Title
Geometry of Local-spectral Expanders

Author
Liu, Siqi

Publication Date
2023

Peer reviewed|Thesis/dissertation

Geometry of Local-spectral Expanders

By

Siqi Liu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Alessandro Chiesa, Chair
Professor Prasad Raghavendra
Professor Tselil Schramm
Professor Avishay Tal

Summer 2023

Geometry of Local-spectral Expanders

Abstract

Geometry of Local-spectral Expanders

by

Siqi Liu

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Alessandro Chiesa, Chair

Expanders are well-connected graphs. They have numerous applications in constructions of error correcting codes, metric embedding, derandomization, sampling algorithms, etc. Local-spectral expanders (HDXes) are a generalization of expander graphs to hypergraphs. They have recently received more attention due to their applications to agreement tests [24], locally testable codes [28, 99, 75, 27], hardness of SoS refutation [25, 59], and connections with local sampling algorithms [5].

In comparison to expanders we have very limited understanding of HDXes: there are abundant random or explicit constructions of sparse expander graphs such as random d-regular graphs [50], algebraic expanders [92, 51], the zig-zag product [101], etc. In contrast, we know only two general constructions of sparse HDXes: the LSV complexes [90] and the coset construction [66]. In this thesis, we take two approaches to tackle the construction problem. The first approach is taking graph products of sparse expander graphs. This is inspired by the zig-zag product. However, this construction fails to give good local-spectral expansion. The second approach is inspired by the following question: does any continuous space have the local-spectral expansion property? We show that the answer is affirmative for high-dimensional spheres. More precisely, we show that 3-uniform hypergraphs sampled randomly over high-dimensional spheres are (relatively sparse) local-spectral expanders.

Furthermore, tight isoperimetric inequalities of local-spectral expanders have remained elusive. Intuitively, isoperimetric inequalities provide a lower bound on the probability that a random walk leaves a subset of vertices in the graph. A tight bound on this probability is crucial for applications to agreement testing. In this thesis, we explore this problem and give an improved bound for good local-spectral expanders.

To my parents for your trust and sacrifices,
and to all my teachers for illuminating the paths ahead.

Contents

Acknowledgments

First and foremost, I would like to thank my advisor, Alessandro Chiesa. During graduate school, I have meandered quite a bit through different research topics, and Ale has always been supportive of my explorations. He is generous with his time and truly cares about our well-being. I am grateful for his wisdom and patience, and for always keeping his students' best interests in mind.

Next I want to thank the rest of my thesis committee. I am very fortunate to have worked closely with Prasad Raghavendra. Prasad's deep commitment to working on fundamental problems has really inspired me to develop my "research compass" that guides me through the at-first overwhelming research world. His preserverance facing a lack of progress teaches me to enjoy even the difficult phases of research.

I started working with Tselil Schramm after my second year. She has taught me a lot about semi-definite programming and is the one who introduced me to geometry. Her great energy during meetings and enthusiasm in learning new subjects greatly influenced me. She is both a motivational mentor and a close friend.

I was a teaching assistant for Avishay Tal's computational complexity course. He guided me through planning discussion sessions and designing instructive problem sets. He also clarified many of my confusions about teaching. I am also grateful for his research advice, some of which I actually learnt from his students!

I would also like to thank my other mentors. I thank Alex Psomas for taking me as a research apprentice when I knew nothing about research and for making my first research experience super smooth and enjoyable. Ever since, I have been inspired to be a mentor like Alex. I thank Tom Gur for introducing me to HDX and for always being cheerful and supportive. I am also grateful to Sanjam Garg for teaching me a lot about cryptography and for giving helpful advice during graduate school application. I also want to thank Ueli Maurer and Chen-Da Liu-Zhang for mentoring me during a summer research internship at ETH.

My research visits to other institutes during graduate school have been eye-opening and refreshing. I would like to thank my hosts Sam Hopkins and Irit Dinur for giving me the opportunities to visit and collaborate.

I am grateful to my close collaborators Sidhanth Mohanty and Elizabeth Yang. Graduate school would not feel the same without the company of them. We spent so many hours together bouncing off ideas, untangling confusions, and checking intuitions. They make me a stronger researcher and a better collaborator. The many meetings at Menchie's, Soda, and Simons are some of my fondest memories in graduate school.

I also want to thank all my other collaborators whose creativity and dedication always humble me. They are Jonathan Bootle, Yotam Dikstein, Louis Golowich, Ziyi Guan, Seri Khoury, Noam Lifshitz, Xinyu Wu, Jeff Xu, and Rachel Zhang.

Berkeley theory group truly feels like family to me. I want to thank the theory faculties for providing such an open and collaborative environment for the students, and thank all the students for sharing their unique perspectives, extensive knowledge, and boundless energy.

They make my graduate school experience fun and stimulating. I want to thank my cohort Sidhanth Mohanty and Seri Khoury. We have exchanged many research related and unrelated conversations during the years. Thank them for sharing their insights about research and about life. I also want to thank my fellow students for making going to Soda so fun: Ishaq Aden-Ali, Omar Alrabiah, Efe Aras, James Bartusek, Zöe Bell, Thiago Bergamaschi, Lynn Chua, Grace Dinh, Arun Ganesh, Louis Golowich, Sam Gunn, Wenshuo Guo, Meghal Gupta, Chris Ikeokwu, Malvika Joshi, Nathan Ju, Tarun Kathuria, Rachel Lawrance, Yunchao Liu, Xin Lyu, Pasin Manurangsi, Jarrod Millman, Ansh Nagda, Chinmay Nirkhe, Orr Paradise, Angelos Pelecanos, Vamsi Policharla, Bhaskar Roberts, Aaron Schild, Jonathan Shafer, Abhishek Shetty, Nick Spooner, Akshayaram Srinivasan, Francisca Vasconcelos, David Wu, Kewen Wu, Hongxun Wu, Elizabeth Yang, Morris Yau, Fred Zhang, and Richard Zhang.

I also want to thank my friends in the theory community beyond Berkeley: Tim Hsieh, Ziyi Guan, Frederic Koehler, Peter Manohar, Xinyu Wu, and Jeff Xu.

Thanks to my friends Vajra, Evon, Ling, Jingwen, Luke, Catherine, Yarong, and Wan Fung for their support during my ups and downs.

Last but not least, many thanks to my parents for taking care of me and protecting me all these years, and for giving me trust and approval. Also thanks to my aunt and grandparents for accompanying me during hard times.

Chapter 1

Introduction

Many functional analysis results over discrete spaces are defined and studied after their continuous counterparts have been proven. Notions including Laplacian operators, derivatives, Fourier decomposition are first defined over continuous spaces such as multivariate Gaussians. They are later generalized to discrete spaces (e.g. the hypercubes $\mathbb{F}_2^n$, the symmetric groups S_n). Moreover, tools like the invariance principle and the Central Limit theorem have been developed to coupled discrete spaces with continuous spaces, and thereby transform isoperimetric type inequalities over the Gaussians to corresponding discrete spaces and vice versa. Examples include the proofs of Log-Sobolev inequalities over the Gaussians [55] and of the majority is the stablest theorem over the hypercubes [94].

Local-spectral expanders (HDXes) are a class of hypergraphs whose spectral expansion can be certified locally. They have recently attracted a lot of attention due to applications to testing, coding, and sampling. The key feature of an HDX is that the subgraphs induced by the neighbors of any hyperedges are all expander graphs. This property is called local-spectral expansion. A natural question is: are there natural continuous spaces that have local-spectral expansion? This is particularly interesting since most known constructions of sparse HDXes are algebraic, and a more geometric view of local-spectral expansion can potentially give more constructions of HDXes and also more intuitions for isoperimetric inequalities over HDXes.

Furthermore due to the abstraction in the definition of HDXes, it has been challenging to obtain tighter isoperimetric inequalities for these hypergraphs. For instance, we do not know good bounds on small set expansion for random walks over HDXes. Known tight bounds for hypercubes rely on orthogonal decompositions for functions over these domains. The main challenge here is that the space of functions over HDXes does not have an explicit orthogonal decomposition.

This thesis addresses the two problems above. This chapter starts with an introduction on expander graphs. We briefly summarize their special properties, their applications in computer science, and some well-known constructions. From there, we generalize the notion of expansion to hypergraphs and define local-spectral expanders. Lastly we motivate and summarize the main results of the thesis: two constructions of HDXes and improved small set expansion results over good HDXes.

1.1 Expanders

Expander graphs (or expanders) are well-connected graphs. More precisely a family of discrete graphs $\{G_n = (V_n = [n], E_n)\}_n$ are expanders if:

- either as $n \to \infty$ the normalized Laplacian matrix L_{G_n}'s smallest eigenvalue is bounded away from 0 (**spectral expansion**),
- or as $n \to \infty$ for all subset of vertices $S \subseteq V_n$ of size $|S| \leqslant |V_n|/2$, a constant fraction of S's adjacency edges are connected to vertices outside S (**combinatorial expansion**).

The expansion parameters are omitted for simplicity.

These two definitions are obtained by generalizing Laplace operators and Cheeger constants to discrete graphs. It was first shown by Cheeger [16] that these two definitions of expansion are roughly equivalent over compact Riemann manifolds. Later, the equivalence was established for discrete graphs by Dodziuk [32]. Some important properties of expander graphs include that random walks over these graphs mix fast, any small set of vertices has most of their neighbors outside the set, and these graphs have no low-distortion embeddings to low-dimensional Euclidean spaces.

Examples of well-known expanders include the complete graphs and boolean hypercubes. While it is easy to construct expanders with large average degree, expanders with constant average degree (also called sparse expanders) have more applications to algorithm derandomization, gap amplification, linear-time encodable codes, etc [57]. Through decades of intensive research, we now have many randomized and explicit constructions of sparse expanders: random d-regular graphs [50], algebraic expanders [92, 51], the zig-zag product [101], to name a few.

1.2 Local-spectral expanders

One important insight in property testing is that a tester T over a set of variables V can be viewed as a hypergraph $G = (V, E)$, where E is the query set of T. We note that if T can check relations over k variables, then E would contain hyperedges of size k. In the case that T only checks binary relations, G is a graph, and T's soundness can be derived from the expansion property of G. To analyze more general testers, one needs to study expansion properties of hypergraphs.

High-dimensional expanders are expanding hypergraphs. Though both spectral expansion and combinatorial expansion can be generalized to hypergraphs, they are not equivalent notions on hypergraphs. In this thesis we focus on local-spectral expanders which are spectrally expanding hypergraphs. Discussions on combinatorially expanding hypergraphs and the comparison between the two definitions are postponed till Chapter 5. In the other chapters HDXes are used interchangeably with local-spectral expanders.

Local-spectral expanders are a generalization of spectral expansion to simplicial complexes [87]. A simplicial complex is a special type of hypergraph whose hyperedges are downward closed. A 1-dimensional simplicial complex is a graph. A general d-dimensional simplicial complex χ consists of hyperedge sets $\chi(0), \dots, \chi(d)$ where $\chi(i)$ contains hyperedges of cardinality $(i+1)$. We define the link of a vertex v to be the $(d-1)$-dimensional hypergraph χ_v with hyperedges $\left\{ f \in \bigcup_{i=0}^{d-1} \chi(i) \mid f \cup \{v\} \in \chi(i+1) \right\}$. So in a 2-dimensional simplicial complex, a link χ_v is a graph over v's neighbors in χ.

A 1-dimensional local-spectral expander is an expander. A d-dimensional local-spectral expander is a simplicial complex that satisfies (1) the global graph $G_\emptyset = (\chi(0), \chi(1))$ is an expander, and (2) for every vertex $v \in \chi(0)$, the link χ_v is a $(d-1)$-dimensional local-spectral expanders. So χ is a 2-dimensional local-spectral expander if it is a 1-dimensional expander and all its vertices' links are also 1-dimensional expanders.

1.2.1 The local-to-global phenomenon

The definition above suggests that to show a d-dimensional χ is an HDX, one should check that the link graphs of *all* hyperedges of size $< (d-1)$ are expanders. However, the trickling-down theorem from [97] states that it suffices to check the links of hyperedges in $\chi(d-2)$ are expanders and the link graphs of all smaller hyperedges are connected.

For a 2-dimensional simplicial complex χ, the trickling-down theorem says that if every vertex's link is a λ-expander (λ is 1 minus the smallest eigenvalue of the graph Laplacian) and if the global graph $G_\emptyset = (\chi(0), \chi(1))$ is connected, then $G_\emptyset$ is a $\left(\frac{\lambda}{1-\lambda}\right)$-expander. Thus χ is a 2-dimensional $\left(\frac{\lambda}{1-\lambda}\right)$ local-spectral expander.

1.2.2 Applications

The local-to-global phenomenon on local-spectral expanders has led to many applications in computer science. Starting with [24], local-spectral expanders have been used to construct agreement tests. As the setup of an agreement test, consider a space Ω, a set of subspaces S, and a code $\mathbb{C}$ that encodes functions $f : \Omega \to \Sigma$ as a collection of local functions $\{f_s : s \to \Sigma\}_{s \in S}$. In an agreement test for $\mathbb{C}$, a randomized tester T is given oracle access to a set of local functions $\{g_s\}_{s \in S}$ and needs to output whether these local functions are close to $\mathbb{C}$ (i.e. close to subspace restrictions of some global function g). T is required to always accept if the local functions are in $\mathbb{C}$ (completeness), and to reject with probability proportional to the distance from the local functions to $\mathbb{C}$ (soundness).

[24] shows that a d-dimensional local-spectral expander χ gives rise to the following agreement tester: let $\Omega = \chi(0)$ and $S = \chi(k)$, then the tester T samples a random $\tau \in \chi(2k+1)$ and two random $s, s' \subset \tau$, and outputs "accept" if and only if f_s and $f_{s'}$ agree

on $s \cap s'$. The local-spectral expansion guarantees that T has soundness. This application illustrates the connection between expansion and testing.

As an example, consider the [24] agreement tester given by a 3-dimensional complete complex X. By definition the hyperedge sets of the complex are given by $X(i) = \binom{[n]}{i+1}$. Let the space $\Omega = [n]$, the alphabet $\Sigma = \mathbb{F}_2$ and $S = X(1)$. So a codeword of $\mathbb{C}$ is a collection of functions $\{f|_s : s \to \mathbb{F}_2 \mid s \in X(1)\}$ that are restrictions of a global function f to hyperedges. The tester T samples a random $\tau \sim X(3)$ and two random $s, s' \subseteq \tau$. Then T outputs whether the input local functions $f_s|_{s\cap s'} = f_{s'}|_{s\cap s'}$.

More recently chain complexes (more general hypergraphs than simplicial complexes) with local-spectral expansion are used to construct locally testable codes with constant rate, distance, and arity [28, 79], and quantum low-density parity-check codes [99, 75, 27].

Markov chain Monte Carlo method (MCMC) is a class of algorithms that sample from exponential size distributions. Whether MCMC can efficiently sample from a distribution depends on the convergence time of the underlying Markov chain. Recently [5] resolved a longstanding open question by showing rapid mixing of matroid basis exchange walks. The result is proved by showing that the matroid basis exchange complexes are local-spectral expanders. Followup works prove more rapid mixing results under this framework [2, 4, 39].

1.2.3 Constructions

In comparison to expanders, we have limited understanding of HDXes. In contrast to the numerous explicit and probabilistic constructions of sparse expanders, we know only a few explicit constructions of bounded-degree HDXes and they are either heavily algebraic [90, 66, 95] or have bounded dimensions [15]. For a while we do not know any combinatorial constructions of HDXes of all constant dimensions. Furthermore, while properties of the Laplace operators of many expanders are well-studied and have been applied to solve problems in combinatorics, computational complexity, and statistical physics, we have yet to understand their counterparts in HDXes.

1.3 Overview

Given many applications of HDXes, we would like to have more intuitive constructions of local-spectral expanders and also find reasons why certain standard approaches fail. Moreover, since certain applications require a more precise characterization of combinatorial expansion of random walks over local-spectral expanders, we need to develop new tools to improve existing analyses. This thesis provides partial answers to these questions.

In Chapter 2, we give an approach of constructing HDXes from expanders. We start with a global graph that is already a sparse expander. Currently the link of a vertex is simply a set of disjoint vertices. Next we start adding edges and triangles to the graph via graph products to make the link graphs of all vertices connected. This approach produces bounded-degree local-spectral expanders whose link graphs are $\left(\frac{1}{2} + \varepsilon\right)$-expanders. This expansion parameter is in the regime where the trickling-down theorem gives a trivial bound of 1. Since many applications hinge on local expansion parameter to be arbitrarily close to 0, this construction is not widely applicable.

In Chapter 3, we take a different approach to the construction problem. We ask if any compact manifolds naturally gives rise to local-spectral expansion. If so, we would be able to find geometric intuition behind analytical properties of HDXes, and also construct HDXes from these manifolds. As a first step towards this direction, we consider the high-dimensional spheres, and study the random 2-dimensional simplicial complexes over the spheres. We obtain polynomial degree HDXes from this model. Though the average degree is not bounded, it beats the average degree of HDXes constructed from random simplicial complex models that do not have any latent geometry.

Lastly in Chapter 4, we characterize non-expanding sets over random walks in HDXes. The main tool is a basis decomposition framework that gives an almost orthogonal decomposition of functions over HDXes. This framework applies generally to spaces that are "locally" close to product spaces. HDXes are examples of such spaces.

1.3.1 Chapter 2: High-dimensional expanders from expanders

We present an elementary way to transform an expander graph into a local-spectral expander where all high order random walks have a constant spectral gap, i.e., they converge rapidly to the stationary distribution. As an upshot, we obtain new constructions, as well as a natural probabilistic model to sample constant degree local-spectral expanders.

In particular, we show that given an expander graph G, adding self loops to G and taking the tensor product of the modified graph with a constant-size local-spectral expander produces a larger local-spectral expander. The resulting local-spectral expanders have local-spectral expansion parameter strictly greater than $\frac{1}{2}$. Though in this regime local-to-global phenomenon does not hold, this is the first combinatorial construction of constant-degree local-spectral expanders of any constant dimension with local expansion independent of the number of vertices and dimension.

We also analyze the various high order random walks over these complexes. Our proof of rapid mixing of high order random walks is based on the decomposable Markov chains framework introduced by Jerrum et al.

This chapter is based on joint work with Mohanty and Yang [83].

1.3.2 Chapter 3: 2-dimensional expanders from random geometric graphs

To achieve local expansion strictly smaller than $\frac{1}{2}$, we construct local-spectral expanders from random geometric complexes.

Consider a 2-dimensional random geometric simplicial complex X sampled as follows: first, sample n vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ uniformly at random on $\mathbb{S}^{d-1}$; then, for each triple $i, j, k \in$

$[n]$, add $\{i, j, k\}$ and all of its subsets to X if and only if $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle \geqslant \tau, \langle \boldsymbol{u}_i, \boldsymbol{u}_k \rangle \geqslant \tau$, and $\langle \boldsymbol{u}_j, \boldsymbol{u}_k \rangle \geqslant \tau$. We prove that for every $\varepsilon > 0$, there exists a choice of $d = \Theta(\log n)$ and $\tau = \tau(\varepsilon, d)$ so that with high probability, X is a local-spectral expander of average degree n^ε in which each vertex's link graph has second eigenvalue smaller than $\frac{1}{2}$.

To our knowledge, this is the first demonstration of a natural distribution over 2-dimensional local-spectral expanders of arbitrarily small polynomial average degree and spectral link expansion better than 1/2. All previously known constructions are algebraic. This distribution also furnishes an example of simplicial complexes for which the trickling-down theorem is nearly tight.

En route, we prove general bounds on the spectral expansion of random induced subgraphs of arbitrary vertex transitive graphs, which may be of independent interest. For example, one consequence is an almost-sharp bound on the second eigenvalue of random n-vertex geometric graphs on $\mathbb{S}^{d-1}$, which was previously unknown for most n, d pairs.

This chapter is based on joint work with Mohanty, Schramm, and Yang [84].

1.3.3 Chapter 4: Global hypercontractivity inequality over ε-product space

A key property of local-spectral expanders is rapid mixing of the up-down walks in every dimension. The k up-down walk in a d-dimensional simplicial complex χ is the two-step walk over the bipartite graph $G_k = (\chi(k), \chi(k+1), E_k)$ induced by the containment relation between $\chi(k)$ and $\chi(k+1)$.

Indeed we know that the k up-down walk has spectral gap $O\left(\frac{1}{k}\right)$ [68]. Therefore by Cheeger's inequality any small set of k-faces has edge expansion $\Omega\left(\frac{1}{k}\right)$ in the walk graph. Though k is a constant in our context, many applications need an expansion that does not depend on k.

However, such expansion cannot hold for all small sets as shown by the following example. Let $v \in \chi(0)$ be a vertex in χ, define the set $S_v(k)$ to contain all k-faces that contains v. $S_v(k)$'s edge expansion is $O\left(\frac{1}{k+1}\right)$. Though a k-independent edge expansion does not hold for all small sets, we note that the counterexamples are highly structured, and one could still hope for a better expansion for all "unstructured sets". This is the result in Chapter 4.

We prove hypercontractivity inequalities for local-spectral expanders. Our inequalities are effective for global functions, which are functions that are not significantly affected by a restriction of a small set of coordinates. As an application, we obtain small-set expansion for local-spectral expanders. It implies that in the k up-down walk of an HDX, the non-expanding sets are precisely those that have large overlaps with sets of the form $\{s \in \chi(k) \mid e \subset f\}$ where e is some lower-dimensional hyperedge. Our approach applies more generally to ε-product spaces. They are multivariate probability spaces satisfying that fixing any constant number of variables make the resulting conditional distribution almost pairwise independent. The distribution given by a local-spectral expander is a natural example of such spaces. The key technique is a new approximate Efron-Stein decomposition for ε-product spaces.

This chapter is based on joint work with Gur and Lifshitz [56].

Chapter 2

High-dimensional Expanders from Expanders

We construct bounded-degree high–dimensional expanders of all constant–sized dimensions, where the high order random walks have a constant spectral gap, and thus mix rapidly. We base our HDX's from existing T-regular one-dimensional constructions, which can be sampled readily from the space of all T-regular graphs. This endows a natural distribution from which we can sample HDX's of our construction as well. After the first version of this paper was written, it was brought to our notice by a reviewer that the construction in this paper has been previously discussed in the community. Nevertheless, a contribution of our work is a rigorous analysis of the expansion properties of this construction.

One sufficient, but not necessary criterion that implies rapid mixing is *spectral*, which comes from the graph theoretic notion below.

Definition 2.0.1 (Informal)**.** A d–dimensional λ–spectral expander is a d–dimensional simplicial complex (i.e. a hypergraph whose faces satisfy downward closure) such that

- (Global Expansion) The vertices and edges (sets of two vertices) of the complex constitute a λ–spectral expander graph,
- (Local Expansion) For every hyperedge E of size $\leqslant d-1$ in the hypergraph, the vertices and edges in the "neighborhood" of E also constitute a λ–spectral expander. (The precise definition of "neighborhood" will be discussed later.)

Most known constructions of bounded-degree high–dimensional spectral expanders are heavily algebraic, rather than combinatorial or randomized. In contrast, there are a wealth of different constructions for bounded-degree (one-dimensional) expander graphs [58]. Some of these are also algebraic, such as the famous LPS construction of Ramanujan graphs [88], but there are also many simple, probabilistic constructions of expanders. In particular, Friedman's Theorem says that with high probability, random d-regular graphs are excellent expanders [49].

Unfortunately, random d–dimensional hypergraphs with low degrees are not d–dimensional expander graphs. For a hypergraph with n vertices, we need a roughly $n\,(\log n/n)^{1/d}$-degree Erdős-Rényi graph to make the neighborhood of every hyperedge of size $\leqslant d-1$ to be connected with high probability. While random low degree hypergraphs are not high–dimensional expanders, our construction provides simple probabilistic high–dimensional expanders of all dimensions.

2.1 Problem background and summary of results

2.1.1 Our results

Construction. We construct an H–dimensional simplicial complex $\mathcal{Q}$ on $n \cdot s$ vertices, from a graph G of n vertices and a (small) H–dimensional complete simplicial complex $\mathcal{B}$ on s vertices. To construct $\mathcal{Q}$, we replace each vertex v of G with a copy of $\mathcal{B}$ which we denote $\mathcal{B}_v$. Denote the copy of a vertex $w \in \mathcal{B}$ in $\mathcal{B}_v$ by (v, w). The faces of $\mathcal{Q}$ are chosen in the following way: for every face $\{w_1, w_2, \ldots, w_k\}$ in $\mathcal{B}$, add it $\{(v_1, w_1), (v_2, w_2), \ldots, (v_k, w_k)\}$ to the complex, where for some edge e in G, the vertices $v_1, \ldots v_k$ are each one of the endpoints of e; in particular there are 2 choices for each v_i. The main punchline of our work is that when G is a (triangle-free) expander graph, the high order random walks on $\mathcal{Q}$ mix rapidly. Specifically, we prove:

Theorem 2.1.1 (Main theorem, informal version of Theorem 2.3.1)**.** *Suppose G is a triangle-free expander graph with two-sided spectral gap ρ. For every k such that $1 \leqslant k < H$, there is a constant C depending on k, H, s, ρ, but independent of n such that the Markov transition matrix for the up-down walk on the k-faces of $\mathcal{Q}$ has two-sided spectral gap C.*

First attempt at proving rapid mixing of high order random walks. [65], which introduced the notions of up-down and down-up random walks, and subsequent works [24, 68, 67, 5] developed and followed the "local-to-global paradigm" to prove rapid mixing of high order random walks. In particular, each of these works would:

A. Establish that all the links of a relevant simplicial complex have "small" second eigenvalue.

B. Prove or cite a statement about how rapid mixing follows from small second eigenvalues of links (such as Theorem 2.1.3).

Then, step A. and step B. together would imply that the up-down and down-up random walks on the simplicial complexes they cared about mixed rapidly. This immediately motivates first bounding the second eigenvalue of the links of our construction, and applying the quantitatively strongest known version of the type of theorem alluded to in step B.. Thus, in Section 2.4 we analyze the second eigenvalue of all links of $\mathcal{Q}$ and prove:

Theorem 2.1.2 (Informal version of Theorem 2.3.3)**.** *The two-sided spectral gap of every link in $\mathcal{Q}$ is bounded by approximately $\frac{1}{2}$.*

And the 'quantitatively strongest' known "local-to-global" theorem is

Theorem 2.1.3 (Informal statement of [68, Theorem 5])**.** *If the second eigenvalue of every link of a simplicial complex $\mathcal{S}$ is bounded by λ, then the up-down walk on k-faces of $\mathcal{S}$, $\mathcal{S}_k^{\uparrow\downarrow}$ satisfies:*

$$\lambda_2(\mathcal{S}_k^{\uparrow\downarrow}) \leqslant \left(1 - \frac{1}{k+1}\right) + k\lambda.$$

Observe that the upper bound on the second eigenvalue of all links must be strictly less than $\frac{1}{k(k+1)}$ to conclude any meaningful bounds on the mixing time of the up-down random walk. Thus, unfortunately, Theorem 2.1.2 in conjunction with Theorem 2.1.3 fails to establish rapid mixing.

Hence, we depart from the local-to-global paradigm and draw on alternate techniques.

Decomposing Markov chains. Each k-face of $\mathcal{Q}$ is either completely contained in a *cluster* $\{(v, ?)\}$ for a single vertex v in G, or straddles two clusters corresponding to vertices connected by an edge, i.e., is contained in $\{(v, ?)\} \cup \{(u, ?)\}$. Consider performing an up-down random walk on the space of k-faces of $\mathcal{Q}$ (henceforth $\mathcal{Q}_k^{\uparrow\downarrow}$). If we record the single cluster or pair of clusters containing the k-face the random walk visits at each timestep, it would resemble:

$$\{17,19\} \to \{17,19\} \to \{17,19\} \to \{17\} \to \{17\} \to \{17,155\} \to \{17,155\} \to \{17,155\} \to \{155\} \to \{155,203\} \to \{155,203\} \to \{155,203\} \to \{155,203\} \to \{155,203\} \to \{203\} \to \{6,203\} \to \{6,203\} \to \cdots$$

In the above illustration of a random walk, let us restrict our attention to the segment of the walk where the k-faces are all contained in, say, the pair of clusters $\{155, 203\}$. Intuitively, we expect the random walk restricted to those k-faces to mix rapidly and also exit the set quickly by virtue of the state space being constant-sized. In particular, if we keep the random walk running for $t \approx C \log n$ steps for some large constant C, it would seem that the number of "exit events"[1] is roughly $\alpha \cdot C \log n$ for some other constant α. The sequence of "exit events" can be viewed as a random walk on the space of edges and vertices of G, and since there are *many* steps in this walk, the expansion properties of G tell us that the location of the random walk after t steps is distributed according to a relevant stationary distribution. In light of these intuitive observations of rapidly mixing in the walks within cluster pairs and also rapidly mixing in a walk on the space of cluster pairs, one would hope that the up-down walk on k-faces mixes rapidly.

This hope is indeed fulfilled and is made concrete in a framework of Jerrum et al. [62]. In their framework, there is a Markov chain M on state space Ω. They show that if Ω can

[1] Transitions like $\{17, 19\} \to \{17\}$, $\{155\} \to \{155, 203\}$, and so on.

be partitioned into $\Omega_1, \ldots, \Omega_\ell$ such that the chain "restricted" (for some formal notion of restricted) to each Ω_i, and an appropriately defined "macro-chain" (where each partition Ω_i is a state) each have a constant spectral gap, then the original Markov chain M has a constant spectral gap as well. Our proof of the fact that $\mathcal{Q}_k^{\uparrow\downarrow}$ has a constant spectral gap utilizes this result of [62]. This framework of decomposable Markov chains is detailed in Section 2.2.2, and the analysis of the spectral gap of the down-up random walk[2] is in Section 2.5.

2.1.2 Related Work

While high–dimensional expanders have been of relatively recent interest, already many different (non-equivalent) notions of high–dimensional expansion have emerged, for a variety of different applications.

The earliest notions of high–dimensional expansion were topological. In this vein of work, [80, 54] introduced coboundary expansion, [37] defined cosystolic expansion, and [37, 64] defined skeleton expansion. To our knowledge, most existing constructions of these types of expanders rely on the Ramanujan complex. We refer the reader to a survey by Lubotzky [87] for more details on these alternate notions of high dimensional expansion and their uses.

To describe notions of high dimensional expansion that are relevant to computer scientists, we need to first highlight a key property of (one-dimensional) expander graphs–that random walks on them mix rapidly to their stationary distribution. The notion of a random walk on graphs was generalized to simplicial complexes in the work of Kaufman and Mass [65] to the "up-down" and "down-up" random walks, whose states are k-faces of a simplicial complex. They were interested in bounded–degree simplicial complexes where the up-down random walk mixed to its stationary distribution rapidly. They then proceed to show that the known construction of Ramanujan complexes from [90] indeed satisfy this property.

A key technical insight in their work that the rapid mixing of up-down random walks follows from certain notions of *local spectral expansion*, i.e., from sufficiently good *two-sided* spectral expansion of the underlying graph of every link. A quantitative improvement between the relationship between the two-sided spectral expansion of links and rapid mixing of random walks was made in [24], and this improvement was used to construct agreement expanders based on the Ramanujan complex construction. Later, [68] showed that one-sided spectral expansion of links actually sufficed to derive rapid mixing of the up-down walk on k-faces.

2.1.3 HDX Constructions

Although this combinatorial characterization of high–dimensional expansion is slightly weaker than some of the topological characterizations mentioned above, few constructions are known

[2] Which is actually equivalent to proving a spectral gap on the up-down random walk but is more technically convenient. See Fact 2.2.29.

for *bounded degree* HDX's with dimension $\geqslant 2$. Most of these rely on heavy algebra. In contrast, for one-dimensional expander graphs, there are a wealth of different constructions, including ones via graph products and randomized ones. [49] states that even a random d-regular graph is an expander with high probability.

The most well-known construction of bounded-degree high–dimensional expanders are the Ramanujan complexes [90]. These require the *Bruhat-Tits building*, which is a high-dimensional generalization of an infinite regular tree. The underlying graph has degree $q^{O(d^2)}$, where q is a prime power satisfying $q \equiv 1 \pmod 4$. The links can be described by *spherical buildings*, which are complexes derived from subspaces of a vector space, and are excellent expanders.

Dinur and Kaufman showed that given any $\lambda \in (0, 1)$, and any dimension d, the d–skeleton of any $d + \lceil 2/\lambda \rceil$–dimensional Ramanujan complex is a d–dimensional λ–spectral expander [24]. Here, the degree of each vertex is $(2/\lambda)^{O((d+2/\lambda)^2)}$. In other words, they "truncate" the Ramanujan complexes, throwing out all faces of size greater than some number k. Their primary motivation was to obtain agreement expanders, which find uses towards PCPs.

Recently, Kaufman and Oppenheim [67] present a construction of one-sided HDXes, which are coset complexes of elementary matrix groups. The construction guarantees that for any $\lambda \in (0, 1)$ and any dimension d, there exists a infinite family of high–dimensional expanders $\{X_i\}_{i\in\mathbb{N}}$, such that (1) every X_i are d–dimensional λ–one–sided–expander; (2) every X_i's 1-skeleton has degree at most $\Theta\left(\sqrt{\frac{(1/\lambda+d-1)^{(d+2)^2}}{2\log(1/\lambda+d-1)}}\right)$; (3) as i goes to infinity the number of vertices in X_i also goes to infinity.

Even more recently, Chapman, Linial, and Peled [15] also provided a combinatorial construction of two-dimensional expanders. They construct an infinite family of (a, b)-regular graphs, which are a-regular graphs whose links with respect to single vertices are b-regular. The primary motivation for their construction comes from the theory of PCPs. They prove an Alon-Boppana type bound on $\lambda_2(G)$ for any (a, b)-regular graph, and construct a family of graphs where this bound is tight. They also build an (a, b)-regular two-dimensional expander using any non-bipartite graph G of sufficiently high girth; they achieve a local expansion only depending on the girth, and the global expansion depending on the spectral gap of G. Like ours, their construction also resembles existing graph product constructions of one-dimensional expanders.

2.2 Preliminaries and Notation

2.2.1 Spectral Graph Theory

While we can describe our constructions combinatorially, our analysis of both the mixing times of certain walks as well as the local expansion will heavily rely on understanding graph spectra.

Definition 2.2.1. For an edge-weighted directed graph G on n vertices, we use $\mathrm{Adj}(G)$ to denote its *(normalized) adjacency matrix*, i.e. the matrix given by

$$\mathrm{Adj}(G)_{(u,v)} = \frac{\mathbf{1}_{(u,v)\in E(G)} \cdot w((u,v))}{\sum_{v:(u,v)\in E(G)} w((u,v))}$$

and write its (right) eigenvalues as

$$1 = \lambda_1(G) \geqslant \lambda_2(G) \geqslant \ldots \geqslant \lambda_n(G) \geqslant -1$$

Let $\mathsf{Spectrum}(G)$ to indicate the set $\{\lambda_i(G)\}$. We write $\mathsf{OneSidedGap}(G)$ for the *spectral gap* of G, which is the quantity $1 - \lambda_2(G)$. Graphs with $\mathsf{OneSidedGap}(G) \geqslant \mu$ are *one-sided μ-expanders.*

Most of the graphs we analyze achieve a stronger condition; that the second largest eigenvalue magnitude is not too large. Formally, we write $|\lambda|_i$ for the i-th largest eigenvalue in absolute value. In particular, $|\lambda|_2 = \max\{|\lambda_2|, |\lambda_n|\}$. The *absolute spectral gap* of G, denoted $\mathsf{TwoSidedGap}(G)$, is the quantity $1 - |\lambda|_2$. Graphs with $\mathsf{TwoSidedGap}(G) \geqslant \mu$ are *two-sided μ-expanders.*

Remark 2.2.2. For an *undirected* weighted graph, we simply have $w((u,v)) = w((v,u))$, and use this to define the adjacency matrix the same way.

Graph Tensors Our construction can roughly be described as a tensor product, defined below.

Definition 2.2.3. The tensor product $G \times H$ of two graphs G and H is given by

1. Vertex set $V(G \times H) = V(G) \times V(H)$,
2. Edge set $E(G \times H) = \{((u_1, v_1), (u_2, v_2)) : (u_1, u_2) \in E(G) \text{ and } (v_1, v_2) \in E(H)\}$.

The adjacency matrix $\mathrm{Adj}(G \times H)$ is the tensor (Kronecker) product $\mathrm{Adj}(G) \otimes \mathrm{Adj}(H)$. Due to this structure, $\mathsf{Spectrum}(G \times H) = \{\lambda_i \mu_j : \lambda_i \in \mathsf{Spectrum} G,\ \mu_j \in \mathsf{Spectrum}(H)\}$. As 1 is the largest eigenvalue of both $\mathrm{Adj}(G)$ and $\mathrm{Adj}(H)$, it follows that both

$$\begin{aligned}
\mathsf{OneSidedGap}(G \times H) &= \min(1 - 1 \cdot \mu_2, 1 - \lambda_2 \cdot 1) \\
&= \min(\mathsf{OneSidedGap}(G), \mathsf{OneSidedGap}(H)) \\
\mathsf{TwoSidedGap}(G \times H) &= \min(1 - 1 \cdot |\mu|_2, 1 - |\lambda|_2 \cdot 1) \\
&= \min(\mathsf{TwoSidedGap}(G), \mathsf{TwoSidedGap}(H))
\end{aligned}$$

2.2.2 Markov Chains

We provide a basic overview of the Markov chain concepts used to analyze our high order walks. We refer to [76] for a detailed and thorough treatment of the fundamentals of Markov chains.

Definition 2.2.4. A *Markov chain* $M = (\Omega, P)$ is given by *states* Ω and a *transition matrix* P where $P[i, j]$ is the probability of going to state j from state i. We may also write this quantity as $M[j \to i]$.

Remark 2.2.5. The literature often defines $P_{i,j}$ as the probability $\mathbf{Pr}(i \to j)$, so their P is the transpose of ours. However, we work with column (right) eigenvectors to analyze the spectrum of P, while this alternate convention uses row (left) eigenvectors, so both conventions yield the same results.

Definition 2.2.6. We can view any Markov chain M as a weighted, directed graph G, defined by $V(G) = \mathsf{States}(M)$, $E(G) \coloneqq \{(i, j) : i, j \in V(G), M[i \to j] > 0\}$, and $w((i, j)) = M[j \to i]$.

The transition matrix of M is $\mathrm{Adj}(G)$, and we also refer to $\mathsf{Spectrum}(G)$ as the *spectrum* of M. Every adjacency matrix has $\lambda_1 = 1$, so transition matrix of M has an eigenvector π_M (normalized so that entries sum to 1) for the eigenvalue 1. We call π_M a *stationary distribution* of M.

Remark 2.2.7. We may use the term "graph" in lieu of "chain" when we want to indicate the random walk defined by the transition matrix $\mathrm{Adj}(G)$.

The next property we introduce is present for every Markov chain we analyze.

Definition 2.2.8. The Markov chain $M = (\Omega, P)$ is *time-reversible* if for any integer $k \geqslant 1$:

$$\pi_M(x_0)M[x_0 \to x_1] \cdots M[x_{k-1} \to x_k] = \pi_M(x_k)M[x_k \to x_{k-1}] \cdots M[x_1 \to x_0]$$

Intuitively, it means that if start at the stationary distribution and run the chain for a sequence of time states, the reverse sequence has the same probability of occurring. Time reversibility helps us compute stationary distributions via the *detailed balance equations*. (This is especially helpful when there are a huge number of symmetric states.)

Fact 2.2.9. *The Markov chain $M = (\Omega, P)$ is time-reversible if and only if it satisfies the detailed balance equations: for all $x, y \in \Omega$,*

$$\pi_M(x)M[x \to y] = \pi_M(y)M[y \to x]$$

Definition 2.2.10. The ε*-mixing time* of a Markov chain M is the smallest t such that for any distribution ν over the states of M,

$$\|\pi_M - P^t\nu\|_1 \leqslant \varepsilon$$

where π_M is the stationary distribution of M.

Theorem 2.2.11. *For any Markov chain M, the ε-mixing time $t_{\mathrm{mix}}(\varepsilon)$ satisfies:*

$$t_{\mathrm{mix}}(\varepsilon) \leqslant \log\left(\frac{1}{\varepsilon \min \pi_M}\right) \cdot \frac{1}{\mathsf{TwoSidedGap}(M)}.$$

Decomposing Markov Chains Consider a finite-state time reversible Markov chain M whose structure gives rise to natural state-space partition, M can be decomposed into a number of restriction chains and a projection chain. [62] show that the spectral gap for the original chain can be lower bounded in terms of the spectral gaps for the restriction and projection chains.

We now formally define the decomposition of a Markov chain. Consider an ergodic Markov chain on finite state space Ω with transition probability $P\colon \Omega^2 \to [0,1]$. Let $\pi\colon \Omega \to [0,1]$ denote its stationary distribution, and let $\{\Omega_i\}_{i\in[m]}$ be a partition of the state space into m disjoint sets, where $[m] \coloneqq \{1,\ldots,m\}$.

The *projection chain* induced by the partition $\{\Omega_i\}$ has state space $[m]$ and transitions

$$\overline{P}(i,j) = \left(\sum_{x\in\Omega_i} \pi(x)\right)^{-1} \sum_{x\in\Omega_i y\in\Omega_j} \pi(x)P(x,y).$$

The above expression corresponds to the probability of moving from any state in Ω_i to any state in Ω_j in the original Markov chain.

For each $i \in [m]$, the *restriction chain* induced by Ω_i has state space Ω_i and transitions

$$P_i(x,y) = \begin{cases} P(x,y), & x \neq y, \\ 1 - \sum_{z\in\Omega_i\setminus\{x\}} P(x,z), & x = y. \end{cases}$$

$P_i(x,y)$ is the probability of moving from state $x \in \Omega_i$ to state y when leaving Ω_i is not allowed.

Regardless of how we define the projection and restriction chains for a time reversible Markov chain, they all inherit one useful property from the original chain.

Fact 2.2.12. *Let $M = (\Omega, P)$ be a time-reversible Markov chain. Then, for any decomposition of M, the projection and restriction chains are also time-reversible.*

We ultimately want to study the spectral gap of random walks. Luckily, the original Markov chain's spectral gap is related to the restriction and projection chains' spectral gaps in the following way:

Theorem 2.2.13 ([62, Theorem 1])**.** *Consider a finite-state time-reversible Markov chain decomposed into a projection chain and m restriction chains as above. Define γ to be maximum probability in the Markov chain that some state leaves its partition block,*

$$\gamma \coloneqq \max_{i\in[m]} \max_{x\in\Omega_i} \sum_{y\in\Omega\setminus\Omega_i} P(x,y).$$

Suppose the projection chain satisfies a Poincaré inequality with constant $\bar{\lambda}$ *, and the restriction chains satisfy inequalities with uniform constant* $\lambda_{\min}$*. Then the original Markov chain satisfies a Poincaré inequality with constant*

$$\lambda := \min\left\{\frac{\bar{\lambda}}{3}, \frac{\bar{\lambda}\lambda_{\min}}{3\gamma + \bar{\lambda}}\right\}.$$

Recall that if λ satisfies a Poincaré inequality, it is a lower bound on the spectral gap (cf. [76]).

2.2.3 High-Dimensional Expanders

The generalization from expander graphs to hypergraphs (more specifically, simplicial complexes) requires great care. We now formally establish the high dimensional notions of "neighborhood", "expansion," and "random walk."

Definition 2.2.14. A *simplicial complex* $\mathcal{S}$ is specified by vertex set $V(\mathcal{S})$ and a collection $\mathcal{F}(\mathcal{S})$ of subsets of $V(\mathcal{S})$, known as *faces*, that satisfy the "downward closure" property: if $A \in \mathcal{F}(\mathcal{S})$ and $B \subseteq A$, then $B \in \mathcal{F}(\mathcal{S})$. Any face $S \in \mathcal{F}(\mathcal{S})$ of cardinality $(k+1)$ is called a *k-face* of $\mathcal{S}$. We use $k\text{-}\mathsf{faces}(\mathcal{S})$ to denote the subcollection of k-faces of $\mathcal{F}(\mathcal{S})$. We say $\mathcal{S}$ has *dimension* d, where $d = \max\{|F| : F \in \mathcal{F}(\mathcal{S})\} - 1$.

Example 2.2.15. A 1-dimensional complex $\mathcal{S}$ is a graph with vertex set $V(\mathcal{S})$ and edge set $1\text{-}\mathsf{faces}(\mathcal{S})$.

Definition 2.2.16. To formally define random walks and Markov chains on a $\mathcal{S}$, we need to associate $\mathcal{S}$ with a weight function $w : \mathcal{F}(\mathcal{S}) \to \mathbb{R}_+$. We want our weight function to be *balanced*, meaning for $F \in k\text{-}\mathsf{faces}(\mathcal{S})$:

$$w(F) = \sum_{J \in (k+1)\text{-}\mathsf{faces}(\mathcal{S}) : J \supset F} w(J)$$

If we restrict ourselves to balanced w, it suffices to only define w over $d\text{-}\mathsf{faces}(\mathcal{S})$ and propagate the weights downward to the lower order faces.

Definition 2.2.17. The *(weighted) k-skeleton* of $\mathcal{S}$ is the complex with vertex set $V(\mathcal{S})$ and all faces in $\mathcal{F}(\mathcal{S})$ of cardinality at most $k+1$, with weights inherited from $\mathcal{S}$.

Example 2.2.18. The 1-skeleton of $\mathcal{S}$ only contains its vertices (0-faces) and edges (1-faces). It can be characterized as a graph with edge weights, so we can also compute $\mathsf{OneSidedGap}(1\text{-}\mathsf{skeleton}(\mathcal{S}))$ and $\mathsf{TwoSidedGap}(1\text{-}\mathsf{skeleton}(\mathcal{S}))$.

Definition 2.2.19. For $S \in k\text{-}\mathsf{faces}(\mathcal{S})$ for $k \leqslant H-1$, we associate a particular $(H-k)$-dimensional complex known as the *link* of S defined below.

$$\mathsf{link}(S) := \{T \setminus S : T \in \mathcal{F}(\mathcal{S}), S \subseteq T\}$$

If $\mathcal{S}$ was equipped with weight function w, then $\mathsf{link}(S)$ "inherits" it. We associate $\mathsf{link}(S)$ with weight function w_S given by $w_S(T) = w(S \cup T)$. If w is balanced, then w_S is also balanced. We call a $\mathsf{link}(S)$ a t-link if $|S|$ has cardinality t.

Example 2.2.20. In a graph, the link of a vertex is simply its neighborhood.

Definition 2.2.21. The *global expansion* of $\mathcal{S}$, denoted $\mathsf{GlobalExp}(\mathcal{S})$, is the expansion of its weighted 1-skeleton.

Definition 2.2.22. The *local expansion* of $\mathcal{S}$, denoted $\mathsf{LocalExp}(\mathcal{S})$ is

$$\mathsf{LocalExp}(\mathcal{S}) := \min_{0 \leqslant k \leqslant H-1} \min_{S \in k\text{-}\mathsf{faces}(\mathcal{S})} \mathsf{TwoSidedGap}(1\text{-}\mathsf{skeleton}(\mathsf{link}(S))).$$

In words, it is equal to the expansion of the worst expanding link.

Example 2.2.23. We use $\mathcal{K}_{H+2}^{(H)}$ to denote the *complete H-dimensional complex* on vertex set $[H+2]$, i.e., the pure H-dimensional simplicial complex obtained by making the set of $(H+1)$-faces equal to all subsets of $[H+2]$ of size $H+1$. The 1-skeleton is then a clique on $H+2$ vertices whose expansion is $1 - \frac{1}{H+1}$ and the 1-skeleton of a t-link is a clique on $H+2-t$ vertices, which has expansion $1 - \frac{1}{H+1-(t-1)}$. As a result, $\mathsf{TwoSidedGap}\left(\mathcal{K}_{H+2}^{(H)}\right) = \frac{1}{2}$.

Remark 2.2.24. We often use $\mathrm{Adj}(\mathcal{S})$ to refer to the adjacency matrix of the 1-skeleton of $\mathcal{S}$, and we may also use $\lambda_i(\mathcal{S})$ to refer to the i-th largest eigenvalue of $\mathrm{Adj}(\mathcal{S})$.

Previously, we mentioned that there are several different notions of high dimensional expansion: some geometric or topological, some combinatorial. We now formally define high dimensional spectral expansion, which is a more combinatorial and graph theoretic notion:

Definition 2.2.25. $\mathcal{S}$ is a *two-sided λ-local spectral expander* if $\mathsf{GlobalExp}(\mathcal{S}) \geqslant \lambda$ and $\mathsf{LocalExp}(\mathcal{S}) \geqslant \lambda$.

High Order Walks on Simplicial Complexes Let $\mathcal{S}$ be a H-dimensional simplicial complex and with weight function $w : k\text{-}\mathsf{faces}(\mathcal{S}) \to \mathbb{R}_{\geqslant 0}$ on the k-faces of $\mathcal{S}$, for $k \leqslant H$. For each $k < H$, we can define a natural (periodic) Markov chain on a state space consisting of k-faces and $(k+1)$-faces of $\mathcal{S}$.

- At a $(k+1)$-face J, there are exactly $(k+2)$ faces $F \in k\text{-}\mathsf{faces}(\mathcal{S})$ such that $F \subset J$, due to the downward closure property. We transition from J to each k-face F with probability $\frac{1}{k+2}$.
- At a k-face F, we transition to each $(k+1)$-face J satisfying $J \supset F$ with probability $\frac{w(J)}{w(F)}$. (Note that w must be balanced for these transitions to be well-defined.)

Restricting the above chain to only odd or even time steps gives us two new random walks: one entirely on $k\text{-}\mathsf{faces}(\mathcal{S})$ and one entirely on $(k+1)\text{-}\mathsf{faces}(\mathcal{S})$.

Definition 2.2.26 (Down-up walk on k-faces of $\mathcal{S}$)**.** = Let $\mathcal{S}_{k+1}^{\downarrow\uparrow}$ be the Markov chain with state space equal to k-$\mathsf{faces}(\mathcal{S})$ and transition probabilities $\mathcal{S}_{\downarrow\uparrow}[J \to J']$ described by the process above, where there is an implicit transition *down* to a k-face and back *up* to a $(k+1)$-face. Then:

$$\mathcal{S}_{\downarrow\uparrow}[J \to J'] = \begin{cases} \dfrac{1}{k+1} \displaystyle\sum_{F \in k\text{-}\mathsf{faces}(\mathcal{S}) : F \subset J} \dfrac{w(J)}{w(F)} & \text{if } J = J' \\ \dfrac{1}{k+1} \cdot \dfrac{w(J')}{w(J \cap J')} & \text{if } J \cap J' \in k\text{-}\mathsf{faces}(\mathcal{S}) \\ 0 & \text{otherwise} \end{cases}$$

Definition 2.2.27 (Up-down walk on k-faces of $\mathcal{S}$)**.** Let $\mathcal{S}_{\uparrow\downarrow}$ be the Markov chain with state space equal to k-$\mathsf{faces}(\mathcal{S})$ and transition probabilities $\mathcal{S}_{\uparrow\downarrow}[F \to F']$ described by the process above, where there is an implicit transition *up* to a $(k+1)$-face and back *down* to a k-face. Then:

$$\mathcal{S}_{\downarrow\uparrow}[F \to F'] = \begin{cases} \dfrac{1}{k+1} & \text{if } F = F' \\ \dfrac{w(F \cup F')}{w(F)} & \text{if } F \cup F' \in (k+1)\text{-}\mathsf{faces}(\mathcal{S}) \\ 0 & \text{otherwise} \end{cases}$$

Remark 2.2.28. In the literature, we also see $\mathcal{S}_{k+1}^{\downarrow\uparrow}$ written as $\mathcal{S}_{k+1}^{\vee}$, and $\mathcal{S}_{k}^{\uparrow\downarrow}$ written as $\mathcal{S}_{k}^{\wedge}$.

We now present some facts about these high order walks without proof. We refer to [68, 5] for proofs of these facts.

Fact 2.2.29. *The transition matrices for $\mathcal{S}_{k+1}^{\downarrow\uparrow}$ and $\mathcal{S}_{k}^{\uparrow\downarrow}$ share the same eigenvalues. The nonzero eigenvalues occur with the same multiplicity. A straightforward but important consequence of this fact is*

$$\mathsf{Spectrum}(\mathcal{S}_{k+1}^{\downarrow\uparrow}) = \mathsf{Spectrum}(\mathcal{S}_{k+1}^{\uparrow\downarrow})$$

Fact 2.2.30. *The Markov chains $\mathcal{S}_{k}^{\downarrow\uparrow}$ and $\mathcal{S}_{k}^{\uparrow\downarrow}$ have the same stationary distribution on k-$\mathsf{faces}(\mathcal{S})$, which is proportional to $w(F)$ for each $F \in k$-$\mathsf{faces}(\mathcal{S})$. We will call this distribution $\pi_k(\cdot)$.*

For the remainder of the paper, we will assume a uniform weight function on d-$\mathsf{faces}(\mathcal{S})$, which is useful for applications like sampling bases of a matroid [5]. When using the uniform weighting scheme, for $F \in k$-$\mathsf{faces}(\mathcal{S})$, there is a natural interpretation of $\pi_k(F)$: the fraction of d-faces that contain F as a subface. (We also note that we will use symbolic variables to represent various weight values, and that it is straightforward to adapt our computations to cases where we have uniform weights over k-$\mathsf{faces}(\mathcal{S})$ for any k.)

2.3 Local Densification of Expanders

For a graph G and H-dimensional simplicial complex $\mathcal{S}$, we give a way to combine the two to produce a bounded-degree H-dimensional complex $\mathsf{LocalDensifier}(G, \mathcal{S})$ of constant expansion. First, construct a graph G' with

1. vertex set equal to $V(G) \times V(\mathcal{S})$, and
2. edge set equal to $\{\{(v_1, b_1), (v_2, b_2)\} : \{b_1, b_2\} \in \mathsf{1\text{-}faces}(\mathcal{S}), \{v_1, v_2\} \in E(G) \text{ or } v_1 = v_2\}$.

$\mathsf{LocalDensifier}(G, \mathcal{S})$ is then defined as the H-dimensional pure complex whose H-faces are all cliques on $H+1$ vertices $\{(v_1, b_1), (v_2, b_2), \ldots, (v_{H+1}, b_{H+1})\}$ such that there exists an edge $\{a, b\}$ in G for which $v_1, \ldots, v_{H+1} \in \{a, b\}$.

To describe a k-face of $\mathsf{LocalDensifier}(G, \mathcal{S})$, we may also use the ordered pair (F, f), where F is a k-face of $\mathcal{S}$, and f is a function mapping each element of F to a vertex of G. Because of the local densifier's tensor structure, $\mathsf{image}(f)$ is either a single vertex, or a pair of vertices that form an edge in G.

Linear algebraically, we can think of this graph construction as adding a self loop to each vertex of G and then taking the tensor product with the 1-skeleton of $\mathcal{S}$.

Our construction is $\mathcal{Q} := \mathsf{LocalDensifier}(G, \mathcal{B})$, where $\mathcal{B}$ is equal to $\mathcal{K}_s^{(H)}$, the H-dimensional complete complex on some constant $s \geqslant H+1$ vertices, and G is a T-regular triangle-free expander graph on n vertices. We endow $\mathcal{Q}$ with a balanced weight function w induced by setting the weights of all H-faces to 1.

As a first step to understanding this construction, we inspect the weights induced on k-faces for $k < H$. Consider a k-face $F := \{(v_1, b_1), \ldots, (v_{k+1}, b_{k+1})\}$. A short calculation reveals that if $v_1, \ldots, v_{k+1}$ are all equal, then $w(F)$ is equal to $w_{J,k} := \binom{s}{H-k} \cdot [T2^{H-k} - (T-1)]$ and otherwise, $w(F)$ is equal to $w_{I,k} := \binom{s}{H-k} \cdot [2^{H-k}]$. Henceforth, write w_J and w_I instead of $w_{J,k}$ and $w_{I,k}$ when k is understood from context.

We now list out what we prove about $\mathcal{Q}$. Most importantly, we show:

Theorem 2.3.1. *For every $1 \leqslant k < H$, the Markov transition matrix $\mathcal{Q}_k^{\downarrow\uparrow}$ for down-up (and equivalently up-down) random walks on the k-faces satisfies:*

$$\mathsf{TwoSidedGap}\left(\mathcal{Q}_k^{\downarrow\uparrow}\right) \geqslant \frac{\mathsf{TwoSidedGap}(G)}{64T^2(k+1)^2(s-k)(2^k-1)} \ .$$

We dedicate Section 2.5 to proving Theorem 2.3.1.

As an immediate corollary of Theorem 2.3.1 and Theorem 2.2.11, we get that

Corollary 2.3.2. *Let N_k denote the number of k-faces in $\mathcal{Q}$. Then the ϵ-mixing time of $\mathcal{Q}_k^{\downarrow\uparrow}$ satisfies:*

$$t(\varepsilon) \leqslant \frac{64T^2(k+1)^2(s-k)(2^k-1)}{\mathsf{TwoSidedGap}(G)} \cdot \log\left(\frac{2N_k}{\varepsilon}\right) \ .$$

We note that $N_k = \Theta(n)$.

We also derive bounds on the expansion of links of $\mathcal{Q}$. In particular, as a direct consequence of Theorem 2.4.2 and the discussion of the expansion properties of the complete complex in Example 2.2.23, we conclude:

Theorem 2.3.3. *We can prove the following bounds on the local and global expansion of $\mathcal{Q}$:*

$$\begin{aligned} \mathsf{GlobalExp}(\mathcal{Q}) &\geqslant \left[\frac{1}{2} - \frac{1}{2 \cdot (T2^H + 1)}\right] \cdot \mathsf{TwoSidedGap}(G), \text{ and} \\ \mathsf{LocalExp}(\mathcal{Q}) &\geqslant \frac{1}{2}. \end{aligned}$$

Remark 2.3.4. Suppose $\boldsymbol{G}$ is a random T-regular (triangle-free) graph and $H \geqslant T$. Then the corresponding (random) simplicial complex $\boldsymbol{\mathcal{Q}}$, as a consequence of Friedman's Theorem [49][3], with high probability satisfies

$$\begin{aligned} \mathsf{TwoSidedGap}\left(\boldsymbol{\mathcal{Q}}_k^{\downarrow\uparrow}\right) &\geqslant \frac{T - 2\sqrt{T-1} - o_n(1)}{64T^3(k+1)^2(s-k)(2^k-1)} \\ \mathsf{GlobalExp}(\mathcal{S}_{\boldsymbol{\mathcal{Q}}}) &\geqslant \frac{T - 2\sqrt{T-1} - o_n(1)}{T+1}, \text{ and} \\ \mathsf{LocalExp}(\mathcal{S}) &\geqslant 1/2. \end{aligned}$$

Thus, $\boldsymbol{\mathcal{Q}}$ endows a natural distribution over simplicial complexes that gives a high-dimensional expander with high probability.

Remark 2.3.5. If G is *strongly explicit*, such as an expander from [101, 10], then $\mathcal{Q}$ is also strongly explicit since the tensor product of two strongly explicit graphs is also strongly explicit.

2.4 Local Expansion

For this entire section, we will mainly work with the complex $\mathsf{LocalDensifier}(G, \mathcal{S})$, so when we use $\mathsf{link}(\cdot)$ without a subscript, it will be with respect to $\mathsf{LocalDensifier}(G, \mathcal{S})$. Next, fix a face $\sigma = (F, f) \in k\text{-}\mathsf{faces}(\mathsf{LocalDensifier}(G, \mathcal{S}))$. In order to study the expansion of the 1-skeleton of $\mathsf{link}(\sigma)$, we need to first compute the weights on its 1-faces.

Let $\tau = \{(v_1, b_1), (v_2, b_2)\} \in 2\text{-}\mathsf{faces}(\mathsf{link}(\sigma))$, where as before, $v_i \in V(G)$ and $b_i \in 1\text{-}\mathsf{faces}(\mathcal{S})$. There are several cases we need to consider:

1. Case 1: $|\mathsf{image}(f)| = 2$.
   Here, $w_\sigma(\tau) = w(\tau \cup \sigma)$, which is proportional to the number of H-faces (F', f') that contain $\tau \cup \sigma$. The face $\tau \cup \sigma$ already has $(k+3)$ vertices, so there are $\binom{S}{H-(k+2)}$ possibilities of F'. There are $2^{H-(k+2)}$ choices for f', since $\mathsf{image}(f')$ must equal $\mathsf{image}(f)$.

[3] Friedman's theorem says that a random T-regular graph, whp, has two-sided spectral gap $(T - 2\sqrt{T-1} - o_n(1))/T$. Additionally, random graphs are triangle-free with constant probability.

2. Case 2: $|\mathsf{image}(f)| = 1$.

   a) Case 2(a): $v_1 = v_2 \in \mathsf{image}(f)$ and $\{b_1, b_2\} \in$ 2-$\mathsf{faces}(\mathsf{link}_{\mathcal{S}}(F))$.
   Again, there are $\binom{S}{H-(k+2)}$ possibilities for F'. Since $v_1 = v_2 \in \mathsf{image}(f)$, we will have $T \cdot [2^{H-(k+2)} - 1] + 1$ choices for f', as v_1 has T neighbors in G, and when f' is not constant on v_1, there are T choices for the other value it can take.

   b) Case 2(b): $v_1 \neq v_2$ but $(v_1, v_2) \in E(G)$, and $\{b_1, b_2\} \in$ 2-$\mathsf{faces}(\mathsf{link}_{\mathcal{S}}(F))$.
   Again, we have $\binom{S}{H-(k+2)}$ possibilities for F, but we only have $2^{H-(k+2)}$ choices for f'; the image of f' must be $\{v_1, v_2\}$.

   c) Case 2(c): $v_1 = v_2 \notin \mathsf{image}(f)$ but $v_1 \cup \mathsf{image}(f) \in E(G)$, and $\{b_1, b_2\} \in$ 2-$\mathsf{faces}(\mathsf{link}_{\mathcal{S}}(F))$.
   The analysis is identical to that of Case 2(b)

For simplicity, we'll assign weights to the elements of 2-$\mathsf{faces}(\mathsf{LocalDensifier}(G, \mathcal{S}))$ as below:

$$w(\{(v_1, b_1), (v_2, b_2)\}) = \begin{cases} w_{S,k} := 2^{H-(k+2)} & \text{for Case 1, 2(b), and 2(c)} \\ w_{C,k} := 1 + T(2^{H-(k+2)} - 1) & \text{for Case 2(a)} \end{cases}$$

(Here, the C and S denote "center" and "satellite," whose meanings will be more natural when discussing $\mathsf{link}(\sigma)$ when $\sigma \neq \emptyset$.)

Remark 2.4.1. Note that if we choose $\sigma = \emptyset$ (so $k = -1$), we simply get the weights of the 1-skeleton of $\mathsf{LocalDensifier}(G, \mathcal{S})$ itself, which will be useful for computing global expansion.

Theorem 2.4.2. *Let G be a triangle-free T-regular graph and let $\mathcal{S}$ be a pure H-dimensional simplicial complex. Then*

$$\mathsf{GlobalExp}(\mathsf{LocalDensifier}(G, \mathcal{S})) = \min\left\{\frac{T2^{H-1}}{T2^H - (T-1)} \cdot \mathsf{TwoSidedGap}(G), \mathsf{GlobalExp}(\mathcal{S})\right\},$$

and $\mathsf{LocalExp}(\mathsf{LocalDensifier}(G, \mathcal{S})) = \mathsf{TwoSidedGap}(\mathcal{S})$.

We omit the proof and refer the readers to section 4 of [83] for more details.

2.5 Spectral Gap of High Order Walks

In this section we omitted the proofs of many lemmas. These proofs can all be found in section 5 of [83].

2.5.1 Offsets and Colors

We now inspect the structure of the k-faces of our construction $\mathcal{Q}$ in more detail.

Definition 2.5.1 (k-faces of $\mathcal{Q}$)**.** The set of k-faces of $\mathcal{Q}$ is exactly equal to the set of tuples (F, f) where F is a k-face of $\mathcal{B}$ and f is a labeling of each element by endpoints of some edge $\{u, v\}$ in G. We call (F, f) *t-offset* if either $|\{x \in F : f(x) = u\}| = t$ or $|\{x \in F : f(x) = v\}| = t$.

Remark 2.5.2. Suppose $t \leqslant k + 1 - t$. Note that a $(k + 1 - t)$-offset state is also t-offset, but we will stick to the convention of describing such states as t-offset. For example, a $(k + 1)$-offset state is also 0-offset, but we will only use the term 0-offset.

Definition 2.5.3 (Coloring of k-faces of $\mathcal{Q}$)**.** We color a k-face (F, f) of $\mathcal{Q}$ with $\mathsf{image}(f)$. Each 0-offset face is then colored with a vertex of G and the remaining faces are each colored with an edge of G.

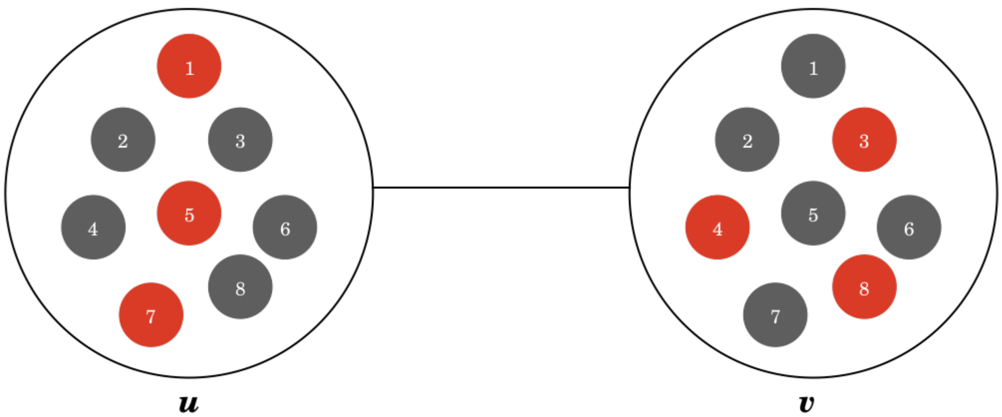


Figure 2.1: A 5-face in $\mathcal{Q}$. Corresponding 5-face in $\mathcal{B}$ is $\{1, 3, 4, 5, 7, 8\}$ is given by red vertices. Labeling is $(1, u), (3, v), (4, v), (5, u), (7, u), (8, v)$. $\{u, v\}$ is an edge in G. Color of 5-face is $\{u, v\}$

In the rest of the section, we study the spectral gap of the Markov chain $\mathcal{Q}_k^{\downarrow\uparrow}$, the down-up random walk on k-faces of $\mathcal{Q}$ induced by certain special weight functions — weight functions $w : k\text{-}\mathsf{faces}(\mathcal{Q}) \to \mathbb{R}_{\geqslant 0}$ with the property that there are two values w_I and w_J such that

$$w((F, f)) = \begin{cases} w_J & \text{if } (F, f) \text{ is 0-offset} \\ w_I & \text{otherwise.} \end{cases}$$

For the sequel, we use D to refer to the quantity $Tw_I + w_J$. The transition probabilities between states (F, f) and (F', f') depends on a number of conditions such as whether they are 0-offset or 1-offset or a different type, whether they arise from the same k-face in $\mathcal{B}$, and the colors of (F, f) and (F', f') respectively. A detailed treatment of the transition probabilities $\mathcal{Q}_k^{\downarrow\uparrow}[(F, f) \to (F', f')]$ can be found in Table 1 in Appendix A of [83]. From the transition probability table we observe that:

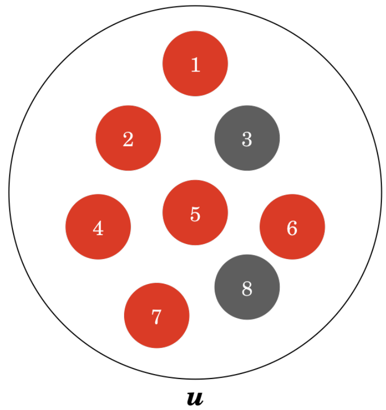


Figure 2.2: A 0-offset 5-face. Color of 5-face is $\{u\}$.

Observation 2.5.4. For all k-faces in $\mathcal{Q}_k^{\downarrow\uparrow}$, the self-loop probability is at least $\frac{1}{s-k} \cdot \frac{w_J}{D}$. Therefore, the smallest eigenvalue of $\mathcal{Q}_k^{\downarrow\uparrow}$ is at least $\frac{1}{s-k} \cdot \frac{w_J}{D} - 1$.

2.5.2 High-Level Picture of $\mathcal{Q}_k^{\downarrow\uparrow}$

As noted in the previous subsection, each k-face can be described by three parameters: a base face $F \in k\text{-faces}(\mathcal{B})$, a "color" set C that is either a single vertex or an edge in $E(G)$, and a function $f : F \to C$. The walk $\mathcal{Q}_k^{\downarrow\uparrow}$ is difficult to analyze directly, but by grouping states based on these three parameters, we can decompose the walk into a projection and restriction chain, and analyze it using the tools from [62].

At the outermost level, we can first group states into subchains based on their color. All subchains whose color is an edge (the rounded rectangles in Figure 2.3) are isomorphic to each other; similarly, all subchains whose color is a single vertex (the circles in Figure 2.3) are also isomorphic to each other. At first, it seems promising to partition $\mathcal{Q}_k^{\downarrow\uparrow}$ into these subchains; however, it is inconvenient that these subchains are not *all* isomorphic. To remedy this, we split the single-vertex-colored subchains into T isomorphic copies (with some changes to transition probabilities), and absorb them into the edge-colored subchains. This is detailed in the next section.

If we use this partition, the projection chain resembles a random walk on the *line graph* of G. Each restriction chain corresponds to all states of a single color C. The states are still represented by any base face $F \in k\text{-faces}(\mathcal{B})$ and any function $f : F \to C$. To analyze each of these restriction chains, it is simplest to apply [62] once more.

Now, we first group states by which base face F they correspond to. The subchains derived from fixing a particular F (the rectangles in Figure 2.4) are all isomorphic to each other, which leads to a much simplified analysis. Using this partition, the projection chain is simply the k-down-up walk on $\mathcal{B}$. Each restriction chain is thus over states corresponding to a fixed base face F and fixed color C, but the function $f : F \to C$ is allowed to vary. At

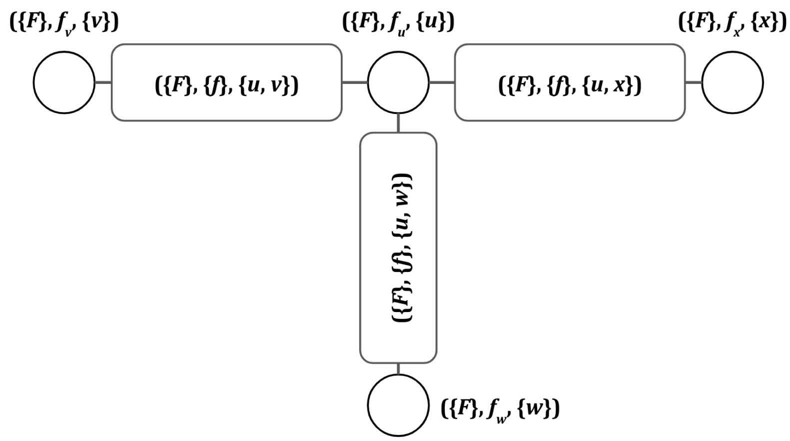


Figure 2.3: This figure illustrates $\mathcal{Q}_k^{\downarrow\uparrow}$, with states clustered by their color. The rounded rectangles correspond to colors that are edges, while circles correspond to colors that are single vertices. In each cluster, the $\{F\}$ indicates that all F could be represented. Similarly, $\{f\}$ indicates that any f with $\mathsf{image}(f)$ as the color set can be represented. We use f_u to denote the constant function on u.

this point, we may assume $|C| = 2$; thus f corresponds to assigning every element of F one of two elements. The inner restriction chain can be modeled by a hypercube.

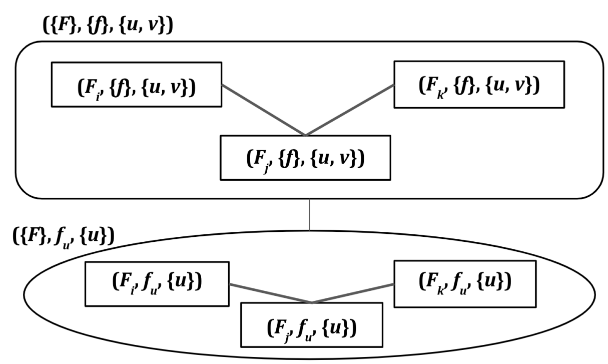


Figure 2.4: This figure illustrates a subchain of $\mathcal{Q}_k^{\downarrow\uparrow}$, for particular color $\{u, v\}$ and $\{u\}$. We can further cluster the states in this subchain by which face F in $\mathcal{B}$ they represent. Again, $\{f\}$ indicates that f can be any function with $\mathsf{image}(f)$ as the color.

Thus, the spectral gap of $\mathcal{Q}_k^{\downarrow\uparrow}$ is a combination of the spectral gaps of (1) the line graph of G, (2) the k-down-up walk on $\mathcal{B}$, and (3) the random walk on a hypercube.

2.5.3 Splitting 0-Offset Vertices

Towards our end goal of lower bounding the spectral gap of $\mathcal{Q}_k^{\downarrow\uparrow}$, we find it convenient to analyze a related Markov chain $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$, since the related chain has a natural partition into isomorphic subchains. $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$ has the property that its spectrum contains that of $\mathcal{Q}_k^{\downarrow\uparrow}$, which lets us translate a lower bound on the spectral gap of $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$ to a lower bound on the spectral gap of $\mathcal{Q}_k^{\downarrow\uparrow}$.

Definition 2.5.5 (Split chain $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$ and coloring of states in $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$)**.** We identify each state in $\mathsf{States}(\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow})$ with a tuple (F, f, c) where (F, f) is a face in $k\text{-}\mathsf{faces}(\mathcal{Q})$ and c is a color.

1. For each 0-offset face (F, f) in $k\text{-}\mathsf{faces}(\mathcal{Q})$, let $\{u\}$ be the color of F, and let the neighbors of u in G be $v_1, \ldots, v_T$. $\mathsf{States}(\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow})$ contains the states $(F, f, \{u, v_1\}), \ldots,$ $(F, f, \{u, v_T\})$ in place of the state (F, f, u).

2. For each remaining k-face (F, f) of $\mathcal{Q}$ (i.e. each k-face that isn't 0-offset), $\mathsf{States}(\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow})$ contains $(F, f, \mathsf{image}(f))$.

For each pair of states $(F, f, c), (F', f', c')$ in $\mathsf{States}(\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow})$,

$$\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}[(F, f, c) \to (F', f', c')] = \begin{cases} \frac{\mathcal{Q}_k^{\downarrow\uparrow}[(F,f)\to(F',f')]}{T} & \text{if } (F', f') \text{ is 0-offset} \\ \mathcal{Q}_k^{\downarrow\uparrow}[(F, f) \to (F', f')] & \text{otherwise.} \end{cases}$$

Intuitively, we want to split any transition to a 0-offset face in $\mathcal{Q}$ into T separate transitions in $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$, since each 0-offset face is also split into T new states.

Definition 2.5.6. We say two k-faces (F, f, e) and (F', f', e') have *identical base k-faces* if $F = F'$ and *different base k-faces* if $F \neq F'$.

Definition 2.5.7. Given a state (F, f, e) such that (F, f) is a 1-offset face, there is a single vertex v such that $f(v)$ is different from $f(u)$ for all u in $F \setminus \{v\}$. We call this vertex v a *lonely* vertex.

In the next lemma, we show that the spectrum of the original Markov chain $\mathcal{Q}_k^{\downarrow\uparrow}$ is contained in that of $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$.

Lemma 2.5.8. $\operatorname{Spec}\left(\mathcal{Q}_k^{\downarrow\uparrow}\right) \subseteq \operatorname{Spec}\left(\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}\right)$*, and therefore,* $\lambda_2(\mathcal{Q}_k^{\downarrow\uparrow}) \leqslant \lambda_2(\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow})$.

2.5.3.1 Stationary Distribution of $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$

If we want to apply the projection and restriction framework to $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$, we first need to compute its stationary distribution. To do this, we take advantage of the time-reversibility of the high order random walks, and apply the detailed balance equations.

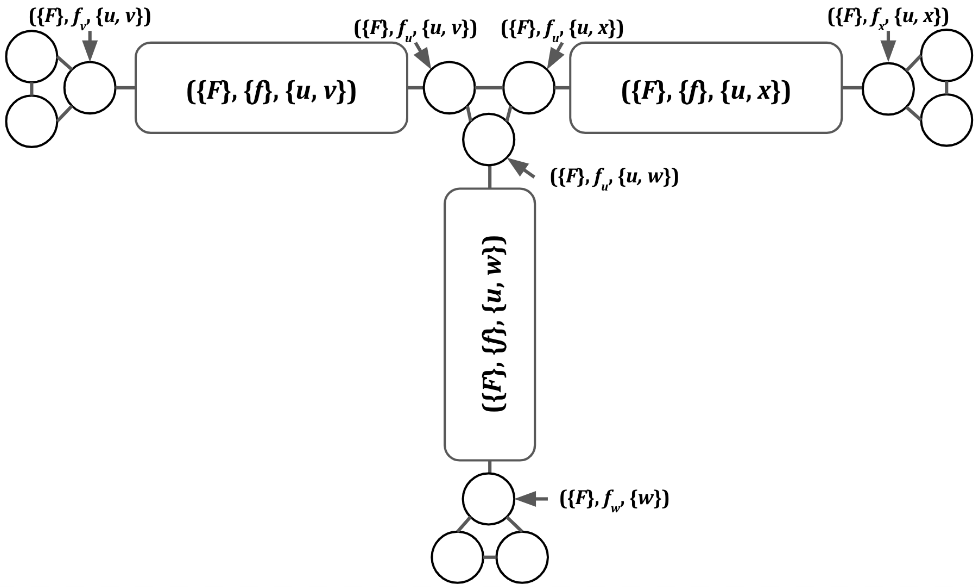


Figure 2.5: This figure illustrates the post-split vertices of Definition 2.5.5. The new vertices can take on any F, but their mappings f will be constant functions.

Lemma 2.5.9. *The stationary distribution of the Markov chain $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$ is given by:*

$$\pi_{\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}}(x) = \begin{cases} \dfrac{1}{|E(G)|} \cdot \dfrac{1}{\binom{s}{k+1}} \cdot \dfrac{1}{2} \cdot \dfrac{w_J}{(2^k - 1)Tw_I + w_J} & \text{for } x \text{ 0-offset} \\ \dfrac{1}{|E(G)|} \cdot \dfrac{1}{\binom{s}{k+1}} \cdot \dfrac{1}{2} \cdot \dfrac{Tw_I}{(2^k - 1)Tw_I + w_J} & \text{otherwise} \end{cases}$$

Proof. Via the detailed balance equations, we first observe that all vertices with the same offset have the same stationary distribution. Now, let x be any 0-offset vertex and y be any 1-offset vertex. Using the detailed balance equations, we have:

$$\pi_{\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}}(x) \cdot \frac{w_I}{(k+1)(s-k)D} = \pi_{\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}}(y) \cdot \frac{w_J}{(k+1)(s-k)DT}$$

Now, let x be any t-offset vertex, with $t \geqslant 1$, and let y be any $(t+1)$-offset vertex. Again, using the detailed balance equations:

$$\pi_{\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}}(x) \cdot \frac{1}{2(k+1)(s-k)} = \pi_{\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}}(y) \cdot \frac{1}{2(k+1)(s-k)}$$

From here, we see that all 0-offset faces have one stationary distribution probability, and all other faces also share the same stationary probability. The relations above tell us that for a 0-offset vertex x, and a t-offset vertex y with $t \geqslant 1$:

$$\frac{\pi_{\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}}(x)}{\pi_{\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}}(y)} = \frac{w_J}{Tw_I}$$

Normalizing so that $\sum_{x \in \widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}} \pi_{\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}}(x) = 1$ gives the desired result. □

2.5.4 Outer Projection and Restriction Chains

Now, we can further decompose $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$ into a projection chain and m isomorphic restriction chains, where $m = |E(G)|$, since we will have one partition element for each edge in G. Formally, we partition $\mathsf{States}(\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow})$ into m disjoint sets $\Omega_1 \cup \cdots \cup \Omega_m$, where $\Omega_i = \{(F, f, c) \mid c = e_i\}$.

2.5.4.1 The Outer Projection Chain

The partition Ω induces a projection chain $([m], P_o)$. The state space is $[m]$. The edge set is

$$E(P_o) = \{\{i, j\} \mid \exists (F, f, e_i) \in \Omega_i \text{ and } (G, g, e_j) \in \Omega_j \text{ s.t. } \widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}[(F, f, e_i) \to (G, g, e_j)] > 0\}$$

In words, we have an edge between i and j if there are transitions from Ω_i to Ω_j.
We obtain the following lower bound on the spectral gap of P_o.

Lemma 2.5.10. *The spectral gap of P_o is*

$$\frac{\mathsf{TwoSidedGap}(G)}{2} \cdot \frac{w_J + T w_I}{w_J + (2^k - 1) T w_I} \geqslant \frac{\mathsf{TwoSidedGap}(G)}{2(2^k - 1)} .$$

2.5.4.2 The Outer Restriction Chain

Each partition block Ω_i induces a restriction chain $R_{o,i}$. We show that all restriction chains $R_{o,i}$ for $i \in [m]$ are isomorphic.

Lemma 2.5.11. *For any $i \neq j$, $i, j \in [m]$, the restriction chains $R_{o,i}$ and $R_{o,j}$ are isomorphic.*

2.5.4.3 Stationary Distribution of $R_{o,1}$

To compute the spectral gap of $R_{o,1}$, we will further decompose the chain in the next section. In order to apply the projection and restriction framework once more to $R_{o,1}$, we must again compute a stationary distribution.

Lemma 2.5.12. *The stationary distribution of the outer restriction chain is given by:*

$$\pi_{R_{o,1}}(x) = \begin{cases} \dfrac{1}{\binom{s}{k+1}} \cdot \dfrac{1}{2} \cdot \dfrac{w_J}{(2^k - 1) T w_I + w_J} & \text{for } x \text{ 0-offset} \\ \dfrac{1}{\binom{s}{k+1}} \cdot \dfrac{1}{2} \cdot \dfrac{T w_I}{(2^k - 1) T w_I + w_J} & \text{otherwise} \end{cases}$$

Proof. By Fact 2.2.12, $R_{o,1}$ is also time-reversible. We proceed using the same analysis we used for Lemma 2.5.9. At the very end, we use a slightly different normalization to get the desired result. □

2.5.5 Inner Projection and Restriction Chains

Now, we are left to study the outer restriction chain, which, for a fixed $e \in E(G)$, is composed of all (F, f, e) in $\mathsf{States}(\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow})$. Again, we further decompose this chain into projection and restriction chains which are easier to analyze.

We group all (F, f, e) with the same $F \in k\text{-}\mathsf{faces}(\mathcal{B})$ into the same restriction state space Ω_F, which induces a projection chain resembling $\mathcal{B}_{\downarrow\uparrow}$, the down-up walk on k-faces of $\mathcal{B}$, and a restriction chain resembling a lazy random walk on a $(k+1)$-dimensional hypercube.

2.5.5.1 The Projection Chain

By defining the projection restriction chains as above, we end up with isomorphic restriction chains for each $F \in k\text{-}\mathsf{faces}(\mathcal{B})$. Thus, we can identify each of the states of the inner projection chain P_I with some face $F \in k\text{-}\mathsf{faces}(\mathcal{S})$. Let $\{F_i\}$ be this partition based on face.

Given $F, F' \in k\text{-}\mathsf{faces}(\mathcal{S})$, we can only transition from F to F' either when $F = F'$, or when $F \cap F' \in (k-1)\text{-}\mathsf{faces}$. This coincides with the feasible transitions in $\mathcal{B}_{\downarrow\uparrow}$.

We are able to obtain the following bounds on the spectral gap of the outer projection chain:

Lemma 2.5.13. $\mathsf{OneSidedGap}(P_I) \geqslant \dfrac{1}{2T(k+1)}$.

2.5.5.2 The Restriction Chain

Each restriction chain R_I can be treated as a $(k+1)$-dimensional hypercube with self loops. To see this, note that each restriction chain is a set of states (F, f, e) in $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$ where both F and e are the same. There are thus 2^{k+1} states in each restriction chain, since for each x, we have two choices for $f(x)$. Associating x where $f(x) = u$ to a 0-coordinate in a hypercube vertex, and x where $f(x) = v$ to a 1-coordinate, gives us a bijection from the restriction chain to the hypercube.

We show:

Lemma 2.5.14. *If we impose uniform weights on the highest order faces,*

$$\mathsf{OneSidedGap}(R_I) \geqslant \frac{w_J}{2Tw_I} \cdot \frac{2w_J}{D(k+1)(s-k)} \geqslant \frac{1}{(k+1)(s-k)} \ .$$

2.5.6 Rapid Mixing for High Order Random Walks

Now we put together the decomposition theorem, the lower bounds for the spectral gaps of the project and restriction chains, and Observation 2.5.4 to obtain the following lower bound on the two-sided spectral gap of $\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}$:

Theorem 2.5.15 (Restatement of Theorem 2.3.1)**.** *The k down-up random walk $\mathcal{Q}_k^{\downarrow\uparrow}$ has one-sided spectral gap,*

$$\mathsf{TwoSidedGap}(\mathcal{Q}_k^{\downarrow\uparrow}) \geqslant \frac{\mathsf{TwoSidedGap}(G)}{64T(k+1)^2(s-k)(2^k-1)} \quad . \tag{2.1}$$

Proof. Use $\mathsf{OneSidedGap}(M)$ to denote the spectral gap of a Markov chain M. We deduce from Lemma 2.5.8 and Theorem 2.2.13 that

$$\begin{aligned}
\mathsf{OneSidedGap}(\mathcal{Q}_k^{\downarrow\uparrow}) &\geqslant \mathsf{OneSidedGap}(\widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}) \\
&\quad (\text{Lemma 2.5.8}) \\
&\geqslant \min\left\{\frac{\mathsf{OneSidedGap}(P_o)}{3}, \frac{\mathsf{OneSidedGap}(P_o)\mathsf{OneSidedGap}(R_{o,1})}{3\gamma_o + \mathsf{OneSidedGap}(P_o)}\right\} \\
&\quad (\text{Theorem 2.2.13 on } \widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}) \\
&\geqslant \min\left\{\frac{\mathsf{OneSidedGap}(P_o)}{3},\right. \\
&\quad \frac{\mathsf{OneSidedGap}(P_o)}{3\gamma_o + \mathsf{OneSidedGap}(P_o)} \cdot \frac{\mathsf{OneSidedGap}(P_I)}{3}, \\
&\quad \left.\frac{\mathsf{OneSidedGap}(P_o)}{3\gamma_o + \mathsf{OneSidedGap}(P_o)} \cdot \frac{\mathsf{OneSidedGap}(P_I)\mathsf{OneSidedGap}(R_I)}{3\gamma_I + \mathsf{OneSidedGap}(P_I)}\right\} \\
&\quad (\text{Theorem 2.2.13 on } R_{o,1}),
\end{aligned}$$

where

$$\gamma_o = \max_{i\in[m]} \max_{x\in\Omega_i} \sum_{y\in\Omega\setminus\Omega_i} \widetilde{\mathcal{Q}}_k^{\downarrow\uparrow}(x,y) < 1$$

$$\gamma_I = \max_{F\in k\text{-}\mathsf{faces}(\mathcal{S})} \max_{x\in V(R_I)} \sum_{y\in V(R_{o,1})\setminus V(R_I)} R_{o,1}(x,y) < 1.$$

Furthermore, Lemma 2.5.10, Lemma 2.5.13 and Lemma 2.5.14 provide lower bounds for $\mathsf{OneSidedGap}(P_o)$, $\mathsf{OneSidedGap}(P_I)$, and $\mathsf{OneSidedGap}(R_I)$. If we substitute the spectral-gap lower bounds, and an upper bound of 1 for both γ_o and γ_I, we obtain a lower bound on $\mathsf{OneSidedGap}(\mathcal{Q}_k^{\downarrow\uparrow})$:

$$\mathsf{OneSidedGap}(\mathcal{Q}_k^{\downarrow\uparrow}) \geqslant \frac{\mathsf{TwoSidedGap}(G)}{64T(k+1)^2(s-k)(2^k-1)}. \tag{2.2}$$

Observation 2.5.4 gives a lower bound on $1 - |\lambda_{\min}(\mathcal{Q}_k^{\downarrow\uparrow})|$ larger than the right hand side of (2.2), which immediately lets us upgrade the statement (2.2) to (2.1), thus proving the theorem. □

Chapter 3

2-dim Expanders from Random Geometric Graphs

The simplest example of a 2-dimensional expander is the complete complex, based on the complete graph K_n. Sparse examples are known as well (e.g. [14, 77, 90, 91, 64]), though at first their existence may seem remarkable: such graphs must be globally sparse, and yet the $O(1)$-sized local neighborhood of every vertex must be densely connected to ensure sufficient expansion. This is a delicate balance, and indeed given the state of our knowledge today the phenomenon of sparse high-dimensional expansion seems "rare," in sharp contrast with the ubiquity of 1-dimensional expansion. Only a few sparse constructions are currently known, and many of these constructions are algebraic, inheriting their expansion properties from the groups used to define them (as discussed further in Section 3.1.2).

A prominent open problem in the area is to identify natural distributions over sparse higher-dimensional expanders [87, 82, 86]; this would be highly beneficial, both for a deeper mathematical understanding and for applications in algorithms and complexity. The simplest distributions immediately fail: random d-regular graphs are locally treelike, and so with high probability $G[N(v)]$ will be an independent set (with $\lambda = 1$) for most $v \in V$. The same is true for an Erdős-Rényi graph $\mathsf{G}(n,p)$ when $p \ll \frac{1}{\sqrt{n}}$. Though a number of distributions have been shown to have some higher-dimensional expansion properties [80, 47, 18, 19, 83, 53], they all fall short in some sense: either they are quite dense (degree $\Omega(\sqrt{n})$) or fail to satisfy the spectral conditon $\lambda < \frac{1}{2}$. In this work, our primary question is the following:

Are there natural, high-entropy distributions over 2-dimensional expanders of average degree $\ll \sqrt{n}$?

We answer this question in the affirmative. We prove that for any $\varepsilon > 0$ and any large enough $n \in \mathbb{N}$, there exists a choice of $d \in \mathbb{N}$ such that a random n-vertex geometric graph on $\mathbb{S}^{d-1}$ with average degree n^ε is a 2-dimensional expander with high probability.

3.1 Problem background and summary of results

3.1.1 Our results

In order to state our results, we first recall some definitions.

Definition 3.1.1 (Simplicial complex)**.** A *k-dimensional simplicial complex* X is a downward-closed collection of subsets of size at most $k+1$ over some ground set X_0, with a downward-closed weight function w.[1] Any $S \in X$ is called a $(|S|-1)$-face, and the restriction of X to sets of size at most $\ell + 1 \leqslant k$ is called the *ℓ-skeleton* of X. The *degree* of $v \in X_0$ is the number of top-level faces that contain it.

For example, the set of all cliques of size at most $k+1$ in a graph G, where the weight of each clique is proportional to the number of $(k+1)$-cliques it occurs in, defines a k-dimensional simplicial complex.

Definition 3.1.2 (Link)**.** Let X be a simplicial complex. For any face S, the *link* of S in X is the simplicial complex X_S with weight function w_S, consisting of all sets in X which contain S, minus S:

$$X_S = \{T \setminus S \mid T \supseteq S, T \in X\}, \qquad w_S(T \setminus S) = w(T) \quad \forall T \in X_S$$

For example, in the simplicial complex whose highest order faces are the triangles in a graph G, the link of a vertex v is the induced graph on the neighbors of v with its isolated vertices removed.

We are interested in simplicial complexes where the links expand enough to trigger a "local-to-global phenomenon" via the trickling-down theorem, stated below in the 2-dimensional case.[2]

Theorem 3.1.3 (Trickling-down theorem [97])**.** *Let X be a 2-dimensional simplicial complex. If its 1-skeleton is connected, and the second eigenvalue of every link's random walk matrix is at most λ, then the second absolute eigenvalue of the random walk matrix of the 1-skeleton of X is at most $\frac{\lambda}{1-\lambda}$.*

This theorem explains the significance of $\lambda = \frac{1}{2}$, since when $\lambda < \frac{1}{2}$, local expansion "trickles down" to imply global expansion. We will show that random geometric graphs, in a carefully-chosen parameter regime, have sufficient link expansion.

[1] Recall that X is called downward-closed if $S \subseteq T$ and $T \in X$ imply $S \in X$, and w is called downward-closed if weights are assigned to maximal faces, and for each non-maximal $S \in X$, we recursively define $w(S) = \sum_{x \in X_0} w(S \cup \{x\})$.

[2] The trickling-down theorem also generalizes to higher dimensions: sufficiently strong local spectral expansion of only the highest-order links implies global spectral expansion.

Definition 3.1.4 (Random geometric graph)**.** A *random geometric graph* $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$ is sampled as follows: for each $i \in [n]$, a vector $\boldsymbol{u}_i$ is drawn independently from the uniform distribution over $\mathbb{S}^{d-1}$ and identified with vertex i. Then, each edge $\{i,j\}$ is included if and only if $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle \geqslant \tau$ where $\tau = \tau(p,d)$ is chosen so that $\mathbf{Pr}_{\mathsf{Geo}_d(n,p)}[(i,j) \in \boldsymbol{G}] = p$.

Definition 3.1.5 (Random geometric complex)**.** The *random geometric k-complex* $\mathsf{Geo}_d^{(k)}(n,p)$ is the distribution defined by sampling $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$ and taking the downward-closure of the complex whose k-faces are the cliques of size $(k+1)$ in $\boldsymbol{G}$.

Our main result proves that there are conditions under which random geometric 2-complexes of degree n^ε are high-dimensional expanders enjoying the trickling-down phenomenon:

Theorem 3.1.6. *For every $0 < \varepsilon < 1$, there exist constants C_ε and $\delta = \exp(-O(1/\varepsilon))$ such that when $\boldsymbol{H} \sim \mathsf{Geo}_d^{(2)}(n, n^{-1+\varepsilon})$ for $d = C_\varepsilon \log n$, with high probability every vertex link of $\boldsymbol{H}$ is a $(\frac{1}{2} - \delta)$-expander, and hence its 1-skeleton is a $(1 - \frac{4\delta}{1+2\delta})$-expander.*

Remark 3.1.7. The complexes arising from Theorem 3.1.6 with high probability have degree bounded by $O(n^{2\varepsilon})$, as the number of triangle a vertex participates in is the square of its degree in the 1-skeleton.

Along the way, we also analyze the spectrum of $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$ directly and obtain sharper control of its second eigenvalue in a more general setting, giving bounds on the spectral norm of random geometric graphs in the full high-dimensional ($d \to_n \infty$) regime. To our knowledge, previous results in this vein are only for $d \sim n^{1/k}$ for fixed integers k [34, 17, 31, 11, 38, 85].

Theorem 3.1.8. *Let $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$ and $\tau \coloneqq \tau(p,d)$. Then with high probability $\boldsymbol{G}$ is a μ-expander, where*

$$\mu \coloneqq (1 + o(1)) \cdot \max\left\{(1 + o_{d\tau^2}(1)) \cdot \tau, \frac{\log^4 n}{\sqrt{pn}}\right\},$$

where $o_{d\tau^2}(1)$ denotes a function that goes to 0 as $d \cdot \tau(p,d)^2 \to \infty$.

In Section 3.8 we show that an eigenvalue close to τ is achieved (for some p, d), so Theorem 3.1.8 is close to sharp. Since in Theorem 3.1.6 we show that the vertex links of $\boldsymbol{G}$ have eigenvalue $\lambda \leqslant \frac{\tau}{1+\tau}$, this implies that the trickling-down theorem is tight

Proposition 3.1.9 (Trickling-down theorem is tight)**.** *For each $\lambda \in (0, \frac{1}{2}]$ and $\eta > 0$ there exists a 2-dimensional expander in which all vertex link eigenvalues are at most λ for which the 1-skeleton is connected with eigenvalue at least $\frac{\lambda}{1-\lambda} - \eta$.*

Spectra of random restrictions. Theorem 3.1.8 (and morally Theorem 3.1.6) is a consequence of a more general theorem that we prove concerning the spectral properties of random restrictions of graphs. We describe this result here, both because it may be of independent interest, and because it may help demystify Theorem 3.1.6.

Random restriction is a procedure for approximating a large graph X by a smaller graph $\boldsymbol{G}$: one selects a random subset of vertices $\boldsymbol{S}$, and then takes $\boldsymbol{G}$ to be the induced graph $X[\boldsymbol{S}]$. The random restriction $\boldsymbol{G}$ is now a smaller (and often sparser) approximation to X; this idea has been useful in a number of contexts in theoretical computer science (e.g. [52, 3, 9, 74, 60]). The core question is: to what extent do random restrictions actually inherit properties of the original graph? We will show that if random walks on X mix rapidly enough, then random restrictions inherit the *spectral* properties of the original graph.

To see the relevance of this result in our context, notice that a random geometric graph on the sphere is a random restriction of the (infinite) graph with vertex set $\mathbb{S}^{d-1}$ and edge set $\{(u, v) \mid \langle u, v\rangle \geqslant \tau\}$. Theorem 3.1.6 is then a consequence of the fact that the sphere is *itself* a 2-dimensional expander.

We state the theorem precisely below.

Definition 3.1.10 (Random restriction)**.** Suppose X is a (possibly infinite) graph, and that the simple random walk on X has unique stationary distribution ρ. We define an *n-vertex random restriction* of X to be a graph $\boldsymbol{G} \sim \mathsf{RR}_n(X)$ sampled by sampling n vertices independently according to ρ, $\boldsymbol{S} \sim \rho^{\otimes n}$, then taking $\boldsymbol{G} = X[\boldsymbol{S}]$ to be the graph induced on those vertices.

We show that if the average degree in $\boldsymbol{G}$ is not too small, $\lambda_2(\boldsymbol{G})$ reflects the rapid mixing of the random walk on X.

Theorem 3.1.11. *Let X be a (possibly infinite) vertex-transitive graph on which the associated simple random walk has a unique stationary distribution ρ, and let $p = \mathbf{Pr}_{\boldsymbol{G}\sim\mathsf{RR}_n(X)}[(i, j) \in E(\boldsymbol{G})]$ be the marginal edge probability of a n-vertex random restriction of X. Suppose there exist $C \geqslant 1$ and $\lambda \in [(np)^{-1/2}, 1]$ such that for any $k \in \mathbb{N}$, k-step walks on X satisfy the following mixing property: for any distribution α over $V(X)$,*

$$\mathrm{d}_{\mathrm{TV}}\left(X^k\alpha, \rho\right) \leqslant C \cdot \lambda^k,$$

where X^k denotes the k-step random walk operator on X, and furthermore suppose $pn \gg C^6 \log^4 n$. Then for any constant $\gamma > 0$,

$$\mathop{\mathbf{Pr}}_{\boldsymbol{G}\sim\mathsf{RR}_n(X)}\left[\left|\lambda_2(\widehat{A}_{\boldsymbol{G}})\right|, \left|\lambda_n(\widehat{A}_{\boldsymbol{G}})\right| \leqslant (1 + o(1)) \cdot \max\left(\lambda, \frac{\log^4 n}{\sqrt{pn}}\right)\right] \geqslant 1 - n^{-\gamma},$$

where $\widehat{A}_{\boldsymbol{G}}$ is the (normalized) adjacency matrix of $\boldsymbol{G}$.

Remark 3.1.12. It is likely that some of the conditions of Theorem 3.1.11 could be weakened. The decay of total variation could plausibly be replaced with a (much weaker) assumption about the spectral gap of X; this would not impact our results for $\mathbb{S}^{d-1}$, but may be useful in other applications. Transitivity is assumed mostly to make the proof of Theorem 3.1.11 go through at this level of generality; to prove Theorem 3.1.6 we re-prove a version of Theorem 3.1.11 for the specific non-transitive case where X is a link of a vector in the sphere (a spherical cap).

3.1.2 Related work

We give a brief overview of related work. While so far we have focused on a spectral notion of high-dimensional expanders (HDX), there are two additional notions: coboundary and cosystolic expansion. These are meant to generalize the Cheeger constant, a cut-based measure of graph expansion.

Distributions over high-dimensional expanders. The existence of natural distributions over sparse HDXs has been a question of interest since sparse HDX were first shown to exist (and this was highlighted as an important open problem in e.g. [87, 86]).

The early work of Linial and Meshulam [80] considered the distribution over 2-dimensional complexes in which all edges $\binom{[n]}{2}$ are included, and each triangle is included independently with probability p; they identified the phase transition at p for coboundary connectivity for this distribution (see also the follow-ups [6, 93, 81]). This distribution has the drawback that the 1-skeleton of these complexes is K_n, and so the resulting complex is far from sparse.

In [47], the authors show that a union of d random partitions of $[n]$ into sets of size $k+1$ with high probability produces a *geometric* expander [54], which is a notion of expansion which measures how much the faces must intersect when the complex is embedded into $\mathbb{R}^k$. The resulting complexes have disconnected links when $d \ll \sqrt{n}$, and so they fail to be spectral HDXs.

The work of [83] introduces a distribution over spectral expanders with expansion exactly $\frac{1}{2}$ by taking a tensor product of a random graph and a HDX; the authors show that down-up walks on these expanders mix rapidly, and [53] introduces a reweighing of these complexes which yields improved mixing time bounds. However, the links in these complexes fail to satisfy $\lambda < \frac{1}{2}$, and so fall outside of the range of the trickling-down theorem. The same drawback applies to [18, 19]: they show that up-down walks mix on random polylogarithmic-degree graphs given by subsampling a random set of generators of a Cayley graph. However, these graphs do not satisfy the conditions of the trickling-down theorem.

Explicit constructions. One of the first constructions of sparse high-dimensional spectral expanders was the Ramanujan complex of [14, 77, 90, 91], which generalize the Ramanujan expander graphs of [89]. Not only are these spectral expanders, but [64, 37] also show that they are co-systolic expanders. These Ramanujan complexes are algebraic by nature, constructed from the Cayley graphs of $\mathrm{PSL}_d(\mathbb{F}_q)$. Other algebraic constructions include that of [66]; the authors analyze the expansion properties of coset complexes for various matrix groups. They achieve sparse spectral expanders, with local expansion arbitrarily close to 0. More recently, [95] extend the coset complex construction to the more general family of Chevalley groups.

A few combinatorial constructions for HDX are also known. [15] prove that objects called (a, b)-expanders are two-dimensional spectral expanders; they give a graph-product-inspired construction of a family of such expanders, and show that other known complexes [14, 77, 90, 91, 66] are also (a, b)-expanders. Their work is extended by [48] to higher dimensions.

Applications of HDX. The local-to-global phenomenon in HDX has already been useful in many settings. [24] use high-dimensional spectral expanders to construct "agreement expanders," whose links give rise to local agreement tests: given "shards" of a function that pass a large fraction of the local agreement tests, the authors can conclude the presence of a "global" function g that stitches the shards together. In coding theory, the locally testable codes of [28] and quantum LDPCs of [98, 75, 26] utilize a common simplicial-complex-like structure called the square Cayley complex, whose local-to-global properties are essential in the analysis of these codes.

The local-to-global phenomenon also implies that "down-up" walks on the associated simplicial complex mix (as made formal in [2]). A k-down-up walk is supported k-faces of the simplicial complex, and transitions occur by dropping down into a random $(k-1)$-face, then transitioning up to a random k-face (one can also define the "up-down" walk analagously). This local-to-global analysis has recently been influential in the study of mixing times of Markov chains. Several well-studied Markov chains can be recast as the k-down-up random walk of a carefully designed simplicial complex. One notable example is the matroid basis exchange walk, which is an algorithm for sampling independent sets of a matroid (e.g. spanning trees in the graphical matroid). [5] were able to obtain an improved mixing time bound for the basis exchange walk–a significant breakthrough that, due to the local-to-global property, was achieved through the analysis of simple, "local" view of the matroids.

Random geometric graphs and random kernel matrices. Random restrictions of metric spaces such as $\mathbb{S}^{d-1}$ and $[-1, 1]^d$ are well-studied in the fixed-dimensional regime, where $d = O(1)$ and $n \to \infty$ (see the survey of Penrose [100]). In our work we are interested in the high-dimensional setting, where $d \to \infty$ with n. The high-dimensional setting was first studied only recently, initiated by [22, 13], and many mysteries remain in this young area of study.

Our Theorem 3.1.8 is related to the study of *kernel random matrices*: random $n \times n$ matrices whose (i, j)-th entry is given by $f_d(\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle)$, for $f_d : \mathbb{R} \to \mathbb{R}$ and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ sampled independently from some distribution over $\mathbb{R}^d$. The special case of $\boldsymbol{u}_i \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ and $f_d(x) = \mathbf{1}[x \geqslant \tau(p, d)]$ yields the adjacency matrix of $\mathsf{Geo}_d(n, p)$. A line of work initiated by [73] studies the spectrum of kernel random matrices [34, 17, 31, 11, 38], and the most recent work [85] characterizes the limiting empirical spectral distribution when $d = \Theta(n^{1/k})$ for k a fixed constant and f can be "reasonably" approximated by polynomials (in a sense that is flexible enough to capture the indicator $f_d(x) = \mathbf{1}[x \geqslant \tau(p, d)]$). In comparison with our results, they characterize the entire empirical spectral distribution, but we do not need to restrict $d \sim n^{1/k}$ for integer k, which is crucial for our applications.

3.1.3 Overview of the proof

We now explain how we prove our main theorem, Theorem 3.1.6, which states that for a complex sampled from $\boldsymbol{H} \sim \mathsf{Geo}_d^{(2)}(n,p)$ for $p = n^{-1+\varepsilon}$ with $0 < \varepsilon < 1$ and $d = C_\varepsilon \log n$, with high probability every link of $\boldsymbol{H}$ is a $\left(\frac{1}{2} - \delta\right)$-expander for some $\delta = \exp(-O(1/\varepsilon))$, and its 1-skeleton is a $\left(1 - \frac{4\delta}{1+2\delta}\right)$-expander. By the trickling-down theorem, it suffices for us to prove:

1. All n vertices' corresponding links in $\boldsymbol{H}$ are $\left(\frac{1}{2} - \delta\right)$-expanders with high probability.

2. The 1-skeleton of $\boldsymbol{H}$ is connected with high probability.

To show Item 2, it is enough to show that some reweighting of the 1-skeleton expands; Item 1 implies that every edge (i,j) must participate in at least one triangle (otherwise the link would contain isolated vertices), so the unweighted 1-skeleton is just the adjacency matrix of an unweighted graph from $\mathsf{Geo}_d(n,p)$. En route to proving Item 1 we'll prove that unweighted random geometric graphs expand, by this logic yielding Item 2 a consequence.

Analyzing link expansion. We establish Item 1 by showing that that each of the n links is a $\left(\frac{1}{2} - \delta\right)$-expander with probability $1 - o(1/n)$, then applying a union bound. We can think of sampling the link of vertex i_w in $\boldsymbol{H}$ by first choosing the number of neighbors $\boldsymbol{r} \sim \mathsf{Binom}(n-1,p)$, then sampling $\boldsymbol{r}$ points $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{\boldsymbol{r}}$ independently and uniformly from a measure-p cap in $\mathbb{S}^{d-1}$ centered at some point w (corresponding to the vector of the link vertex i_w), placing an edge between every i,j such that $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle \geqslant \tau(p,d)$. Finally, we remove any isolated vertices; here, we'll show that the graph expands with high probability before removing these isolated vertices, which implies that no isolated vertices have to be removed. For the remainder of the overview, let $\tau = \tau(p,d)$. We'll show that:

Theorem 3.1.13 (Informal version of Theorem 3.5.1)**.** *Let $\boldsymbol{G}$ be the link of some point $w \sim \mathbb{S}^{d-1}$ induced by $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m \sim \mathrm{cap}_p(w)$. Then with high probability $\boldsymbol{G}$ is a μ-expander where*

$$\mu := (1 + o(1)) \cdot \max\left\{ \frac{\tau}{\tau+1}, \frac{\log^4 m}{\sqrt{qm}} \right\} + o_d(1).$$

Here $q = \mathbf{Pr}_{u,v\sim\mathbb{S}^{d-2}}\left[\langle u, v\rangle \geqslant \frac{\tau}{\tau+1}\right]$.

Links are essentially random geometric graphs in one lower dimension. Since most of the measure of the cap lies close to its boundary, intuitively the link is distributed *almost* like a random geometric graph with points drawn independently from the cap boundary, i.e. the shell $\mathrm{shell}_p(w) := \{x : \langle x, w\rangle = \tau\}$. Our proof of Theorem 3.1.13 must pay attention to the fluctuations in $\langle \boldsymbol{v}_i, w\rangle - \tau$, but to simplify our current discussion we assume each link is in fact a random geometric graph on $\mathrm{shell}_p(w)$, and address the fluctuations later in the overview.

Observe that a uniformly random $\boldsymbol{v}$ from $\text{shell}_p(w)$ is distributed as $\tau \cdot w + \sqrt{1-\tau^2} \cdot \boldsymbol{u}$ where $\boldsymbol{u}$ is a uniformly random unit vector orthogonal to w. Using this decomposition, we see that $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle \geqslant \tau$ if and only if $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle \geqslant \frac{\tau}{1+\tau}$. Thus, under our simplifying assumption, the link is distributed exactly like a random geometric graph on $\mathbb{S}^{d-2}$ with inner product threshold $\frac{\tau}{1+\tau}$. Hence (up to the difference between $\text{cap}_p(w)$ and $\text{shell}_p(w)$), to understand link expansion we can study the second eigenvalue of a random geometric graph on the sphere.

Remark 3.1.14 (Requiring $d = \Theta(\log n)$)**.** In light of Theorem 3.1.13 (and even the heuristic discussion above), it turns out that $d = \Theta(\log n)$ is the only regime for which the links can be connected while the 1-skeleton has average degree $\ll \sqrt{n}$. To see this, we consider the relationship between p, τ, and d; we have that

$$p = \Pr_{\boldsymbol{v}, \boldsymbol{v}' \sim \mathbb{S}^{d-1}} [\langle \boldsymbol{v}, \boldsymbol{v}' \rangle \geqslant \tau] = \Theta\left(\tfrac{1}{\tau d}\right) \cdot \left(1 - \tau^2\right)^{\frac{d-1}{2}} \approx \exp(-d\tau^2/2). \tag{3.1}$$

See Lemma 3.2.8 for a formal argument.[3] Note that the arguments above in conjunction with (3.1) imply that the probability that two vertices within the link are connected is also roughly

$$q = \Pr_{\boldsymbol{u}, \boldsymbol{u}' \sim \mathbb{S}^{d-2}} \left[\langle \boldsymbol{u}, \boldsymbol{u}' \rangle \geqslant \tfrac{\tau}{1+\tau}\right] = \Theta\left(\tfrac{1}{\tau d}\right) \cdot \left(1 - \tfrac{\tau^2}{(1+\tau)^2}\right)^{\frac{d-2}{2}},$$

since the link is like a random geometric graph on $\text{shell}_p(w)$.

Connectivity within the links in conjunction with sparsity now requires us to have $d \in \Theta(\log n)$: The number of vertices inside each link concentrates around $m = np$, so the average degree inside the link is $qm \approx qpn$; we must have the average link degree $qpn \geqslant 1$, otherwise the link is likely disconnected. Now, if $\tau = o(1)$, then $\tau \approx \frac{\tau}{1+\tau}$ and $p \approx q$, so $qpn \geqslant 1 \implies p^2 n \gtrapprox 1 \implies p \gtrapprox n^{-1/2}$, ruling out a 1-skeleton with average degree $\ll \sqrt{n}$. Hence we need $\tau = \Omega(1)$. Given that $\tau = \Omega(1)$, (3.1) implies that to have the average 1-skeleton degree $\sqrt{n} \geqslant pn \geqslant 1$ we need $d \in \Theta(\log n)$.

Spectral expansion in random geometric graphs. We now explain how to prove near-sharp second eigenvalue bounds for random geometric graphs.

Theorem (Restatement of Theorem 3.1.8)**.** *Let $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$ and $\tau := \tau(p,d)$. Then with high probability $\boldsymbol{G}$ is a μ-expander, where*

$$\mu := (1 + o(1)) \cdot \max\left\{(1 + o_{d\tau^2}(1)) \cdot \tau, \frac{\log^4 n}{\sqrt{pn}}\right\},$$

where $o_{d\tau^2}(1)$ denotes a function that goes to 0 as $d \cdot \tau(p,d)^2 \to \infty$.

[3] Heuristically, it makes sense that $p = \Pr[\langle \boldsymbol{v}, \boldsymbol{v}' \rangle \geqslant \tau] \approx \exp(-\Theta(\tau^2 d))$, because $\langle \boldsymbol{v}, \boldsymbol{v}' \rangle$ is approximately $\mathcal{N}(0, \frac{1}{d})$.

As mentioned above, Theorem 3.1.8 is a consequence of the more general Theorem 3.1.11 about the second eigenvalue of random restrictions of vertex-transitive graphs, and the inner product threshold $\tau = \tau(p, d)$ appears as the mixing rate of the random walk on $\mathbb{S}^{d-1}$ where a step originating at v walks to a random vector in $\mathrm{cap}_p(v)$. Via standard concentration arguments applied to the vertex degrees, to prove the above it suffices to bound $\|A_{\boldsymbol{G}} - \mathbf{E}\, A_{\boldsymbol{G}}\| \leqslant \mu \cdot pn$, where $A_{\boldsymbol{G}}$ is the (unnormalized) adjacency matrix of $\boldsymbol{G}$. We'll focus on the regime where $pn \gg \mathrm{poly}\log n$, so that $\mu \approx \tau$.

Trace method for random geometric graphs. To bound $\|A_{\boldsymbol{G}} - \mathbf{E}\, A_{\boldsymbol{G}}\|$, we employ the trace method, bounding the expected trace of a power of $A_{\boldsymbol{G}} - \mathbf{E}\, A_{\boldsymbol{G}}$. This is sufficient for the following reason: for convenience, let $\overline{A}_{\boldsymbol{G}} = A_{\boldsymbol{G}} - \mathbf{E}\, A_{\boldsymbol{G}}$, and let ℓ be any non-negative, even integer. Since ℓ is even,

$$\left\|\overline{A}_{\boldsymbol{G}}\right\|^{\ell} = \left\|\overline{A}_{\boldsymbol{G}}^{\ell}\right\| \leqslant \mathrm{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right),$$

And so applying Markov's inequality,

$$\mathbf{Pr}\left(\left\|\overline{A}_{\boldsymbol{G}}\right\| \geqslant e^{\varepsilon}\left(\mathbf{E}\,\mathrm{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right)\right)^{1/\ell}\right) = \mathbf{Pr}\left(\left\|\overline{A}_{\boldsymbol{G}}\right\|^{\ell} \geqslant e^{\varepsilon\ell}\,\mathbf{E}\,\mathrm{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right)\right) \leqslant \exp(-\varepsilon\ell).$$

Thus, our goal reduces to bounding the expectation of $\mathrm{tr}(\overline{A}_{\boldsymbol{G}}^{\ell})$ for a sufficiently large even ℓ; in particular, if we choose $\ell \gg \log n$, then since $\overline{A}_{\boldsymbol{G}}$ has n eigenvalues, $\mathrm{tr}(\overline{A}_{\boldsymbol{G}}^{\ell})^{1/\ell}$ is a good "soft-max" proxy for $\|\overline{A}_{\boldsymbol{G}}\|$, and we will obtain high-probability bounds.

We now explain why properties of random walks on $\mathbb{S}^{d-1}$ naturally arise when applying the trace method. Concretely, $\mathrm{tr}(\overline{A}_{\boldsymbol{G}}^{\ell})$ is a sum over products of entries of $\overline{A}_{\boldsymbol{G}}$ corresponding to closed walks of length ℓ in the complete graph K_n on n vertices:

$$\mathrm{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right) = \sum_{i_0,\dots,i_{\ell-1}\in[n]} \prod_{t=0}^{\ell-1} (\overline{A}_{\boldsymbol{G}})_{i_t i_{t+1 \bmod \ell}},$$

The walk $i_0, i_2, \dots, i_{\ell-1}, i_0$ can be represented as a directed graph. When we take the expectation, the symmetry of the distribution means that all sequences $i_0, \dots, i_{\ell-1}$ which result in the same graph (up to relabeling) give the same value. That is, letting $\mathcal{W}_\ell$ be the set of all such graphs, and for each $W \in \mathcal{W}_\ell$ letting N_W be the number of ways it can arise in the sum above,

$$\mathbf{E}\,\mathrm{tr}\left(\overline{A}_{\boldsymbol{G}}^{\ell}\right) = \sum_{W\in\mathcal{W}_\ell} N_W \cdot \mathbf{E} \prod_{(i,j)\in W} (\overline{A}_{\boldsymbol{G}})_{ij}. \tag{3.2}$$

To bound this sum, we must bound the expectation contributed by each $W \in \mathcal{W}_\ell$. For the sake of this overview we will consider only the case when $W = C_\ell$, the cycle on ℓ vertices, as it requires less accounting than the other cases; however it is reasonable to restrict our attention to this case for now, as bounding it already demonstrates our main ideas, and because this term roughly dominates the sum with $N_{C_\ell} \gg N_{W'}$ for all other $W' \in \mathcal{W}_\ell$ at $\ell = \operatorname{poly}\log n$ and $pn \gg \operatorname{poly}\log n$.[4]

We now bound the expectation for the case $W = C_\ell$; readers uninterested in the finer details may skip to the conclusion in (3.4). We expand the product using that $(\overline{A}_{\boldsymbol{G}})_{ij} = \boldsymbol{A}_{ij} - p$ (since $\mathbf{E}[\boldsymbol{A}_{ij}] = p$):

$$\begin{aligned}\mathbf{E}\prod_{i=1}^{\ell}(\boldsymbol{A}_{i,i+1} - p) &= \sum_{T\subseteq[\ell]}(-p)^{\ell-|T|}\,\mathbf{E}\prod_{i\in T}\boldsymbol{A}_{i,i+1}\\ &= \sum_{T\subseteq[\ell]}(-p)^{\ell-|T|}\,\mathbf{Pr}[\{(i,i+1) : i\in T\} \text{ is subgraph of } \boldsymbol{G}]. \qquad (3.3)\end{aligned}$$

and thus our focus is to understand subgraph probabilities in a random geometric graph. It is not too hard to see that when the edges specified by T form a forest, its subgraph probability is $p^{|T|}$, identical to its counterpart in an Erdős–Rényi graph; the nontrivial correlations introduced by the geometry only play a role when T has cycles. Hence, the sum (3.3) simplifies,

$$\begin{aligned}\mathbf{E}\prod_{i=1}^{\ell}(\overline{A}_{\boldsymbol{G}})_{i,i+1} &= \sum_{T\subsetneq[\ell]}(-p)^{\ell-|T|}p^{|T|} + \mathbf{Pr}[C_\ell \text{ is subgraph of } \boldsymbol{G}]\\ &= \mathbf{Pr}[C_\ell \text{ is subgraph of } \boldsymbol{G}] - p^\ell, \qquad (3.4)\end{aligned}$$

where we used that the binomial sum is equal to $(p-p)^\ell = 0$.

Hence it remains to estimate the subgraph probability of a length-ℓ cycle. We will now see how subgraph probabilities are related to the mixing rate of a random walk on $\mathbb{S}^{d-1}$.

Subgraph probability of a cycle in a random geometric graph. For the cycle $C_\ell = 0, 1, \ldots, \ell-1, 0$, by Bayes' rule:

$$\begin{aligned}\mathbf{Pr}[C_\ell \in \boldsymbol{G}] &= \prod_{i=0}^{\ell-1}\mathbf{Pr}[(i,i+1)\in\boldsymbol{G} \mid \forall j<i, (j,j+1)\in\boldsymbol{G}]\\ &= p^{\ell-1}\cdot\mathbf{Pr}[(\ell-1,0)\in\boldsymbol{G} \mid 0,1,\ldots\ell-1\in\boldsymbol{G}],\end{aligned}$$

since in all but the step $i+1=\ell$, the graph in question is a forest. Identifying each i with a point $\boldsymbol{x}_i$ on $\mathbb{S}^{d-1}$, for any choice of $\boldsymbol{x}_0$ the above probability can equivalently be written as

$$p^{\ell-1}\cdot\mathbf{Pr}[\langle\boldsymbol{x}_{\ell-1},\boldsymbol{x}_0\rangle \geqslant \tau \mid \langle\boldsymbol{x}_i,\boldsymbol{x}_{i+1}\rangle \geqslant \tau : 0\leqslant i\leqslant \ell-2].$$

Denoting with P the transition kernel of the random walk we alluded to earlier, where in one step we walk from a point x to a uniformly random point in $\operatorname{cap}_p(x)$, we can write the

[4] Briefly, this is because whenever $i_0, \ldots, i_{\ell-1}$ are all distinct elements of $[n]$, the resulting walk's graph is a cycle, and when $\ell = \operatorname{poly}\log n$, ℓ indices sampled at random from $[n]$ are all distinct with high probability.

distribution of $\boldsymbol{x}_\ell \mid \{\boldsymbol{x}_0, \langle \boldsymbol{x}_i, \boldsymbol{x}_{i+1}\rangle \geqslant \tau : 0 \leqslant i \leqslant \ell-2\}$ as $P^{\ell-1}\delta_{\boldsymbol{x}_0}$ where $\delta_{\boldsymbol{x}_0}$ refers to the point mass probability distribution supported at $\boldsymbol{x}_0$. In turn, we can write the subgraph probability as:

$$p^{\ell-1} \cdot \mathop{\mathbf{Pr}}_{\boldsymbol{x}_{\ell-1} \sim P^{\ell-1}\delta_{\boldsymbol{x}_0}} \left[\boldsymbol{x}_{\ell-1} \in \mathrm{cap}_p(\boldsymbol{x}_0)\right].$$

If $\boldsymbol{x}_{\ell-1}$ were sampled from the uniform distribution ρ on $\mathbb{S}^{d-1}$ then the probability of landing in $\mathrm{cap}_p(\boldsymbol{x}_0)$ would be p, which lets us upper bound the subgraph probability by:

$$p^{\ell-1} \cdot \left(p + \mathrm{d_{TV}}\left(P^{\ell-1}\delta_{\boldsymbol{x}_0}, \rho\right)\right).$$

The terms for more complicated subgraphs $W' \in \mathcal{W}_\ell$ also similarly depend on the mixing properties of P via subgraph probabilities. Our next goal then is to understand the mixing properties of P.

Remark 3.1.15. To prove Theorem 3.1.11 about random restrictions, the same strategy is used to relate subgraph probabilities with mixing rate of the random walk on the original graph we start with.

Mixing properties of P. We show that the walk over $\mathbb{S}^{d-1}$ with transition kernel P contracts the TV distance by coupling this discrete walk with the continuous Brownian motion U_t over $\mathbb{S}^{d-1}$. Then via a known log-Sobolev inequality for Brownian motion on spheres, we can prove the following contraction property for P.

Theorem 3.1.16 (Informal version of Theorem 3.4.6)**.** *For any probability measure α over $\mathbb{S}^{d-1}$ and integer $k \geqslant 0$,*

$$\mathrm{d_{TV}}\left(P_p^k \alpha, \rho\right) \leqslant \left((1 + o_{d\tau^2}(1)) \cdot \tau\right)^{k-1} \cdot \sqrt{\frac{1}{2}\log\frac{1}{p}},$$

where P_p denotes the transition kernel in which every $x \in \mathbb{S}^{d-1}$ walks to a uniformly random point in the measure-p cap around it and $o_{d\tau^2}(1)$ denotes a function that goes to 0 as $d\tau^2 \to \infty$.

We leave the details to Section 3.4, but in brief, the reason we are able to execute this coupling is that the probability mass in $P\delta_{\boldsymbol{x}_0}$ concentrates around $\mathrm{shell}_{=\tau}(\boldsymbol{x}_0)$, and most of the $(\frac{1}{d-1}\log\frac{1}{\tau})$-step Brownian motion starting from $\boldsymbol{x}_0$ concentrates at $\mathrm{shell}_{=\tau}(\boldsymbol{x}_0)$, so when $t = \frac{1}{d-1}\log\frac{1}{\tau}$ the operators P and U_t have similar action.

We can now apply Theorem 3.1.16 to bound $\mathrm{d_{TV}}\left(P^{\ell-1}\delta_{\boldsymbol{x}_0}, \rho\right)$ with $\alpha = \delta_{\boldsymbol{x}_0}$ and $k = \ell-1$:

$$\mathrm{d_{TV}}\left(P^{\ell-1}\delta_{\boldsymbol{x}_0}, \rho\right) \leqslant \left((1 + o(1))\tau\right)^{\ell-2}\sqrt{\tfrac{1}{2}\log\tfrac{1}{p}}.$$

Spectral norm of random geometric graph. We now return to bounding the expected trace of $\overline{A}_{\boldsymbol{G}}^{\ell}$; putting together the above, we have the bound

$$\begin{aligned}\mathbf{E}\prod_{(i,j)\in C_\ell}(\overline{A}_{\boldsymbol{G}})_{ij} &\leqslant \mathbf{Pr}[C_\ell \in \boldsymbol{G}] - p^\ell \\ &\leqslant p^{\ell-1}\left(p + \mathrm{d}_{\mathrm{TV}}\left(P^{\ell-1}\delta_{\boldsymbol{x}_0}, \rho\right)\right) - p^\ell \\ &\leqslant p^{\ell-1}((1+o(1))\tau)^{\ell-2}\sqrt{\tfrac{1}{2}\log\tfrac{1}{p}}.\end{aligned}$$

The coefficient N_{C_ℓ} in front of the $W = C_\ell$ term in (3.2) is the number of sequences $i_1, \ldots, i_\ell \in [n]$ which yield an ℓ-cycle graph; this happens if and only if all of the indices are distinct, so $N_{C_\ell} = \ell! \cdot \binom{n}{\ell} \leqslant n^\ell$. Hence the contribution of the ℓ-cycle to the sum is at most $((1+o(1))np\tau)^{\ell-2} \cdot \mathrm{poly}(n)$ when $p > 1/n$. By a careful accounting similar to the above for all graphs $W \in \mathcal{W}_\ell$, one can show that in the parameter regime $pn \gg \mathrm{poly}\log(n)$ and $\ell = \mathrm{poly}\log n$, the term $W = C_\ell$ contains $(1-o(1))$ of the total value of this sum, so we obtain the bound

$$\left[\mathbf{E}\,\mathrm{tr}(\overline{A}_{\boldsymbol{G}}^{\ell})\right]^{1/\ell} \leqslant \left((1+o(1))\cdot((1+o(1))np\tau)^{\ell-2}\cdot\mathrm{poly}(n)\right)^{1/\ell} = (1+o(1))np\tau,$$

when we choose $\ell = \omega(\log n)$. Applying Markov's inequality we conclude that $\|\overline{A}_{\boldsymbol{G}}\| \leqslant (1+o(1))np\tau$ with high probability, and normalizing by the degrees (which concentrate well around np) we conclude our upper bound of τ in Theorem 3.1.8.

Adapting the spectral norm bound to links. Up until now, we have pretended that the link of i_w is a random geometric graph, where the vertices are identified with vectors in $\mathrm{shell}_{=\tau}(w)$, rather than $\mathrm{cap}_{\geqslant\tau}(w)$. While it is true that most of the probability mass in $\mathrm{cap}_{\geqslant\tau}(w)$ is close to the boundary, some $\frac{1}{\mathrm{poly}(m)}$-fraction of the vertices j in the link will have $\langle \boldsymbol{v}_j, w\rangle = \boldsymbol{\kappa}_j > (1+\delta)\tau$ for some $\delta > 0$. And within the link, these vertices will have higher expected degree: for $\boldsymbol{v}_i, \boldsymbol{v}_j$ having $\langle \boldsymbol{v}_i, w\rangle = \boldsymbol{\kappa}_i$ and $\langle \boldsymbol{v}_j, w\rangle = \boldsymbol{\kappa}_j$, following a similar calculation to the one above,

$$q_{ij} \coloneqq \mathbf{Pr}[i \sim j] = \mathop{\mathbf{Pr}}_{\boldsymbol{u}_i,\boldsymbol{u}_j\sim\mathbb{S}^{d-2}}\left[\langle \boldsymbol{u}_i, \boldsymbol{u}_j\rangle \geqslant \tfrac{\tau - \boldsymbol{\kappa}_i\boldsymbol{\kappa}_j}{\sqrt{(1-\boldsymbol{\kappa}_i^2)(1-\boldsymbol{\kappa}_j^2)}}\right] \tag{3.5}$$

And this quantity is $\gg q = \mathbf{Pr}_{\boldsymbol{u}_i,\boldsymbol{u}_j\sim\mathbb{S}^{d-2}}[\langle \boldsymbol{u}_i, \boldsymbol{u}_j\rangle \geqslant \frac{\tau}{1+\tau}]$ when $\boldsymbol{\kappa}_i > (1+\delta)\tau$ and $\boldsymbol{\kappa}_j \geqslant \tau$. Hence, vertex degrees are not as well concentrated within each link as they are (around pn) in the entire graph $\boldsymbol{H}$.

As a result, if we let $\boldsymbol{G}_w$ now stand for the link and $A_{\boldsymbol{G}_w}$ now stand for the adjacency matrix of the link, it is no longer the case that $\|A_{\boldsymbol{G}_w} - \mathbf{E}\,A_{\boldsymbol{G}_w}\|$ is small: $\mathbf{E}\,A_{\boldsymbol{G}_w}$ still has every entry equal to q, but the top eigenvector of $A_{\boldsymbol{G}}$ will not be close to the all-1 vector.

To contend with this, we analyze the spectral norm of $A_{\boldsymbol{G}}$ *conditioned* on the shells that the points in $\mathrm{cap}_p(w)$ are in. Letting $\boldsymbol{\kappa} \in [\tau, 1]^m$ be such that $\boldsymbol{\kappa}_i = \langle \boldsymbol{v}_i, w\rangle$, vertex degrees concentrate in $\boldsymbol{G}_w$ conditioned on $\boldsymbol{\kappa}$, and we can readily bound the spectral norm of $\overline{A}_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa} = A_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa} - \mathbf{E}[A_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa}]$.

The analysis of the spectral norm of $\overline{A}_{\boldsymbol{G}_w}$ is then not so different from that of $\overline{A}_{\boldsymbol{G}}$ for $\boldsymbol{G}$ a random geometric graph; the main difference is that now, instead of working with the walk P in which we walk from $\boldsymbol{u}_i$ to a random point in $\mathrm{cap}_{\geqslant\tau}(\boldsymbol{u}_i)$, at each step of the walk we must adjust the volume of the cap: when considering the probability that the edge i, j is present, we apply the operator $P_{q_{ij}}$ for $q_{ij}(\boldsymbol{\kappa}_i, \boldsymbol{\kappa}_j)$ as defined in (3.5), which walks from $\boldsymbol{u}_i$ to a random point in $\mathrm{cap}_{q_{ij}}(\boldsymbol{u}_i)$. This requires some additional accounting, but one can show that the slowest mixing occurs when $\boldsymbol{\kappa}_i = \boldsymbol{\kappa}_j = \tau$ and $q_{ij} = \frac{\tau}{1+\tau}$, from which we obtain the desired bound on $\|\overline{A}_{\boldsymbol{G}_w} \mid \kappa\|$. For details, see Section 3.5.

One additional complication is that $\mathbf{E}\, A_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa}$ is not a rank-1 matrix, so bounding $\|\overline{A}_{\boldsymbol{G}_w} \mid \kappa\|$ does not directly imply a bound on the second eigenvalue of $A_{\boldsymbol{G}_w}$. However, it turns out that $\mathbf{E}\, A_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa}$ is sufficiently close to a rank-1 matrix $R_{\boldsymbol{G}_w}$ (the matrix whose (i, j)th entry is the product of the expected degrees conditioned on $\boldsymbol{\kappa}$) that we can apply the triangle inequality:

$$\|(A_{\boldsymbol{G}_w} - R_{\boldsymbol{G}_w}) \mid \boldsymbol{\kappa}\| \leqslant \|A_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa} - \mathbf{E}[A_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa}]\| + \|\mathbf{E}[A_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa}] - R_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa}\|,$$

the first term we bound using the trace method as described above. The second term we bound via more-or-less direct calculation: because all but an $o(1)$ fraction of $\boldsymbol{\kappa}_i \approx \tau$, when ignoring an $o(1)$ fraction of rows and columns, the rows of $\mathbf{E}\, A_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa}$ are almost constant multiples of each other, and further these $o(1)$ fraction of rows and columns represent an $o(1)$ fraction of the total absolute value of $\mathbf{E}[A_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa}]$. (This is because the high-degree vertices in $\boldsymbol{G}_w$ represent an $o(1)$ fraction of the total edges in $\boldsymbol{G}_w$.) Now, thinking of $\mathbf{E}[A_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa}]$ as a transition operator of a Markov chain, we are able to use this to argue that the Markov chain mixes so rapidly that $\mathbf{E}[A_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa}]$ must be close to $R_{\boldsymbol{G}_w} \mid \boldsymbol{\kappa}$, yielding the desired bound. For details, see Section 3.6.

3.2 Preliminaries

Notation. For a self-adjoint matrix M, we denote its eigenvalues in decreasing order as $\lambda_1(M) \geqslant \ldots \geqslant \lambda_n(M)$, the absolute values of its eigenvalues as $|\lambda|_1(M) \geqslant \ldots \geqslant |\lambda|_n(M)$, and $\lambda_{\max}(M)$ and $|\lambda|_{\max}(M)$ to denote $\lambda_1(M)$ and $|\lambda|_1(M)$ respectively. Given a sequence of matrices $M_1, \ldots, M_T$ we use $\prod_{i=1}^T M_i$ to denote the matrix $M_T \cdot M_{T-1} \cdots M_1$.

For a graph G, we use $V(G)$ to refer to its vertex set and $E(G)$ to refer to its edge set. For a vertex $v \in V(G)$, we use $N(v)$ to denote the set of neighbors of v.

For a probability distribution $\mathcal{D}$, we use $\Phi_{\mathcal{D}}(x)$ to denote the CDF of $\mathcal{D}$ at x, and $\overline{\Phi}_{\mathcal{D}}(x) \coloneqq 1 - \Phi_{\mathcal{D}}(x)$ to denote the tail of $\mathcal{D}$ at x. For any point x, we use δ_x to denote the delta distribution at x .

3.2.1 Linear algebra

The following articulates how one gets a handle on the second eigenvalue of a matrix after subtracting a rank-1 term, which will be used in Section 3.3 and Section 3.5.

Fact 3.2.1. *For any $n \times n$ symmetric matrix M and rank-1 PSD matrix R, $|\lambda|_2(M) \leqslant \|M - R\|$.*

Proof. By Cauchy's interlacing theorem, $\lambda_2(M) \leqslant \lambda_1(M - R) \leqslant \|M - R\|$ and $-\lambda_n(M) \leqslant -\lambda_n(M - R) \leqslant \|M - R\|$. The desired inequality is then true since

$$|\lambda|_2(M) \leqslant \max\{\lambda_2(M), -\lambda_n(M)\}.$$

□

Establishing second eigenvalue bounds in Section 3.3 and Section 3.5 also involves bounding the spectral norm of some matrices via the "trace method" articulated below.

Claim 3.2.2 (Trace Method)**.** Let $\boldsymbol{M}$ be a symmetric (random) matrix. Then for any even integer $\ell \geqslant 0$,

$$\mathbf{Pr}\left[\|\boldsymbol{M}\| \geqslant e^{\varepsilon} \cdot \mathbf{E}\left[\operatorname{tr}\left((\boldsymbol{M})^{\ell}\right)\right]^{1/\ell}\right] \leqslant \exp(-\varepsilon\ell).$$

Proof. By Markov's inequality, $\mathbf{Pr}\left[\|\boldsymbol{M}\| \geqslant t\right] \leqslant t^{-\ell}\,\mathbf{E}\left(\|\boldsymbol{M}\|^{\ell}\right)$. The claim then follows because for any self-adjoint matrix M, $\lambda_{\max}\left(M^{\ell}\right) \leqslant \operatorname{tr}(M^{\ell})$ when ℓ is even. □

We will also require the following bound on the spectrum of a matrix, which is a special case of the Gershgorin circle theorem.

Claim 3.2.3 (Row sum bound)**.** For any matrix M, $|\lambda|_{\max}(M) \leqslant \max_i \|M[i,*]\|_1$.

Proof. Let v be the eigenvector achieving $\lambda = |\lambda|_{\max}(M)$. Then letting k be the index maximizing $|v_k|$, we have

$$|\lambda v_k| = |(Mv)_k| = \left|\sum_j M_{kj} v_j\right| \leqslant |v_k| \sum_j |M_{kj}| \leqslant |v_k| \max_i \|M[i,*]\|_1,$$

and dividing through by $|v_k|$ gives the conclusion. □

3.2.2 Probability

Definition 3.2.4. The *total variation distance* between probability distributions μ and ν is defined as:

$$\mathrm{d}_{\mathrm{TV}}(\mu, \nu) := \max_{\mathcal{E}} |\mu(\mathcal{E}) - \nu(\mathcal{E})|.$$

Fact 3.2.5. *When ρ is a nonnegative measure such that μ and ν are absolutely continuous with respect to ρ, then:*

$$\mathrm{d_{TV}}(\mu,\nu) = \frac{1}{2}\int\left|\frac{d\mu}{d\rho}(x) - \frac{d\nu}{d\rho}(x)\right| d\rho(x) = \int\left(\frac{d\mu}{d\rho}(x) - \frac{d\nu}{d\rho}(x)\right)\cdot \mathbf{1}\left[\frac{d\mu}{d\rho}(x) > \frac{d\nu}{d\rho}(x)\right] d\rho(x).$$

When μ and ν are supported on $[n]$, then:

$$\mathrm{d_{TV}}(\mu,\nu) = \frac{1}{2}\|\mu-\nu\|_1 = \sum_{i=1}^{n}(\mu(i)-\nu(i))\cdot \mathbf{1}[\mu(i)>\nu(i)]$$

where μ and ν are the vectors of probabilities.

We describe a Markov chain via its transition operator P where $P(i,j)$ denotes the probability of transitioning from state i to state j.

We call the joint distribution $\omega(\mu,\nu)$ a coupling between two distributions μ and ν if $\mu = \omega(\cdot,\nu)$ and $\mu = \omega(\mu,\cdot)$. In other words, the marginals of ω correspond to μ and ν.

Fact 3.2.6. *Let x and y be two arbitrary states in a Markov chain over state space Ω with transition operator P, and sample $X \sim P(x,\cdot)$ and $Y \sim P(y,\cdot)$, where $P(z,\cdot)$ denotes the distribution over Ω given by a single step of the walk starting from state z. Then, there exists a coupling of X and Y such that $X = Y$ with probability $1-\varepsilon$ if and only if $\mathrm{d_{TV}}(P(x,\cdot),P(y,\cdot)) \leqslant \varepsilon$.*

3.2.3 The uniform distribution over the unit sphere

We use ρ to denote the uniform distribution on $\mathbb{S}^{d-1}$.

Let $v \in \mathbb{S}^{d-1}$ and $\boldsymbol{w} \sim \rho$. Then the distribution $\mathsf{D_{ip}}(d)$ of $\langle \boldsymbol{w}, v\rangle$ is invariant under the choice of v, is supported on $[-1,1]$ and has probability density function:

$$\psi_d(x) = \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-1}{2}\right)\sqrt{\pi}}\cdot\left(1-x^2\right)^{(d-3)/2}.$$

Henceforth, we use Z_d to denote the normalizing constant $\frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-1}{2}\right)\sqrt{\pi}}$.

Fact 3.2.7. *$Z_d \leqslant O(\sqrt{d})$.*

In addition, we will rely heavily on the following sharp estimate of the tail of $\mathsf{D_{ip}}(d)$.

Lemma 3.2.8. *Let $\Phi_{\mathsf{D_{ip}}(d)}(t) \coloneqq \mathbf{Pr}_{X\sim \mathsf{D_{ip}}(d)}[X \geqslant t]$. Then, when $t \geqslant 0$:*

$$\frac{Z_d}{t(d-1)}\cdot\left(1-t^2\right)^{(d-1)/2}\cdot\left(1-\frac{4\log(1+d\cdot t^2)}{d\cdot t^2}\right) \leqslant \overline{\Phi}_{\mathsf{D_{ip}}(d)}(t) \leqslant \frac{Z_d}{t(d-1)}\cdot\left(1-t^2\right)^{(d-1)/2}.$$

Proof. It suffices to upper and lower bound $\int_t^1 (1-x^2)^{(d-3)/2}$. We first obtain an upper bound.

$$\begin{aligned}\int_t^1 \left(1-x^2\right)^{(d-3)/2} dx &= \frac{1}{t}\int_t^1 t\left(1-x^2\right)^{(d-3)/2} dx \\ &\leqslant \frac{1}{t}\int_t^1 x\left(1-x^2\right)^{(d-3)/2} dx \\ &= -\frac{1}{t(d-1)}\cdot\left(1-x^2\right)^{(d-1)/2}\Bigg|_t^1 \\ &= \frac{1}{t(d-1)}\cdot\left(1-t^2\right)^{(d-1)/2}\end{aligned}$$

Now we prove the lower bound. For any $\varepsilon > 0$ such that $t\cdot\sqrt{1-\varepsilon+\frac{\varepsilon}{t^2}} \leqslant 1$, and defining $\delta := \frac{\varepsilon}{t^2} - \varepsilon$, we have the following.

$$\begin{aligned}\int_t^1 (1-x^2)^{(d-3)/2} dx &\geqslant \frac{1}{t\sqrt{1+\delta}}\int_t^{t\sqrt{1+\delta}} \left(t\sqrt{1+\delta}\right)\left(1-x^2\right)^{(d-3)/2} dx \\ &\geqslant \frac{1-\delta}{t}\int_t^{t\sqrt{1+\delta}} x\left(1-x^2\right)^{(d-3)/2} dx \\ &= -\frac{1-\delta}{t(d-1)}\cdot\left(1-x^2\right)^{(d-1)/2}\Bigg|_t^{t\sqrt{1+\delta}} \\ &= \frac{1-\delta}{t(d-1)}\cdot\left(1-t^2\right)^{(d-1)/2}\cdot\left(1-(1-\varepsilon)^{(d-1)/2}\right)\end{aligned}$$

where the second inequality uses $\frac{1}{\sqrt{1+\delta}} \geqslant 1-\delta$ and the last equality uses $1-t^2(1+\delta) = (1-t^2)(1-\varepsilon)$. Choosing $\varepsilon = \frac{2\log\left(1+dt^2\right)}{d-1}$ yields:

$$\int_t^1 \left(1-x^2\right)^{(d-3)/2} \geqslant \frac{1}{t(d-1)}\cdot\left(1-t^2\right)^{(d-1)/2}\cdot\left(1-\frac{4\log(1+dt^2)}{dt^2}\right). \qquad \square$$

We use $\mathsf{D}_{\mathsf{ip}}(d)|_{\geqslant\tau}$ to represent $\mathsf{D}_{\mathsf{ip}}(d)$ conditioned on lying in $[\tau, 1]$.

Definition 3.2.9. For a vector y, we use $\mathrm{cap}_p(y)$ and $\mathrm{cap}_{\geqslant\tau(p)}(y)$ interchangeably to denote the measure-p spherical cap around y:

$$\mathrm{cap}_p(y) = \mathrm{cap}_{\geqslant\tau(p)}(y) := \left\{u : \langle u, y\rangle \geqslant \tau(p), u \in \mathbb{S}^{d-1}\right\}.$$

We use $\underline{\mathrm{cap}}_p(y)$ and $\underline{\mathrm{cap}}_{\geqslant\tau(p)}(y)$ to denote the uniform measure over the set $\mathrm{cap}_p(y)$. We denote the boundary of $\mathrm{cap}_p(y)$ by $\mathrm{shell}_p(y)$ or $\mathrm{shell}_{=\tau(p)}(y)$. That is,

$$\mathrm{shell}_p(y) := \left\{u : \langle u, y\rangle = \tau(p), u \in \mathbb{S}^{d-1}\right\}.$$

3.3 The second eigenvalue of random restrictions

In this section we prove Theorem 3.1.11. Let X be a (possibly infinite) vertex-transitive graph with a unique stationary measure ρ. Let $\boldsymbol{G} \sim \mathsf{RR}_n(X)$ be a random restriction of X as defined in Definition 3.1.10, and let $p = \mathbf{Pr}_{\boldsymbol{G}\sim\mathsf{RR}_n(X)}[(i,j) \in E(\boldsymbol{G})]$ be the marginal edge probability in $\boldsymbol{G}$. Suppose furthermore that

$$\exists\, C, \lambda \text{ with } C \geqslant 1 \text{ and } \frac{1}{\sqrt{pn}} \leqslant \lambda \leqslant 1 \quad \text{s.t. for any distribution } \alpha \text{ on } V(X),$$
$$\mathrm{d_{TV}}\left(X^k\alpha, \rho\right) \leqslant C\lambda^k. \quad (3.6)$$

We overload notation and use X to denote the transition operator for the simple random walk on X, and for $H \subseteq V(X)$ we also use H to denote the indicator vector of the set H.

We denote its adjacency matrix by $A_{\boldsymbol{G}}$, the diagonal degree matrix by $D_{\boldsymbol{G}}$, the centered adjacency matrix by $\overline{A}_{\boldsymbol{G}} = A_{\boldsymbol{G}} - \mathbf{E}\,A_{\boldsymbol{G}}$, and the normalized adjacency matrix by $\widehat{A}_{\boldsymbol{G}} = D_{\boldsymbol{G}}^{-1/2}A_{\boldsymbol{G}}D_{\boldsymbol{G}}^{-1/2}$. Then we'll show the following.

Theorem 3.3.1. *As long as $pn \gg C^6\log^8 n$, for any constant $\gamma > 0$, with probability at least $1 - n^{-\gamma}$,*

$$|\lambda|_2\left(\widehat{A}_{\boldsymbol{G}}\right) \leqslant (1 + o(1)) \cdot \max\left(\lambda, \frac{\log^4 n}{\sqrt{pn}}\right).$$

Proof. By Fact 3.2.1, for any rank-1 PSD matrix R, $|\lambda|_2\left(\widehat{A}_{\boldsymbol{G}}\right) \leqslant \left\|\widehat{A}_{\boldsymbol{G}} - R\right\|$. Thus we turn our attention to bounding $\left\|\widehat{A}_{\boldsymbol{G}} - R\right\|$ for appropriately chosen R. Setting $R_{\boldsymbol{G}} = pD_{\boldsymbol{G}}^{-1/2}JD_{\boldsymbol{G}}^{-1/2}$ where J is the all-ones matrix and using submultiplicativity of the operator norm, we see:

$$\left\|\widehat{A}_{\boldsymbol{G}} - R_{\boldsymbol{G}}\right\| \leqslant \left\|D_{\boldsymbol{G}}^{-1/2}\right\|^2 \cdot \|A_{\boldsymbol{G}} - pJ\|.$$

Now, observe that $\left\|D_{\boldsymbol{G}}^{-1/2}\right\|^2 = \left\|D_{\boldsymbol{G}}^{-1}\right\|$. To bound this quantity, we'll use the concentration of the vertex degrees (the entries of the diagonal of $D_{\boldsymbol{G}}$). For every vertex, the marginal distribution of the degree is $\mathsf{Binom}(n,p)$. So by Hoeffding's inequality and the union bound, when $pn \gg \log^8 n$, for any fixed $\gamma > 0$, $|(D_{\boldsymbol{G}})_{ii} - pn| \leqslant \sqrt{pn\log^2 n}$ for all $i \in [n]$ with probability at least $1 - n^{\gamma}$. So with probability at least $1 - n^{-\gamma}$, $D_{\boldsymbol{G}}^{-1} = \frac{1}{pn}I + \Delta$ for Δ a diagonal matrix with entries with absolute value of order $\sqrt{\log^2 n/(pn)^3}$. Thus, $\left\|D_{\boldsymbol{G}}^{-1}\right\| \leqslant \frac{1}{pn} \cdot \left(1 + \frac{\log n}{\sqrt{pn}}\right)$.

Next, $\|A_{\boldsymbol{G}} - pJ\| \leqslant \left\|\overline{A}_{\boldsymbol{G}}\right\| + p$, where recall $\overline{A}_{\boldsymbol{G}} = A_{\boldsymbol{G}} - \mathbf{E}\,A_{\boldsymbol{G}}$. We will show:

$$\left\|\overline{A}_{\boldsymbol{G}}\right\| \leqslant (1 + o(1)) \cdot \max\left\{\lambda pn, \sqrt{pn}\log^4 n\right\}.$$

Putting these bounds together gives:

$$|\lambda|_2\left(\widehat{A}_{\boldsymbol{G}}\right) \leqslant (1+o(1)) \cdot \max\left\{\lambda, \frac{\log^4 n}{\sqrt{pn}}\right\}.$$

Finally, we devote the rest of the proof to bounding $\left\|\overline{A}_{\boldsymbol{G}}\right\|$. By Claim 3.2.2, it suffices to bound $\mathbf{E}\,\mathrm{tr}((\overline{A}_{\boldsymbol{G}})^\ell)$ for a large enough even ℓ.

For an $n \times n$ matrix M, $\mathrm{tr}(M^\ell)$ can be written as a sum over length-ℓ closed walks on the complete graph $\mathcal{K}_n$, with each walk W weighted according to $\prod_{(i,j)\in W} M_{ij}$. The exchangeability of entries in $\overline{A}_{\boldsymbol{G}}$ means that the walks can be partitioned into equivalence classes based on their topology as graphs, where the members of each class contribute identically to the summation.

Definition 3.3.2. We use $\mathcal{W}_\ell$ to denote the collection of length-ℓ walks in $\mathcal{K}_n$, the complete graph on n vertices. For $W \in \mathcal{W}_\ell$, we use $G(W) = (V(W), E(W))$ to denote the simple graph induced by edges walked on in W. We let the *multiplicity* of e in W, $m(e)$, be the number of times e occurs in W.

We can then write:

$$\mathbf{E}\,\mathrm{tr}\left((A_{\boldsymbol{G}} - \mathbf{E}\,A_{\boldsymbol{G}})^\ell\right) = \sum_{W\in\mathcal{W}_\ell} \mathbf{E} \prod_{e\in E(W)} (\mathbf{1}[e\in\boldsymbol{G}] - p)^{m(e)} \tag{3.7}$$

We now focus on understanding each term of the above summand in terms of the properties of $G(W)$. Our first step is to handle leaves.

Definition 3.3.3. We use $G_2(W) = (V_2(W), E_2(W))$ to denote the 2-*core* of $G(W)$, the graph obtained by recursively deleting degree-1 vertices from $G(W)$. We denote the graph induced on the edges deleted in this process as $G_1(W)$.

Observation 3.3.4. We have $G(W) = G_1(W) \cup G_2(W)$. Further, every vertex in $G_2(W)$ has degree at least 2, and $G_1(W)$ is a forest where each connected component has at most one vertex in $G_2(W)$.

Notice that if F is a forest, then $\mathbf{Pr}[F \in \boldsymbol{G}] = p^{|E(F)|}$, and further if F is a forest sharing at most one vertex with a graph H, then the events $\{H \in \boldsymbol{G}\}$ and $\{F \in \boldsymbol{G}\}$ are independent. Hence, with the above decomposition in hand, we can "peel off" the one-core and for any $W \in \mathcal{W}_\ell$ we can write:

$$\begin{aligned}
&(3.7)\\
&= \mathop{\mathbf{E}}_{\substack{\boldsymbol{u}_i\\ i\in V_2(W)}} \mathop{\mathbf{E}}_{\substack{\boldsymbol{u}_j\\ j\in V_1(W)\setminus V_2(W)}} \prod_{e\in E(W)} (\mathbf{1}[e\in\boldsymbol{G}] - p)^{m(e)}\\
&= \prod_{e\in E_1(W)} \mathbf{E}\Big((\mathbf{1}[e\in\boldsymbol{G}] - p)^{m(e)}\Big) \cdot \mathop{\mathbf{E}}_{\substack{\boldsymbol{u}_i\\ i\in V_2(W)}} \prod_{e\in E_2(W)} (\mathbf{1}[e\in\boldsymbol{G}] - p)^{m(e)}
\end{aligned}$$

$$= \prod_{e\in E_1(W)} \mathbf{E}\big(\mathbf{1}[e\in \boldsymbol{G}]\big((1-p)^{m(e)} - (-p)^{m(e)}\big) + (-p)^{m(e)}\big) \cdot \mathop{\mathbf{E}}_{\substack{\boldsymbol{u}_i \\ i\in V_2(W)}} \prod_{e\in E_2(W)} (\mathbf{1}[e\in\boldsymbol{G}] - p)^{m(e)}$$

$$\leqslant \left| \prod_{e\in E_1(W)} \big(p(1-p)^{m(e)} + (1-p)(-p)^{m(e)}\big) \right| \cdot \left| \mathop{\mathbf{E}}_{\boldsymbol{u}_i : i\in V_2(W)} \prod_{e\in E_2(W)} (\mathbf{1}[e\in\boldsymbol{G}] - p)^{m(e)} \right|, \tag{3.8}$$

where in the third line we've used that $(\mathbf{1}[e\in\boldsymbol{G}]-p)^k = \mathbf{1}[e\in\boldsymbol{G}]((1-p)^k-(-p)^k)+(-p)^k$.

It now remains to handle the 2-core $G_2(W)$. To simplify the expression, we'll exploit the following fact: if J is a subset of vertices in $G_2(W)$, conditional on an assignment of $\boldsymbol{u}_i$ for all $i\in J$, the existence of edges in regions of $G_2(W)$ separated by J are independent. We'll take advantage of this fact by splitting $G_2(W)$ into regions separated by the set of vertices in $G_2(W)$ of degree at least 3, leaving us to bound a collection of paths and cycles.

Definition 3.3.5 (Junction vertices)**.** We use $J(W)$ to denote the set of *junction vertices* of $G_2(W)$, which are vertices with degree-$\geqslant 3$ in $G_2(W)$, or in the case that $G_2(W)$ only has vertices of degree-2, we choose an arbitrary vertex ot $G_2(W)$ and add it to $J(W)$. We use $G_J(W) = (J(W), E_J(W))$ to denote the *junction graph* of $G_2(W)$, which is a multigraph obtained by starting with $G_2(W)$ and contracting to an edge all walks $\gamma = u_0 \dots u_t$ satisfying the following conditions:

1. u_0 and u_t are (possibly identical) junction vertices,
2. $u_1,\dots,u_{t-1}$ are distinct vertices with degree-2 in $G_2(W)$.

For an edge $f\in E_J(W)$, we use $\gamma(f) = u_0,\dots,u_t$ to identify the walk from which f arose in $G_2(W)$, $s(f)$ to denote the "start" vertex u_0 of $\gamma(f)$, and $t(f)$ to denote the "terminal" vertex u_t of $\gamma(f)$.

Then we can bound the contribution of the 2-core in terms of the contribution of the walk $\gamma(f)$ corresponding to each edge f in the junction graph:

$$\begin{aligned}
\left| \mathop{\mathbf{E}}_{\substack{\boldsymbol{u}_i \\ i\in V_2(W)}} \prod_{e\in E_2(W)} (\mathbf{1}[e\in\boldsymbol{G}]-p)^{m(e)} \right| &= \left| \mathop{\mathbf{E}}_{\boldsymbol{u}_i : i\in J(W)} \mathop{\mathbf{E}}_{\boldsymbol{u}_i : i\notin J(W)} \prod_{e\in E_2(W)} (\mathbf{1}[e\in\boldsymbol{G}]-p)^{m(e)} \right| \\
&= \left| \mathop{\mathbf{E}}_{\boldsymbol{u}_i : i\in J(W)} \prod_{f\in E_{J(W)}} \mathop{\mathbf{E}}_{\boldsymbol{u}_i : i\in\gamma(f)\setminus J(W)} \prod_{e\in\gamma(f)} (\mathbf{1}[e\in\boldsymbol{G}]-p)^{m(e)} \right| \\
&\leqslant \mathop{\mathbf{E}}_{\boldsymbol{u}_i : i\in J(W)} \prod_{f\in E_{J(W)}} \left| \mathop{\mathbf{E}}_{\boldsymbol{u}_i : i\in\gamma(f)\setminus J(W)} \prod_{e\in\gamma(f)} (\mathbf{1}[e\in\boldsymbol{G}]-p)^{m(e)} \right|.
\end{aligned} \tag{3.9}$$

We now focus on understanding the innermost expected value, the expectation over the internal vertices along a path, conditioned on the endpoints. Again using $(\mathbf{1}[e\in\boldsymbol{G}]-p)^k =$

$\mathbf{1}[e \in \boldsymbol{G}]((1-p)^k - (-p)^k) + (-p)^k$,

$$\begin{aligned}
&\left| \mathop{\mathbf{E}}_{\boldsymbol{u}_i : i \in \gamma(f) \setminus J(W)} \prod_{e \in \gamma(f)} (\mathbf{1}[e \in \boldsymbol{G}] - p)^{m(e)} \right| \\
&\quad = \left| \mathop{\mathbf{E}}_{\boldsymbol{u}_i : i \in \gamma(f) \setminus J(W)} \prod_{e \in \gamma(f)} \left(\mathbf{1}[e \in \boldsymbol{G}] \cdot \left((1-p)^{m(e)} - (-p)^{m(e)}\right) + (-p)^{m(e)} \right) \right|, \\
&\quad = \left| \sum_{T \subseteq \gamma(f)} \mathop{\mathbf{E}}_{\boldsymbol{u}_i : i \in \gamma(f) \setminus J(W)} \prod_{e \in T} \mathbf{1}[e \in \boldsymbol{G}] \cdot \left((1-p)^{m(e)} - (-p)^{m(e)}\right) \prod_{e \in \gamma(f) \setminus T} (-p)^{m(e)} \right|
\end{aligned}$$

Now, using the independence of edges in a forest, we can bound terms where $T \neq \gamma(f)$ simply, and the term $T = \gamma(f)$ in terms of the probability that a $|\gamma(f)|$-length walk in X starting at $\boldsymbol{u}_{s(f)}$ ends at $\boldsymbol{u}_{t(f)}$ (which is where properties of the random walk in X will enter into the bound):

$$\begin{aligned}
&= \Bigg| \sum_{\substack{T \subseteq \gamma(f) \\ T \neq \gamma(f)}} \prod_{e \in T} p \cdot \left((1-p)^{m(e)} - (-p)^{m(e)}\right) \cdot \prod_{e \in \gamma(f) \setminus T} (-p)^{m(e)} \\
&\qquad + \prod_{e \in \gamma(f)} \left((1-p)^{m(e)} - (-p)^{m(e)}\right) \cdot p^{|\gamma(f)|-1} \cdot \left\langle N(\boldsymbol{u}_{s(f)}), X^{|\gamma(f)|-1} \delta_{\boldsymbol{u}_{t(f)}} \right\rangle \Bigg|,
\end{aligned}$$

where $N(\boldsymbol{u}_{s(f)})$ is the neighborhood of $\boldsymbol{u}_{s(f)}$ in X, and $\delta_{\boldsymbol{u}_{t(f)}}$ is the point mass at $\boldsymbol{u}_{t(f)}$. Now adding and subtracting $\prod_{e \in \gamma(f)} \left(\left((1-p)^{m(e)} - (-p)^{m(e)}\right) + (-p)^{m(e)}\right) \cdot p^{|\gamma(f)|}$, we complete the first summation and from the triangle inequality we obtain the bound

$$\begin{aligned}
&\leqslant \left| \prod_{e \in \gamma(f)} \left(p(1-p)^{m(e)} + (1-p)(-p)^{m(e)}\right) \right| \\
&\qquad + \left| \prod_{e \in \gamma(f)} \left((1-p)^{m(e)} - (-p)^{m(e)}\right) \cdot p^{|\gamma(f)|-1} \cdot \left(\left\langle N(\boldsymbol{u}_{s(f)}), X^{|\gamma(f)|-1} \delta_{\boldsymbol{u}_{t(f)}} \right\rangle - p \right) \right|.
\end{aligned} \tag{3.10}$$

We bound (3.10) based on the graphical properties of $\gamma(f)$.

Definition 3.3.6. We say an edge e is a *singleton edge* if $m(e) = 1$ and a *duplicative edge* otherwise.

If $\gamma(f)$ contains any singleton edges, then the first term of (3.10) is 0; otherwise it is bounded by

$$\prod_{e \in \gamma(f)} (p(1-p)^2 + (1-p)p^2) \leqslant \prod_{e \in \gamma(f)} p(1-p) \leqslant p^{|\gamma(f)|}.$$

The second term can always be bounded by

$$\prod_{e\in\gamma(f)}\left((1-p)^{m(e)}+p^{m(e)}\right)\cdot p^{|\gamma(f)|-1}\cdot\left|\left\langle N(\boldsymbol{u}_{s(f)}),X^{|\gamma(f)|-1}\delta_{\boldsymbol{u}_{t(f)}}\right\rangle-p\right|\leqslant p^{|\gamma(f)|-1}\cdot\left|\left\langle N(\boldsymbol{u}_{s(f)}),X^{|\gamma(f)|-1}\delta_{\boldsymbol{u}_{t(f)}}\right\rangle-p\right|.$$

Using $D_J(W)$ to denote the collection of edges f in G_J such that $\gamma(f)$ contains no singleton edges, and $S_J(W)$ to use the collection of edges f in G_J such that $\gamma(f)$ contains a singleton edge, and plugging the above bounds into (3.9) tells us:

$$(3.9)\leqslant\mathop{\mathbf{E}}_{\substack{\boldsymbol{u}_i\\ i\in J(W)}}\prod_{f\in D_J(W)}p^{|\gamma(f)|-1}\cdot\left(\left|\left\langle N(\boldsymbol{u}_{s(f)}),X^{|\gamma(f)|-1}\delta_{\boldsymbol{u}_{t(f)}}\right\rangle-p\right|+p\right)\cdot \prod_{f\in S_J(W)}p^{|\gamma(f)|-1}\cdot\left|\left\langle N(\boldsymbol{u}_{s(f)}),X^{|\gamma(f)|-1}\delta_{\boldsymbol{u}_{t(f)}}\right\rangle-p\right|.$$

If $G_J(W)$ were a tree, we could recursively take the expectation over leaf vertices to bound the quantity above, as we did to get rid of G_1. However, it is not a tree, so we'll pick an arbitrary spanning tree $T_J(W)$ of $G_J(W)$, and bound edges outside of the spanning tree directly. For $f\in E_J(W)\setminus T_J(W)$, we use Assumption 3.6 to conclude that $\mathrm{d_{TV}}\left(X^{|\gamma(f)|-1}\delta_{\boldsymbol{u}_{t(f)}},\rho\right)\leqslant C\lambda^{|\gamma(f)|-1}$, which thus implies that

$$\left|\left\langle N(\boldsymbol{u}_{s(f)}),X^{|\gamma(f)|-1}\delta_{\boldsymbol{u}_{t(f)}}\right\rangle-p\right|\leqslant C\lambda^{|\gamma(f)|-1},\tag{3.11}$$

because $\left\langle N(\boldsymbol{u}_{s(f)}),X^{|\gamma(f)|-1}\delta_{\boldsymbol{u}_{t(f)}}\right\rangle$ represents the probability that a point sampled at random from the measure $X^{|\gamma(f)|-1}\delta_{\boldsymbol{u}_{t(f)}}$ lands in $N(\boldsymbol{u}_{s(f)})$, which is a set of measure p under ρ. We now prove the following by induction.

Claim 3.3.7. We have the following bound on the contribution of $f\in T_J(W)$:

$$\mathop{\mathbf{E}}_{\boldsymbol{u}_i:i\in J(W)}\prod_{f\in T_J(W)}p^{|\gamma(f)|-1}\cdot\left(\left|\left\langle N(\boldsymbol{u}_{s(f)}),X^{|\gamma(f)|-1}\delta_{\boldsymbol{u}_{t(f)}}\right\rangle-p\right|+p\cdot\mathbf{1}[f\in D_J(W)]\right)\leqslant\prod_{f\in T_J(W)}p^{|\gamma(f)|}\cdot\left(2C\lambda^{|\gamma(f)|}+\mathbf{1}[f\in D_J(W)]\right).$$

Proof. We fix an order for $i\in J(W)$, $i_0,\ldots,i_t$ such that i_j is a leaf in $T_J^{(j)}(W)$, the graph obtained by taking $T_J(W)$ and deleting $i_{j+1},\ldots,i_t$. We use f_j to denote the unique edge incident to i_j in $T_J^{(j)}(W)$. Then if we define

$$a_j:=\mathop{\mathbf{E}}_{\boldsymbol{u}_i:i\in V(T_J^{(j)}(W))}\prod_{f\in T_J^{(j)}(W)}p^{|\gamma(f)|-1}\cdot\left(\left|\left\langle N(\boldsymbol{u}_{s(f)}),X^{|\gamma(f)|-1}\delta_{\boldsymbol{u}_{t(f)}}\right\rangle-p\right|+p\cdot\mathbf{1}[f\in D_J(W)]\right)$$

Because f_j is independent of $f_{j'}$ for $j' < j$ we can write:

$$a_j := \mathop{\mathbf{E}}_{\boldsymbol{u}_{i_0}} \cdots \mathop{\mathbf{E}}_{\boldsymbol{u}_{i_{j-1}}} \prod_{f \in T_J^{(j-1)}(W)} p^{|\gamma(f)|-1} \cdot \left(\left| \left\langle N(\boldsymbol{u}_{s(f)}), X^{|\gamma(f)|-1} \delta_{\boldsymbol{u}_{t(f)}} \right\rangle - p \right| + p \cdot \mathbf{1}[f \in D_J(W)] \right) \cdot$$
$$\mathop{\mathbf{E}}_{\boldsymbol{u}_{i_j}} p^{|\gamma(f_j)|-1} \cdot \left(\left| \left\langle N(\boldsymbol{u}_{s(f_j)}), X^{|\gamma(f_j)|-1} \delta_{\boldsymbol{u}_{t(f_j)}} \right\rangle - p \right| + p \cdot \mathbf{1}[f_j \in D_J(W)] \right)$$

Without loss of generality we can assume $i_j = t(f_j)$, and because $N(\boldsymbol{u}_{s(f_j)}) = pX\delta_{\boldsymbol{u}_{s(f_j)}}$,

$$\begin{aligned} \mathop{\mathbf{E}}_{\boldsymbol{u}_{i_j}} \left| \left\langle N(\boldsymbol{u}_{s(f_j)}), X^{|\gamma(f_j)|-1} \delta_{\boldsymbol{u}_{t(f_j)}} \right\rangle - p \right| &= \mathop{\mathbf{E}}_{\boldsymbol{u}_{i_j}} \left| \left\langle N(\boldsymbol{u}_{s(f_j)}), X^{|\gamma(f_j)|-1} \delta_{\boldsymbol{u}_{i_j}} \right\rangle - p \right| \\ &= p \mathop{\mathbf{E}}_{\boldsymbol{u}_{i_j}} \left| \left\langle X^{|\gamma(f_j)|} \delta_{\boldsymbol{u}_{s(f_j)}}, \delta_{\boldsymbol{u}_{i_j}} \right\rangle - 1 \right| \\ &= 2p \cdot \mathrm{d_{TV}} \left(X^{|\gamma(f_j)|} \delta_{\boldsymbol{u}_{s(f_j)}}, \rho \right) \\ &\leqslant p \cdot 2C\lambda^{|\gamma(f_j)|}. \end{aligned}$$

This gives us the inequality:

$$\alpha_j \leqslant \alpha_{j-1} \cdot p\big(2C\lambda^{|\gamma(f_j)|} + \mathbf{1}[f_j \in D_J(W)]\big).$$

The above inequality combined with the fact that $\alpha_0 = 1$ yields the claim. □

We use $e(W)$ to denote $|E(W)|$ and $\mathrm{sing}(W)$ to denote the number of singleton edges in $G_2(W)$,[5]. For any graph H we use $\mathrm{exc}(H)$ to denote the *excess* of H, which is $|E(H)| - |V(H)| + 1$, the number of edges H has over a tree.

Observation 3.3.8. $\mathrm{exc}(G(W)) = \mathrm{exc}(G_2(W)) = \mathrm{exc}(G_J(W))$. Thus, we denote this quantity as $\mathrm{exc}(W)$.

Observation 3.3.9. $|E_J(W)| \leqslant 3\mathrm{exc}(W)$.

Proof. We use Observation 3.3.8 to write:

$$2\mathrm{exc}(W) - 2 = 2|E_J(W)| - 2|V_J(W)| = \sum_{v \in V_J} (\deg_{G(J)}(v) - 2) \geqslant |V_J(W)| - 1,$$

where the degree a self-loop incurs on a vertex is 2, and the -1 on the right-hand side is to capture the possibility that $|J(W)| = 1$ when $G_2(W)$ has no degree-3 vertices. Adding $\mathrm{exc}(W)$ to both sides gives:

$$3\mathrm{exc}(W) \geqslant |E_J(W)|.$$ □

[5]Note $\mathrm{sing}(W)$ is the same as the number of singleton edges in $G(W)$ since $G_1(W)$ cannot have singleton edges, as it is the multigraph induced by a closed walk of length ℓ.

Using the bound on the non-tree edges from (3.11) and Claim 3.3.7, we get:

$$\begin{aligned}(3.9) \leqslant \prod_{f \in T_J(W)} & p^{|\gamma(f)|} \cdot \left(2C\lambda^{|\gamma(f)|} + \mathbf{1}[f \in D_J(W)]\right) \cdot \\ \prod_{f \in E_J(W) \setminus T_J(W)} & p^{|\gamma(f)|-1} \cdot \left(C\lambda^{|\gamma(f)|-1} + p \cdot \mathbf{1}[f \in D_J(W)]\right)\end{aligned}$$

Now, we bound separately the contribution of singleton and duplicative edges. For each $f \in S_J(W)$, we pull out a factor of $(p\lambda)^{|\gamma(f)|}2C$ if the edge was in the tree, and a factor $(p\lambda)^{|\gamma(f)|-1}C$ if the edge was not in the tree; this fully accounts for the contributions of singleton edges. For each $f \in D_J(W)$, we upper bound its contribution by $p^{|\gamma(f)|}3C$ if the edge was in the tree, and a factor $p^{|\gamma(f)|-1}3C$ otherwise; this is potentially loose because we don't keep the factors of λ, but it is a valid upper bound because $C \geqslant 1$ and $p, \lambda \leqslant 1$. We thus have a factor of p from $|E_2(W)| - \mathrm{exc}(W)$ edges, a factor of λ from $\mathrm{sing}(W) - \mathrm{exc}(W)$ edges, and a factor of at most $3C$ from each edge in $E_J(W)$. Summarizing,

$$\leqslant p^{|E_2(W)|-\mathrm{exc}(W)} \lambda^{\mathrm{sing}(W)-\mathrm{exc}(W)} \cdot (3C)^{|E_J(W)|},$$

and by Observation 3.3.9, the above is bounded by:

$$\leqslant p^{|E_2(W)|-\mathrm{exc}(W)} \lambda^{\mathrm{sing}(W)-\mathrm{exc}(W)} \cdot (3C)^{3\mathrm{exc}(W)}.$$

Since $m(e) \geqslant 2$ for every edge in $e \in E_1$ (otherwise the walk cannot be closed), by an analysis identical to that of the first term of (3.10), we have:

$$\begin{aligned}(3.8) &\leqslant p^{|E_1(W)|} \cdot p^{|E_2(W)|-\mathrm{exc}(W)} \lambda^{\mathrm{sing}(W)-\mathrm{exc}(W)} \cdot (3C)^{3\mathrm{exc}(W)} \\ &= p^{e(W)-\mathrm{exc}(W)} \lambda^{\mathrm{sing}(W)} \left(\frac{27C^3}{\lambda}\right)^{\mathrm{exc}(W)}.\end{aligned}$$

Finally, we can bound the trace power (3.7) as follows.

$$\begin{aligned}(3.7) &\leqslant \sum_{W \in \mathcal{W}_\ell} p^{e(W)-\mathrm{exc}(W)} \lambda^{\mathrm{sing}(W)} \left(\frac{27C^3}{\lambda}\right)^{\mathrm{exc}(W)} \\ &= \sum_{a=1}^{\ell} \sum_{b=1}^{\ell} \sum_{c=1}^{\ell} \sum_{\substack{W \in \mathcal{W}_\ell \\ e(W)=a,\, \mathrm{sing}(W)=b,\, \mathrm{exc}(W)=c}} p^{a-c} \lambda^b \left(\frac{27C^3}{\lambda}\right)^c \\ &= \sum_{a=1}^{\ell} \sum_{b=1}^{\ell} \sum_{c=1}^{\ell} p^{a-c} \lambda^b \left(\frac{27C^3}{\lambda}\right)^c \cdot |\{W \in \mathcal{W}_\ell : e(W) = a,\, \mathrm{sing}(W) = b,\, \mathrm{exc}(W) = c\}|\end{aligned} \tag{3.12}$$

To finish bounding the trace power, it remains to count length-ℓ closed walks with a specified number of edges, excess edges, and singleton edges.

Claim 3.3.10. The number of walks W such that $e(W) = a$, $\text{sing}(W) = b$, and $\text{exc}(W) = c$ is at most:

$$n^{a-c+1} \cdot \ell^{2(\ell-b)} \cdot \ell^{2c}.$$

Proof. Observe that W has $a - c + 1$ vertices. Then the following information about W is sufficient to reconstruct it:

- The labels of the visited vertices in $[n]$ in the order in which they are visited. *There are at most n^{a-c+1} labelings.*
- The timestamps when the edge walked on is not a singleton edge. *There are at most $\ell^{\ell-b}$ possibilities.*
- The timestamps when W takes a step uv such that the edge $\{u, v\}$ has not been previously covered by W, but v has been previously visited, along with the timestamp of when v was visited for the first time. *There are c such steps, and hence there are at most ℓ^{2c} possibilities.*
- The timestamps when W takes a step uv such that the edge $\{u, v\}$ has been previously covered by W along with the timestamp of when $\{u, v\}$ was covered the first time. *There are at most $\frac{\ell-b}{2}$ such steps, and hence there are at most $\ell^{\ell-b}$ possibilities.*

Putting the above bounds together completes the proof. □

Observation 3.3.11. Any walk with b singleton edges and c excess edges has at most $\frac{\ell+b}{2}$ edges.

Proof. Each nonsingleton edge must be visited at least twice. There are at most $\ell - b$ nonsingleton steps. So, there are at most $\frac{\ell-b}{2}$ nonsingleton edges, and the total number of edges is at most $\frac{\ell+b}{2}$. □

Now we can continue bounding the trace power.

$$\begin{aligned}
(3.12) &\leqslant \sum_{a=1}^{\ell}\sum_{b=1}^{\ell}\sum_{c=1}^{\ell} p^{a-c}\lambda^b \left(\frac{27C^3}{\lambda}\right)^c \cdot n^{a-c+1} \cdot \ell^{3(\ell-b)} \cdot \ell^{2c} \\
&= n\sum_{a=1}^{\ell}\sum_{b=1}^{\ell}\sum_{c=1}^{\ell} (pn)^a \lambda^b \left(\frac{27C^3\ell^2}{\lambda pn}\right)^c \cdot \ell^{2(\ell-b)} \\
&\leqslant n\ell \cdot \max\left\{1, \left(\frac{27C^3\ell^2}{\lambda pn}\right)^\ell\right\} \cdot \sum_{a=1}^{\ell}\sum_{b=1}^{\ell} (\lambda pn)^b \cdot (pn)^{a-b} \cdot \ell^{2(\ell-b)}
\end{aligned}$$

By Observation 3.3.11 and the assumption on λ from Assumption 3.6, we can bound the total edges a and hence the below.

$$\leqslant n\ell \cdot \max\left\{1, \left(\frac{27C^3\ell^2}{\sqrt{pn}}\right)^\ell\right\} \cdot \sum_{a=1}^{\ell}\sum_{b=1}^{\ell} (\lambda pn)^b \cdot (pn)^{\frac{\ell-b}{2}} \cdot \ell^{2(\ell-b)}$$

$$\leqslant n\ell^3 \cdot \max\left\{1, \left(\frac{27C^3\ell^2}{\sqrt{pn}}\right)^\ell\right\} \max\left\{(\lambda pn)^\ell, (pn\ell^4)^{\ell/2}\right\}$$

By Claim 3.2.2,

$$\mathbf{Pr}\left[\|\overline{A}_{\boldsymbol{G}}\| \geqslant e^\varepsilon \cdot \left(n^{1/\ell}\ell^{3/\ell}\right) \max\left\{1, \frac{27C^3\ell^2}{\sqrt{pn}}\right\} \max\left\{\lambda pn, \sqrt{pn}\ell^2\right\}\right] \leqslant \exp(-\varepsilon\ell),$$

and choosing $\ell = \log^2 n$, $\varepsilon = \log\log n / \log n$, for any constant γ, we get:

$$\|A_{\boldsymbol{G}} - \mathbf{E}\, A_{\boldsymbol{G}}\| \leqslant (1 + o(1)) \cdot \left(1 + \frac{27C^3\log^4 n}{\sqrt{pn}}\right) \cdot \max\left\{\lambda pn, \sqrt{pn}\log^4 n\right\} \tag{3.13}$$

with probability at least $1 - n^{-\gamma}$. Now, by the assumption of the theorem, $pn \gg C^6 \log^8 n$, so $1 + \frac{27C^3\log^4 n}{\sqrt{pn}} = 1 + o(1)$. □

3.4 Analyzing the discrete walk with Brownian motion

In this section we quantify the extent to which convolving a measure α over $\mathbb{S}^{d-1}$ with a spherical cap of measure p brings α closer to uniform, provided that α satisfies a certain monotonicity property. We now define this monotonicity property, establishing a couple of additional definitions along the way.

Definition 3.4.1. We say a distribution on $\mathbb{S}^{d-1}$ with relative density α is *symmetric about* $y \in \mathbb{S}^{d-1}$ if there exists a function $\ell_\alpha : [-1, 1] \to \mathbb{R}$ such that $\alpha(z) = \ell_\alpha(\langle z, y\rangle)$. We note that ℓ_α is also the density α projected onto the line defined by y relative to the projection of the uniform distribution, so that

$$\ell_\alpha(t) = \frac{\int_{\mathbb{S}^{d-1}} \mathbf{1}[\langle z, y\rangle = t] \cdot \ell_\alpha(t)\, d\rho(z)}{\int_{\mathbb{S}^{d-1}} \mathbf{1}[\langle z, y\rangle = t]\, d\rho(z)} = \frac{\int_{\mathbb{S}^{d-1}} \mathbf{1}[\langle z, y\rangle = t] \cdot \alpha(z)\, d\rho(z)}{\int_{\mathbb{S}^{d-1}} \mathbf{1}[\langle z, y\rangle = t]\, d\rho(z)}.$$

Notice that $\ell_\rho = 1$.

Definition 3.4.2. A measure α over $\mathbb{S}^{d-1}$ which is symmetric about some $y \in \mathbb{S}^{d-1}$ is said to be *spherically monotone* if ℓ_α is monotone non-decreasing.

An alternate characterization of spherically monotone distributions is that their relative densities can be written as a non-negative combination of spherical caps. Recall that we use $\underline{\mathrm{cap}}_p(y)$ and $\underline{\mathrm{cap}}_{\geqslant\tau(p)}(y)$ interchangeably to denote the uniform measure over $\mathrm{cap}_p(y)$.

Claim 3.4.3. A density $\alpha : \mathbb{S}^{d-1} \to \mathbb{R}$ which is symmetric about $y \in \mathbb{S}^{d-1}$ is spherically monotone if and only if there is a distribution r on $[-1, 1]$ such that:

$$\alpha = \int \underline{\mathrm{cap}}_{\geqslant \theta} \, dr(\theta).$$

We call the above way of writing α as the *cap decomposition* of α. Further, $\ell_\alpha = \int \ell_{\underline{\mathrm{cap}}_{\geqslant \theta}} \, dr(\theta)$.

We give the straightforward proof later. Notice that in writing the expression for ℓ_α we have replaced $\ell_{\underline{\mathrm{cap}}_{\geqslant \theta}(y)}$ with $\ell_{\underline{\mathrm{cap}}_{\geqslant \theta}}$; this is because $\ell_{\underline{\mathrm{cap}}_{\geqslant \theta}(y)}$ does not depend on y.

Definition 3.4.4. Given a measure μ over $\mathbb{S}^{d-1}$ which is symmetric about some $y \in \mathbb{S}^{d-1}$, its *spherical kernel* P_μ is the transition operator of the random walk on $\mathbb{S}^{d-1}$ where a single step, starting from $x \in \mathbb{S}^{d-1}$, samples $\boldsymbol{a} \sim \ell_\mu$ and then walks from x to a uniformly random $\boldsymbol{w} \in \mathbb{S}^{d-1}$ satisfying $\langle \boldsymbol{w}, x \rangle = \boldsymbol{a}$. Equivalently, the density of $P_\mu \alpha$ is $\mu * \alpha$ for $*$ denoting convolution.

Remark 3.4.5. For brevity, we will use P_p as a shorthand for $P_{\underline{\mathrm{cap}}_p}$.

The main result of this section, proved after developing some tools, is the following:

Theorem 3.4.6. *If a probability distribution α over $\mathbb{S}^{d-1}$ is symmetric and spherically monotone, then for any integer $k \geqslant 0$,*

$$\mathrm{d_{TV}}\left(P_p^k \alpha, \rho\right) \leqslant \left((1 + o_{d\tau^2}(1)) \cdot \tau\right)^k \cdot \sqrt{\tfrac{1}{2} \mathrm{D}(\alpha \| \rho)},$$

where $o_{d\tau^2}(1)$ denotes a function that goes to 0 as $d\tau^2 \to \infty$.

As an immediate corollary, we obtain the following version which can be used in conjunction with Theorem 3.1.11 to conclude a bound on the second eigenvalue of random geometric graphs.

Corollary 3.4.7. *For any probability distribution α over $\mathbb{S}^{d-1}$,*

$$\mathrm{d_{TV}}\left(P_p^k \alpha, \rho\right) \leqslant \left((1 + o_{d\tau^2}(1)) \cdot \tau\right)^{k-1} \cdot \sqrt{\frac{1}{2} \cdot \log \frac{1}{p}}.$$

Proof. We write α as a convex combination of (symmetric, spherically monotone) point masses δ_x. Then we apply Theorem 3.4.6 in conjunction with the triangle inequality and the fact that $P_p^k \delta_x = P_p^{k-1} \underline{\mathrm{cap}}_p(x)$ and $\mathrm{D}(\underline{\mathrm{cap}}_p(x) \| \rho) = \log \frac{1}{p}$. □

Our proof of Theorem 3.4.6 will relate the action of P_p to the action of the *Brownian motion kernel*.

Definition 3.4.8 (Brownian motion on $\mathbb{S}^{d-1}$)**.** Let $(\boldsymbol{B}_t)_{t\geqslant 0}$ be standard Brownian motion in $\mathbb{R}^d$. We define Brownian motion on $\mathbb{S}^{d-1}$ starting at some point $V_0 \in \mathbb{S}^{d-1}$ as the process $(\boldsymbol{V}_t)_{t\geqslant 0}$ via the following stochastic differential equation:

$$d\boldsymbol{V}_t = \sqrt{2}\big(\mathbb{1} - \boldsymbol{V}_t\boldsymbol{V}_t^\top\big)\, d\boldsymbol{B}_t - (d-1)\boldsymbol{V}_t\, dt.$$

Definition 3.4.9. For any $t \geqslant 0$, let the *time-t Brownian motion kernel* U_t be the transition operator of a random walk on $\mathbb{S}^{d-1}$ where a single step samples runs a time-t Brownian motion on the sphere. Equivalently, $U_t = P_{\beta_t}$ for β_t the (spherically symmetric) density of a t-step Brownian motion.

For any $y \in \mathbb{S}^{d-1}$, $P_p y$ is highly concentrated near the boundary of the cap of measure p around y. As we will show in Section 3.4.1, the same is true for $U_t y$; it is highly concentrated near the boundary of a cap of measure $q = q(t)$ around y. So, choosing $T > 0$ so that $q(T) \approx p$, we will argue that U_T and P_p have similar action on spherically monotone measures.

We can then take advantage of the contractive properties of U_T in order to prove that P_p is contractive. The Brownian motion kernel satisfies the following mixing condition (which can be obtained, e.g., as a corollary of [8, Theorem 5.2.1] and [33, Corollary 2]):

Theorem 3.4.10 (Mixing of Brownian motion on $\mathbb{S}^{d-1}$)**.** *For any probability distribution ϕ on $\mathbb{S}^{d-1}$,*

$$\mathrm{D}(U_t\phi \,\|\, \rho) \leqslant \exp(-2(d-1)t) \cdot \mathrm{D}(\phi \,\|\, \rho)$$

As a corollary of the above and Pinsker's inequality, for any $t > 0$ and measure α over $\mathbb{S}^{d-1}$,

$$2\left(\mathrm{d_{TV}}\left(U_t\alpha, \rho\right)\right)^2 \leqslant \mathrm{D}(U_t\alpha \,\|\, \rho) \leqslant \exp(-2(d-1)t) \cdot \mathrm{D}(\alpha \,\|\, \rho). \tag{3.14}$$

Armed with (3.14), we can pass to working exclusively with the 1-dimensional projection of the measures in question onto the direction y.

Claim 3.4.11. For any spherically symmetric distribution with relative density γ, $\mathrm{d_{TV}}\left(\gamma, \rho\right) = \mathrm{d_{TV}}\left(\ell_\gamma, \ell_\rho\right)$.

Proof. We express the total variation distance in terms of the ℓ_1 norm:

$$\begin{aligned} 2\mathrm{d_{TV}}\left(\gamma, \rho\right) = \int_{z\in\mathbb{S}^{d-1}} |\gamma(z) - 1| \; d\rho(z) &= \int_{z\in\mathbb{S}^{d-1}} |\ell_\gamma(\langle z, y\rangle) - 1| \; d\rho(z) \\ &= \int_{t\in[-1,1]} |\ell_\gamma(t) - 1| \; d\ell_\rho(t) = 2\mathrm{d_{TV}}\left(\ell_\gamma, \ell_\rho\right). \end{aligned}$$ □

Note that if α is spherically symmetric about y then so is $U_t\alpha$, by the rotational invariance of Brownian Motion on the sphere. Hence combining Claim 3.4.11 with (3.14), we have that

$$\mathrm{d_{TV}}\left(\ell_{U_t\alpha}, \ell_\rho\right) \leqslant \sqrt{\frac{1}{2} \cdot \exp(-(d-1)t) \cdot \mathrm{D}(\alpha \| \rho)}.$$

Now, we'll show that for a well-chosen $T > 0$, $\ell_{U_T\alpha}$ nearly stochastically dominates $\ell_{P_p\alpha}$, and that $P_p\alpha$ and $U_T\alpha$ are both spherically monotone, and that this furthermore implies that $\mathrm{d_{TV}}\left(\ell_{U_T\alpha}, \ell_\rho\right)$ and $\mathrm{d_{TV}}\left(\ell_{P_p\alpha}, \ell_\rho\right)$ are related. Specifically, we show the following lemmas:

Lemma 3.4.12. *If ν and μ are spherically monotone densities and $\ell_\nu \preceq_{\mathrm{st}} \ell_\mu$, then*[6]

$$\mathrm{d_{TV}}\left(\ell_\nu, \ell_\rho\right) \leqslant \mathrm{d_{TV}}\left(\ell_\mu, \ell_\rho\right).$$

We prove the lemma below, but intuitively, a spherically monotone distribution can be realized as a non-negative combination of spherical caps; the uniform distribution has all of its mass on the largest cap (of measure 1). If $\ell_\nu \preceq_{\mathrm{st}} \ell_\mu$, then the total probability mass within any radius θ of the mode of μ exceeds that of ν, witnessing a larger total variation distance.

Lemma 3.4.13. *Let μ, ν, α be spherically monotone densities over $\mathbb{S}^{d-1}$, with $\ell_\nu \preceq_{\mathrm{st}} \ell_\mu$. Then*

1. *$P_\mu\alpha$ is spherically monotone (as is $P_\nu\alpha$),*
2. *$\ell_{P_\alpha\nu} \preceq_{\mathrm{st}} \ell_{P_\alpha\mu}$, and*
3. *$\ell_{P_\nu\alpha} \preceq_{\mathrm{st}} \ell_{P_\mu\alpha}$.*

We will prove this lemma below as well; the crux of the proof of Part 1 is to realize that because α, μ are spherically monotone, they can be decomposed as a non-negative combination of spherical caps. Then, by linearity of P_μ and by the commutativity of convolution, Part 1 reduces to showing that the convolution of two spherical caps is spherically monotone (this is a statement that we find intuitive, and it is easy to verify by directly examining the expression for $\ell_{P_{\geqslant\theta}\underline{\mathrm{cap}}_{\geqslant\psi}}$). To show Part 2, we observe that by decomposing α in its cap decomposition, it is then enough to compare $\ell_{P_{\geqslant\theta}\nu}$ with $\ell_{P_{\geqslant\theta}\mu}$ for each θ. Here, when $\ell_\mu \succeq_{\mathrm{st}} \ell_\nu$, a straightforward coupling demonstrates that $\ell_{P_{\geqslant\theta}\mu} \succeq_{\mathrm{st}} \ell_{P_{\geqslant\theta}\nu}$. Part 3 is a consequence of Part 2 and commutativity of convolution.

Our aim is to now apply these lemmas with $\nu \approx \underline{\mathrm{cap}}_p(y)$ and $\mu = \beta_T$ (note that $P_\nu\alpha = P_p\alpha$ and $P_\mu\alpha = U_T\alpha$). We now verify that these densities meet the conditions above. The density $\underline{\mathrm{cap}}_p(y)$ is spherically monotone because it is the same as ρ conditoned on being closer to y; we now show that β_t is indeed spherically monotone.

Claim 3.4.14. The density of a time-t Brownian motion, β_t, is spherically monotone.

Proof. Since Brownian motion on $\mathbb{S}^{d-1}$ can be realized as a sequence of random steps within spherical caps of infintesimally small measure ds, the measure of a t-step Brownian motion starting from $y \in \mathbb{S}^{d-1}$ is achieved by iteratively applying $P_{\underline{\mathrm{cap}}_{ds}}$ to the point mass at y. The proof is then complete by noting that $\ell_{\underline{\mathrm{cap}}_p}$ is spherically monotone for every p, then applying Part 1 of Lemma 3.4.13. □

[6] As will be apparent from the proof, one may replace ℓ_ν, ℓ_μ with any monotone non-decreasing densities on $[-1, 1]$.

Next, we argue that for $T = T(p)$, there is some small δ for which $(1-\delta)\ell_{\underline{\mathrm{cap}}_p} + \delta\ell_\rho \preceq_{\mathrm{st}} \ell_{\beta_T}$; that is, the linear projection of the p-cap is almost stochastically dominated by the linear projection of Brownian motion run for the proper amount of time. In order to do this, we first establish that almost all of the probability mass of β_T is in a cap of radius close to p. In Section 3.4.1, we'll prove the following lemma:

Lemma 3.4.15. *Let $(\boldsymbol{V}_t)_{t\geqslant 0}$ be a Brownian motion on $\mathbb{S}^{d-1}$ starting at V_0. Then for any time $t \geqslant 0$,*

$$\mathbf{Pr}\left[|\langle V_0, \boldsymbol{V}_t\rangle - \exp\left(-(d-1)t\right)| \geqslant x\right] \leqslant 2\exp\left(-\tfrac{d-1}{2}\frac{x^2}{1-e^{-2(d-1)t}}\right).$$

From this lemma, we can show that almost all of the mass of the cap decomposition of ℓ_{β_T} is contained inside a $(\geqslant \tau)$-cap:

Claim 3.4.16. Let $\nu > 0$, $T \coloneqq \frac{1}{d-1}\left(\log\frac{1}{\nu} - 2\varepsilon\right)$, and $\varepsilon \in \left[0, \frac{1}{2}\log\frac{1}{\nu}\right]$. Then the total mass of ℓ_{β_T} outside of $\mathrm{cap}_{\geqslant(1+\varepsilon)\nu}(V_0)$ for V_0 the starting point of the Brownian motion is bounded:

$$\int_{-1}^{(1+\varepsilon)\tau} d\ell_{\beta_T}(x) \leqslant \delta(\varepsilon) \coloneqq 2\exp\left(-\frac{(d-1)\varepsilon^2\nu^2}{2(1-\nu^2)}\right).$$

Proof. We let $(\boldsymbol{V}_t)_{t\geqslant 0}$ be a Brownian motion on the sphere, $\boldsymbol{A}_t = \langle \boldsymbol{V}_t, V_0\rangle$, and $\boldsymbol{A}_t = \exp\left(-(d-1)t\right) + \boldsymbol{R}_t$. At time T, we have

$$\begin{aligned} \boldsymbol{A}_T = \exp\left(-(d-1)\cdot T\right) + \boldsymbol{R}_T = \nu\cdot\exp\left(2\varepsilon\right) + \boldsymbol{R}_T \\ \geqslant \nu\cdot(1+2\varepsilon) + \boldsymbol{R}_T \geqslant \nu\cdot(1+2\varepsilon) + \boldsymbol{R}_T. \end{aligned}$$

The event that $\boldsymbol{A}_T \leqslant \nu\cdot(1+\varepsilon)$ implies $\boldsymbol{R}_T < -\varepsilon\nu$, so it suffices to upper bound the probability that $|\boldsymbol{R}_T| > \varepsilon\nu$. Applying Lemma 3.4.15,

$$\mathbf{Pr}[|\boldsymbol{R}_T| \geqslant \varepsilon\nu] \leqslant 2\exp\left(-\tfrac{d-1}{2}\frac{\varepsilon^2\nu^2}{1-e^{-2(d-1)\cdot T}}\right) = 2\exp\left(-\frac{\varepsilon^2\nu^2(d-1)}{2(1-\nu^2)}\right)$$ □

Now, we are ready to establish the stochastic domination of the combination.

Claim 3.4.17. Let $p \in \left(0, \frac{1}{2}\right)$ and $\nu = \tau(p) + \frac{4}{\sqrt{d}}$. For $T = \frac{1}{d-1}(\log\frac{1}{\nu} - 2\varepsilon)$ with $\varepsilon \in \left[\frac{5}{(d-1)\nu^2}, \frac{1}{2}\log\frac{1}{\nu}\right]$,

$$\ell_{\beta_T} \succeq_{\mathrm{st}} (1-2\delta(\varepsilon))\ell_{\underline{\mathrm{cap}}_p} + 2\delta(\varepsilon)\ell_\rho,$$

for $\delta(\varepsilon)$ as defined in the statement of Claim 3.4.16.

Proof. Using Claim 3.4.3, we write $\beta_T = \int_{-1}^{1} c_\theta \cdot \ell_{\underline{\mathrm{cap}}_{\geqslant\theta}} d\theta$, with $\int c_\theta d\theta = 1$. Let $\tau' \in [-1, 1]$ be such that

$$\int_{-1}^{\tau'} c_\theta d\theta = 2\delta(\varepsilon), \quad \text{and} \quad \int_{\tau'}^{1} c_\theta d\theta = 1 - 2\delta(\varepsilon). \tag{3.15}$$

The proof strategy is to show that the conclusion follows if $\tau' \geqslant \tau(p)$, and then establish that inequality.

First observe that if α and $\{\gamma_x\}_{x\in X}$ are measures satisfying $\gamma_x \succeq_{\text{st}} \alpha$ for all $x \in X$, then a convex combination $\int c_x\gamma_x dx \succeq_{\text{st}} \alpha$ as well, from which the conclusion follows. Now, writing

$$\beta_T = \int_{-1}^{\tau'} c_\theta \cdot \ell_{\underline{\text{cap}}_{\geqslant\theta}} d\theta + \int_{\tau'}^{1} c_\theta \cdot \ell_{\underline{\text{cap}}_{\geqslant\theta}} d\theta,$$

we see that the first term on the right-hand-side stochastically dominates $2\delta(\varepsilon)\cdot \ell_{\underline{\text{cap}}_{\geqslant -1}} = 2\delta(\varepsilon)\cdot\ell_\rho$ since for every $\theta \in [-1,\tau']$, $\theta \geqslant -1$ and therefore $\ell_{\underline{\text{cap}}_{\geqslant\theta}} \succeq_{\text{st}} \ell_{\underline{\text{cap}}_{\geqslant -1}} = \ell_\rho$. By identical reasoning, the second term stochastically dominates $\ell_{\underline{\text{cap}}_{\geqslant\tau(p)}}$ since for every $\theta \in [\tau',1]$, $\theta \geqslant \tau(p)$ and therefore $\ell_{\underline{\text{cap}}_{\geqslant\theta}} \succeq_{\text{st}} \ell_{\underline{\text{cap}}_{\geqslant\tau}} = \ell_{\underline{\text{cap}}_p}$.

We now show that the τ' satisfying (3.15) is at least τ, for which it is sufficient to show $\tau' \geqslant \nu$. Let $\kappa = \int_{-1}^{\nu} c_\theta\, d\theta$; $\tau' \geqslant \nu$ is equivalent to showing that $\kappa \leqslant 2\delta(\varepsilon)$. Using Claim 3.4.16, we know that $\mathbf{Pr}_{\boldsymbol{v}\sim\beta_T}[\boldsymbol{v} \in \text{cap}_{\geqslant(1+\varepsilon)\nu}(V_0)] \geqslant 1-\delta(\varepsilon)$.

$$\begin{aligned}
&1-\delta(\varepsilon)\\
&\leqslant \mathop{\mathbf{Pr}}_{\boldsymbol{v}\sim\beta_T}\left[\boldsymbol{v}\in \text{cap}_{\geqslant(1+\varepsilon)\nu}(V_0)\right]\\
&= \int_{-1}^{(1+\varepsilon)\nu} c_\theta \cdot \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\ell_{\underline{\text{cap}}_{\geqslant\theta}}}[\boldsymbol{x}\geqslant(1+\varepsilon)\nu]\,d\theta + \int_{(1+\varepsilon)\nu}^{1} c_\theta\,d\theta\\
&\leqslant \int_{-1}^{\nu} c_\theta \cdot \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\ell_{\underline{\text{cap}}_{\geqslant\theta}}}[\boldsymbol{x}\geqslant(1+\varepsilon)\nu]\,d\theta + \int_{\nu}^{1} c_\theta\,d\theta\\
&\leqslant \left(\max_{\theta\in[-1,\nu]}\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\ell_{\underline{\text{cap}}_{\geqslant\theta}}}[\boldsymbol{x}\geqslant(1+\varepsilon)\nu]\right)\cdot\kappa + \int_{\nu}^{1} c_\theta\,d\theta = \mathop{\mathbf{Pr}}_{x\sim\ell_{\underline{\text{cap}}_{\geqslant\nu}}}[\boldsymbol{x}\geqslant(1+\varepsilon)\nu]\cdot\kappa + \int_{\nu}^{1} c_\theta\,d\theta.
\end{aligned}$$

Using Lemma 3.2.8 and $\nu \geqslant 4/\sqrt{d}$,

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\ell_{\underline{\text{cap}}_{\geqslant\nu}}}[\boldsymbol{x}\geqslant(1+\varepsilon)\nu] &= \frac{\rho(\text{cap}_{\geqslant(1+\varepsilon)\nu})}{\rho(\text{cap}_{\geqslant\nu})}\\
&\leqslant \frac{3\nu\left(1-((1+\varepsilon)\nu)^2\right)^{(d-1)/2}}{2\nu(1+\varepsilon)\left(1-\nu^2\right)^{(d-1)/2}}\\
&\leqslant \frac{3}{2}\cdot\left(\frac{1-(1+\varepsilon)^2\nu^2}{1-\nu^2}\right)^{(d-1)/2}\\
&= \frac{3}{2}\cdot\left(1-\frac{2\varepsilon\nu^2+\varepsilon^2\nu^2}{1-\nu^2}\right)^{(d-1)/2}\\
&\leqslant \frac{3}{2}\cdot\left(1-2\varepsilon\nu^2\right)^{(d-1)/2}\\
&\leqslant \frac{3}{1+(d-1)\varepsilon\nu^2}.
\end{aligned}$$

The final quantity is smaller than $\frac{1}{2}$ given our lower bound on ε, and $\int_\tau^1 c_\theta \, d\theta = 1 - \kappa$. Plugging into the above, we have that

$$1 - \delta(\varepsilon) \leqslant \frac{1}{2}\kappa + 1 - \kappa \implies \kappa \leqslant 2\delta(\varepsilon),$$

which completes the proof. □

Finally, we will need the following claim to transfer the statement about the stochastic domination of a linear combination of $\ell_{P_p\alpha}$ and ℓ_ρ to just $\ell_{P_p\alpha}$:

Lemma 3.4.18. *Suppose μ and ν are spherically monotone distributions, then for any $\eta \in [0, 1)$,*

$$\mathrm{d}_{\mathrm{TV}}\left(\ell_\mu, \ell_\rho\right) \leqslant \frac{1}{1-\eta} \mathrm{d}_{\mathrm{TV}}\left((1-\eta)\ell_\mu + \eta\ell_\nu, \ell_\rho\right).$$

Proof. Let $s \in [-1, 1]$ be such that:

$$\mathrm{d}_{\mathrm{TV}}\left(\ell_\mu, \ell_\rho\right) = \int_s^1 (\ell_\mu(x) - 1) d\rho_{\mathrm{1D}} = \int_{-1}^s (1 - \ell_\mu(x)) d\rho_{\mathrm{1D}},$$

where ρ_{1D} is the density of the 1-dimensional projection of $\underline{\mathrm{cap}}_1(y)$. The choice of s satisfying the above is the one satisfying $\ell_\mu(s) = 1$. If $\ell_\nu(s) \geqslant 1$, by spherical monotonicity $\ell_\nu(x) \geqslant 1$ on $[s, 1]$ and:

$$\begin{aligned}\mathrm{d}_{\mathrm{TV}}\left((1-\eta)\ell_\mu + \eta\ell_\nu, \ell_\rho\right) &\geqslant \int_s^1 \left((1-\eta)(\ell_\mu(x) - 1) + \eta(\ell_\nu(x) - 1)\right) d\rho_{\mathrm{1D}} \\ &\geqslant (1-\eta) \int_s^1 (\ell_\mu(x) - 1) d\rho_{\mathrm{1D}} = (1-\eta) \cdot \mathrm{d}_{\mathrm{TV}}\left(\ell_\mu, \ell_\rho\right).\end{aligned}$$

On the other hand, if $\ell_\nu(s) \leqslant 1$, by an identical argument we know:

$$\mathrm{d}_{\mathrm{TV}}\left((1-\eta)\ell_\mu + \eta\ell_\nu, \ell_\rho\right) \geqslant (1-\eta) \int_{-1}^s (1 - \ell_\mu(x)) d\rho_{\mathrm{1D}} = (1-\eta) \cdot \mathrm{d}_{\mathrm{TV}}\left(\ell_\mu, \ell_\rho\right).$$

The desired statement follows from rearranging the above inequality. □

We are now ready to prove Theorem 3.4.6, following the reasoning above, in combination with induction on k, the number of applications of P_p. We state and prove a more refined version of Theorem 3.4.6 below.

Theorem 3.4.19. *If a probability distribution α over $\mathbb{S}^{d-1}$ is symmetric and spherically monotone, then for any integer $k \geqslant 0$ and for $\nu = \tau(p, d) + \frac{4}{\sqrt{d}}$,*

$$\mathrm{d}_{\mathrm{TV}}\left(P_p^k \alpha, \rho\right) \leqslant \nu^k \left(\frac{\exp\left(\frac{4}{(d-1)^{1/4}\sqrt{\nu}}\right)}{\sqrt{1 - 2\exp(-\nu\sqrt{d-1})}} \right).$$

Note that when $\tau^2 d \to \infty$, the parenthesized term is $1 + o(1)$ and $\nu = \tau \cdot (1 + o(1))$.

Proof. Suppose $\tau(p) \geqslant 1 - 1/(d-1)^{1/4}$, then the statement is vacuously true. Thus, we assume from now on $\tau(p) < 1 - 1/(d-1)^{1/4}$.

Let $\nu = \tau(p) + 4/\sqrt{d}$, let $t = \frac{1}{d-1}\left(\log\frac{1}{\nu} - 2\varepsilon\right)$, and $\delta = 2\exp\left(-\frac{(d-1)\varepsilon^2\nu^2}{2(1-\nu^2)}\right)$ for $\varepsilon = \frac{\sqrt{2-2\nu^2}}{(d-1)^{1/4}\sqrt{\nu}}$; note that for d sufficiently large, $\varepsilon \in \left[\frac{5}{(d-1)\nu^2}, \frac{1}{2}\log\frac{1}{\nu}\right]$. For convenience's sake, define $P_{p,\delta} = (1-2\delta)P_p + 2\delta P_1$. We will prove that

$$\ell_{P_{p,\delta}^k \alpha} \preceq_{\mathrm{st}} \ell_{U_t^k \alpha}, \qquad \text{and} \qquad U_t^k \alpha,\ P_{p,\delta}^k \alpha \quad \text{are spherically monotone.} \tag{3.16}$$

Given this, the proof of the theorem will follow: by the linearity of the projection onto the line defined by y, and by the commutativity of convolution,

$$\ell_{P_{p,\delta}^k \alpha} = \sum_{j=0}^{k} (1-2\delta)^{k-j}(2\delta)^j \binom{k}{j} \ell_{P_p^{k-j} P_1^j \alpha},$$

So from Claim 3.4.11, Lemma 3.4.18, (3.16), and Lemma 3.4.12,

$$\mathrm{d_{TV}}\left(P_p^k \alpha, \rho\right) = \mathrm{d_{TV}}\left(\ell_{P_p^k \alpha}, \ell_\rho\right) \leqslant \frac{1}{(1-2\delta)^k}\mathrm{d_{TV}}\left(\ell_{P_{p,\delta}^k \alpha}, \ell_\rho\right) \leqslant \frac{1}{(1-2\delta)^k}\mathrm{d_{TV}}\left(\ell_{U_t^k \alpha}, \ell_\rho\right). \tag{3.17}$$

Then we can apply Claim 3.4.11 to get that

$$\mathrm{d_{TV}}\left(\ell_{U_t^k \alpha}, \ell_\rho\right) = \mathrm{d_{TV}}\left(U_t^k \alpha, \rho\right), \tag{3.18}$$

and finally using that $U_t^k = U_{k\cdot t}$ in conjunction with Lemma 3.4.10, we have that

$$\mathrm{d_{TV}}\left(U_t^k \alpha, \rho\right) = \mathrm{d_{TV}}\left(U_{k\cdot t} \alpha, \rho\right) \leqslant \sqrt{\frac{1}{2}\exp(-2(d-1)tk) \cdot \mathrm{D}(\alpha \| \rho)}, \tag{3.19}$$

So combining (3.17), (3.18), and (3.19), we have that

$$\mathrm{d_{TV}}\left(P_p^k \alpha, \rho\right) \leqslant \sqrt{\frac{1}{2(1-2\delta)^k}\exp(-2(d-1)tk) \cdot \mathrm{D}(\alpha \| \rho)}.$$

In our case, $\delta = \exp(-\nu\sqrt{d-1})$, $t = \frac{1}{d-1}\left(\log\frac{1}{\nu} - \frac{\sqrt{2-2\nu^2}}{(d-1)^{1/4}\sqrt{\nu}}\right)$, so combining these estimates,

$$\mathrm{d_{TV}}\left(P_p^k \alpha, \rho\right) \leqslant \nu^k \cdot \left(\frac{\exp\left(\frac{4}{(d-1)^{1/4}\sqrt{\nu}}\right)}{\sqrt{1 - 2\exp(-\nu\sqrt{d-1})}}\right)^k \cdot \sqrt{\tfrac{1}{2}\mathrm{D}(\alpha \| \rho)},$$

as desired.

Now we prove (3.16). The proof is by induction on k; when $k = 0$, there is nothing to prove. Suppose now that the statement holds true for k; we shall prove it for $k+1$. By Claim 3.4.14, the density of a time-t spherical Brownian motion β_t is spherically monotone about its starting point, and clearly, any convex combination of caps is spherically monotone. Hence we can apply Lemma 3.4.13, Part 1 in conjunction with the induction hypothesis to conclude that both $P_{p,\delta}^{k+1}\alpha = P_{p,\delta}(P_{p,\delta}^k\alpha)$ and $U_t^{k+1}\alpha = U_t(U_t^k\alpha)$ are spherically monotone, giving the second part of the induction hypothesis.

By our induction hypothesis $U_t^k\alpha$ and $P_{p,\delta}^k\alpha$ are spherically monotone with $\ell_{U_t^k\alpha} \succeq_{\mathrm{st}} \ell_{P_{p,\delta}^k\alpha}$, and so we can apply Lemma 3.4.13, Part 2 in conjunction with Claim 3.4.14 to conclude that

$$\ell_{U_t^{k+1}\alpha} = \ell_{P_{\beta_t}(U_t^k\alpha)} \succeq_{\mathrm{st}} \ell_{P_{\beta_t}(P_{p,\delta}^k\alpha)},$$

and then apply Lemma 3.4.13, Part 3 in conjunction with Claim 3.4.17 to conclude that

$$\ell_{P_{\beta_t}(P_{p,\delta}^k\alpha)} \succeq_{\mathrm{st}} \ell_{P_{p,\delta}(P_{p,\delta}^k\alpha)} = \ell_{P_{p,\delta}^k\alpha},$$

completing the proof. □

Now, we fill in the proofs of the lemmas from above.

Claim (Restatement of Claim 3.4.3)**.** A density $\alpha : \mathbb{S}^{d-1} \to \mathbb{R}$ which is symmetric about $y \in \mathbb{S}^{d-1}$ is spherically monotone if and only if there is a distribution r on $[-1,1]$ such that:

$$\alpha = \int \underline{\mathrm{cap}}_{\geqslant\theta}\, dr(\theta).$$

We call the above way of writing α as the *cap decomposition* of α. Further, $\ell_\alpha = \int \ell_{\underline{\mathrm{cap}}_{\geqslant\theta}}\, dr(\theta)$.

Proof of Claim 3.4.3. We first prove the "only if" direction. Since α is spherically symmetric about y, $\alpha(v) = \ell_\alpha(\langle v, y\rangle)$. Let $d\ell_\alpha$ be the distributional derivative of ℓ_α, and set $dr(\theta) = \rho(\mathrm{cap}_{\geqslant\theta}(y))\, d\ell_\alpha(\theta)$.

$$\int (\underline{\mathrm{cap}}_{\geqslant\theta}(y))(v)\, dr(\theta) = \int \frac{\mathbf{1}[\langle v,y\rangle \geqslant \theta]}{\rho(\mathrm{cap}_{\geqslant\theta}(y))} \cdot \rho(\mathrm{cap}_{\geqslant\theta}(y))\, d\ell_\alpha(\theta) = \int \mathbf{1}[\langle v,y\rangle \geqslant \theta]\, d\ell_\alpha(\theta) = \alpha(v).$$

To see that the measure dr indeed gives a probability distribution, first observe that $dr(\theta) \geqslant 0$ for every θ due to the monotonicity of ℓ_α, and next observe that

$$\begin{aligned} 1 = \int_{v\in\mathbb{S}^{d-1}} \alpha(v)\, d\rho(v) &= \int_{v\in\mathbb{S}^{d-1}} \int_{-1}^{1} (\underline{\mathrm{cap}}_{\geqslant\theta}(y))(v)\, dr(\theta)\, d\rho(v) \\ &= \int_{-1}^{1} \int_{v\in\mathbb{S}^{d-1}} (\underline{\mathrm{cap}}_{\geqslant\theta}(y))(v)\, d\rho(v)\, dr(\theta) = \int_{-1}^{1} dr(\theta). \end{aligned}$$

In summary, since r is a positive measure which integrates to 1, it is a probability distribution. The claim regarding ℓ_α follows because the line projection onto y is a linear operation.

Now we prove the converse. Suppose $\alpha = \int_{-1}^{1} \underline{\mathrm{cap}}_{\geqslant\theta}(y)\, dr(\theta)$. By linearity of projection onto the line defined by y, $\ell_\alpha = \int_{-1}^{1} \ell_{\underline{\mathrm{cap}}_{\geqslant\theta}}\, dr(\theta)$. Since $\ell_{\underline{\mathrm{cap}}_{\geqslant\theta}}$ is monotone for every θ, and a non-negative combination of monotone functions is monotone, ℓ_α is also monotone, concluding the proof. □

Lemma (Restatement of Lemma 3.4.12)**.** *If ν and μ are spherically monotone densities and $\ell_\nu \preceq_{\mathrm{st}} \ell_\mu$, then*[7]

$$\mathrm{d}_{\mathrm{TV}}\left(\ell_\nu, \ell_\rho\right) \leqslant \mathrm{d}_{\mathrm{TV}}\left(\ell_\mu, \ell_\rho\right).$$

Proof of Lemma 3.4.12. First, observe that $\ell_\nu \succeq_{\mathrm{st}} \ell_\rho$ and $\ell_\mu \succeq_{\mathrm{st}} \ell_\rho$ by the assumption that μ, ν are spherically monotone. Thus, $\ell_\mu \succeq_{\mathrm{st}} \ell_\nu \succeq_{\mathrm{st}} \ell_\rho$. Further, if measures a, b on $[-1, 1]$ satisfy $a \succeq_{\mathrm{st}} b$, then their CDFs G_a and G_b satisfy $G_a(s) \leqslant G_b(s)$ for every s. Hence,

$$G_{\ell_\mu}(s) \leqslant G_{\ell_\nu}(s) \leqslant G_{\ell_\rho}(s) \quad \forall s \in [-1, 1].$$

By definition of the total variation distance, for any non-decreasing density $\gamma : [-1, 1] \to \mathbb{R}$,

$$\mathrm{d}_{\mathrm{TV}}\left(\gamma, \ell_\rho\right) = \max_{s \in [-1,1]} G_{\ell_\rho}(s) - G_\gamma(s).$$

Thus,

$$\mathrm{d}_{\mathrm{TV}}\left(\ell_\nu, \ell_\rho\right) = G_{\ell_\rho}(s^*) - G_{\ell_\nu}(s^*) \leqslant G_{\ell_\rho}(s^*) - G_{\ell_\mu}(s^*) \leqslant \mathrm{d}_{\mathrm{TV}}\left(\ell_\mu, \ell_\rho\right),$$

which completes the proof. □

We'll now prove Lemma 3.4.13.

Lemma (Restatement of Lemma 3.4.13)**.** *Let μ, ν, α be spherically monotone densities over $\mathbb{S}^{d-1}$, with $\ell_\nu \preceq_{\mathrm{st}} \ell_\mu$. Then*

1. *$P_\mu \alpha$ is spherically monotone (as is $P_\nu \alpha$),*
2. *$\ell_{P_\alpha \nu} \preceq_{\mathrm{st}} \ell_{P_\alpha \mu}$, and*
3. *$\ell_{P_\nu \alpha} \preceq_{\mathrm{st}} \ell_{P_\mu \alpha}$.*

Proof of Lemma 3.4.13. We first prove Part 1. We can write α and μ in terms of their *cap decompositions* as shown in Claim 3.4.3, $\alpha = \int_{-1}^{1} \underline{\mathrm{cap}}_{\geqslant\theta}(y)\, dr(\theta)$ and $\mu = \int_{-1}^{1} \underline{\mathrm{cap}}_{\geqslant\psi}(z)\, ds(\psi)$ for some $z \in \mathbb{S}^{d-1}$. P_μ is a linear operator, so $P_\mu \alpha = \int P_\mu \underline{\mathrm{cap}}_{\geqslant\theta}(y)\, dr(\theta)$. Further, by the commutativity of convolution, $P_\mu \underline{\mathrm{cap}}_{\geqslant\theta}(y) = P_{\geqslant\theta} \mu_y$, where μ_y denotes the version of μ centered at y. Hence,

$$P_\mu \alpha = \int P_\mu \underline{\mathrm{cap}}_{\geqslant\theta}(y) dr(\theta) = \int P_{\geqslant\theta} \mu_y\, dr(\theta) = \int\int P_{\geqslant\theta} \underline{\mathrm{cap}}_{\geqslant\psi}(y)\, ds(\psi)\, dr(\theta).$$

[7] As will be apparent from the proof, one may replace ℓ_ν, ℓ_μ with any monotone non-decreasing densities on $[-1, 1]$.

Each $P_{\geqslant\theta}\underline{\mathrm{cap}}_{\geqslant\psi}(y)$ is clearly spherically symmetric about y. Since the projection onto the line defined by y is a linear operation, $\ell_{P_\mu\alpha} = \int\int \ell_{P_{\geqslant\theta}\underline{\mathrm{cap}}_{\geqslant\psi}} ds(\psi)dr(\theta)$, and because a non-negative combination of monotone functions is monotone, it suffices to prove that for any $\theta, \psi \in [-1,1]$, $\ell_{P_{\geqslant\theta}\underline{\mathrm{cap}}_{\geqslant\psi}}$ is monotone. By definition,

$$\begin{aligned}
\ell_{P_{\geqslant\theta}\underline{\mathrm{cap}}_{\geqslant\psi}(y)}(t) &= \mathop{\mathbf{E}}_{\boldsymbol{v}\sim\rho}\left[\left(P_{\geqslant\theta}\underline{\mathrm{cap}}_{\geqslant\psi}(y)\right)(\boldsymbol{v}) \mid \langle \boldsymbol{v}, y\rangle = t\right] \\
&= \mathop{\mathbf{E}}_{\boldsymbol{v}\sim\rho}\left[\mathop{\mathbf{E}}_{\boldsymbol{w}\sim\underline{\mathrm{cap}}_{\geqslant\theta}(\boldsymbol{v})}\left[\left(\underline{\mathrm{cap}}_{\geqslant\psi}(y)\right)(\boldsymbol{w})\right] \mid \langle \boldsymbol{v}, y\rangle = t\right] \\
&= \mathop{\mathbf{E}}_{\boldsymbol{v}\sim\rho}\left[\mathop{\mathbf{E}}_{\boldsymbol{w}\sim\underline{\mathrm{cap}}_{\geqslant\theta}(\boldsymbol{v})}\left[\frac{\mathbf{1}[\langle \boldsymbol{w}, y\rangle \geqslant \psi]}{\rho(\mathrm{cap}_{\geqslant\psi})}\right] \mid \langle \boldsymbol{v}, y\rangle = t\right] \\
&= \frac{\mathbf{Pr}_{\boldsymbol{v},\boldsymbol{w}\sim\rho}\left[\langle \boldsymbol{w}, y\rangle \geqslant \psi \mid \langle \boldsymbol{w}, \boldsymbol{v}\rangle \geqslant \theta, \langle \boldsymbol{v}, y\rangle = t\right]}{\mathbf{Pr}_{\boldsymbol{w}\sim\rho}\left[\langle \boldsymbol{w}, y\rangle \geqslant \psi\right]}.
\end{aligned}$$

This ratio is monotone increasing in t, completing the proof of (1).

Now we show Part 2. Claim 3.4.3 shows that by the spherical monotonicity of α, we can express α in its cap decomposition,

$$\alpha = \int_0^1 \underline{\mathrm{cap}}_q \, dr(q),$$

and now by the linearity of convolution, $P_\alpha = \int_0^1 P_{\underline{\mathrm{cap}}_q} dr(q)$, and $P_\alpha\mu = \int P_q\mu \, dr(q)$, $P_\alpha\nu = \int P_q\nu dr(q)$. So, to show that $\ell_{P_\alpha\mu} \succeq_{\mathrm{st}} \ell_{P_\alpha\nu}$, it suffices to argue "slice-by-slice" that for every $q \in [0,1]$, $\ell_{P_q\mu} \succeq_{\mathrm{st}} \ell_{P_q\nu}$.

This follows from the following coupling argument: we sample $(\boldsymbol{x}, \boldsymbol{y})$ from $(\ell_{P_q\mu}, \ell_{P_q\nu})$ in a coupled manner as follows: first, sample $(\boldsymbol{a}_\mu, \boldsymbol{a}_\nu) \sim (\ell_\mu, \ell_\nu)$ in a coupled manner so that $\boldsymbol{a}_\mu \geqslant \boldsymbol{a}_\nu$; such a coupling is guaranteed because $\ell_\mu \succeq_{\mathrm{st}} \ell_\nu$. Next, choose $(\boldsymbol{v}_\mu, \boldsymbol{v}_\nu)$ at random in $\mathbb{S}^{d-1}$ conditioned on $\langle \boldsymbol{v}_\mu, y\rangle = \boldsymbol{a}_\mu$ and $\langle \boldsymbol{v}_\nu, y\rangle = \boldsymbol{a}_\nu$. Now, let $\boldsymbol{\theta}_\mu$ be the random variable $\langle y, \boldsymbol{u}_\mu\rangle$ for $\boldsymbol{u}_\mu \sim \underline{\mathrm{cap}}_q(\boldsymbol{v}_\mu)$, and $\boldsymbol{\theta}_\nu = \langle y, \boldsymbol{u}_\nu\rangle$ for $\boldsymbol{u}_\nu \sim \underline{\mathrm{cap}}_q(\boldsymbol{v}_\nu)$. Note that the marginal over $\boldsymbol{\theta}_\mu$ is $\ell_{P_q\mu}$ and the marginal over $\boldsymbol{\theta}_\nu$ is $\ell_{P_q\nu}$. The probability $\mathbf{Pr}[\theta_\mu > t]$ is proportional to the measure of the intersection of $\mathrm{cap}_{\geqslant t}(y)$ and $\mathrm{cap}_q(\boldsymbol{u}_\mu)$, and similarly the probability $\mathbf{Pr}[\theta_\nu > t]$ is proportional to the measure of the intersection of $\mathrm{cap}_{\geqslant t}(y)$ and $\mathrm{cap}_q(\boldsymbol{u}_\nu)$. By our choice of coupling, the angle between $\boldsymbol{u}_\mu$ and y is smaller than the angle between $\boldsymbol{u}_\nu$ and y, so for every $t \in [-1,1]$,

$$\mathbf{Pr}[\theta_\mu > t] \geqslant \mathbf{Pr}[\theta_\nu > t],$$

and hence we may couple $\boldsymbol{\theta}_\mu$ and $\boldsymbol{\theta}_\nu$ so that $\boldsymbol{\theta}_\mu \geqslant \boldsymbol{\theta}_\nu$ always. Taking $\boldsymbol{x} = \boldsymbol{\theta}_\mu$ and $\boldsymbol{y} = \boldsymbol{\theta}_\nu$ in this coupling gives our conclusion.

Finally, observe that by the commutativity of convolution, $P_\mu\alpha = P_\alpha\mu$ and $P_\nu\alpha = P_\alpha\nu$, and so Part 3 follows from Part 2. □

3.4.1 Concentration of spherical Brownian Motion within a cap

In this section, we study the concentration of Brownian Motion on $\mathbb{S}^{d-1}$ in the spherical cap around its starting point.

Lemma (Restatement of Lemma 3.4.15)**.** *Let $(\boldsymbol{V}_t)_{t\geqslant 0}$ be a Brownian motion on $\mathbb{S}^{d-1}$ starting at V_0. Then for any time $t \geqslant 0$,*

$$\mathbf{Pr}\left[|\langle V_0, \boldsymbol{V}_t\rangle - \exp\left(-(d-1)t\right)| \geqslant x\right] \leqslant 2\exp\left(-\tfrac{d-1}{2}\frac{x^2}{1-e^{-2(d-1)t}}\right).$$

Proof of Lemma 3.4.15. Letting $\boldsymbol{A}_t = \langle V_0, \boldsymbol{V}_t\rangle$ be the correlation of the motion at step t with the starting point, $(\boldsymbol{B}_t)_{t\geqslant 0}$ be standard Brownian motion on $\mathbb{R}^d$, $(\boldsymbol{B}'_t)_{t\geqslant 0}$ be standard Brownian motion on $\mathbb{R}$, and $\theta = d-1$,

$$\begin{aligned} d\boldsymbol{A}_t = \langle V_0, d\boldsymbol{V}_t\rangle &= -\theta\cdot \boldsymbol{A}_t\,dt + \sqrt{2}\big\langle V_0, \left(\mathbb{1} - \boldsymbol{V}_t\boldsymbol{V}_t^\top\right) d\boldsymbol{B}_t\big\rangle \\ &= -\theta\cdot \boldsymbol{A}_t\,dt + \sqrt{2}\left\langle \left(\mathbb{1} - \boldsymbol{V}_t\boldsymbol{V}_t^\top\right)V_0, d\boldsymbol{B}_t\right\rangle \\ &= -\theta\cdot \boldsymbol{A}_t\,dt + \sqrt{2}\sqrt{1-\boldsymbol{A}_t^2}\,d\boldsymbol{B}'_t \end{aligned}$$

The solution to the deterministic differential equation $dx_t = -\theta x_t$ with initial condition $x_0 = 1$ is $x_t = \exp(-\theta t)$. To this end, it's convenient to split $\boldsymbol{A}_t$ up into a deterministic and a random part:

$$\boldsymbol{A}_t = \exp\left(-\theta t\right) + \boldsymbol{R}_t,$$

with the initial condition $R_0 = 0$. Then via calculation,

$$d\boldsymbol{R}_t = -\theta\boldsymbol{R}_t dt + \sqrt{2}\sqrt{1-\boldsymbol{A}_t^2}\,d\boldsymbol{B}'_t. \tag{3.20}$$

We now relate $\boldsymbol{R}_t$ to a stochastic process without drift, as is done, for example, in the analysis of the Ornstein-Uhlenbeck process. Consider $\boldsymbol{R}_t\exp(\theta t)$. Note that

$$\begin{aligned} d(\boldsymbol{R}_t\exp(\theta t)) &= \exp(\theta t)\,d\boldsymbol{R}_t + \boldsymbol{R}_t\,\theta\exp(\theta t)\,dt \\ &= -\boldsymbol{R}_t\,\theta\exp(\theta t)\,dt + \sqrt{2}\exp(\theta t)\sqrt{1-\boldsymbol{A}_t^2}\,d\boldsymbol{B}'_t + \boldsymbol{R}_t\,\theta\exp(\theta t)\,dt \\ &= \sqrt{2}\exp(\theta t)\sqrt{1-\boldsymbol{A}_t^2}\,d\boldsymbol{B}'_t, \end{aligned}$$

a process without drift.

The following version of the Azuma–Hoeffding inequality will allow us to argue that this driftless process concentrates.

Lemma 3.4.20. *Let $(\boldsymbol{X}_t)_{t\geqslant 0} \subset \mathbb{R}$ be a stochastic process adapted to the filtration $\mathcal{F}_t$ with $\mathbf{E}[e^{r\,d\boldsymbol{X}_t} \mid \mathcal{F}_t] < \exp\left(r^2\sigma_t^2 dt\right)$, for all t, r. Then for all $s, x > 0$,*

$$\mathbf{Pr}[|\boldsymbol{X}_s - \boldsymbol{X}_0| \geqslant x] \leqslant 2\exp\left(\frac{-x^2}{4\int_0^s \sigma_t^2 dt}\right).$$

Versions of this lemma are known (c.f. [21] and references therein).
We apply Lemma 3.4.20 to prove that

$$\mathbf{Pr}[|\boldsymbol{R}_s| \geqslant x] \leqslant 2\exp\left(-C' \cdot \frac{x^2 d\theta}{1-\exp(-\theta s)}\right).$$

Indeed, $\boldsymbol{R}_t \exp(\theta t)$ is a stochastic process without drift and satisfies that

$$\begin{aligned}\mathbf{E}[\exp(r\, d(\boldsymbol{R}_t \exp(\theta t)))] &= \mathbf{E}\left[\exp\left(\sqrt{2}r\exp(\theta t)\sqrt{1-\boldsymbol{A}_t^2}\, d\boldsymbol{B}_t'\right)\right] \\ &\leqslant \exp\left(r^2\exp(2\theta t)\cdot\left(1-\boldsymbol{A}_t^2\right)dt\right) \\ &\leqslant \exp\left(r^2\exp(2\theta t)\, dt\right),\end{aligned}$$

Since $\boldsymbol{A}_t$ is real-valued. So we can apply Lemma 3.4.20 to the process and derive that

$$\begin{aligned}\mathbf{Pr}[|\boldsymbol{R}_s| \geqslant x] &= \mathbf{Pr}[|\boldsymbol{R}_s\exp(\theta s)| \geqslant x\exp(\theta s)] \\ &\leqslant 2\exp\left(-\frac{x^2\exp(2\theta s)}{4\int_0^s \exp(2\theta t)dt}\right) \\ &= 2\exp\left(-\frac{\theta x^2}{2\cdot(1-\exp(-2\theta s))}\right),\end{aligned}$$

and plugging in $\theta = d-1$ concludes the proof. □

3.5 The second eigenvalue of links

In this section we analyze the links of the random geometric complex. Each link is a random geometric graph in a cap centered around some $w \in \mathbb{S}^{d-1}$ on $\boldsymbol{m}$ vertices where $\boldsymbol{m} \sim \mathsf{Binom}(n,p)$. We are interested in obtaining a high probability bound on the second eigenvalue of $\widehat{A}_{\boldsymbol{G}} := D_{\boldsymbol{G}}^{-1/2} A_{\boldsymbol{G}} D_{\boldsymbol{G}}^{-1/2}$, the normalized adjacency matrix of link graph $\boldsymbol{G}$, where $A_{\boldsymbol{G}}$ and $D_{\boldsymbol{G}}$ denote its adjacency matrix and diagonal degree matrix. Since the number of vertices $\boldsymbol{m}$ concentrates well in our setting, throughout this section we treat the number of vertices m as fixed and handle the variation in $\boldsymbol{m}$ in Section 3.7. We also specialize the parameters to the regime relevant in proving Theorem 3.1.6 in Section 3.7 — in particular, the relationship between n, p and d is such that $\lim_{n\to\infty}\tau(p,d)$ is a constant in $(0,1)$, np is a polynomially large function of n, and $d = \Omega(\log n)$.

Theorem 3.5.1. *Let $0 < \tau < 1$ be a constant. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \sim \mathrm{cap}_{\geqslant\tau}(w)$ and $\boldsymbol{G} := \mathsf{gg}_\tau(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_m)$. Then for $q := \overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d)}\left(\frac{\tau}{1+\tau}\right)$, suppose $qm \gg \log^8 m \cdot \log^{3/2}\frac{1}{q}\cdot\left(\frac{1+\tau}{\tau}\right)^3$ and $d \geqslant C\cdot\log m$ for any constant $C>0$, then for any constant $\gamma > 0$,*

$$\mathbf{Pr}\left[|\lambda|_2\left(\widehat{A}_{\boldsymbol{G}}\right) > \frac{\tau}{1+\tau} + o_{d,m}(1)\right] \leqslant O\left(m^{-\gamma}\right).$$

To prove Theorem 3.5.1, by Fact 3.2.1 it suffices to bound $\left\| \widehat{A}_{\boldsymbol{G}} - R \right\|$ for any rank-1 PSD matrix R. For a given $\boldsymbol{G}$, the minimizing R for $\left\| \widehat{A}_{\boldsymbol{G}} - R \right\|$ is $R = R_{\boldsymbol{G}} = \frac{D_{\boldsymbol{G}}^{1/2} J D_{\boldsymbol{G}}^{1/2}}{\mathrm{tr}(D_{\boldsymbol{G}})}$ where J is the all-ones matrix.

One challenge in directly performing the trace method on $\widehat{A}_{\boldsymbol{G}} - R_{\boldsymbol{G}}$ is that the degree of any vertex i is a random variable that depends on the locations of all the vectors, and hence introduces extra correlations. In Section 3.3, this issue was resolved because the degrees concentrated very well, and hence $D_{\boldsymbol{G}}^{-1/2}$ and $D_{\boldsymbol{G}}^{1/2}$ were close to scalar multiples of identity. However, in the links the degrees of vertices in $\boldsymbol{G}$ no longer concentrate around a single value, and even the behavior of the *expected* degree of vertex i depends on which "shell" $\boldsymbol{v}_i$ is contained in around w, $\langle \boldsymbol{v}_i, w \rangle$. To better control the degrees, we will study the spectral norm of $\widehat{A}_{\boldsymbol{G}} - R_{\boldsymbol{G}} | \boldsymbol{\kappa}$ conditioned on the shells $\boldsymbol{\kappa} := \{\boldsymbol{\kappa}_i := \langle w, \boldsymbol{v}_i \rangle\}_{i=1}^m$.

Let $D_{\boldsymbol{\kappa}} \in \mathbb{R}^{m \times m}$ be the conditional expected diagonal degree matrix with $D_{\boldsymbol{\kappa}}[i,i] = \mathbf{E}[\deg_{\boldsymbol{G}}(i) \mid \boldsymbol{\kappa}]$. Then we define the new normalized matrix $\underline{A}_{\boldsymbol{G}} = D_{\boldsymbol{\kappa}}{}^{-1/2} A_{\boldsymbol{G}} D_{\boldsymbol{\kappa}}{}^{-1/2}$ and the new conditional rank-1 PSD matrix $R_{\boldsymbol{\kappa}} = \frac{D_{\boldsymbol{\kappa}}{}^{1/2} J D_{\boldsymbol{\kappa}}{}^{1/2}}{\mathrm{tr}(D_{\boldsymbol{\kappa}})}$. Then by optimality of $R_{\boldsymbol{G}}$:

$$\begin{aligned}
\left\| \widehat{A}_{\boldsymbol{G}} - R_{\boldsymbol{G}} \right\| &\leqslant \left\| \widehat{A}_{\boldsymbol{G}} - \left(D_{\boldsymbol{G}}^{-1/2} D_{\boldsymbol{\kappa}}{}^{1/2} \right) R_{\boldsymbol{\kappa}} \left(D_{\boldsymbol{\kappa}}{}^{1/2} D_{\boldsymbol{G}}^{-1/2} \right) \right\| \\
&= \left\| \left(D_{\boldsymbol{G}}^{-1/2} D_{\boldsymbol{\kappa}}{}^{1/2} \right) (\underline{A}_{\boldsymbol{G}} - R_{\boldsymbol{\kappa}}) \left(D_{\boldsymbol{\kappa}}{}^{1/2} D_{\boldsymbol{G}}^{-1/2} \right) \right\| \\
&\leqslant \| \underline{A}_{\boldsymbol{G}} - R_{\boldsymbol{\kappa}} \| \cdot \left\| D_{\boldsymbol{G}}^{-1/2} D_{\boldsymbol{\kappa}}{}^{1/2} \right\|^2.
\end{aligned}$$

Since $\left\| D_{\boldsymbol{G}}^{-1/2} D_{\boldsymbol{\kappa}}{}^{1/2} \right\|^2 = \left\| D_{\boldsymbol{G}}^{-1} D_{\boldsymbol{\kappa}} \right\|$, this is equivalent to bounding

$$\| \underline{A}_{\boldsymbol{G}} - R_{\boldsymbol{\kappa}} \| \cdot \left\| D_{\boldsymbol{G}}^{-1} D_{\boldsymbol{\kappa}} \right\| \leqslant \| \underline{A}_{\boldsymbol{G}} - R_{\boldsymbol{\kappa}} \| \cdot \max_{i \in [m]} \frac{D_{\boldsymbol{\kappa}}[i,i]}{D_{\boldsymbol{G}}[i,i]}$$

Now, in the trace method it is convenient to work with $A_{\boldsymbol{G}} - \mathbf{E}[A_{\boldsymbol{G}}] \mid \boldsymbol{\kappa}$, which is not a rank-1 matrix. So, applying the triangle inequality,

$$\leqslant \left(\| \underline{A}_{\boldsymbol{G}} - \mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \boldsymbol{\kappa}] \| + \| \mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \boldsymbol{\kappa}] - R_{\boldsymbol{\kappa}} \| \right) \cdot \max_{i \in [m]} \frac{D_{\boldsymbol{\kappa}}[i,i]}{D_{\boldsymbol{G}}[i,i]}, \quad (3.21)$$

It then suffices to bound $\| \mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \boldsymbol{\kappa}] - R_{\boldsymbol{\kappa}} \|$, $\max_{i \in [m]} \frac{D_{\boldsymbol{\kappa}}[i,i]}{D_{\boldsymbol{G}}[i,i]}$, and $\| \underline{A}_{\boldsymbol{G}} - \mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \boldsymbol{\kappa}] \|$ to complete the proof of Theorem 3.5.1.

In Section 3.6 we'll show that $\mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \boldsymbol{\kappa}]$ is close to $R_{\boldsymbol{G}}$ in spectral norm:

Lemma 3.5.2. *If $d \geqslant C \cdot \log m$ for some constant $C > 0$ and the constant $\tau \in (0,1)$ satisfies $qm \gg \log^8 m$, then*

$$\| \mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \boldsymbol{\kappa}] - R_{\boldsymbol{\kappa}} \| \leqslant O\left(\sqrt{\frac{\log^2 d}{d}} \right)$$

with probability at least $1 - o(m^{-\gamma})$ for any constant $\gamma > 0$.

And the remainder of this section will be devoted to bounding the other two quantities, as follows:

Lemma 3.5.3. *For any* $0 < \alpha < 1$,

$$\max_{i \in [m]} \frac{D_\kappa[i,i]}{D_{\boldsymbol{G}}[i,i]} \leqslant \frac{1}{1-\alpha},$$

with probability at least $1 - m \cdot \exp\left(-\frac{\alpha^2 q(m-1)}{4}\right)$.

Lemma 3.5.4. *For any* κ, q *and* m *and* $qm \gg \log^8 m \cdot \log^{3/2} \frac{1}{q} \cdot \left(\frac{1+\tau}{\tau}\right)^3$

$$\|\underline{A}_{\boldsymbol{G}} - \mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \kappa]\| \leqslant (1 + o_m(1)) \cdot \frac{\tau}{1+\tau}.$$

In service of proving Lemma 3.5.2, Lemma 3.5.3 and Lemma 3.5.4, we need the following fact that arises in studying random geometric graphs with shifted edge-connectivity thresholds.

Definition 3.5.5. We define the bivariate function $T(x,y) := \frac{\tau - xy}{\sqrt{(1-x^2)(1-y^2)}}$ as the *shifted threshold* function, defined so that

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x},\boldsymbol{y} \sim \mathbb{S}^{d-2}} [\langle \boldsymbol{x}, \boldsymbol{y} \rangle \geqslant T(x,y)] = \mathop{\mathbf{Pr}}_{\boldsymbol{u},\boldsymbol{v} \sim \mathbb{S}^{d-1}} [\langle \boldsymbol{u}, \boldsymbol{v} \rangle \geqslant \tau \mid \langle \boldsymbol{u}, w \rangle = x, \langle \boldsymbol{v}, w \rangle = y].$$

Claim 3.5.6. The shifted threshold function $T(x,y) := \frac{\tau - xy}{\sqrt{(1-x^2)(1-y^2)}}$ on the domain $x, y \in [\tau, 1]$ is maximized when $x = y = \tau$, and achieves value $\frac{\tau}{1+\tau}$. Additionally $\partial_x T(x,y)$ and $\partial_y T(x,y)$ are both negative.

Proof. The derivatives $\partial_y T(x,y) = \frac{\tau}{\sqrt{1-x^2}} \cdot g(y) - \frac{x}{\sqrt{1-x^2}} \cdot h(y)$ and $\partial_x T(x,y) = \frac{\tau}{\sqrt{1-y^2}} \cdot g(x) - \frac{y}{\sqrt{1-y^2}} \cdot h(x)$, where $g(z) := \frac{z}{(1-z^2)^{3/2}}$ and $h(z) := \frac{1}{\sqrt{1-z^2}} + \frac{z^2}{(1-z^2)^{3/2}}$. Since $g(z) < h(z)$ for $z \in (0,1]$, then for $x, y \geqslant \tau$ we deduce that $\partial_y T, \partial_x T < 0$. Therefore, T achieves the maximum value $\frac{\tau - \tau^2}{1-\tau^2} = \frac{\tau}{1+\tau}$ when $x = y = \tau$. □

Now we prove Lemma 3.5.3.

Proof of Lemma 3.5.3. For any $\alpha \in (0,1)$, consider the event that $\max_{i \in [m]} \frac{D_\kappa[i,i]}{D_{\boldsymbol{G}}[i,i]} > \frac{1}{1-\alpha}$. We can bound the probability that this event happens by union bound and Bernstein's inequality:

$$\begin{aligned}\mathbf{Pr}\left[\exists i \in [m],\ D_{\boldsymbol{G}}[i,i] \leqslant (1-\alpha) D_\kappa[i,i]\right] &\leqslant \sum_{i=1}^m \mathbf{Pr}[D_{\boldsymbol{G}}[i,i] \leqslant (1-\alpha) D_\kappa[i,i]] \\ &\leqslant m \cdot \max_i \exp\left(-\frac{1}{2} \cdot \frac{\alpha^2 D_\kappa[i,i]^2}{(\alpha+1) D_\kappa[i,i]}\right)\end{aligned}$$

$$\leqslant m \cdot \max_i \exp\left(-\frac{\alpha^2 D_\kappa[i,i]}{4}\right)$$

Observe that $D_\kappa[i,i] = \sum_{j\neq i} \overline{\Phi}_{\mathsf{D_{ip}}(d-1)}(T(\kappa_i,\kappa_j))$. By Claim 3.5.6, $T(\kappa_i,\kappa_j) \leqslant \frac{\tau}{1+\tau}$, so $\overline{\Phi}_{\mathsf{D_{ip}}(d-1)}(T(\kappa_i,\kappa_j)) \geqslant \overline{\Phi}_{\mathsf{D_{ip}}(d-1)}\left(\frac{\tau}{1+\tau}\right) = q$. Consequently, $D_\kappa[i,i] \geqslant q(m-1)$ from which the desired statement follows. □

3.5.1 Spectral norm bound for centered links

In the rest of the section, we prove Lemma 3.5.4 by bounding the expected trace

$$\mathbf{E}\left[\operatorname{tr}\left((\underline{A}_{\boldsymbol{G}} - \mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \kappa])^\ell\right)\right],$$

for $\kappa \in [\tau,1]^m$ a fixed configuration of shells. The proof will be almost identical to the one in Section 3.3, but here we have to deal with the fact that the graph is not vertex-transitive.

Proof of Lemma 3.5.4. First observe:

$$\begin{aligned}\mathbf{E}\left[\operatorname{tr}\left((\underline{A}_{\boldsymbol{G}} - \mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \kappa])^\ell\right)\right] &= \mathbf{E}\left[\operatorname{tr}\left(\left({D_\kappa}^{-1/2} A_{\boldsymbol{G}} {D_\kappa}^{-1/2} - {D_\kappa}^{-1/2}\,\mathbf{E}[A_{\boldsymbol{G}}|\kappa]{D_\kappa}^{-1/2}\right)^\ell\right)\right] \\ &= \mathbf{E}\left[\operatorname{tr}\left(\left({D_\kappa}^{-1} A_{\boldsymbol{G}} - {D_\kappa}^{-1}\,\mathbf{E}[A_{\boldsymbol{G}}|\kappa]\right)^\ell\right)\right].\end{aligned}$$

We rewrite the expression in terms of ${D_\kappa}^{-1}A_{\boldsymbol{G}}$ which approximates the transition matrix of the random walk on $\boldsymbol{G}$.[8] Next, we expand the expression in terms of walks in $\mathcal{K}_m$.

Following the convention of Section 3.3, we use $\mathcal{W}_\ell$ to denote the collection of length-ℓ walks in $\mathcal{K}_m$. For every $W \in \mathcal{W}_\ell$, use $G(W) = (V(W), E(W))$ to denote the multigraph obtained by the vertices and edges used in W. Use $m(e)$ to denote the number of times that an edge e appears in the walk W.

Definition 3.5.7. We also introduce the following notation. Let $d_W(\kappa) := \prod_{(i_t,i_{t+1})\in W} {D_\kappa}^{-1}[i_t,i_t]$ denote the normalization constant along the path W conditioned on the shells κ. Also define $p_e = \mathbf{E}[\mathbf{1}[e \in \boldsymbol{G}] \mid \kappa]$ to be the probability that an edge e exists conditioned on κ.

Then:

$$\mathbf{E}\left[\operatorname{tr}\left((\underline{A}_{\boldsymbol{G}} - \mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \kappa])^\ell\right) \mid \kappa\right] = \sum_{W\in\mathcal{W}_\ell} d_W(\kappa)\cdot \mathbf{E}\left[\prod_{e\in E(W)} (\mathbf{1}[e\in\boldsymbol{G}] - p_e)^{m(e)} \mid \kappa\right] \tag{3.22}$$

Next we apply the decomposition in Section 3.3 to $G(W)$ and obtain the 2-core graph $G_2(W)$ and the forest graph $G_1(W)$. Since conditioned on the vectors $\boldsymbol{v}_i \in V_2(W)$ the events

[8] If D_κ were not the *expected* degree matrix but rather the exact degree matrix of $\boldsymbol{G}$, we would have a true transition matrix here.

$e \in \boldsymbol{G}$ are independent for all $e \in E_1(W)$, the expectation in (3.22) can be decomposed into two parts:

$$\mathbf{E}\left[\prod_{e\in E(W)} (\mathbf{1}[e \in \boldsymbol{G}] - p_e)^{m(e)} \mid \kappa\right]$$
$$= \prod_{e\in E_1(W)} \mathbf{E}\Big[(\mathbf{1}[e \in \boldsymbol{G}] - p_e)^{m(e)} \mid \kappa\Big] \mathop{\mathbf{E}}_{\substack{\boldsymbol{v}_i\\ i\in V_2(W)}}\left[\prod_{e\in E_2(W)} (\mathbf{1}[e \in \boldsymbol{G}] - p_e)^{m(e)} \mid \kappa\right] \tag{3.23}$$

We bound the contribution from the edges in $E_2(W)$ by further spliting $G_2(W)$ into paths consisting of degree-2 vertices and the junction graph $G_J(W) = (J(W), E_J(W))$ as defined in Definition 3.3.5. As in Section 3.3 the key observation here is that conditioned on vertices in $J(W)$ the contributions from the paths of degree-2 vertices are all independent from each other:

$$\left|\mathop{\mathbf{E}}_{\substack{\boldsymbol{v}_i\\ i\in V_2(W)}}\left[\prod_{e\in E_2(W)} (\mathbf{1}[e \in \boldsymbol{G}] - p_e)^{m(e)} \mid \kappa\right]\right| \leqslant$$
$$\mathop{\mathbf{E}}_{\substack{\boldsymbol{v}_i\\ i\in J(W)}} \prod_{f\in E_{J(W)}} \left|\mathop{\mathbf{E}}_{\substack{\boldsymbol{v}_i\\ i\in \gamma(f)\setminus J(W)}}\left[\prod_{e\in\gamma(f)} (\mathbf{1}[e \in \boldsymbol{G}] - p_e)^{m(e)} \mid \kappa\right]\right| \tag{3.24}$$

Now, let $X_{\kappa,\kappa'}$ be the transition operator for the random step that walks from vector v in $\text{shell}_{=\kappa}(w)$ to a uniformly random vector $\boldsymbol{v'}$ in $\text{shell}_{=\kappa'}(w) \cap \text{cap}_{\geqslant\tau}(v)$. Like in Section 3.3, we use $\gamma(f) = (f_0, f_1, \ldots, f_{\ell(f)})$ to identify the walk in $G_2(W)$ corresponding to the edge $f \in E_{J(W)}$. We denote the edge (f_i, f_{i+1}) with $\gamma_i(f)$. We simplify the contribution from each path $\gamma(f)$ where $f \in E_J(W)$ as follows:

$$\left|\mathop{\mathbf{E}}_{\boldsymbol{v}_i : i\in\gamma(f)\setminus J(W)}\left[\prod_{e\in\gamma(f)} (\mathbf{1}[e \in \boldsymbol{G}] - p_e)^{m(e)} \mid \kappa\right]\right|$$
$$=\left|\mathop{\mathbf{E}}_{\boldsymbol{v}_i : i\in\gamma(f)\setminus J(W)}\left[\prod_{e\in\gamma(f)} \Big(\mathbf{1}[e \in \boldsymbol{G}] \cdot \big((1-p_e)^{m(e)} - (-p_e)^{m(e)}\big) + (-p_e)^{m(e)}\Big) \mid \kappa\right]\right|$$
$$=\left|\sum_{T\subseteq\gamma(f)} \mathop{\mathbf{E}}_{\boldsymbol{v}_i : i\in\gamma(f)\setminus J(W)}\left[\prod_{e\in T} \mathbf{1}[e \in \boldsymbol{G}] \cdot \big((1-p_e)^{m(e)} - (-p_e)^{m(e)}\big) \prod_{e\in\gamma(f)\setminus T} (-p_e)^{m(e)} \mid \kappa\right]\right|$$
$$\leqslant\left|\prod_{e\in\gamma(f)} \big(p_e(1-p_e)^{m(e)} + (1-p_e)(-p_e)^{m(e)}\big)\right|$$

$$+\left|\prod_{e\in\gamma(f)}\left((1-p_e)^{m(e)}-(-p_e)^{m(e)}\right)\cdot\prod_{i=0}^{\ell(f)-2}p_{\gamma_i(f)}\cdot\left(\left\langle\prod_{i=0}^{\ell(f)-2}X_{\kappa_{f_i},\kappa_{f_{i+1}}}\delta_{\boldsymbol{v}_{f_0}},\mathrm{cap}_{p_{\gamma_{\ell(f)-1}(f)}}(\boldsymbol{v}_{f_{\ell(f)}})\right\rangle-p_{\gamma_{\ell(f)-1}(f)}\right)\right| \tag{3.25}$$

Let $S_J(W)\subseteq E_J(W)$ be the set of edges f such $m(e)=1$ for some $e\in\gamma(f)$, and $D_J(W)=E_J(W)\setminus S_J(W)$. For any $f\in S_J(W)$ the first term in (3.25) vanishes, while for any $f\in D_J(W)$

$$\left|\prod_{e\in\gamma(f)}\left(p_e(1-p_e)^{m(e)}+(1-p_e)(-p_e)^{m(e)}\right)\right|\leqslant\prod_{e\in\gamma(f)}\left(p_e(1-p_e)^2+(1-p_e)p_e^2\right)\leqslant\prod_{e\in\gamma(f)}p_e.$$

Therefore we can derive the following bound on the contribution from the 2-core graph.

$$\begin{aligned}&(3.24)\\&\leqslant\mathop{\mathbf{E}}_{\substack{\boldsymbol{v}_i\\ i\in J(W)}}\Bigg[\prod_{f\in D_J(W)}\prod_{i=0}^{\ell(f)-2}p_{\gamma_i(f)}\cdot\left(\left|\left\langle\prod_{i=0}^{\ell(f)-2}X_{\kappa_{f_i},\kappa_{f_{i+1}}}\delta_{\boldsymbol{v}_{f_0}},\mathrm{cap}_{p_{\gamma_{\ell(f)-1}(f)}}(\boldsymbol{v}_{f_{\ell(f)}})\right\rangle-p_{\gamma_{\ell(f)-1}(f)}\right|\right.\\&\quad\left.+p_{\gamma_{\ell(f)-1}(f)}\right)\cdot\prod_{f\in S_J(W)}\prod_{i=0}^{\ell(f)-2}p_{\gamma_i(f)}\cdot\left|\left\langle\prod_{i=0}^{\ell(f)-2}X_{\kappa_{f_i},\kappa_{f_{i+1}}}\delta_{\boldsymbol{v}_{f_0}},\mathrm{cap}_{p_{\gamma_{\ell(f)-1}(f)}}(\boldsymbol{v}_{f_{\ell(f)}})\right\rangle-p_{\gamma_{\ell(f)-1}(f)}\right|\\&\quad\mid\kappa\Bigg]\end{aligned}$$

To bound the absolute value terms, we take an arbitrary spanning tree $T_J(W)$ of $G_J(W)$, and bound the absolute value differently depending on whether $f\in T_J(W)$ or not.

To bound this expectation, let $T_J(W)$ be a spanning tree of $G_J(W)$. For every edge not in $T_J(W)$, we apply a worst-case bound. To state this bound, we define $C\coloneqq\sqrt{\frac{1}{2}\log\frac{1}{q}}\cdot\left(\frac{1+\tau}{\tau}\right)$ and $\lambda\coloneqq\frac{\tau}{1+\tau}$.

Claim 3.5.8. For every shell configuration $\kappa\in[\tau,1]^m$ and non-tree edge $f\in E_J(W)\setminus T_J(W)$, we have that

$$\left|\left\langle\prod_{i=0}^{\ell(f)-2}X_{\kappa_{f_i},\kappa_{f_{i+1}}}\delta_{\boldsymbol{v}_{f_0}},\mathrm{cap}_{p_{\gamma_{\ell(f)-1}(f)}}(\boldsymbol{v}_{f_{\ell(f)}})\right\rangle-p_{\gamma_{\ell(f)-1}(f)}\right|\leqslant C\cdot\lambda^{\ell(f)-1}$$

Proof. To prove the claim, we first need to understand the random variable

$$\left\langle\prod_{i=0}^{\ell(f)-2}X_{\kappa_{f_i},\kappa_{f_{i+1}}}\delta_{\boldsymbol{v}_{f_0}},\mathrm{cap}_{p_{\gamma_{\ell(f)-1}(f)}}(\boldsymbol{v}_{f_{\ell(f)}})\right\rangle.$$

Recall that at time step i the operator $X_{\kappa_{f_i},\kappa_{f_{i+1}}}$ denotes the random step that takes a vector $\boldsymbol{v}_{f_i} = \kappa_{f_i} \cdot w + \sqrt{1-\kappa_{f_i}^2} \cdot \boldsymbol{z}_i$ and outputs $\boldsymbol{v}_{f_{i+1}} := \kappa_{f_{i+1}} \cdot w + \sqrt{1-\kappa_{f_{i+1}}^2} \cdot \boldsymbol{z}_{i+1}$ where $\boldsymbol{z}_{i+1}$ is a uniformly random unit vector orthogonal to w such that

$$\left\langle \boldsymbol{v}_{f_i}, \boldsymbol{v}_{f_{i+1}} \right\rangle \geqslant \tau.$$

This is equivalent to

$$\kappa_{f_i}\kappa_{f_{i+1}} + \sqrt{(1-\kappa_{f_i}^2)(1-\kappa_{f_{i+1}}^2)} \cdot \langle \boldsymbol{z}_i, \boldsymbol{z}_{i+1} \rangle \geqslant \tau,$$

which can then be rearranged as

$$\langle \boldsymbol{z}_i, \boldsymbol{z}_{i+1} \rangle \geqslant T\left(\kappa_{f_i}\kappa_{f_{i+1}}\right) := \frac{\tau - \kappa_{f_i}\kappa_{f_{i+1}}}{\sqrt{(1-\kappa_{f_i}^2)(1-\kappa_{f_{i+1}}^2)}}.$$

In particular, we are choosing $\boldsymbol{z}_{i+1}$ in the $p_{\gamma_i(f)}$-cap of $\boldsymbol{z}_i$ within the $d-2$ dimensional unit sphere orthogonal to w. So the operator $\prod_{i=0}^{\ell(f)-2} X_{\kappa_{f_i},\kappa_{f_{i+1}}}$ can be decomposed into its action in the span of w and that in the space orthogonal to w. The action in the span of w conditioned on κ is deterministic. Orthogonal to w, it is the operator $\prod_{i=0}^{\ell(f)-2} P_{p_{\gamma_i(f)}}$ on $\mathbb{S}^{d-2}$. Thus the quantity we are interested in understanding is the same as

$$\left\langle \prod_{i=0}^{\ell(f)-2} P_{p_{\gamma_i(f)}} \delta_{\boldsymbol{z}_0}, \mathrm{cap}_{p_{\gamma_{\ell(f)-1}}}(\boldsymbol{z}_{\ell(f)}) \right\rangle.$$

Now, observe that:

$$\left| \left\langle \prod_{i=0}^{\ell(f)-2} P_{p_{\gamma_i(f)}} \delta_{\boldsymbol{z}_0}, \mathrm{cap}_{p_{\gamma_{\ell(f)-1}}}(\boldsymbol{z}_{\ell(f)}) \right\rangle - p_{\gamma_{\ell(f)-1}} \right| \leqslant \mathrm{d_{TV}} \left(\prod_{i=0}^{\ell(f)-2} P_{p_{\gamma_i(f)}} \delta_{\boldsymbol{z}_0}, \rho \right).$$

Recall that $p_{\gamma_i(f)} = \Phi_{\mathsf{D_{ip}}(d-1)}\left(\tau_{\kappa_i,\kappa_{i+1}}\right)$, which by Claim 3.5.6 is minimized when $\tau_{\kappa_i,\kappa_{i+1}} = \frac{\tau}{1+\tau}$, which means $p_{\gamma_i(f)} \geqslant q$. Thus, by Claim 3.4.11, Lemma 3.4.13, and Lemma 3.4.12, which make concrete the intuition that applying P_q should only mix slower than applying $P_{q'}$ for $q' \geqslant q$, we have:

$$\begin{aligned} \left| \left\langle \prod_{i=0}^{\ell(f)-2} P_{p_{\gamma_i(f)}} \delta_{\boldsymbol{z}_0}, \mathrm{cap}_{p_{\gamma_{\ell(f)-1}}}(\boldsymbol{z}_{\ell(f)}) \right\rangle - p_{\gamma_{\ell(f)-1}} \right| &\leqslant \mathrm{d_{TV}} \left(P_q^{\ell(f)-1} \delta_{\boldsymbol{z}_0}, \rho \right) \\ &= \mathrm{d_{TV}} \left(\frac{1}{q} P_q^{\ell(f)-2} \mathrm{cap}_q \boldsymbol{z}_0, \rho \right). \end{aligned}$$

Then by Theorem 3.4.6, the above is

$$\leqslant \sqrt{\frac{1}{2}\log\frac{1}{q}} \cdot \left(\frac{\tau}{1+\tau}\right)^{\ell(f)-2} = \sqrt{\frac{1}{2}\log\frac{1}{q}} \cdot \frac{1+\tau}{\tau} \cdot \left(\frac{\tau}{1+\tau}\right)^{\ell(f)-1} = C \cdot \lambda^{|\gamma(f)|-1},$$

which completes the proof. $\square$

Next we bound the contribution of a tree edge $f \in T_J(W)$ using the following claim whose proof is identical to that of Claim 3.3.7.

Claim 3.5.9. For every shell vector κ and tree edge $f \in T_J(W)$, we have that

$$\mathop{\mathbf{E}}_{\substack{\boldsymbol{v}_i \\ i\in J(W)}} \prod_{f\in T_J(W)} \prod_{i=0}^{\ell(f)-2} p_{\gamma_i(f)} \cdot \left(\left| \left\langle \prod_{i=0}^{\ell(f)|-1} X_{\kappa_{f_i},\kappa_{f_{i+1}}} \delta_{\boldsymbol{v}_{f_0}}, \mathrm{cap}_{p_{\gamma_{\ell(f)-1}(f)}}(\boldsymbol{v}_{f_{\ell(f)}}) \right\rangle - p_{\gamma_{\ell(f)-1}(f)} \right| \right.$$

$$\left. + p_{\gamma_{\ell(f)-1}(f)} \cdot \mathbf{1}[f \in D_J(W)] \right)$$

$$\leqslant \prod_{f\in T_J(W)} \prod_{i=0}^{\ell(f)-1} p_{\gamma_i(f)} \cdot \left(2C\lambda^{\ell(f)} + \mathbf{1}[f \in D_J(W)]\right).$$

Combining the two bounds for different edges in $G_J(W)$ to obtain the simplified bound for (3.24):

$$(3.24) \leqslant \prod_{f\in T_J(W)} \prod_{i=0}^{\ell(f)-1} p_{\gamma_i(f)} \left(2C\lambda^{\ell(f)} + \mathbf{1}[f \in D_J(W)]\right) \cdot$$

$$\cdot \prod_{f\in E_J(W)\setminus T_J(W)} \prod_{i=0}^{\ell(f)-2} p_{\gamma_i(f)} \cdot \left(C\lambda^{\ell(f)-1} + p_{\gamma_{\ell(f)-1}(f)-1} \mathbf{1}[f \in D_J(W)] \right)$$

We now recall some notation from Section 3.3. We use $e(W)$ to denote $|E(W)|$, $\mathrm{sing}(W)$ for the number of singleton edges in $G_2(W)$, and $\mathrm{exc}(G)$ for $|E(G)| - (|V(G)| - 1)$, the number of edges G has more than a tree. The relations between these variable are already shown in Observation 3.3.9 and Claim 3.3.10. So here we directly apply these results to get that

$$(3.24) \leqslant \prod_{e\in E_2(W)} p_e \cdot \prod_{f\in E_J(W)\setminus T_J(W)} p^{-1}_{\gamma_{\ell(f)-1}(f)} \cdot \lambda^{\mathrm{sing}(W)-\mathrm{exc}(W)} \cdot (3C)^{|E_J(W)|}$$

$$\leqslant \prod_{e\in E_2(W)} p_e \cdot \prod_{f\in E_J(W)\setminus T_J(W)} p^{-1}_{\gamma_{\ell(f)-1}(f)} \cdot \lambda^{\mathrm{sing}(W)-\mathrm{exc}(W)} \cdot (3C)^{3\mathrm{exc}(W)}$$

last inequality by Observation 3.3.9

Therefore

$$(3.23) \leqslant \prod_{e\in E_1(W)} \mathbf{E}\left[(\mathbf{1}[e \in \boldsymbol{G}] - p_e)^{m(e)} \mid \kappa \right] \cdot \prod_{e\in E_2(W)} p_e \cdot$$

$$\prod_{f\in E_J(W)\setminus T_J(W)} p^{-1}_{\gamma_{\ell(f)-1}(f)} \cdot \lambda^{\mathrm{sing}(W)-\mathrm{exc}(W)} \cdot (3C)^{3\mathrm{exc}(W)}$$

$$\leqslant \prod_{e\in E_1(W)} p_e \cdot \prod_{e\in E_2(W)} p_e \cdot \prod_{f\in E_J(W)\setminus T_J(W)} p^{-1}_{\gamma_{\ell(f)-1}(f)} \cdot \lambda^{\mathrm{sing}(W)-\mathrm{exc}(W)} \cdot (3C)^{3\mathrm{exc}(W)}$$

$$\leqslant \prod_{e \in E(W)} p_e \cdot \prod_{f \in E_J(W) \setminus T_J(W)} p^{-1}_{\gamma_{\ell(f)-1}(f)} \cdot \lambda^{\mathrm{sing}(W)} \left(\frac{27C^3}{\lambda} \right)^{\mathrm{exc}(W)} \tag{3.26}$$

Before finally bounding the trace power, we make the following observations.

Observation 3.5.10. As a consequence of Claim 3.5.6 for all $i \in [m]$, the expected degree of vertex i satisfies

$$D_\kappa[i,i] = \mathbf{E}[\deg_{\boldsymbol{G}}(i) \mid \kappa] \geqslant (m-1) \cdot q$$

We define Struc_ℓ to be the set of distinct unlabelled walks of length ℓ. Then as a corollary of Claim 3.3.10, we have

Corollary 3.5.11. *The number of unlabelled walks $U \in \mathsf{Struc}_\ell$ such that $e(U) = a, s(U) = b$, and $\mathrm{exc}(U) = c$ is at most:*

$$\ell^{3(\ell-b)} \cdot \ell^{2c}.$$

The result follows by observing that since U is unlabelled, we can remove the m^{a-c+1} term that counts the number of distinct labelings in Claim 3.3.10.

For an unlabeled walk U and labeled walk W, we say $W \sim U$ if W is a labeling of U in $[m]$.

Claim 3.5.12. For any unlabelled walk U we have that

$$\sum_{W \sim U} d_W(\kappa) \prod_{e \in E(W)} p_e \cdot \prod_{f \in E_J(W) \setminus T_J(W)} p^{-1}_{\gamma_{\ell(f)-1}(f)} \leqslant (m \cdot q)^{-\mathrm{exc}(U) - \frac{\ell - s(U)}{2}}$$

Proof. For each $W \sim U$ we use $i_1, \ldots, i_a \in [m]$ to denote the label of each vertex in W in the order of visit. Then we construct the canonical spanning tree $T(W)$ by adding each directed edge in the order of W as long as the edge goes to an unvisited vertex. Use $\mathrm{Par}(i_j)$ to denote the parent of vertex i_j. Then the j-th edge of $T(W)$ is $(\mathrm{Par}(i_{j+1}), i_{j+1})$, and use $T(W)^{(j)}$ to denote the tree consisting of the first j edges of $T(W)$. Then i_{j+1} is always a leaf in $T(W)^{(j)}$.

$T(W)$ gives rise to a canonical spanning tree $T_J(W)$ in the contracted graph $G_J(W)$: an edge f is in $T_J(W)$ if and only if every edge in the path $\gamma(f)$ is in $T(W)$. From this fact we can deduce that

$$T(W) = E(W) \setminus \left\{ \gamma_{\ell(f)-1}(f) : f \in E_J(W) \setminus T_J(W) \right\}.$$

Therefore, using Observation 3.5.10 we can take a loose upper bound on the contribution of edges outside of $T(W)$ and write

$$\begin{aligned} &\sum_{W \sim U} d_W(\kappa) \prod_{e \in E(W)} p_e \cdot \prod_{f \in E_J(W) \setminus T_J(W)} p^{-1}_{\gamma_{\ell(f)-1}(f)} \\ &\leqslant \sum_{W \sim U} ((m-1) \cdot q)^{-\ell + (|V(U)|-1)} \prod_{(i,j) \in T(W)} \frac{p_{i,j}}{D_\kappa[i,i]} \end{aligned}$$

$$\leqslant ((m-1)\cdot q)^{-\ell+(|V(U)|-1)} \sum_{i_1,\ldots,i_a \in [m]} \prod_{j=2}^{a} \frac{p_{\mathrm{Par}(i_j),i_j}}{D_\kappa[\mathrm{Par}(i_j),\mathrm{Par}(i_j)]} \tag{3.27}$$

where $a = |V(U)|$. Next we show by induction on a that

$$\sum_{i_1,\ldots,i_a} \prod_{j=2}^{a} \frac{p_{(\mathrm{Par}(i_j),i_j)}}{D_\kappa[\mathrm{Par}(i_j),\mathrm{Par}(i_j)]} = 1$$

The base case $a = 1$ is true by definition. Suppose this is true for $a-1$. Then:

$$\sum_{i_1,\ldots,i_a} \prod_{j=2}^{a} \frac{p_{(\mathrm{Par}(i_j),i_j)}}{D_\kappa[\mathrm{Par}(i_j),\mathrm{Par}(i_j)]} = \sum_{i_1,\ldots,i_{a-1}} \prod_{j=2}^{a-1} \frac{p_{(\mathrm{Par}(i_j),i_j)}}{D_\kappa[\mathrm{Par}(i_j),\mathrm{Par}(i_j)]} \cdot \sum_{i_a=1}^{m} \frac{p_{(\mathrm{Par}(i_a),i_a)}}{D_\kappa[\mathrm{Par}(i_a),\mathrm{Par}(i_a)]}$$

By definition $\sum_{i_a} \frac{p_{(\mathrm{Par}(i_a),i_a)}}{D_\kappa[\mathrm{Par}(i_a),\mathrm{Par}(i_a)]} = 1$, so we have:

$$= \sum_{i_1,\ldots,i_{a-1}} \prod_{j=2}^{a-1} \frac{p_{(\mathrm{Par}(i_j),i_j)}}{D_\kappa[\mathrm{Par}(i_a),\mathrm{Par}(i_a)]} \cdot 1 = 1$$

Finally, observe that $\ell - (|V(U)|-1) \geqslant \mathrm{exc}(U) + \frac{\ell - s(U)}{2}$ is at least the number of steps that use a previously walked-on edge. The way to see this is to observe that the quantity $\ell - (|V(U)|-1)$ counts the number of steps to a previously visited vertex. Such a step can either (1) use an excess edge for the first time, of which there are $\mathrm{exc}(U)$ steps, or (2) use a previously walked-on edge, which must be at least half the steps that do not use a singleton edge, i.e. at least $\frac{\ell - s(U)}{2}$ steps. Thus we conclude that (3.27) $\leqslant ((m-1)\cdot q)^{-\ell+(|V(U)|-1)} \leqslant ((m-1)\cdot q)^{-\mathrm{exc}(U) - \frac{\ell - s(U)}{2}}$. □

Now we are finally already to bound the expected trace power. Plugging (3.26) into (3.22) gives:

$$\begin{aligned}
(3.22) &= \sum_{U \in \mathsf{Struc}_\ell} \sum_{W \sim U} d_W(\kappa) \prod_{e \in E(W)} p_e \cdot \prod_{f \in E_J(W) \setminus T_J(W)} p^{-1}_{\gamma_{\ell(f)-1}(f)} \cdot \lambda^{\mathrm{sing}(W)} \left(\frac{27C^3}{\lambda}\right)^{\mathrm{exc}(W)} \\
&\leqslant \sum_{U \in \mathsf{Struc}_\ell} ((m-1)\cdot q)^{-\mathrm{exc}(U) - \frac{\ell - s(U)}{2}} \cdot \lambda^{s(U)} \left(\frac{27C^3}{\lambda}\right)^{\mathrm{exc}(U)} \qquad \text{by Claim 3.5.12} \\
&= \sum_{a=1}^{\ell} \sum_{b=1}^{\ell} \sum_{c=1}^{\ell} \sum_{\substack{U \in \mathsf{Struc}_\ell \\ e(U)=a,\, s(U)=b,\, \mathrm{exc}(U)=c}} ((m-1)\cdot q)^{-\frac{\ell-b}{2}} \cdot \lambda^b \left(\frac{27C^3}{\lambda q(m-1)}\right)^c \\
&= \sum_{a=1}^{\ell} \sum_{b=1}^{\ell} \sum_{c=1}^{\ell} ((m-1)\cdot q)^{-\frac{\ell-b}{2}} \cdot \lambda^b \left(\frac{27C^3}{\lambda q(m-1)}\right)^c \cdot \ell^{2(\ell-b)} \cdot \ell^{2c} \qquad \text{by Claim 3.5.11}
\end{aligned}$$

$$= \ell \sum_{b=1}^{\ell} \sum_{c=1}^{\ell} \left(\frac{\ell^2}{\sqrt{(m-1) \cdot q}} \right)^{\ell - b} \cdot \lambda^b \left(\frac{27C^3\ell^2}{\lambda q(m-1)} \right)^c$$
$$= \ell^3 \max\left(1, \left(\frac{27C^3\ell^2}{\lambda q(m-1)} \right)^{\ell}\right) \cdot \max\left(\lambda^{\ell}, \left(\frac{\ell^2}{\sqrt{(m-1) \cdot q}} \right)^{\ell}\right)$$

By choosing $\ell = \log^2 m$, we can conclude that with probability at least $1 - m^{-\gamma}$,

$$\|\underline{A}_{\boldsymbol{G}} - \mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \kappa]\| \leqslant (1 + o(1)) \cdot \left(1 + \frac{27C^3 \log^4 m}{\lambda q m}\right) \cdot \max\left\{\lambda, \frac{\log^4 m}{\sqrt{qm}}\right\}$$

Since $qm \gg \log^8 m \cdot \log^{3/2} \frac{1}{q} \cdot \left(\frac{1+\tau}{\tau}\right)^3$, we have:

$$\|\underline{A}_{\boldsymbol{G}} - \mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \kappa]\| \leqslant (1 + o(1)) \cdot \lambda.$$

□

3.6 The second eigenvalue of the shell walk

The goal of this section is to prove Lemma 3.5.2, and in particular bound $\|\mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \boldsymbol{\kappa}] - R_{\boldsymbol{\kappa}}\|$ where $\boldsymbol{\kappa} \sim (\mathsf{D}_{\mathsf{ip}}(d)_{\geqslant \tau})^{\otimes m}$ is a configuration of m shells, and we have conditioned on $\kappa_i = \langle w, \boldsymbol{v}_i \rangle$ for all $i \in [m]$.

To make the matrix more amenable to analysis via the coupling-based techniques we use here, we first observe that the spectral norm we are interested in bounding is equal to the largest eigenvalue of $\mathbf{E}[D_{\boldsymbol{\kappa}}{}^{-1} A_{\boldsymbol{G}} \mid \boldsymbol{\kappa}] - \vec{1}\boldsymbol{\pi}^{\top}$ where $\boldsymbol{\pi} := \frac{D_{\boldsymbol{\kappa}}}{\operatorname{tr}(D_{\boldsymbol{\kappa}})}\vec{1}$ is the stationary distribution of the Markov chain described by the transition matrix $\mathbf{E}[D_{\boldsymbol{\kappa}}{}^{-1} A_{\boldsymbol{G}} \mid \boldsymbol{\kappa}]$. Indeed:

$$\|\mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \boldsymbol{\kappa}] - R_{\boldsymbol{\kappa}}\| = |\lambda|_{\max}(\mathbf{E}[\underline{A}_{\boldsymbol{G}} \mid \boldsymbol{\kappa}] - R_{\boldsymbol{\kappa}}) = |\lambda|_{\max}\left(\mathbf{E}[D_{\boldsymbol{\kappa}}{}^{-1} A_{\boldsymbol{G}} \mid \boldsymbol{\kappa}] - 1\boldsymbol{\pi}^{\top}\right)$$

where the first equality uses symmetry of the matrix and the second equality uses the fact that the spectra of M and $D_{\boldsymbol{\kappa}}{}^{-1/2} M D_{\boldsymbol{\kappa}}{}^{1/2}$ are identical. For convenience, let $Q = \mathbf{E}[A_{\boldsymbol{G}} \mid \boldsymbol{\kappa}]$ and let $\underline{Q} = D_{\boldsymbol{\kappa}}{}^{-1} Q$. The following main result of this section implies Lemma 3.5.2.

Lemma 3.6.1. *There exists a constant $C > 0$ such that for any $d \geqslant C \log m$, any threshold $\tau \in (0,1)$ such that $q := \overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d)}\left(\frac{\tau}{1+\tau}\right) \gg \log^8 m / m$, and any constant $\gamma > 0$, with probability at least $1 - o(m^{-\gamma})$ over the shells $\boldsymbol{\kappa} \sim (\mathsf{D}_{\mathsf{ip}}(d)_{\geqslant \tau})^{\otimes m}$,*

$$|\lambda|_{\max}\left(\underline{Q} - \vec{1}\boldsymbol{\pi}^{\top}\right) \leqslant O\left(\sqrt{\frac{\log^2 d}{d}}\right).$$

In service of proving Lemma 3.6.1, we show:

Lemma 3.6.2. *There exists a constant $C > 0$ such that for any $d \geqslant C \log m$, any threshold $\tau \in (0,1)$ such that $qm \gg \log^8 m$, and any constant $\gamma > 0$, with probability at least $1 - o(m^{-\gamma})$ over the shells $\boldsymbol{\kappa} \sim (\mathsf{D}_{\mathsf{ip}}(d)_{\geqslant \tau})^{\otimes m}$,*

$$\max_{i,j \in [n]} \left\| \left(\underline{Q}^2\right)_{i,*} - \left(\underline{Q}^2\right)_{j,*} \right\|_1 \leqslant O\left(\frac{\log^2 d}{d}\right),$$

where $\left(\underline{Q}^2\right)_{i,*}$ denotes the i-th row of the matrix $\underline{Q}^2$.

We show how to prove Lemma 3.6.1 using Lemma 3.6.2 and then dedicate the rest of the section to proving Lemma 3.6.2.

Proof of Lemma 3.6.1. First, observe that $|\lambda|_{\max}\left(\underline{Q} - \vec{1}\boldsymbol{\pi}^\top\right) = \sqrt{|\lambda|_{\max}\left(\underline{Q}^2 - \vec{1}\boldsymbol{\pi}^\top\right)}$. Via the row sum bound from Claim 3.2.3, Lemma 3.6.2 and the fact that $\boldsymbol{\pi}$ is the stationary distribution of $\underline{Q}$:

$$\begin{aligned}
|\lambda|_{\max}\left(\underline{Q}^2 - \vec{1}\boldsymbol{\pi}^\top\right) &\leqslant \max_{i \in [n]} \left\| \left(\underline{Q}^2\right)_{i,*} - \boldsymbol{\pi}^\top \right\|_1 \\
&= \max_{i \in [n]} \left\| \left(\underline{Q}^2\right)_{i,*} - \sum_{j \in [n]} \boldsymbol{\pi}_j \left(\underline{Q}^2\right)_{j,*} \right\|_1 \\
&= \max_{i \in [n]} \left\| \sum_{j \in [n]} \boldsymbol{\pi}_j \left(\left(\underline{Q}^2\right)_{i,*} - \left(\underline{Q}^2\right)_{j,*} \right) \right\|_1 \\
&\leqslant \max_{i,j \in [n]} \left\| \left(\underline{Q}^2\right)_{i,*} - \left(\underline{Q}^2\right)_{j,*} \right\|_1.
\end{aligned}$$

□

3.6.1 Coupling for the shell walk

In this section we give the proof of Lemma 3.6.2 assuming a few key lemmas. The proofs for the key lemmas are deferred to the next section.

3.6.1.1 A high-probability condition for $\boldsymbol{\kappa}$

To simplify the upcoming computations for Lemma 3.6.2, we will condition on the following high-probability event $\mathcal{E}_\gamma$ over the sample space of the shells $\boldsymbol{\kappa}$:

Definition 3.6.3. Let $\mathcal{E}_\gamma$ be the event that for all m shells $\boldsymbol{\kappa}_i \in \boldsymbol{\kappa}$ in the link, $\boldsymbol{\kappa}_i \leqslant \eta$, where

$$\eta = \tau\left(m^{-2\gamma-1} \cdot \overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d)}(\tau), d\right).$$

Note that the outermost $\tau(\cdot)$ refers to the threshold function, rather than the value of the threshold such that $\overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d)}(\tau) = p$.

Claim 3.6.4. The event $\mathcal{E}_\gamma$ occurs with probability at least $1 - m^{-2\gamma}$.

Proof. By definition, for any shell $\boldsymbol{\kappa}_i$: $\mathbf{Pr}_{\boldsymbol{\kappa}_i \sim \mathsf{D}_{\mathsf{ip}}(d)|_{\geqslant \tau}}[\boldsymbol{\kappa}_i \geqslant \eta] \leqslant m^{-2\gamma-1}$. Our conclusion follows from taking a union bound over all m shells. □

The conditioning on $\mathcal{E}_\gamma$ can be folded into high-probability guarantee over $\boldsymbol{\kappa}$ in Lemma 3.6.2. Thus, for the remainder of the section, we can assume that $\boldsymbol{\kappa}$ obeys event $\mathcal{E}_\gamma$. This will be especially relevant in the analysis of the outlier shells (Section 3.6.2.2).

Claim 3.6.5. If $d \geqslant C \log m$ for some constant $C > 0$, then $\eta \leqslant 1 - \varepsilon_\gamma$, where $\varepsilon_\gamma > 0$ is a constant depending only on γ.

Proof. Since τ is a constant bounded away from 1, and $d = \Omega(\log m)$, by the lower bound in Lemma 3.2.8, the quantity $m^{-2\gamma-1} \cdot \overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d)}(\tau)$ is at least $\exp(-C_\gamma d)$ for some constant C_γ depending on γ. By the upper bound in Lemma 3.2.8, there is a constant $\varepsilon_\gamma > 0$ such that $\overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d)}(1 - \varepsilon_\gamma) \leqslant \exp(-C_\gamma d)$. Since $\overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d)}$ is a decreasing function, $\eta \leqslant 1 - \varepsilon_\gamma$. □

3.6.1.2 "Typical" and "outlier" shells

In the proof of Lemma 3.6.2, we analyze the contributions of "typical" and "outlier" shells separately.

Definition 3.6.6. We say that a shell κ_i is "typical" if $\kappa_i \in [\tau, \tau(1+\alpha)]$, for $\alpha = \frac{36 \log d}{\tau^2 (d-3)(1-\eta)}$.

Remark 3.6.7. α is chosen so that $\underline{Q}$, when restricted to typical rows and columns, will resemble a rank-1 matrix. For our eventual choices of d and m, the event that every shell is typical *does not* occur with high probability; we will inevitably need to deal with outlier shells.

3.6.1.3 Total variation bound from similarity of typical rows and scarcity of outlier columns

To obtain the desired row-sum bound in Lemma 3.6.2, we will prove the following two lemmas about the matrix $\underline{Q}$. The first shows that outlier columns do not contribute much to the total row sum of any row:

Lemma 3.6.8. *For any $d \geqslant C \log m$ for some constant $C > 0$ and any $\tau \in [0, 1]$ such that $qm \gg \log^8 m$, if $\kappa_i \leqslant \eta$,*

$$\sum_{k=1}^{m} \underline{Q}_{i,k} \cdot \mathbf{1}[k \text{ outlier}] \leqslant O\left(\frac{1}{d}\right)$$

with probability $1 - o(m^{-\log m})$.

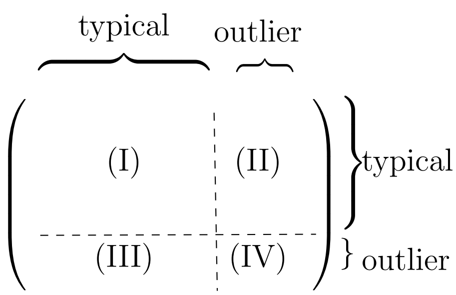
The second shows that typical rows are similar at indices corresponding to typical columns:

Lemma 3.6.9. *For any dimension d and any threshold $\tau \in (0,1)$, if κ_i, κ_j correspond to typical shells, then for all ℓ such that κ_ℓ is typical,*

$$\underline{Q}_{i,\ell} \in \left(1 \pm O\left(\frac{\log^2 d}{d}\right)\right) \underline{Q}_{j,\ell}$$

These lemmas are both proven by direct calculation, and we leave their proofs to Section 3.6.2.2 and Section 3.6.2.3 respectively.

To illustrate these statements, we provide a schematic of the matrix Q below, organized into its typical and outlier rows and columns. Lemma 3.6.9 states that the sub-rows in area (I) of the matrix are all nearly equal to each other. Lemma 3.6.8 says that the sum of its entries in area (II) or area (IV) is a $O\left(\frac{1}{d}\right)$ fraction of the total row sum.



One straightforward corollary of Lemma 3.6.9 and Lemma 3.6.8 is that the ℓ_1 norms of the differences between any two *typical* rows of $\underline{Q}$ is at most $O\left(\frac{\log^2 d}{d}\right)$. More formally:

Corollary 3.6.10. *For any $d \geqslant C \log m$ for some constant $C > 0$ and any threshold $0 < \tau \leqslant 1$ such that $qm \gg \log^8 m$, let i, j be rows of $\underline{Q}$ corresponding to typical shells κ_i, κ_j. Then:*

$$\left\| \left(\underline{Q}\right)_{i,*} - \left(\underline{Q}\right)_{j,*} \right\|_1 \leqslant O\left(\frac{\log^2 d}{d}\right)$$

with probability $1 - o(m^{-\log m})$.

Proof. We split $\left\| \left(\underline{Q}\right)_{i,*} - \left(\underline{Q}\right)_{j,*} \right\|_1$ based on its contributions from typical columns and outlier columns.

$$\left\| \left(\underline{Q}\right)_{i,*} - \left(\underline{Q}\right)_{j,*} \right\|_1 = \sum_{\ell \text{ typical}} \left| \underline{Q}_{i,\ell} - \underline{Q}_{j,\ell} \right| + \sum_{\ell \text{ outlier}} \left| \underline{Q}_{i,\ell} - \underline{Q}_{j,\ell} \right|$$

Lemma 3.6.8 and the triangle inequality tell us that with probability $1 - o(m^{-\log m})$:

$$\sum_{\ell \text{ outlier}} \left| \underline{Q}_{i,\ell} - \underline{Q}_{j,\ell} \right| \leqslant \sum_{\ell \text{ outlier}} \left| \underline{Q}_{i,\ell} \right| + \sum_{\ell \text{ outlier}} \left| \underline{Q}_{j,\ell} \right| \leqslant O\left(\frac{1}{d}\right)$$

Lemma 3.6.9 tells us that for some constant $C > 0$:

$$\sum_{\ell \text{ typical}} \left|\underline{Q}_{i,\ell} - \underline{Q}_{j,\ell}\right| \leqslant \sum_{\ell \text{ typical}} \left[1 + \left(\frac{C\log^2 d}{d} - 1\right)\right] \underline{Q}_{j,\ell}$$
$$= \frac{C\log^2 d}{d} \sum_{\ell \text{ typical}} \underline{Q}_{j,\ell} \leqslant \frac{C\log^2 d}{d}$$

Combining the bounds on $\sum_{\ell \text{ typical}} \left|\underline{Q}_{i,\ell} - \underline{Q}_{j,\ell}\right|$ and $\sum_{\ell \text{ outlier}} \left|\underline{Q}_{i,\ell} - \underline{Q}_{j,\ell}\right|$ gives the desired result. □

We can furthermore use Lemma 3.6.8, Corollary 3.6.10, and a coupling argument, to prove Lemma 3.6.2.

Proof of Lemma 3.6.2. We may assume event $\mathcal{E}_\gamma$ (that all shells $\kappa_i \leqslant \eta$), and the outcomes of Lemma 3.6.8 and Corollary 3.6.10. The union of these three events occur with probability $1 - o(m^{-\gamma} + m^{-\log m})$.

Let $(\boldsymbol{X}_a^{(t)})_{t\geqslant 0}$ be the trajectory of Markov chain $\underline{Q}$ starting at vertex a. For any pair of vertices $i, j \in [n]$, we couple $\boldsymbol{X}_i^{(2)}$ and $\boldsymbol{X}_j^{(2)}$ such that they are equal with probability $1 - O\left(\frac{\log^2 d}{d}\right)$, and so by Fact 3.2.6:

$$\left\|\text{pmf}\left(\boldsymbol{X}_i^{(2)}\right) - \text{pmf}\left(\boldsymbol{X}_j^{(2)}\right)\right\|_{\text{TV}} = \frac{1}{2}\left\|\left(\underline{Q}^2\right)_{i,*} - \left(\underline{Q}^2\right)_{j,*}\right\|_1 \leqslant O\left(\frac{\log^2 d}{d}\right)$$

from which the desired result follows.

We now exhibit such a coupling between $\boldsymbol{X}_i^{(2)}$ and $\boldsymbol{X}_j^{(2)}$. Observe that $\boldsymbol{X}_i^{(1)}$ and $\boldsymbol{X}_j^{(1)}$ are distributed according to $\left(\underline{Q}\right)_{i,*}$ and $\left(\underline{Q}\right)_{j,*}$. When κ_i and κ_j are both typical shells, we can couple $\boldsymbol{X}_i^{(1)}$ and $\boldsymbol{X}_j^{(1)}$ such that they are equal with probability $1 - O\left(\frac{\log^2 d}{d}\right)$ using Corollary 3.6.10 and Fact 3.2.6. As a result $\boldsymbol{X}_i^{(2)}$ and $\boldsymbol{X}_j^{(2)}$ can be coupled so that they agree with probability $1 - O\left(\frac{\log^2 d}{d}\right)$. When either κ_i or κ_j is an outlier shell, though $\boldsymbol{X}_i^{(1)}$ and $\boldsymbol{X}_j^{(1)}$ may have TV distance greater than $O\left(\frac{\log^2 d}{d}\right)$, by Lemma 3.6.8 for both random variables $1 - O\left(\frac{1}{d}\right)$-fraction of the probability mass is over the typical shells. Due to that, we can couple $\boldsymbol{X}_i^{(2)}$ and $\boldsymbol{X}_j^{(2)}$ with probability $\left(1 - O\left(\frac{1}{d}\right)\right) \cdot \left(1 - O\left(\frac{\log^2 d}{d}\right)\right)$ by Lemma 3.6.8. Thereby we complete the proof. □

3.6.2 Relating typical rows and bounding outlier columns

Our next step is to prove Lemma 3.6.8 and Lemma 3.6.9. Throughout this section, instead of working with $\underline{Q}$, we will work with $Q = \mathbf{E}[\boldsymbol{A_G} \mid \kappa]$; it will be simpler to operate on the entries of Q and later relate them to $\underline{Q}$. We first characterize the entries of Q using the $\mathsf{D}_{\mathsf{ip}}(d-1)$ distribution.

3.6.2.1 The conditional expected adjacency matrix

For each pair of vertices $i, j \in [m]$, we have

$$Q_{i,j} = q_{\kappa_i}(\kappa_j) \coloneqq \mathop{\mathbf{Pr}}_{\boldsymbol{v}_i,\boldsymbol{v}_j\sim\rho_w}\left[\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle \geqslant \tau \mid \langle \boldsymbol{v}_i, w\rangle = \kappa_i, \langle \boldsymbol{v}_j, w\rangle = \kappa_j\right].$$

Though $q_x(y)$ is symmetric in its inputs x, y, we choose this notation because we will often work with the function $q_x(\cdot)$, where the input is any value in $[\tau, 1]$.

Claim 3.6.11. The quantity $q_x(y)$ is exactly a tail probability of the $\mathsf{D}_{\mathsf{ip}}(d-1)$ distribution:

$$q_x(y) = \overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d-1)}(T(x,y))$$

Proof. Conditional on $\langle \boldsymbol{v}_i, w\rangle = x$ and $\langle \boldsymbol{v}_j, w\rangle = y$, $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ are distributed as $\boldsymbol{v}_i = x \cdot w + \sqrt{1-x^2}\cdot \boldsymbol{u}_i$ and $\boldsymbol{v}_j = y\cdot w + \sqrt{1-y^2}\cdot \boldsymbol{u}_j$, for $\boldsymbol{u}_i, \boldsymbol{u}_j$ uniformly random unit vectors orthogonal to w. Now, observe that the condition $\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle \geqslant \tau$ is equivalent to $\langle \boldsymbol{u}_i, \boldsymbol{u}_j\rangle \geqslant \frac{\tau - xy}{\sqrt{(1-x^2)(1-y^2)}}$, and thus the desired statement follows since $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$ are sampled from a space isometric to $\mathbb{S}^{d-2}$. □

3.6.2.2 The contribution of outlier columns

The goal of this section is to prove Lemma 3.6.8.

Lemma (Restatement of Lemma 3.6.8)**.** *For any $d \geqslant C\log m$ for some constant $C > 0$ and any $\tau \in [0,1]$ such that $qm \gg \log^8 m$, if $\kappa_i \leqslant \eta$,*

$$\sum_{k=1}^{m} \underline{Q}_{i,k} \cdot \mathbf{1}[k \text{ outlier}] \leqslant O\left(\frac{1}{d}\right)$$

with probability $1 - o(m^{-\log m})$.

The lemma statement is equivalent to the following about Q: for all $i \in [m]$,

$$\frac{\sum_{k=1}^{m} Q_{i,k}\cdot \mathbf{1}[k \text{ outlier}]}{\sum_{k=1}^{m} Q_{i,k}} \leqslant O\left(\frac{1}{d}\right)$$

with probability $1 - o(m^{-\log m})$. Recalling that we use Z to denote the normalizing constant from Section 3.2.3, we can define:

$$N(x) \coloneqq \int_{\tau(1+\alpha)}^{1} q_x(y)\cdot Z^{-1}(1-y^2)^{(d-3)/2}dy = Z^{-1}\int_{\tau(1+\alpha)}^{1}(1-y^2)^{(d-3)/2}\cdot\overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d-1)}\left(T(x,y)\right)dy$$

$$D(x) \coloneqq Z^{-1}\int_{\tau}^{1} q_x(y)\cdot(1-y^2)^{(d-3)/2}dy = Z^{-1}\int_{\tau}^{1}(1-y^2)^{(d-3)/2}\cdot\overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d-1)}\left(T(x,y)\right)dy$$

By our definitions of $N(x)$ and $D(x)$, and recalling that we condition on $\mathcal{E}_\gamma$ (Definition 3.6.3) throughout this section,

$$N(\kappa_i) = \mathop{\mathbf{E}}_{\boldsymbol{\kappa}_\ell}[Q_{i,\ell} \cdot \mathbf{1}[\ell \text{ outlier}]], \quad D(\kappa_i) = \mathop{\mathbf{E}}_{\boldsymbol{\kappa}_\ell}[Q_{i,\ell}]$$

The $Z^{-1}(1-y^2)^{(d-3)/2}$ expression in each integrand comes from the probability density over shells.

First, when $\kappa_i \leqslant \eta$, we establish that the ratio of the *expected* sum of outlier $Q_{i,k}$ and typical $Q_{i,k}$ is of the desired magnitude of $O\left(\frac{1}{d}\right)$.

Lemma 3.6.12. *For any $d \geqslant C \log m$ for some constant $C > 0$, any constant $\tau \in [0,1]$, and any $x \leqslant \eta$,*

$$\frac{N(x)}{D(x)} \leqslant O\left(\frac{1}{d}\right).$$

The proof is omitted here and can be found in Appendix B of [84]. We are now ready to prove Lemma 3.6.8.

Proof of Lemma 3.6.8. For convenience, we use N and D as shorthand for $N(\kappa_i)$ and $D(\kappa_i)$. We compute a high probability lower bound for the numerator $\sum_{k=1}^m Q_{i,k} \cdot \mathbf{1}[k \text{ outlier}]$ and a high probability upper bound for the denominator $\sum_{k=1}^m Q_{i,k}$.

Concentration of the numerator: We first show that $\sum_{k=1}^m Q_{i,k} \cdot \mathbf{1}[k \text{ outlier}]$ concentrates around Nm. First, each $Q_{i,k} \leqslant 1$. Then, $\mathbf{Var}(Q_{i,k}) \leqslant \mathbf{E}[Q_{i,k}^2 \cdot \mathbf{1}[k \text{ outlier}]] \leqslant \mathbf{E}[Q_{i,k} \cdot \mathbf{1}[k \text{ outlier}]] = N$. Applying Bernstein's Inequality, we obtain:

$$\begin{aligned}
&\mathbf{Pr}\left[\sum_{k=1}^m Q_{i,k} \cdot \mathbf{1}[k \text{ outlier}] \geqslant Nm + \left(\sqrt{Nm}+1\right)\log^2 m\right] \\
&\leqslant \exp\left(-\frac{\left(\sqrt{Nm}+1\right)^2 \log^4 m}{\frac{1}{2}\cdot Nm + \frac{1}{3}\cdot\left(\sqrt{Nm}+1\right)\log^2 m}\right) \\
&\leqslant m^{-\log m}
\end{aligned}$$

Concentration of the denominator: We next show that $\sum_{k=1}^m Q_{i,k}$ concentrates around Dm. Using a similar bound on variance as above, and applying Bernstein's inequality again:

$$\begin{aligned}
\mathbf{Pr}\left[\sum_{k=1}^m Q_{i,k} \leqslant Dm - \left(\sqrt{Dm}+1\right)\log m\right] &\leqslant \exp\left(-\frac{\left(\sqrt{Dm}+1\right)^2 \log^2 m}{\frac{1}{2}\cdot Dm + \frac{1}{3}\cdot\left(\sqrt{Dm}+1\right)\log m}\right) \\
&\leqslant m^{-\log m}
\end{aligned}$$

Thus, with probability greater that $1 - 2m^{-\log m}$, the ratio $\frac{\sum_{k=1}^{n} Q_{i,k} \cdot \mathbf{1}[k \text{ outlier}]}{\sum_{k=1}^{m} Q_{i,k}}$ is at most

$$\frac{Nm + \left(\sqrt{Nm} + 1\right)\log^2 m}{Dm - \left(\sqrt{Dm} + 1\right)\log m}$$

We can upper bound this by $\leqslant O\left(\frac{1}{d}\right)$, as Lemma 3.6.12 tells us $\frac{N}{D} \leqslant O\left(\frac{1}{d}\right)$, and since $Dm \geqslant qm \gg \log^8 m$ the first terms in the ratio dominate. □

3.6.2.3 Relating typical rows

Our goal for this section is to prove:

Lemma (Restatement of Lemma 3.6.9)**.** *For any dimension d and any threshold $\tau \in (0,1)$, if κ_i, κ_j correspond to typical shells, then for all ℓ such that κ_ℓ is typical,*

$$\underline{Q}_{i,\ell} \in \left(1 \pm O\left(\frac{\log^2 d}{d}\right)\right)\underline{Q}_{j,\ell}$$

We will translate Lemma 3.6.9 into a statement about Q first. Let $Q_{i,*}$ and $Q_{j,*}$ be rows of Q corresponding to typical shells κ_i, κ_j. We will prove that $Q_{i,*}$ and $Q_{j,*}$, when restricted to typical columns, are nearly constant scalings of each other. Formally, we will prove:

Lemma 3.6.13. *For any dimension d and any threshold $\tau \in (0,1)$, let $\kappa_i, \kappa_j, \kappa_\ell$ be typical shells. Then,*

$$\frac{q_{\kappa_i}(\kappa_\ell)}{q_{\kappa_j}(\kappa_\ell)} \cdot \left(\frac{q_{\kappa_i}(\tau)}{q_{\kappa_j}(\tau)}\right)^{-1} \in 1 \pm O\left(\frac{\log^2 d}{d}\right)$$

In other words, this establishes that for any typical shells $\kappa_i, \kappa_j, \kappa_\ell$,

$$\frac{q_{\kappa_i}(\kappa_\ell)}{q_{\kappa_j}(\kappa_\ell)} \approx \frac{q_{\kappa_i}(\tau)}{q_{\kappa_j}(\tau)},$$

where the quantity on the right is a constant c_{ij} depending only on κ_i and κ_j (not κ_ℓ).

Proof of Lemma 3.6.9 using Lemma 3.6.13. By the definition of $\underline{Q}$,

$$\underline{Q}_{i,\ell} = \frac{Q_{i,\ell}}{\sum_{k=1}^{m} Q_{i,k}}, \; \underline{Q}_{j,\ell} = \frac{Q_{j,\ell}}{\sum_{k=1}^{m} Q_{j,k}}$$

It suffices to prove that $\frac{\underline{Q}_{i\ell}}{\underline{Q}_{j,\ell}}$ is close to 1. Expanding $\frac{\underline{Q}_{i\ell}}{\underline{Q}_{j,\ell}}$, we can upper bound:

$$\frac{\underline{Q}_{i,\ell}}{\underline{Q}_{j,\ell}} = \frac{Q_{i,\ell}}{\sum_{k=1}^{m} Q_{i,k}} \cdot \frac{\sum_{k=1}^{m} Q_{j,k}}{Q_{j,\ell}}$$

$$
\begin{aligned}
&= \frac{Q_{i,\ell}}{Q_{j,\ell}} \cdot \frac{\sum_{k=1}^m Q_{j,k}}{\sum_{k=1}^m Q_{i,k}} \\
&\leqslant \left(1 + \frac{C\log^2 d}{d}\right)\left(\frac{q_{\kappa_i}(\tau)}{q_{\kappa_j}(\tau)}\right) \cdot \frac{\sum_{k \text{ typical}} Q_{j,k} + \sum_{k \text{ outlier}} Q_{j,k}}{\sum_{k \text{ typical}} Q_{i,k} + \sum_{k \text{ outlier}} Q_{i,k}} \\
&\leqslant \left(1 + \frac{C\log^2 d}{d}\right)\left(\frac{q_{\kappa_i}(\tau)}{q_{\kappa_j}(\tau)}\right) \cdot \frac{\left(1 + \frac{C'}{d}\right)\sum_{k \text{ typical}} Q_{j,k}}{\sum_{k \text{ typical}} Q_{i,k}} \\
&\leqslant \left(1 + \frac{C\log^2 d}{d}\right)\left(\frac{q_{\kappa_i}(\tau)}{q_{\kappa_j}(\tau)}\right) \cdot \frac{\left(1 + \frac{C'}{d}\right)\left(1 + \frac{C\log^2 d}{d}\right)\sum_{k \text{ typical}} \left(\frac{q_{\kappa_j}(\tau)}{q_{\kappa_i}(\tau)}\right) \cdot Q_{i,k}}{\sum_{k \text{ typical}} Q_{i,k}} \\
&\leqslant 1 + \frac{C''\log^2 d}{d}
\end{aligned}
$$

The first inequality uses Lemma 3.6.13 to bound $\frac{Q_{i,\ell}}{Q_{j,\ell}}$. The second inequality uses the fact that the outlier entries of $Q_{i,*}$ and $Q_{j,*}$ only occupy an $O\left(\frac{1}{d}\right)$ fraction of the ℓ_1 norm of each row (Lemma 3.6.8). The third inequality again comes from an application of Lemma 3.6.13 to relate $Q_{i,k}$ to $Q_{j,k}$ when k is typical. The lower bound follows analogously. □

Proof of Lemma 3.6.13. By definition:

$$
\frac{q_{\kappa_i}(\kappa_\ell)}{q_{\kappa_j}(\kappa_\ell)} \cdot \left(\frac{q_{\kappa_i}(\tau)}{q_{\kappa_j}(\tau)}\right)^{-1} = \frac{\overline{\Phi}_{\mathsf{D_{ip}}(d-1)}\left(T(\kappa_i,\kappa_\ell)\right)}{\overline{\Phi}_{\mathsf{D_{ip}}(d-1)}\left(T(\kappa_j,\kappa_\ell)\right)} \cdot \frac{\overline{\Phi}_{\mathsf{D_{ip}}(d-1)}\left(T(\kappa_j,\tau)\right)}{\overline{\Phi}_{\mathsf{D_{ip}}(d-1)}\left(T(\kappa_i,\tau)\right)} \tag{3.28}
$$

Since $\kappa_i, \kappa_j, \kappa_\ell \in [\tau, \tau(1+\alpha)]$, by Claim 3.5.6, all terms of the form $T(x,y)$ in the above are lower bounded by $T(\tau(1+\alpha), \tau(1+\alpha))$, which is lower bounded by a constant for large enough d. Thus, by Lemma 3.2.8:

$$
(3.28) = \left(1 \pm O\left(\frac{\log d}{d}\right)\right) \cdot \frac{T(\kappa_j,\kappa_\ell)\cdot T(\kappa_i,\tau)}{T(\kappa_i,\kappa_\ell)\cdot T(\kappa_j,\tau)} \cdot \left(\frac{A}{B}\right)^{(d-1)/2}
$$

where $A \coloneqq \left(1 - T(\kappa_i,\kappa_\ell)^2\right)\left(1 - T(\kappa_j,\tau)^2\right)$ and $B \coloneqq \left(1 - T(\kappa_j,\kappa_\ell)^2\right)\left(1 - T(\kappa_i,\tau)^2\right)$. We show:

$$
\left|\frac{T(\kappa_j,\kappa_\ell)\cdot T(\kappa_i,\tau)}{T(\kappa_i,\kappa_\ell)\cdot T(\kappa_j,\tau)} - 1\right| \leqslant O(\alpha^2) \tag{3.29}
$$

$$
\left|\frac{A}{B} - 1\right| \leqslant O(\alpha^2) \tag{3.30}
$$

where (3.29) and (3.30) are proved in Appendix C of [84]. Consequently,

$$
\frac{q_a(x)}{q_b(x)} \cdot \frac{q_b(\tau)}{q_a(\tau)} = \left(1 \pm O\left(\frac{\log d}{d}\right)\right) \cdot \left(1 \pm O(\alpha^2)\right) \cdot \left(1 \pm O(d\alpha^2)\right)
$$

The term of order $d\alpha^2$ dominates, and because $\alpha = O(\frac{\log d}{d})$ we conclude the desired result. □

3.7 2-dimensional expansion of the random geometric complex

In this section we prove Theorem 3.1.6.

Theorem 3.7.1. *For every $0 < \varepsilon < 1$, $0 < \eta < 2\varepsilon$ and $d = \eta \log_{4/3} n$, if $\boldsymbol{H} \sim \mathsf{Geo}_d^{(2)}(n, n^{-1+\varepsilon})$, then every link of $\boldsymbol{H}$ is a $\left(\frac{1}{2} - \delta\right)$-expander, and its 1-skeleton is a $\left(1 - \frac{4\delta}{1+2\delta}\right)$-expander with high probability where $\delta = \frac{1}{2} \cdot \frac{1-\sqrt{1-\exp\left(-2\log\frac{4}{3}\cdot(1-\varepsilon)/\eta\right)}}{1+\sqrt{1-\exp\left(-2\log\frac{4}{3}\cdot(1-\varepsilon)/\eta\right)}} - o_n(1)$.*

One of the ingredients in the proof of Theorem 3.7.1 is the spectral expansion of random geometric graphs, which is a corollary of Theorem 3.1.11 and Corollary 3.4.7:

Theorem (Restatement of Theorem 3.1.8)**.** *Let $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$ and $\tau := \tau(p,d)$. Then with high probability $\boldsymbol{G}$ is a μ-expander, where*

$$\mu := (1 + o(1)) \cdot \max\left\{(1 + o_{d\tau^2}(1)) \cdot \tau, \frac{\log^4 n}{\sqrt{pn}}\right\},$$

where $o_{d\tau^2}(1)$ denotes a function that goes to 0 as $d \cdot \tau(p,d)^2 \to \infty$.

The second ingredient is a bound on the second eigenvalue of the links, proved in Section 3.5:

Theorem (Restatement of Theorem 3.5.1)**.** *Let $0 < \tau < 1$ be a constant. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \sim \mathrm{cap}_{\geqslant\tau}(w)$ and $\boldsymbol{G} := \mathsf{gg}_\tau(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m)$. Then for $q := \overline{\Phi}_{\mathsf{D_{ip}}(d)}\left(\frac{\tau}{1+\tau}\right)$, suppose $qm \gg \log^8 m \cdot \log^{3/2}\frac{1}{q} \cdot \left(\frac{1+\tau}{\tau}\right)^3$ and $d \geqslant C \cdot \log m$ for any constant $C > 0$, then for any constant $\gamma > 0$,*

$$\mathbf{Pr}\left[|\lambda|_2\left(\widehat{A}_{\boldsymbol{G}}\right) > \frac{\tau}{1+\tau} + o_{d,m}(1)\right] \leqslant O\left(m^{-\gamma}\right).$$

Proof of Theorem 3.7.1. To show that the links expand, we apply Theorem 3.5.1 in combination with a union bound over all links. The second eigenvalue bound for the 1-skeleton is then proved using Theorem 3.1.3, the trickling-down theorem. Let $p = n^{-1+\varepsilon}$, $d = \eta \log_{4/3} n$ and $\tau = \tau(p,d)$.

Let $\boldsymbol{G} := \mathsf{gg}_\tau(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ to denote the geometric graph of the collection of vectors used to generate $\boldsymbol{H}$. The number of vertices that fall in the neighborhood of a vertex v within $\boldsymbol{G}$ is $\boldsymbol{m}_v \sim \mathsf{Binom}(n,p)$, and hence $\boldsymbol{m}_v \geqslant m := pn - 2\sqrt{pn\log n}$ except with probability $o(1/n)$. For the rest of the proof, we condition on the event that $\boldsymbol{m}_v \geqslant m$ for all v, which happens with probability $1 - o(1)$ by the union bound.

The link $\boldsymbol{H}_v$ of a vertex v is obtained by taking $\boldsymbol{G}_v$, the subgraph of $\boldsymbol{G}$ induced by the neighborhood of v, and then removing the isolate vertices. Note that the isolated vertices need to be removed since when sampling a random complex, we remove all edges that are not in any triangles. Our goal is to control the second eigenvalue of all the links in $\boldsymbol{H}$, and we do so by showing bounds on the second eigenvalue of $\boldsymbol{G}_v$ for all v. The second eigenvalue bounds show that with high probability, for all v, the graph $\boldsymbol{G}_v$ is connected, and hence has no isolated vertices. Consequently, $\boldsymbol{H}_v$ is in fact equal to $\boldsymbol{G}_v$ and the second eigenvalue bounds port over.

As a first step, we show that $\boldsymbol{G}_v$ satisfies the hypothesis of Theorem 3.5.1. In particular, we show for $q := \Phi_{\mathsf{D}_{\mathsf{ip}}(d-1)}\left(\frac{\tau}{1+\tau}\right)$

$$q \cdot m_v \gg \log^4 m_v \cdot \log^2 \frac{1}{q} \cdot \left(\frac{1+\tau}{\tau}\right)^4. \tag{3.31}$$

Using Lemma 3.2.8 and the fact that the tail function of a probability distribution is monotone decreasing, we can lower bound q:

$$q \geqslant \overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d-1)}\left(\frac{1}{2}\right) \geqslant \frac{2Z_{d-1}}{d-2} \cdot \left(\frac{3}{4}\right)^{(d-2)/2} \cdot \left(1 - \frac{16\log d}{d-1}\right) \geqslant \Omega\left(\frac{1}{\sqrt{d}}\right) \cdot n^{-\eta/2},$$

where the first inequality holds since $\tau \in (0,1]$, and the last inequality holds by definitions of d and Z_d. We now lower bound τ by a constant. By Lemma 3.2.8:

$$\overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d)}\left(\sqrt{1 - \exp\left(-\frac{(1-\varepsilon)\log\frac{4}{3}}{\eta}\right)}\right) \geqslant \Omega\left(\frac{1}{\sqrt{d}}\right) \cdot \sqrt{p} \cdot \left(1 - O\left(\frac{\log d}{d}\right)\right) \geqslant p = \overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d)}(\tau),$$

where the first inequality holds by definition of Z_d. Since $\overline{\Phi}_{\mathsf{D}_{\mathsf{ip}}(d)}$ is a decreasing function,

$$\tau \geqslant \sqrt{1 - \exp\left(-\frac{(1-\varepsilon)\log\frac{4}{3}}{\eta}\right)}.$$

Consequently:

$$\log^4 m_v \cdot \log^2 \frac{1}{q} \cdot \left(\frac{1+\tau}{\tau}\right)^4 \leqslant \log^6 n.$$

On the other hand, $qm_v \geqslant \Omega\left(\frac{1}{\sqrt{d}}\right) \cdot n^{\varepsilon-\eta/2} \gg \log^6 n$, which establishes (3.31). By (3.31) and Theorem 3.5.1 with $\gamma = 2/\varepsilon$, with probability at least $1 - O(1/n^2)$:

$$|\lambda|_2\left(\widehat{A}_{\boldsymbol{G}_v}\right) \leqslant \frac{\tau}{1+\tau} + o_n(1).$$

By the union bound over all vertices, with probability $1 - O(1/n)$:

$$|\lambda|_2\left(\widehat{A}_{\boldsymbol{G}_v}\right) \leqslant \frac{\tau}{1+\tau} + o_n(1) \qquad \forall v \in [n].$$

Henceforth, we condition on the above. Since $\frac{\tau}{1+\tau} < 1$, for all $v \in [n]$, each $\boldsymbol{G}_v$ is connected and has no isolated vertices and hence $\boldsymbol{H}_v = \boldsymbol{G}_v$. Consequently

$$|\lambda|_2\left(\widehat{A}_{\boldsymbol{H}_v}\right) \leqslant \frac{\tau}{1+\tau} + o_n(1) \qquad \forall v \in [n].$$

Assuming the 1-skeleton $\boldsymbol{H}^{(1)}$ is connected, by the trickling-down theorem (Theorem 3.1.3) it satisfies:

$$|\lambda|_2\left(\widehat{A}_{\boldsymbol{H}^{(1)}}\right) \leqslant \frac{\frac{\tau}{1+\tau} + o_n(1)}{1 - \frac{\tau}{1+\tau} - o_n(1)} = \tau + o_n(1).$$

It remains to bound τ, $\tau/(1+\tau)$ and show $\boldsymbol{H}^{(1)}$ is connected. By Lemma 3.2.8, the following inequality must be satisfied:

$$p \leqslant \frac{Z_d}{\tau(d-1)} \cdot \left(1-\tau^2\right)^{(d-1)/2}.$$

Since the right hand side of the above is a decreasing function of τ and plugging in $\sqrt{1-\exp\left(-\frac{2(1-\varepsilon)\log\frac{4}{3}}{\eta}\right)}$ yields a value smaller than p, we know:

$$\tau \leqslant \sqrt{1-\exp\left(-\frac{2(1-\varepsilon)\log\frac{4}{3}}{\eta}\right)} = 1 - \frac{4\delta}{1+2\delta}. \tag{3.32}$$

The function $\tau/(1+\tau)$ is an increasing function and hence:

$$\frac{\tau}{1+\tau} \leqslant \frac{\sqrt{1-\exp\left(-\frac{2(1-\varepsilon)\log\frac{4}{3}}{\eta}\right)}}{1+\sqrt{1-\exp\left(-\frac{2(1-\varepsilon)\log\frac{4}{3}}{\eta}\right)}} + o_n(1) = \frac{1}{2} - \delta.$$

Finally, to show $\boldsymbol{H}^{(1)}$ is connected, it suffices to illustrate $\widetilde{\boldsymbol{H}}^{(1)}$, a reweighted version of $\boldsymbol{H}^{(1)}$, whose normalized adjacency matrix has a spectral gap. We use $\boldsymbol{G}$ as our reweighting of $\boldsymbol{H}^{(1)}$, which is valid since all edges in $\boldsymbol{G}$ occur in $\boldsymbol{H}^{(1)}$ with probability $1-o_n(1)$. Indeed, for every vertex v and neighbor w the vertex w has some neighbor w' within $\boldsymbol{G}_v$, which means $\{v, w, w'\}$ is a triangle in $\boldsymbol{H}$ causing $\{v, w\}$ to appear in $\boldsymbol{H}^{(1)}$. By our choice of parameters, the lower and upper bounds on τ shown in (3.7) and (3.32) respectively, and Theorem 3.1.8, we know $|\lambda|_2(A_{\boldsymbol{G}}) < \tau + o_n(1) < 1$, which implies $\boldsymbol{H}^{(1)}$ is connected, which completes our proof. □

3.8 Tightness of the tricking-down theorem

In this section we will show that the trickle-down theorem is tight:

Proposition (Restatement of Proposition 3.1.9)**.** *For each $\lambda \in (0, \frac{1}{2}]$ and $\eta > 0$ there exists a 2-dimensional expander in which all vertex link eigenvalues are at most λ for which the 1-skeleton is connected with eigenvalue at least $\frac{\lambda}{1-\lambda} - \eta$.*

We prove the proposition by showing that a random geometric graph's adjacency matrix (when weighted in a regular way) has second eigenvalue at least τ, and then prove that the random geometric complex indeed satisfies that regularity condition.

Lemma 3.8.1. *Let $\boldsymbol{G} \sim \mathsf{Geo}_d(n,p)$ generated by vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$, and let W be the transition matrix of any time-reversible Markov chain on $\boldsymbol{G}$ with stationary distribution π. Then with high probability $\lambda_2(W) \geqslant \tau - o_n(1) - O(\mathrm{d_{TV}}(\pi, U_n)^2)$ where U_n is the uniform distribution on $[n]$.*

Proof. When $\mathrm{d_{TV}}(\pi, U_n) \geqslant 0.1$, the statement is vacuously true. Thus, we assume $\mathrm{d_{TV}}(\pi, U_n) < 0.1$ for the rest of this proof. We see that:

$$1 - \lambda_2(W) = \min_{\substack{f: V(\boldsymbol{G}) \to \mathbb{R}^d \\ f \text{ non-constant}}} \frac{\mathbf{E}_{x \sim_W y} \|f(x) - f(y)\|^2}{\mathbf{E}_{x,y \sim \pi} \|f(x) - f(y)\|^2} \leqslant \frac{\mathbf{E}_{x \sim_W y} \|\boldsymbol{v}_x - \boldsymbol{v}_y\|^2}{\mathbf{E}_{x,y \sim \pi} \|\boldsymbol{v}_x - \boldsymbol{v}_y\|^2} \leqslant \frac{2(1 - \tau(p,d))}{\mathbf{E}_{x,y \sim \pi} \|\boldsymbol{v}_x - \boldsymbol{v}_y\|^2} \tag{3.33}$$

where the last inequality uses that for adjacent x, y, $\langle \boldsymbol{v}_x, \boldsymbol{v}_y \rangle \geqslant \tau(p,d)$. To lower bound the denominator, observe:

$$\begin{aligned}
\mathop{\mathbf{E}}_{x,y\sim\pi} \|\boldsymbol{v}_x - \boldsymbol{v}_y\|^2 &= \sum_{x,y \in [n]} \pi(x)\pi(y)(2 - 2\langle \boldsymbol{v}_x, \boldsymbol{v}_y \rangle) = 2\left(1 - \sum_{x,y\in[n]} \langle \pi(x)\boldsymbol{v}_x, \pi(y)\boldsymbol{v}_y \rangle\right) \\
&= 2\left(1 - \left\|\sum_{x\in[n]} \pi(x)\boldsymbol{v}_x\right\|^2\right) = 2\left(1 - \left\|\sum_{x\in[n]} \frac{1}{n}\boldsymbol{v}_x + \sum_{x\in[n]} \left(\pi(x) - \frac{1}{n}\right)\boldsymbol{v}_x\right\|^2\right) \\
&\geqslant 2\left(1 - \left\|\frac{1}{n}\sum_{x\in[n]} \boldsymbol{v}_x\right\|^2 - 4\left\|\frac{1}{n}\sum_{x\in[n]} \boldsymbol{v}_x\right\| \cdot \mathrm{d_{TV}}(\pi, U_n) - 4\mathrm{d_{TV}}(\pi, U_n)^2\right).
\end{aligned}$$

By standard concentration arguments, $\left\|\frac{1}{n}\sum_{x\in[n]} \boldsymbol{v}_x\right\|$ is $o_n(1)$ with high probability. Plugging in the lower bound into (3.33) tells us:

$$1 - \lambda_2(W) \leqslant 1 - \tau(p,d) + o_n(1) + O\big(\mathrm{d_{TV}}(\pi, U_n)^2\big),$$

which can be rearranged into the desired inequality. □

Armed with this lemma we can prove Proposition 3.1.9.

Proof of Proposition 3.1.9. Let $\tau = \frac{\lambda}{1-\lambda}$, which is in $(0,1)$ for $\lambda \in \left(0, \frac{1}{2}\right)$. Using the bounds from Lemma 3.2.8 we can choose n and $d = \Theta(\log n)$ such that for $p = \overline{\Phi}_{\mathsf{D_{ip}}(d)}(\tau)$, we have $\frac{np^2}{2} \gg \operatorname{poly}\log n$.

Let $\boldsymbol{H} \sim \mathsf{Geo}_d^{(2)}(n,p)$. Since each link contains $\mathsf{Binom}(n-1,p)$ vertices, and $(n-1)p \gg \operatorname{poly}\log n$, every link has $(n-1)p(1 \pm o_n(1)) \geqslant m := np/2$ vertices with probability $1 - O(n^{-1})$. Also, $\overline{\Phi}_{\mathsf{D_{ip}}(d)}\left(\frac{\tau}{1+\tau}\right) \geqslant \overline{\Phi}_{\mathsf{D_{ip}}(d)}(\tau)$, so $m \cdot \overline{\Phi}_{\mathsf{D_{ip}}(d)}\left(\frac{\tau}{1+\tau}\right) \geqslant \frac{np^2}{2} \gg \operatorname{poly}\log m$, and so the conditions of Theorem 3.5.1 are met so that by a union bound we can conclude that all links have second eigenvalue at most $\frac{\tau}{1+\tau} + o(1) = \lambda + o(1)$.

Simultaneously, for any pair of vertices the number of triangles they participate in are within a multiplicative factor of $1 \pm \frac{\log n}{\sqrt{p^2 n}} = 1 \pm o_n(1)$ of each other, as we argue in the next paragraph. Since the stationary distribution π of the random walk on $\boldsymbol{G}$ weighted according to $\boldsymbol{H}^{(1)}$, the 1-skeleton of $\boldsymbol{H}$, puts mass on vertex v proportional to the number of triangles v participates in, it must be the case that $\pi(v) = (1 \pm o_n(1)) \cdot \frac{1}{n}$. Consequently, $\mathrm{d_{TV}}(\pi, U_n) = o_n(1)$, and by Lemma 3.8.1, $\lambda_2(\boldsymbol{H}^{(1)}) \geqslant \tau - o_n(1) = \frac{\lambda}{1-\lambda} - o_n(1)$.

We now show concentration for the number of triangles that contain a vertex. Indeed, the number of triangles that a vertex v participates in is equal to the number of edges in its link. Using, $\boldsymbol{m}_v$ to denote the number of vertices in the link of v, $\deg(u)$ to denote the degree of a vertex u within the link of v, and $\boldsymbol{\kappa}$ to denote the collection of shells that vertices in the link of v lie in, we have:

$$|E(\mathrm{Link}(v))| = \frac{1}{2} \sum_{i=1}^{\boldsymbol{m}_v} \deg_v(u).$$

Henceforth we condition on $\boldsymbol{m}_v$ achieving some value in $(1 \pm o_n(1))p(n-1)$. The average degree of a vertex u within the link of v is at least $\frac{np^2}{2}$, and hence by Bernstein's inequality each $\deg_v(u) = (1 \pm o_n(1))\, \mathbf{E}[\deg_v(u)|\boldsymbol{\kappa}_u]$ except with probability $O(n^{-4})$ since $\deg_v(u)|\boldsymbol{\kappa}_u$ is a sum of independent indicator random variables. The random variables $\mathbf{E}[\deg_v(u)|\boldsymbol{\kappa}_u]$ are independent and distributed as $p(\boldsymbol{\kappa}_u)\boldsymbol{m}_v$ where $p(\boldsymbol{\kappa}_u)$ is the probability that a uniformly random vector in $\mathrm{cap}_p(v)$ falls in $\mathrm{cap}_p(u)$ where $\langle u, v\rangle = \boldsymbol{\kappa}_u$. We can show with Bernstein's inequality that:

$$\sum_{u=1}^{\boldsymbol{m}_v} \mathbf{E}[\deg_v(u)|\boldsymbol{\kappa}_u] = (1 \pm o(1))\, \mathbf{E}[p(\boldsymbol{\kappa}_u)]\boldsymbol{m}_v^2$$

except with probability $O(n^{-4})$. By the union bound, with probability $O(n^{-1})$ for all $v \in [n]$,

$$|E(\mathrm{Link}(v))| = \frac{1 \pm o_n(1)}{2}\, \mathbf{E}[p(\boldsymbol{\kappa}_u)]\boldsymbol{m}_v^2,$$

which completes the proof. □

Chapter 4

Hypercontractive inequalities over epsilon product spaces

In this chapter, we focus on analysis of Boolean functions on high dimensional expanders, whose systematic study was recently initiated by Dikstein et al. [23]. This continues a long line of investigation of Fourier analysis of Boolean functions on extended domains beyond the Boolean hypercube, such as the Boolean slice [96, 42, 43, 46], the Grassmann scheme [29, 72, 36], the symmetric group [45, 40, 20], the p-biased cube [35, 78, 41], and the multi-slice [44, 12]. The foregoing extended domains arise naturally throughout theoretical computer science, and indeed, the study of analysis of Boolean functions on extended domains has recently led to a breakthrough regarding the unique games conjecture [71, 30, 29, 72].

4.1 Problem background and summary of results

Hypercontractive inequalities are amongst the most powerful technical tools in Fourier analysis, yielding a plethora of applications in algorithms, complexity, learning theory, statistical physics, social choice, and beyond (see [94] and references therein). Loosely speaking, such statements assert that functions of low Fourier degree are "well behaved" in terms of their distribution around their mean. Concretely, in the Boolean hypercube, the simplest example of a hypercontractive inequality is Bonami's lemma, which states that for every function $f\colon \{0,1\}^n \to \mathbb{R}$ of Fourier degree at most d, it holds that $\|f\|_4 \leqslant \sqrt{3^d}\|f\|_2$.

Alas, in the setting of high dimensional expanders, where the domain is not a product space and the induced measure is biased, general strong hypercontractivity cannot hold. The heart of the problem is that some highly local functions, such as dictators (i.e., $f(x) = x_i$), provide strong counterexamples to hypercontractivity. A similar phenomenon also occurs in several prominent extended domains, such as the p-biased cube and the Grassmann scheme.

Fortunately, as observed in the setting of the p-biased cube [70], all of the aforementioned examples are local, in the sense that a small number of coordinates can significantly influence the output of the function. This led to the definition of 'global' functions. For Boolean valued functions, these are functions wherein a small number of coordinates can change the output of the function only with a negligible probability. For real valued functions, this is captured by the 2-norm remaining roughly the same when restricting O(1) coordinates of the input. More precisely, consider the setting of a general product measure. Let (V_i, μ_i) be probability spaces, let $V_S = \prod_{i \in S} V_i$ and equip V_S with the product measure, which we denote by μ_S. Every function $f \in L^2\left(V_{[k]}, \mu\right)$ is equipped with an orthogonal decomposition $\sum_{S \subseteq [n]} f^{=S}$ known as the Efron–Stein decomposition. The function $f^{=S}$ in the Efron–Stein decomposition plays a similar role to the function $\widehat{f}(S)\chi_S$ in the Boolean cube. Using that analogy we write

$$f^{\leqslant d} = \sum_{|S| \leqslant d} f^{=S},$$

and f is said to be of degree d if $f = f^{\leqslant d}$. Keevash et al. [69] introduced the following notions. The *Laplacians of f* are given by

$$L_S[f] = \sum_{T \supseteq S} (-1)^{|T|} f^{=T}.$$

For $x \in V_S$ the *derivatives* are given by restricting the laplacians

$$D_{S,x} f = L_S[f](x, \cdot),$$

and the (S,x)-*influence* of f is defined as

$$I_{S,x}[f] = \|D_{S,x}[f]\|_2^2.$$

In this setting, a function f is (r,δ)-*global* if $\|f(x,\cdot)\|^2_{2,\mu_{[n]\setminus S}} \leqslant \delta$ for each $|S| \leqslant r$. We remark that here, being (r,δ)-global for a small $\delta > 0$ is, in a sense, equivalent to having $I_{S,x}[f] \leqslant \delta'$ for a small δ' for all $|S| \leqslant r$ and all x. In fact, δ, δ' can be taken to be within a factor of 2^r of one another.

In [69], it was shown that if $f \in L^2(V, \mu)$ is of degree d, then the following hypercontractive inequality holds:

$$\|f\|_4^4 \leqslant 1000^d \sum_S \mathbb{E}_{x \sim \mu_S} I_{S,x}[f]^2. \tag{4.1}$$

This allowed them to deduce if a function f of degree d is (d,δ)-*global*, then

$$\|f\|_4^4 \leqslant \delta 8000^d \|f\|_2^2.$$

Here when setting $\delta = 100\|f^{\leqslant d}\|_2^2$ one gets the statement $\|f\|_4 \leqslant C^d \|f\|_2$, which replicates the behavior in the Boolean cube. Moreover, the statement is useful even for larger values of δ.

In this work, we raise the following question.

Does hypercontractivity hold for high dimensional expanders?

4.1.1 Main results

We answer the question above in the affirmative. Namely, our main contribution is a hypercontractive inequality for functions on the k-faces of an ϵ-HDX. We denote by $X(k)$ the k-faces of a simplicial complex X, and denote by μ the uniform measure on its k-faces. We define the influences $I_{S,x}^{\leqslant d}$ and the degree restriction operator $(\cdot)^{\leqslant d}$ analogously to their definition on the p-biased cube (see Section 4.4 for precise definition). We then prove the following hypercontractive statement for high dimensional expanders in the spirit of (4.1).

Theorem 4.1.1. *Let X be an ϵ-HDX, and let $f \in L^2\left(X(k), \mu\right)$. We have*

$$\|f^{\leqslant d}\|_4^4 \leqslant 20^d \sum_{|S| \leqslant d} (4d)^{|S|}\, \mathbb{E}_{x\sim\mu_S} I_{S,x}^{\leqslant d}\left[f\right]^2 + O_k\left(\epsilon^2\right) \|f\|_2^2 \|f\|_\infty^2.$$

In the setting of ϵ-HDX, we say that a function f is (d,δ)-global if for each $|S| \leqslant d$, we have $\|f\left(x,\cdot\right)\|_{L^2(V_x,\mu_x)} \leqslant \delta$. We show that we can bound the infinity norm of global functions and obtain the following strong hypercontractive inequality for global functions on ϵ-HDX.

Corollary 4.1.2. *For each $\zeta, d, k > 0$, there exists $\epsilon_0 = \epsilon_0\left(\zeta,k,d\right), \delta_0 = \delta_0\left(\zeta,d\right)$, such that the following holds. Let $\epsilon \leqslant \epsilon_0, \delta \leqslant \delta_0$, let X be an ϵ-HDX, and let $f \in L^2\left(X(k),\mu\right)$. If f is (d,δ)-global, then we have*

$$\|f^{\leqslant d}\|_4^4 \leqslant \zeta \|f\|_2^2.$$

We remark that, in fact, we prove our results in a slightly more general setting, to which we refer as ϵ-product measures. See Section 4.7 for details.

4.1.2 Applications

As corollaries of our hypercontractive inequality for high dimensional expanders, we obtain several applications, which we discuss below. See Section 4.8 for more details.

4.1.2.1 Fourier spectrum concentration theorem

Fourier concentration results are widely useful in complexity theory and learning theory. Our first application is a Fourier concentration theorem for HDX. Namely, the following theorem shows that global Boolean functions on ϵ-HDX are concentrated on the high degrees, in the sense that the 2-norm of the restriction of a function to its low-degree coefficients only constitutes a tiny fraction of its total 2-norm.

Theorem 4.1.3. *For each $\zeta, d, k > 0$, there exists $\epsilon_0 = \epsilon_0\left(\zeta,k,d\right), \delta_0 = \delta_0\left(\zeta,d\right)$, such that the following holds. Let $\epsilon \leqslant \epsilon_0, \delta \leqslant \delta_0$, let X be an ϵ-HDX, and let $f\colon X(k) \to \{0,1\}$ be (d,δ)-global. Then*

$$\|f^{\leqslant d}\|_2^2 \leqslant \zeta \|f\|_2^2.$$

4.1.2.2 Small set expansion theorem

Small set expansion is a fundamental property that is prevalent in combinatorics and complexity theory. In the setting of the ρ-noisy Boolean hypercube, the small set expansion theorem of Kahn, Kalai, and Linial [63] gives an upper bound on $\mathrm{Stab}_\rho(1_A) = \langle 1_A, T_\rho 1_A \rangle$ for indicators 1_A of small sets A. The noise stability $\mathrm{Stab}_\rho(1_A)$ captures the probability that a random edge (x, y) of the ρ-noisy hypercube has both its endpoints in A. Hence, an inequality of the form $\mathrm{Stab}_\rho(1_A) \leqslant \zeta \|1_A\|_2^2$ for an arbitrarily small ζ and sufficiently small A implies that that small sets are expanding in the sense that the random walk makes you leave them with probability $\geqslant 1-\zeta$. Our second application is a small set expansion theorem for global functions on ϵ-HDX, captured via bounding the natural noise operator in this setting. Let $\rho \in (0, 1)$ be a noise-rate parameter. The noise operator is given by

$$\mathrm{T}_\rho f\left(x\right) := \sum_{S \subseteq [k]} \rho^{|S|} \left(1-\rho\right)^{k-|S|} \mathbb{E}_{y\sim\mu}\left[f\left(y\right)|y_S = x_S\right].$$

In other words, T_ρ corresponds to the random walk that starts with x chooses a ρ-biased random $S \subseteq [k]$, keeps x_S, and re-randomises x given x_S. Our small set expansion theorem tells us that if we start with a small subset $A \subseteq X\left(k\right)$ and we apply one step of the random walk, then we leave A with probability 0.99.

Theorem 4.1.4. *For each $\zeta, d, k > 0$, there exists $\epsilon_0 = \epsilon_0\left(\zeta, k, d\right), \delta_0 = \delta_0\left(\zeta, d\right)$, such that the following holds. Let $\epsilon \leqslant \epsilon_0, \delta \leqslant \delta_0$, and let X be an ϵ-HDX. If $f\colon V_{[k]} \to \{0,1\}$ is (d, δ)-global, then*

$$\|\mathrm{T}_\rho f\|_2^2 \leqslant \zeta \|f\|_2^2.$$

4.1.2.3 Kruskal–Katona theorem

Our last application is an analogue of the Kruskal–Katona theorem in the setting of high dimensional expanders. The Kruskal-Katona theorem is a fundamental and widely-applied result in extremal combinatorics, which gives a lower bound on the size of the lower shadow $\partial\left(A\right)$ of a k-uniform hypergraph A on n vertices. The *lower shadow* is defined to be the family of all $(k-1)$-sets that are contained in an edge of A. More generally, if $A \subseteq X\left(k\right)$, then we similarly let $\partial\left(A\right)$ be the family of all $k-1$-faces that are contained in a k-face of A.

Filmus et al. [45] used their hypercontractivity theorem to prove a stability result for the Kruskal–Katona theorem. We prove a similar stability result for ϵ-HDX.

Theorem 4.1.5. *Let X be an ϵ-HDX, for a sufficiently small $\epsilon > 0$. Let $\delta \leqslant \left(200d\right)^{-d}$, and let $A \subseteq X\left(k-1\right)$ be (d, δ)-global. Then*

$$\mu\left(\partial\left(A\right)\right) \geqslant \mu\left(A\right)\left(1 + \frac{d}{2k}\right).$$

4.1.3 Techniques

Conceptually, one can view the theory of expanders and pseudorandom graphs in the following perspective: Given a pseudorandom regular graph $G = (V, E)$ and $(x, y) \sim E$, the goal is to show that x, y behave similarly to independent random variables $x, y \sim V$, i.e., as an approximation of a product space.

In the theory of high dimensional expanders, we are given a distribution μ on $(k+1)$-tuples by choosing a random k-face $(x_1, \ldots, x_{k+1})$ of a sparse simplicial complex, and the goal is again to show that the variables $\{x_i\}$ approximately behave as though they were independent. Thus, our main objective is to generalise results from the product space setting, where the x_i's are independent, to the setting of HDX, where we only have local spectral information about the links. However, such a generalisation yields significant challenges.

One of the fundamental tools for studying the product space setting is the aforementioned Efron–Stein decomposition. Its role in the analysis of product spaces is that it allows us to easily generalise techniques from the Boolean cube by replacing the Fourier expression $\widehat{f}(S)\chi_S$ with the function $f^{=S}$.

Our high-level proof strategy is to develop new Efron–Stein decompositions for HDX. We show that despite the more involved setting, and despite the fact that we only have mere local spectral information, we can still obtain similar structural properties as in product spaces. We now list a few of the challenges that we are facing, which require fundamentally new ideas and techniques.

Dikstein et al. [23] gave a decomposition of the form $f = \sum_{d=0}^{k} f^{=d}$. We provide a new decomposition $\left\{f^{=S}\right\}_{S\subseteq[k]}$ such that $f = \sum_{S\subseteq[k]} f^{=S}$, and despite not having orthogonality, we can still show that the inner product $\left\langle f^{=S}, f^{=T}\right\rangle$ is negligible compared to $\|f\|_2^2$. This allows us to generalise the Laplacians, derivatives and influences, but we have to deal with the following problems:

- Let $\mathcal{F} \subseteq [k]$ be a small set. We would like to say that $g = \sum_{S\in\mathcal{F}} f^{=S}$ is supported on $\mathcal{F}$, but we have no way of knowing that looking at $\left\{g^{=S}\right\}_{S\subseteq[k]}$, as $g^{=S}$ may be nonzero even for $S \notin \mathcal{F}$. This leads to the problem of how to even define the degree of a function. We would like to say that $f^{\leqslant d} := \sum_{|S|\leqslant d} f^{=S}$ is of degree at most d, and that f is of degree d if $f = f^{\leqslant d}$. Alas, according to this definition the function $f^{\leqslant d}$ is not of degree d.

- We can and do define the derivatives $D_{S,x}$ to be the restrictions of the Laplacians. In the product case the derivatives decrease the degree by $|S|$, and this is a very desirable property as our proof goes by induction on d. However, this is no longer true in the HDX setting.

- We may define the influences by taking 2-norms of the derivatives. However, now it is no longer true that having small influences is equivalent to being global. This leads us to the following problem which is the source for all of the difficulty.

- The spectral information tells us that HDX should behave similarly to product spaces with respect to the L^2-norm. However, we care about L_4 information when bounding $\|f\|_4$, and we deal with L_∞-hypothesis as the globalness notion is about **all** the restrictions. There is no reason for HDX to behave well with respect to L_4 and even more so for L_∞.

At first, the above, and especially the last point, seem as fundamental barriers to this approach.

Nevertheless, we overcome this barrier by developing an alternative notion, which we call the *approximate Efron–Stein decomposition*. Our new notion has the following properties that fix all of the above problems.

- If $\{f_S\}_{S\subseteq[k]}$ is an approximate Efron–Stein decomposition, then crucially, $\{f_S\}_{S\in\mathcal{F}}$ is an approximate Efron–Stein decomposition for $\sum_{S\in\mathcal{F}} f^{=S}$.
- If f is approximately of degree d, in the sense that $\{f_S\}$ is an approximate Efron-Stein decomposition for f, then the derivative $D_{S,x}[f]$ may be L_4-approximated by $D_{S,x}[f]^{\leqslant d-|S|}$.
- We find a way of proving an inequality of the form
$$\mathbb{E}_{x\sim\mu_S} I^2_{S,x}[f] \leqslant \delta \mathbb{E}_{x\sim\mu_S} I_{S,x}[f],$$
without having the traditional hypothesis $\max_x I_{S,x}[f] \leqslant \delta$ at our disposal.
- We show that we may move freely between different approximate Efron–Stein decomposition up to a small L_4-norm error term.

We believe that our approximate Efron–Stein decomposition provides the desired comfortable platform for analysing functions on HDX in the same way one would analyze a product space.

See Section 4.4 for a detailed exposition of our approximate Efron–Stein decomposition, and see Section 4.5 for a more detailed proof overview of our main hypercontractivity results, which build on the aforementioned decomposition.

4.1.4 Related work

Simultaneously and independently to this work, Bafna, Hopkins, Kaufman, and Lovett [7] also obtained hypercontractive inequalities for high dimensional expanders. We remark that while the main hypercontractive inequalities in both papers achieve essentially the same parameters, the techniques are completely different. Namely, in [7] the proof strategy follows the approach of analogous results in the setting of the Grassmann graph, whereas our approach generalises Efron–Stein decompositions and hypercontractivity for general product spaces. We further note that our approximate Efron–Stein decomposition extends approximate Fourier decompositions that appeared in several recent works [67, 68, 23, 1, 61].

4.1.5 Organisation

The rest of the paper is organised as follows. We start in Section 4.2, where we recall the notions of hypercontractivity and globalness in general product spaces, as well as provide an alternative proof of a slightly weaker hypercontractive inequality that is more amenable for generalisation to non-product spaces. In Section 4.3, we present the framework of ϵ-product spaces, of which high dimensional expanders are a special case, and we also define key operators in this setting and show some basic properties they satisfy. Next, in Section 4.4, which is introducing a new approximate Efron–Stein decomposition and developing a framework for proving hypercontractivity results using this decomposition. Then, in Section 4.5, we give a detailed proof overview of our hypercontractive inequalities for high dimensional expanders, which build on the foregoing framework. In Section 4.6, we define the notions of laplacians, derivatives and influences in the setting of ϵ-measures, give bounded approximated Efron–Stein decompositions related to the Laplacians, define globalness, and show that it implies small influences.. Then, we provide the full proof of our main hypercontractivity results in Section 4.7. Finally, in Section 4.8, we show how to derive the applications from our hypercontractive inequalities.

Acknowledgments

We are grateful to Alessandro Chiesa for participating in early stages of this research. We thank Yuval Filmus and Dor Minzer for insightful discussions. We further thank Yuval Filmus for insightful comments on a previous version of this work.

4.2 Recalling globalness and hypercontractivity in the product space setting

We begin by recalling the Efron–Stein decomposition, as well as derivatives and Laplacians in the setting of general product spaces, and state the hypercontractivity inequalities for product spaces that were shown in [69]. We then give a proof, inspired by [36], of a slightly weaker hypercontractivity inequality that we will later generalise to approximate product spaces.

4.2.1 Efron-Stein decomposition

Let $(V_1, \mu_1), \ldots, (V_k, \mu_k)$ be a probability space. Let μ be the corresponding product measure $\mu_1 \otimes \cdots \otimes \mu_k$. For a set $S \subseteq [k]$, we write $V_S = \prod_{i \in S} V_i$, and we write μ_S for the product measure $\mu_S = \bigotimes_{i \in S} \mu_i$. The *Efron–Stein decomposition* is a decomposition of $L^2\left(V_{[k]}, \mu\right)$ into 2^k orthogonal spaces $\{W_S\}_{S \subseteq [k]}$. Every function $f \in L^2\left(V_{[k]}, \mu\right)$ can then be decomposed as $f = \sum_{S \subseteq [k]} f^{=S}$, where $f^{=S}$ is the projection of f to W_S. The Efron–Stein decomposition is characterised by the orthogonality of $\{W_S\}$, the fact that $\sum_S W_S = L^2(V, \mu)$, and the fact that the space W_S is composed of functions depending only on S.

The functions $f^{=S}$ also have an explicit formula for $x \in V_S$, where we denote

$$A_S f(x) = \mathbb{E}_{y \sim \left(V_{\overline{S}}, \mu_{\overline{S}}\right)}[f(x, y)],$$

where $\bar{S} = [k] \setminus S$. We then write

$$f^{=S} = \sum_{T \subseteq S} (-1)^{|S \setminus T|} A_T f.$$

The function $A_S[f]$ then has the following neat Efron–Stein decomposition

$$A_S[f] = \sum_{T \subseteq S} f^{=T}.$$

See [94, Chapter 8] for more details.

4.2.2 Notations

We write $a = b \pm \epsilon$ to indicate that $a \in (b - \epsilon, b + \epsilon)$. We use $a \leqslant O(b)$ to denote that the inequality holds up to an absolute constant, and $a \leqslant O_k(b)$ to denote that the inequality holds up to a constant only depending on k.

4.2.3 Derivatives and Laplacians

Let $\mu = \mu_1 \otimes \cdots \otimes \mu_k$ be a product measure. Let $f \in L^2\left(V_{[k]}, \mu\right)$, $S \subseteq [n]$. The *Laplacian* is given by the formula

$$L_S[f] = \sum_{T \supseteq S} f^{=T} = \sum_{T \subseteq S} (-1)^{|T|} A_{[k] \setminus T} f.$$

For $S \subseteq [n]$ and $x \in V_S$ the *derivative* $D_{S,x} \in L^2(V_{\overline{S}}, \mu_{\overline{S}})$ is defined by

$$D_{S,x} f = L_S[f](x, \cdot).$$

For convenience, we also write $D_{\varnothing} f = f$. The (S, x)-*influence* of f is defined as

$$I_{S,x}[f] = \|D_{S,x}[f]\|_2^2.$$

This includes the case $S = \varnothing$, where we have $I_{\varnothing}[f] := \|f\|_2^2$.

We now state a few facts from [69] that we generalise. The following lemma, which appears in [69], shows that the notion of small influences corresponds to small 2-norms of the restriction of f.

Lemma 4.2.1. *Suppose that $I_{S,x}[f] \leqslant \delta$ for each set S of size at most r. Then $\|f(x,\cdot)\|^2_{2,\mu_{[n]\setminus S}} \leqslant \delta 4^r$ for each S and $x \in V_S$. Conversely, if $\|f(x,\cdot)\|^2_{2,\mu_{[n]\setminus S}} \leqslant \delta$ for each $|S| \leqslant r$ and $x \in V_S$, then $I_{S,x}[f] \leqslant \delta 4^r$ for each S of size at most r and $x \in V_S$.*

For the above reason they gave the following definition.

Definition 4.2.2. A function f is said to be (r,δ)-*global* if $I_{S,x}[f] \leqslant \delta$ for each $|S| \leqslant r$.

The *degree* of a function is the largest S, such that $f^{=S} \neq 0$. The derivatives decrease the degrees for the following reason.

Lemma 4.2.3. *$D_{S,x}\left[f^{=T}\right]$ is 0 unless $S \subseteq T$, and if $S \subseteq T$, then*

$$D_{S,x}\left[f^{=T}\right] \in W^{T\setminus S}.$$

Consequently, if $f = \sum_{|S|\leqslant d} f^{=S}$ is of degree d, then $D_{S,x}[f]$ is of degree $d - |S|$.

4.2.4 Hypercontractivity

The following result is by [69].

Theorem 4.2.4. *If $f \in L^2(V,\mu)$ is of degree d, then*

$$\|f\|_4^4 \leqslant 1000^d \sum_S \mathbb{E}_x I_{S,x}[f]^2.$$

To show the implication of the theorem for global functions they use the following inequality.

Lemma 4.2.5.

$$\sum_S \mathbb{E}_x I_{S,x}[f] \leqslant 2^d \|f\|_2^2.$$

Proof. The right hand side is equal to

$$\sum_S \|L_S[f]\|_2^2 = \sum_S \sum_{T\supseteq S, |T|\leqslant d} \|f^{=T}\|_2^2 \leqslant 2^d \sum_T \|f^{=T}\|_2^2 = 2^d \|f\|_2^2.$$

□

Combining Theorem 4.2.4 and Lemma 4.2.5, we obtain the following corollary.

Corollary 4.2.6. *If f of degree d is (d,δ)-global. Then $\|f^{\leqslant d}\|_4^4 \leqslant \delta 2000^d \|f\|_2^2$.*

Proof. We have

$$\begin{aligned}\|f^{\leqslant d}\|_4^4 &\leqslant 1000^d \sum_{S\subseteq[n]} \mathbb{E}_{x\sim\mu_S}\|I_{S,x}\left[f\right]\|_2^4 \\ &\leqslant \delta 1000^d \sum_{S} \mathbb{E}_{x\sim\mu_S}\|I_{S,x}\left[f\right]\|_2^2 \\ &\leqslant \delta 2000^d \|f\|_2^2.\end{aligned}$$

□

4.2.5 An alternative proof of hypercontractivity on product spaces

We give an alternative proof of the following slightly weaker version of Theorem 4.2.4. The proof is inspired by a future work by Ellis, Kindler, and the second author [36], who show that the same idea works in the Grassmann setting. In this paper we show that it generalises to HDX as well.

Theorem 4.2.7. *Let $f \in L^2(V,\mu)$ be of degree d. Then*

$$\|f\|_4^4 \leqslant 2\cdot 9^d \sum_{|T|\leqslant d} (9d)^{|T|}\,\mathbb{E}_x\left[I_{S,x}\left[f\right]^2\right].$$

4.2.5.1 Proof overview.

Before providing the full proof, we first describe the high-level approach for proving Theorem 4.2.7. The strategy is to first show a lemma that gives the following bound

$$\|f\|_4^4 \leqslant C^d\|f\|_2^4 + \sum_{S\subseteq[n]} (4d)^{|S|}\,\|L_S\left[f\right]\|_4^4, \tag{4.2}$$

for a constant C. Using this lemma, we can give an inductive proof by first noting that $\|L_S\left[f\right]\|_4^4 = \mathbb{E}_x\|D_{S,x}\|_4^4$, and then applying induction using the fact that $D_{S,x}$ is of degree $d-|S|$. Finally, using the fact that $D_{S,x}D_{T,y} = D_{S\cup T,(x,y)}$, we can get our desired hypercontractive statement.

Hence, the key step is to prove the aforementioned lemma. To this end, we first use the fact that

$$\mathbb{E}\left[f^4\right] = \sum_S \|\left(f^2\right)^{=S}\|_2^2.$$

We then expand the summands of $\left(f^2\right)^{=S}$ as sums of terms of the form $\left(f^{=T_1}f^{=T_2}\right)^{=S}$. Next, we note that the nonzero terms either satisfy $T_1\cap T_2\cap S \neq \varnothing$ or satisfy $T_1\Delta T_2 = S$.

Terms of the first kind are cancelled out by $L_i[f]^4$ for an $i \in T_1 \cap T_2 \cap S$ on the right hand side of (4.2). (The terms $\|L_S[f]\|_4^4$ appear because of over counting, which we resolve by inclusion exclusion.) Terms of the latter kind correspond to the situation in the Boolean cube where $f^{=T} = \widehat{f}(T)\chi_T$ and $\chi_T\chi_S = \chi_{T\Delta S}$. We then upper bound $\|\left(f^{=T_1}f^{=T_2}\right)^{=S}\|_2$ by $\|f^{=T_1}\|_2\|f^{=T_2}\|_2$. This allows us to translate the problem of upper bounding the terms of the first kind to the problem of upper bounding the 4-norm of a low degree function on the Boolean cube. Namely, the function

$$\sum_{|T|\leqslant d}\|f^{=T}\|_2\chi_T.$$

Finally, we use hypercontractivity to upper bound the 4-norm by its 2-norm, which is equal to the 2-norm of f. This concludes the proof overview.

4.2.5.2 Proof of hypercontractivity on product spaces

We now give a formal proof of Theorem 4.2.7. We shall first need the following key lemma, which admits the inductive approach.

Lemma 4.2.8. *Let $f \in L^2(V,\mu)$ be of degree d. Then*

$$\frac{1}{2}\|f\|_4^4 \leqslant 9^d\|f\|_2^4 + \sum_{T\neq\varnothing}(4d)^{|T|}\|L_T[f]\|_4^4.$$

We are now ready to prove the lemma.

Proof of Lemma 4.2.8. By Parseval we have

$$\|f\|_4^4 = \sum\|\left(f^2\right)^{=S}\|_2^2.$$

We bound each term $\|\left(f^2\right)^{=S}\|_2^2$ individually. By expanding and using the linearity of the $\cdot^{=S}$ operator we have

$$\left(f^2\right)^{=S} = \sum_{T_1,T_2}\left(f^{=T_1}f^{=T_2}\right)^{=S}.$$

We now divide the pairs (T_1,T_2) into three sums.

1. We let I_1 be the set of pairs (T_1,T_2) such that $T_1\cap T_2\cap S\neq\varnothing$. If i is in $T_1\cap T_2\cap S$, then the summand $\left(f^{=T_1}f^{=T_2}\right)^{=S}$ appears as a summand when expanding $\left(L_i[f]^2\right)^{=S}$. This explains the role of the Laplacians in the right hand side.

2. We let I_2 be the set of pairs such that $T_1\Delta T_2 = S$. These kind of pairs have a similar behavior to the one in the Boolean cube. There $f^{=S} = \widehat{f}(S)\chi_S$ and

$$f^{=S}f^{=T} = \widehat{f}(S)\widehat{f}(T)\chi_{S\Delta T}.$$

We show that the contribution from the pairs in I_2 is $\leqslant C^d\|f\|_2^2$.

3. We let $I_3 = (T_1, T_2)$ such that either $(T_1 \Delta T_2) \setminus S \neq \varnothing$ or $S \setminus (T_1 \cup T_2) \neq \varnothing$. We show that in this case $\left(f^{=T_1} f^{=T_2}\right)^{=S} = 0$.

It is easy to verify that each pair (T_1, T_2) belongs to at least one of the sets I_1, I_2, I_3. We additionally have $I_1 \cap I_2 = \varnothing$.

Upper bounding the contribution from I_1

Let us start by upper bounding the contribution from pairs corresponding to I_1. For a nonempty $T \subseteq S$ write $I_1(T)$ for the pairs (T_1, T_2), such that $T_1 \cap T_2 \supseteq T$. Then

$$\left(L_T[f]^2\right)^{=S} = \sum_{(T_1,T_2)\in I_1(T)} \left(f^{=T_1} f^{=T_2}\right)^{=S}.$$

Now $I_1 = \bigcup_{i\in S} I_1(i)$, so as a multiset inclusion-exclusion shows that we have

$$I_1 = \sum_{T\subseteq S} (-1)^{|T|-1} \bigcap_{i\in T} I_1(i) = \sum_{T\subseteq S} (-1)^{|T|-1} I_1(T).$$

We therefore have the equality:

$$\sum_{(T_1,T_2)\in I_1} \left(f^{=T_1} f^{=T_2}\right)^{=S} = \sum_{T\subseteq S, T\neq\varnothing} (-1)^{|T|-1} \left(L_T[f]^2\right)^{=S}.$$

By the triangle inequality and Cauchy–Schwarz, we obtain that

$$\begin{aligned}\| \sum_{(T_1,T_2)\in I_1} \left(f^{=T_1} f^{=T_2}\right)^{=S} \|_2^2 &\leqslant \left(\sum_{i=1}^{|S|} \binom{|S|}{i} (4|S|)^{-i}\right)\left(\sum_{T\subseteq S} (4|S|)^{|T|} \| \left(L_T[f]^2\right)^{=S} \|_2^2\right) \\ &\leqslant \sum_{T\subseteq S} (4|S|)^{|T|} \| \left(L_T[f]^2\right)^{=S} \|_2^2.\end{aligned}$$

Summing over all S we have

$$\sum_S \| \sum_{(T_1,T_2)\in I_1} \left(f^{=T_1} f^{=T_2}\right)^{=S} \|_2^2 \leqslant \sum_T (4d)^{|T|} \|L_T[f]\|_4^4.$$

Upper bounding the contribution from I_2

We now upper bound the contribution from I_2. Let $T_1 \Delta T_2 = S$. Then for each $S' \subsetneq S$, we assert that $A_{S'}\left(f^{=T_1} f^{=T_2}\right) = 0$. Let $i \in S \setminus S'$. Then $i \in T_1 \Delta T_2$. Assume without loss of generality that $i \in T_1$. Then

$$A_{S'\cup T_2}\left(f^{=T_1} f^{=T_2}\right) = f^{=T_2} A_{S'\cup T_2} f^{=T_1} = 0.$$

This shows that $A_{S'}\left[f^{=T_1}f^{=T_2}\right]=0$. Hence,

$$\left(f^{=T_1}f^{=T_2}\right)^{=S}=A_S\left(f^{=T_1}f^{=T_2}\right)=\left\langle f^{=T_1}\left(x,\cdot\right),f^{=T_2}\left(x,\cdot\right)\right\rangle_{L^2\left(\mu_{\overline{S}}\right)}.$$

By Cauchy–Schwarz we have

$$\begin{aligned}\|\sum_{(T_1,T_2)\in I_2}\left(f^{=T_1}f^{=T_2}\right)^{=S}\|_2^2&=\sum_{T_1\Delta T_2=T_3\Delta T_4=S}\left\langle\left(f^{=T_1}f^{=T_2}\right)^{=S},\left(f^{=T_3}f^{=T_4}\right)^{=S}\right\rangle\\&\leqslant\sum_{T_1\Delta T_2=T_3\Delta T_4=S}\left\|\left(f^{=T_1}f^{=T_2}\right)^{=S}\right\|_2\left\|\left(f^{=T_3}f^{=T_4}\right)^{=S}\right\|_2.\end{aligned}$$

Now, for each $(T_1,T_2)\in I_2$ we have

$$\begin{aligned}\left\|\left(f^{=T_1}f^{=T_2}\right)^{=S}\right\|_2^2&=\mathbb{E}_{x\sim\mu_S}\left\langle f^{=T_1}\left(x,\cdot\right),f^{=T_2}\left(x,\cdot\right)\right\rangle_{L^2\left(\mu_{\overline{S}}\right)}^2\\&\leqslant\mathbb{E}_{x\sim\mu_S}\left[\|f^{=T_1}\left(x,\cdot\right)\|_{L^2\left(\mu_{\overline{S}}\right)}^2\|f_x^{=T_2}\|_{L^2\left(\mu_{\overline{S}}\right)}^2\right]\\&=\mathbb{E}_{x\sim\mu_S}\|f_x^{=T_1}\|_2^2\mathbb{E}_{x\sim\mu_S}\|f_x^{=T_2}\|_2^2\\&=\|f^{=T_1}\|_2^2\|f^{=T_2}\|_2^2,\end{aligned}$$

where in the second equality we used the fact that $\|f^{=T}\left(x,\cdot\right)\|_{L^2\left(\mu_{\overline{S}}\right)}^2$ depends only on $x_{T\cap S}$, so these are independent for $T=T_1$ and $T=T_2$. This establishes

$$\mathbb{E}\left[\left(f^{=T_1}f^{=T_2}\right)^{=S}\left(f^{=T_3}f^{=T_4}\right)^{=S}\right]\leqslant\|f^{=T_1}\|_2\|f^{=T_2}\|_2\|f^{=T_3}\|_2\|f^{=T_4}\|_2$$

Summing over all S, we obtain

$$\begin{aligned}\sum_S\|\sum_{(T_1,T_2)\in I_2}\left(f^{=T_1}f^{=T_2}\right)^{=S}\|_2^2&\leqslant\mathbb{E}_{\left(\{0,1\}^n,\mu_{\frac{1}{2}}\right)}\left(\sum_{S\subseteq[n]}\|f^{=S}\|_2\chi_S\right)^4\\&\leqslant 9^d\mathbb{E}\left[\left(\sum\|f^{=S}\|_2\chi_S\right)^2\right]^2\\&=9^d\|f\|_2^4.\end{aligned}$$

Here the first inequality follows by expanding both terms and the second is a well known consequence of hypercontractivity in the uniform cube.

Showing that there is no contribution from I_3

We recall that I_3 consist of the pairs with either $(T_1\Delta T_2)\setminus S\neq\varnothing$ or $S\setminus(T_1\cup T_2)\neq\varnothing$. Then we claim that $\left(f^{=T_1}f^{=T_2}\right)^{=S}=0$. If $T_1\cup T_2$ does not contain S, then

$$f^{=T_1}f^{=T_2}=A_{T_1\cup T_2}\left(f^{=T_1}f^{=T_2}\right)=\sum_{S'\subseteq S}\left(f^{=T_1}f^{=T_2}\right)^{=S'}.$$

The uniqueness of the Efron–Stein decomposition shows that $\left(f^{=T_1} f^{=T_2}\right)^{=S} = 0$. Suppose now that there exists $i \in (T_1 \Delta T_2) \setminus S$. Without loss of generality $i \in T_1$. We then have

$$A_{[k]\setminus\{i\}}\left(f^{=T_1} f^{=T_2}\right) = f^{=T_2} \cdot A_{[k]\setminus\{i\}} f^{=T_1} = 0.$$

In particular, $\left(f^{=T_1} f^{=T_2}\right)^{=S} = 0$ as for each $S \subseteq [k] \setminus \{i\}$ we have

$$\left(f^{=T_1} f^{=T_2}\right)^{=S} \left[A_{[k]\setminus\{i\}}\left(f^{=T_1} f^{=T_2}\right)\right]^{=S} = 0$$

Combining the contributions from I_1 and I_2.

The lemma now follows by Cauchy–Schwarz. We have

$$\begin{aligned}
\|f\|_4^4 &\leqslant \sum_S \|\left(f^2\right)^{=S}\|_2^2 \\
&\leqslant \sum_S \left(2 \sum_{(T_1,T_2)\in I_1} \|\left(f^2\right)^{=S}\|_2^2 + 2 \sum_{(T_1,T_2)\in I_2} \|\left(f^2\right)^{=S}\|_2^2 \right) \\
&\leqslant 2 \sum_T (4d)^{|T|} \|L_T[f]\|_4^4 + 2 \cdot 9^d \|f\|_2^4.
\end{aligned}$$

□

Finally, using Lemma 4.2.8, we can derive Theorem 4.2.7 as follows.

Proof of Theorem 4.2.7. The proof is by induction on d. Assume the theorem holds for all degrees $\leqslant d-1$. Since $D_{T,x}[f]$ is of degree $d - |T| \leqslant d-1$, we have

$$\begin{aligned}
\frac{1}{2}\|f\|_4^4 &\leqslant 9^d \|f\|_2^4 + \sum_{T\neq\varnothing} (4d)^{|T|} \|L_T[f]\|_4^4 \\
&= 9^d \|f\|_2^4 + \sum_{T\neq\varnothing} (4d)^{|T|} \mathbb{E}_{x\sim\mu_T} \|D_{T,x}[f]\|_4^4 \\
&\leqslant 9^d \|f\|_2^4 + \sum_{T\neq\varnothing} 2 \cdot 9^{d-|T|} (4d)^{|T|} \sum_{T'\subseteq[n]\setminus T} (8d)^{|T'|} \mathbb{E}_{x\sim\mu_{T\cup T'}} I_{T\cup T',x}^2 \\
&= 9^d \|f\|_2^4 + \sum_{T\cap T'=\varnothing} 2^{|T'|+1} 9^{d-|T|} (4d)^{|T\cup T'|} \mathbb{E}_{x\sim\mu_{T\cup T'}} \|D_{T'\cup T,x}[f]\|_2^4 \\
&\leqslant 9^d \sum_{T\subseteq S} (9d)^{|T|} \mathbb{E}_{x\sim\mu_T} I_{T,x}^2.
\end{aligned}$$

□

4.3 Epsilon product spaces and the projection operators

In this section, we present the framework of ϵ-product spaces, of which high dimensional expanders are a special case. We also define key operators in this setting and show some basic properties that they satisfy.

4.3.1 Complexes having ϵ-pseudorandom links.

It is useful for us to consider measures on $V_1 \times \cdots \times V_k$ rather than pure $(k-1)$-dimensional complexes, which can be identified with subsets $S \subseteq V^k$. Instead we identify a set with the uniform measure over it.

Projected complexes

Let μ be a probability measure on $V_1 \times \cdots \times V_k$. We say that μ is a, weighted k-partite, $(k-1)$-dimensional complex. Let $S \subseteq [k]$ we write μ_S for the projection of μ on S. We write μ_i rather than $\mu_{\{i\}}$. We write V_S for the support of μ_S inside $\prod_{i \in S} V_i$. We write $\overline{S}$ for the complement of S.

Restricted complexes

Let $x \in V_S$. We write μ_x for the measure on $V_{\overline{S}}$ given by

$$\mu_x(y) = \frac{\mu(x, y)}{\mu_S(x)}.$$

We write V_x for the support of μ_x. We refer to (V_x, μ_x) as the *link* of μ on x.

ϵ-pseudorandom weighted graphs

Let V_1, V_2 be finite sets. A measure μ on $V_1 \times V_2$ can be thought of as a weighted bipartite graph. We say that μ is ϵ-pseudorandom if for each $f_1 \colon V_1 \to \mathbb{R}$, $f_2 \colon V_2 \to \mathbb{R}$ we have

$$\left|\mathbb{E}_{(x_1,x_2)\sim\mu}\left[f_1(x_1) f_2(x_2)\right] - \mathbb{E}_{x_1\sim\mu_1}\left[f_1(x_1)\right] \mathbb{E}_{x_2\sim\mu_2}\left[f_2(x_2)\right]\right| \leqslant \epsilon\sqrt{\operatorname{Var}_{x_1\sim\mu_1}\left[f_1(x_1)\right] \operatorname{Var}_{x_2\sim\mu_2}\left[f_2(x_2)\right]}.$$

We let A_{12} be the operator from $L^2(V_1, \mu_1)$ to $L^2(V_2, \mu_2)$ given by

$$A_{12} f(x) = \mathbb{E}_{y\sim\mu_x}\left[f(y)\right].$$

We have the following standard lemma.

Lemma 4.3.1. *The following are equivalent.*

1. *μ is ϵ-pseudorandom*
2. *$\|A_{12} - \mathbb{E}\|_{2\to 2} \leqslant \epsilon$.*
3. *The second eigenvalue of $\mathrm{A}_{12}^{*}\mathrm{A}_{12}$ is $\leqslant \epsilon^2$.*

ϵ-pseudorandom links

Now let μ on $V_1 \times \cdots \times V_k$. We say that μ has *ϵ-pseudorandom skeletons* if for each S of size 2 the measure μ_S is ϵ-pseudorandom.

We say that μ is *ϵ-product* if for each $S \subseteq [k]$ of size $\leqslant k-2$ and each $x \in V_S$ the link μ_x has ϵ-pseudorandom skeletons.

In all that follows we assume that μ is an ϵ-product measure on $V_1 \times \cdots \times V_k$.

Inheritance

The definition of ϵ-product makes it easy for inductive type argument for the following reason.

Lemma 4.3.2. *Let μ on $\prod_{i=1}^{k} V_i$ be ϵ-product. Let $S, T \subseteq [k]$ be disjoint. Then for each $x \in V_S$, the probability measure $(\mu_x)_T = (\mu_{S\cup T})_x$ is ϵ-product.*

Proof. All the skeletons of links of $(\mu_{S\to x})_T$ are also skeletons of links of μ. □

Pseudorandomness as a measure of independence

Let $S, T \subseteq [n]$. Then we have an operator $A_{S,T} : L^2(V_S, \mu_S) \to L^2(V_T, \mu_T)$. The operator is given by

$$A_{S,T} f(y) = \mathbb{E}_{x\sim\mu}\left[f(x_S) \,|\, x_T = y\right].$$

We write $A_{S,T}^{\mu}$ to stress that the operator is taken with respect to μ. We write A_S for $A_{[k],S}$, the operator given by restricting S and taking expectation.

When S, T are disjoint we expect $A_{S,T} f$ to be close to $\mathbb{E}[f]$, as in the product case $A_{S,T}$ is equal to the expectation. In fact, we do have the following.

Lemma 4.3.3. *Let μ be ϵ-product. Let $S, T \subseteq [k]$ be disjoint, and let $f \in L^2(V_S, \mu_S)$. We have*

$$\|A_{S,T} f - \mathbb{E}[f]\|_2^2 \leqslant |S|\,|T|\,\epsilon^2 \|f\|_2^2$$

Proof. We prove it by induction on k. The case where $k = 2$ is Lemma 4.3.1, so we assume $k > 2$. Given a probability space (Ω, μ) we write $1^{\perp}$ for the subspace of $L^2(\Omega, \mu)$ consisting of functions that are orthogonal to the constant function 1. We write $\widetilde{\|}A_{S,T}\widetilde{\|}$ for the L^2 operator norm of $A_{S,T}$ as an operator from $1^\perp$ to $1^\perp$. I.e.

$$\widetilde{\|}A_{S,T}\widetilde{\|} = \max_{f\in 1^\perp} \frac{\|A_{S,T}f\|_2}{\|f\|_2}.$$

Our goal is to show that

$$\widetilde{\|}A_{S,T}\widetilde{\|} \leqslant \sqrt{|S|\,|T|}\epsilon.$$

Discarding the trivial cases

If $T = \varnothing$, then $A_{S,T} = \mathbb{E}$ and the result is trivial. If $S \cup T \neq [k]$ the result follows by working with the space $(V_{S\cup T}, \mu_{S\cup T})$ rather then $\left(V_{[k]}, \mu_{[k]}\right)$. We also have $\widetilde{\|}A_{S,T}\widetilde{\|} = \widetilde{\|}A^*_{S,T}\widetilde{\|}$ as $A_{S,T}1 = 1$. As $A^*_{S,T} = A_{T,S}$ we may assume that $|T| \leqslant |S|$. As $k > 2$ we may therefore assume that $|T| \geqslant 2$.

Completing the proof in the case where $|T| > 1$

Assume without loss of generality that $1 \in T$.

Let $f \in 1^\perp$. Using the fact that the equality

$$\|X\|_2^2 = \mathbb{E}\left[X\right]^2 + \|X - \mathbb{E}\left[X\right]\|_2^2$$

holds for every random variable X we have

$$\begin{aligned}
\|A_{S,T}f\|_2^2 &= \mathbb{E}_{y\sim\mu_T}\mathbb{E}^2_{x\sim\mu_y}\left[f\left(x_S\right)\right] \\
&= \mathbb{E}_{a\sim\mu_1}\|A^{\mu_a}_{S,T\setminus\{1\}}f\|_2^2 \\
&= \mathbb{E}_{a\sim\mu_1}\left[\mathbb{E}^2_{\mu_a}f + \|A^{\mu_a}_{S,T\setminus\{1\}}f - \mathbb{E}_{\mu_a}f\|_2^2\right]. \\
&= \mathbb{E}_{a\sim\mu_1}\left[A_{S,1}f^2\left(a\right) + \|A^{\mu_a}_{S,T\setminus\{1\}}f - \mathbb{E}_{\mu_a}f\|_2^2\right]
\end{aligned}$$

By induction we may upper bound the right hand side we have

$$\begin{aligned}
RHS &\leqslant \mathbb{E}_{a\sim\mu_1}\left[A_{S,1}f^2\left(a\right) + |S|\,|T-1|\,\epsilon^2\|f\|^2_{L^2(\mu_a)}\right]. \\
&= \|A_{S,1}f\|_2^2 + |S|\,|T-1|\,\epsilon^2\|f\|_2^2. \\
&\leqslant |S| + |S|\,|T-1|\,\epsilon^2\|f\|_2^2 \\
&= |S|\,|T|\,\epsilon^2\|f\|_2^2
\end{aligned}$$

□

Understanding the operators $A_{S,T}$ and their compositions

We now deduce that we have a similar upper bound of the form

$$\|A_{S,T} - A_{S,S\cap T}\|_{2\to 2} \leqslant \sqrt{|S|\,|T|}\epsilon.$$

Corollary 4.3.4. *Let $S,T \subseteq [k]$, and let $f \in L^2(\mu_S)$. Then*

$$\|A_{S,T}f - A_{S,S\cap T}f\|_2^2 \leqslant |S|\,|T|\,\epsilon^2\|f\|_2^2.$$

Proof. Lemma 4.3.3 covers the case $S \cap T = \varnothing$. This shows that the corollary is true in μ_x for each $x \in V_{S\cap T}$. Therefore

$$\begin{aligned}\|A_{S,T}f - A_{S,S\cap T}f\|_2^2 &= \mathbb{E}_{x\sim\mu_{S\cap T}}\|A^{\mu_x}_{S\setminus T,T\setminus S}f - A^{\mu_x}_{S\setminus T,\varnothing}f\|^2_{L^2(\mu_x)} \\ &\leqslant |S|\,|T|\,\epsilon^2\mathbb{E}_x\|f\|^2_{L^2(\mu_x)} \\ &= |S|\,|T|\,\epsilon^2\|f\|_2^2.\end{aligned}$$

□

We now show that compositions behave similarly to the product space setting.

Lemma 4.3.5. *We have*

$$\|A_{T_2}A_{T_1} - A_{T_1\cap T_2}\|_{2\to 2} \leqslant |T_1|\,|T_2|\,\epsilon.$$

Proof. We may assume that $T_1 \cap T_2 = \varnothing$. Indeed, if the lemma holds for $T_1 \cap T_2 = \varnothing$ then it holds in general. Indeed, write

$$\widetilde{T}_1 = T_1 \setminus T_2, \widetilde{T}_2 = T_2 \setminus T_1, A = [k] \setminus (T_1 \cap T_2).$$

Let $x \in V_{T_1\cap T_2}$. Then we have

$$(A_{T_2}A_{T_1}f)(x,\cdot) = \left(A^{\mu_x}_{\widetilde{T}_2}A^{\mu_x}_{\widetilde{T}_1}\right)(f(x,\cdot))$$

and

$$A_{T_1\cap T_2}f(x,\cdot) = \mathbb{E}_{y\sim\mu_x}[f(x,y)].$$

Therefore once we prove the case $T_1 \cap T_2 = \varnothing$ it would imply that for each x

$$\mathbb{E}_{y\sim\mu_x}\left(A_{T_2}A_{T_1}f(x,y) - A_{T_1\cap T_2}f(x,y)\right)^2 \leqslant |T_1|\,|T_2|\,\epsilon^2\mathbb{E}_{y\sim\mu_x}f(x,y)^2.$$

The lemma will then follow by taking expectations over x.

Let us now settle the case $T_1 \cap T_2 = \varnothing$. Write $\mathrm{T} = A_{T_2}A_{T_1}$. Then

$$\mathrm{T} = A_{T_1,T_2}A_{T_1}.$$

Write $g = A_{T_1}f$. We have $\|g\|_2 \leqslant \|f\|_2$ by Cauchy–Schwarz. By Lemma 4.3.3 we have

$$\begin{aligned}\|\mathrm{T}f - \mathbb{E}[f]\|_2^2 &= \|A_{T_1,T_2}g - \mathbb{E}g\|_2^2 \\ &\leqslant |T_1|\,|T_2|\,\epsilon^2\|g\|_2^2 \\ &\leqslant |T_1|\,|T_2|\,\epsilon^2\|f\|_2^2.\end{aligned}$$

□

4.4 Efron–Stein decompositions for link expanders

In this section, we introduce a new approximate Efron–Stein decomposition for high dimensional expanders. In fact, it is more convenient to state and prove our results in the more general setting of ϵ-product spaces, of which high dimensional expanders are a special case. We proceed to discuss this setting below.

We first define the Efron–Stein decomposition via the usual formula for it.

Definition 4.4.1. Let $f \in L^2(V, \mu)$ and $S \subseteq [n]$. We write

$$f^{=S} = \sum_{T \subseteq [S]} (-1)^{|S \setminus T|} A_T f.$$

The functions $f^{=S}$ are defined in terms of the operators A_T. L^2-wise the composition of the operators $\{A_T\}_{T \subseteq [k]}$ behave similarly to the compositions in the product case setting. We satrt this section by making use of that and showing that many known facts from the product setting generalize to the ϵ-product setting up to a small error.

4.4.1 L^2-approximations for the Efron–Stein decomposition

Thinking of ϵ as tending to 0 in a much quicker pace than $\frac{1}{k}$. Our goal is now to show that if μ is ϵ-product, then we have:

1. $$\left| \|f\|_2^2 - \sum_{S \subseteq [k]} \|f^{=S}\|_2^2 \right| = o\left(\|f\|_2^2\right),$$

2. and more generally
$$\left| \langle f, g \rangle - \sum_{S} \left\langle f^{=S}, g^{=S} \right\rangle \right| = o\left(\|f\|_2 \|g\|_2\right).$$

One main tool involves the notion of a junta. We say that $g \colon V \to \mathbb{R}$ is a *T-junta* if $g(x)$ depends only on x_T. Equivalently, g is a T-junta if $A_T g = g$.

Our first step towards the proof is a near orthogonality result between $f^{=T}$ and $g^{=S}$ for $T \neq S$.

We start by a Fourier formula that holds exactly, this is unlike most of the results in this section that only generalize the situation from the product space setting up to a small error term.

Lemma 4.4.2. *We have*

$$A_S[f] = \sum_{T \subseteq S} f^{=T}(x).$$

In particular $f = \sum_{S \subseteq [k]} f^{=S}$.

Proof. We have

$$\begin{aligned}\sum_{T\subseteq S} f^{=T} &= \sum_{T\subseteq S}\sum_{T'\subseteq T} (-1)^{|T\setminus T'|} A_{T'}f \\ &= \sum_{T'\subseteq S} A_{T'}f \sum_{T'\subseteq T\subseteq S} (-1)^{|T\setminus T'|} \\ &= A_S f,\end{aligned}$$

where the last equality follows from the fact that whenever $T' \neq S$ and $i \in S \setminus T'$ the pairs

$$(T, T\Delta\{i\})$$

contribute opposing signs to the sum $\sum_{T'\subseteq T\subseteq S} (-1)^{|T\setminus T'|}$. The 'in particular' part follows by taking $S = [k]$. □

The following lemma holds even without assuming that μ is ϵ-product.

Lemma 4.4.3. *We have* $\|A_{S,T}\|_{2\to 2} \leqslant 1$ *and*

$$\|f^{=S}\|_2 \leqslant 2^{|S|}\|f\|_2.$$

Proof. The triangle inequality implies that it suffices to prove the former claim. Now by Cauchy–Schwarz we have

$$\begin{aligned}\|A_{S,T}f\|_2^2 &= \mathbb{E}_{x\sim\mu_T} A_{S,T}f(x)^2 \\ &= \mathbb{E}_{x\sim\mu_T}\left(\mathbb{E}_{y\sim(\mu_x)_T} f(y)\right)^2 \\ &\leqslant \mathbb{E}_{x\sim\mu_T}\mathbb{E}_{y\sim(\mu_x)_T} f(y)^2 \\ &= \|f\|_2^2.\end{aligned}$$

□

Lemma 4.4.4. *Let* $f\colon V \to \mathbb{R}$*,* T *be a set not containing* S*, and* g *be a* T*-junta. Then*

$$\left\langle f^{=S}, g\right\rangle \leqslant \epsilon\sqrt{|S|\,|T|}2^{|S|}\|f\|_2\|g\|_2.$$

Proof. As A_T is the dual to the inclusion operator $L^2(V_T) \to L^2\left(V_{[k]}\right)$ we have

$$\left\langle f^{=S}, g\right\rangle = \left\langle A_T f^{=S}, g\right\rangle.$$

By Cauchy–Schwarz it is sufficient to show that

$$\|A_T f^{=S}\|_2 \leqslant \epsilon\,|S|\,|T|\,2^{|S|}\|f\|_2.$$

Now

$$A_T f^{=S} = \sum_{S'\subseteq S} (-1)^{|S\setminus S'|} A_T A_{S'} f.$$

Roughly speaking, we rely on Lemma 4.4.2, which says that $\|A_T A_{S'} - A_{T\cap S'}\|_{2\to 2}$ is small together with the fact that

$$\sum_{S'\subseteq S} (-1)^{|S\setminus S'|} A_{T\cap S'} f = 0. \tag{4.3}$$

The equality follows by choosing an arbitrary $i \in S\setminus T$ and noting that the sets $(S', S'\Delta\{i\})_{S'\subseteq S}$ correspond to the same term $A_{[k],T\cap S'}$, while appearing with opposite signs. This shows that we have

$$A_T f^{=S} = \sum_{S'\subseteq S} (-1)^{|S\setminus S'|} \left(A_T A_{S'} f - A_{T\cap S'} f\right).$$

By Lemma 4.3.5 we have

$$\|A_T A_{S'} f - A_{T\cap S'} f\|_2 \leqslant \sqrt{|T|\,|S|}\epsilon\|f\|_2 \leqslant \sqrt{|S|\,|T|}\epsilon\|f\|_2.$$

Hence,

$$\|A_T f^{=S}\|_2 \leqslant \sqrt{|S|\,|T|}2^{|S|}\epsilon.$$

□

Proof of our near orthogonality result

Corollary 4.4.5. *Let $T \neq S$. Then $\left\langle f^{=S}, g^{=T}\right\rangle \leqslant 2^{2|S|+2|T|}\epsilon\|f\|_2\|g\|_2$.*

Proof. The function $g^{=T}$ is a T-junta. By Lemmas 4.4.4 and 4.4.3 we therefore have the following chain of inequalities if T does not contain S.

$$\left\langle f^{=S}, g^{=T}\right\rangle \leqslant \epsilon\sqrt{|S|\,|T|}2^{|S|}\|f\|_2\|g^{=T}\|_2 \leqslant \epsilon 2^{2|S|+2|T|}\|f\|_2\|g\|_2.$$

A similar chain of inequalities holds when S does not contain T. □

Parseval holds approximately for the Efron–Stein decomposition

Lemma 4.4.6. *We have*

$$\left|\langle f, g\rangle - \sum_{S\subseteq[k]} \left\langle f^{=S}, g^{=S}\right\rangle\right| \leqslant 2^{4k}\epsilon\|f\|_2\|g\|_2.$$

Moreover, if f is a T-junta, then

$$\left|\langle f, g\rangle - \sum_{S\subseteq T} \left\langle f^{=S}, g^{=S}\right\rangle\right| = 2^{4|T|}\epsilon\|f\|_2\|g\|_2.$$

Proof. We have $\langle f, g\rangle = \sum_{S\subseteq[k]} \left\langle f^{=S}, g^{=S}\right\rangle + \sum_{S\neq T} \left\langle f^{=S}, g^{=T}\right\rangle$. By corollary 4.4.5 we have

$$\sum_{S\neq T} \left\langle f^{=S}, g^{=T}\right\rangle \leqslant 2^{4k}\epsilon\|f\|_2\|g\|_2.$$

For the 'moreover' part note that if f is a T-junta, then

$$\langle f, g\rangle = \langle f, A_T g\rangle_{L^2(\mu_T)}.$$

We may then apply the first part of the lemma in μ_T noting that $(A_T g)^{=T'} = g^{=T'}$ for each $T' \subseteq T$. □

$\left(f^{=S}\right)^{=S}$ is L^2-close to $f^{=S}$.

In the product space setting we have $\left(f^{=S}\right)^{=T} = \begin{cases} f^{=S} & T = S \\ 0 & T\neq S\end{cases}$. Here we have the following instead:

Lemma 4.4.7. *Let $g = f^{=S}$. Then:*

1. *If $S \neq T$, then*
$$\|g^{=T}\|_2^2 \leqslant 2^{8k}\epsilon^2\|f\|_2^2$$

2.
$$\|g^{=S} - g\|_2^2 \leqslant 2^{10k}\epsilon^2\|f\|_2^2.$$

Proof. We have

$$g^{=T} = \sum_{T'\subseteq T, S'\subseteq S} (-1)^{|S\setminus S'|+|T\setminus T'|} A_{T'}A_{S'}f.$$

Write

$$h = \sum_{T'\subseteq T, S'\subseteq S} (-1)^{|S\setminus S'|+|T\setminus T'|} A_{T'\cap S'}f.$$

By Lemma 4.3.4 we therefore have

$$\|h - g^{=T}\|_2 \leqslant 2^{2k} \max_{T',S'} \|A_{T'}A_{S'} - A_{T'\cap S'}\|_{2\to 2}\|f\|_2 \leqslant 2^{4k}\epsilon\|f\|_2.$$

Now we claim that $h = 0$. Indeed, assume without loss of generality that T is not contained in S and let $i \in T\setminus S$. Then the terms $A_{T'\cap S'}$ appears with opposing sums for the pairs T' and $T'\Delta\{i\}$.

(2)-follows by the fact that

$$\|g^{=S} - g\|_2 = \left\|\sum_{T\neq S} g^{=T}\right\|_2 \leqslant \sum_{T\neq S}\|g^{=T}\|_2 \leqslant 2^{5k}\epsilon\|f\|_2.$$

□

4.4.2 Approximate Efron-Stein decomposition

Again think of ϵ as tending to 0 much more quickly than $\frac{1}{k}$. We now define a notion of (α, ϵ')-approximate Efron–Stein decomposition. We show that a version of Lemma 4.4.6 still holds for these approximate Efron–Stein decompositions.

Motivation

One reason that demonstrates our need for an approximate Efron–Stein decomposition is as follows. Let $f^{\leqslant d} = \sum_{|S|<d} f^{=S}$. Then we do not have

$$\left(f^{\leqslant d}\right)^{=S} = \begin{cases} f^{=S} & |S| \leqslant d \\ 0 & |S| > d \end{cases},$$

but we would nevertheless like to work with the decomposition $\left\{f^{=S}\right\}_{|S|\leqslant d}$ as an approximate Efron–Stein decomposition for f. We capture that notion as follows.

Defining the (α, ϵ')-approximate Efron–Stein decomposition

Definition 4.4.8. We say that $\{f_S\}_{S\subseteq[k]}$ is an (α, ϵ')-approximate Efron–Stein decomposition if

1. $\|f\|_2 \leqslant \alpha$.
2. $$\|f - \sum_S f_S\|_2 < \epsilon',$$
3. For each S there exists h_S with $\|h_S\|_2 \leqslant \alpha$ and $$\|h_S^{=S} - f_S\|_2 \leqslant \epsilon'.$$

It turns out that we have an approximate Parseval theorem for every approximate Efron–Stein decomposition.

Lemma 4.4.9. *Let $\alpha_1, \alpha_2, \epsilon_1, \epsilon_2 > 0$. Suppose that f has an (α_1, ϵ_1)-bounded approximate Efron–Stein decomposition $\{f_S\}$ and g has an (α_2, ϵ_2)-bounded Efron–Stein decomposition $\{g_S\}$. Then*

$$\left|\langle f, g\rangle - \sum_S \langle f_S, g_S\rangle\right| \leqslant 2^{6k}\left(\epsilon_1\alpha_2 + \epsilon_2\alpha_1 + \epsilon\alpha_1\alpha_2\right).$$

Proof. For each $S \subseteq [k]$ let $\widetilde{f}_S, \widetilde{g}_S$ be with $\|\widetilde{f}_S\|_2 \leqslant \alpha_1, \|\widetilde{g}_S\|_2 \leqslant \alpha_2$

$$\|\widetilde{f}_S^{=S} - f_S\|_2 \leqslant \epsilon_1,$$

and

$$\|\widetilde{g}_S^{=S} - g_S\|_2 \leqslant \epsilon_2.$$

Let

$$f'_S = \widetilde{f}_S^{=S}, g'_S = \widetilde{g}_S^{=S},$$

$$f' = \sum_{S \subseteq [k]} f'_S$$

and

$$g' = \sum_{S \subseteq [k]} g'_S.$$

By Lemma 4.4.5 we have

$$\begin{aligned} \langle f', g' \rangle &= \sum_S \langle f'_S, g'_S \rangle + \sum_{S \neq T \subseteq [k]} \langle f'_S, g'_T \rangle \\ &= \sum_S \langle f'_S, g'_S \rangle \pm 2^{6k} \epsilon \alpha_1 \alpha_2. \end{aligned}$$

Now by Cauchy–Schwarz

$$\begin{aligned} \langle f, g \rangle &= \langle f', g' \rangle + \langle f', g - g' \rangle + \langle f - f', g \rangle \\ &= \sum_S \langle f'_S, g'_S \rangle \pm \left(2^{6k} \epsilon \alpha_1 \alpha_2 + \|f'\|_2 \|g - g'\|_2 + \|f - f'\|_2 \|g\|_2\right) \\ &= \sum_S \langle f'_S, g'_S \rangle \pm \left(2^{6k} \epsilon \alpha_1 \alpha_2 + 2^{2k} \alpha_1 \epsilon_2 + \epsilon_1 \alpha_2\right), \end{aligned}$$

where the last equality used

$$\|f'\|_2 \leqslant \sum \|f'_S\|_2 \leqslant 2^{k+|S|} \alpha_1 \leqslant 2^{2k} \alpha_1,$$

which follows from Lemma 4.4.3.

To complete the proof we note that we similarly have

$$\begin{aligned} \langle f'_S, g'_S \rangle &= \langle f_S, g_S \rangle \pm \|f_S\|_2 \|g_S - g'_S\|_2 + \|f'_S - f_S\|_2 \|g'_S\|_2 \\ &= \langle f_S, g_S \rangle \pm \alpha \epsilon_2 + 2^k \epsilon_1 \alpha_2. \end{aligned}$$

□

The above approximate Efron–Stein decomposition works well when we care about L_2-norms. We actually care about closeness in higher norms specifically 4-norms. Our strategy when wishing to upper bound $\|f - f'\|_4$ is to use the inequality

$$\|f - f'\|_4^4 \leqslant \|f - f'\|_2^2 \left(\|f\|_\infty + \|f'\|_\infty\right).$$

Where we hope that the L^2-closeness is sufficient to overcome the loss of using infinity norms. We would therefore like everything to have a relatively small infinity norm.

Definition 4.4.10. We say that $\{f_S\}$ is a $(\beta, \alpha, \epsilon')$-*bounded approximate Efron-Stein* decomposition if it is an (α, ϵ')-approximate Efron–Stein decomposition and moreover for each S:

$$\|h_S^{=S}\|_\infty, \|f_S\|_\infty, \|f\|_\infty$$

are all $\leqslant \beta$. Here $h_S^{=S}$ is as in Definition 4.4.8.

We now show that the different Efron–Stein decompositions of a function f are all close in L_4.

Lemma 4.4.11. *Suppose that $\{f_S\}, \{f'_S\}$ are $(\beta, \alpha, \epsilon')$-bounded approximate Efron–Stein decompositions for f. Then*

1. $$\|f_S - f'_S\|_2^2 \leqslant O_k\left(\epsilon'\right)^2 + O_k\left(\epsilon\alpha^2\right),$$

2. $$\|f_S - f'_S\|_4^4 \leqslant O_k\left(\epsilon'^2\beta^2\right) + O_k\left(\epsilon\alpha^2\beta^2\right),$$

3. $$\|\sum_S \left(f_S - f'_S\right)\|_4^4 \leqslant O_k\left(\epsilon'^2\beta^2\right) + O_k\left(\epsilon^2\alpha^2\beta^2\right),$$

4. *and*
$$\|f - \sum_{S\subseteq[k]} f_S\|_4^4 \leqslant O_k\left(\epsilon^2\beta^2\right)\left(\alpha^2 + \|f\|_2^2\right).$$

Proof. (3) is an immediate corollary of (2). (4) also follows immediately from (3) by setting $f'_S = f^{=S}$ while applying it with $2^k\beta$ rather than β. Indeed, $\|f^{=S}\|_\infty \leqslant 2^k\|f\|_\infty \leqslant 2^k\beta$. Therefore $\left\{f^{=S}\right\}_{S\subseteq[k]}$ is a $\left(2^k\beta, \alpha, 0\right)$-approximate Efron–Stein decomposition for f. (2) follows immediately from (1) as we have

$$\|f_S - f'_S\|_4^4 \leqslant \|f_S - f'_S\|_2^2\|f_S - f'_S\|_\infty^2$$

and $\|f_S - f'_S\|_\infty^2 \leqslant 4\beta^2$.

We now prove (1).

Reducing to the case that $f'_S = f^{=S}$

First we assert that we may assume that $f'_S = f^{=S}$ for each S. Indeed, $\left\{f^{=S}\right\}$ is a $(\beta, \alpha, 0)$-Efron–Stein decomposition. By the triangle inequality we have

$$\|f_S - f'_S\|_2 \leqslant \|f_S - f^{=S}\|_2 + \|f^{=S} - f'_S\|_2,$$

which implies (by Hólder) that

$$\|f_S - f'_S\|_2^2 \leqslant 2\|f_S - f^{=S}\|_2^2 + 2\|f^{=S} - f'_S\|_2^2.$$

This shows that it is sufficient to prove the theorem when $\{f_S\} = \left\{f^{=S}\right\}$ and when $\{f'_S\} = \left\{f^{=S}\right\}$. Without loss of generality we may assume that $f'_S = f^{=S}$.

Reducing to the case that $f_S = h_S^{=S}$

Let h_S be with $\|h_S\|_2 \leqslant \alpha$ and $\|f_S - h_S^{=S}\|_2 < \epsilon'$. Setting $\widetilde{f_S} = h_S^{=S}$ we obtain by the triangle inequality that $\left\{\widetilde{f_S}\right\}_{S\subseteq[k]}$ is a $\left(\beta, \alpha, \left(2^k + 1\right)\epsilon'\right)$-bounded approximate Efron–Stein decomposition for f. We have

$$\|f_S - f^{=S}\|_2^2 \leqslant 2\|\widetilde{f_S} - f_S\|_2^2 + 2\|\widetilde{f_S} - f^{=S}\|_2 \leqslant 2\epsilon' + 2\|\widetilde{f_S} - f^{=S}\|_2.$$

Therefore it is sufficient to prove (1) when f_S is replaced by $\widetilde{f_S}$.

Proving the lemma when $f_S = h_S^{=S}$ and $f'_S = f^{=S}$

By Cauchy–Schwarz and Corollary 4.4.5 we have:

$$\begin{aligned}\left\langle f_S - f^{=S}, f\right\rangle &= \sum_{T\subseteq[k]} \left\langle f_S - f^{=S}, f^{=T}\right\rangle \\ &= \left\langle f_S - f^{=S}, f^{=S}\right\rangle + \sum_{T\neq S}\left\langle f^{=S} - h_S^{=S}, f^{=T}\right\rangle \\ &= \left\langle f_S - f^{=S}, f^{=S}\right\rangle + O_k\left(\epsilon\alpha^2\right).\end{aligned}$$

Again by Corollary 4.4.5 and Cauchy–Schwarz we have:

$$\begin{aligned}\left\langle f_S - f^{=S}, f\right\rangle &= \left\langle f_S - f^{=S}, \sum_T f_T\right\rangle + \left\langle f_S - f^{=S}, f - \sum_T f_T\right\rangle \\ &= \left\langle f_S - f^{=S}, f_S\right\rangle + O_k\left(\epsilon\alpha^2\right) + \|f_S - f^{=S}\|_2\epsilon'.\end{aligned}$$

Rearranging we obtain,

$$\|f_S - f^{=S}\|_2^2 \leqslant O_k\left(\epsilon'\right)\left(\|f^{=S} - f_S\|_2\right) + O_k\left(\epsilon\alpha^2\right).$$

This shows that

$$\|f_S - f^{=S}\|_2^2 \leqslant O_k\left(\epsilon'\right)^2 + O_k\left(\epsilon\alpha^2\right).$$

□

4.5 Proof overview

Building on the framework we established in Section 4, we can now give a proof overview for our hypercontactive inequality on high dimensional expanders. Recall that in the setting of direct products, we first prove a key lemma, (Lemma 4.2.8) and then use it to derive the theorem via an inductive argument. We now give a sketch of how to generalise this approach to the ϵ-product setting.

4.5.1 Generalising Lemma 4.2.8

Recall that we would like to show a lemma of the form

$$\|f\|_4^4 \leqslant C^d \|f\|_2^4 + \sum_S (4d)^{|S|} \left\| L_S\left[f\right] \right\|_4^4.$$

We instead show a similar lemma that holds up to a small error term of $O_k\left(\epsilon \|f\|_2^2 \|f\|_\infty^2\right)$:

$$\frac{1}{2}\|f^{\leqslant d}\|_4^4 \leqslant 9^d \|f^{\leqslant d}\|_2^4 + 4 \sum_{0<|T|\leqslant d} (4d)^{|T|} \left\| L_T^{\leqslant d}\left[f\right] \right\|_4^4 + O_k\left(\epsilon\right) \|f\|_2^2 \|f\|_\infty^2. \tag{4.4}$$

However, first note that we do not have a useful notion of a low degree function. Instead we work with

$$f^{\leqslant d} = \sum_{|S|\leqslant d} f^{=S}.$$

In turn, instead of $L_S\left[f\right]$ we have

$$L_S^{\leqslant d}\left[f\right] = \sum_{T\supseteq S, |T|\leqslant d} f^{=T}.$$

We show that when expanding

$$\left(\left(f^{\leqslant d}\right)^2\right)^{=S} = \sum_{T_1,T_2} \left(f^{=T_1} f^{=T_2}\right)^{=S},$$

there are three kinds of terms: (1) terms that vanish in the product space setting, but here they do not; (2) terms with $T_1 \cap T_2 \cap S \neq \varnothing$; and (3) terms with $T_1 \Delta T_2 = S$.

Our high-level approach is to show that the same proof as in the setting of product spaces works up to an error term. We accomplish that by expressing everything in terms of our operators $\{A_S\}$, and we then replace equalities that hold in the product space by L_2-approximation of the form

$$\|A_S A_T - A_{S\cap T}\|_{2\to 2} \leqslant O_k\left(\epsilon\right).$$

At first glance, it might appear that this approach would not suffice, as we eventually would like to upper bound 4-norms of terms, or 2-norms of expressions involving the product of two functions such as $\left(f^{=T_1}f^{=T_2}\right)^S$. Nevertheless, we are able to accomplish that via inequalities of the form

$$\|f\|_4^4 \leqslant \|f\|_2^2\|f\|_\infty^2.$$

We then use the fact that all our terms are bounded by $O_k\left(\|f\|_\infty\right)$, and our L_2-approximations involve ϵ, and therefore beat the $O_k\left(1\right)$-terms. This allows us to generalise Lemma 4.2.8 and prove (4.4).

4.5.2 Applying induction

After having an inequality of the form

$$\|f^{\leqslant d}\|_4^4 \leqslant C^d\|f^{\leqslant d}\|_2^2 + \sum_S \left(4d\right)^{|S|}\|L_S^{\leqslant d}\left[f\right]\|_4^4,$$

we would like to use a similar idea to the one we used in the product space setting; that is, restrict S to some $x \in V_S$, and then apply induction for the function $L_S^{\leqslant d}\left[f\right]\left(x,\cdot\right)$. The problem is that the restricted function $L_S^{\leqslant d}\left[f\right]\left(x,\cdot\right)$ is no longer of degree $d-|S|$, and hence we can no longer use induction.

We overcome this problem by using the notion of our approximate Efron–Stein decompositions. Namely, we show that $L_S^{\leqslant d}$ has two different approximate Efron–Stein decomposition. The first one is

$$\left\{f^{=T}\right\}_{T\supseteq S,|T|\leqslant d},$$

and the other one replaces $f^{=T}$ by the function f_T

$$\left(x,y\right) \mapsto \left(L_S\left[f\right]\left(x,\cdot\right)\right)^{=T\setminus S}\left(y\right).$$

We then obtain that $\sum_{|T|\supseteq S,|T|\leqslant d} f_T\left(x,\cdot\right)$ is of the form $D_{S,x}^{\leqslant d-|S|}$, which allows us to use induction similarly as in the product space setting.

After applying induction we get the compositions of two derivatives, and we are again able to translate them back to expressions of the form $\mathbb{E}_{x\sim\mu_S} I_{S,x}^2$ by showing that $D_{S,x}D_{T,y}$ and $D_{S\cup T,(x,y)}$ are both approximate Efron–Stein decompositions of the same expression.

The remaining step is to upper bound the influences. We achieve that by generalising the inequality

$$\mathbb{E}_{x\sim\mu_S}\left[I_{S,x}^2\right] \leqslant \delta\|L_S\left[f\right]\|_2^2$$

from the product space setting, where crucially, we obtain that without upper bounding $\|I_{S,x}\|_\infty$.

4.6 Laplacians, influences, and globalness on epsilon measures

In this section, we define the notions of laplacians, derivatives and influences in the setting of ϵ-measures, give bounded approximated Efron–Stein decompositions related to the Laplacians, define globalness, and show that it implies small influences.

4.6.1 Defining the Laplacians, derivatives and influences

Definition 4.6.1. We define the Laplacians via the formula

$$L_i[f] = f - A_{[k]\setminus\{i\}}f.$$

Lemma 4.6.2. *We have*

$$L_i[f] = \sum_{S\ni i} f^{=S}.$$

Proof. This follows immediately from Lemma 4.4.2, which shows that

$$A_{[k]\setminus\{i\}}[f] = \sum_{S\subseteq[k]\setminus\{i\}} f^{=S}.$$

□

Definition 4.6.3. We define $L_S[f] = \sum_{T\supseteq S} f^{=T}$. Alternatively,

$$L_S[f] = \sum_{T\subseteq S} (-1)^{|T|} A_{[k]\setminus T}f.$$

Let $x \in V_S$. We let $D_{S,x} = L_S[f](x,\cdot)$, i.e. the function in $L^2(V_x,\mu_x)$ obtained by plugging in x in the S coordinates. We let

$$I_{S,x}[f] = \|D_{S,x}[f]\|_{L^2(V_x,\mu_x)}.$$

4.6.2 Bounded approximated Efron–Stein decompositions related to the Laplacians

Lemma 4.6.4. . *There exists $C = O_k(1)$, such that $\{f^{=T}\}_{T\supseteq S}$ is a $(C\|f\|_\infty, C\|f\|_2, 0)$-bounded approximate Efron–Stein decomposition for $L_S[f]$.*

Proof. We have

$$\|f^{=S}\|_\infty \leqslant \sum_{T\subseteq S} \|A_T f\|_\infty \leqslant 2^{|S|}\|f\|_\infty$$

and

$$\|L_S[f]\|_\infty \leqslant \sum_{T\subseteq S} \|A_{\overline{T}} f\|_\infty \leqslant 2^{|S|}\|f\|_\infty.$$

The other properties are easy to verify. □

Let $f \in L^2\left(V_{[k]}, \mu_{[k]}\right)$ and let $g_T \in L^2\left(V_T, \mu_T\right)$ be given by

$$g_T(x) = I_{T,x}[f].$$

Then g_T can be interpreted interms of the Laplacians and the averaging operators as

$$g_T = A_T\left(L_T[f]^2\right).$$

Suppose that $\{f_S\}_{S\subseteq[k]}$ is a (β,α,ϵ')-bounded approximate Efron–Stein decompositions for f and set $\widetilde{L_T[f]} = \sum_{S\supseteq T} f_S$. The following lemma essentially shows that the function $A_T\left(\widetilde{L_T[f]}^2\right)$ is a good L_2-approximation for the function g_T. This can be interpreted by saying that the generalised influences could be computed via any (β,α,ϵ')-bounded approximate Efron–Stein decomposition for f.

Lemma 4.6.5. *Let $\{f_S\}$ and $\{f'_S\}$ be (β,α,ϵ')-bounded Efron-Stein decompositions for f. Then*

$$\|A_T\left(\sum_{S\supseteq T} f_S\right)^2\|_2^2 \leqslant 2\|A_T\left(\sum_{S'\supseteq T} f'_S\right)^2\|_2^2 + O_k\left(\epsilon'^2\beta^2\right) + O_k\left(\epsilon\alpha^2\beta^2\right).$$

Proof. By Cauchy–Schwarz we have

$$\left(\sum_{S\supseteq T} f_S\right)^2 \leqslant 2\left(\sum_{S\supseteq T} f'_S\right)^2 + 2\left(\sum_{S\supseteq T} f'_S - f_S\right)^2.$$

Therefore

$$\|A_T\left(\sum_{S\supseteq T} f_S\right)^2\|_2^2 \leqslant 2\|A_T\left(\sum_{S'\supseteq T} f'_S\right)^2\|_2^2 + 2\|A_T\left(\sum_{S\supseteq T} f'_S - f_S\right)^2\|_2^2.$$

Now since A_T contracts 2-norms (Lemma 4.4.3). We have

$$2\|A_T\left(\sum_{S\supseteq T} f'_S - f_S\right)^2\|_2^2 \leqslant 2\|\sum_{S\supseteq T} f'_S - f_S\|_4^4.$$

Lemma 4.4.11 now completes the proof. □

We now show that Lemma 4.6.4 is a special case of a more general phenomenon. Whenever $\{f_S\}_{S\subseteq[k]}$ is a (β,α,ϵ')-bounded approximate Efron–Stein decomposition for f, we obtain that $\{f_T\}_{T\supseteq S}$ is a $\left(\widetilde{\beta},\widetilde{\alpha},\widetilde{\epsilon}\right)$-bounded approximate Efron–Stein decomposition for suitable values of $\widetilde{\beta},\widetilde{\alpha},\widetilde{\epsilon}$. We show the following.

Lemma 4.6.6. *There exists $C = O_k(1)$, such that the following holds. Suppose that $\{f_T\}_{T\subseteq[k]}$ is a (β,α,ϵ')-Approximate Efron–Stein decomposition for f. Then $\{f_T\}_{T\supseteq S}$ is a $(C\beta,\alpha,C(\epsilon'+\alpha\sqrt{\epsilon}))$-Approximate Efron–Stein decomposition for $L_S[f]$.*

Proof. The only requirements that are not automatically inherited from f are the upper bounds on $\|L_S[f]\|_\infty$, and on $\|L_S f - \sum_{T\supseteq S} f_T\|_2$. The former inequality follows from the inequality

$$\|L_S[f]\|_\infty \leqslant 2^{|S|}\|f\|_\infty \leqslant 2^k\beta.$$

While the latter follows from Lemma 4.4.11 and the triangle inequality:

$$\begin{aligned}\|L_S f - \sum_{T\supseteq S} f_T\|_2 &= \|\sum_{T\supseteq S}\left(f^{=T} - f_T\right)\|_2.\\ &\leqslant \sum_{T\supseteq S}\|f^{=T} - f_T\|_2\\ &\leqslant O_k(\epsilon') + O_k\left(\alpha\sqrt{\epsilon}\right).\end{aligned}$$

□

4.6.3 Low degree functions and truncations

Definition 4.6.7. We define the *low degree part* of f by setting

$$f^{\leqslant d} = \sum_{|S|\leqslant d} f^{=S}$$

we define the *low degree Laplacians* of f by setting

$$L_T^{\leqslant d}[f] = \sum_{S\supseteq T, |S|\leqslant d} f^{=T}.$$

We now show that if $\{f_T\}_{T\subseteq[k]}$ is a (β,α,ϵ')bounded approxiate Efron–Stein decomposition for f, then we may turn it into an Efron–Stein decomposition for $f^{\leqslant d}$ and $L_S^{\leqslant d}[f]$ in the obvious way.

Lemma 4.6.8. *There exists $C = O_k(1)$, such that the following holds. Suppose that $\{f_T\}_{T\subseteq[k]}$ is a (β,α,ϵ')-Approximate Efron–Stein decomposition for f. Then*

1. *The functions* $\{f_T\}_{|T|\leqslant d}$ *are a* $(C\beta, \alpha, C\epsilon + C\epsilon')$*-approximate Efron–Stein decomposition for* $f^{\leqslant d}$*.*

2. *the functions* $\{f_T\}_{|T|\supseteq S, |T|\leqslant d}$ *are a* $(C\beta, \alpha, C\epsilon + C\epsilon')$*-approximate Efron–Stein decomposition for* $L_S^{\leqslant d}[f]$*.*

Proof. It is sufficient to prove (2) as (1) is the special case where $S = \varnothing$. By Lemma 4.4.11, we have

$$\|f_T - f^{=T}\|_2 \leqslant O_k(\epsilon') + O_k\left(\sqrt{\epsilon}\alpha\right).$$

Hence, by the triangle inequality we have

$$\begin{aligned}\|L_S^{\leqslant d}[f] - \sum_{T\supseteq S, |T|\leqslant d} f_T\|_2 &\leqslant \sum_{T\supseteq S, |T|\leqslant d} \|f^{=T} - f_T\|_2 \\ &= O_k(\epsilon') + O_k\left(\sqrt{\epsilon}\alpha\right).\end{aligned}$$

Moreover,

$$\begin{aligned}\|L_S^{\leqslant d}[f]\|_\infty &= \left\|\sum_{T\supseteq S, |T|\leqslant d} f^{=T}\right\|_\infty \\ &\leqslant 2^k \max_S \|f^{=S}\|_\infty \\ &\leqslant 4^k\|f\|_\infty \leqslant 4^k\beta.\end{aligned}$$

□

4.6.4 Globalness

Unlike the product space setting the two possible definitions of globalness are not equivalent. It turns out to be more convenient to work with the notion concerning the restrictions.

Definition 4.6.9. We say that f is (d, δ)-global if for each $|S| \leqslant d$ and each $x \in V_S$ we have

$$\|f(x, \cdot)\|_{L^2(V_x, \mu_x)} \leqslant \delta.$$

Claim 4.6.10. If f is (d, δ)-global and ϵ is sufficiently small, then for each T of size $\leqslant d$ we have

$$\|f^{=T}\|_\infty \leqslant 2^{|T|}\delta.$$

Proof. This follows from the triangle inequality once we show that $\|A_{T'}f\|_\infty \leqslant \delta$ for each $T' \subseteq T$. Indeed, for each x we have

$$A_{T'}f(x) = \mathbb{E}_{(V_x, \mu_x)} f(x, \cdot) \leqslant \|f(x, \cdot)\|_{L^2(V_x, \mu_x)} \leqslant \delta.$$

□

Lemma 4.6.11. *Suppose that f is (d,δ)-global. Then $\left\{f^{=S}\right\}_{|S|\leqslant d}$ is a $\left(k^d\delta, \|f\|_2, 0\right)$-bounded Efron–Stein decomposition for $f^{\leqslant d}$.*

Proof. We have $\|f^{=S}\|_\infty \leqslant 2^d\delta$ by Claim 4.6.10. We also have

$$\|f\|_\infty \leqslant \sum \|f^{=S}\|_\infty \leqslant k^d\delta.$$

The rest of the conditions hold automatically. □

Definition 4.6.12. We say that f is of (β,α)-degree d if $f = \sum_{|S|\leqslant d} f_S$ and $f_S = h_S^{=S}$ where $\|h_S\|_2 \leqslant \alpha$, $\|h_S^{=S}\|_\infty \leqslant \beta$ and $\|f\|_2 \leqslant \alpha, \|f\|_\infty \leqslant \beta$.

If $f = \sum_{|S|\leqslant d} f_S$ is of (β,α)-degree d as above, then $\{f_S\}$ is one $(\beta,\alpha,0)$-bounded Efron–Stein decomposition for f. We now show that in this case the canonical $\left\{f^{=S}\right\}_{|S|\leqslant d}$ is also $(\beta',\alpha',\epsilon')$-bounded Efron–Stein decomposition for the right parameters.

4.6.5 Other approximate Efron-Stein decompositions for $L_T[f], L_T^{\leqslant d}[f]$

Definition 4.6.13. We define the low degree derivatives for $T \subseteq [k]$ and $x \in V_T$

$$D_{T,x}^{\leqslant d} \colon L^2(\mu) \to L^2(V_x, \mu_x)$$

via

$$D_{T,x}^{\leqslant d}[f] = L_T^{\leqslant d}[f](x,\cdot)$$

The low degree influences for $T \subseteq [n]$ and $x \in V_T$ are defined by

$$I_{T,x}^{\leqslant d}[f] = \|L_T^{\leqslant d}[f](x,\cdot)\|_{L^2(V_x,\mu_x)}^2.$$

We now move on to the critical lemma for our inductive approach. In the product space setting our inductive approach relied on the fact that $D_{T,x}f$ is of degree $\leqslant d-|T|$ whenever f is of degree d. Here we show that $L_T[f]$ has an alternative (β,α,ϵ)-bounded approximate Efron–Stein decompositions $\{f_S\}_{S\supseteq T}$ that gives rise to a function

$$\widetilde{L_T^{\leqslant d}[f]} = \sum_{S\supseteq T, |S|\leqslant d} f^{=S}$$

with the property that for each x $\widetilde{L_T^{\leqslant d}[f]}(x,\cdot)$ is of degree $d-|T|$.

Lemma 4.6.14. *Let $f_S(x,y) = D_{T,x}^{=S\setminus T}[f](y)$. Then for each f:*

1. The set $\{f_S\}_{S\supseteq T}$ is a $(C\|f\|_\infty, C\|f\|_2, C\epsilon\|f\|_2)$-bounded approximate Efron–Stein decomposition for $L_T[f]$.

2. The set $\{f_S\}_{S\supseteq T,|S|\leqslant d}$ is a $(C\|f\|_\infty, C\|f\|_2, C\epsilon\|f\|_2)$-bounded approximate Efron–Stein decomposition for $L_T^{\leqslant d}[f]$.

3. If $T'\supseteq T$, then the set $\{f_S\}_{S\supseteq T'}$ is a $(C\|f\|_\infty, C\|f\|_2, C\epsilon\|f\|_2)$-bounded approximate Efron Stein decomposition for $L_{T'}[f]$.

4. The set $\{f_S\}_{S\supseteq T',|S|\leqslant d}$ is a $(C\|f\|_\infty, C\|f\|_2, C\epsilon\|f\|_2)$-bounded approximate Efron–Stein decomposition for $L_{T'}^{\leqslant d}[f]$.

5. If f is (d,δ)-global. Then $\{f_S\}_{S\supseteq T',|S|\leqslant d}$ is a $(C\delta, C\|f\|_2, C\epsilon\|f\|_2)$-bounded approximate Efron–Stein decomposition for $L_{T'}^{\leqslant d}[f]$.

Proof. Due to Lemma 4.6.8 (1) implies (2)-(4). By Lemma 4.4.3 all the operators A_I contract ∞-norms. We therefore have

$$\|f_S\|_\infty \leqslant \max_x 2^{|S\setminus T|}\|D_{T,x}f\|_\infty = 2^{|S\setminus T|}\|L_T f\|_\infty \leqslant 2^{|S|}\|f\|_\infty.$$

To complete the proof it is sufficient to show that

$$\|f_S - f^{=S}\|_2 \leqslant O_k\left(\epsilon\|f\|_2\right)$$

as this will also imply that

$$\begin{aligned}\|\sum_{S\supseteq T} f_S - L_T\left[f\right]\|_2 &= \|\sum_{S\supseteq T} f_S - \sum_{S\supseteq T} f^{=S}\|_2 \\ &= O_k\left(\epsilon\|f\|_2\right).\end{aligned}$$

We have

$$\begin{aligned} f_S &= \sum_{S'\subseteq S}(-1)^{|S\setminus S'|} A_{S'}L_T f. \\ &= \sum_{S'\subseteq S}\sum_{T'\subseteq T}(-1)^{|S\setminus S'|+|T'|} A_{S'}A_{[k]\setminus T'}f. \end{aligned}$$

Write

$$h = \sum_{S'\subseteq S}\sum_{T'\subseteq T}(-1)^{|S\setminus S'|+|T'|} A_{S'\setminus T'}f$$

Then by Lemma 4.3.4 we have

$$\|f_S - h\|_2 \leqslant O_k\left(\epsilon\right)\|f\|_2.$$

We then observe that whenever $S'\not\supseteq T$ the inner sum corresponding to it is 0. In this case there is some $i\in T\setminus S'$ and $T', T'\cup\{i\}$ appear with alternating signs and correspond to the same term $A_{S'\setminus T'}$. Therefore we have

$$h = \sum_{T\subseteq S'\subseteq S}\sum_{T'\subseteq T}(-1)^{|S\setminus S'|+|T'|} A_{S'\setminus T'}f = \sum_{S''\subseteq S}(-1)^{|S\setminus S''|} A_{S''}f = f^{=S}.$$

This shows that $\|f_S - f^{=S}\|_2 \leqslant O_k\left(\epsilon\right)\|f\|_2$, which completes the proof. ∎

4.6.6 Globalness implies small influences

In the product space setting we had $\|I_{T,x}\|_\infty \leqslant 4^d\delta^2$ and we used it via the inequality

$$\mathbb{E}_{x\sim\mu_T}\left[I_{T,x}^2\right] \leqslant \mathbb{E}_{x\sim\mu_T}\left[I_{T,x}\right] 4^d\delta^2 = 4^d\delta^2 \|L_T\left[f\right]\|_2^2. \tag{4.5}$$

See the proof of Corollary 4.2.6. Here we find a convoluted way of proving an analogue of (4.5) without having any upper bound on $\|I_{T,x}\|_\infty$ at our disposal.

Lemma 4.6.15. *Suppose that $f\colon V_{[k]} \to \mathbb{R}$ is (d,δ)-global, and let $|T| \leqslant d$. Then*

$$\mathbb{E}_{x\sim\mu_T}\left[\left(I_{T,x}\left[f\right]\right)^2\right] \leqslant 2^{d+1}\delta^2\mathbb{E}_{x\sim\mu_T} I_{T,x} + O_k\left(\epsilon^2\|f\|_4^4\right).$$

Proof. Write $g\left(x\right) = I_{T,x}\left[f\right]$. Then $g = A_T\left[\left(L_T\left[f\right]\right)^2\right]$. We would like to upper bound $\|g\|_2^2$. We accomplish that by upper bounding $\|g\|_2^2$ by $\mathbb{E}\left[gg''\right]$ for a function g'' with a small ∞-norm.

By Cauchy–Schwarz we have

$$L_T\left[f\right]^2 \leqslant 2^{|T|}\sum_{T'\supseteq\overline{T}} A_{T'}\left[f\right]^2.$$

This shows that

$$g \leqslant 2^{|T|}\sum_{T'\subseteq T} A_T\left[\left(A_{T'}f\right)^2\right], \tag{4.6}$$

on all x. Let us denote by g' the right hand side of 4.6. Also let

$$g'' = 2^{|T|}\sum_{T'\subseteq T} A_{T\cap T'}\left[\left(A_{T'}f\right)^2\right].$$

Then in the product space setting the functions g', g'' would have been equal. Here we have an L^2-approximation between them.

Claim 4.6.16. $\|g' - g''\|_2 \leqslant O_k\left(\epsilon\right)\|f\|_4^2$.

Proof. As $\left(A_{T'}f\right)^2$ is a T'-junta we have

$$\left(A_{T'}f\right)^2 = A_{T'}\left[\left(A_{T'}f\right)^2\right].$$

By Lemma 4.3.4 we have

$$\|A_T A_{T'} - A_{T\cap T'}\|_{2\to 2} \leqslant O_k\left(\epsilon\right).$$

We therefore have

$$\begin{aligned}\|A_T\left[\left(A_{T'}f\right)^2\right] - A_{T\cap T'}\left[\left(A_{T'}f\right)^2\right]\|_2^2 &\leqslant O_k\left(\epsilon^2\right)\|\left(A_{T'}f\right)^2\|_2^2\\ &\leqslant O_k\left(\epsilon^2\right)\|f\|_4^4,\end{aligned}$$

as $A_{T'}$ contracts 4-norms. Therefore,

$$\|g' - g''\|_2 \leqslant O_k\left(\epsilon\right)\|f\|_4^2.$$

□

As $0 \leqslant g \leqslant g'$ we have

$$\mathbb{E}\left[g^2\right] \leqslant \mathbb{E}\left[g'g\right] \leqslant \mathbb{E}\left[g''g\right] + \mathbb{E}\left[\left(g' - g''\right)g\right].$$

By Cauchy–Schwarz we have

$$\mathbb{E}\left[\left(g' - g''\right)g\right] \leqslant \|g' - g''\|_2 \|g\|_2 \leqslant O_k\left(\epsilon\right)\|f\|_4^2 \|g\|_2.$$

Now either

$$\mathbb{E}\left[g^2\right] \leqslant 2\mathbb{E}\left[g''g\right] \tag{4.7}$$

or

$$\mathbb{E}\left[g^2\right] \leqslant 2\mathbb{E}\left[\left(g' - g''\right)g\right] \leqslant O_k\left(\epsilon\right)\|f\|_4^2 \|g\|_2.$$

In the latter case we have

$$\|g\|_2^2 \leqslant O_k\left(\epsilon^2\right)\|f\|_4^4. \tag{4.8}$$

after rearranging. We can now sum the upper bounds of (4.7) and (4.8) corresponding to each of the cases to obtain the upper bound

$$\mathbb{E}\left[g^2\right] \leqslant 2\mathbb{E}\left[g''g\right] + O_k\left(\epsilon^2\right)\|f\|_4^4.$$

that is true in both cases. The following claim completes the proof.

Claim 4.6.17. $\|g''\|_\infty \leqslant 2^d\delta^2$.

Proof. By Cauchy–Schwarz we point-wise have $\left(A_{T'}f\right)^2 \leqslant A_{T'}\left(f^2\right)$. We therefore have

$$A_{T\cap T'}\left[\left(A_{T'}f\right)^2\right] \leqslant A_{T\cap T'}A_{T'}\left(f^2\right) = A_{T\cap T'}\left(f^2\right) \leqslant \delta^2.$$

This shows that $\|g''\|_\infty \leqslant 2^d\delta^2$. □

□

The same proof works for the truncated influences.

Lemma 4.6.18. *Suppose that $f\colon V_{[k]} \to \mathbb{R}$ is (d,δ)-global. Suppose additionally that $\epsilon \leqslant \epsilon_0\left(k\right)$. Then we have*

$$\mathbb{E}_{x\sim\mu_T}\left[\left(I_{T,x}^{\leqslant d}\left[f\right]\right)^2\right] \leqslant 2^{d+4}\delta^2\mathbb{E}_{x\sim\mu_T}\left[I_{T,x}^{\leqslant d}\right] + O_k\left(\epsilon^2\|f\|_\infty^2\|f\|_2^2\right).$$

Proof. Write

$$g_1\left(x\right) = \|\left(D_{T,x}\left[f\right]\right)^{\leqslant d-|T|}\|_2^2.$$

We now proceed with the following steps.

Upper bounding $\mathbb{E}_{x\sim\mu_T}\left[\left(I_{T,x}^{\leqslant d}[f]\right)^2\right]$ in terms of g_1

By Lemma 4.6.14 the functions

$$\left\{D_{T,x}[f]^{=S}\right\}_{|S|\leqslant d-|T|},$$

is an alternative $(O_k\|f\|_\infty, O_k\|f\|_2, O_k(\epsilon\|f\|_2))$-bounded approximate Efron–Stein decomposition for $L_T^{\leqslant d}[f]$. Therefore by Lemma 4.6.5 we have

$$\mathbb{E}_{x\sim\mu_T}\left[I_{T,x}^{\leqslant d}[f]^2\right]\leqslant 2\mathbb{E}_{x\sim\mu_T}\left[g_1(x)^2\right]+O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right). \tag{4.9}$$

Repeating the proof of Lemma 4.6.18

Now by Lemma 4.4.9 we have $g_1(x)\leqslant 2I_{T,x}$ for each x, provided that ϵ is sufficiently small. Write $g_2(x)=I_{T,x}[f]$. Similarly to the proof of Lemma 4.6.18 we let

$$g_2'=2^{|T|}\sum_{T'\subseteq T}A_T\left[(A_{T'}f)^2\right]$$

and let

$$g_2''=2^{|T|}\sum_{T'\subseteq T}A_{T\cap T'}\left[(A_{T'}f)^2\right].$$

By Cauchy–Schwarz we have

$$\|g_1\|_2^2\leqslant 2\mathbb{E}[g_1g_2]\leqslant 2\mathbb{E}[g_1g_2']\leqslant 2\mathbb{E}[g_1g_2'']+2\mathbb{E}[g_1(g_2'-g_2'')].$$

Now either

$$\|g_1\|_2^2\leqslant 4\mathbb{E}[g_2''g_1],$$

which would imply

$$\|g_1\|_2^2\leqslant 4\mathbb{E}[g_2''g_1]\leqslant 2^{d+2}\delta\mathbb{E}[g_1]$$

by Claim 4.6.17, or

$$\|g_1\|_2^2\leqslant 4\mathbb{E}[g_1(g_2'-g_2'')]\leqslant 4\|g_1\|_2\|g_2'-g_2''\|_2$$

and rearranging, we obtain

$$\|g_1\|_2^2\leqslant 16\|g_2'-g_2''\|_2^2\leqslant O_k\left(\epsilon^2\|f\|_4^4\right)$$

by Claim 4.6.16. This shows that

$$\|g_1\|_2^2\leqslant 2^{d+2}\delta\mathbb{E}[g_1]+O_k\left(\epsilon^2\|f\|_4^4\right). \tag{4.10}$$

Moving back from g_1 to $I_{T,x}^{\leqslant d}$

By Lemmas 4.6.14 we have

$$\| \left(D_{T,x}\left[f\right]\right)^{\leqslant d-|T|} - D_{T,x}^{\leqslant d}\left[f\right] \|_2^2 \leqslant O_k\left(\epsilon^2 \|f\|_2^2\right)$$

yielding

$$\begin{aligned} \mathbb{E}\left[g_1\right] &\leqslant 2\|D_{T,x}^{\leqslant d}\left[f\right]\|_2^2 + O_k\left(\epsilon^2\|f\|_2^2\right) \\ &= 2\mathbb{E}_{x\sim\mu_T} I_{T,x}^{\leqslant d}\left[f\right] + O_k\left(\epsilon^2\right)\|f\|_2^2 \end{aligned} \tag{4.11}$$

by the triangle inequality and Cauchy–Schwarz. By combining (4.9), (4.10) with (4.11) we obtain

$$\begin{aligned} \mathbb{E}_{x\sim\mu_T}\left[I_{T,x}^{\leqslant d}\left[f\right]^2\right] &\leqslant 2\|g_1\|_2^2 + O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right). \\ &\leqslant 2^{d+3}\delta^2\mathbb{E}\left[g_1\right] + O_k\left(\epsilon^2\|f\|_2^2\|f\|_\infty^2\right) + O_k\left(\epsilon^2\|f\|_4^4\right) \\ &\leqslant 2^{d+3}\delta^2\mathbb{E}\left[g_1\right] + O_k\left(\epsilon^2\|f\|_2^2\|f\|_\infty^2\right). \end{aligned}$$

The lemma now follows by putting everything together. □

4.7 Proving hypercontractivity for epsilon product measures

We suggest revisiting Section 2 before reading this section. Our strategy is the same as in the product case, and we deal with the differences by appealing to the tools developed in Sections 3-6.

4.7.1 Upper bounding $\|f^{\leqslant d}\|_4^4$ by 4-norms of non-trivial Laplacians and $\|f^{\leqslant d}\|_2^4$

We now move on to preparing the ground for the proof of our hypercontractive inequality.

Lemma 4.7.1. *Let f be (d,δ)-global. Suppose that $\epsilon \leqslant \epsilon_0(k)$. Then we have*

$$\frac{1}{2}\|f^{\leqslant d}\|_4^4 \leqslant 9^d\|f^{\leqslant d}\|_2^4 + 4\sum_{0<|T|\leqslant d}(4d)^{|T|}\|L_T^{\leqslant d}\left[f\right]\|_4^4 + O_k(\epsilon)\|f\|_2^2\|f\|_\infty^2.$$

Proof. Let $g = f^{\leqslant d}$. By Lemma 4.4.9 we have

$$\|g\|_4^4 \leqslant 2\sum_S \|\left(g^2\right)^{=S}\|_2^2.$$

We now upper bound $\| \left(g^2\right)^{=S} \|_2^2$. We have

$$\left(g^2\right)^{=S} = \sum_{|T_1|\leqslant d, |T_2|\leqslant d} \left(f^{=T_1} f^{=T_2}\right)^{=S}.$$

Let

1. $I_1 = \{(T_1, T_2) : T_1 \cap T_2 \cap S \neq \varnothing\}$.
2. $I_2 = ((T_1, T_2) : T_1 \Delta T_2 = S)$
3. $I_3 = (T_1 \Delta T_2) \setminus S \neq \varnothing$ or $S \setminus (T_1 \cup T_2) \neq \varnothing$.

Our first step is to show that the contribution from I_3 is negligible. This is to be expected as in the product space setting we were able to show that the contribution from I_3 is 0.

Claim 4.7.2. Let $(T_1, T_2) \in I_3$. Then

$$\| \left(f^{=T_1} f^{=T_2}\right)^{=S} \|_2^2 \leqslant O_k\left(\epsilon^2\right) \|f\|_2^2 \|f\|_\infty^2.$$

Proof. Suppose first that $(T_1 \Delta T_2) \setminus S \neq \varnothing$. Then without loss of generality we may assume that there is some $i \in T_1 \setminus (T_2 \cup S)$. By Lemma 4.4.9 we have

$$\| \left(f^{=T_1} f^{=T_2}\right)^{=S} \|_2^2 \leqslant 2 \|A_{[k]\setminus\{i\}} \left(f^{=T_1} f^{=T_2}\right) \|_2^2.$$

Now

$$A_{[k]\setminus\{i\}} \left(f^{=T_1} f^{=T_2}\right) = \left(A_{[k]\setminus\{i\}} \left(f^{=T_1}\right)\right) f^{=T_2}.$$

By Lemma 4.4.7 we have

$$\|A_{[k]\setminus\{i\}} f^{=T_1}\|_2 \leqslant \sum_{T' \not\ni i} \| \left(f^{=T_1}\right)^{=T'} \|_2 \leqslant O_k(\epsilon) \|f\|_2.$$

This shows that

$$\begin{aligned} \|A_{[k]\setminus\{i\}} \left(f^{=T_1} f^{=T_2}\right) \|_2^2 &\leqslant \|A_{[k]\setminus\{i\}} f^{=T_1}\|_2^2 \|f^{=T_2}\|_\infty^2 \\ &\leqslant O_k\left(\epsilon^2\right) \|f\|_\infty^2 \|f\|_2^2. \end{aligned}$$

Suppose now that $S \setminus (T_1 \cup T_2) \neq \varnothing$. Let $i \in S \setminus (T_1 \cup T_2)$. Then the function $g = f^{=T_1} f^{=T_2}$ is a $T_1 \cup T_2$-junta. This shows that $g = A_{T_1 \cup T_2} g$.

Hence by Lemma 4.4.7 and the triangle inequality we have

$$\begin{aligned} \|g^{=S}\|_2 &= \| \left(A_{T_1 \cup T_2} g\right)^{=S} \|_2 \\ &\leqslant \| \sum_{T \subseteq T_1 \cup T_2} \left(g^{=T}\right)^{=S} \|_2 \\ &\leqslant O_k(\epsilon) \|g\|_2. \end{aligned}$$

It now remains to note that $\|g\|_2 \leqslant \|f^{=T_1}\|_2 \|f^{=T_2}\|_\infty \leqslant 2^{2k} \|f\|_2 \|f\|_\infty$. □

We now move on to our next step of upper bounding the contribution from the pairs in I_1.

Claim 4.7.3. $\sum_{(T_1,T_2)\in I_1}\left(f^{=T_1}f^{=T_2}\right)^{=S} = \sum_{T\subseteq S}(-1)^{|T|+1}\|\left(L_T^{\leqslant d}[f]^2\right)^{=S}\|_2^2$.

Proof. The proof is exactly the same as in the product case so we omit it. ☐

It now remains to consider the contribution from I_2, i.e. the case $T_1\Delta T_2 = S$. Here just like the product case it is sufficient to show the following claim

Claim 4.7.4. Let $T_1\Delta T_2 = S$. Then we have

$$\|\left(f^{=T_1}f^{=T_2}\right)^{=S}\|_2 \leqslant 2\|f^{=T_1}\|_2\|f^{=T_2}\|_2 + O_k(\epsilon)\|f\|_2\|f\|_\infty,$$

provided that ϵ is sufficiently small.

Proof. First let $S'\subsetneq S$. As $T_1\Delta T_2 = S$, there exists $i\in(T_1\Delta T_2)\setminus S'$. Without loss of generality $i\in T_1$. By Lemmas 4.4.9, 4.4.7, and 4.4.2 we have

$$\begin{aligned}\|A_{S'}\left(f^{=T_1}f^{=T_2}\right)\|_2 &\leqslant 2\|A_{S'\cup T_2}\left(f^{=T_1}f^{=T_2}\right)\|_2\\ &\leqslant 2\|A_{S'\cup T_2}f^{=T_1}\|_2\|f^{=T_2}\|_\infty\\ &\leqslant O_k(\epsilon)\|f\|_2\|f\|_\infty.\end{aligned}$$

By the triangle inequality this shows that

$$\begin{aligned}\|\left(f^{=T_1}f^{=T_2}\right)^{=S}\|_2 &\leqslant \sum_{S'}(-1)^{|S\setminus S'|}A_{S'}\left(f^{=T_1}f^{=T_2}\right).\\ &\leqslant \|A_S\left(f^{=T_1}f^{=T_2}\right)\|_2 + O_k(\epsilon)\|f\|_2\|f\|_\infty\end{aligned}$$

We now upper bound $\|A_S\left(f^{=T_1}f^{=T_2}\right)\|_2$.

By Cauchy–Schwarz for $x\in V_S$ we have

$$\begin{aligned}A_S\left(f^{=T_1}f^{=T_2}\right)(x) &= \left\langle f^{=T_1}(x,\cdot), f^{=T_2}(x,\cdot)\right\rangle_{L^2(V_x,\mu_x)}\\ &\leqslant \|f^{=T_1}(x,\cdot)\|_{L^2(V_x,\mu_x)}\|f^{=T_2}(x,\cdot)\|_{L^2(V_x,\mu_x)}.\end{aligned}$$

This shows that

$$\|A_S\left(f^{=T_1}f^{=T_2}\right)\|_2^2 \leqslant \mathbb{E}_{x\sim\mu_S}\left[\|f^{=T_1}(x,\cdot)\|_{L^2(V_x,\mu_x)}^2\|f^{=T_2}(x,\cdot)\|_{L^2(V_x,\mu_x)}^2\right]. \tag{4.12}$$

We have

$$\|f^{=T_1}(x,\cdot)\|_{L^2(V_x,\mu_x)}^2 = A_S\left[\left(f^{=T_1}\right)^2\right] = A_SA_{T_1}\left(f^{=T_1}\right)^2$$

By Lemma 4.3.4

$$\|A_SA_{T_1} - A_{S\cap T_1}\|_{2\to 2} \leqslant O_k(\epsilon).$$

Hence,

$$\|A_S A_{T_1}\left(f^{=T_1}\right)^2 - A_{S\cap T_1}\left[\left(f^{=T_1}\right)^2\right]\|_2^2 \leqslant O_k\left(\epsilon^2\|f^{=T_1}\|_4^4\right)$$
$$\leqslant O_k\left(\epsilon^2\|f\|_2^2\|f\|_\infty^2\right).$$

By Cauchy–Schwarz this shows that

$$\begin{aligned}\text{RHS of (4.12)} &= \left\langle A_S A_{T_1}\left(f^{=T_1}\right)^2, A_S\left(f^{=T_2}\right)^2\right\rangle_{L^2(V_S,\mu_S)} \\ &= \left\langle A_{S\cap T_1}\left(f^{=T_1}\right)^2, A_S\left(f^{=T_2}\right)^2\right\rangle_{L^2(V_S,\mu_S)} \\ &+ O_k\left(\epsilon\|f\|_2\|f\|_\infty\right)\|A_S\left(f^{=T_2}\right)^2\|_2.\end{aligned}$$

As we have

$$\|A_S\left(f^{=T_2}\right)^2\|_2^2 \leqslant O_k\left(\|f\|_4^4\right) \leqslant O_k\left(\|f\|_2^2\|f\|_\infty^2\right).$$

Therefore,

$$\begin{aligned}\|A_S\left(f^{=T_1}f^{=T_2}\right)\|_2^2 &\leqslant \left\langle A_{S\cap T_1}\left(f^{=T_1}\right)^2, A_S\left(f^{=T_2}\right)^2\right\rangle_{L^2(V_S,\mu_S)} + O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right) \\ &= \left\langle A_{S\cap T_1}\left(f^{=T_1}\right)^2, A_{S\cap T_1}\left(f^{=T_2}\right)^2\right\rangle_{L^2(\mu)} + O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right).\end{aligned}$$

Now

$$A_{S\cap T_1}\left(f^{=T_2}\right)^2 = A_{S\cap T_1}A_{T_2}\left(f^{=T_2}\right)^2$$

and $\|A_{S\cap T_1}A_{T_2} - \mathbb{E}\|_{2\to 2} \leqslant \epsilon$ by Lemma 4.3.4. Therefore we similarly have

$$\begin{aligned}\left\langle A_{S\cap T_1}\left(f^{=T_1}\right)^2, A_{S\cap T_1}\left(f^{=T_2}\right)^2\right\rangle_{L^2(\mu)} &= \left\langle A_{S\cap T_1}\left(f^{=T_1}\right)^2, \|f^{=T_2}\|_2^2\right\rangle + O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right). \\ &= \|f^{=T_1}\|_2^2\|f^{=T_2}\|_2^2 + O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right).\end{aligned}$$

This completes the proof of the claim. □

The rest of the proof is the exactly the same as in the product case setting. □

Now the only thing to remains is to apply the inductive hypothesis.

Theorem 4.7.5. *We have* $\|f^{\leqslant d}\|_4^4 \leqslant 20^d \sum_{|S|\leqslant d}(4d)^{|S|}\mathbb{E}_{x\sim\mu_S} I_{S,x}^{\leqslant d}[f]^2 + O_k(\epsilon^2)\|f\|_2^2\|f\|_\infty^2$.

Proof. The proof is by induction on d. By Lemma 4.7.1 we have

$$\|f^{\leqslant d}\|_4^4 \leqslant 2\cdot 9^d\|f^{\leqslant d}\|_2^4 + 2\cdot\sum_{S\neq\varnothing}(4d)^{|S|}\|L_S^{\leqslant d}[f]\|_4^4 + O_k(\epsilon)\|f\|_2^2\|f\|_\infty^2. \tag{4.13}$$

Write $g_{S,x}(y) = \left(D_{S,x}[f]\right)^{\leqslant d-|T|}(y)$. Then by Lemma 4.6.14 and Lemma 4.4.11 we have:

$$\mathbb{E}_x\|D_{S,x}^{\leqslant d}[f]\|_4^4 \leqslant 2\mathbb{E}_x\|g_{S,x}\|_4^4 + O_k\left(\epsilon^2\right)\|f\|_2^2\|f\|_\infty^2.$$

By induction, we have

$$\|g_{S,x}\|_4^4 \leqslant 20^{d-|S|} \sum_{T\cap S=\varnothing,|T|\leqslant d-|S|} (4d)^{|T|}\,\mathbb{E}_{y\sim\mu_T} I_{T,y}^2\left[g_{S,x}\right] + O_k\left(\epsilon\right)\|g_{S,x}\|_2^2\|g_{S,x}\|_\infty^2. \tag{4.14}$$

By Lemma 4.6.14 we have $\|g_{S,x}\|_\infty = O_k\left(\|f\|_\infty\right)$. By Lemmas 4.6.14, 4.4.11, and 4.4.9 we have

$$\begin{aligned}\mathbb{E}_{x\sim\mu_S}\|g_{S,x}\|_2^2 &\leqslant 2\mathbb{E}_{x\sim\mu_S}\|D_{S,x}^{\leqslant d} f\|_2^2 + O_k\left(\epsilon^2\right)\|f\|_2^2 \\ &\leqslant O_k\left(\|f\|_2^2\right)\end{aligned} \tag{4.15}$$

Taking expectations over (4.14), and plugging in (4.15) we obtain:

$$\mathbb{E}_x\|g_{S,x}\|_4^4 \leqslant 2\sum_{T\cap S=\varnothing,|T|\leqslant d-|S|}(4d)^{|T|}\,\mathbb{E}_{(x,y)\sim\mu_{S\cup T}} I_{T,y}^2\left[g_{S,x}\right] + O_k\left(\epsilon\right)\|f\|_2^2\|f\|_\infty^2.$$

By Lemmas 4.6.14 and 4.4.11 we have

$$\mathbb{E}_{(x,y)\sim\mu_{S\cup T}} I_{T,y}^2\left[g_{S,x}\right] = \mathbb{E}_{z\sim\mu_{T\cup S}} I_{T\cup S,z}^2\left[g\right] + O_k\left(\epsilon^2\right)\|f\|_2^2\|f\|_\infty^2.$$

Hence,

$$\mathbb{E}_x\|g_{S,x}\|_4^4 \leqslant 20^{d-|S|}\sum_{S'\supseteq S|S'|\leqslant d}(4d)^{|S'\setminus S|}\,\mathbb{E}_{z\sim\mu_{S'}}\left(I_{S',z}^{\leqslant d}\left[f\right]\right)^2 + O_k\left(\epsilon\right)\|f\|_2^2\|f\|_\infty^2.$$

This gives

$$\mathbb{E}_x\|D_{S,x}^{\leqslant d}\left[f\right]\|_4^4 \leqslant 2\cdot 20^{d-|S|}\sum_{S'\supseteq S|S'|\leqslant d}(4d)^{|S'\setminus S|}\,\mathbb{E}_{z\sim\mu_{S'}}\left(I_{S',z}^{\leqslant d}\left[f\right]\right)^2 + O_k\left(\epsilon\right)\|f\|_2^2\|f\|_\infty^2$$

The proof is now completed by plugging this inequality in (4.13). Indeed, we have

$$\begin{aligned}\|f^{\leqslant d}\|_4^4 &\leqslant 2\cdot 9^d\|f\|_2^4 + O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right) \\ &\quad+ \sum_{0<|S|\leqslant d}(4d)^{|S|}\cdot 2\cdot 20^{d-|S|}\sum_{S'\supseteq S|S'|\leqslant d}(4d)^{|S'\setminus S|}\,\mathbb{E}_{z\sim\mu_{S'}}\left(I_{S',z}^{\leqslant d}\left[f\right]\right)^2 \\ &\leqslant 20^d\sum_{|S'|\leqslant d}(4d)^{|S|}\,\mathbb{E}_{z\sim\mu_{S'}}\left(I_{S',z}^{\leqslant d}\left[f\right]\right)^2 + O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right).\end{aligned}$$

□

4.7.2 The case where $\|f\|_\infty$ is large

Here we show a hypercontractive inequality whose error term does not include the factor $\|f\|_\infty$. This may be useful when $\|f\|_\infty$ is significantly larger than δ.

Theorem 4.7.6. *Suppose that f is (d,δ)-global, then*

$$\|f^{\leqslant d}\|_4^4 \leqslant 20^{d+1} \sum_{|S|\leqslant d} (4d)^{|S|}\, \mathbb{E}_{x\sim\mu_S} I_{S,x}^{\leqslant d}[f]^2 + O_k\left(\epsilon^2\delta^2\right)\|f\|_2^2.$$

Proof. By applying Theorem 4.7.7 with $f^{\leqslant d}$ rather then f and using $\|f^{\leqslant d}\|_\infty \leqslant \delta$ we obtain

$$\|\left(f^{\leqslant d}\right)^{\leqslant d}\|_4^4 \leqslant 20^d \sum_{|S|\leqslant d} (4d)^{|S|}\, \mathbb{E}_{x\sim\mu_S} I_{S,x}^{\leqslant d}\left[f^{\leqslant d}\right]^2 + O_k\left(\epsilon\delta^2\right)\|f\|_2^2.$$

The theorem now follows from Lemmas 4.6.11, 4.4.11 and 4.6.5. □

Theorem 4.7.7. *Let $\epsilon \leqslant \epsilon_0(k)$ be sufficiently small. Suppose that f is (d,δ)-global. Then we have*

$$\|f^{\leqslant d}\|_4^4 \leqslant (100d)^d\,\delta^2\|f^{\leqslant d}\|_2^2 + O_k\left(\delta^2\epsilon^2\|f\|_2^2\right).$$

Proof. By Theorem 4.7.6, Lemma 4.6.18, and 4.4.9 we have

$$\begin{aligned}
\|f^{\leqslant d}\|_4^4 &\leqslant 20^d \sum_{|S|\leqslant d} (4d)^d\, \mathbb{E}_{x\sim\mu_S}\left(I_{S,x}^{\leqslant d}[f]\right)^2 + O_k\left(\epsilon^2\|f\|_2^2\|f\|_\infty^2\right)\\
&\leqslant 20^d \sum_{|S|\leqslant d} (8d)^{d+2}\,\delta^2\mathbb{E}_x\left[I_{S,x}^{\leqslant d}[f]\right] + O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right)\\
&\leqslant 20^d \sum_{|S|\leqslant d} (8d)^{d+2}\,\delta^2 \sum_{T\supseteq S, |T|\leqslant d} \|f^{=T}\|_2^2 + O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right)\\
&\leqslant (40d)^d\,\delta^2 \sum_{|T|\leqslant d} \|f^{=T}\|_2^2 + O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right)\\
&\leqslant 2\,(40d)^d\,\delta^2\|f^{\leqslant d}\|_2^2 + O_k\left(\epsilon\|f\|_2^2\|f\|_\infty^2\right).
\end{aligned}$$

□

4.8 Applications

In this section, we show our applications of the hypercontractive inequality on high dimensional expanders, which we have shown in the previous section. The applications follow in a fairly straightforward way, and hence we present them with brevity.

4.8.1 Global Boolean functions are concentrated on the high degrees.

Fourier concentration results are widely useful in complexity theory and learning theory. Our first application is a Fourier concentration theorem for HDX. Namely, the following theorem shows that global Boolean functions on ϵ-HDX are concentrated on the high degrees, in the sense that the 2-norm of the restriction of a function to its low-degree coefficients only constitutes a tiny fraction of its total 2-norm.

Corollary 4.8.1. *If $f\colon V_{[k]}\to\{0,1\}$ is (d,δ)-global and ϵ is sufficiently small. Then*

$$\|f^{\leqslant d}\|_2^2 \leqslant \left(O_k\left(\sqrt{\epsilon}\right)+(200d)^d\,\delta^{\frac{1}{2}}\right)\|f\|_2^2.$$

Proof. By Lemma 4.4.9 we have

$$\|f^{\leqslant d}\|_2^2 = \left\langle f^{\leqslant d}, f\right\rangle - O_k\left(\epsilon\right)\|f\|_2^2.$$

We also have by Theorem 4.7.7

$$\begin{aligned}\left\langle f^{\leqslant d}, f\right\rangle &\leqslant \|f^{\leqslant d}\|_4\|f\|_{\frac{4}{3}}\\ &\leqslant (100d)^d\,\delta^{\frac{1}{2}}\sqrt{\|f^{\leqslant d}\|_2}\|f\|_{\frac{4}{3}} + O_k\left(\sqrt{\epsilon}\|f\|_2^{\frac{1}{2}}\|f\|_\infty^{\frac{1}{2}}\|f\|_{\frac{4}{3}}\right).\\ &\leqslant 2\,(100d)^d\,\delta^{\frac{1}{2}}\sqrt{\|f\|_2}\|f\|_{\frac{4}{3}} + O_k\left(\sqrt{\epsilon}\|f\|_2^2\right)\\ &\leqslant (200d)^d\,\delta^{\frac{1}{2}}\|f\|_2^2 + O_k\left(\sqrt{\epsilon}\|f\|_2^2\right)\end{aligned}$$

□

The Corollary completes the proof of Theorem 4.1.3.

4.8.2 Small-set expansion theorem

Small set expansion is a fundamental property that is prevalent in combinatorics and complexity theory. In the setting of the ρ-noisy Boolean hypercube, the small set expansion theorem gives an upper bound on $\mathrm{Stab}_\rho(1_A) = \langle 1_A, T_\rho 1_A\rangle = E[1_A(x)1_A(y)]$ for indicators 1_A of small sets A, which captures the probability that a random walk starting at a point $x\in A$ remains in A, hence showing that small sets are expanding. Our second application is a small set expansion theorem for global functions on ϵ-HDX, captured via bounding the natural noise operator in this setting.

Definition 4.8.2. Let $\rho\in(0,1)$. Given $x\in V_{[k]}$ we let $N_\rho(x)$ be the distribution where $y\sim N_\rho(x)$ is chosen by choosing a random set S where each i is in S independently with probability ρ, then choosing $z\sim\mu_{x_S}$ and setting $y=(x_S,z)$. We then set

$$\mathrm{T}_\rho f\left(x\right) = \mathbb{E}_{y\sim N_\rho(x)} f.$$

Alternatively we can use the averaging operators to give the following equivalent definition:

$$\mathrm{T}_\rho := \sum_{S\subseteq[k]} \rho^{|S|}\left(1-\rho\right)^{k-|S|} A_S\left[f\right].$$

We have the following formula for the noise operator, which is similar to the one in the product space setting.

Claim 4.8.3. We have $\mathrm{T}_\rho f = \sum_S \rho^{|S|} f^{=S}$.

Proof. We have

$$\begin{aligned}\mathrm{T}_\rho f &= \sum_{S\subseteq[k]} \rho^{|S|}\left(1-\rho\right)^{k-|S|} A_S\left[f\right]\\ &= \sum_{S\subseteq[k]} \rho^{|S|}\left(1-\rho\right)^{k-|S|} \sum_{T\subseteq S} f^{=T}\\ &= \sum_{T\subseteq[k]}\sum_{S\supseteq T} \rho^{|S|}\left(1-\rho\right)^{k-|S|}\\ &= \sum_{T\subseteq[k]} \rho^{|T|} f^{=T}.\end{aligned}$$

□

Via a standard argument we have the following bound on the noise operator.

Lemma 4.8.4. *We have*

$$\|\mathrm{T}_\rho f\|_2^2 \leqslant \|f^{\leqslant d}\|_2^2 + \left(\rho^d + O_k\left(\epsilon\right)\right)\|f\|_2^2.$$

Proof. This is immediate from Lemmas 4.8.3 and 4.4.9. □

Our small set expansion applications are as follows.

Corollary 4.8.5 (Small set expansion theorem)**.** *If* $f\colon V_{[k]}\to\{0,1\}$ *is* (d,δ)*-global. Then*

$$\|\mathrm{T}_\rho f\|_2^2 \leqslant \left(\rho^d + \left(100d\right)^d\delta^2 + O_k\left(\sqrt{\epsilon}\right)\right)\|f\|_2^2.$$

Proof. This follows immediately from Lemma 4.8.4 and Corollary 4.8.1. □

4.8.3 Kruskal–Katona theorem

Our last application is an analogue of the Kruskal–Katona theorem in the setting of high dimensional expanders. The Kruskal-Katona theorem is a fundamental and widely-applied result in algebraic combinatorics, which gives a lower bound on the size of the lower shadow of a set A, denoted $\partial(A) = \{x\colon y \prec x, \text{ for some } y\in A\}$.

We first consider the natural up-down walk in our setting.

Definition 4.8.6. The operator corresponding to up-down random walk is

$$\mathrm{T} = \frac{1}{k}\sum_{i=1}^{k} A_{[k]\setminus\{i\}}\left[f\right] = \sum_S \frac{k-|S|}{k} f^{=S}.$$

By applying the approximate Parseval inequality (Lemma 4.4.9), we obtain the following claim.

Claim 4.8.7. We have

$$\langle f - \mathrm{T}f, f\rangle \geqslant \frac{d}{k}\|f^{\geqslant d}\|_2^2 - O_k(\epsilon)\|f\|_2^2.$$

By our Fourier concentration theorem, (Corollary 4.8.1), we have the following lower bound on the 2-norm of the high degree part of f.

Claim 4.8.8. Let $\delta \leqslant (200d)^{-2d}$, and $\epsilon \leqslant \epsilon_0(k)$ be sufficiently small. If $f\colon V_{[k]} \to \{0,1\}$ is (d,δ)-global. Then

$$\|f^{\geqslant d}\|_2^2 \geqslant \frac{1}{2}\|f\|_2^2.$$

Combining the above claims we get the following.

Claim 4.8.9. Let $\delta \leqslant (200d)^{-2d}$. We have

$$\langle f - \mathrm{T}f, f\rangle \geqslant \frac{d}{2k}\|f\|_2^2$$

We are now ready to prove the Kruskal–Katona theorem in the setting of high dimensional expanders.

Corollary 4.8.10. *Let X be an ϵ-HDX, for a sufficiently small $\epsilon > 0$. Let $\delta \leqslant (200d)^{-d}$, and let $A \subseteq X(k-1)$ be (d,δ)-global. Then*

$$\mu(\partial(A)) \geqslant \mu(A)\left(1 + \frac{d}{2k}\right).$$

Proof. Let $f = 1_A$. We have

$$\begin{aligned}\langle f - \mathrm{T}f, f\rangle &= \mathop{\mathbf{Pr}}_{\sigma\sim X(k-1)}\mathop{\mathbf{Pr}}_{\tau_1,\tau_2\supset\sigma}[\tau_1 \in A, \tau_2 \notin A].\\ &\leqslant \mathbf{Pr}[\sigma \in \partial(A), f(\tau_2) \notin A]\\ &= \mu(\partial(A)) - \mu(A).\end{aligned}$$

□

Chapter 5

Conclusion and future directions

So far in the thesis we explore the geometry of HDXes by constructing HDXes from random complexes with latent geometry and showing isoperimetric inequalities for HDXes. In this chapter we look at a few open directions related to geometry and applications of HDXes.

5.1 Geometry and high-dimensional expansion

5.1.1 Other manifolds/complexes

In Chapter 3, the random geometric model over $\mathbb{S}^{d-1}$ fits in the broader framework of *random restrictions* of simplicial complexes: starting with a dense high-dimensional expander X, we sample a subset of vertices $\boldsymbol{S}$ of X to produce the sparser induced complex $X[\boldsymbol{S}]$.

We have shown that $X[\boldsymbol{S}]$ inherits the spectral properties of X itself, and we've leveraged this to show that for any polynomial average degree, one can produce a 2-dimensional expander by taking a random restriction of X the sphere in a particular dimension and with a particular connectivity distance. We hope that this framework might help us identify additional natural distributions over sparser and/or higher-dimensional complexes. More specifically,

> *Is there a simplicial complex X whose random restrictions yield high-dimensional expanders whose links have eigenvalue $< \frac{1}{2}$, of sub-polynomial or polylogarithmic degree?*

5.1.2 Other approaches to discretization

One interpretation of the construction in Chapter 3 is that the random geometric model $\mathsf{Geo}_d(n, p)$ gives a random discretization of high-dimensional spheres that preserves local and global expansions of the continuous spaces.

However this random discretization approach fails to give bounded-degree local-spectral expanders. Even to ensure that the resulting random discrete graphs are connected, we need to set the average degree of this model to polylogarithmic in the number of vertices. The average degree needs to be even higher to further guarantee global and local expansions.

Another drawback of this approach is the hardness in generalizing the spectral analysis to dimensions greater than 2. In order to bound the local expansion of 2-dimensional complexes, we have to carefully use concentration of random points in a geodesic ball. Extending the analysis to higher-dimensional links would require showing similar concentrations for more complicated shapes. So we ask,

> *Are there randomized or deterministic constructions of local-spectral expanders that avoid the two problems above?*

5.2 Coboundary and cosystolic expanders

Just as local-spectral expanders are the generalization of spectral expansion to simplicial complexes, coboundary expanders are the generalization of combinatorial expansion to simplicial complexes. Although there are many known examples of coboundary expanders such as the complete complexes, spherical buildings, and random clique complexes, all these d-dimensional coboundary expanders are polynomially dense when $d > 2$.

> *Can we construct bounded-degree d-dimensional coboundary expanders for any constant d?*

Many known coboundary expanders are face-transitive. Current technique for showing coboundary expansion hinges on both this symmetry and the fact that these complexes have poly(n) average degree. So to prove the property over asymmetric constructions and/or sparser complexes, one should consider new approaches for such problems. One natural question is

> *Do any of the known constructions of local-spectral expanders have non-trivial coboundary expansion? For instance, does the random geometric model yield 2-dimensional coboundary expanders?*

Cosystolic expansion is a relaxation of coboundary expansion that turns out to be useful for constructing codes and explicit hard instances for Sum-of-Square algorithms [28, 99, 75, 27, 25, 59]. Many of the known algebraic constructions of sparse local-spectral expanders are also known to be cosystolic expanders [90, 66].

One known connection between different types of HDXes is that a d-dimensional simplicial complex is a cosystolic expander if it is a local-spectral expander and every vertex's link is a $(d-1)$-dimensional coboundary expander [37]. Some applications in testing require a complex to have both local-spectral expansion and cosystolic expansion. One open question is

> *Is there a simpler characterization for complexes with both expansion properties?*

5.3 Beyond constant-dimensional simplicial complexes

5.3.1 Superconstant dimension

We have mentioned many constructions of HDXes. So far when computing sparsity and expansion, we have treated the dimension parameter d as a constant. Thus by "bounded-degree complex", we mean that the number of d-dimensional faces in the complex is $O_d(n)$. However, when considering using HDXes to construct tests for relations over superconstant size alphabet, one needs to take into consideration the dependency on d. The question here is

> *For* $d = \mathrm{polylog}(n)$*, are there* d*-dimensional local-spectral or coboundary expanders with* $n^{O(k)}$ k*-dimensional faces for every* $k \in \{0, \ldots, d\}$*?*

5.3.2 Chain complexes

Simplicial complexes have been the focus of this thesis. However, high-dimensional expansion can be easily defined for more general chain complexes. One such instance is the Grassmann complexes. Recall that in a complete simplicial complex χ, the face set $\chi(i)$ is all size-$(i+1)$ subsets of $\{1, \ldots, n\}$. In a Grassmann complex X, the face set $X(i)$ is all $(i+1)$-dimensional subspaces of a vector space $\mathbb{F}_p^d$. We call any subcomplex of a Grassmann complex a Grassmann-type complex. Notions defined on simplicial complexes such as link and local-spectral expansion generalize to Grassmann-type complexes.

Grassmann-type complexes have a natural correspondence with polynomial codes. Indeed sparse Grassmann-type complexes with local-spectral expansions are candidates of locally testable polynomial codes with good rate and distance. In this direction the major open question is

> *Can we find locally testable polynomial codes with constant rate, relative distance, and arity via constructing locally expanding Grassmann-type complexes?*

Bibliography

[1] Vedat Levi Alev, Fernando Granha Jeronimo, and Madhur Tulsiani. "Approximating constraint satisfaction problems on high-dimensional expanders". In: *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2019, pp. 180–201.

[2] Vedat Levi Alev and Lap Chi Lau. "Improved analysis of higher order random walks and applications". In: *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*. 2020, pp. 1198–1211.

[3] Noga Alon et al. "Random sampling and approximation of MAX-CSPs". In: *Journal of computer and system sciences* 67.2 (2003), pp. 212–243.

[4] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. "Spectral independence in high-dimensional expanders and applications to the hardcore model". In: *SIAM Journal on Computing* 0 (2021), FOCS20–1.

[5] Nima Anari et al. "Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. 2019, pp. 1–12.

[6] Eric Babson, Christopher Hoffman, and Matthew Kahle. "The fundamental group of random 2-complexes". In: *Journal of the American Mathematical Society* 24.1 (2011), pp. 1–28.

[7] Mitali Bafna et al. "Hypercontractivity on High Dimensional Expanders". In: *(personal communication)* (2021).

[8] Dominique Bakry, Ivan Gentil, Michel Ledoux, et al. *Analysis and geometry of Markov diffusion operators*. Vol. 103. Springer, 2014.

[9] Boaz Barak et al. "Subsampling mathematical relaxations and average-case complexity". In: *Proceedings of the twenty-second annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2011, pp. 512–531.

[10] Avraham Ben-Aroya and Amnon Ta-Shma. "A combinatorial construction of almost-Ramanujan graphs using the zig-zag product". In: *SIAM Journal on Computing* 40.2 (2011), pp. 267–290.

[11] Charles Bordenave. "On Euclidean random matrices in high dimension". In: *Electronic Communications in Probability* 18 (2013), pp. 1–8.

[12] Mark Braverman et al. "An Invariance Principle for the Multi-slice, with Applications". In: *arXiv preprint arXiv:2110.10725* (2021).

[13] Sébastien Bubeck et al. "Testing for high-dimensional geometry in random graphs". In: *Random Structures & Algorithms* 49.3 (2016), pp. 503–532.

[14] Donald I Cartwright, Patrick Solé, and Andrzej Żuk. "Ramanujan geometries of type $\widetilde{A}_n$". In: *Discrete mathematics* 269.1-3 (2003), pp. 35–43.

[15] Michael Chapman, Nati Linial, and Yuval Peled. "Expander graphs – both local and global". In: *Combinatorica* 40.4 (2020), pp. 473–509.

[16] J CHEEGER. "A lower bound for the smallest eigenvalue of the Laplacian". In: *Problems in Analysis* (1970), pp. 195–199.

[17] Xiuyuan Cheng and Amit Singer. "The spectrum of random inner-product kernel matrices". In: *Random Matrices: Theory and Applications* 2.04 (2013), p. 1350010.

[18] David Conlon. "Hypergraph expanders from Cayley graphs". In: *Israel Journal of Mathematics* 233.1 (2019), pp. 49–65.

[19] David Conlon, Jonathan Tidor, and Yufei Zhao. "Hypergraph expanders of all uniformities from Cayley graphs". In: *Proceedings of the London Mathematical Society* 121.5 (2020), pp. 1311–1336.

[20] Neta Dafni et al. "Complexity measures on symmetric group and beyond". In: *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. 2021.

[21] Amir Dembo. "Moderate deviations for martingales with bounded jumps". In: (1996).

[22] Luc Devroye et al. "High-dimensional random geometric graphs and their clique number". In: *Electronic Journal of Probability* 16 (2011), pp. 2481–2508.

[23] Yotam Dikstein et al. "Boolean Function Analysis on High-Dimensional Expanders". In: *LIPIcs-Leibniz International Proceedings in Informatics*. Vol. 116. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2018.

[24] Irit Dinur and Tali Kaufman. "High dimensional expanders imply agreement expanders". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2017, pp. 974–985.

[25] Irit Dinur et al. "Explicit SoS lower bounds from high-dimensional expanders". In: *arXiv preprint arXiv:2009.05218* (2020).

[26] Irit Dinur et al. "Good Quantum LDPC Codes with Linear Time Decoders". In: *arXiv preprint arXiv:2206.07750* (2022).

[27] Irit Dinur et al. "Good quantum LDPC codes with linear time decoders". In: *arXiv preprint arXiv:2206.07750* (2022).

[28] Irit Dinur et al. "Locally testable codes with constant rate, distance, and locality". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 2022, pp. 357–374.

[29] Irit Dinur et al. “On Non-Optimally Expanding Sets in Grassmann Graphs”. In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*. 2018, pp. 940–951.

[30] Irit Dinur et al. “Towards a Proof of the 2-to-1 Games Conjecture?” In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*. 2018, pp. 376–389.

[31] Yen Do and Van Vu. “The spectrum of random kernel matrices: universality results for rough and varying kernels”. In: *Random Matrices: Theory and Applications* 2.03 (2013), p. 1350005.

[32] Jozef Dodziuk. “Difference equations, isoperimetric inequality and transience of certain random walks”. In: *Transactions of the American Mathematical Society* 284.2 (1984), pp. 787–794.

[33] Jean Dolbeault et al. “Sharp interpolation inequalities on the sphere: new methods and consequences”. In: *Partial Differential Equations: Theory, Control and Approximation*. Springer, 2014, pp. 225–242.

[34] Noureddine El Karoui. “The spectrum of kernel random matrices”. In: *The Annals of Statistics* 38.1 (2010), pp. 1–50.

[35] David Ellis, Nathan Keller, and Noam Lifshitz. “On a biased edge isoperimetric inequality for the discrete cube”. In: *Journal of Combinatorial Theory, Series A* 163 (2019), pp. 118–162.

[36] David Ellis, Guy Kindler, and Noam Lifshitz. “Hypercontractivity for global functions in the bilinear scheme”. In: *In preparation* (2021).

[37] Shai Evra and Tali Kaufman. “Bounded degree cosystolic expanders of every dimension”. In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. 2016, pp. 36–48.

[38] Zhou Fan and Andrea Montanari. “The spectral norm of random inner-product kernel matrices”. In: *Probability Theory and Related Fields* 173.1 (2019), pp. 27–85.

[39] Weiming Feng et al. “Rapid mixing from spectral independence beyond the Boolean domain”. In: *ACM Transactions on Algorithms (TALG)* 18.3 (2022), pp. 1–32.

[40] Yuval Filmus. *Almost linear Boolean functions on S n are almost unions of cosets*. Manuscript. 2021.

[41] Yuval Filmus. *FKN theorems for the biased cube and the slice*. Manuscript. 2021.

[42] Yuval Filmus. “Friedgut-Kalai-Naor Theorem for Slices of the Boolean Cube”. In: *Chicago J. Theor. Comput. Sci.* 2016 (2016).

[43] Yuval Filmus and Elchanan Mossel. “Harmonicity and Invariance on Slices of the Boolean Cube”. In: *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan*. 2016, 16:1–16:13.

[44] Yuval Filmus, Ryan O'Donnell, and Xinyu Wu. "A log-Sobolev inequality for the multislice, with applications". In: *Proceedings of the 10th Innovations in Theoretical Computer Science conference (ITCS'19)*. 2019.

[45] Yuval Filmus et al. "Hypercontractivity on the symmetric group". In: *arXiv preprint arXiv:2009.05503* (2020).

[46] Yuval Filmus et al. "Invariance Principle on the Slice". In: *TOCT* 10.3 (2018), 11:1–11:37.

[47] Jacob Fox et al. "Overlap properties of geometric expanders". In: *Journal für die reine und angewandte Mathematik (Crelles Journal)* 2012.671 (2012), pp. 49–83.

[48] Ehud Friedgut and Yonatan Iluz. "Hyper-regular graphs and high dimensional expanders". In: *arXiv preprint arXiv:2010.03829* (2020).

[49] Joel Friedman. "A proof of Alon's second eigenvalue conjecture". In: *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. ACM. 2003, pp. 720–724.

[50] Joel Friedman. "On the second eigenvalue and random walks in random d-regular graphs". In: *Combinatorica* 11.4 (1991), pp. 331–362.

[51] Ofer Gabber and Zvi Galil. "Explicit constructions of linear-sized superconcentrators". In: *Journal of Computer and System Sciences* 22.3 (1981), pp. 407–420.

[52] Oded Goldreich, Shafi Goldwasser, and Dana Ron. "Property testing and its connection to learning and approximation". In: *Journal of the ACM (JACM)* 45.4 (1998), pp. 653–750.

[53] Louis Golowich. "Improved Product-Based High-Dimensional Expanders". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2021.

[54] Mikhail Gromov. "Singularities, expanders and topology of maps. Part 2: From combinatorics to topology via algebraic isoperimetry". In: *Geometric and Functional Analysis* 20.2 (2010), pp. 416–526.

[55] Leonard Gross. "Logarithmic sobolev inequalities". In: *American Journal of Mathematics* 97.4 (1975), pp. 1061–1083.

[56] Tom Gur, Noam Lifshitz, and Siqi Liu. "Hypercontractivity on high dimensional expanders". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 2022, pp. 176–184.

[57] Shlomo Hoory, Nathan Linial, and Avi Wigderson. "Expander graphs and their applications". In: *Bulletin of the American Mathematical Society* 43.4 (2006), pp. 439–561.

[58] Shlomo Hoory, Nathan Linial, and Avi Wigderson. "Expander graphs and their applications". In: *Bulletin of the American Mathematical Society* 43.4 (2006), pp. 439–561.

[59] Max Hopkins and Ting-Chun Lin. "Explicit Lower Bounds Against $\Omega(n)$-Rounds of Sum-of-Squares". In: *arXiv preprint arXiv:2204.11469* (2022).

[60] Samuel B Hopkins et al. "The power of sum-of-squares for detecting hidden structures". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2017, pp. 720–731.

[61] Fernando Granha Jeronimo, Shashank Srivastava, and Madhur Tulsiani. "Near-linear time decoding of Ta-Shma's codes via splittable regularity". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 1527–1536.

[62] Mark Jerrum et al. "Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains". In: *The Annals of Applied Probability* 14.4 (2004), pp. 1741–1765.

[63] Jeff Kahn, Gil Kalai, and Nathan Linial. *The Influence of Variables on Boolean Functions*. IEEE, 1988.

[64] Tali Kaufman, David Kazhdan, and Alexander Lubotzky. "Ramanujan complexes and bounded degree topological expanders". In: *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*. IEEE. 2014, pp. 484–493.

[65] Tali Kaufman and David Mass. "High dimensional random walks and colorful expansion". In: *arXiv preprint arXiv:1604.02947* (2016).

[66] Tali Kaufman and Izhar Oppenheim. "Construction of new local spectral high dimensional expanders". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 2018, pp. 773–786.

[67] Tali Kaufman and Izhar Oppenheim. "Construction of new local spectral high dimensional expanders". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*. 2018, pp. 773–786.

[68] Tali Kaufman and Izhar Oppenheim. "High order random walks: Beyond spectral gap". In: *Combinatorica* (2020), pp. 1–37.

[69] Peter Keevash et al. "Forbidden intersections for codes". In: *arXiv preprint arXiv:2103.05050* (2021).

[70] Peter Keevash et al. "Hypercontractivity for global functions and sharp thresholds". In: *arXiv preprint arXiv:1906.05568* (2019).

[71] Subhash Khot, Dor Minzer, and Muli Safra. “On Independent Sets, 2-to-2 Games, and Grassmann Graphs”. In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*. 2017, pp. 576–589.

[72] Subhash Khot, Dor Minzer, and Muli Safra. “Pseudorandom Sets in Grassmann Graph Have Near-Perfect Expansion”. In: *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*. 2018, pp. 592–601.

[73] Vladimir Koltchinskii and Evarist Giné. “Random matrix approximation of spectra of integral operators”. In: *Bernoulli* (2000), pp. 113–167.

[74] James R Lee, Prasad Raghavendra, and David Steurer. “Lower bounds on the size of semidefinite programming relaxations”. In: *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*. 2015, pp. 567–576.

[75] Anthony Leverrier and Gilles Zémor. “Quantum tanner codes”. In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2022, pp. 872–883.

[76] David A Levin and Yuval Peres. *Markov chains and mixing times*. Vol. 107. American Mathematical Soc., 2017.

[77] Wen-Ching Winnie Li. “Ramanujan hypergraphs”. In: *Geometric & Functional Analysis GAFA* 14.2 (2004), pp. 380–399.

[78] Noam Lifshitz and Dor Minzer. “Noise sensitivity on the p-biased hypercube”. In: *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2019, pp. 1205–1226.

[79] Ting-Chun Lin and Min-Hsiu Hsieh. “c 3-locally testable codes from lossless expanders”. In: *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2022, pp. 1175–1180.

[80] Nathan Linial and Roy Meshulam. “Homological connectivity of random 2-complexes”. In: *Combinatorica* 26.4 (2006), pp. 475–487.

[81] Nathan Linial and Yuval Peled. “On the phase transition in random simplicial complexes”. In: *Annals of mathematics* (2016), pp. 745–773.

[82] Nati Linial. *Some Geometric Perspectives on Combinatorics: High-Dimensional, Local and Local-to-Global II*. Simons Institute big data bootcamp workshop, https://www.youtube.com/watch?v=Kt9SWmIPsH8. 2013.

[83] Siqi Liu, Sidhanth Mohanty, and Elizabeth Yang. “High-Dimensional Expanders from Expanders”. In: *11th Innovations in Theoretical Computer Science Conference*. 2020.

[84] Siqi Liu et al. “Local and global expansion in random geometric graphs”. In: *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. 2023, pp. 817–825.

[85] Yue M Lu and Horng-Tzer Yau. "An Equivalence Principle for the Spectrum of Random Inner-Product Kernel Matrices". In: *arXiv preprint arXiv:2205.06308* (2022).

[86] Alex Lubotzky. *High-Dimensional Expanders and Property Testing*. STOC workshop on advances in coding theory, https://youtu.be/P5vs4ARRQjU?t=2967. 2022.

[87] Alexander Lubotzky. "High dimensional expanders". In: *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*. World Scientific. 2018, pp. 705–730.

[88] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. "Ramanujan graphs". In: *Combinatorica* 8.3 (1988), pp. 261–277.

[89] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. "Ramanujan graphs". In: *Combinatorica* 8.3 (1988), pp. 261–277.

[90] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. "Explicit constructions of Ramanujan complexes of type Ad". In: *European Journal of Combinatorics* 26.6 (2005), pp. 965–993.

[91] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. "Ramanujan complexes of typeÃ d". In: *Israel journal of Mathematics* 149.1 (2005), pp. 267–299.

[92] Grigorii Aleksandrovich Margulis. "Explicit constructions of concentrators". In: *Problemy Peredachi Informatsii* 9.4 (1973), pp. 71–80.

[93] Roy Meshulam and Nathan Wallach. "Homological connectivity of random k-dimensional complexes". In: *Random Structures & Algorithms* 34.3 (2009), pp. 408–417.

[94] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[95] Ryan O'Donnell and Kevin Pratt. "High-Dimensional Expanders from Chevalley Groups". In: *arXiv preprint arXiv:2203.03705* (2022).

[96] Ryan O'Donnell and Karl Wimmer. "KKL, Kruskal–Katona, and Monotone Nets". In: *SIAM Journal on Computing* 42.6 (2013), pp. 2375–2399.

[97] Izhar Oppenheim. "Local spectral expansion approach to high dimensional expanders part I: Descent of spectral gaps". In: *Discrete & Computational Geometry* 59.2 (2018), pp. 293–330.

[98] Pavel Panteleev and Gleb Kalachev. "Asymptotically good quantum and locally testable classical LDPC codes". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 2022, pp. 375–388.

[99] Pavel Panteleev and Gleb Kalachev. "Quantum LDPC codes with almost linear minimum distance". In: *IEEE Transactions on Information Theory* 68.1 (2021), pp. 213–229.

[100] Mathew Penrose. *Random geometric graphs*. Vol. 5. OUP Oxford, 2003.

[101] Omer Reingold, Salil Vadhan, and Avi Wigderson. "Entropy waves, the zig-zag graph product, and new constant-degree expanders". In: *Annals of mathematics* (2002), pp. 157–187.