Chernoff Bound for High-Dimensional Expanders

Tali Kaufman
Department of Computer Science, Bar-Ilan University, Ramat Gan, Israel
kaufmant@mit.edu

Ella Sharakanski
Department of Computer Science, Bar-Ilan University, Ramat Gan, Israel
Ella10004@gmail.com

Abstract

We generalize the expander Chernoff bound to high-dimensional expanders. The expander Chernoff bound is an essential property of expanders, first proved by Gillman [9]. Given a graph G and a function f on the vertices, it states that the probability of f's mean sampled via a random walk on G to deviate from its actual mean, has a bound that depends on the spectral gap of the walk and decreases exponentially as the walk's length increases.

We are interested in obtaining an analog Chernoff bound for high order walks on high-dimensional expanders. A naive generalization of the expander Chernoff bound from expander graphs to high-dimensional expanders gives a very poor bound due to obstructions that occur in high-dimensional expanders and are not present in (one-dimensional) expander graphs. Because of these obstructions, the spectral gap of high-order random walks is inherently small.

A natural question that arises is how to get a meaningful Chernoff bound for high-dimensional expanders. In this paper, we manage to get a strong Chernoff bound for high-dimensional expanders by looking beyond the spectral gap.

First, we prove an expander Chernoff bound that depends on a notion that we call the "shrinkage of a function" instead of the spectral gap. In one-dimensional expanders, the shrinkage of any function with zero-mean is bounded by $\lambda(M)$. Therefore, the spectral gap is just the one-dimensional manifestation of the shrinkage.

Next, we show that in good high-dimensional expanders, the shrinkage of functions that "do not come from below" is good. A function does not come from below if from any local point of view (called "link") its mean is zero.

Finally, we prove a high-dimensional Chernoff bound that captures the expansion of the complex. When the function on the faces has a small variance and does not "come from below", our bound is better than the naive high-dimensional expander Chernoff bound.

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains; Mathematics of computing → Spectra of graphs; Mathematics of computing → Hypergraphs

Keywords and phrases High Dimensional Expanders, Random Walks, Tail Bounds

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2020.25

Category RANDOM

1 Introduction

Expander graphs have been studied extensively for several decades. Essentially, these are graphs that are sparse yet highly connected. They have countless applications in computer science, mathematics and physics (see e.g. the surveys of [13] and [19]).

Since expanders are such a marvelous tool, it is believed that their generalization to higher dimensions is promising. Expanders consist of edges, which are sets of two vertices. In contrast, high-dimensional (abbreviated as HD) expanders consist of faces, which are sets of any amount of vertices, like edges, triangles, pyramids, and more. Since HD expanders expand at *all dimensions*, they are much stronger objects than expander graphs, which only expand at one dimension.

Recently, HD expanders have attracted a lot of attention. A notable example is the paper of [1] which solves a major open problem on counting the bases of matroids. The algorithms and proofs in their paper build on recent results about *random walks on HD expanders* (e.g. [16] and [15]).

In this paper, we study further aspects of random walks on HD expanders. In particular, we generalize the expander Chernoff bound to HD expanders. Generally speaking, this bound states that a random walk on an expander acts "as expected".

Let us present the (one-dimensional) expander Chernoff bound in a more precise way. There are many forms for this bound (see e.g. [10, 14, 18, 21, 12, 20, 5]), the form we present appears in [10] and is considered standard.

Let G be an undirected connected graph on vertices $V \coloneqq \{1,\dots,n\}$ with positive weights on its edges. The graph is equivalent to a connected random walk with transition matrix M. Let π be the stationary distribution of M and let $\{w_i\}_{i=1}^{t}$ be a random walk according to M with initial distribution π. Let $f : V \to [0,1]$ be a function on the vertices.

The general form of the graph generalization of Chernoff bound is as follows: for all $0 \le \epsilon \le 1$,

$$\mathbb{P}\Big[\big|\frac{1}{t}\sum_{i=1}^{t} f(w_i) - \mathbb{E}_{\pi}[f]\big| \ge \epsilon\Big] \le 4 \cdot \exp\big(-\Omega(\epsilon^2 st)\big). \tag{1}$$

Note that for a better bound, we want s to be bigger.

The expander Chernoff bound states that inequality (1) holds for $s = 1 - \lambda_2(M)$, where $\lambda_2(M)$ is the second largest eigenvalue of M. This bound works best for graphs with λ_2 as close to 0 as possible. By Definition 7, these are exactly (one-sided) expander graphs.

As thoroughly explained in [21], this bound is a very useful tool in theoretical computer science. It has applications in a variety of areas like construction of efficient error-correcting codes, hardness of approximation, deterministic amplification, security amplification in cryptography, extractor construction, and more.

Since high-dimensional expanders are very promising, and the Chernoff bound for expander graphs is fundamental and extremely useful, we see great potential in further generalizing it to higher dimensions.

Such a generalization should provide a Chernoff bound for random walks on high-dimensional faces, that is similar to the bound we have for walks on vertices. Since high-dimensional expanders are better than one-dimensional, one could expect to obtain a generalization simply by applying the existing expander Chernoff bound to high-order walks.

However, this method results in a surprisingly poor bound, due to natural obstructions that occur even in the best high-dimensional expanders, and do not happen for one-dimensional expanders. Because of these obstructions, the spectral gap $1 - \lambda(M)$ of high-order random walks is inherently small.

In this paper, we manage to overcome this problem by looking beyond the spectral gap. First, we prove an expander Chernoff bound that depends on the "shrinkage of the function" f (see Definition 3) instead of the spectral gap. In one-dimensional expanders, the shrinkage of any function with zero-mean is bounded by $\lambda(M)$. Therefore, the spectral gap is just the one-dimensional manifestation of the shrinkage.

Next, we show that in good high-dimensional expanders, the shrinkage of functions that "do not come from below" (see Definition 1) is good. A function does not come from below if its mean is zero from any local point of view (called "link", see Definition 11). This criterion naturally generalizes zero-mean functions since there is only one local point of view on vertices and it includes all of them.

Thus, we obtain a high-dimensional expander Chernoff bound that works well for functions that have a small variance and do not come from below.

1.1 Background

Let us provide a simplified version of the definitions we need in order to present our results. The detailed definitions are deferred to Section 2.

Given a weighted undirected connected graph G, its normalized second largest eigenvalue in absolute value is denoted $\lambda(G)$. We say that G is a (two-sided) λ-*spectral expander* if $\lambda(G) < \lambda$.

We study a generalization of the above definition to higher dimensions, based on *simplicial complexes*. A pure d-dimensional simplicial complex X consists of sets of size $d+1$ and all of their subsets. These sets are called faces, and $X(i)$ denotes the set of faces of size $i+1$. Our complexes are equipped with a probability distribution π_d over the d-faces, which induces probability distributions π_k over k-faces. In addition, each π_k induces an inner-product on the subspace of k-cochains (functions from $X(k)$ to $\mathbb{R}$) denoted $C^k(X)$.

There are several different ways to define high-dimensional expanders. We use a notion called (two-sided) local-spectral expander, defined in [7], since it is the generalization of spectral-expanders to higher dimensions. Given a face $\tau \in X(i)$, the *link* of τ, denoted X_τ, is its local point of view. Formally, it is a simplicial complex obtained by taking the faces of X that contain τ and removing τ from them. A pure complex is a γ-*local-spectral expander* if for any link, the underlying graph is a γ-spectral expander (see Definition 13).

The high-order random walk we study is called the lazy upper random walk on k-faces, defined in [15] and denoted M_k^+. This walk naturally generalizes the standard lazy walk on graphs. It can move between two k-faces if they are both contained in the same $(k+1)$-face.

Given a function $f : X(k) \to \mathbb{R}$ and a face $\tau \in X(j)$, the localization of f on X_τ is the function $f_\tau : X_\tau(k-j-1) \to \mathbb{R}$ defined by $f_\tau(\sigma) = f(\tau \cup \sigma)$.

▶ **Definition 1.** *We say that a function $f : X(k) \to \mathbb{R}$ is* from level $j+1$ *if its expectation on the links of j-faces is zero. Formally, f is from level $j+1$ if $\mathbb{E}_{\pi_\tau}[f_\tau] = 0$ for all $\tau \in X(j)$.*

Informally, we say that f *does not come from below* if it is from a high level (e.g. from level k). Our high-dimensional Chernoff bound works best for functions f that do not come from below and have a small variance.

1.2 Our Contribution: Approach and Considerations

A bound for high-order random walks. Observe that the graph Chernoff bound is actually a bound for *random walks* on graphs. Hence, what we are looking for is a bound for *random walks* on high-dimensional expanders. Specifically, we study the lazy upper random walk, since it is highly related to the structure of the HD expander (see Definition 17).

Capturing the expansion. Given a HD γ-local-spectral expander X (see Definition 13), we want a bound that captures the expansion of X. We can apply the graph Chernoff bound (1) to the random walk M_k^+ and get a bound where $s = 1 - \lambda(M_k^+)$ (see Definitions 7). Unfortunately, this value of s is bounded away from 1 (in fact, it behaves like $1/k$, so it approaches 0 as k increases) and hence does not capture the expansion of X well.

According to [17], $\lambda(M_k^+)$ is large due to natural obstructions that occur in HD expanders and not in one-dimensional expander graphs. To bypass the obstructions, our Chernoff bound takes into account the entire spectrum of M_k^+.

Generalizing a one-dimensional Chernoff bound. We prove our high-dimensional Chernoff bound by generalizing a one-dimensional Chernoff bound. Given an expander graph G and a function f on the vertices (as in (1)), our key idea is to only look at eigenvalues λ_i that correspond to eigenvectors that are part of the orthogonal decomposition of f into eigenspaces. We prove a Chernoff bound that depends on the shrinkage of f, denoted $\lambda(f) = \max_i |\lambda_i|$ (see Definition 3), instead of $\lambda(G)$. When f is "good" (not an obstruction), $\lambda(f)$ is a small eigenvalue of G. By applying our one-dimensional Chernoff bound to M_k^+, we obtain a high-dimensional Chernoff bound that can see beyond $\lambda(M_k^+)$ and therefore capture the expansion of X well.

A bound for functions that do not come from below. The standard one-dimensional expander Chernoff bound depends on $\lambda(G)$, which is the largest eigenvalue in absolute value except for 1. Namely, the bound only considers eigenspaces orthogonal to span$\{\mathbf{1}\}$. Equivalently, it considers eigenspaces of zero-mean functions. These are exactly functions that their expectation on the links of (-1)-faces is zero (since the only (-1)-face is the empty face and its link is the entire set of vertices), which we call functions from level 0. Therefore, the existing bound is in fact an expander Chernoff bound for functions from level 0. Similarly, our HD expander Chernoff bound applies to functions from level j.

Moreover, as we will prove in Corollary 37, given a good enough γ-local-spectral expander, a function is from level $j+1$ if and only if its orthogonal decomposition into eigenspaces of M_k^+ only includes eigenvectors matching eigenvalues smaller than $\frac{k-j}{k+2} + o_\gamma(1)$. Namely, a function does not come from below if and only if its shrinkage is small. Hence, considering functions that do not come from below is highly natural.

Irregularity matters. Some expander Chernoff bounds apply only to regular graphs (see e.g. [12, 8]). In the case of HD expanders, even when the weights of the largest faces are uniform, the weights of other k-dimensional faces are usually not uniform, and hence M_k^+ is usually irregular. Therefore, it is important that our bound applies to irregular graphs too.

1.3 Results and Technique

Our main result is a Chernoff bound for HD expanders. When the function on the k-faces has a small variance and does not "come from below", our bound captures the expansion of the complex well. Namely, we present a Chernoff bound for M_k^+ in the form of (1), where s approaches 1 as k increases, as opposed to the naive bound where s approaches 0.

▶ **Theorem 2** (Main theorem, HD expander Chernoff bound, informal, for formal see Theorem 48)**.** *Let X be a d-dimensional γ-local-spectral expander, let $0 \le k < d$, and assume that γ is small enough. Let $\pi = \pi_k$ be the probability distribution over $X(k)$ and denote $\pi_* := \min \pi_i$. Let $w_1, \dots, w_t$ be a random walk according to M_k^+ with initial distribution π. Let $f : X(k) \to \mathbb{R}$ such that f is from level $j+1$ for some $-1 \le j < k$, and assume $\|f\|_\infty = 1$. Then the following hold:*

1. *If $\mathrm{Var}_\pi(f) \le \pi_* \epsilon^2 c$ for a constant $c \in \mathbb{R}$ then inequality (1) holds for $s = \left(1 - \dfrac{k-j}{k+2} + o_\gamma(1)\right)^2$. In particular, for $j = k-1$, inequality (1) holds for $s = \left(1 - \dfrac{1}{k+2} + o_\gamma(1)\right)^2$.*
2. *If $\mathrm{Var}_\pi(f) \le \pi_* \epsilon^2 c$ such that $4c \le \left(\dfrac{k+2}{k-j+o_\gamma(1)}\right)^2$, then inequality (1) holds for $s = 1$.*

   In particular, for $j = k-1$, if $c \le \dfrac{(k+2)^2}{4 + o_\gamma(1)}$ then inequality (1) holds for $s = 1$.

Note that $o_\gamma(1)$ refers to a value that approaches 0 as γ goes to 0.

Extensions of the theorem. In the detailed version of this theorem, stated in Theorem 48, we also provide a bound for larger $\mathrm{Var}_\pi(f)$. Additionally, in Remark 49 we provide a bound for cases where f does come from below, but its mass on the lower eigenspaces is small.[1]

Comparison with the naive bound. One could try to achieve a similar bound by applying to M_k^+ a standard expander Chernoff bound like [10]. However, this standard bound depends on the spectral gap, which is low even in the best HD expanders. This method would give $s \approx 1/(k+2)$, which approaches 0 as k increases. We successfully overcome this problem by considering the shrinkage of f in functions that do not come from below, instead of the spectral gap. Indeed, we achieve a value of s that is either 1 or it approaches 1 as k grows.

Proof overview. To prove this theorem, we first define the *shrinkage* of a vector by an operator (Definition 3). The shrinkage is key to overcoming the obstructions which make $\lambda(M_k^+)$ large. We then prove a (one-dimensional) expander Chernoff bound that depends on the shrinkage of the function on the vertices, instead of on the graph's expansion (Theorem 4). Next, we apply to M_k^+ our expander Chernoff bound for shrinking functions. Finally, we bound the shrinkage of k-cochains (Proposition 5) to prove the HD expander Chernoff bound.

1.3.1 Defining the Shrinkage of a Vector

Let W be a finite-dimensional vector-space with inner-product $\langle \cdot, \cdot \rangle$. Let $M : W \to W$ be a self-adjoint operator. By the spectral theorem, $W = \bigoplus_i E_{\lambda_i}$ where E_{λ_i} is the eigenspace of M associated with the eigenvalue $\lambda_i \in \mathbb{R}$. Moreover, there exists an orthonormal basis of W that consists of eigenvectors of M.

▶ **Definition 3.** *Let $w \in W$. The* shrinkage *of w is $\lambda_M(w) := \max_i\{|\lambda_i| \mid w \not\perp E_{\lambda_i}\}$. Equivalently, $\lambda_M(w)$ is the largest eigenvalue of M (in absolute value) that matches an eigenvector that spans w. Formally, let B be any orthonormal basis of W consisting of eigenvectors ϕ_i corresponding to eigenvalues λ_i respectively. Then $\lambda_M(w) := \max_i\{|\lambda_i| \mid w \not\perp \phi_i\}$. For ease of notation, when M is clear from context, we simply write $\lambda(w)$.*

1.3.2 Expander Chernoff Bound for Shrinking Functions

We prove an expander Chernoff bound that depends on the shrinkage of the function on the vertices, $\lambda(f)$, instead of on the graph's expansion, $\lambda(G)$. As far as we are aware, this is the first bound of this kind. The bound is a generalization of a similar bound from [8].

▶ **Theorem 4** (Expander Chernoff bound that depends on $\lambda(f)$ instead of $\lambda(G)$, informal, for formal see Theorem 29)**.** *In the setting of inequality (1), assume the function $f : V \to \mathbb{R}$ satisfies $\mathbb{E}_\pi[f] = 0$ and $\|f\|_\infty = 1$, and denote $\pi_* := \min \pi_i$. Then the following hold:*

1. *If $\mathrm{Var}_\pi(f) \le \pi_* \epsilon^2 c$ for some constant $c \in \mathbb{R}$ then inequality (1) holds for $s = (1 - \lambda(f))^2$.*
2. *Moreover, if $\mathrm{Var}_\pi(f) \le \pi_* \epsilon^2 c$ such that $c \le \dfrac{1}{4\lambda(f)^2}$ then inequality (1) holds for $s = 1$.*

In the detailed version of this theorem, stated in Theorem 29, we also provide a bound for when $\mathrm{Var}_\pi(f)$ is larger.

[1] Namely, denote $P_{k,j} := P_{(\mathrm{Im}\, d_{j\nearrow k})^\perp}$ the orthogonal projection on the upper subspaces and $\ell := \|f - P_{k,j} f\|_\pi$. Then the bound above holds if $\ell < \epsilon\sqrt{\pi_*}/2$ and $\mathrm{Var}_\pi(f) \le (\epsilon\sqrt{\pi_*} - 2\ell)^2 c$.

The proof is by showing a generic reduction from the problem of concentration for random walks to the well-studied problem of concentration for martingales (see e.g. [4]). This reduction is a generalization of [8, Theorem 1.6] and it may be of independent interest. As in [8], we then use Azuma's inequality to finish the proof.

1.3.3 Bounding the Shrinkage of k-cochains

The following proposition shows that functions that do not "come from below" shrink well, thereby relating the structure of a complex and its spectral properties.

▶ **Proposition 5** (Shrinkage of k-cochains, informal, for formal see Proposition 41)**.** *Let X be a d-dimensional γ-local-spectral expander, let $0 \le k < d$, and assume γ is small enough. Let $f : X(k) \to \mathbb{R}$ such that f is from level $j+1$ for some $-1 \le j < k$. Then*

$$\lambda_{M_k^+}(f) \le \frac{k-j}{k+2} + o_\gamma(1).$$

We prove the proposition by showing that the spectral decomposition (see Definition 24) and the combinatorial decomposition (see Definition 23) of $C^k(X)$, both originally defined in [17], are identical (see Theorem 36).

Furthermore, we calculate the number of functions that do not come from below and hence shrink well (see Corollary 45), to show that they are very common.

1.4 Related Work and Comparison

First of all, let us compare our expander Chernoff bound, stated in Theorem 29, to other related results. Compared to the bound from [8, Theorem 1.6], our bound improves and generalizes their proof in a few ways. First, our proof applies to any stationary distribution, not necessarily a uniform one. Secondly, our bound depends on $\lambda(f)$ (see Definition 3) instead of $\lambda(G)$. Thirdly, we provide a better bound for the case where $\mathrm{Var}_\pi(f)$ is small enough. Lastly, we maintain the dependence of the bound on $\mathrm{Var}_\pi(f)/\pi_*$ instead of bounding it by n. Note that the proof of [8] applies to vector-valued functions $f : V \to \mathbb{R}^N$, but we only prove the case of $N = 1$ for simplicity.

Compared to other similar bounds, the first advantage of our bound is that it depends on $\lambda(f)$ instead of $\lambda(G)$. Secondly, our proof has the advantage of being elementary and not use perturbation theory. Lastly, our bound includes a special case concerning $\mathrm{Var}_\pi(f) \le \pi_*\epsilon^2/4\lambda(f)^2$, in which the bound does not depend on the spectral gap at all and therefore it is better than any other bound that we know of.

Since our bound depends on $\ln(4\,\mathrm{Var}_\pi(f)/\epsilon^2\pi_*)$, it is a better fit for cases where $\mathrm{Var}_\pi(f)$ is small enough compared to ϵ (or, equivalently, ϵ is large enough), and where the stationary distribution is close to being uniform.

For example, compared to [10] (see (1)), the main advantage of our bound is that it depends on $(1-\lambda(f))^2$, which can sometimes be much larger than $1-\lambda_2(G)$. Moreover, our proof is elementary whereas their proof is based on perturbation theory. However, their bound does not depend on $\mathrm{Var}_\pi(f)$ at all, so it is better when $\mathrm{Var}_\pi(f)/\pi_*$ is large compared to ϵ.

Compared to [18, Theorem 1.1], both bounds are influenced by the variance of f, but our bound depends on $\ln(4\,\mathrm{Var}_\pi(f)/\epsilon^2\pi_*)$ whereas their bound depends on $\|f\|_2^2$. So for most values of $\|f\|_2$, π_* and ϵ, e.g. when $\|f\|_2 > 8$, $\pi_* = 1/|V|$ and $\epsilon > 0.01$, our bound is better. Other than that, as with [10], their bound has $1-\lambda_2(G)$ while our bound has $(1-\lambda(f))^2$, and their proof is based on the non-elementary perturbation theory while our method has the advantage of being elementary.

Compared to [18, Section 4], their bound depends on $\|f\|_2^2$, whereas our bound depends on $\ln(4\operatorname{Var}_\pi(f)/\epsilon^2\pi_*)$, which is mostly better when ϵ is not too small. Moreover, their bound is only applicable for $\epsilon \le \|f\|_2^2$. Lastly, as before, their bound depends on $1-\lambda(G)$, and our bound depends on $(1-\lambda(f))^2$ which can be significantly smaller.

Secondly, let us compare our Chernoff bound for high-dimensional expanders, stated in Theorem 48, to related results. As far as we know, there are no existing high-dimensional Chernoff bounds. However, we can still compare our bound to the bounds we get by applying the existing (one-dimensional) graph Chernoff bounds to M_k^+. This comparison is covered in depth by the above paragraphs. In short, if f has a small enough variance and does not "come from below", our bound is better since it depends on $\lambda(f)$ instead of $\lambda(M_k^+)$ or $\lambda_2(M_k^+)$. Note that since all of the eigenvalues of M_k^+ are non-negative, $\lambda(f) \le \lambda(M_k^+) = \lambda_2(M_k^+)$.

Finally, our work in Section 5 improves some results from [17]. Corollary 39 improves [17, Theorem 5.9] (stated here as Proposition 19). It shows that the set containing the spectrum of M_k^+ is even smaller than what was known before. As a consequence, corollary 40, which gives an upper bound on $\|(M_k^+)^i\phi\|$, improves [17, Corollary 5.11].

1.5 Organization

This paper is organized as follows. Section 2 provides the relevant background in detail. In Section 3 we study the shrinkage of a vector denoted $\lambda(w)$. In Section 4 we provide a Chernoff bound for expander graphs which depends on the shrinkage of the function on the vertices. In Section 5 we study the shrinkage of vectors by high-order random walks and also improve some existing results. In Section 6 we prove our main result regarding Chernoff bound for high-dimensional expanders. In Appendix A we prove the the bound from Section 4. Due to space restrictions, some of the proofs are omitted and can be found in the full version of the paper.

2 Preliminaries

In this section, we present the relevant background about expander graphs, high-dimensional expanders, random walks, and decompositions of the subspace of k-cochains. We provide more details in the full version of this paper.

2.1 Expander Graphs and Random Walks

In this subsection, we define walk operators and the inner product-space they act on. Then, we define λ-spectral expander graphs.

In this paper, we will consider only finite state-spaces and only *time-homogeneous* Markov chains. A *random walk* M is a finite reversible Markov chain. A sequence of random variables according to M is also called random walk. Throughout this paper, for ease of presentation and connection with graphs, we say that a random walk is *connected* instead of irreducible. By the book of [11], a connected random walk has a positive and unique stationary distribution.

▶ **Definition 6** (Walk operator, $l_2(\pi)$). *Let M be a connected random walk over state-space V and let π be its stationary distribution. We denote by $l_2(\pi)$ the vector-space of functions $f : V \to \mathbb{R}$ with inner product $\langle f, g\rangle_\pi = \sum_{i\in V} \pi(i)f(i)g(i)$. It is indeed an inner product since π is positive. The* walk operator *$M : l_2(\pi) \to l_2(\pi)$ is defined by $(Mf)(i) := \sum_{j\in V} M(i,j)f(j)$.*

It is easy to see that a connected random walk operator M is self-adjoint. Therefore, by the spectral theorem, there is an orthonormal basis of $l_2(\pi)$ consisting of eigenvectors of M, and each eigenvalue of M is real. We are now ready to define expanders:

▶ **Definition 7** (λ-spectral expander graph)**.** *Let G be an undirected connected graph with n vertices and let $\lambda_1 \geq \cdots \geq \lambda_n$ be its (real) eigenvalues. G is a* λ-spectral expander *if $\lambda(G) \leq \lambda$ where $\lambda(G) := \max\{|\lambda_2|, |\lambda_n|\}$. Alternatively, G is a* one-sided λ-spectral expander *if $\lambda_2 \leq \lambda$.*

A small value of λ means that the graph is a better expander. λ also controls how well M shrinks functions, since $\|Mf\|_\pi \leq \lambda\|f\|_\pi$ for any f such that $\mathbb{E}_\pi[f] = 0$ (again by the spectral theorem). Therefore good expanders *shrink functions* well. The following proposition is well-known:

▶ **Proposition 8.** *Let G be an undirected connected graph. G is non-bipartite if and only if $\lambda(G) < 1$.*

2.2 High-Dimensional Expanders

High-dimensional expanders are a special type of abstract simplicial complexes. They are similar to graphs, but they consist of faces instead of just edges. Faces can be edges, but they can also be vertices, triangles, pyramids, etc. Let us provide the formal definitions.

A *face* τ is a set of vertices. The *dimension* of τ is defined by $\dim(\tau) := |\tau| - 1$. Let X be a set of faces. X is an *abstract simplicial complex* if for every face $\tau \in X$, if $\sigma \subset \tau$ then $\sigma \in X$. We denote by $X(k)$ the set of faces of dimension $k \geq -1$. The *dimension* of X is $\max\{\dim(\tau) \mid \tau \in X\}$.

▶ **Definition 9** (Pure simplicial complex)**.** *Let X be a d-dimensional abstract simplicial complex. X is called* pure *if for every $\sigma \in X$, there is a d-dimensional face τ such that $\sigma \subset \tau$.*

As a generalization of weighted graphs, we have weighted simplicial complexes. The weight of smaller faces is induced by the weight of larger faces to comply with the high-dimensional structure.

▶ **Definition 10** (Weighted simplicial complex)**.** *Let X be a pure d-dimensional simplicial complex. Let π_d be a positive probability distribution over its d-dimensional faces. For each $-1 \leq k \leq d$, the induced probability distribution over the k-faces is denoted π_k, where $\pi_k(\sigma)$ is the probability to choose a d-face τ according to π_d and then get σ when uniformly choosing a k-face contained by τ. A pure simplicial complex equipped with a distribution is called a* weighted simplicial complex*.*

There are a few related but not equivalent definitions of HD expanders. Each one of them generalizes a different aspect of expander graphs. We stick to the definition that is related to random walks, called *local-spectral expander*. Basically, it says that from any local point of view (or *link*), the complex should be expanding.

▶ **Definition 11** (Link)**.** *Let X be a pure d-dimensional simplicial complex. Let $\tau \in X(k)$ be a k-dimensional face. The* link *of τ, denoted X_τ, is a pure $(d-k-1)$-dimensional simplicial complex obtained by taking every face that contains τ and removing τ's vertices from it. Formally speaking, $X_\tau := \{\sigma \setminus \tau \mid \tau \subseteq \sigma \in X\}$.*

If X is weighted, we give each top face in X_τ a probability proportional to its probability in X, normalized so that the sum of the probabilities is 1. The probability is well-defined since every face in X_τ is also a face in X.

▶ **Definition 12** (Skeleton). *Let X be an abstract simplicial complex. The* skeleton *of X, also called the* underlying graph*, is an undirected graph with $X(0)$ as its vertices and $X(1)$ as its edges. If X is weighted, the skeleton inherits its weights.*

▶ **Definition 13** (High-dimensional expander). *Let X be a pure complex. If the skeleton of every one of its links is a γ-spectral expander then X is a γ-*local-spectral expander*.*

▶ **Definition 14** (Degree). *Let X be an abstract simplicial complex and let $v \in X(0)$ be a vertex in X. The* degree *of v is the number of faces that contain it, namely, $|X_v|$. The* degree *of X is $\max_{v \in X(0)} |X_v|$.*

As with expander graphs, we are interested in infinite families of d-dimensional bounded degree γ-local-spectral expanders. Explicit constructions of such families exist, see [7, Lemma 1.5] for example. Moreover, we know that explicit constructions with a D-flat top distribution exist:

▶ **Definition 15** (D-flat). *Let π be a probability distribution over V and let $D \in \mathbb{N}$. We say that π is D-flat if there exists some $R \in \mathbb{R}$ such that $\pi_v \in \{\frac{1}{R}, \frac{2}{R}, \ldots, \frac{D}{R}\}$ for all $v \in V$.*

▶ **Proposition 16** ([6, Claim 6.2]). *For every $\gamma > 0$ and $d \in \mathbb{N}$ there exists an explicit infinite family of bounded degree d-dimensional complexes which are γ-local-spectral expanders. They have top distribution π_d that is D-flat, for $D \leq (1/\gamma)^{O(d^2/\gamma^2)}$.*

2.2.1 Random Walks on High-Dimensional Expanders

One can define many different random walks on a given HD expander. We are specifically interested in a walk that captures the expanding structure of the complex. Such walks were defined for the first time by [15].

▶ **Definition 17** (Lazy upper random walk). *Let X be a weighted pure d-dimensional simplicial complex. Let $0 \leq k < d$. Define M_k^+ to be the* lazy upper random walk *on the k-dimensional faces: For each $\sigma, \tau \in X(k)$, $M_k^+(\sigma, \tau)$ is the probability to choose a $(k+1)$-face $\eta \supset \sigma$ according to π_{k+1} and then get τ when uniformly choosing a k-face contained by η. Equivalently, $M_k^+ : X(k) \times X(k) \to \mathbb{R}$ is:*

$$M_k^+(\sigma, \tau) := \begin{cases} \dfrac{1}{k+2} & \sigma = \tau \\ \dfrac{\pi_{k+1}(\sigma \cup \tau)}{(k+2)\pi_k(\sigma)} & \sigma \cup \tau \in X(k+1) \\ 0 & \textit{otherwise} \end{cases}$$

This walk is called lazy *since $M(\sigma, \sigma) > 0$ so it considers standing-still a possible move.*

Let us provide the following definition from [17, Theorem 5.6]:

▶ **Definition 18.** *Let X be a weighted pure d-dimensional γ-local-spectral expander. Define $\epsilon_0 = \gamma$ and $\epsilon_k = 2k(1 + 2k\sqrt{k})\epsilon_{k-1} + (k+1)\gamma$ for all $0 < k < d$. We say that γ is* small enough with respect to d *if $\epsilon_{k-1} \leq \frac{1}{2(1+2k\sqrt{k})}$ for all $0 < k < d$, and $\epsilon_{d-1} < \dfrac{1}{2\sqrt{d}}$.*

Note that $\lim_{\gamma \to 0} \epsilon_k = 0$. Therefore, we can construct a γ-local-spectral expander with a small enough γ by using the explicit construction from Proposition 16 with a sufficiently small γ.

The following proposition shows that even though the second largest eigenvalue of M_k^+ goes to 1 as k increases, we can go beyond that eigenvalue to achieve much better results. For example, when γ is small enough, the smallest eigenvalue of M_k^+ goes to 0 as k increases.

▶ **Proposition 19** ([17, Theorem 5.9]). *Let X be a weighted d-dimensional γ-local-spectral expander. If γ is small enough with respect to d then for every $0 \le k < d$:*

$$\mathrm{Spec}(M_k^+) \subseteq \{1\} \cup \bigcup_{j=0}^{k} \left[\frac{k+1-j}{k+2} - \frac{\sqrt{k+1}}{k+2}\epsilon_k, \frac{k+1-j}{k+2} + \frac{\sqrt{k+1}}{k+2}\epsilon_k \right]$$

where $\mathrm{Spec}(M_k^+)$ *is the set of eigenvalues of* M_k^+.

2.2.2 Decompositions of the Subspace of k-Cochains

Let X be an d-dimensional γ-local-spectral expander. For all $-1 \le k \le d$, define $C^k(X)$ to be the subspace of k-cochains, i.e., functions $\phi : X(k) \to \mathbb{R}$. Let $\pi = \pi_k$ be the probability distribution over $X(k)$. Let the inner-product of $C^k(X)$ be $\langle \phi, \psi \rangle \coloneqq \sum_{i \in X(k)} \pi(i)\phi(i)\psi(i)$ and let $\|\cdot\|$ be the induced norm. Note that as a walk operator, $M_k^+ : C^k(X) \to C^k(X)$.

For all $-1 \le k \le d$, define $C_0^k(X) \coloneqq \{\phi \in C^k(X) \mid \phi \perp \mathbf{1}\}$ where $\mathbf{1}$ is the constant 1 function. For all $-1 \le j \le k \le d$, the operator $d_{j\nearrow k} : C^j(X) \to C^k(X)$ is defined by:

$$d_{j\nearrow k}\phi(\sigma) \coloneqq \sum_{\substack{\tau \in X(j) \\ \tau \subseteq \sigma}} \phi(\tau)$$

for all $\phi \in C^j(X)$ and $\sigma \in C^k(X)$. Note that $d_{k\nearrow k}$ is the identity function. In addition, for $k < d$, we denote $d_k \coloneqq d_{k\nearrow k+1}$.

▶ **Lemma 20** ([17, Section 5.3]). *Let $0 \le j \le k$. Then $d_{j-1\nearrow k}(C_0^{j-1}(X)) \subseteq d_{j\nearrow k}(C_0^j(X))$.*

This operator is highly related to the lazy upper random walk:

▶ **Lemma 21** ([17, Corollary 3.7]). *$d_k^* d_k = (k+2)M_k^+$ for all $-1 \le k < d$.*

In addition, $d_{j\nearrow k}$ and d_j are related in the following way:

▶ **Lemma 22** ([17, Proposition 5.5]). *$d_{j\nearrow k} = \frac{1}{(k-j)!} d_{k-1} \dots d_j$ for all $-1 \le j < k \le d$.*

The following definition implies a combinatorial orthogonal decomposition of $C^k(X)$.

▶ **Definition 23.** *For all $0 \le j \le k \le d$, let U_k^j be a subspaces of C_0^k that contains k-cochains that come from j-cochains but not from $(j-1)$-cochains. Formally,*

$$U_k^j \coloneqq d_{j\nearrow k}(C_0^j(X)) \cap (d_{j-1\nearrow k}(C_0^{j-1}(X)))^{\perp}$$

By [17, Section 5.3], the following is an orthogonal decomposition:

$$C^k(X) = \mathrm{span}\{\mathbf{1}\} \oplus U_k^k \oplus U_k^{k-1} \oplus \cdots \oplus U_k^0.$$

This decomposition is *combinatorial* in the sense that each subspace U_k^j is related to the structure of X. Let us provide the definition of another orthogonal decomposition of $C^k(X)$, which is a *spectral* decomposition, relying on Proposition 19.

▶ **Definition 24.** *For all $0 \le j \le k < d$, we define the* neighborhoods *of M_k^+ to be:*

$$W_k^j \coloneqq \mathrm{span}\left\{ \phi \;\middle|\; M_k^+\phi = \mu\phi, \mu \in \left[\frac{k+1-j}{k+2} - \frac{\sqrt{k+1}}{k+2}\epsilon_k, \frac{k+1-j}{k+2} + \frac{\sqrt{k+1}}{k+2}\epsilon_k \right] \right\}$$

If γ is small enough then these subspaces intersect trivially and hence:

$$C^k(X) = \mathrm{span}\{\mathbf{1}\} \oplus W_k^k \oplus W_k^{k-1} \oplus \cdots \oplus W_k^0.$$

3 The Shrinkage of a Vector

Let W be a finite-dimensional vector-space with inner-product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$.

Given $u, w \in W$, we say that u is perpendicular to w and denote by $u \perp w$, if $\langle u, w \rangle = 0$. Similarly, $u \not\perp w$ if $\langle u, w \rangle \neq 0$. Moreover, given a subspace U, we denote $w \perp U$ if $w \perp u$ for all $u \in U$. We denote $w \not\perp U$ if there exists a vector $u \in U$ such that $w \not\perp u$.

Let $M : W \to W$ be a self-adjoint operator. By the spectral theorem, there exists an orthonormal basis of W that consists of eigenvectors of M.

▶ **Lemma 25.** *Let $w \in W$ and $k \geq 0$. Then $\lambda(M^k w) = \lambda(w)$.*

Proof. Let $k \geq 0$. By the definition of $\lambda(w)$ it is enough to show that $w \perp \phi \iff (M^k w) \perp \phi$ where ϕ is an eigenvector of M with eigenvalue $\lambda \neq 0$. This is true because $\langle M^k w, \phi \rangle = \langle w, M^k \phi \rangle = \langle w, \lambda^k \phi \rangle = \lambda^k \langle w, \phi \rangle$ since M^k is self-adjoint. ◀

The following lemmas explains why the shrinkage of a vector is called this way. They show that $\lambda(w)$ is an upper bound for how much M shrinks w.

▶ **Lemma 26.** *Let $w \in W$. Then $\|Mw\| \leq \lambda(w)\|w\|$.*

▶ **Lemma 27.** *Let $w \in W$ and $k \geq 0$. Then $\|M^k w\| \leq \lambda(w)^k \|w\|$.*

Proof. The proof is by induction on k together with Lemma 25 and Lemma 26. ◀

▶ **Lemma 28.** *Let $w \in W$ and let $B = \{\phi_i\}$ be an orthogonal basis of eigenvectors of M matching eigenvalues λ_i respectively. Let $C \subseteq B$ such that $w \in \operatorname{span}(C)$. Then $\lambda(w) \leq \max_{\phi_i \in C} |\lambda_i|$.*

Proof. By definition $\lambda(w) = |\lambda_j|$ for some index j such that $w \not\perp \phi_j$. This implies that $\phi_j \in C$ since $w \in \operatorname{span}(C)$. Therefore $\lambda(w) = |\lambda_j| \leq \max_{\phi_i \in C} |\lambda_i|$. ◀

4 Expander Chernoff Bound for Shrinking Functions

In the following section, we show an expander Chernoff bound that depends on the shrinkage of the function on the vertices $\lambda(f)$, instead of the graph's expansion. It allows us to see beyond the spectral gap and instead consider the entire spectrum of the expander. As far as we are concerned, it is the only bound of this kind. The proof is deferred to the Appendix.

▶ **Theorem 29** (Expander Chernoff bound that depends on $\lambda(f)$ instead of $\lambda(M)$)**.** *Let $M : l_2(\pi) \to l_2(\pi)$ be a connected non-bipartite random walk operator over state-space $V = \{1, \ldots, n\}$ with a stationary distribution π. Let $w_1, \ldots, w_t$ be a random walk according to M with initial distribution π and denote $\pi_* := \min_i \pi_i$. Let $f \in l_2(\pi)$ be a function satisfying $\mathbb{E}_\pi[f] = 0$. Then for every $\epsilon > 0$:*

$$\mathbb{P}[|\frac{1}{t}\sum_{i=1}^{t} f(w_i)| > \epsilon] \leq \begin{cases} 2\exp\Big(-\frac{t\epsilon^2}{32\|f\|_\infty^2}\Big) & \operatorname{Var}_\pi(f) \leq \frac{\epsilon^2 \pi_*}{4\lambda(f)^2} \\ 2\exp\Big(-\frac{t\epsilon^2}{32\|f\|_\infty^2 \mathcal{L}^2}\Big) & \textit{otherwise} \end{cases}$$

where $\mathcal{L} := \left\lceil \frac{\ln(4\operatorname{Var}_\pi(f)/\epsilon^2 \pi_*)}{1 - \lambda(f)} \right\rceil$.

▶ Remark 30. We can prove a more general version of Theorem 29, appropriate for functions that have a small mass on eigenspaces associated with a large eigenvalue. The proof is the same, except we use the generalized version of Proposition 52 which appears in Remark 62. Define P as the orthogonal projection on eigenspaces associated with small eigenvalues, and denote $\ell := \|f - Pf\|_\pi$. If $\ell < \epsilon\sqrt{\pi_*}/2$, then we get the following expander Chernoff bound. For all $\epsilon > 0$:

$$\mathbb{P}[|\frac{1}{t}\sum_{i=1}^{t} f(w_i)| > \epsilon] \leq \begin{cases} 2\exp\Big(-\dfrac{t\epsilon^2}{32\|f\|_\infty^2}\Big) & \mathrm{Var}_\pi(f) \leq \Big(\dfrac{\epsilon\sqrt{\pi_*} - 2\ell}{2\lambda(Pf)}\Big)^2 \\ 2\exp\Big(-\dfrac{t\epsilon^2}{32\|f\|_\infty^2 \mathcal{L}^2}\Big) & \text{otherwise} \end{cases}$$

where $\mathcal{L} := \left\lceil \dfrac{\ln(4\,\mathrm{Var}_\pi(f)/(\epsilon\sqrt{\pi_*} - 2\ell)^2)}{1 - \lambda(Pf)} \right\rceil$.

5 Shrinking Functions Do Not Come From Below

In this section, we discuss which functions M_k^+ shrinks well. We prove that these are exactly functions that do not come from below. In addition, we find what it means to "come from below" and how rare it is, by studying the two decompositions of $C_0^k(X)$ and the spectrum of M_k^+.

First, we provide an equivalent definition for functions from level $j+1$ (and for functions that do not come from below). The following lemmas relate functions that do not come from below to the subspace $(\mathrm{Im}\, d_{j\nearrow k})^\perp$.

▶ **Lemma 31.** *Let X be a d-dimensional γ-local-spectral expander. Let $-1 \leq j \leq k \leq d$ and $f \in C^k(X)$. Then $d^*_{j\nearrow k} f(\tau) = \mathbb{E}_{\pi_\tau}[f_\tau]$ for all $\tau \in X(j)$.*

Proof. We prove by induction on j. For $j = k$, we have that $d^*_{k\nearrow k} f(\tau) = f(\tau)$. On the other hand, $X_\tau = \varnothing$ and $\pi_\tau(\varnothing) = 1$. Therefore, $\mathbb{E}_{\pi_\tau}[f_\tau] = f_\tau(\varnothing) = f(\tau \cup \varnothing) = f(\tau) = d^*_{k\nearrow k} f(\tau)$ as needed. Next, we assume that the lemma holds for all $j < m \leq k$ and prove for j. Note that

$$\mathbb{E}_{\pi_\tau}[f_\tau] = \sum_{\sigma \in X_\tau(k-j-1)} \pi_\tau(\sigma) f_\tau(\sigma) = \sum_{\nu \in X(k), \tau \subseteq \nu} \frac{\pi(\nu)}{\pi(\tau)} f(\nu)$$

And by the induction assumption,

$$\begin{aligned} d^*_{j\nearrow k} f(\tau) &= \frac{1}{k-j} \cdot d^*_j d^*_{j+1\nearrow k} f(\tau) = \frac{1}{k-j} \cdot \sum_{\eta \in X(j+1), \tau \subseteq \eta} \frac{\pi(\eta)}{\pi(\tau)} d^*_{j+1\nearrow k} f(\eta) \\ &= \frac{1}{k-j} \cdot \sum_{\eta \in X(j+1), \tau \subseteq \eta} \frac{\pi(\eta)}{\pi(\tau)} \mathbb{E}_{\pi_\eta}[f_\eta] \\ &= \frac{1}{k-j} \cdot \sum_{\eta \in X(j+1), \tau \subseteq \eta} \frac{\pi(\eta)}{\pi(\tau)} \sum_{\nu \in X(k), \eta \subseteq \nu} \frac{\pi(\nu)}{\pi(\eta)} f(\nu) \\ &= \sum_{\nu \in X(k), \tau \subseteq \nu} \frac{\pi(\nu)}{\pi(\tau)} f(\nu) = \mathbb{E}_{\pi_\tau}[f_\tau]. \end{aligned}$$

◀

▶ **Lemma 32.** *Let X be a d-dimensional γ-local-spectral expander. Let $-1 \leq j \leq k \leq d$ and $f \in C^k(X)$. Then $f \in (\mathrm{Im}\, d_{j\nearrow k})^\perp$ if and only if f is from level $j+1$.*

Proof. First note that $f \in (\operatorname{Im} d_{j\nearrow k})^{\perp}$ if and only if $f \in \ker d^*_{j\nearrow k}$, if and only if $d^*_{j\nearrow k} f = 0$. Therefore, by Lemma 31, $\mathbb{E}_{\pi_\tau}[f_\tau] = 0$ for all $\tau \in X(j)$, which means f is from level $j+1$. ◀

Next, we provide a lemma that formulates the relation between functions that do not "come from below" and the subspaces U_k^j. But first, we need the following lemma.

▶ **Lemma 33.** *Let X be a d-dimensional γ-local-spectral expander. For all $-1 \le j \le k \le d$:*

$$C_0^k(X) \cap (d_{j\nearrow k} C_0^j(X))^{\perp} = (\operatorname{Im} d_{j\nearrow k})^{\perp}$$

Note that $\operatorname{Im} d_{j\nearrow k} := d_{j\nearrow k}(C^j(X)) = \{d_{j\nearrow k}(\psi) \mid \psi \in C^j(X)\}$ is the subspace of all k-cochains that come from j-cochain.

Proof. Since $(C_0^k(X) \cap (d_{j\nearrow k} C_0^j(X))^{\perp})^{\perp} = \operatorname{span}\{\mathbf{1}_k\} + d_{j\nearrow k} C_0^j(X)$, the lemma follows from the following equality: $d_{j\nearrow k}\mathbf{1}_j = \binom{k+1}{j+1}\mathbf{1}_k$. ◀

The following lemma shows that functions belong to $\bigoplus_{j<m\le k} U_k^m$ if and only if they are from level $j+1$.

▶ **Lemma 34.** *Let X be a d-dimensional γ-local-spectral expander. For all $-1 \le j \le k \le d$:*

$$\bigoplus_{j<m\le k} U_k^m = (\operatorname{Im} d_{j\nearrow k})^{\perp}$$

(we define the left-hand side to be the zero subspace when $j = k$).

Proof. The Lemma follows from Lemma 20, the definition of U_k^j, and Lemma 33. ◀

Now, let us study how well M_k^+ shrinks functions that do not come from below. By the above lemmas, we already know that if a function does not come from below, it does not belong to U_k^j where j is low. Additionally, by the definition of W_k^j, we know that M_k^+ shrinks $f \in W_k^j$ well when j is high. Therefore, our next step is to show how the two decompositions of $C_0^k(X)$: the spectral decomposition denoted W_k^j, and the combinatorial decomposition denoted U_k^j, are related.

From the work of [17] we already know that the projection of ϕ on W_k^j can be approximated by its projection on U_k^j:

▶ **Proposition 35** ([17, Theorem 5.10, simplified])**.** *Let X be a d-dimensional γ-local-spectral expander, let $0 \le k < d$, and assume γ is small enough. For all $0 \le j \le k$ and $\phi \in C_0^k(X)$:*

$$(1 - b(k)\gamma)\|P_{U_k^j}\phi\| \le \|P_{W_k^j}\phi\| \le (1 + b(k)\gamma)\|P_{U_k^j}\phi\|$$

where $b(k)$ is an explicit positive constant such that $b(k)\gamma < 1$, and P is the orthogonal projection.

Let us prove that the two decompositions are actually identical:

▶ **Theorem 36** (Spectral decomposition and combinatorial decomposition are identical)**.** *Let X be a d-dimensional γ-local-spectral expander, let $0 \le k < d$, and assume γ is small enough. Then $W_k^j = U_k^j$ for all $0 \le j \le k$.*

Proof. Let $0 \le j \le k$. We will prove that $U_k^j = W_k^j$ by showing that $(U_k^j)^{\perp} = (W_k^j)^{\perp}$. Let $\phi \in (U_k^j)^{\perp}$. Then $P_{U_k^j}\phi = 0$ and by Proposition 35 we get that $P_{W_k^j}\phi = 0$, and hence $\phi \in (W_k^j)^{\perp}$. Conversely, assume that $\phi \in (W_k^j)^{\perp}$, so $P_{W_k^j}\phi = 0$. By Proposition 35:

$$\frac{1}{1 + b(k)\gamma}\|P_{W_k^j}\phi\| \le \|P_{U_k^j}\phi\| \le \frac{1}{1 - b(k)\gamma}\|P_{W_k^j}\phi\|.$$

Hence $\|P_{U_k^j}\phi\| = 0$, and therefore $\phi \in (U_k^j)^{\perp}$. ◀

▶ **Corollary 37.** *Let X be a d-dimensional γ-local-spectral expander and assume γ is small enough. Let $-1 \leq j < k \leq d$ and $f \in C^k(X)$. Then f is from level $j+1$ if and only if $f \in W_k^{j+1} \oplus \cdots \oplus W_k^k$.*

Proof. The Corollary follows from Lemma 34, Lemma 32, and Theorem 36. ◀

Therefore, functions from "good neighborhoods" do not "come from below". By good neighborhoods we mean the subspaces W_k^j related to small eigenvalues of M_k^+.

We will now use Theorem 36 to improve some existing results. First, the following result improves the spectral decomposition of $C^k(X)$ for the case where X is a good enough expander. It easily follows from [17, Corollary 5.8]

▶ **Proposition 38.** *Let X be a d-dimensional γ-local-spectral expander, let $0 \leq k < d$, and assume γ is small enough. Then for all $0 \leq j \leq k$:*

$$W_k^j = \operatorname{span}\left\{\phi \;\middle|\; M_k^+\phi = \mu\phi, \mu \in \left[\frac{k+1-j}{k+2} - \frac{\epsilon_k}{k+2}, \frac{k+1-j}{k+2} + \frac{\epsilon_k}{k+2}.\right]\right\}$$

Next, the following corollary improves [17, Theorem 5.9].

▶ **Corollary 39.** *Let X be a d-dimensional γ-local-spectral expander, let $0 \leq k < d$, and assume γ is small enough. Then*

$$\operatorname{Spec}(M_k^+) \subseteq \{1\} \cup \bigcup_{j=0}^{k} \left[\frac{k+1-j}{k+2} - \frac{\epsilon_k}{k+2}, \frac{k+1-j}{k+2} + \frac{\epsilon_k}{k+2}\right].$$

Finally, the following corollary of Proposition 38 and Theorem 36 improves [17, Corollary 5.11].

▶ **Corollary 40.** *Let X be a d-dimensional γ-local-spectral expander, let $0 \leq k < d$, and assume that γ is small enough. Then for all $i \in \mathbb{N}$ and $\phi \in C_0^k(X)$:*

$$\sum_{j=0}^{k} \left(\frac{k+1-j}{k+2} - \frac{\epsilon_k}{k+2}\right)^{2i} \left\|P_{U_k^j}\phi\right\|^2 \leq \|(M_k^+)^i\phi\|^2 \leq \sum_{j=0}^{k} \left(\frac{k+1-j}{k+2} + \frac{\epsilon_k}{k+2}\right)^{2i} \left\|P_{U_k^j}\phi\right\|^2$$

where P is the orthogonal projection.

Therefore if ϕ does not come from below then M_k^+ will shrink it better. For example, the best shrinking functions are from U_k^k.

Now, let us bound the shrinkage of f in a way that depends on its neighborhood.

▶ **Proposition 41.** *Let X be a d-dimensional γ-local-spectral expander, let $0 \leq k < d$, and assume γ is small enough. Let $f \in (\operatorname{Im} d_{j \nearrow k})^\perp$ for some $-1 \leq j < k$. Then*

$$\lambda_{M_k^+}(f) \leq \frac{k-j}{k+2} + \frac{\epsilon_k}{k+2}.$$

Proof. By Lemma 34, Theorem 36 and Proposition 38:

$$\begin{aligned} f \in (\operatorname{Im} d_{j \nearrow k})^\perp &= \bigoplus_{j<m\leq k} U_k^m = \bigoplus_{j<m\leq k} W_k^m \\ &= \operatorname{span}\left\{\phi \;\middle|\; M_k^+\phi = \mu\phi, \mu \in \left[\frac{1}{k+2} - \frac{\epsilon_k}{k+2}, \frac{k-j}{k+2} + \frac{\epsilon_k}{k+2}\right]\right\}. \end{aligned}$$

Lemma 28 finishes the proof. ◀

Next, our goal is to prove that there are many functions that shrink well. Therefore, we want to calculate the dimension of each W_k^j. First, let us prove the following lemmas.

▶ **Lemma 42.** *Let X be a d-dimensional γ-local-spectral expander and let $-1 \le k < d$. If γ is small enough then d_k is one-to-one.*

Proof. Let μ be the smallest eigenvalue of M_k^+. By Corollary 39, $\mu \ge \dfrac{1-\epsilon_k}{k+2}$. Since γ is small enough, by Definition 18, $\epsilon_k < 1$ and therefore $\mu > 0$. Hence, M_k^+ is invertible. By Lemma 21, $d_k^* d_k = (k+2)M_k^+$ and therefore d_k is one-to-one. ◀

▶ **Lemma 43.** *Let X be a d-dimensional γ-local-spectral expander and let $-1 \le j \le k \le d$. If γ is small enough then $\dim \operatorname{Im} d_{j\nearrow k} = \dim C^j(X)$ and $\dim(\operatorname{Im} d_{j\nearrow k})^\perp = \dim C^k(X) - \dim C^j(X)$.*

Proof. By Lemma 22, $d_{j\nearrow k} = \frac{1}{(k-j)!} d_{k-1} \dots d_j$. Therefore, since γ is small enough, by Lemma 42, $d_{j\nearrow k}$ is one-to-one. Hence, $\dim \ker d_{j\nearrow k} = 0$. By the rank-nullity theorem, $\dim \operatorname{Im} d_{j\nearrow k} + \dim \ker d_{j\nearrow k} = \dim C^j(X)$ and therefore $\dim \operatorname{Im} d_{j\nearrow k} = \dim C^j(X)$. Next, $(\operatorname{Im} d_{j\nearrow k})^\perp \oplus \operatorname{Im} d_{j\nearrow k} = C^k(X)$ implies that $\dim(\operatorname{Im} d_{j\nearrow k})^\perp = \dim C^k(X) - \dim \operatorname{Im} d_{j\nearrow k}$. Therefore, $\dim(\operatorname{Im} d_{j\nearrow k})^\perp = \dim C^k(X) - C^j(X)$. ◀

Since the combinatorial and spectral decompositions are equal, we can calculate the dimension of U_k^j instead of finding the dimension of W_k^j.

▶ **Proposition 44.** *Let X be a d-dimensional γ-local-spectral expander and let $0 \le j \le k \le d$. If γ is small enough then $\dim U_k^j = |X(j)| - |X(j-1)|$.*

Proof. By [2, Proposition 2.43], $\dim(U_k^j \oplus \cdots \oplus U_k^k) = \dim U_k^j + \dim(U_k^{j+1} \oplus \cdots \oplus U_k^k)$. Applying Lemma 34, we get: $\dim(\operatorname{Im} d_{(j-1)\nearrow k})^\perp = \dim U_k^j + \dim(\operatorname{Im} d_{j\nearrow k})^\perp$. And by Lemma 43: $\dim C^k(X) - \dim C^{j-1}(X) = \dim U_k^j + \dim C^k(X) - \dim C^j(X)$. Therefore, $\dim U_k^j = \dim C^j(X) - \dim C^{j-1}(X) = |X(j)| - |X(j-1)|$. ◀

▶ **Corollary 45.** *Let X be a d-dimensional γ-local-spectral expander and let $0 \le j \le k < d$. If γ is small enough then $\dim W_k^j = |X(j)| - |X(j-1)|$.*

The above corollary shows that the amount of eigenvalues of M_k^+ in each neighborhood is not affected by the dimension k. We conclude that as j increases, W_k^j becomes much bigger, and hence most of the functions do not come from below.

Finally, we show how neighborhoods of different dimensions are connected.

▶ **Proposition 46.** *Let X be a d-dimensional γ-local-spectral expander and let $0 \le j \le k < d$. If γ is small enough then $d_k U_k^j = U_{k+1}^j$ and $d_k^* U_{k+1}^j = U_k^j$.*

The following corollary unravels what it means to "come from below" and provides a simpler decomposition of $C^k(X)$.

▶ **Corollary 47.** *Let X be a d-dimensional γ-local-spectral expander and let $0 \le j \le m \le k < d$. If γ is small enough then:*

$$d_{m\nearrow k} W_m^j = d_{m\nearrow k} U_m^j = U_k^j = W_k^j \qquad \text{and} \qquad d_{m\nearrow k}^* W_k^j = d_{m\nearrow k}^* U_k^j = U_m^j = W_m^j.$$

In particular, for $m = j$, $W_k^j = d_{j\nearrow k} \ker d_{j-1}^*$ and therefore

$$C^k(X) = \operatorname{span}\{\mathbf{1}\} \oplus d_{0\nearrow k} \ker d_{-1}^* \oplus d_{1\nearrow k} \ker d_0^* \oplus \cdots \oplus d_{k-1\nearrow k} \ker d_{k-2}^* \oplus \ker d_{k-1}^*.$$

6 Chernoff Bound for High-Dimensional Expanders

In this section, we apply our findings regarding high-order random walks and the Chernoff bound, to get a Chernoff bound for high-dimensional expander graphs. For functions that do not "come from below" and have a small variance, our bound is better than the naive HD Chernoff bound.

▶ **Theorem 48** (HD expander Chernoff bound). *Let X be a weighted d-dimensional γ-local-spectral expander, let $0 \leq k < d$, and assume that γ is small enough. Let $\pi = \pi_k$ be the probability distribution over $X(k)$ and denote $\pi_* := \min_i \pi_i$. Let $w_1, \dots, w_t$ be a random walk according to M_k^+ with initial distribution π. Let $f : X(k) \to \mathbb{R}$ such that f is from level $j+1$ (see Definition 1) for some $-1 \leq j < k$. Then for all $\epsilon > 0$:*

$$\mathbb{P}[|\frac{1}{t}\sum_{i=1}^{t} f(w_i)| > \epsilon] \leq \begin{cases} 2\exp\left(-\frac{t\epsilon^2}{32\|f\|_\infty^2}\right) & \mathrm{Var}_\pi(f) \leq \frac{\epsilon^2\pi_*}{4h_k(j)^2} \\ 2\exp\left(-\frac{t\epsilon^2}{32\|f\|_\infty^2\mathcal{L}^2}\right) & \textit{otherwise} \end{cases}$$

where $\mathcal{L} := \left\lceil \frac{\ln(4\,\mathrm{Var}_\pi(f)/\epsilon^2\pi_*)}{1-h_k(j)} \right\rceil$, and $h_k(j) := \frac{k-j}{k+2} + \frac{\epsilon_k}{k+2}$ where $\lim_{\gamma\to 0}\epsilon_k = 0$ (see Definition 18).

Proof. First note that by Lemma 32, $f \in (\mathrm{Im}\, d_{j\nearrow k})^\perp$. By Lemma 34, $f \in C_0^k(X)$. By definition of $C_0^k(X)$, we know that $\mathbb{E}_\pi[f] = 0$. Hence, by Theorem 29, for every $\epsilon > 0$:

$$\mathbb{P}[|\frac{1}{t}\sum_{i=1}^{t} f(w_i)| > \epsilon] \leq \begin{cases} 2\exp\left(-\frac{t\epsilon^2}{32\|f\|_\infty^2}\right) & \mathrm{Var}_\pi(f) \leq \frac{\epsilon^2\pi_*}{4\lambda(f)^2} \\ 2\exp\left(-\frac{t\epsilon^2}{32\|f\|_\infty^2\mathcal{Q}^2}\right) & \text{otherwise} \end{cases}$$

where $\mathcal{Q} := \left\lceil \frac{\ln(4\,\mathrm{Var}_\pi(f)/\epsilon^2\pi_*)}{1-\lambda(f)} \right\rceil$. Now, by Proposition 41: $\lambda(f) \leq \frac{k-j}{k+2} + \frac{\epsilon_k}{k+2} = h_k(j)$. Therefore $\mathcal{Q} \leq \mathcal{L}$ and the theorem follows. ◀

▶ Remark 49. We can prove a more general version of this theorem, which applies to cases where f does come from below, but its mass on the lower subspaces is small. Denote $P_{k,j} := P_{(\mathrm{Im}\, d_{j\nearrow k})^\perp}$ the orthogonal projection on the upper subspaces. Denote $\ell := \|f - P_{k,j}f\|_\pi$ and assume $\ell < \epsilon\sqrt{\pi_*}/2$. Then for all $\epsilon > 0$:

$$\mathbb{P}[|\frac{1}{t}\sum_{i=1}^{t} f(w_i)| > \epsilon] \leq \begin{cases} 2\exp\left(-\frac{t\epsilon^2}{32\|f\|_\infty^2}\right) & \mathrm{Var}_\pi(f) \leq \left(\frac{\epsilon\sqrt{\pi_*}-2\ell}{2h_k(j)}\right)^2 \\ 2\exp\left(-\frac{t\epsilon^2}{32\|f\|_\infty^2\mathcal{L}^2}\right) & \text{otherwise} \end{cases}$$

where $\mathcal{L} := \left\lceil \frac{\ln(4\,\mathrm{Var}_\pi(f)/(\epsilon\sqrt{\pi_*}-2\ell)^2)}{1-h_k(j)} \right\rceil$.

References

1 Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 1–12. ACM, 2019. doi:10.1145/3313276.3316385.

2 Sheldon Jay Axler. *Linear algebra done right*, volume 2. Springer, third edition, 1997.

3 Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Math. J. (2)*, 19(3):357–367, 1967. `doi:10.2748/tmj/1178243286`.

4 Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Mathematics*, 3(1):79–127, 2006.

5 Kai-Min Chung, Henry Lam, Zhenming Liu, and Michael Mitzenmacher. Chernoff-Hoeffding bounds for markov chains: Generalized and simplified. In *29th International Symposium on Theoretical Aspects of Computer Science, STACS 2012, February 29th - March 3rd, 2012, Paris, France*, pages 124–135, 2012. `doi:10.4230/LIPIcs.STACS.2012.124`.

6 Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon Ta Shma. List decoding with double samplers. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2134–2153. SIAM, 2019.

7 Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 974–985, 2017. `doi:10.1109/FOCS.2017.94`.

8 Ankit Garg, Yin Tat Lee, Zhao Song, and Nikhil Srivastava. A matrix expander chernoff bound. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 1102–1114, New York, NY, USA, 2018. ACM. `doi:10.1145/3188745.3188890`.

9 D Gillman. A chernoff bound for random walks on expander graphs. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 680–691. IEEE, 1993.

10 David Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

11 Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes*, pages 225, 227. Oxford University Press, third edition, 2001.

12 Alexander D Healy. Randomness-efficient sampling within nc. *Computational Complexity*, 17(1):3–37, 2008.

13 Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

14 Nabil Kahale. Large deviation bounds for markov chains. *Combinatorics, Probability and Computing*, 6(4):465–474, 1997.

15 Tali Kaufman and David Mass. High dimensional random walks and colorful expansion. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

16 Tali Kaufman and Izhar Oppenheim. High Order Random Walks: Beyond Spectral Gap. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, volume 116 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 47:1–47:17, 2018.

17 Tali Kaufman and Izhar Oppenheim. High order random walks: Beyond spectral gap. *Combinatorica*, pages 1–37, 2020.

18 Pascal Lezaud. Chernoff-type bound for finite markov chains. *Annals of Applied Probability*, pages 849–867, 1998.

19 Alexander Lubotzky. Expander graphs in pure and applied mathematics. *Bulletin of the American Mathematical Society*, 49(1):113–162, 2012.

20 Roy Wagner. Tail estimates for sums of variables sampled by a random walk. *Combinatorics, Probability and Computing*, 17(2):307–316, 2008.

21 Avi Wigderson and David Xiao. A randomness-efficient sampler for matrix-valued functions and applications. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 397–406. IEEE, 2005.

A Proving the Expander Chernoff Bound

Our proof is a generalization of [8]. As in [8], the proof is by reduction to the well-studied subject of martingales. First, let us provide the relevant definitions.

A.1 Preliminaries

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability-space: Ω is the sample space which is the set of possible outcomes of an experiment; $\mathcal{F}$ is a σ-algebra that is a set of events, where each event is a set of outcomes; and $\mathbb{P}$ is a probability measure function that assigns a probability to each event.

For example, in our case, the experiment is a random walk of length t on a graph $G = (V, E)$. A single outcome is a sequence of t vertices. The sample space is the set of possible sequences. The σ-algebra of our probability-space is the power set of the sample space. A set in this power set is an event (a set of walks), and the probability of an event is the probability that one of the walks in that event has occurred.

A probability-space can be equipped with a *filtration*, which is a monotone sequence of sub-σ-algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}$. Intuitively, each sub-$\sigma$-algebra $\mathcal{F}_i$ is the information known at time i (which includes what was known before that).

In our case, we are interested in the filtration generated by a random walk on the graph. Let $w_1, \ldots, w_t$ denote a random walk. Each w_i is a random variable from our sample space Ω to $V = \{1, \ldots, n\}$. The value of w_i given a sequence from Ω, is the i-th vertex in that sequence. The σ-algebra generated by $w_1, \ldots, w_i$ represents the information known after i steps and is denoted by $\sigma(w_1, \ldots, w_i)$. This σ-algebra is generated by the events $E_{v_1,\ldots,v_i}$ for all possible walks $(v_1, \ldots, v_i) \in V^i$, where $E_{v_1,\ldots,v_i}$ denotes the event of all possible outcomes that start with $v_1, \ldots, v_i$.

A sequence $\{X_n\}$ of random variables is said to be *adapted* to a filtration $\{\mathcal{F}_n\}$ if X_n is $\mathcal{F}_n$-measurable for all n. Informally, X_n is *$\mathcal{F}_n$-measurable* if X_n is known at time n. A random variable w_i is $\mathcal{F}_j$-measurable if $w_i^{-1}(v) \in \mathcal{F}_j$ for all $v \in V$, which means that w_i is known after j steps. Note that w_i is $\mathcal{F}_j$-measurable for $i \leq j$, and therefore $w_1, \ldots, w_t$ is adapted to the filtration $\{\sigma(w_1, \ldots, w_i)\}_{i=1}^t$.

We can now provide the definition of martingales and martingale difference sequences.

▶ **Definition 50** (Martingale, MDS). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability-space and let $\{\mathcal{F}_n\}$ be a filtration. Let $\{X_n\}$ be a sequence of random variables such that $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$ and $\mathbb{E}[|X_n|] < \infty$ for all n.*

1. *If $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$ for all n then $\{X_n\}$ is a* martingale *with respect to $\{\mathcal{F}_n\}$.*
2. *If $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = 0$ for all n then $\{X_n\}$ is an* MDS *(martingale difference sequence) with respect to the filtration $\{\mathcal{F}_n\}$.*

The following inequality is a generalization of the Chernoff bound to martingales.

▶ **Proposition 51** (Azuma's inequality [3]). *Let $X_0, \ldots, X_t$ be a martingale and suppose there is $c > 0$ such that $|X_i - X_{i-1}| \leq c$ for all $1 \leq i \leq t$ a.s. Then for all $\epsilon > 0$:*

$$\mathbb{P}(|X_t - X_0| \geq \epsilon) \leq 2 \exp\left(\frac{-\epsilon^2}{2tc^2}\right).$$

A.2 Reduction to Martingales

In this subsection we prove the following reduction from the problem of concentration for random walks to the problem of concentration for martingales. The proof is a generalization of the proof from [8, Theorem 1.6].

▶ **Proposition 52** (Reduction to martingales). *Let $M : l_2(\pi) \to l_2(\pi)$ be a connected non-bipartite random walk operator over state-space $V = \{1, \ldots, n\}$ with a stationary distribution π. Let $w_1, \ldots, w_t$ be a random walk according to M with initial distribution π and denote $\pi_* := \min_i \pi_i$. Let $f \in l_2(\pi)$ be a function satisfying $\mathbb{E}_\pi[f] = 0$. Then for every $\epsilon > 0$, there is a martingale difference sequence $Z_1, \ldots, Z_t$ (see Definition 50) with respect to the filtration $\{\sigma(w_1, \ldots, w_i)\}_{i=1}^t$ such that $\frac{1}{t}\sum_{i=1}^t f(w_i) = W + \frac{1}{t}\sum_{i=1}^t Z_i$, where $|W| \le \epsilon$ and each term Z_i satisfies $|Z_i| \le 2\|f\|_\infty \left\lceil \max\left\{\frac{\ln(\|f\|_\pi/\epsilon\sqrt{\pi_*})}{-\ln(\lambda(f))}, 1\right\}\right\rceil$.*

In order to prove this proposition, we first need to prove several lemmas. Let $K \in \mathbb{N}$. We will set it to its optimal value later in the subsection.

Our goal is to show that the sum $\sum_{i=1}^t f(w_i)$ is a martingale difference sequence with respect to the filtration $\{\sigma(w_1, \ldots, w_i)\}_{i=1}^t$, with a small error W. Ideally, we would have wanted $\mathbb{E}[f(w_{i+1}) \mid w_i]$ to be equal to 0. But in fact, the following holds:

▶ **Lemma 53.** *Let $1 \le k \le K$ and $1 \le i \le t-k$. Then $(M^k f)(w_i) = \mathbb{E}\big[(M^{k-1}f)(w_{i+1}) \bigm| w_i\big]$.*

Proof. By definition of conditional expectation with respect to an event, we get that: $\mathbb{E}\big[(M^{k-1}f)(w_{i+1}) \bigm| w_i\big] = \sum_{j=1}^n \mathbb{P}[w_{i+1} = j \mid w_i](M^{k-1}f)(j) = \sum_{j=1}^n M(w_i, j)(M^{k-1}f)(j) = \big(M(M^{k-1}f)\big)(w_i) = (M^k f)(w_i)$. ◀

Therefore, to create an MDS out of $\sum_{i=1}^t f(w_i)$ we define the following random variables: $Y_1^{(k)} := (M^{k-1}f)(w_1)$ for all $1 \le k \le K$, and $Y_i^{(k)} := (M^{k-1}f)(w_i) - (M^k f)(w_{i-1})$ for all $1 < i \le t - k + 1$.

▶ **Lemma 54.** *The following holds: $\sum_{i=1}^t f(w_i) = \sum_{k=1}^K \sum_{i=1}^{t-k+1} Y_i^{(k)} + \sum_{i=1}^{t-K} (M^K f)(w_i)$.*

Proof. Let $1 \le m \le t$ and $1 \le k \le K$. Then $\sum_{i=1}^m (M^{k-1}f)(w_i) = \sum_{i=1}^m Y_i^{(k)} + \sum_{i=1}^{m-1} (M^k f)(w_i)$, and the lemma is easy to prove by induction on K. ◀

The following corollary separates $\sum_{i=1}^t f(w_i)$ to an MDS part and an error part W. Later in this subsection, we will show that $\{Z_i\}$ is indeed an MDS and that W is small.

▶ **Corollary 55.** *Denote $W := \frac{1}{t}\sum_{i=1}^{t-K} (M^K f)(w_i)$, and $Z_i := \sum_{k=1}^{\min\{t-i+1,K\}} Y_i^{(k)}$ for all $1 \le i \le t$. Then $\frac{1}{t}\sum_{i=1}^t f(w_i) = W + \frac{1}{t}\sum_{i=1}^t Z_i$.*

Proof. By interchanging the order of summation,

$$\sum_{k=1}^K \sum_{i=1}^{t-k+1} Y_i^{(k)} = \sum_{i=1}^t \sum_{k=1}^{\min\{t-i+1,K\}} Y_i^{(k)} = \sum_{i=1}^t Z_i.$$

Lemma 54 finishes the proof. ◀

Our next step is to show that $\{Z_i\}$ is a martingale difference sequence with respect to the filtration $\{\sigma(w_1, \ldots, w_i)\}_{i=1}^t$. We prove it by showing that $\{Y_i\}$ is an MDS.

▶ **Lemma 56.** *For all $1 \le k \le K$, $Y_1^{(k)}, \dots, Y_{t-k+1}^{(k)}$ is an MDS with respect to the filtration $\{\sigma(w_1, \dots, w_i)\}_{i=1}^{t-k+1}$.*

Proof. First note that for all $0 \le k \le K$ and $1 \le i \le t-k$:
1. $(M^k f)(w_i)$ is $\sigma(w_1, \dots, w_i)$-measurable.
2. By the Markov property, $\mathbb{E}\big[(M^k f)(w_{i+1}) \bigm| w_1, \dots, w_i\big] = \mathbb{E}\big[(M^k f)(w_{i+1}) \bigm| w_i\big]$.

Let $1 \le k \le K$. By 1, $Y_i^{(k)}$ is $\sigma(w_1, \dots, w_i)$-measurable.

Secondly, we should prove that $Y_1^{(k)}, \dots, Y_{t-k+1}^{(k)}$ has the MDS property. Let $1 \le i \le t-k$. By 1, $(M^k f)(w_i)$ is $\sigma(w_1, \dots, w_i)$-measurable, hence by the stability property of conditional expectation:

$$\mathbb{E}\big[(M^k f)(w_i) \bigm| w_1, \dots, w_i\big] = (M^k f)(w_i). \tag{2}$$

Therefore, we have $\mathbb{E}\big[Y_{i+1}^{(k)} \bigm| w_1, \dots, w_i\big] = \mathbb{E}\big[(M^{k-1} f)(w_{i+1}) - (M^k f)(w_i) \bigm| w_1, \dots, w_i\big] = \mathbb{E}\big[(M^{k-1} f)(w_{i+1}) \bigm| w_i\big] - \mathbb{E}\big[(M^k f)(w_i) \bigm| w_1, \dots, w_i\big] = (M^k f)(w_i) - (M^k f)(w_i) = 0$ where the second equality is by 2 and the third equality is by Lemma 53 and (2). ◀

▶ **Corollary 57.** *$Z_1, \dots, Z_t$ is an MDS with respect to $\{\sigma(w_1, \dots, w_i)\}_{i=1}^{t}$.*

Now, let us bound the size of each Z_i. We need the following lemma:

▶ **Lemma 58.** *Let M be a connected random walk and let $f \in l_2(\pi)$. Then $\|Mf\|_\infty \le \|f\|_\infty$.*

Proof. Let $1 \le i \le n$. Then we have $|(Mf)(i)| \le \sum_j |M(i,j) f(j)| = \sum_j M(i,j)|f(j)| \le \sum_j M(i,j)\|f\|_\infty = \|f\|_\infty$, and therefore $\|Mf\|_\infty \le \|f\|_\infty$. ◀

▶ **Corollary 59.** *Let $1 \le i \le t$. Then $|Z_i| \le 2K\|f\|_\infty$.*

Proof. Since $|Z_i| = |\sum_{k=1}^{\min\{t-i+1,K\}} Y_i^{(k)}| \le \sum_{k=1}^{\min\{t-i+1,K\}} |Y_i^{(k)}| \le \sum_{k=1}^{K} |Y_i^{(k)}|$, we have

$$\begin{aligned} |Z_i| &\le \sum_{k=1}^{K} \big|(M^{k-1} f)(w_i) - (M^k f)(w_{i-1})\big| \le \sum_{k=1}^{K} \big(|(M^{k-1} f)(w_i)| + |(M^k f)(w_{i-1})|\big) \\ &\le \sum_{k=1}^{K} (\|M^{k-1} f\|_\infty + \|M^k f\|_\infty) \le \sum_{k=1}^{K} (\|f\|_\infty + \|f\|_\infty) = 2K\|f\|_\infty \end{aligned}$$

where the last inequality is by Lemma 58. ◀

Our last step before proving Proposition 52 is to bound the error $|W|$. First note that

▶ **Lemma 60.** *Let $1 \le i \le n$. Then $\sqrt{\pi_i}|(M^K f)(i)| \le \lambda(f)^K \|f\|_\pi$.*

Proof. By Lemma 27, we have that $\pi_i (M^K f)(i)^2 \le \sum_{j=1}^{n} \pi_j (M^K f)(j)^2 = \|M^K f\|_\pi^2 \le (\lambda(f)^K \|f\|_\pi)^2$ as needed. ◀

Let us now bound the error W.

▶ **Proposition 61.** $|W| \le \dfrac{\lambda(f)^K \|f\|_\pi}{\sqrt{\pi_*}}$.

Proof. From Lemma 60 we have:

$$|W| \le \frac{1}{t}\sum_{i=1}^{t-K}|(M^K f)(w_i)| \le \frac{1}{t}\sum_{i=1}^{t-K}\frac{1}{\sqrt{\pi_{w_i}}}\sqrt{\pi_{w_i}}|(M^K f)(w_i)|$$
$$\le \frac{1}{t}\sum_{i=1}^{t-K}\frac{1}{\sqrt{\pi_{w_i}}}\cdot\lambda(f)^K\|f\|_\pi \le \frac{t-K}{t}\lambda(f)^K\|f\|_\pi \max_i \frac{1}{\sqrt{\pi_i}} \le \frac{\lambda(f)^K\|f\|_\pi}{\sqrt{\pi_*}}$$

where the last inequality is by definition of π_*. ◀

We can now prove Proposition 52.

Proof of Proposition 52. By Corollary 55: $\displaystyle\frac{1}{t}\sum_{i=1}^{t} f(w_i) = W + \frac{1}{t}\sum_{i=1}^{t} Z_i$ where

1. $W := \displaystyle\frac{1}{t}\sum_{i=1}^{t-K}(M^K f)(w_i)$.
2. $Z_i := \displaystyle\sum_{k=1}^{\min\{t-i+1,K\}} Y_i^{(k)}$ for all $1 \le i \le t$.

By Proposition 61, $|W| \le \dfrac{\lambda(f)^K\|f\|_\pi}{\sqrt{\pi_*}}$. In order for $|W| \le \epsilon$ to apply, we should choose K such that $\dfrac{\lambda(f)^K\|f\|_\pi}{\sqrt{\pi_*}} \le \epsilon$. Note that by Proposition 8, $\lambda(f) < 1$. Therefore

$$K \ge \frac{\ln(\|f\|_\pi/\epsilon\sqrt{\pi_*})}{-\ln(\lambda(f))}. \tag{3}$$

By Corollary 57, $Z_1, \dots, Z_t$ is an MDS with respect to $\{\sigma(w_1, \dots, w_i)\}_{i=1}^t$. Moreover, by Corollary 59, $|Z_i| \le 2K\|f\|_\infty$ for all $1 \le i \le t$. So to get the best bound on $|Z_i|$ we should choose the smallest K possible. Therefore, to comply with (3): $K := \left\lceil \max\left\{\dfrac{\ln(\|f\|_\pi/\epsilon\sqrt{\pi_*})}{-\ln(\lambda(f))}, 1\right\}\right\rceil$ which yields the advertised bound on $|Z_i|$. ◀

▶ Remark 62. We can prove a more general version of this bound that applies to functions that most of their mass is on eigenspaces of small eigenvalues, and a small portion of their mass is on eigenspaces associated with large eigenvalues.

Let P be an orthogonal projection to a direct sum of some of the eigenspaces, such that $\lambda(Pf)$ is small and $\ell := \|f - Pf\|_\pi$ is small too. Then $\|M^K f\|_\pi \le \|M^K Pf\| + \|M^K(I-P)f\|_\pi \le \lambda(Pf)^K\|Pf\|_\pi + \|(I-P)f\|_\pi \le \lambda(Pf)^K\|f\|_\pi + \ell$. Therefore, in Lemma 60 we would get $\sqrt{\pi_i}|(M^K f)(i)| \le \|M^K f\|_\pi \le \lambda(Pf)^K\|f\|_\pi + \ell$. This means we would get that $|W| \le \dfrac{\lambda(Pf)^K\|f\|_\pi + \ell}{\sqrt{\pi_*}}$ and therefore, for $\ell < \epsilon\sqrt{\pi_*}$ we get $|Z_i| \le 2\|f\|_\infty \left\lceil \max\left\{\dfrac{\ln(\|f\|_\pi/(\epsilon\sqrt{\pi_*}-\ell))}{-\ln(\lambda(f))}, 1\right\}\right\rceil$.

A.3 Proof of Theorem 29

Proof. We prove an Expander Chernoff bound that depends on $\lambda(f)$. Let $\epsilon > 0$. Applying Proposition 52 with $\epsilon/2$ we get that there is an MDS $Z_1, \dots, Z_t$ with respect to the filtration $\{\sigma(w_1, \dots, w_i)\}_{i=1}^t$ such that $\frac{1}{t}\sum_{i=1}^t f(w_i) = W + \frac{1}{t}\sum_{i=1}^t Z_i$, where $|W| \le \epsilon/2$ and each term

Z_i satisfies $|Z_i| \le 2\|f\|_\infty \left\lceil \max\left\{\frac{\ln(2\|f\|_\pi/\epsilon\sqrt{\pi_*})}{-\ln(\lambda(f))}, 1\right\}\right\rceil$. In order to use Azuma's inequality, we first need to define a martingale out of our MDS $\{Z_i\}$. Let $X_0 = 0$ and $X_i \coloneqq X_{i-1} + Z_i$ for all $i \ge 1$. Then

$$|X_i - X_{i-1}| = |Z_i| \le 2\|f\|_\infty \left\lceil \max\left\{\frac{\ln(2\|f\|_\pi/\epsilon\sqrt{\pi_*})}{-\ln(\lambda(f))}, 1\right\}\right\rceil, \tag{4}$$

$$\sum_{j=1}^{i} Z_j = \sum_{j=1}^{i}(X_j - X_{j-1}) = \sum_{j=1}^{i} X_j - \sum_{j=0}^{i-1} X_j = X_i - X_0. \tag{5}$$

▶ **Lemma 63.** *$X_1, \dots, X_t$ is a martingale with respect to $\{\sigma(w_1, \dots, w_i)\}_{i=1}^t$.*

Proof. First we'll prove that $X_1, \dots, X_t$ is adapted to the filtration, so we should prove that for all $i \ge 1$, X_i is $\sigma(w_1, \dots, w_i)$-measurable. Since $X_i = \sum_{j=1}^i Z_j$, it's enough to show that each Z_j is $\sigma(w_1, \dots, w_i)$-measurable when $j \le i$. From 57 we know that $Z_1, \dots, Z_t$ is an MDS with respect to our filtration, hence each Z_j is $\sigma(w_1, \dots, w_j)$-measurable. In addition, $\sigma(w_1, \dots, w_j) \subseteq \sigma(w_1, \dots, w_i)$ so each Z_j is also $\sigma(w_1, \dots, w_i)$-measurable. Therefore $X_1, \dots, X_t$ is adapted to the filtration.

Next we will prove that $X_1, \dots, X_t$ has the martingale property. Let $i \ge 1$. Since X_i is $\sigma(w_1, \dots, w_i)$-measurable, $\mathbb{E}[X_i \mid w_1, \dots, w_i] = X_i$ by the stability property of conditional expectation. In addition, since by Corollary 57 $Z_1, \dots, Z_t$ is an MDS, $\mathbb{E}[Z_{i+1} \mid w_1, \dots, w_i] = 0$. Therefore, $\mathbb{E}[X_{i+1} \mid w_1, \dots, w_i] = \mathbb{E}[X_i + Z_{i+1} \mid w_1, \dots, w_i] = \mathbb{E}[X_i \mid w_1, \dots, w_i] + \mathbb{E}[Z_{i+1} \mid w_1, \dots, w_i] = X_i$. ◀

Continuing with the proof of Theorem 29, by equation (5) we have that

$$\begin{aligned}\mathbb{P}\Big[\Big|\frac{1}{t}\sum_{i=1}^{t} f(w_i)\Big| > \epsilon\Big] &= \mathbb{P}\Big[\Big|W + \frac{1}{t}\sum_{i=1}^{t} Z_i\Big| > \epsilon\Big] \le \mathbb{P}\Big[|W| + \Big|\frac{1}{t}\sum_{i=1}^{t} Z_i\Big| > \epsilon\Big] \\ &\le \mathbb{P}\Big[\Big|\frac{1}{t}\sum_{i=1}^{t} Z_i\Big| > \frac{\epsilon}{2}\Big] = \mathbb{P}\Big[|\sum_{i=1}^{t} Z_i| > \frac{t\epsilon}{2}\Big] = \mathbb{P}\Big[|X_t - X_0| > \frac{t\epsilon}{2}\Big],\end{aligned}$$

Let $c \coloneqq 2\|f\|_\infty \left\lceil \max\left\{\frac{\ln(2\|f\|_\pi/\epsilon\sqrt{\pi_*})}{-\ln(\lambda(f))}, 1\right\}\right\rceil$. By (4) we know that $|X_i - X_{i-1}| \le c$. Furthermore, by Lemma 63, X_t is a martingale. Therefore we can apply Azuma's inequality to get: $\mathbb{P}[|\frac{1}{t}\sum_{i=1}^{t} f(w_i)| > \epsilon] \le 2\exp\left(\frac{-(\frac{t\epsilon}{2})^2}{2tc^2}\right) = 2\exp\left(-\frac{t^2\epsilon^2}{8tc^2}\right) = 2\exp\left(-\frac{t\epsilon^2}{8c^2}\right)$.

Since $\mathbb{E}_\pi[f] = 0$, we get $\mathrm{Var}_\pi(f) \coloneqq \sum_i \pi_i(f(i) - \mathbb{E}_\pi[f])^2 = \sum_i \pi_i f(i)^2 = \|f\|_\pi^2$, so

$$c = 2\|f\|_\infty \left\lceil \max\left\{\frac{\ln(2\sqrt{\mathrm{Var}_\pi(f)}/\epsilon\sqrt{\pi_*})}{-\ln(\lambda(f))}, 1\right\}\right\rceil = 2\|f\|_\infty \left\lceil \max\left\{\frac{\ln(4\,\mathrm{Var}_\pi(f)/\epsilon^2\pi_*)}{-2\ln(\lambda(f))}, 1\right\}\right\rceil.$$

Note that $c = 2\|f\|_\infty$ if $\ln(4\,\mathrm{Var}_\pi(f)/\epsilon^2\pi_*) \le -2\ln(\lambda(f))$, or $\mathrm{Var}_\pi(f) \le \epsilon^2\pi_*/4\lambda(f)^2$. Moreover, $c \le 2\|f\|_\infty \left\lceil \max\left\{\frac{\ln(4\,\mathrm{Var}_\pi(f)/\epsilon^2\pi_*)}{1-\lambda(f)}, 1\right\}\right\rceil$ since $2\ln(r) \le r - 1$ for all $0 < r \le 1$. Hence,

$$\mathbb{P}[|\frac{1}{t}\sum_{i=1}^{t} f(w_i)| > \epsilon] \le \begin{cases} 2\exp\Big(-\frac{t\epsilon^2}{32\|f\|_\infty^2}\Big) & \mathrm{Var}_\pi(f) \le \frac{\epsilon^2\pi_*}{4\lambda(f)^2} \\ 2\exp\Big(-\frac{t\epsilon^2}{32\|f\|_\infty^2\mathcal{L}^2}\Big) & \text{otherwise.}\end{cases}$$ ◀