}{N} \rceil}} \sum_{m \geq 0} \frac{a_1 + m}{k} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a} + m \mathbf{f}_1}^r) \\
&\lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_2 \leq \frac{C}{2} \frac{rk}{N} \\ a_1 = \lceil \frac{C}{2} \frac{rk}{N} \rceil}} \sum_{m \geq 0} \frac{\frac{rk}{N} + m}{k} \frac{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)}{2^m} \\
&\lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_2 \leq \frac{C}{2} \frac{rk}{N} \\ a_1 = \lceil \frac{C}{2} \frac{rk}{N} \rceil}} \frac{r}{N} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \sum_{m \geq 0} \left(1 + \frac{m}{rk/N}\right) 2^{-m} \\
&\lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_2 \leq \frac{C}{2} \frac{rk}{N} \\ a_1 = \lceil \frac{C}{2} \frac{rk}{N} \rceil}} \frac{r}{N} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \sum_{m \geq 0} (m 2^{-m}) \\
&\lesssim \frac{r}{N} \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_2 \leq \frac{C}{2} \frac{rk}{N} \\ a_1 = \lceil \frac{C}{2} \frac{rk}{N} \rceil}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \\
&\leq \frac{r}{N} \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 = \lceil \frac{C}{2} \frac{rk}{N} \rceil}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \\
&\leq \frac{r}{N} e^{-\Omega(rk/N)} \tag{2.325}
\end{aligned}$$

by (2.324).

Returning to (2.314), the lemma will be complete once we bound $\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_j > \frac{k}{6}}} \mathbb{P}_E(\mathbb{T}_{\mathbf{a}}^r)$ by $\frac{r}{N} e^{-\Omega(rk/N)}$ for $j \in \{1, 2\}$ and $E \in \{\mathbb{S}'_k, \mathbb{S}''_k\}$. As in (2.316) the $j = 1$ and $j = 2$ terms are handled identically so it will suffice to bound the $j = 1$ term. Moreover, as in (2.317), it will suffice to achieve a bound on $\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_j > \frac{k}{6}}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)$ uniform in

$\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}'_k \cup \mathbb{S}''_k$, leaving us the task of bounding

$$\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > \frac{k}{6}}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r)$$

Choosing $c > 0$ sufficiently small, we can ensure that $\lceil \frac{k}{6} \rceil - 1 \geq \frac{C}{2} \frac{rk}{N}$ so that, as in (2.323), for all $\mathbf{a}$ with $a_1 = \lceil \frac{k}{6} \rceil - 1$, we have

$$\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \leq e^{-\Omega(rk/N)} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}} \left([D_{\leq r}^0, D_{\leq r}^1, D_{\leq r}^2] = \mathbf{a} - \left\lfloor \frac{C}{4} \frac{rk}{N} \right\rfloor (\mathbf{e}_1 - \mathbf{e}_0) \right). \quad (2.326)$$

Moreover, as before, we sum (2.326) over all $\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3$ such that $a_1 = \lceil \frac{k}{6} \rceil - 1$ to achieve

$$\begin{aligned} \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 = \lceil \frac{k}{6} \rceil - 1}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) &\leq e^{-\Omega(rk/N)} \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 = (\lceil \frac{k}{6} \rceil - 1) - \lfloor \frac{C}{4} \frac{rk}{N} \rfloor}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \\ &= e^{-\Omega(rk/N)} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}} \left(\bigcup_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 = (\lceil \frac{k}{6} \rceil - 1) - \lfloor \frac{C}{4} \frac{rk}{N} \rfloor}} \mathbb{T}_{\mathbf{a}}^r \right) \\ &\leq e^{-\Omega(rk/N)} \end{aligned} \quad (2.327)$$

Because $a_1 \gtrsim k$ for such an $\mathbf{a}$, (2.319) implies that $\partial_{\mathbf{e}_j - \mathbf{e}_0}^{\mathbf{a}} L_{\mathbf{b}}(\mathbf{a}) \leq -\ln\left(\frac{r}{N}\right) + O(1)$. Thus, if $a_0 > 0$, then

$$\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}^r) \lesssim \frac{r}{N} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r). \quad (2.328)$$

In particular, we have

$$\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 = \lceil \frac{k}{6} \rceil}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}+\mathbf{f}_1}^r) \lesssim \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1 = \lceil \frac{k}{6} \rceil - 1}} \frac{r}{N} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r). \quad (2.329)$$

Then (2.327), (2.329), and the exponential decay with respect to a_1 coming from inductive applications of (2.320) yield

$$\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > \frac{k}{6}}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) = \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 = \lceil \frac{k}{6} \rceil}} \sum_{m \geq 0} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}+m\mathbf{f}_1}^r)$$

$$\begin{aligned}
&\leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 = \lceil \frac{k}{6} \rceil}} \sum_{m \geq 0} \frac{\mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{\mathbf{a}}^r)}{2^m} \\
&\leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 = \lceil \frac{k}{6} \rceil}} \mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{\mathbf{a}}^r) \\
&\leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 = \lceil \frac{k}{6} \rceil - 1}} \frac{r}{N} \mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_{\mathbf{a}}^r) \\
&\leq \frac{r}{N} e^{-\Omega(rk/N)}, \tag{2.330}
\end{aligned}$$

finishing off the proof. $\square$

Remark 2.9.8. *The condition $D_{\leq r}^j > C \frac{rk}{N}$ is essentially sharp for $j = 1$ but not necessarily if $j > 1$. If $b_j^k \geq 1$, the analogous condition should look more like $D_{\leq r}^j > C \frac{rb_j^k}{N} \sim C \frac{rk^j}{N^j}$. In our present context in which $d = 2$, this changes nothing for $j = 1$ but for $j = 2$ it is a weaker requirement and thus a more useful condition. However, we do not need the weakened requirement for the present case so we do not prove it. In general, if $d > 2$, it very well might be necessary to prove and use the weakened requirement to get an analogous bound when $D_{\leq r}^j$ is substantially larger than $\frac{rb_j^k}{N}$ for some index $j \in \{1, \dots, d\}$.*

At this stage, most of the pieces of the proof of Proposition 2.9.1 have fallen into place.

Proof of Proposition 2.9.1. Let $c > 0$ be so small and $C > 1$ be so large that for any eigenvector $\xi \in \widehat{G}^N$ with $\|\xi\|_{freq}$, we have

$$\left| \widehat{S'_k}(\xi) \right|, \left| \widehat{S''_k}(\xi) \right| \leq C e^{-c(rk/N)}, \tag{2.331}$$

which is justified by Corollary 2.9.4. Then modify the constants as needed to ensure that, whenever $r < cN$, $\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3$,

$$\mathbb{P}_{\mathbb{S}'_k} \left(D_{\leq r}^1 \leq c \frac{rk}{N} \right), \mathbb{P}_{\mathbb{S}''_k} \left(D_{\leq r}^1 \leq c \frac{rk}{N} \right) \leq C e^{-c(rk/N)}, \tag{2.332}$$

$$\left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \leq C \left(\frac{a_1 + a_2}{k} + \frac{r}{N} \right) \text{ whenever } a_1 + 2a_2 \leq \frac{k}{2}, \quad (2.333)$$

$$\left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \leq C \frac{r}{N} \text{ whenever } a_1, a_2 \leq \frac{1}{c} \frac{rk}{N}, \text{ and} \quad (2.334)$$

$$\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > \frac{1}{c} \frac{rk}{N}}} |\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r) - \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r)| \leq C \frac{r}{N} e^{-c(rk/N)} \text{ whenever } \frac{rk}{N} \geq \delta, \quad (2.335)$$

where $\delta > 0$ is a constant to be determined later. Note that (2.332), (2.333), (2.334), and (2.335) are justified by Lemma 2.9.3, Lemma 2.9.5, Corollary 2.9.6, and Lemma 2.9.7 respectively.

If $r \geq cN$, then the bound $\frac{r}{N} e^{-\Omega(rk/N)}$ is equivalent up to a factor of c to $e^{-\Omega(rk/N)}$. Therefore, if $r \geq cN$, (2.331) is sufficient to prove the proposition. Hereafter we are free to assume that $r < cN$. We have already done the bulk of hard work in the case $\frac{rk}{N} = \Omega(1)$. More concretely we will assume $\frac{rk}{N} \geq \delta$. As in the proof of Corollary 2.9.4, we may assume without loss of generality that

$$\xi = \left(\bigotimes_{n=1}^r \xi_{x_n} \right) \otimes \left(\mathbf{1}^{\otimes(N-r)} \right).$$

where $|\widehat{P}_1(\xi_n)| \leq \lambda < 1$ for all $n \in \{1, \dots, r\}$. Here λ is the spectral bound in the definition of a radial space in §2.1.1 (Definition 2.1.1, axiom (12)). We then decompose

$$\begin{aligned} & \left| \widehat{S'_k}(\xi) - \widehat{S''_k}(\xi) \right| \\ &= \left| \sum_{\mathbf{x} \in \{0,1,2\}^r} \left(\prod_{n=1}^r \widehat{P}_{x_n}(\xi_n) \right) \left(\mathbb{P}_{S'_k}(\mathbf{D}_{\leq r} = \mathbf{x}) - \mathbb{P}_{S''_k}(\mathbf{D}_{\leq r} = \mathbf{x}) \right) \right| \\ &\leq \sum_{\mathbf{x} \in \{0,1,2\}^r} \left(\prod_{n=1}^r |\widehat{P}_{x_n}(\xi_n)| \right) \left| \frac{\mathbb{P}_{S'_k}(\mathbf{D}_{\leq r} = \mathbf{x})}{\mathbb{P}_{S''_k}(\mathbf{D}_{\leq r} = \mathbf{x})} - 1 \right| \left(\mathbb{P}_{S''_k}(\mathbf{D}_{\leq r} = \mathbf{x}) \right) \\ &\leq \sum_{\mathbf{x} \in \{0,1,2\}^r} \lambda^{|\{m: x_m=1\}|} \left| \frac{\mathbb{P}_{S'_k}(\mathbf{D}_{\leq r} = \mathbf{x})}{\mathbb{P}_{S''_k}(\mathbf{D}_{\leq r} = \mathbf{x})} - 1 \right| \left(\mathbb{P}_{S''_k}(\mathbf{D}_{\leq r} = \mathbf{x}) \right) \\ &= \sum_{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3} \sum_{\substack{\mathbf{x} \in \{0,1,2\}^r \\ |\{m: x_m=j\}|=a_j \\ j \in \{0,1,2\}}} \lambda^{a_1} \left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \left(\mathbb{P}_{S''_k}(\mathbf{D}_{\leq r} = \mathbf{x}) \right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3} \lambda^{a_1} \left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \\
&\leq \left(\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 \leq c \frac{rk}{N}}} + \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > c \frac{rk}{N} \\ a_1 + a_2 \leq \frac{1}{c} \frac{rk}{N}}} + \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > \frac{1}{c} \frac{rk}{N}}} \right) \lambda^{a_1} \left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \quad (2.336)
\end{aligned}$$

We bound the three sums in (2.336) individually. We apply (2.332) and (2.334) to bound the first sum.

$$\begin{aligned}
\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 \leq c \frac{rk}{N}}} \lambda^{a_1} \left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) &\leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 \leq c \frac{rk}{N}}} \left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \\
&\leq C \frac{r}{N} \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 \leq c \frac{rk}{N}}} \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \\
&\leq C \frac{r}{N} C e^{-c(rk/N)} \\
&\lesssim \frac{r}{N} e^{-\Omega(rk/N)}. \quad (2.337)
\end{aligned}$$

We apply (2.334) to bound the second sum.

$$\begin{aligned}
\sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > c \frac{rk}{N} \\ a_1 + a_2 \leq \frac{1}{c} \frac{rk}{N}}} \lambda^{a_1} \left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) &\leq \lambda^{c(rk/N)} \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > c \frac{rk}{N} \\ a_1 + a_2 \leq \frac{1}{c} \frac{rk}{N}}} \left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \\
&\leq C \frac{r}{N} \lambda^{c(rk/N)} \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > c \frac{rk}{N} \\ a_1 + a_2 \leq \frac{1}{c} \frac{rk}{N}}} \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \\
&= C \frac{r}{N} \lambda^{c(rk/N)} \mathbb{P}_{S''_k} \left(\bigcup_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 > c \frac{rk}{N} \\ a_1 + a_2 \leq \frac{1}{c} \frac{rk}{N}}} \mathbb{T}_{\mathbf{a}}^r \right) \\
&\leq C \frac{r}{N} \lambda^{c(rk/N)}
\end{aligned}$$

$$\lesssim \frac{r}{N} e^{-\Omega(rk/N)}. \quad (2.338)$$

We apply (2.335) to bound the third (and final) sum.

$$\begin{aligned} \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > \frac{1}{c} \frac{rk}{N}}} \lambda^{a_1} \left| \frac{\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) &\leq \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > \frac{1}{c} \frac{rk}{N}}} \left| \frac{\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) \\ &= \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 > \frac{1}{c} \frac{rk}{N}}} \left| \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r) - \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) \right| \\ &\leq C \frac{r}{N} e^{-c(rk/N)} \\ &\lesssim \frac{r}{N} e^{-\Omega(rk/N)}. \end{aligned} \quad (2.339)$$

We simply apply (2.337), (2.338), and (2.339) to (2.336) to prove the proposition for the case of $\frac{rk}{N} \geq \delta$.

To handle the $\frac{rk}{N} < \delta$ case we observe that the exponential portion of the bound is trivial so proposition devolves into bounding $\left| \widehat{S}'_k(\xi) - \widehat{S}''_k(\xi) \right|$ by $\frac{r}{N}$. Looking back at (2.336), we compute

$$\begin{aligned} &\left| \widehat{S}'_k(\xi) - \widehat{S}''_k(\xi) \right| \\ &= \left| \sum_{\mathbf{x} \in \{0,1,2\}^r} \left(\prod_{n=1}^r \widehat{P_{x_n}}(\xi_n) \right) \left(\mathbb{P}_{\mathbb{S}'_k}(\mathbf{D}_{\leq r} = \mathbf{x}) - \mathbb{P}_{\mathbb{S}''_k}(\mathbf{D}_{\leq r} = \mathbf{x}) \right) \right| \\ &\leq \sum_{\mathbf{x} \in \{0,1,2\}^r} \left| \mathbb{P}_{\mathbb{S}'_k}(\mathbf{D}_{\leq r} = \mathbf{x}) - \mathbb{P}_{\mathbb{S}''_k}(\mathbf{D}_{\leq r} = \mathbf{x}) \right| \\ &= \sum_{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3} \sum_{\substack{\mathbf{x} \in \{0,1,2\}^r \\ |\{m: x_m=j\}|=a_j \\ j \in \{0,1,2\}}} \left| \mathbb{P}_{\mathbb{S}'_k}(\mathbf{D}_{\leq r} = \mathbf{x}) - \mathbb{P}_{\mathbb{S}''_k}(\mathbf{D}_{\leq r} = \mathbf{x}) \right| \\ &= \sum_{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3} \left| \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r) - \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) \right| \\ &= \sum_{m=0}^r \sum_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 = m}} \left| \mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r) - \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) \right| \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{m=0}^r (m+1) \left(\max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1+a_2=m}} \left| \mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r) - \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \right| \right) \\
&= \left(\sum_{m=0} + \sum_{1 \leq m \leq \frac{k}{4}} + \sum_{\frac{k}{4} < m \leq \min(r,k)} \right) \left[(m+1) \left(\max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1+a_2=m}} \left| \mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r) - \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \right| \right) \right] \\
&\lesssim \left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{r\mathbf{e}_0}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{r\mathbf{e}_0}^r)} - 1 \right| \mathbb{P}_{S''_k}(\mathbb{T}_{r\mathbf{e}_0}^r) + \\
&\quad \sum_{1 \leq m \leq \frac{k}{4}} m \left(\max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1+a_2=m}} \left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r)} - 1 \right| \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \right) + \\
&\quad \sum_{\frac{k}{4} < m \leq \min(r,k)} m \left(\max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1+a_2=m}} \left| \mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r) - \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \right| \right) \\
&\lesssim \left| \frac{\mathbb{P}_{S'_k}(\mathbb{T}_{r\mathbf{e}_0}^r)}{\mathbb{P}_{S''_k}(\mathbb{T}_{r\mathbf{e}_0}^r)} - 1 \right| + \\
&\quad \sum_{1 \leq m \leq \frac{k}{4}} m \left(\max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1+a_2=m}} \frac{m}{k} \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \right) + \tag{2.340}
\end{aligned}$$

$$\begin{aligned}
&\quad \sum_{\frac{k}{4} < m \leq \max(r,k)} m \left(\max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1+a_2=m}} \left(\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r) + \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \right) \right) \\
&\leq O\left(\frac{r}{N}\right) + \tag{2.341}
\end{aligned}$$

$$\sum_{1 \leq m \leq \frac{k}{4}} \frac{m^2}{k} \left(\max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1+a_2=m}} \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \right) + \tag{2.342}$$

$$\sum_{\frac{k}{4} < m \leq \min(r,k)} m \left(\max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1+a_2=m}} \left(\mathbb{P}_{S'_k}(\mathbb{T}_{\mathbf{a}}^r) + \mathbb{P}_{S''_k}(\mathbb{T}_{\mathbf{a}}^r) \right) \right) \tag{2.343}$$

where both lines (2.340) and (2.341) used Lemma 2.9.5 directly. Before returning to (2.342) and (2.343), we recall from (2.319) that

$$\partial_{\mathbf{e}_j - \mathbf{e}_0}^{\mathbf{a}} L_{\mathbf{b}}(\mathbf{a}) \leq \ln\left(\frac{r}{N}\right) - \ln\left(\frac{a_j + 1}{k}\right) + O(1) \tag{2.344}$$

$$= \ln\left(\frac{rk}{N}\right) - \ln(a_j + 1) + O(1). \tag{2.345}$$

Integrating and exponentiating (2.345) for a fixed $\mathbb{T}_{\mathbf{b}} \in \mathbb{S}'_k \cup \mathbb{S}''_k$ shows that for

some $\alpha \in (0, 1)$,

$$\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \lesssim \left(\frac{rk}{N}\right)^{\Omega(a_1+a_2)} \leq \alpha^{a_1+a_2} \quad (2.346)$$

provided that δ is chosen small enough to compensate for the implicit constants in (2.345). Moreover, we can exploit the bound

$$\partial_{\mathbf{e}_j - \mathbf{e}_0}^{\mathbf{a}} L_{\mathbf{b}}(\mathbf{a}) \leq \ln\left(\frac{r}{N}\right) - \ln\left(\frac{a_j + 1}{k}\right) + O(1)$$

from (2.344) as we integrate over an interval of length $\Theta(1)$ in a region where $a_j \sim k$ to gain a decay factor of $\frac{r}{N}$. In particular, if we use the exponential decay from (2.345) everywhere else in the region of integration, we have that for any $\mathbf{a}$ with $a_1 \sim k$ or $a_2 \sim k$ (the condition $a_1 + a_2 > \frac{k}{4}$ will suffice to ensure at least one of these conditions),

$$\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \lesssim \frac{r}{N} \alpha^{a_1+a_2}. \quad (2.347)$$

Then (2.347) almost immediately provides the necessary bound for (2.343) as follows:

$$\begin{aligned} & \sum_{\frac{k}{4} < m \leq \min(r, k)} m \left(\max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1+a_2=m}} \left(\mathbb{P}_{\mathbb{S}'_k}(\mathbb{T}_{\mathbf{a}}^r) + \mathbb{P}_{\mathbb{S}''_k}(\mathbb{T}_{\mathbf{a}}^r) \right) \right) \\ & \lesssim \sum_{\frac{k}{4} < m \leq \min(r, k)} m \left(\max_{\substack{\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}'_k \cup \mathbb{S}''_k \\ \mathbf{a} \in \mathcal{D}_r^{\mathbf{b}} \cap \mathbb{N}^3 \\ a_1+a_2=m}} \mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{\mathbf{a}}^r) \right) \\ & \lesssim \sum_{m > \frac{k}{4}} \frac{r}{N} m \alpha^{-m} \\ & \lesssim \frac{r}{N}. \end{aligned} \quad (2.348)$$

Bounding (2.342) is slightly subtler, but similar. We compute for any $\mathbb{T}_{\mathbf{b}} \subset \mathbb{S}''_k$, for $j \in \{1, 2\}$,

$$\frac{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{(r-1)\mathbf{e}_0 + \mathbf{e}_j}^r)}{\mathbb{P}_{\mathbb{T}_{\mathbf{b}}}(\mathbb{T}_{r\mathbf{e}_0}^r)} = \frac{r \binom{N-r}{\mathbf{b} + (1-r)\mathbf{e}_0 - \mathbf{e}_j}}{\binom{N-r}{\mathbf{b} - r\mathbf{e}_0}} = \frac{rb_j}{b_0 - r + 1}. \quad (2.349)$$

Because we are free to choose δ arbitrarily small, we may enforce the bounds $b_0 = N - b_1 - b_2 \geq N - k \geq \frac{N}{2}$ and $r < \frac{N}{4}$. Moreover, because $b_j \lesssim k$, we may conclude from (2.349) that

$$\frac{\mathbb{P}_{\mathbb{T}_b} \left(\mathbb{T}_{(r-1)\mathbf{e}_0 + \mathbf{e}_j}^r \right)}{\mathbb{P}_{\mathbb{T}_b} \left(\mathbb{T}_{r\mathbf{e}_0}^r \right)} \lesssim \frac{rk}{N}. \quad (2.350)$$

for $j \in \{1, 2\}$ and, in particular that

$$\mathbb{P}_{\mathbb{T}_b} \left(\mathbb{T}_{(r-1)\mathbf{e}_0 + \mathbf{e}_1}^r \right), \mathbb{P}_{\mathbb{T}_b} \left(\mathbb{T}_{(r-1)\mathbf{e}_0 + \mathbf{e}_2}^r \right) \lesssim \frac{rk}{N}. \quad (2.351)$$

As noted above, (2.345) implies that $\mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_a^r)$ decays exponentially with respect to $a_1 + a_2$ so, (2.351) proves that, increasing $\alpha \in (0, 1)$ as necessary,

$$\max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 = m}} \mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_a^r) \lesssim \frac{rk}{N} \alpha^m. \quad (2.352)$$

Thus we find that as long as $m \geq 1$,

$$\begin{aligned} \max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 = m}} \mathbb{P}_{\mathbb{S}_k''}(\mathbb{T}_a^r) &\leq \max_{\substack{\mathbb{T}_b \subset \mathbb{S}_k'' \\ \mathbf{a} \in \mathcal{D}_r^b \cap \mathbb{N}^3 \\ a_1 + a_2 = m}} \mathbb{P}_{\mathbb{T}_b}(\mathbb{T}_a^r) \\ &\lesssim \left(\frac{rk}{N} \alpha^m \right) \end{aligned} \quad (2.353)$$

Applying (2.353), we can bound (2.342) as follows:

$$\begin{aligned} \sum_{1 \leq m \leq \frac{k}{4}} \frac{m^2}{k} \left(\max_{\substack{\mathbf{a} \in \mathcal{D}_r^k \cap \mathbb{N}^3 \\ a_1 + a_2 = m}} \mathbb{P}_{\mathbb{S}_k''}(\mathbb{T}_a^r) \right) &\lesssim \sum_{1 \leq m \leq \frac{k}{4}} \frac{m^2}{k} \left(\frac{rk}{N} \alpha^m \right) \\ &\leq \frac{r}{N} \sum_{m=1}^{\infty} m^2 \alpha^m \\ &\lesssim \frac{r}{N}. \end{aligned} \quad (2.354)$$

Plugging (2.348) and (2.354) into (2.343) and (2.342) proves that

$$\left| \widehat{S'_k}(\xi) - \widehat{S''_k}(\xi) \right| \lesssim \frac{r}{N}.$$

thus concluding the proposition and thereby completing the central Theorem 2.1.5. $\square$

CHAPTER 3

Dimensionally Unbounded Maximal Operators

We present two examples of families of graphs whose members' Cartesian powers' spherical maximal operators are not only dimensionally unbounded in the L^p norm for suitable p , but grow exponentially with respect to dimension. Neither is a standard definition in the literature to the best of our knowledge, however both can be readily defined in standard terms.

Although it seems quite possible that the family of global antipode graphs (to be introduced in §3.1) is entirely contained in the family of non-mean distance regular graphs (to be introduced in §3.2), the proofs are different in nature. §3.1 has the advantages of being more elementary, concrete in its bound, and conducive to concise summary. Moreover the lower bound established there applies to all $p < \infty$. §3.2 has the advantages of being more familiar in its techniques and definitions, heuristically broader in scope, and stronger in the sense that it proves a lower bound for the ball as well as the spherical maximal averaging operators.

3.1 Global Antipode Graphs

3.1.1 The Exponential Lower Bound

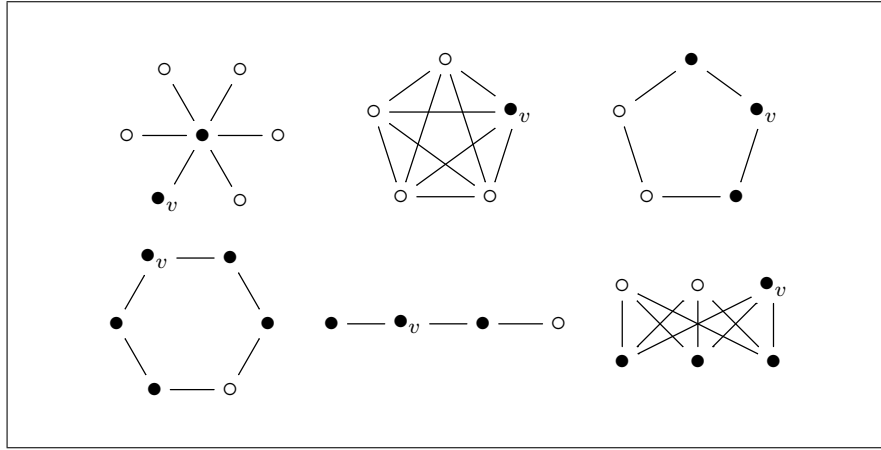
Because we will view global antipode graphs as a generalization of trees, we introduce some notation for convenience. Recall from Definition 1.3.5 that when we use the term “tree,” we mean a standard tree *with at least 3 vertices*.

Notation 7 (Leaf Set). We define the leaf set of a tree T by

$$\text{Leaf}(T) := \{x \in T : \deg(x) = 1\}$$

Definition 3.1.1 (Antipode). Let G be a graph and $x \in V(G)$ be a vertex. The antipode of x in G , denoted $\text{Ant}_G(x)$, is the (nonempty) sphere centered at x of maximal radius with respect to d_G .

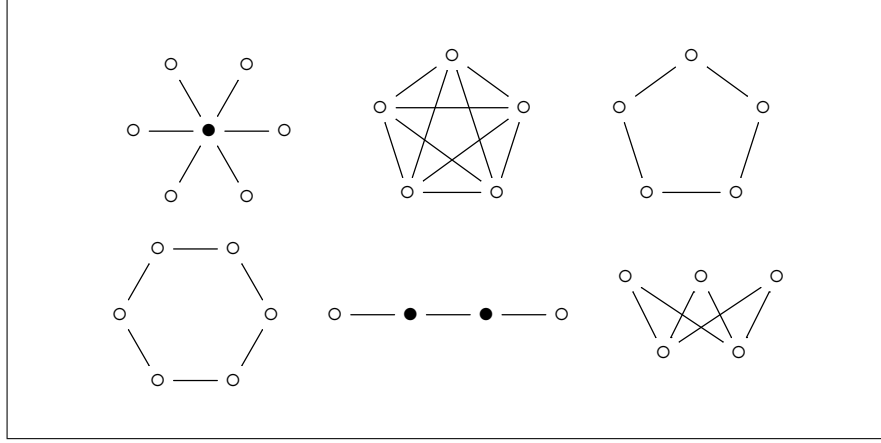
Example 3 (Examples of Antipodes). In the graphs below, the vertices in the antipode of vertex v are marked by empty circles:



Definition 3.1.2 (Global Antipode). If G is a graph, the global antipode of G is the subset of $V(G)$ given by

$$\text{Ant}(G) := \bigcup_{x \in V(G)} \text{Ant}_G(x)$$

Example 4. In the graphs below, the vertices in the global antipode are marked by empty circles:



Notice that for four of the graphs above, the global antipode is the entire vertex set. However, for the star and the path, this is not the case, leading to a key definition.

Definition 3.1.3 (Global Antipode Graph). *We call a graph G a global antipode graph (hereafter GAG) if its global antipode is a proper subset of $V(G)$, i.e. $\text{Ant}(G) \subsetneq V(G)$.*

For a tree T , it is straightforward to verify that for any vertex $x \in V(T)$, $\text{Ant}_T(x) \subset \text{Leaf}(T)$. Therefore any tree T is a GAG and $\text{Ant}(T) \subset \text{Leaf}(T) \subsetneq V(T)$. We will discuss examples of GAGs with pathological properties in §3.1.2.

Lemma 3.1.4. *If G and H are GAGs, then the Cartesian product $G \square H$ and the Cartesian power G^N are also GAGs. Moreover, $\text{Ant}(G \square H) = \text{Ant}(G) \times \text{Ant}(H)$ and $\text{Ant}(G^N) = \text{Ant}(G)^N$.*

Proof. Given GAGs G and H , it suffices to prove only the claims concerning $G \square H$ because the claims concerning G^N immediately follow by induction. Furthermore, it suffices to prove that $\text{Ant}(G \square H) = \text{Ant}(G) \times \text{Ant}(H)$ because, following Definition 3.1.3, if $\text{Ant}(G) \subsetneq V(G)$ and $\text{Ant}(H) \subsetneq V(H)$, then trivially $\text{Ant}(G) \times \text{Ant}(H) \subsetneq V(G \square H)$.

Recall from Remark 1.1.3 that

$$d_{G \square H}((g, h), (g', h')) = d_G(g, g') + d_H(h, h').$$

$$\mathbb{S}_{a+b}^{G \square H}(g, h) = \mathbb{S}_a^G(g) \times \mathbb{S}_b^H(h),$$

$$\mathbb{S}_{a+c}^{G \square H}(g, h) = \emptyset \text{ for } c > b, \text{ and}$$

$$\mathbb{S}_{d+b}^{G \square H}(g, h) = \emptyset \text{ for } d > a.$$

The upshot of the computation is that

$$\mathbb{S}_r^{G \square H}(g, h) = \text{Ant}_G(g) \times \text{Ant}_H(h) \text{ for } r = a + b \text{ and}$$

$$\mathbb{S}_r^{G \square H}(g, h) = \emptyset \text{ for } r > a + b$$

so $\text{Ant}_{G \square H}((g, h)) = \text{Ant}_G(g) \times \text{Ant}_H(h)$. Finally we observe that

$$\begin{aligned} \text{Ant}(G \square H) &= \bigcup_{\substack{g \in G \\ h \in H}} \text{Ant}_{G \square H}((g, h)) \\ &= \bigcup_{\substack{g \in G \\ h \in H}} \text{Ant}_G(g) \times \text{Ant}_H(h) \\ &= \left(\bigcup_{g \in G} \text{Ant}_G(g) \right) \times \left(\bigcup_{h \in H} \text{Ant}_H(h) \right) \\ &= \text{Ant}(G) \times \text{Ant}(H). \end{aligned}$$

proving the claim. □

The remainder of this section is dedicated to proving a dimensionally exponential lower bound for the L^p norm for the spherical maximal operator on a GAG.

Lemma 3.1.5. *For any graph G , $M \mathbb{1}_{\text{Ant}(G)}$ is identically equal to 1.*

Proof. For any $x \in V(G)$, by definition $\text{Ant}_G(x) \subset \text{Ant}(G)$. Thus, because $\text{Ant}_G(x)$ is a nonempty sphere centered at x ,

$$M^G \mathbb{1}_{\text{Ant}(G)}(x) \geq \frac{1}{|\text{Ant}_G(x)|} \left| \sum_{y \in \text{Ant}_G(x)} \mathbb{1}_{\text{Ant}(G)}(y) \right| = 1$$

The opposite inequality, $M^G \mathbb{1}_{\text{Ant}(G)}(x) \leq 1$, is nothing but the observation that averaging operators cannot increase maxima, i.e. $\|M^G\|_\infty = 1$. □

The exponential lower bound follows quickly.

Theorem 3.1.6. *If G is a GAG and M^N is the spherical maximal operator on G^N , then for all $p < \infty$, M^N satisfies the an exponential lower bound in the L^p operator norm, namely*

$$\|M^N\|_p^p \geq \left(\frac{|G|}{|\text{Ant}(G)|} \right)^N = \mathbb{P}(\text{Ant}(G))^{-N}.$$

Proof. By Lemma 3.1.4, $\text{Ant}(G^N) = \text{Ant}(G)^N$ so $|\text{Ant}(G^N)| = |\text{Ant}(G)|^N$. Thus, $\|1_{\text{Ant}(G^N)}\|_p^p = |\text{Ant}(G)|^N$ and by Lemma 3.1.5,

$$\|M^N 1_{\text{Ant}(G^N)}\|_p^p = \|1\|_p^p = |G^N| = |G|^N$$

Therefore we immediately have the desired lower bound:

$$\|M^N\|_p^p \geq \left(\frac{\|M^N 1_{\text{Ant}(G^N)}\|_p}{\|1_{\text{Ant}(G^N)}\|_p} \right)^p = \left(\frac{|G|}{|\text{Ant}(G)|} \right)^N$$

and the analogous probabilistic formulation in the theorem statement. $\square$

Theorem 3.1.7. *Let T be a tree. Then, if M^N is the spherical maximal operator on T^N , for any $p < \infty$,*

$$\|M^N\|_p^p \geq \left(\frac{|T|}{|\text{Leaf}(T)|} \right)^N = \mathbb{P}(\text{Leaf}(T))^{-N}.$$

Proof. Because any tree T is a GAG and $\text{Ant}(T) \subset \text{Leaf}(T)$ (in particular $|\text{Ant}(T)| \leq |\text{Leaf}(T)|$), Theorem 3.1.7 follows immediately from Theorem 3.1.6. $\square$

3.1.2 Pathological Examples

In all graph diagrams in this section, the vertices in the global antipode will be marked by empty circles. Moreover, unless otherwise specified, a “minimal” graph with respect to a property P is a graph G with property P such that any other graph with property P has at least as many vertices *and* at least as many edges as G . A “strongly minimal” graph with respect to a property P is a graph G with property P such that any other graph with property P has strictly more vertices *and* strictly more edges than G . We recall the following standard notation:

Notation 8. For a graph G , $\epsilon(x) = \epsilon_G(x)$ (we generally suppress the subscript) is the eccentricity of $x \in V(G)$. That is,

$$\epsilon(x) = \max\{r : \mathbb{S}_r^G(x) \neq \emptyset\}.$$

Also, $\text{rad}(G)$ and $\text{diam}(G)$ are the radius and diameter of G respectively. That is,

$$\text{rad}(G) = \min_{x \in G} \epsilon(x) \leq \max_{x \in G} \epsilon(x) = \text{diam}(G).$$

Notice that $\text{Ant}_G(x)$ is the sphere of radius $\epsilon(x)$ centered at x .

Remark 3.1.8. The proofs in this section will mostly use only elementary graph theory and are not intended to be direct contributions to the overarching project of characterizing asymptotic maximal bounds for Cartesian powers of finite graphs. Rather, they are presented to provide intuition for the relationship between structure and asymptotic maximal bounds. For this reason we present intuitive proofs rather than mechanically computing minimal examples.

Definition 3.1.9 (Eccentric Graph). We call a graph G an eccentric graph (or EG) if not all vertices have the same eccentricity. More concisely, G is eccentric if and only if $\text{rad}(G) < \text{diam}(G)$.

Notice that if G is an EG, there exist r and $x, y \in V(G)$ such that $\mathbb{S}_r^G(x) = \emptyset \neq \mathbb{S}_r^G(y)$ so that $|\mathbb{S}_r^G(x)| = 0 < |\mathbb{S}_r^G(y)|$. In particular, EGs are never DDRGs.

Lemma 3.1.10. GAGs are eccentric.

Proof. Let G be a GAG, $x \in V(G) \setminus \text{Ant}(G)$, and $y \in \text{Ant}_G(x)$. By the definition of an antipode, $d(x, y) = \epsilon(y)$ if and only if $x \in \text{Ant}_G(y)$. Because $\text{Ant}_G(y) \subset \text{Ant}(G)$, $x \notin \text{Ant}_G(y)$ so we have the inequality $d(x, y) < \epsilon(y)$. However, because $y \in \text{Ant}_G(x)$, $\text{rad}(G) \leq \epsilon(x) = d(x, y) < \epsilon(y) \leq \text{diam}(G)$. $\square$

Corollary 3.1.11. GAGs cannot be distance degree regular, distance regular, vertex transitive, or Cayley graphs.

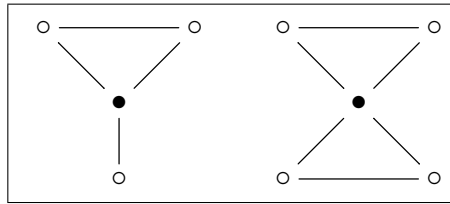
Proof. It is straightforward to verify from definitions that distance regular and vertex transitive graphs are distance degree regular. Because eccentric graphs cannot be distance regular, by Lemma 3.1.10, GAGs cannot be DDRGs. Because Cayley graphs are vertex transitive, the entire corollary follows. $\square$

Although the only examples of GAGs we have discussed so far are trees, there are many GAGs that are not trees. Heuristically, GAGs are similar to trees in that some of their vertices are more central (analogous to a root) while some are more peripheral (analogous to leaves). It is easily verified that the minimal GAG is the path P_3 :



However, the class of GAGs is strictly larger than the class of trees.

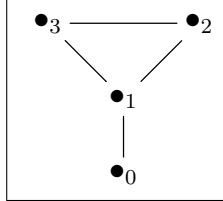
Proposition 3.1.12. *The kite graph (left) is a strongly minimal non-tree GAG and the hourglass graph (right) is a minimal GAG with no leaf vertices (strictly minimal with respect to edges):*



Proof. Calculating global antipodes is straightforward so we will only prove minimality.

Kite: The only 3 vertex graphs are P_3 and K_3 so we first prove that there are no 4 vertex non-tree GAGs other than the kite graph above. To this end, suppose G is non-tree GAG with 4 vertices. If there is no vertex of degree 3, G

is a path and thus a tree so the will let 1 be a vertex with degree 3. Notice that $\text{Ant}_G(x)$ consists of all other vertices because $\epsilon(x) = 1$. Letting 0, 2, and 3 denote the remaining 3 vertices, there must be at least one edge between 0, 2, and 3 or else G would be a tree. Without loss of generality 2 and 3 are adjacent:

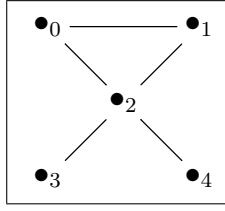


If 2 or 3 is adjacent to 0 (without loss of generality 2 is adjacent to 0), then $\epsilon(2) = 1$ so

$$V(G) = \text{Ant}_G(1) \cup \text{Ant}_G(2) \subset \text{Ant}(G)$$

and thus G is not a GAG. Therefore the kite is the only GAG with 4 vertices. It follows that there are no other non-tree GAGs with 4 edges because, if G were such a graph, it would have 4 edges and at least 5 vertices so it could only be P_5 , a tree.

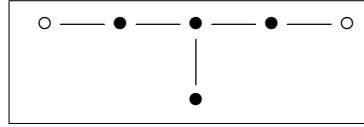
Hourglass: Suppose G is a GAG with no vertices of degree 1 (i.e. leaves). Because G is not a tree and it cannot be the kite graph, we just showed it must have at least 5 vertices and at least 5 edges. Suppose G has 5 vertices. Because each vertex has degree at least 2, any pair of non-adjacent vertices must share a neighbor simply by the pigeonhole principle so $\text{diam}(G) = 2$. By Corollary 3.1.10, $\text{rad}(G) = 1$ so there exists a vertex, which we call 2, connected to all other vertices. Moreover, because no vertices have degree 1, there must be at least (actually strictly more than) one edge connecting two of the remaining vertices so without loss of generality we assume vertices 1 and 3 are adjacent. Therefore G contains the following subgraph:



The only graph with at most one extra edge containing this subgraph is the hourglass, thus proving that all other GAGs without a vertex of degree 1 have more edges or more vertices. Notice that the only graphs with more than 5 vertices that have at most 6 edges are either trees (and thus have vertices of degree 1) or the cycle C_6 . Because cycles are vertex transitive, by Corollary 3.1.11, C_6 is not a GAG. Thus there cannot be another GAG with 6 edges and no leaves. $\square$

We observe also that the global antipode of a tree may be strictly smaller than its leaf set.

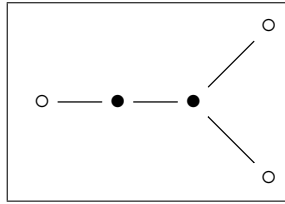
Proposition 3.1.13. *The graph below is the strongly minimal tree whose leaf set is strictly larger than its global antipode:*



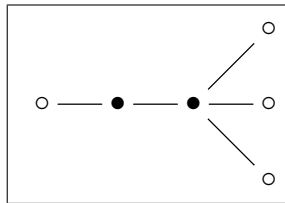
Proof. For path and star graphs, the antipode of any leaf is precisely the set of all other leaves. Thus the global antipode must include all leaves, so if G is strongly minimal among trees with the property that $\text{Leaf}(G) \neq \text{Ant}(G)$, it can be neither a star nor a path.

Because it is not a path it must have a vertex of degree at least 3 and, because it is not a star, there must be more than 3 other vertices so $|V(G)| \geq 5$. If G has 5 vertices, its vertices must have degree at most 3 as it is not a star. The only 5

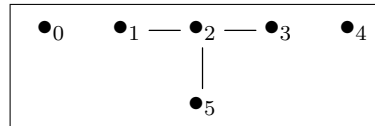
vertex tree with a vertex of degree 3 is the following:



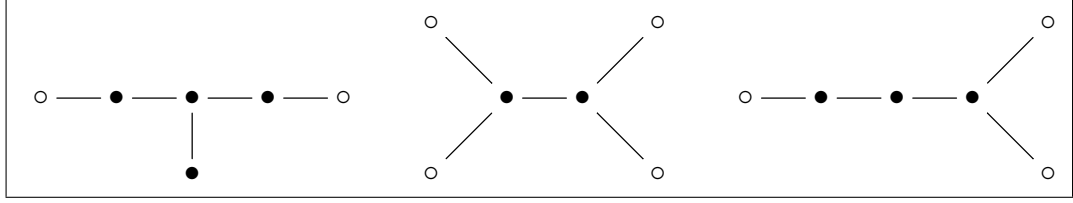
Because the leaf set and the global antipode of the above graph are equal, it must be the case that $|V(G)| \geq 6$. If G has 6 vertices, it cannot have a vertex of degree 5 as it would be a star. Moreover, the only tree with a vertex of degree 4 is the following:



Therefore the highest vertex degree in G must be 3 so it contains the following (disconnected) subgraph:



At least one of vertices 0 and 4, without loss of generality 4, must connect to the central connected component containing 1, 2, 3, and 5. Without loss of generality 4 connects to 3. Connecting 0 to one of the remaining vertices yields a graph isomorphic to one of the following:

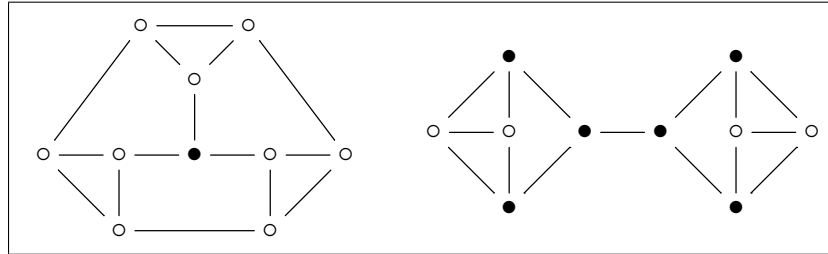


Because only the leftmost graph has a leaf outside its global antipode, this proves the claim. $\square$

Although GAGs cannot be sphere regular, they can in fact be regular.

Proposition 3.1.14. *All regular GAGs have at least 10 vertices. Moreover, there are multiple non-isomorphic examples that are minimal in both vertices and edges among regular GAGs.*

Proof. It is straightforward to verify that the following graphs are 3-regular GAGs with 10 vertices:



Moreover, they are certainly not isomorphic, as seen by noting the graph on the left has a 9 element global antipode while that on the right has a 4 element global antipode. Alternatively, one can note that the graph on the left has radius 2 and diameter 3 while that on the right has radius 3 and diameter 5.

We will now show that there cannot exist regular GAGs with fewer than 10 vertices. Notice first that any 2-regular graph is a cycle and, as in Proposition 3.1.12, cannot be a GAG. This also means that once we show there are no regular

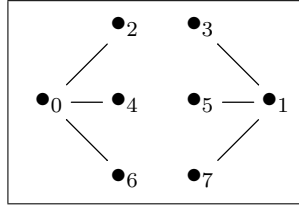
GAGs with fewer than 10 vertices, the 3-regular 10 vertex GAGs above must have minimal edges among regular GAGs.

Let G be a regular GAG with a minimal size vertex set among regular GAGs. Because we know G must be at least 3-regular, immediately we know $|V(G)| \geq 4$. Because a complete graph is vertex transitive, G cannot be K_4 . If $|V(G)| = 5$, then G must be at least 4-regular because the sum of its vertices' degrees must be even. However this cannot hold because G cannot be K_5 .

Following this line of reasoning, if $|V(G)| = 6$, G must be 3 or 4 regular. Either way, by the pigeonhole principle, if $x, y \in V(G)$ are not connected, they must share a neighbor so for any $x, y \in V(G)$, $d(x, y) \leq 2$. Therefore $\text{diam}(G) \leq 2$. However, because G is 3 or 4 regular, no vertex is connected to every other vertex and so $\text{rad}(G) \geq 2 \geq \text{diam}(G) \geq \text{rad}(G)$ so G is not eccentric. Thus, by Lemma 3.1.10, G cannot be a GAG.

Because 7 is odd, if $|V(G)| = 7$, G must be 4 or 6 regular. 6 is impossible because G is not complete and the same pigeonhole argument for the $|V(G)| = 6$ case above shows that G is not eccentric if it is 4-regular so it cannot be a GAG. Similarly if $|V(G)| = 9$, the same argument shows that it cannot be 4, 6, or 8 regular. Once again, if $|V(G)| = 8$ then G cannot be 4 (or more) regular because this would prevent it from being eccentric. Thus, as long as G is not 3-regular with 8 vertices, it must contain at least 10 vertices as claimed.

Suppose G is 3-regular and $|V(G)| = 8$. As before, $\text{diam}(G) \geq 3$ so there exist vertices 0 and 1 such that $d_G(0, 1) \geq 3$. Because 0 and 1 cannot share a neighbor, by the pigeonhole principle G must contain the following (disconnected) subgraph:



We will refer to the vertices with odd labels as “odd vertices” and even labels as “even vertices.” Because G is connected, there must be an edge between an even and an odd vertex and thus, $d_G(0, 1) = 3$. Because this holds for any pair of vertices that share no neighbors (in place of 0 and 1), $\text{diam}(G) = 3$. Therefore any vertex with eccentricity 3 is in the global antipode (in particular, 0 and 1).

Notice that the (possibly disconnected) subgraph S of G induced by the vertex set $V := \{2, 3, 4, 5, 6, 7\}$ is 2-regular because G is 3 regular and each vertex in V connects to exactly one vertex outside of V . Thus S is either composed of two 3-cycles or one 6-cycle.

Because G is a GAG, there must be a vertex, without loss of generality 2, that is not in $\text{Ant}(G)$. However, if S is composed of two 3-cycles, there must exist at least one odd vertex, without loss of generality 7, that is not in the same 3-cycle as 2 within S . This means 2 and 7 do not share a neighbor so $\epsilon_G(2) \geq 3 = \text{diam}(G)$ so $2 \in \text{Ant}(G)$, a contradiction.

This leaves only the possibility that S consists of a single 6-cycle and we can therefore partition V into pairs that are mutually antipodal in S . Suppose that the antipode of 2 in S is an odd vertex, without loss of generality 1 so $d_S(2, 3) = 3$. Then any path between 2 and 3 in G that is not contained in S must contain either 0 or 1 and will therefore have length at least 3 because $d_G(2, 1) = 2$ and $d_G(0, 3) \geq 2$. Therefore the antipode of 2 in S is even. Without loss of generality, it is 4.

Notice that $d_G(2, 4) = 2$ because they do not share an edge in $E(G) \setminus E(S)$

and, by merit of being an antipodal pair in S , $d_S(2, 4) = 3$. Because $2 \notin \text{Ant}(G)$ and thus $2 \notin \text{Ant}_G(4)$, there must be a vertex in $V(G)$ that is distance 3 (in G) away from vertex 4. This cannot be an even vertex because the even vertices are all connected to 0 or 0 itself. Moreover, because 2 is the unique antipodal vertex of 4 in S , vertices 3, 5, and 7 are no further than 2 from 4 in S and therefore in the larger graph G as well. This only leaves the possibility that $d_G(4, 1) = 3$.

However, if $d_G(4, 1) = 3$, 4 cannot share a neighbor with 1 so it cannot have an odd neighbor. Thus, by the pigeonhole principle, the neighbors of 4 are 0, 2, and 6. In particular, this implies that $d_G(2, 4) = 1$, contradicting our earlier observation that $d_G(2, 4) = 2$. $\square$

3.2 Non-Mean Distance Regular Graphs

Throughout this section, we use the notation $G = (V, E)$ for a generic graph and think of N as fixed, proving bounds independent of it. We use the notations $\{X_1, \dots, X_N\}$ and $\{Y_1, \dots, Y_N\}$ for i.i.d. sequences of uniformly chosen random vertices in V and note that therefore $\mathbf{X} := (X_1, \dots, X_N)$ and $\mathbf{Y} := (Y_1, \dots, Y_N)$ are uniformly distributed random vertices in G^N . Unless otherwise stated, capital letters refer to random variables while lower case letters signify deterministic variables. Also, unless otherwise stated, we will use numerical subscripts for vertices in G , defaulting to the subscript 1, to distinguish them from vertices in G^N , to which we give no subscripts. All implicit constants are allowed to depend on G and we will only explicitly state dependence on G in Theorem 3.2.4 to facilitate external reference. For all vertices $y_1 \in V$, let

$$X_n^{y_1} := \mathbb{1}_{\{X_n = y_1\}} \text{ and}$$

$$X^{y_1} := \sum_{n=1}^N X_n^{y_1}.$$

Lemma 3.2.1. Fix $\epsilon > 0$. Then

$$\mathbb{P} \left(\bigcup_{y_1 \in V} \left\{ \left| \frac{X^{y_1}}{N} - \frac{1}{|V|} \right| > \epsilon \right\} \right) \lesssim e^{-\Omega_\epsilon(N)}.$$

Proof. By the central limit theorem, for any fixed $y_1 \in V$,

$$\mathbb{P} \left(\left| \frac{X^{y_1}}{N} - \frac{1}{|V|} \right| > \epsilon \right) = e^{-\Omega_\epsilon(N)}. \quad (3.1)$$

Applying the union bound yields

$$\mathbb{P} \left(\bigcup_{y_1 \in V} \left\{ \left| \frac{X^{y_1}}{N} - \frac{1}{|V|} \right| > \epsilon \right\} \right) \leq |V| e^{-\Omega_\epsilon(N)} \lesssim e^{-\Omega_\epsilon(N)}, \quad (3.2)$$

the desired result. $\square$

Lemma 3.2.2. Fix $\epsilon > 0$ and $x \in V^N$. For all $y_1 \in V$, let

$$p_{y_1}^x = \frac{|\{n : x_n = y_1\}|}{N}.$$

Then we have the bound

$$\mathbb{P} \left(\frac{d(x, \mathbf{Y})}{N} < \sum_{y_1 \in V} p_{y_1}^x \mathbb{E} d(y_1, X_1) - \epsilon \right) = e^{-\Omega_\epsilon(N)}.$$

Proof. For a fixed $y_1 \in V$ such that $p_{y_1}^x > \epsilon$, $\sum_{n: x_n=y_1} \frac{d(y_1, Y_n)}{N}$ is a normalized sum of $Np_{y_1}^x \geq \epsilon N = \Omega_\epsilon(N)$ i.i.d. random variables with expectation $\mathbb{E} d(y_1, X_1)$. Thus when $p_{y_1}^x > \epsilon$ the central limit theorem implies

$$\mathbb{P} \left(\sum_{n: x_n=y_1} \frac{d(y_1, Y_n)}{N} < p_{y_1}^x \mathbb{E} d(y_1, X_1) - \epsilon \right) \lesssim e^{-\Omega_\epsilon(N)} \quad (3.3)$$

and, applying a union bound and (3.3), we find

$$\begin{aligned} & \mathbb{P} \left(\sum_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} \sum_{n: x_n=y_1} \frac{d(y_1, Y_n)}{N} < \sum_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} p_{y_1}^x \mathbb{E} d(y_1, X_1) - \epsilon \right) \\ & \leq \mathbb{P} \left(\bigcup_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} \left\{ \sum_{n: x_n=y_1} \frac{d(y_1, Y_n)}{N} < p_{y_1}^x \mathbb{E} d(y_1, X_1) - \frac{\epsilon}{|V|} \right\} \right) \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} \mathbb{P} \left(\sum_{n: x_n = y_1} \frac{d(y_1, Y_n)}{N} < p_{y_1}^x \mathbb{E} d(y_1, X_1) - \frac{\epsilon}{|V|} \right) \\
&\leq \sum_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} e^{-\Omega_\epsilon(N)} \\
&\leq |V| e^{-\Omega_\epsilon(N)} \\
&\lesssim e^{-\Omega_\epsilon(N)}
\end{aligned} \tag{3.4}$$

Note also that

$$\begin{aligned}
\sum_{y_1 \in V} p_{y_1}^x \mathbb{E} d(y_1, X_1) &= \sum_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} p_{y_1}^x \mathbb{E} d(y_1, X_1) + \sum_{\substack{y_1 \in V \\ p_{y_1}^x < \epsilon}} p_{y_1}^x \mathbb{E} d(y_1, X_1) \\
&\leq \sum_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} p_{y_1}^x \mathbb{E} d(y_1, X_1) + [|V|\epsilon] [\mathbb{E} d(y_1, X_1)] \\
&\leq \sum_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} p_{y_1}^x \mathbb{E} d(y_1, X_1) + \epsilon(|V| \text{diam}(G))
\end{aligned} \tag{3.5}$$

We compute,

$$\begin{aligned}
\frac{d(x, \mathbf{Y})}{N} &= \sum_{n=1}^N \frac{d(x_n, Y_n)}{N} \\
&= \sum_{y_1 \in V} \sum_{n: x_n = y_1} \frac{d(y_1, Y_n)}{N} \\
&\geq \sum_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} \sum_{n: x_n = y_1} \frac{d(y_1, Y_n)}{N}.
\end{aligned} \tag{3.6}$$

Combining (3.5) and (3.6), we can infer the following implication:

$$\begin{aligned}
\frac{d(x, \mathbf{Y})}{N} &< \sum_{y_1 \in V} p_{y_1}^x \mathbb{E} d(y_1, X_1) - \epsilon(|V| \text{diam}(G) + 1) \\
\implies \sum_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} \sum_{n: x_n = y_1} \frac{d(y_1, Y_n)}{N} &< \sum_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} p_{y_1}^x \mathbb{E} d(y_1, X_1) - \epsilon.
\end{aligned} \tag{3.7}$$

Using implication (3.7) for a probabilistic bound on the respective events, we have

$$\begin{aligned}
& \mathbb{P} \left(\frac{d(x, \mathbf{Y})}{N} < \sum_{y_1 \in V} p_{y_1}^x \mathbb{E} d(y_1, X_1) - \epsilon (|V| \text{diam}(G) + 1) \right) \\
& \leq \mathbb{P} \left(\sum_{\substack{y_1 \in V \\ p_{y_1}^x \geq \epsilon}} \sum_{n: x_n = y_1} \frac{d(y_1, Y_n)}{N} < \sum_{y_1 \in V} p_{y_1}^x \mathbb{E} d(y_1, X_1) - \epsilon \right) \\
& \lesssim e^{-\Omega_\epsilon(N)}
\end{aligned} \tag{3.8}$$

by (3.4). Scaling ϵ by a factor of $(|V| \text{diam}(G) + 1)$ and renaming proves the lemma. $\square$

Definition 3.2.3 (Mean Distance Regular Graph). *We call $G = (V, E)$ a mean distance regular graph (hereafter MDRG) if the quantity*

$$\mathbb{E} d(x, X_1)$$

is independent of $x_1 \in V$.

Stated a bit more intuitively, an MDRG is a graph such that the expected distance of a random vertex from a fixed vertex x is the same regardless of x .

Theorem 3.2.4. *If a graph $G = (V, E)$ is not an MDRG and M_B^N is the ball maximal operator on G^N , then for some $p > 1$, M_B^N satisfies an exponential lower bound (with respect to N) depending G in the L^p operator norm, i.e.*

$$\|M_B^N\|_p \geq e^{\Omega_G(N)}.$$

Proof. Fix a small $\epsilon > 0$ only depending on G to be chosen later. By Lemma 3.2.1, we have the bound

$$\begin{aligned}
& \mathbb{P} \left(\bigcap_{y_1 \in V} \left\{ \left| \frac{X^{y_1}}{N} - \frac{1}{|V|} \right| \leq \epsilon \right\} \right) \\
& = 1 - \mathbb{P} \left(\bigcup_{y_1 \in V} \left\{ \left| \frac{X^{y_1}}{N} - \frac{1}{|V|} \right| > \epsilon \right\} \right) = 1 - o(1) = \Omega(1).
\end{aligned} \tag{3.9}$$

Thus, in the language of Lemma 3.2.2, the probability that for all $y_1 \in V$, $\left|p_{y_1}^{\mathbf{X}} - \frac{1}{|V|}\right| < \epsilon$ is asymptotically almost certain and, in particular, has positive (lower) density. Let $x = (x_1, \dots, x_N) \in V^N$ be such a vertex, i.e.

$$\frac{1}{|V|} - \epsilon < p_{y_1}^x < \frac{1}{|V|} + \epsilon \quad (3.10)$$

for all $y_1 \in V$.

By Lemma 3.2.2,

$$\begin{aligned} & \mathbb{P} \left(\frac{d(x, \mathbf{Y})}{N} < \sum_{y_1 \in V} \frac{1}{|V|} \mathbb{E} d(y_1, X_1) - [|V| \text{diam}(G) + 1] \epsilon \right) \\ & \leq \mathbb{P} \left(\frac{d(x, \mathbf{Y})}{N} < \sum_{y_1 \in V} \left[\frac{1}{|V|} - \epsilon \right] \mathbb{E} d(y_1, X_1) - \epsilon \right) \\ & \leq \mathbb{P} \left(\frac{d(x, \mathbf{Y})}{N} < \sum_{y_1 \in V} p_{y_1}^x \mathbb{E} d(y_1, X_1) - \epsilon \right) \\ & = e^{-\Omega_\epsilon(N)}. \end{aligned} \quad (3.11)$$

Moreover, if

$$\begin{aligned} r &:= \sum_{y_1 \in V} \frac{1}{|V|} \mathbb{E} d(y_1, X_1) - [|V| \text{diam}(G) + 1] \epsilon \\ &= \mathbb{E} d(Y_1, X_1) - [|V| \text{diam}(G) + 1] \epsilon \end{aligned}$$

we may rephrase (3.11) in the language of balls in the G^N to conclude

$$\mathbb{P}(\mathbb{B}_{rN}(x)) = e^{-\Omega_\epsilon(N)}. \quad (3.12)$$

Because G is not mean distance regular, there exists a vertex $z_1 \in V$ such that

$$\mathbb{E} d(z_1, X_1) < \mathbb{E} d(Y_1, X_1) \quad (3.13)$$

Defining $z := (z_1, \dots, z_1) \in V^N$, we use (3.10) to compute

$$d(x, z) = \sum_{n=1}^N d(x_n, z_1)$$

$$\begin{aligned}
&= \sum_{y_1 \in V} |\{n : x_n = y_1\}| d(y_1, z_1) \\
&= \sum_{y_1 \in V} N p_{y_1}^x d(y_1, z_1) \\
&< \sum_{y_1 \in V} N \left(\frac{1}{|V|} + \epsilon \right) d(y_1, z_1) \\
&\leq N \sum_{y_1 \in V} \frac{d(y_1, z_1)}{|V|} + N\epsilon |V| \text{diam}(G) \\
&= N [\mathbb{E} d(z_1, Y_1) + \epsilon |V| \text{diam}(G)]
\end{aligned} \tag{3.14}$$

or, more succinctly, defining

$$r' := \mathbb{E} d(z_1, Y_1) + \epsilon |V| \text{diam}(G),$$

we can conclude that

$$z \in \mathbb{B}_{r'N}(x). \tag{3.15}$$

Looking back on (3.13), we observe that if ϵ sufficiently small relative to $|V|$ and $\text{diam}(G)$ (in particular, ϵ will depend only on G) then $r' \leq r$ so (3.15) implies

$$z \in \mathbb{B}_{rN}(x). \tag{3.16}$$

Thus, if M^B is the ball maximal averaging operator on G^N then (3.12) implies

$$\begin{aligned}
M^B \delta_z(x) &= \max_{0 \leq k \leq N \text{diam}(G)} \mathbb{E}_{\mathbb{B}_k(x)} \delta_z \\
&= \max_{0 \leq k \leq N \text{diam}(G)} \mathbb{1}_{\{d(x,z) \leq k\}} \mathbb{E}_{\mathbb{B}_k(x)} \delta_z \\
&= \max_{0 \leq k \leq N \text{diam}(G)} \mathbb{1}_{\{d(x,z) \leq k\}} \frac{\mathbb{P}(\{z\})}{\mathbb{P}(\mathbb{B}_k(x))} \\
&\geq \frac{|V|^{-N}}{\mathbb{P}(\mathbb{B}_r(x))} \\
&\geq \frac{|V|^{-N}}{e^{-\Omega_\epsilon(N)}} \\
&= \frac{e^{\Omega_\epsilon(N)}}{|V|^N}.
\end{aligned} \tag{3.17}$$

Recall from (3.9) that the set

$$S := \left\{ x \in V^N : \forall y_1 \in V, \frac{1}{|V|} - \epsilon < p_{y_1}^x < \frac{1}{|V|} + \epsilon \right\}$$

has positive density, i.e. $|S| \sim |V|^N$. Thus, (3.17) implies that for any $p \in [1, \infty)$,

$$\|M^B \delta_z\|_p^p \geq \left(\frac{e^{\Omega_\epsilon(N)}}{|V|^N} \right)^p |S| = |V|^{N(1-p)} e^{\Omega_\epsilon(N)} \quad (3.18)$$

so for some $A > 1$ depending only on G ,

$$\|M^B \delta_z\|_p^p \geq |V|^{N(1-p)} A^N = \left(\frac{A}{|V|^{p-1}} \right)^N. \quad (3.19)$$

If we choose $p > 1$ sufficiently small that $A > |V|^{p-1}$, we have

$$\|M^B \delta_z\|_p \geq \left[\left(\frac{A}{|V|^{p-1}} \right)^{1/p} \right]^N = e^{\Omega_G(N)}, \quad (3.20)$$

thus proving the theorem. $\square$

Corollary 3.2.5. *If G is not an MDRG and M^N is its spherical maximal averaging operator, then for some $p > 1$, we have the lower bound*

$$\|M^N\|_p \geq e^{\Omega_G(N)}.$$

Proof. This is nothing but Theorem (3.2.4) in conjunction with the standard observation that the spherical maximal operator pointwise dominates the ball maximal operator. $\square$

It is straightforward to see that DDRGs (Definition 1.3.2) are MDRGs. It is also straightforward to calculate whether or not a given graph is mean distance regular. For the sake of concreteness, we present an explicit family of non-MDRGs. For this short proof, we lift our convention on vertex subscripts because there is no discussion of, and therefore no ambiguity due to, Cartesian powers.

Proposition 3.2.6. *If $G = (V, E)$ is a graph with at least 3 vertices and has a vertex of degree 1, it is not an MDRG.*

Proof. Let $y \in V$ be a degree 1 vertex and let x be its (unique) neighbor. For any $z \in V \setminus \{x, y\}$, let $(z_0, \dots, z_k)$ be a minimal path from y to z , i.e. $z_0 = y$, $z_k = z$, and $d(y, x) = k$. Because x is the only neighbor of $y = z_0$, it must be the case that $x = z_1$ and therefore $d(x, z) \leq k - 1 < k$. Then, if X_1 is a uniformly random vertex in V ,

$$\begin{aligned}
\mathbb{E} d(y, X_1) &= \sum_{z \in V} \frac{d(y, z)}{|V|} \\
&= \sum_{z \in V \setminus \{x, y\}} \frac{d(y, z)}{|V|} + \frac{d(y, y)}{|V|} + \frac{d(y, x)}{|V|} \\
&= \sum_{z \in V \setminus \{x, y\}} \frac{d(y, z)}{|V|} + 0 + \frac{1}{|V|} \\
&= \sum_{z \in V \setminus \{x, y\}} \frac{d(y, z)}{|V|} + \frac{d(x, y)}{|V|} + \frac{d(x, x)}{|V|} \\
&> \sum_{z \in V \setminus \{x, y\}} \frac{d(x, z)}{|V|} + \frac{d(x, y)}{|V|} + \frac{d(x, x)}{|V|} \\
&= \sum_{z \in V} \frac{d(x, z)}{|V|} \\
&= \mathbb{E} d(x, X_1),
\end{aligned}$$

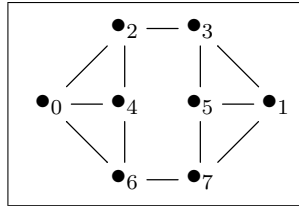
thus forcing G to fail the criterion of mean distance regularity. $\square$

Corollary 3.2.7. *All trees are non-MDRGs.*

Proof. Because we defined trees to have at least 3 vertices in §1.3 and all trees have some vertices of degree 1, this follows immediately from Proposition 3.2.6. $\square$

Proposition 3.2.8. *There exist regular graphs that are not mean distance regular. Moreover all such graphs have at least 8 vertices and 12 edges.*

Proof. Direct computation reveals that in the graph



the expected distance from 0 is $\frac{13}{7}$ while the expected distance from 2 is $\frac{12}{7}$ and thus it is not mean distance regular. Note that distance degree regularity immediately implies mean distance regularity so any regular non-MDRG must be regular of degree at least 3, as otherwise it would be a cycle and thus a DDRG. In the proof of Proposition 3.1.14 we concluded that any non-cycle regular graph of degree below 8 must have diameter at most 2. Because cliques are clearly mean distance regular, we immediately disqualify the diameter 1 case.

If a diameter 2 graph $G = (V, E)$ is regular of degree a , then for any vertex $x \in V$ has

$$|\mathbb{S}_2(x)| = |V| - |\mathbb{S}_1(x)| - |\mathbb{S}_0(x)| = |V| - a - 1.$$

In particular, this means all vertices have the same number of vertices any fixed distance in $\{0, 1, 2\}$, which exhausts all possible distances. Thus G is distance degree and therefore mean distance regular. Therefore, any regular non-MDRG must have at least 8 vertices and be regular of degree at least 3 and thus must have at least 12 edges. This concludes the proof. $\square$

We finish §3.2 by recalling from §3.1 that the family of GAGs contains all trees and some regular graphs, thus providing two distinct proofs of the extreme failure of trees and the weakness of regularity in containing Cartesian powers' maximal averaging operators' norms.

3.3 L^p Asymptotics for Other Graphs

We briefly mention those graphs whose Cartesian powers' maximal operators are substantially better behaved than those of GAGs and non-MDRGs, but not as well behaved as the graphs satisfying the conditions set forth in Chapter 2. The natural example is the family of distance degree regular graphs (Definition 1.3.2).

If G is an arbitrary DDRG, almost immediately from the definition, it follows that the spherical averaging operators are L^p contractions for all $p \in [1, \infty]$. In fact it is straightforward to show that this is an equivalent definition. Because the spherical averaging operators on the powers G^N are convex combinations of tensor products of the spherical averaging operators on G , it follows that they too are all L^p contractions. This quickly provides a dimensionally linear bound on the L^1 norm of the spherical maximal averaging operator.

Proposition 3.3.1. *If G is a DDRG of diameter d and M^{G^N} is the spherical maximal averaging operator on G^N , then*

$$\left\| M^{G^N} f \right\|_1 \lesssim (dN + 1) \|f\|_1$$

for all $f \in L^1(G^N)$.

Proof. It suffices to prove the claim for $f \geq 0$, in which case we have the pointwise bound

$$\begin{aligned} M^{G^N} f(x) &= \max_{0 \leq k \leq \text{diam}(G^N)} f(x) \\ &= \max_{0 \leq k \leq dN} f(x) \\ &= \sum_{k=0}^{dN} S_k f(x) \end{aligned}$$

Taking L^1 norms and applying the triangle inequality then reveals

$$\left\| M^{G^N} f \right\|_1 \leq \sum_{k=0}^{dN} \|S_k f\|_1. \quad (3.21)$$

Because G is a DDRG and therefore, as noted above, S_k is an L^1 contraction for all k , (3.21) implies

$$\left\| M^{G^N} f \right\|_1 \leq \sum_{k=0}^{dN} \|f\|_1 = (dN + 1) \|f\|_1$$

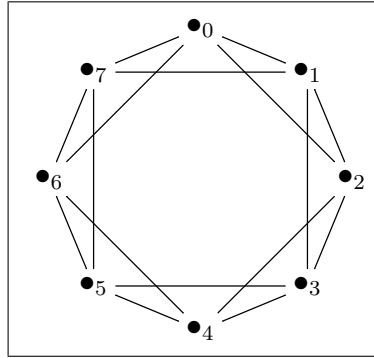
as claimed. □

Corollary 3.3.2. *If G is a DDRG and M^{G^N} is the spherical maximal averaging operator on G^N , then*

$$\left\| M^{G^N} \right\|_p \lesssim_G N^{\frac{1}{p}}.$$

Proof. We simply linearize M^G as we did in the final proof of Theorem 2.1.5 and interpolate the L^1 bound from Proposition 3.3.1 against the trivial L^∞ bound. $\square$

While Proposition 3.3.1 and Corollary 3.3.2 are neither particularly difficult to prove nor illuminating in their own right, they provide the only established dimensional asymptotics for graph spherical (or ball for that matter) maximal averaging operators that lie between exponential growth and a constant bound. Moreover, there is a subfamily of DDRGs of particular interest in our present discussion, namely those whose spherical averaging operators are pairwise commutative. Such graphs need not be distance regular; for instance in §2.1.3 we mentioned the Cayley graph of the group $\mathbb{Z}_8$ with respect to the generator set $\{1, 2, 6, 7\}$:



Because $\mathbb{Z}_8$ is abelian, it follows that its spherical averaging operators commute with one another. However, it is straightforward to verify that it is not distance regular. Although distance regularity was an important property in operator semigroup machinery of §2.5, it far played a less pronounced role in the spectral machinery of §2.7, §2.8, and §2.9. The crucial fact about distance regular graphs

in those sections was that their spherical averaging operators are simultaneously diagonalizable and that the radius 1 spherical averaging operator has a spectral gap. Both of these properties are satisfied by the graph above.

Although distance regularity is quite possibly necessary for a dimension-independent bound on the smooth spherical maximal operator

$$M_S f(x) := \max_{0 \leq K \leq \text{diam}(G^N)} \frac{1}{K+1} \sum_{k=0}^K S_k f(x),$$

we can quickly attain a logarithmic bound simply using the pointwise domination

$$\begin{aligned} M_S^{G^N} f(x) &= \max_{0 \leq K \leq \text{diam}(G^N)} \frac{1}{K+1} \sum_{k=0}^K S_k f(x) \\ &= \max_{0 \leq K \leq \text{diam}(G^N)} \sum_{k=0}^K \frac{1}{k+1} S_k f(x) \end{aligned}$$

and proceeding as in Proposition 3.3.1. If commuting spherical averaging operators and a spectral gap are indeed sufficient to carry over the spectral arguments of §2.7, §2.8, and §2.9, it seems entirely possible that the dimensionally logarithmic bound on $M_S^{G^N}$ could be extended to the spherical maximal operator M^{G^N} . This would be a substantial and interesting improvement on the bounds of Proposition 3.3.1 and Corollary 3.3.2 for this case. However, this would require a very close examination of the computationally heavy spectral arguments, without which we are not sufficiently confident to positively affirm the validity of such a logarithmic bound.

CHAPTER 4

Future Questions

We organize the future questions by our best estimate as to the size of the project. That said, keep in mind that the labels are relative. There are many other relevant questions to this thesis, but we only list a few that we deemed particularly natural, accessible, and/or interesting.

4.1 Small Questions

Question 1. *Which DRGs' Cartesian powers' spherical and/or ball maximal averaging operators' L^2 norms are bounded by Theorem 1.4.1?*

We verified that C_5 and the Petersen graph fall under the umbrella of Theorem 1.4.1. It is fairly straightforward to program a computer to test other graphs.

Question 2. *Does there exist an MDRG that is not a DDRG?*

If the answer is yes, we would expect the example to have relatively few vertices because more vertices imply more constraints. Therefore it is quite possible that if the answer to Question 2 is yes, then one could explicitly find an example simply by programming a computer to test the two properties and compare for small graphs. If one is unable to find such an example, then it is a good idea to simply try and prove the answer is no from elementary graph properties. Our guess is that the answer is no, but we are not confident in that assessment.

Question 3. *Do all non-DDRGs' Cartesian powers' spherical and/or ball maximal averaging operators' L^p norms grow exponentially in dimension for all $1 \leq p < \infty$?*

Although we have not yet shown that the exponential growth phenomenon holds for *all* $p < \infty$ in the special non-MDRG case, the arguments of §3.2 are not optimized and it seems unlikely that we have used their full power. If we are able to extend the result for all $p < \infty$ in the non-MDRG case, it is plausible that such an argument could be extended to general DDRGs. Moreover, if the answer to Question 2 is no, then MDRGs and DDRGs are equivalent so the distinction is moot.

Question 4. *Do all graphs' Cartesian powers' spherical maximal averaging operators' weak type $(1, 1)$ norms grow on the order $\Omega(\sqrt{N})$?*

We are confident that the answer to Question 4 is yes. The idea is that in a DDRG the sphere $\mathbb{S}_k$ of maximal probability in G^N should have $\mathbb{P}(\mathbb{S}_k) \sim \sqrt{N}$ by log concavity. This immediately settles the question simply by applying the maximal operator to an appropriate (in the DDRG case, arbitrary) δ -mass. See [3], Proposition 4.1 for more details on this method.

Question 5. *What are the sharp weak-type $(1, 1)$ asymptotics for DDRGs' Cartesian powers' spherical maximal averaging operators?*

A solution to Question 5 is likely to be comprised of purely combinatorial estimates, but we do not have a strong sense of where to start.

4.2 Medium Questions

Question 6. *Does the approach of bounding Mf by $M_S f$ and an L^2 bounded square function carry over to graphs of diameter greater than 2?*

The answer to Question 6 seems very likely to be yes. Throughout Chapter 2 we laid out much of the necessary machinery. Indeed the only parts that need active adaptation to diameter greater than 2 are §2.9 and the latter half of §2.8. The task is substantial but more straightforward than the other questions in §4.2.

Question 7. *If a graphs' Cartesian powers' spherical maximal averaging operators satisfy dimension-independent L^p bounds for some $p < \infty$, must the spherical averaging operators of the base graph commute with one another?*

We expect the answer to Question 7 to be yes. However, this is based on the importance of a commuting spherical averaging operators to the spectral techniques that we used in §2.7 rather than the existence of a natural method of proof.

Question 8. *Does there exist a graph whose Cartesian powers' spherical and/or ball maximal averaging operators satisfy dimension-independent L^2 bounds but have dimensionally unbounded L^p norms for some $p \in (1, 2)$?*

We suspect the answer to Question 8 is no. The spectral machinery of §2.7 in conjunction with the complex interpolation techniques used in [3], [7], [9], and [10] should carry over to the context of any graph whose spherical averaging operators commute with one another. However, if this hunch is correct, it would require a more sophisticated refinement of §2.8 in order to achieve spectral bounds on higher order differences.

Question 9. *Can a statement analogous to Corollary 2.9.4 (and, by extension, Proposition 2.9.1) be proven in the absence of a spectral gap on the base graph (i.e. for bipartite graphs)?*

We strongly suspect the answer to Question 9 is yes mostly based on such an argument for the canonical case of K_2 in [8]. That said, the fact that allowing the diameter of the base graph to exceed 1 allows for a mix of “oscillating,” i.e.

eigenvalue -1 , and “decaying,” i.e. eigenvalue in $(-1, 1)$, eigenvectors makes this more of a technical challenge than the authors encountered in [8]. Still, we expect that an argument can be built from some combination of those in §2.9 and [8], Lemma 7.

It is also possible that maximal bounds, at least for balls, for bipartite DRGs can be achieved by halving (see [2], page 25) them and reducing the problem to one about non-bipartite graphs. We are not especially optimistic about this approach, however it has the potential to circumvent a great deal of heavy combinatorial estimates and may therefore be worth further examination before attempting to adapt Corollary 2.9.4 and Proposition 2.9.1 to the non-bipartite case.

Question 10. *Does there exist a DRG whose Cartesian powers’ spherical and/or ball maximal averaging operators have unbounded norms in L^2 ?*

This is a very interesting question and likely the most difficult in §4.2. Recall that dimension-independent L^2 bounds on the smooth maximal operators for a distance regular graph were crucial in our argument for Theorem 1.4.1 in Chapter 2. In §2.4, §2.5, and §2.6 we settled on a criterion for such a bound based on whether the (unique) radial noise process can be obtained by a family of natural noise processes. This relationship between radial and natural noise operators may well be a crucial property and perhaps even an important piece of Question 11 below. However, there are other noise semi-groups to compare with the radial noise process. Moreover, it is not obvious that the realizability of the radial noise process by natural noise processes is essential for the main bound.

4.3 Main Question

Question 11. *Which graphs’ Cartesian powers’ spherical and/or ball maximal averaging operators satisfy dimension-independent L^p bounds for all $p \geq 2$?*

Question 11 is ultimately at the heart of this thesis. Although our ostensible goal was to characterize the relationship between a graph and the dimensional asymptotics of its spherical and/or ball maximal averaging operators' norms, the project was set into motion by [8], a dimension-independent bound, and remains focused on this phenomenon as the most interesting and non-intuitive part of our study of dimensional asymptotics.

The answer may well be the set of all distance regular graphs. Indeed, this would be quite an elegant and (subjectively, of course) desirable resolution to our investigation. Unfortunately, whether or not this is correct, we are not yet within striking distance of Question 11.

If the answer to Question 10 is yes, that would immediately invalidate the conjecture that the answer to Question 11 is the set of distance regular graphs. If Question 10 is unresolved, Questions 6 and 9 would be prerequisites to an affirmative answer to the conjecture.

Even with Questions 6 and 9 resolved, we would still need to deal with finding an optimal way to compare the radial and natural noise processes or find an alternative method, as suggested in our response to Question 10. This still leaves open the question of whether there could be a non-DRG whose Cartesian powers' spherical and/or ball maximal averaging operators satisfy dimension-independent L^p bounds for some $p < \infty$. Keep in mind that many of the techniques from Chapter 2 extend to wider classes of graphs. Indeed, a great deal of our proofs concerning bounding the smooth maximal operator are sufficiently combinatorial that they would hold for any DDRG. Although §2.7, §2.8, and §2.9 require further structure, many of those arguments apply to any DDRG whose spherical averaging operators commute with one another, as we alluded to in §3.3.

In short, many hurdles remain between our current understanding of the dimension-independence phenomenon and answering the central Question 11. In spite of that, we feel that Theorem 1.4.1 and, even more so, the tools used to

prove it represent a substantial step in the right direction. We adapted a great deal of the combinatorial machinery necessary for dealing with the concentration of mass in spheres in Cartesian powers of DDRGs with diameter greater than 1 and established the behavior the radial noise mass. More profoundly, we have opened up the use of unnormalized semigroups as a means to bound this elusive radial noise process and built up the probabilistic approach of §2.8 to establish spectral smoothness in the absence of the minimal diameter and maximal symmetry of the clique as a base graph. We hope that future work can harness the power of these tools to settle Question 11 in its entirety.

REFERENCES

- [1] Abramowitz, M.; Stegun, I. Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables. National Bureau of Standards, Applied Mathematics Series, 55. U.S. Government Printing Office, Washington D.C., 1972.
- [2] Brouwer, A. E.; Cohen, A. M.; Neumaier, A. Distance Regular Graphs. Ergebnisse der Mathematik und ihrer Grenzgebiete, 18. Springer-Verlag, 1989.
- [3] Krause, B. Dimension-Free L^p -Maximal Inequalities for Spherical Means in the Hypercube. <http://arxiv.org/abs/1309.4466>.
- [4] Ceccherini-Silberstein, T.; Scarabotti, F.; Tolli, F. Harmonic Analysis on Finite Groups. Cambridge Studies in Advanced Mathematics, 108. Cambridge University Press, Cambridge, 2008.
- [5] Dunford, N.; Schwartz, J. T. Linear Operators. Part I. General theory. With the assistance of William G. Bade and Robert G. Bartle. Reprint of the 1958 original. Wiley Classics Library. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1988.
- [6] Greenblatt, J. Dimensionally Exponential Lower Bounds on the L^p Norms of the Spherical Maximal Operator for Cartesian Powers of Finite Trees and Related Graphs. <http://arxiv.org/abs/1509.02843>.
- [7] Greenblatt, J.; Kolla, A.; Krause, B. Dimension-Free L^p Maximal Inequalities in $\mathbb{Z}_{m+1}^N$. <http://arxiv.org/abs/1209.4148>.
- [8] Harrow, A.; Kolla, A.; Schulman, L. Dimension-Free L^2 Maximal Inequality for Spherical Means in the Hypercube. Theory of Computing (2014), vol. 10, 55-75
- [9] Nevo, A.; Stein, E. M. A Generalization of Birkhoff's Pointwise Ergodic Theorem. Acta Math. 173 (1994), no. 1, 135-154.
- [10] Stein, E. M. On the Maximal Ergodic Theorem. Proc. Nat. Acad. Sci. U.S.A. 47 1961 1894-1897.
- [11] Stein, E. M.; Strömberg, J.O. Behavior of maximal functions in $\mathbb{R}^n$ for large n . Ark. Mat. 21 (1983), no. 2, 259-269.