Concentration and Anti-concentration for Markov Chains

by

Shravas Rao

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science

New York University

May, 2019

Oded Regev

Dedication

To Raghu Mama.

Acknowledgements

This thesis would not be possible without the help of many individuals besides myself, so I would like to take the time to acknowledge them. First and foremost, I would like to thank my advisor Oded Regev for guiding me throughout my time in graduate school. It was Oded who introduced me to the background of this thesis and helped me get started doing research. When I started to work on problems independently of him, Oded listened to me when I got stuck, and suggested resources to consult. And most importantly, Oded showed me how to approach research problems by using intuition and understanding to help carry through the more difficult parts.

I would also like to thank Jop Briët. I first met Jop when he was a postdoc at NYU when he somehow thought that I might be a good person to collaborate with. As time went on, we started working together more, and this continued even after Jop moved away to CWI. Jop was the source of many new ideas and problems, and served almost like a second advisor.

I would also like to thank the various other mentors I have had during grad school, including Richard Cole, Igor Shinkar, and Noah Stephens-Davidowitz. Richard agreed to supervise a project on algorithmic game theory, and despite the fact that I decided this was not the area for me, still gave me valuable advice that only someone as senior as him could give. Igor was the postdoc next door who was always eager to tell me and anyone else about his ideas and questions, one of which turned into my first paper in graduate school. Noah probably thinks I'm being silly by calling him a mentor, but Noah helped a lot both in disproving many of my claims about basic math, and also helping me navigate the social side of academia.

I would also like to thank Eyal Lubetzky and Assaf Naor (along with Oded) for serving on my depth qualification exam committee, and Afonso Bandeira (along with Jop and Oded) for serving on my thesis proposal committee. I would also like to thank Subhash Khot and Joel Spencer (along with Jop, Oded, and Noah) for serving on my thesis defense committee.

I would like to thank the staff in the computer science department, including Rosemary Amico and Santiago Pizzini, for making sure the department ran smoothly, answering the questions that none of the faculty seemed to know the answers to, and tolerating my use of my office as a storage unit while I was away for three months.

I have been lucky to have been part of an active theory group at NYU, and would like to thank Divesh Aggarwal, Huck Bennett, Jop Briët, Laura Florescu, Chaya Ganesh, Aleksandr Golovnev, Siyao Guo, Siddharth Krishna, Igor Shinkar, Noah Stephens-Davidowitz, Devanathan Thiruvenkatachari, Omri Weinstein, and Aravindan Vijayaraghavan for providing me with many reading groups, student seminars, and mini-courses to attend. I would especially like to thank Huck for joining me in many two-person reading groups over the years, making the summers when everyone else was gone a little less lonely.

Outside of the theory group, I would like to thank Emily Denton, Siddharth Krishna, Rahul Krishnan, and Kevin Munger for reading over my many manuscripts and applications for fellowships and internships, and for pointing out when I wasn't making any sense. I would also like to thank Sandro Coretti, Francisca Gil, Matthew Morse, Zvonimir Pavlinovic, Mukund Sudarshan, Davi Tozoni, Ioanna Tzailla, along with everyone else, for helping provide me with an active social life.

I would also like to thank Rediet Abebe, Justin Holmgren, and Minseon Shin for being there to talk to when I needed the perspective of a grad student not at NYU. It turns out, surprisingly, that grad students everywhere share similar ups and downs!

Before I was at NYU, I was at MIT, and I would like to thank Scott Aaronson and especially Michael Sipser for providing me with my very first research experiences and for giving me the confidence that graduate school could be in my future. I would also like to thank David Vogan. After asking him to supervise an independent study, a mysterious anonymous donation reversed my rejection from the directed reading program and I was able to spend my last IAP learning about group theory. I would also like to thank Jing Jian, Eric Shyu, and JinHao Wan (along with Rediet Abebe, Justin Holmgren, and Minseon Shin) for taking many math and computer science courses with me, and letting me work with them on completing problem sets and studying for tests.

Even before MIT, I was a high school student in Reynoldsburg Ohio, and I would like to thank my teachers Tony DeGennaro, Harry Gee, and my guidance counselor Amy Cox for providing me the support and encouragement to pursue my interests, even when it took extra effort on their part. I would also like to thank Don Piele, Rob Kolstad, and all of the coaches that helped out with the USA Computing Olympiad when I was a student. This was my very first experience learning about theoretical computer science, and if it weren't for the availability of resources and openness of the program, I would not be here today.

Finally I would like to thank Amma, Baba, and Sujit for their constant support throughout the years, and for cheering me up whenever I had doubts about my abilities. I finally did it!

Abstract

We study tail bounds and small ball probabilities for sums of random variables obtained from a Markov chain. In particular, we consider the following sum

$$S_n = f_1(Y_1) + \cdots + f_n(Y_n)$$

where $\{Y_i\}_{i=1}^{\infty}$ is a Markov chain with state space $[N]$, transition matrix A, and stationary distribution μ such that Y_1 is distributed as μ, and $f_i : [N] \to \mathbb{R}$. We also consider settings in which $f_i(Y_i)$ is vector-valued.

In all results, the bounds are in terms of the spectral gap of the Markov chain. In almost all of the results in this thesis, when the transitions are independent and the spectral gap is 1, the bounds match the corresponding bounds for independent random variables up to constant factors.

- We first obtain tail bounds in the case that only the pth moment of the random variable $f_i(Y_i)$ is bounded. This is a Markov chain version of a corollary of the Marcinkiewicz–Zygmund inequality. Using this, we also obtain tail bounds for S_n when the $f_i(Y_i)$ are elements of an ℓ_q space.

- Next, we obtain tail bounds in the case that the $f_i(Y_i)$ are bounded in the range $[-a_i, a_i]$ for each i. This is a Markov chain version of the Hoeffding inequality. This improves upon previously known bounds in that the dependence is on $\sqrt{a_1^2 + \cdots + a_n^2}$ rather than $\max_i\{a_i\}\sqrt{n}$. Using this, we obtain tail bounds for certain types of random variables in which the $f_i(Y_i)$ are elements of any Banach space.

- Next, we obtain sharp tail bounds when the random variables $f_i(Y_i)$ are bounded and the expected value of S_n is small. This is a Markov chain version of a Poisson approximation to sums of independent random variables. As an application, we explain how such tail bounds can be used to construct simple and explicit resilient functions that match the non-constructive functions shown to exist due to the work of Ajtai and Linial.

- Finally, we show that if the $f_i(Y_i)$ take on values $\{-v_i, v_i\}$ with equal probability and the v_i are Euclidean vectors with norm at least 1, the probability that S_n lies in a ball of volume 1 is small. This is a Markov chain version of the Littlewood-Offord inequality. We also construct a new pseudorandom generator for the Littlewood-Offord problem.

Contents

Chapter 1

Introduction

It is well known by the central limit theorem that in many cases, the probability distribution of the sum of independent random variables tends to that of a normal distribution as the number of terms increases. This is even if the random variables themselves do not have a probability distribution that is similar to a normal distribution. Expanding upon this, one can ask what properties of normal distributions also hold for sums of independent random variables. These can include upper bounds on the tails of a Gaussian (concentration), and on any short interval (anti-concentration). In particular, it is known that for a Gaussian random variable X with mean 0 and variance 1 that

$$\Pr[|X| \geq u] \leq 2\exp(-u^2/2) \tag{1.1}$$

and

$$\Pr\left[|X| \leq u\right] \leq u. \tag{1.2}$$

Thus, one can ask if sums of independent random variables also exhibit concentration and anti-concentration.

Concentration for sums of independent random variables: We start by discussing concentration bounds for sums of random variables. If $X_1, \ldots, X_n$ are independent random variables each with mean 0 and bounded in absolute value by 1, Chernoff bounds tell us that

$$\Pr[|X_1 + \ldots + X_n| \geq u\sqrt{n}] \leq 2\exp(-u^2/2), \tag{1.3}$$

and thus the property of Gaussian random variables exhibited in Eq. (1.1) also applies to sums of independent variables.

If just the pth moment of the X_i are bounded, that is $\mathbb{E}[|X_i|^p] \leq 1$, the following bound on the pth moment of $|X_1 + \cdots + X_n|$ is a corollary of the Marcinkiewicz–Zygmund inequality (MZ37), for some constant C independent of n and p.

$$\mathbb{E}[|X_1 + \cdots + X_n|^p]^{1/p} \leq C\sqrt{pn}. \tag{1.4}$$

If u is small enough, then one can obtain tail bounds as in Eq. (1.3) with just a change in the constant factor of the exponent.

Additionally, if each X_i is bounded in absolute value by a_i, the Hoeffding inequality (Hoe63) states that

$$\Pr\left[|X_1 + \cdots + X_n| \geq u\left(\sum_{i=1}^{n} a_i^2\right)^{1/2}\right] \leq 2\exp(-u^2/2). \tag{1.5}$$

We also point out that in some cases, it is possible to show that the probability distribution of a sum of independent random variables tends to that of a Poisson random variable as the number of terms increases. In particular, it is known that for a Poisson random variable X with mean μ,

$$\Pr[X \geq u] \leq \exp\left(u - \mu - u\ln\frac{u}{\mu}\right). \tag{1.6}$$

If $X_1, \ldots, X_n$ are independent random variables in the range $[0, 1]$ and $\mathbb{E}[X_1 + \cdots + X_n] = \Phi$, then

$$\Pr[X_1 + \cdots + X_n \geq u] \leq \exp\left(u - \Phi - u \ln \frac{u}{\Phi}\right). \tag{1.7}$$

exhibiting the same property as found in Poisson random variables.

Finally in recent years, much attention has been placed on concentration inequalities for vector-valued random variables. That is, rather than assuming that the X_i are scalars, we can assume that they are matrices, or elements of some other Banach space.

One example of such an inequality is due to Ahlswede and Winter (AW02), who considered the case of matrices under the Schatten-∞ norm. They showed that if $X_1, \ldots, X_n$ are random $d \times d$ matrices such that $\|X_i\| \leq 1$ and $\mathbb{E}[X_i] = 0$ for all i, then

$$\Pr[\|X_1 + \cdots + X_n\|_{S_\infty} \geq u\sqrt{n}] \leq d \exp(-cu^2).$$

This inequality has been extended to many more settings in a similar manner to many of the above extensions of the regular Chernoff bound in the scalar case. For a more complete treatment of tail bounds for sums of matrices, see the monograph by Tropp (Tro15).

For general Banach spaces, Naor (Nao12) obtained tail bounds for sums of random variables from a Banach space satisfying certain properties. Before stating the corresponding tail bound, we define a quantity called the modulus of uniform smoothness.

Definition 1. *The modulus of uniform smoothness of a Banach space* $(X, \|\cdot\|)$ *is*

$$\rho_X(\tau) = \sup\left\{\frac{\|x + \tau y\| + \|x - \tau y\|}{2} - 1 : x, y \in X, \|x\| = \|y\| = 1\right\}.$$

Let $(X, \|\cdot\|)$ be a Banach space so that $\rho_X(\tau) \leq s\tau^2$ for some s and all $\tau > 0$. When the elements of the Markov chain are independent, for $f_i : [N] \to \{x \in X : \|x\| \leq a_i\}$ and such that $\mathbb{E}[f_i(Y_i)] = 0$, it was shown in (Nao12) that

$$\Pr\left[\|f_1(Y_1) + \cdots + f_n(Y_n)\| \geq u \left(\sum_{i=1}^{n} a_i^2\right)^{1/2}\right] \leq \exp\left(s + 2 - cu^2\right) \tag{1.8}$$

for some universal constant c.

Anti-concentration for sums of independent random variables: In terms of anti-concentration for sums of random variables, we will focus for the most part on the following problem first posed by Littlewood and Offord. If $v_1, \ldots, v_n \in \mathbb{R}^d$ are fixed vectors of Euclidean length at least 1, and $X_1, \ldots, X_n$ are independent Rademacher random variables, so that $\Pr[X_i = 1] = \Pr[X_i = -1] = 1/2$ for all i, then

$$\Pr[X_1 v_1 + \cdots + X_n v_n \in B] \leq \frac{C}{\sqrt{n}} \tag{1.9}$$

for an open Euclidean ball B with radius 1, and a constant C. The case $d = 1$ was proved by Erdős (Erd45), and for general d by Kleitman (Kle70), and was subsequently improved by series of work (Sal83; Sal85; FF88; TV12). When $d = 1$, Eq. (1.9) shows that this sum of random variables exhibits the property of Gaussian random variables in Eq. (1.2).

Random variables obtained from a Markov chain

In this thesis, we investigate to what extent independence is needed for the sums of random variables to exhibit similar behavior to Gaussian and Poisson random variables. In particular, we consider the case that the random variables are obtained from a Markov chain as follows. Consider a Markov chain $\{Y_i\}_{i=1}^{\infty}$ with state space $[N]$, transition matrix A, and stationary distribution μ such that Y_1 is distributed as π, and let $f_1, \ldots, f_n : [N] \to \mathbb{R}$ be functions on the state space. Then we consider the random variable $f_1(Y_1) + \cdots + f_n(Y_n)$.

For the most part, we will show that the bounds in the independent case also hold when the random variables are obtained from a Markov chain, with an additional dependence on the spectral gap which we define as follows. Let E_μ be the associated averaging operator defined by $(E_\mu)_{ij} = \mu_j$, so that for $v \in \mathbb{R}^N$ $E_\mu v = \mathbb{E}_\mu[v]\mathbf{1}$ where $\mathbf{1}$ is the vector whose entries are all 1. We define $\lambda(Y)$ to be

$$\lambda(Y) = \|A - E_\pi\|_{L_2(\mu)\to L_2(\mu)},$$

and the spectral gap to be $1 - \lambda(Y)$. The averaging operator E_μ is the transition matrix of a Markov chain whose transitions are independent and whose stationary distribution is μ. Thus, in some sense $\lambda(Y)$ represents how close the transitions of $\{Y_i\}_{i=1}^\infty$ are to being independent.

Concentration inequalities for sums of random variables obtained from a Markov chain were first obtained by Gillman, who gave the following tail bound in the case that $f_1 = \cdots f_n = f$, for reversible Markov chains.

$$\Pr[|f(Y_1) + \cdots + f(Y_n)| \geq u\sqrt{n}] \leq 2\exp\left(\frac{-u^2(1-\lambda)}{20}\right). \tag{1.10}$$

These bounds were improved by a series of work, including (Din95; Kah97; Lez98; LP04), where the result due to León and Perron is tight for reversible Markov chains. In particular, this result yields a Markov chain version of both Eq. (1.3) and Eq (1.7).

We remark that the dependence on λ in both Eq. (1.10) is optimal, which was shown in (LP04). In particular, one can consider the Markov chain on two states with the transition matrix

$$\begin{bmatrix} \frac{1+\lambda}{2} & \frac{1-\lambda}{2} \\ \frac{1-\lambda}{2} & \frac{1+\lambda}{2} \end{bmatrix}$$

so that $f_i(1) = 1$ and $f_i(2) = -1$ for all i. Intuitively, the random variable $f_1(Y_1)+\cdots+f_n(Y_n)$ is similar to the sum of $n(1-\lambda)$ geometric random variables with parameter $1-\lambda$. Thus, each is close to $1/(1-\lambda)$ or close to $-1/(1-\lambda)$ with equal probability.

Markov chain versions of Bennett's and Bernstein's inequality, a version of Eq. (1.3) with extra dependencies on the variance of the individual random variables, were obtained in (Lez98; Wag08; Pau15). Healey developed a simpler proof that does not require the use of perturbation theory, and allows for differing functions f_i (Hea08). The case of non-reversible Markov chains was first studied in (CLLM12). In work done independently of this thesis, a version of Hoeffding's inequality for general Markov chains was shown by (FJS18), and a version of Bernstein's inequality for general Markov chains was shown by (JSF18).

In the setting of vector-valued random variables, a tail bound was obtained for ℓ_2 spaces by (Kar07), with the assumption that the functions are bounded, that is, $\|f(v)\|_{\ell_2} \le L$ for all $v \in [N]$ for some constant L. In particular, it was shown that when $f : [N] \to \mathbb{R}^d$,

$$\Pr[\|f(Y_1) + \cdots + f(Y_n)\|_{\ell_2} \ge u\sqrt{n}] \le 3 \cdot 2^{d/2} \exp\left(-\alpha \frac{u^2}{L}\right),$$

where $\mathbb{E}[\langle f(Y_1), u\rangle^2]$, for constants C_1 and C_2, α is a constant that depends on λ. As λ goes to 1, it behaves like $O(1-\lambda)$, which matches the exponent for the same setting in Eq (1.10).

Finally, Garg, Lee, Song and Srivastava (GLSS17) were able to generalize the matrix Chernoff bound due to Ahlswede and Winter (AW02) to Markov chains. In particular it was shown that for a Markov chain $\{Y_s\}_{i=1}^{\infty}$ and function $f : [N] \to \mathbb{R}^{d\times d}$ satisfying $\mathbb{E}[f(Y_1)] = 0$ and $\|f(v)\|_{S_\infty} \le 1$ for all $v \in [N]$,

$$\Pr[\|f(Y_1) + \cdots + f(Y_n)\|_{S_\infty} \ge u\sqrt{n}] \le 2d \exp\left(-c(1-\lambda)u^2\right). \tag{1.11}$$

As far as we are aware, this thesis is the first to investigate anti-concentration inequalities for sums of random variables obtained from a Markov chain.

Overview of Results

In this thesis we obtain the following concentration and anti-concentration results for Markov chains, along with applications.

In Chapter 3, we prove a corollary of the Marcinkiewicz-Zygmund inequality for Markov chains, thus generalizing Eq. (1.4) to Markov Chains. Additionally, we use this bound to obtain moment bounds for sums of random variables from an ℓ_p space, generalizing Eq. (1.8) to Markov chains for such spaces. This chapter is based on work with Assaf Naor and Oded Regev in (NRR17).

In Chapter 4, we prove a Hoeffding inequality for Markov chains, generalizing Eq. (1.5). Additionally, we use this bound to prove tail bounds for certain types of random variables from a general Banach space. As an application, we are able to analyze the Schatten-∞ norm of random matrices whose entries are not necessarily independent. This chapter appears in (Rao19a).

In Chapter 5, we prove a Poisson approximation for sums of random variables obtained from a Markov chain and different functions, thus generalizing Eq. (1.7) to Markov chains. As an application, we are able to construct explicit resilient functions matching the non-constructive functions shown to exist due to Ajtai and Linial. Informally, resilient functions are functions for which any not too large subset of the inputs cannot affect the output of the function when the other inputs are set randomly. This chapter is based on joint work with Oded Regev in (RR17).

Finally, in Chapter 6, we prove a Littlewood-Offord inequality for Markov chains, generalizing Eq. (1.9). However, our bound has an extra factor in the dimension of the vectors v_i. As an application, we are able to obtain new pseudorandom generators for the Littlewood-Offord problem. This chapter appears in (Rao19b).

Chapter 2

Background and Preliminaries

In this chapter we introduce the notation to be used throughout the thesis. We also give background on Markov chains and the special case of random walks on expander graphs which will be used in the applications of the results in this thesis. We also present some claims that are used throughout this thesis to bound the norms of products of matrices. Finally, we give a proof of Gillman's original bound using similar techniques as the results in this thesis.

2.1 Notation

In this thesis we will use the following notation.

Given vectors $v, \mu \in \mathbb{R}^N$ so that μ has positive entries, (typically μ will be a distribution over $[N]$), let

$$\|v\|_{L_p(\mu)}^p = \sum_{i=1}^{N} \mu_i |v_i|^p.$$

We define the inner product for two vectors $u, v \in \mathbb{R}^N$ and $\mu \in \mathbb{R}^N$ with positive entries to be

$$\langle u, v \rangle_{L_2(\mu)} \sum_{i=1}^{N} \mu_i u_i v_i.$$

Additionally, we let the operator norm of a matrix $A \in \mathbb{R}^{N \times N}$ be defined as

$$\|A\|_{L_p(\mu) \to L_q(\mu)} = \max_{v:\|v\|_{L_p(\mu)}=1} \|Av\|_{L_q(\mu)}.$$

We will use ℓ_p in place of $L_p(\mathbf{1})$ where $\mathbf{1}$ is the vector whose entries are all 1.

The Schatten p-norm of a matrix $A \in \mathbb{R}^{N \times N}$ is defined to be

$$\|A\|_{S_p}^p = \sum_{i=1}^{N} s_i^p$$

where $s_1, \dots, s_N$ are the singular values of A.

For a vector v, we let $\mathrm{diag}(v)$ be the diagonal matrix where $\mathrm{diag}(v)_{i,i} = v_i$.

2.2 Markov Chains

A Markov chain on a finite state space $[N]$ is a random sequence, $\{Y_n\}_{n=1}^{\infty}$ whose elements are from the state space $[N]$ in which each Y_n depends only on Y_{n-1}. Additionally, this dependency on the previous element of the sequence is independent of n. Thus, a Markov chain can be completely defined by its transition matrix, $A \in \mathbb{R}^{N \times N}$ where $A_{i,j}$ is the probability that $Y_n = j$, given that $Y_{n-1} = i$, for all n. Such a matrix must be stochastic, that is, the entries on each row must sum to 1. Thus $\mathbf{1}$, the vector whose entries are all 1, is a right eigenvector of A.

In this thesis, we focus only on Markov chains with a stationary distribution $\mu \in \mathbb{R}^N$, where μ is a vector with positive entries that sum to 1, and is a left eigenvector with eigenvalue 1 of the transition matrix A, so that $\mu^t A = \mu^t$. One interpretation of the stationary distribution is that given that Y_n is distributed as μ, it is also the case that Y_{n+1} is distributed as μ, for all n. The stationary distribution is not always unique and some Markov chains have many stationary distributions — for example, the Markov chain whose transition matrix is the identity matrix. This will not affect any of the results in this thesis. For such Markov chains, a stationary distribution can be chosen arbitrarily when needed.

We define the averaging operator $E_\mu \in \mathbb{R}^{N \times N}$ of a Markov chain to be the matrix defined as $(E_\mu)_{i,j} = \mu_j$, so that $E_\mu = \mathbf{1}\mu^t$, where $\mathbf{1}$ is the vector whose entries are 1, and μ is a stationary distribution. Such an operator is also the transition matrix of a Markov chain whose transitions are completely independent, that is, Y_n does not even depend on Y_{n-1}. It is also helpful to note that $E_\mu A = AE_\mu = E_\mu^2 = E_\mu$.

For a Markov chain $\{Y_i\}_{i=1}^\infty$, we define the quantity $\lambda(Y)$ to be

$$\lambda(Y) = \|A - E_\mu\|_{L_2(\mu)}.$$

Informally, this represents how close the transitions of the Markov chain are to being completely independent. Sometime we refer to $1 - \lambda(Y)$ as the spectral gap. For example, if the transition matrix A is E_μ, the transitions are completely independent, $\lambda(Y) = 0$ and the spectral gap is 1. On the other hand, if the transition matrix A is I, the transitions are completely dependent on the previous state, $\lambda(Y) = 1$, and the spectral gap is 0.

We note that in the literature, it is common to define $\lambda(Y)$ as the quantity

$$\lambda(Y) = \max_{v:\langle v,\mathbf{1}\rangle_{L_2(\mu)}=0} \frac{\|vA\|_{L_2(\mu)}}{\|v\|_{L_2(\mu)}}.$$

This definition is equivalent, which can be seen from the fact that $\mu(A - E_\mu) = 0$. Informally, the v that maximizes $\|v(A - E_\mu)\|_{L_2(\mu)}$ is orthogonal to μ.

It is also the case that

$$\lambda(Y) = \|A - (1 - \lambda(Y))E_\mu\|_{L_2(\mu)},$$

as the operator $A - \lambda(Y)E_\mu$ maps the space of vectors orthogonal to the all-1's vectors to itself. This fact will be used in many of the proofs in this thesis.

In many results, it is often assumed that that the Markov chain be reversible, that is, $\mu_i A_{i,j} = \mu_j A_{j,i}$. In other words, running the Markov chain in reverse yields the same distribution of sequences. The assumption of reversibility is natural; as explained in the next section, the common example of random walks on undirected graphs are reversible Markov chains. Practically, the assumption of reversibility implies that the matrix A is symmetric with respect to the space $L_2(\mu)$, and thus has an orthonormal basis of eigenvectors along with real eigenvalues. This further implies that $\lambda(Y)$ is the second largest absolute value of an eigenvalue of A.

The results in this thesis apply to general Markov chains, without requiring much extra effort to handle the case of irreversible Markov chains. However, they will all be stated for reversible Markov chains. Because $\lambda(Y)$ does not necessarily correspond to the absolute value of an eigenvalue, it can be difficult to gain an intuition about the spectral gap of non-reversible Markov chains. Additionally, there are examples of irreversible Markov chains where $\lambda(Y) = 1$ that mix well, and thus it is unclear if $\lambda(Y)$ is an appropriate quantity to consider when analyzing irreversible Markov chains. For a more complete treatment of irreversible Markov chains, see the survey due to to Montenegro and Tetali (MT06)

2.3 Expander Graphs

Many of the applications of the results in this thesis use the existance of expander graphs.

A graph $G = (V, E)$ consists of a set of vertices V and a set of edges $E \subseteq V \times V$. For each vertex $v \in V$, we let its out-degree $d(v)$ be the number of edges e such that $e = (v, u)$ for any vertex u. Given a graph G, we can define a random walk to be a random sequence of vertices $\{Y_i\}_{i=1}^{\infty}$ such that for every n and every edge $(v, u) \in E$, we have that

$$\Pr[Y_n = u \mid Y_{n-1} = v] = \frac{1}{d(v)}.$$

In particular, a random walk on a graph is an example of a Markov chain. Thus, all the results in this thesis can also be applied to random walks on graphs.

We let the normalized adjacency matrix A of a graph G be defined by $A_e = 1/d(e_1)$ for each edge $e \in E$. This is also the transition matrix of a random walk on the graph G. Additionally, we can refer to the spectral gap of a graph as the spectral gap of the Markov chain representing the random walk on the graph.

We say that a graph is undirected if $(u, v) \in E$ implies that $(v, u) \in E$. We say that a graph is d-regular if $d(v) = d$ for all $v \in E$. The stationary distribution of undirected d-regular graphs is the one that is uniform over all vertices. Thus, the spectral gap of random walks on these graphs can be expressed as

$$1 - \|A - J\|_{\ell_2 \to \ell_2}$$

where J is the $|V| \times |V|$ matrix whose entries are all $1/|V|$.

We say that a family of graphs $\mathcal{G}$ is an expander if $\lambda(G)$ is bounded above by a constant less than 1 for every $G \in \mathcal{G}$ or if the spectral gap is bounded below by a constant greater than 0. It is well known that there exist several explicit infinite families of undirected d-regular graphs that are expanders. We state the following result due to Lubotzky, Phillips, and Sarnak.

Theorem 1. *For every d such that $d - 1$ is a prime congruent to $1 \bmod 4$, there exists an infinite family of undirected d-regular graphs of increasing size such that $\lambda(G) \leq 2\sqrt{d-1}/d$.*

These graphs have the optimal dependence on λ and degree d due to the Alon-Boppana theorem, which we state below.

Theorem 2. *For every constant $d \in \mathbb{N}$, any undirected d-regular graph $G = (V, E)$ satisfies $\lambda(G) \geq 2\sqrt{d-1}/d - o(1)$, where the $o(1)$ term vanishes as $|V| \to \infty$ and d is held constant.*

The benefit of expander graphs can be seen in considering the seed-length. For a set V, define a *sampler* from V of length n as a function whose range is V^n. The *seed length* of a sampler is defined as the logarithm of the cardinality of the function's domain, and can be seen as the number of random bits necessary to sample from the function. Consider a sampler that is a list of vertices visited in a random walk of a d-regular graph $G = (V, E)$. It has seed length $\log_2 |V| + (n-1)\log_2 d$, and in particular, if G is a constant degree expander, the seed length of H is $\log_2 |V| + O(n)$. On the other hand, if one samples independently, the seed length is $n \log_2 |V|$.

The fact that expander walks have small seed length while having a large spectral gap has led them to many useful applications in theoretical computer science, among other areas of mathematics. In particular, Gillman's original bound has been used in randomness reduction for randomized algorithms (AKK99; AS99), proving the inapproximability of NP-hard problems (Zuc07), and decoding algorithms for error correcting codes (HOW15). For a more complete treatment of expander graphs and expander walks and their uses in theoretical computer science, we direct the reader to the survey by Hoory, Linial, and Wigderson (HLW06).

In this thesis, we apply some of our new results on concentration and anti-concentration for Markov chains to problems in theoretical computer science, via the use of expander walks. In Chapter 5, we construct explicit resilient functions matching the non-constructive functions shown to exist by Ajtai and Linial (AL93). Informally, resilient functions are functions for which any not too large subset of the inputs cannot affect the output of the function when the other inputs are set randomly. In Chapter 6, we construct an explicit pseudorandom generator for the Littlewood-Offord problem. This is an explicit subset D of $\{-1,1\}^n$ of size $|D| \leq 2^{C_1\sqrt{n}}$, so that sampling uniformly from D gives the same anti-concentration properties up to constant factors for the Littlewood-Offord problem, as sampling uniformly from $\{-1,1\}^n$.

2.4 Preliminary Claims

We present a few preliminary results that will be used throughout the thesis. These will mostly be used to bound expressions of the form $\mathbb{E}[f_{j_1}(Y_{j_1})\cdots f_{j_p}(Y_{j_p})]$ for $j_1,\ldots,j_p \in [n]$.

The following claim shows how the averaging operator can help break up expressions.

Claim 1. *For all $k \geq 1$, matrices $R_1,\ldots,R_k \in \mathbb{R}^{N\times N}$, and distributions μ over $[N]$*

$$\langle \mathbf{1}, R_1 E_\mu R_2 E_\mu \cdots E_\mu R_k \mathbf{1}\rangle_{L_2(\mu)} = \prod_{i=1}^{k} \langle \mathbf{1}, R_i\mathbf{1}\rangle_{L_2(\mu)} \leq \prod_{i=1}^{k} \|R_i\mathbf{1}\|_{L_1(\mu)} .$$

The following is Hölder's inequality.

Lemma 1 (Hölder's inequality)**.** *Let f and g be vectors, and $r,p,q \geq 1$ be such that $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. Then*

$$\|f\circ g\|_{L_r(\mu)} \leq \|f\|_{L_p(\mu)}\|g\|_{L_q(\mu)}$$

where $f\circ g$ is the coordinate-wise product of f and g.

We use Hölder's inequality in the next claim to again break up expressions.

Claim 2. *Let $u_1,\ldots,u_{k+1} \in \mathbb{R}^N$, let $U_i = \operatorname{diag}(u_i)$, and let $T_1,\ldots,T_k \in \mathbb{R}^{N\times N}$. Then for* *any $q > k-1$ and a probability measure μ on $[N]$,*

$$\left\| \left(\prod_{j=1}^{k} U_j T_j \right) U_{k+1} \mathbf{1} \right\|_{L_1(\mu)} \leq$$

$$\|u_1\|_{L_{\frac{2q}{q-k+1}}(\mu)} \|u_{k+1}\|_{L_{\frac{2q}{q-k+1}}(\mu)} \left(\prod_{j=2}^{k} \|u\|_{L_q(\mu)}^{k-1} \right) \left(\prod_{j=1}^{k} \|T_j\|_{L_{\frac{2q}{q+k+1-2j}}(\mu) \to L_{\frac{2q}{q+k+1-2j}}(\mu)} \right) \ .$$

If $u := u_1 = \cdots = u_{k+1}$ and $q \geq k+1$, this is at most

$$\|u\|_{L_q(\mu)}^{k+1} \prod_{j=1}^{k} \|T_j\|_{L_{\frac{2q}{q+k+1-2j}}(\mu) \to L_{\frac{2q}{q+k+1-2j}}(\mu)} \ . \tag{2.1}$$

For $q = \infty$, we obtain the bound

$$\|u_1\|_{L_2(\mu)} \|u_{k+1}\|_{L_2(\mu)} \left(\prod_{j=2}^{k} \|u_j\|_{L_\infty(\mu)}^{k-1} \right) \left(\prod_{j=1}^{k} \|T_j\|_{L_2(\mu) \to L_2(\mu)} \right) \tag{2.2}$$

We note that the reciprocals of the norms appearing in the product above in the main statement. form an arithmetic progression with difference $1/q$ centered around $1/2$.

Proof. By Hölder's inequality and the definition of operator norm,

$$\begin{aligned}
1\left\|\left(\prod_{j=1}^{k} UT_j\right) u\right\|_{L_1(\mu)} &\le \|u\|_{L_{\frac{2q}{q-k+1}}(\mu)} \left\|T_1\left(\prod_{j=2}^{k} UT_j\right) u\right\|_{L_{\frac{2q}{q+k-1}}(\mu)\to L_{\frac{2q}{q+k-1}}(\mu)} \\
&\le \|u\|_{L_{\frac{2q}{q-k+1}}(\mu)} \|T_1\|_{L_{\frac{2q}{q+k-1}}(\mu)\to L_{\frac{2q}{q+k-1}}(\mu)} \left\|\left(\prod_{j=2}^{k} UT_j\right) u\right\|_{L_{\frac{2q}{q+k-1}}(\mu)\to L_{\frac{2q}{q+k-1}}(\mu)} \\
&\le \|u\|_{L_{\frac{2q}{q-k+1}}(\mu)} \|T_1\|_{L_{\frac{2q}{q+k-1}}(\mu)\to L_{\frac{2q}{q+k-1}}(\mu)} \|u\|_{L_q(\mu)} \\
&\qquad \|T_2\|_{L_{\frac{2q}{q+k-3}}(\mu)\to L_{\frac{2q}{q+k-3}}(\mu)} \left\|\left(\prod_{j=3}^{k} UT_j\right) u\right\|_{L_{\frac{2q}{q+k-3}}(\mu)\to L_{\frac{2q}{q+k-3}}(\mu)} \\
&\le \|u\|^2_{L_{\frac{2q}{q-k+1}}(\mu)} \|u\|^{k-1}_{L_q(\mu)} \left(\prod_{j=1}^{k} \|T_j\|_{L_{\frac{2q}{q+k+1-2j}}(\mu)\to L_{\frac{2q}{q+k+1-2j}}(\mu)}\right),
\end{aligned}$$

as desired. □

The next claim helps rearrange products involving the averaging operator and diagonal matrices.

Claim 3. *Let $\mu \in \mathbb{R}^N$ be a distribution, and let E_μ be the associated averaging operator. Then for any $u \in \mathbb{C}^N$,*

$$E_\mu \operatorname{diag}(u) E_\mu = \langle u, \mu\rangle_{L_2(\mu)} E_\mu.$$

Proof.

$$(E_\mu \operatorname{diag}(u) E_\mu)_{i,j} = \sum_{k\in[N]} (E_\mu)_{i,k} u_k (E_\mu)_{k,j} = \langle u, \mu\rangle_{L_2(\mu)} \mu_j.$$

□

Finally, we combine all of the above claims in the following Lemma 2, that will be used throughout this thesis.

Lemma 2 uses notation which we define below. Let S_k be the set of $\{0,1\}^k$ of vectors s

with no two consecutive 0s and so that $s_k = 1$. For a given vector $s \in S_k$ and index j for which $s_j = 1$, we define $p(s, j) \geq 1$ in the following way. Consider the consecutive run of 1s in which j is located, and let i_1 and i_2 be the first and last indices of this run. (Formally, $i_1 = \max\{j < i : s_j = 0\} + 1$ or 1 if no such j exists, and $i_2 = \min\{j > i : s_j = 0\} - 1$ or k if no such j exists.) Then we define $p(s, j) := 2q/(q + i_2 + i_1 - 2j)$. In other words, for each consecutive run of 1s, we assign norms whose reciprocals form an arithmetic progression with difference $1/q$ centered around $1/2$, as in Claim 2.

Lemma 2. *Let $k \geq 1$ be an integer and $q \geq k + 1$. Let $S_k \subseteq \{0, 1\}^k$ be as above. Let μ be a probability measure on $[N]$, $u_1, \ldots, u_{k+1} \in \mathbb{R}^N$ be an N-dimensional vectors, let $U_i = \operatorname{diag}(u_i)$ for all i, and let $T_1, \ldots, T_k \in \mathbb{R}^{N \times N}$. We have the following upper bounds on the quantity*

$$\left| \langle \mathbf{1}, U_1(T_1 + E_\mu)U_2(T_2 + E_\mu)U_3 \cdots U_k(T_k + E_\mu)U_{k+1}\mathbf{1}\rangle_{L_2(\mu)} \right| \tag{2.3}$$

- *If $u = u_1 = \cdots = u_{k+1}$ and $\langle u, \mu\rangle_{L_2(\mu)} = 0$, then we obtain an upper bound of*

$$\|u\|_{L_q(\mu)}^{k+1} \sum_{s \in S_k} \prod_{j : s_j = 1} \|T_j\|_{L_{p(s,j)}(\mu) \to L_{p(s,j)}(\mu)} \ . \tag{2.4}$$

- *If $\langle u_j, \mu\rangle_{L_2(\mu)} = 0$ for all j, then we obtain an upper bound of*

$$\|u_1\|_{L_\infty(\mu)} \|u_2\|_{L_\infty(\mu)} \cdots \|u_{k+1}\|_{L_\infty(\mu)} \sum_{s \in S_k} \prod_{j : s_j = 1} \|T_j\|_{L_2(\mu) \to L_2(\mu)} \ . \tag{2.5}$$

- *Let $\omega_j = \langle u_j, \mu\rangle_{L_2(\mu)}$. If the coordinates of u_j are in the range $[0, 1]$ for all j, then we obtain an upper bound of*

$$\sum_{s \in \{0,1\}^k} \left(\prod_{j : s_j = 0} \sqrt{\omega_j \omega_{j+1}} \right) \left(\prod_{j : s_j = 1} \|T_j\|_{L_2(\mu) \to L_2(\mu)} \right) \ . \tag{2.6}$$

Finally, for $s \in \{0,1\}^k$, let $\overline{s} := (0, s, 0) \in \{0,1\}^{k+2}$ and $t(s) := \{j \ : \ \overline{s}_j = \overline{s}_{j-1} = 0\}$. Then if $\|u_i\|_{L_\infty(\mu)} \leq 1$ for all i,

$$\left\|U_1(T_1 + (1-\lambda)E_\mu)U_2(T_2 + (1-\lambda)E_\mu)U_3 \cdots U_k(T_k + (1-\lambda)E_\mu)U_{k+1}\mathbf{1}\right\|_{L_1(\mu)} \leq \sum_{s \in \{0,1\}^k} \left(\prod_{j : s_j = 1} \|T_j\|_{L_2(\mu) \to L_2(\mu)} \right) \left(\prod_{j : s_j = 0} (1-\lambda) \right) \left(\prod_{j \in t(s)} \left| \langle u_j, \mu \rangle_{L_2(\mu)} \right| \right) . \tag{2.7}$$

Proof. We start with the first three statements. For $j = 1, \ldots, k$, let $T_{j,0} = E_\mu$ and $T_{j,1} = T_j$. Then using the triangle inequality, the left-hand side of (2.3) is at most

$$\sum_{s \in \{0,1\}^k} \left\| \left(\prod_{j=1}^k UT_{j,s_j} \right) U\mathbf{1} \right\|_{L_1(\mu)} . \tag{2.8}$$

For the first two statements in particular, we have that this is equal to

$$\sum_{s \in S_k} \left\| \left(\prod_{j=1}^k UT_{j,s_j} \right) U\mathbf{1} \right\|_{L_1(\mu)} , \tag{2.9}$$

since the terms corresponding to vectors s with two consecutive zeros or with $s_k = 0$ are equal to 0 because in these cases the term $E_\mu U E_\mu = 0$ (or $E_\mu U\mathbf{1} = 0$) appears.

Fix an $s \in S_k$ for the first and second bound, or an $s \in \{0,1\}^k$ for the third, and let $1 \leq r_1 < r_2 < \cdots < r_\ell < k$ be the indices of s that are 0. By Claim 1, the term corresponding to s in Eq. (2.8) or Eq. (5.4) is at most

$$\|UT_1UT_2 \cdots T_{r_1 - 1}U\mathbf{1}\|_{L_1(\mu)} \cdot \|UT_{r_1+1}UT_{r_1+2} \cdots T_{r_2-1}U\mathbf{1}\|_{L_1(\mu)} \cdots \|UT_{r_\ell+1}UT_{r_\ell+2} \cdots T_kU\mathbf{1}\|_{L_1(\mu)} .$$

The lemma for the first three cases now follows by applying Claim 2. In particular, Eq. (2.1) from Claim 2 gives us the first bound, and Eq. (2.2) gives us the second and third. For the second bound, we note that because μ is a probability distribution, $\|u_j\|_{L_2(\mu)} \leq \|u_j\|_{L_\infty(\mu)}$

for all j. For the third bound, we note that $\|u_j\|_{L_\infty(\mu)} \leq 1$ for all j, and that $\|u_j\|^2_{L_2(\mu)} \leq \|u_j\|_{L_1(\mu)} = \omega_j$, as the entries of u_j are in the range $[0,1]$.

For the final bound, let $U'_i = U_i$ if $i \in t(s)$, and $U'_i = I$ otherwise. Again, let $T_{j,0} = E_\mu$ and $T_{j,1} = T_j$. Then the left-hand side of Eq. (5.4) is at most

$$\sum_{s\in\{0,1\}^k} \left\| \left(\prod_{j=1}^{k} U_j T_{j,s_j} \right) U_{k+1}\mathbf{1} \right\|_{L_1(\mu)} =$$

$$\sum_{s\in\{0,1\}^k} \left\| \left(\prod_{j}^{k} U'_j T_{j,s_j} \right) U'_{k+1}\mathbf{1} \right\|_{L_1(\mu)} \left(\prod_{j:s_j=0} (1-\lambda) \right) \left(\prod_{j\in t(s)} \left| \langle \mu, u\rangle_{L_2(\mu)} \right| \right) \quad (2.10)$$

where the equality follows from Claim 3 and the fact that $E^2_\mu = E_\mu$. We finish by fixing an $s \in \{0,1\}^k$ and proceeding as in the proofs of the second bound, Eq. (2.5). Note that $\|u_j\|_{L_\infty(\mu)} \leq 1$ for all j □

2.5 The Basic Chernoff Bound for Markov Chains

In this section, we present a proof of the Chernoff bound for Markov Chains as originally proved in (Gil98). We state our version of the bound below, which is has different constant factors. Our proof uses the preliminary claims developed in Section 2.4

Theorem 3. *Let $\{Y_i\}_{i=1}^\infty$ be a stationary reversible Markov chain with state space $[N]$ and stationary probability measure μ, and so that Y_1 is distributed according to μ. Let $f_1, \ldots, f_n : [N] \to [-1,1]$ be such that $\mathbb{E}[f_i(Y_i)] = 0$ for all i. Then,*

$$\Pr[f_1(Y_1) + \cdots + f_n(Y_n) \geq u\sqrt{n}] \leq 2\exp(-u^2/(32e(1-\lambda)))$$

We start by bounding $\mathbb{E}[|f(Y_1) + \cdots + f(Y_n)|^q]$ by a sum of $L_1(\mu)$-norms of vectors that can expressed as a product of the transition matrix of the Markov chain as well as matrices defined by the function f. We note that Lemma 3 will also be used in Chapter 3 in the corollary of the Marcinkiewicz-Zygmund inequality for Markov chains.

Lemma 3. *Let $\{Y_i\}_{i=1}^{\infty}$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A and stationary probability measure μ, and so that Y_1 is distributed according to μ, and let $f_1, \ldots, f_n : [N] \to \mathbb{R}$ Then for any even integer $q \geq 2$ and any integer $n \geq 1$,*

$$\mathbb{E}[|f_1(Y_1) + \cdots + f_n(Y_n)|^q] \leq q! \sum_{\substack{v_0,\ldots,v_{q-1} \geq 0: \\ v_0 + \cdots + v_{q-1} \leq n-1}} \|U_1 A^{v_1} U_2 A^{v_2} \cdots U_q A^{v_{q-1}} U\mathbf{1}\|_{L_1(\mu)},$$

where u_i is the vector such that $u_i(v) = f_i(v)$ and $U_i = \operatorname{diag}(u_i)$.

Proof. Because q is even, the left-hand side can be rewritten as

$$\mathbb{E}\left[\sum_{w \in [n]^q} \prod_{i=1}^{q} f(W_{w_i})\right]. \tag{2.11}$$

Let V_q be the set of vectors in $w \in [n]^q$ so that $1 \leq w_1 \leq w_2 \leq \cdots \leq w_q \leq n$. Then Eq. (2.11) is bounded above by

$$q! \sum_{w \in V_q} \mathbb{E}\left[\prod_{i=1}^{q} f(W_{w_i})\right].$$

The lemma follows by noting that

$$\mathbb{E}\left[\prod_{i=1}^{q} f(W_{w_i})\right] \leq \|UA^{w_1 - w_2} UA^{w_3 - w_2} \cdots UA^{w_q - w_{q-1}} U\mathbf{1}\|_{L_1(\mu)}$$

and that the map sending any $w \in V_q$ to the vector $(w_1 - 1, w_2 - w_1, \ldots, w_q - w_{q-1})$ (whose coordinates are non-negative and sum to at most $n-1$) is injective. □

We proceed by applying Lemma 3 to each term in the sum to obtain a bound on the qth moment.

Lemma 4. *Let $\{Y_i\}_{i=1}^{\infty}$ be a stationary reversible Markov chain with state space $[N]$ and stationary probability measure μ, and so that Y_1 is distributed according to μ. Let $f_1, \ldots, f_n : [N] \to [-1, 1]$ be such that $\mathbb{E}[f_i(Y_i)] = 0$ for all i. Then for any integer $n \geq 1$ and any even $2 \leq q$,*

$$\mathbb{E}[|f_1(Y_1) + \cdots + f_n(Y_n)|^q] \leq 4^q q^{q/2} n^{q/2} \left(\frac{1}{1-\lambda}\right)^{q/2}.$$

Proof. When $q > n$, we have

$$\mathbb{E}[(f(W_1) + \cdots + f(W_n))^q] \leq n^q \leq n^{q/2} q^{q/2}.$$

By Lemma 3, we can bound the left-hand side from above by

$$q! \sum_{\substack{v_0,\ldots,v_{q-1}\geq 0: \\ v_0+\cdots+v_{q-1}\leq n-1}} \|U_1 A^{v_1} U_2 A^{v_2} \cdots U_q A^{v_{q-1}} U\mathbf{1}\|_{L_1(\mu)},$$

where A is the transition matrix of Y and u_i is the vector such that $u_i(v) = f_i(v)$ and $U_i = \operatorname{diag}(u_i)$. For each j, let $T_j = A^{v_j} - E_\mu$. By Eq. (2.5) of Lemma 2, this is bounded above by

$$q! \sum_{\substack{v_0,\ldots,v_{q-1}\geq 0: \\ v_0+\cdots+v_{q-1}\leq n-1}} \sum_{s\in S_{q-1}} \prod_{j:s_j=1} \lambda^{v_j} \tag{2.12}$$

using the fact that

$$\|A^{v_j} - E_\mu\|_{L_2(\mu)\to L_2(\mu)} = \|(A - E_\mu)^{v_j}\|_{L_2(\mu)\to L_2(\mu)} \leq (\|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)})^{v_j}.$$

Let $\lambda_0 = (1 - q/(2n))\lambda$, and notice that since $(1 - q/(2n))^n \geq 2^{-q}$, Eq. (2.12) is at most

$$q! 2^q \sum_{\substack{v_0,\ldots,v_{q-1}\geq 0: \\ v_0+\cdots+v_{q-1}\leq n-1}} \sum_{s\in S_{q-1}} \prod_{j:s_j=1} \lambda_0^{v_j} \tag{2.13}$$

Fix some $s \in S_{q-1}$, and let q_0 (respectively q_1) be the number of coordinates of s that are 0 (respectively 1). Notice that $q_0 + q_1 = q - 1$ and that by definition of S_{q-1}, $q_1 \geq q/2$. Consider the $q_0 + 1$ variables v_j for which either $j = 0$ or $s_j = 0$. There are

$$\binom{q_0 + n}{q_0 + 1} \leq \frac{(2n)^{q_0+1}}{(q_0 + 1)!}$$

ways to set them so that their sum is at most $n - 1$. Therefore, the term corresponding to s in (3.5) is at most

$$\begin{aligned} q!2^q \frac{(2n)^{q_0+1}}{(q_0+1)!} \prod_{j:s_j=1} \sum_{i=0}^{\infty} \lambda_0^i &= q!2^q \frac{(2n)^{q_0+1}}{(q_0+1)!} \left(\frac{1}{1-\lambda_0}\right)^{q_1} \\ &\leq q!2^q \frac{(2n)^{q/2}}{q^{q_1-q/2}(q_0+1)!} \left(\frac{1}{1-\lambda_0}\right)^{q/2} \\ &\leq 2^q q^{q/2} (2n)^{q/2} \left(\frac{1}{1-\lambda}\right)^{q/2}. \end{aligned}$$

The first inequality above follows by noting that $q_1 \geq q/2$ and that $\lambda_0 \leq 1 - q/(2n)$ and thus $1/(1 - \lambda_0) \leq 2n/q$. The second follows by noting that because $q_1 \geq q/2$, and $q_0 + q_1 = q - 1$, we have $q^{q_1-q/2}(q_0 + 1)! \geq (q/2)!$, and that $(1 - \lambda_0)^{-1} \leq (1 - \lambda)^{-1}$. □

We prove Theorem 3 by using Markov's inequality.

Proof of Theorem 3. If $u^2/(16e(1 - \lambda)) < 2$, the right-hand side is greater than 1 and the theorem holds trivially.

Otherwise, by Markov's inequality,

$$\begin{aligned} \Pr[f_1(Y_1) + \cdots + f_n(Y_n) \geq u\sqrt{n}] &= \Pr[(f_1(Y_1) + \cdots + f_n(Y_n))^q \geq u^q n^{q/2}] \\ &\leq \frac{\mathbb{E}[(f_1(Y_1) + \cdots + f_n(Y_n))^q]}{u^q n^{q/2}}. \end{aligned}$$

By Lemma 4 and letting q be an even integer in the range $[u^2/(32e(1 - \lambda)), u^2/(16e(1 - \lambda))]$

we obtain the upper bound of $\exp(-q/2) \leq \exp(-u^2/(32e(1-\lambda))$. The Theorem follows by repeating the above for the left tail. $\square$

Chapter 3

A Marcinkiewicz-Zygmund Inequality for Markov Chains

Let $X_1, \ldots, X_n$ be independent and identically distributed real-valued random variables such that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[|X_i|^q] \leq 1$ for some $q \geq 1$. Then it is well known that

$$\mathbb{E}[|X_1 + \cdots + X_n|^q]^{1/q} \lesssim \sqrt{qn}, \tag{3.1}$$

for example by the Marcinkiewicz-Zygmund inequality (MZ37) applied in the case that all variables are identically distributed. After applying Markov's inequality, Eq. (3.1) allows one to obtain sharper tail bounds than what is given by Chebyshev's inequality, without requiring a bound on the moment generating function $\mathbb{E}[\exp(tX_1)]$, which is necessary to apply Chernoff bounds.

We obtain the following analogue of Eq. (3.1).

Theorem 4. *Let $\{Y_i\}_{i=1}^{\infty}$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A and stationary probability measure μ, so that Y_1 is distributed according to μ. Let $\lambda = \|A - E_\mu\|_{\sigma \to L_2(\mu)}$. Let $f : [N] \to \mathbb{R}$ and $q \geq 2$ so that $\mathbb{E}[f(Y_1)] = 0$ and*

$\mathbb{E}[|f(Y_1)|^q] \leq 1$. *Then for any integer* $n \geq 1$,

$$\mathbb{E}[|f(Y_1) + \cdots + f(Y_n)|^q]^{1/q} \lesssim \sqrt{\frac{qn}{1-\lambda}}.$$

By a simple application of Markov's inequality, one can obtain the following tail bound. See Section 3.4 for the proof.

Corollary 1. *Assume the setting of Theorem 4. Then there exists a universal constant* C *such that for any* $u \geq 0$,

$$\Pr[|f(Y_1) + \cdots + f(Y_n)| \geq u\sqrt{n}] \leq \left(\frac{Cq}{(1-\lambda)u^2}\right)^{q/2}.$$

In particular, if $q \geq u^2(1-\lambda)/(Ce)$, *then*

$$\Pr[|f(Y_1) + \cdots + f(Y_n)| \geq u\sqrt{n}] \leq \exp\left(2 - \frac{u^2(1-\lambda)}{2Ce}\right).$$

Vector-valued random variables: Using Theorem 4 we also obtain moment bounds when the function f is vector–valued, in particular, when f takes on values in ℓ_p for $p \geq 2$. In the case of *independent* random variables, the following moment bound follows from (Nao12, Theorem 2.1). In particular, if $X_1, \ldots, X_n$ are independent and identically distributed random variables such that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[\|X_i\|_{\ell_p}^q] \leq 1$, then

$$\mathbb{E}[\|X_1 + \cdots + X_n\|_{\ell_p}^q]^{1/q} \lesssim \sqrt{\max\{p,q\}n}. \tag{3.2}$$

When the random variables are not independent, we obtain the following moment bound.

Theorem 5. *Let* $\{Y_i\}_{i=1}^{\infty}$ *be a stationary reversible Markov chain with state space* $[N]$, *transition matrix* A *and stationary probability measure* μ, *so that* Y_1 *is distributed according to* μ. *Let* $\lambda = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$. *Let* $f : [N] \to \mathbb{R}^m$ *and* $p, q \geq 2$ *so that* $\mathbb{E}[f(Y_1)] = 0$ *and*

$\mathbb{E}[\|f(Y_1)\|_{\ell_p}^q] \leq 1$. *Then,*

$$\mathbb{E}[\|f(Y_1) + \cdots + f(Y_n)\|_{\ell_p}^q]^{1/q} \lesssim \sqrt{\frac{\max\{p,q\}n}{1-\lambda}}.$$

As in the scalar case, a simple application of Markov's inequality yields the following tail bound.

Corollary 2. *Assume the setting of Theorem 5. Then there exists a universal constant C such that for any $u \geq 0$,*

$$\Pr[\|f(Y_1) + \cdots + f(Y_n)\|_{\ell_p} \geq u\sqrt{n}] \leq \left(\frac{C\max\{p,q\}}{(1-\lambda)u^2}\right)^{q/2}.$$

In particular, if $q \geq u^2(1-\lambda)/(Ce)$, then

$$\Pr[\|f(Y_1) + \cdots + f(Y_n)\|_{\ell_p} \geq u\sqrt{n}] \leq \exp\left(p - \frac{u^2(1-\lambda)}{2Ce}\right).$$

We remark that the dependence on $\max\{p,q\}$ in Eq. (3.2) (and similarly Theorem 5), rather than just q as in the scalar case, is in fact necessary. To see this, consider the following construction for the case p even, $n = p$ and $q = 1$. Let $v_1, \ldots, v_p \in \mathbb{R}^{2^p}$ be random vectors such that $(v_i)_j = -1^{s_{i,j}} \cdot (\varepsilon_i/2)$ where $s_{i,j}$ is the ith digit in the binary representation of j, and ε_i are independent Rademacher random variables. Then it always holds that $\|v_i\|_{\ell_p} = 1$ for all i. Additionally,

$$\mathbb{E}[\|v_1 + \cdots + v_p\|_{\ell_p}] \geq \mathbb{E}\left[\max_j (v_1 + \cdots + v_p)_j\right] = \frac{p}{2}.$$

3.1 Preliminaries

The following preliminaries are specific to this chapter.

When $v \in (\mathbb{R}^d)^N$ and $\| \cdot \|$ is a norm on $\mathbb{R}^d$, we let

$$\|v\|_{L_p(\mu;\|\cdot\|)}^p = \sum_{i=1}^{N} \mu_i \|v_i\|^p.$$

As stated in the introduction, our bounds are in term of the quantity $\lambda = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$. We will use the Riesz-Thorin interpolation theorem (Rie27; Tho48) to bound $\|A - E_\mu\|_{L_p(\mu)\to L_p(\mu)}$ in terms of λ.

Theorem 6 (Riesz-Thorin interpolation theorem). *Let $1 \le p_0 \le p_1 \le \infty$ and $1 \le q_0 \le q_1 \le \infty$. Let μ and μ' be any probability measures and let T be a linear operator such that $\|T\|_{L_{p_0}(\mu)\to L_{q_0}(\mu')}$ and $\|T\|_{L_{p_1}(\mu)\to L_{q_1}(\mu')}$ are finite. Let $1/p_\theta = (1-\theta)/p_0 + \theta/p_1$ and $1/q_\theta = (1-\theta)/q_0 + \theta/q_1$. Then*

$$\|T\|_{L_{p_\theta}(\mu)\to L_{q_\theta}(\mu')} \le \|T\|_{L_{p_0}(\mu)\to L_{q_0}(\mu')}^{1-\theta} \|T\|_{L_{p_1}(\mu)\to L_{q_1}(\mu')}^{\theta}.$$

We use the Riesz-Thorin interpolation theorem to prove the following bound on $\|A - E_\mu\|_{L_p(\mu)\to L_p(\mu)}$ in terms of $\lambda(A - E_\mu) = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$.

Claim 4. *Let A be a stochastic matrix, and let E_μ be the averaging operator on its stationary distribution. Then for all $1 \le p \le \infty$,*

$$\|A - E_\mu\|_{L_p(\mu)\to L_p(\mu)} \le 2\|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}^{1-|1-2/p|}.$$

Proof. Because A is a stochastic matrix, $\|A\|_{L_1(\mu)\to L_1(\mu)}$ and $\|A\|_{L_\infty(\mu)\to L_\infty(\mu)}$ are bounded above by 1, and the same is true for E_μ. Therefore, $\|A - E_\mu\|_{L_1(\mu)\to L_1(\mu)}, \|A - E_\mu\|_{L_\infty(\mu)\to L_\infty(\mu)} \le 2$. When $p \le 2$, by Theorem 6

$$\|A - E_\mu\|_{L_p(\mu)\to L_p(\mu)} \le 2\|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}^{2(1-1/p)},$$

and similarly when $p > 2$,

$$\|A - E_\mu\|_{L_p(\mu)\to L_p(\mu)} \leq 2\|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}^{2/p}$$

as desired. □

3.2 Calculating moment bounds in the scalar case

We start proving Theorem 4 by considering even moments. Thus, we can use combining Lemmas 3 from Chapter 2 used in the proof of Gillman's original bound, and apply Eq. (2.4) from 2, also from Chapter 2. Finally, we can use Claim 4 to bound the norms of the matrices that appear in the corresponding expression. In particular, we point out the left-hand side of the inequality in Lemma 5 below appears in the right-hand side of the inequality in Lemma 3.

Lemma 5. *Let $A \in \mathbb{R}^{N\times N}$ be a stochastic matrix, and let $\lambda = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$ for some probability measure μ over $[N]$. Let u be a vector such that $\sum_i \mu_i u_i = 0$ and let $U = \mathrm{diag}(u)$. Then for all $n \geq 1$ and all even $q \geq 2$,*

$$\sum_{\substack{v_0,\ldots,v_{q-1}\geq 0:\\ v_0+\cdots+v_{q-1}\leq n-1}} \|UA^{v_1}UA^{v_2}\cdots UA^{v_{q-1}}U\mathbf{1}\|_{L_1(\mu)} \leq C^q(1-\lambda)^{-q/2}\|u\|_{L_q(\mu)}^q \frac{n^{q/2}}{(q/2)!} \tag{3.3}$$

for some universal constant C.

Proof. We start by applying Eq. (2.4) from Lemma 2 in Chapter 2 with $k = q-1$ and $T_j = A^{v_j} - E_\mu$. Because $AE_\mu = E_\mu A = E_\mu$ we have that $A^{v_j} - E_\mu = (A - E_\mu)^{v_j}$ and thus $\|T_j\|_{L_2(\mu)\to L_2(\mu)} = \lambda^{v_j}$. By Claim 4, $\|T_j\|_{L_{p(s,j)}(\mu)\to L_{p(s,j)}(\mu)} \leq 2\lambda^{v_j(1-|1-2/p(s,j)|)}$, and thus we can bound the left-hand side of (3.3) from above by

$$\|u\|_{L_q(\mu)}^q \sum_{s\in S_{q-1}} \sum_{\substack{v_0,\dots,v_{q-1}\geq 0:\\ v_0+\dots+v_{q-1}\leq n-1}} \prod_{j:s_j=1} 2\lambda^{v_j(1-|1-2/p(s,j)|)}. \tag{3.4}$$

Let $\lambda_0 = (1-q/(2n))\lambda$, and notice that since $(1-q/(2n))^n \geq 2^{-q}$, Eq. (3.4) is at most

$$\begin{aligned} 2^q\|u\|_{L_q(\mu)}^q \sum_{s\in S_{q-1}} \sum_{\substack{v_0,\dots,v_{q-1}\geq 0:\\ v_0+\dots+v_{q-1}\leq n-1}} \prod_{j:s_j=1} 2\lambda_0^{v_j(1-|1-2/p(s,j)|)} \leq \\ 4^q\|u\|_{L_q(\mu)}^q \sum_{s\in S_{q-1}} \sum_{\substack{v_0,\dots,v_{q-1}\geq 0:\\ v_0+\dots+v_{q-1}\leq n-1}} \prod_{j:s_j=1} \lambda_0^{v_j(1-|1-2/p(s,j)|)}\ . \end{aligned} \tag{3.5}$$

Fix some $s \in S_{q-1}$, and let q_0 (respectively q_1) be the number of coordinates of s that are 0 (respectively 1). Notice that $q_0 + q_1 = q-1$ and that by definition of S_{q-1}, $q_1 \geq q/2$. Consider the q_0+1 variables v_j for which either $j=0$ or $s_j=0$. There are

$$\binom{q_0+n}{q_0+1} \leq \frac{(2n)^{q_0+1}}{(q_0+1)!}$$

ways to set them so that their sum is at most $n-1$. Therefore, the term corresponding to s in (3.5) is at most

$$\frac{(2n)^{q_0+1}}{(q_0+1)!} \prod_{j:s_j=1} \sum_{i=0}^{\infty} \lambda_0^{i(1-|1-2/p(s,j)|)}\ . \tag{3.6}$$

The sum in (3.6) is bounded above by

$$\sum_{i=0}^{\infty} \lambda_0^{i(1-|1-2/p(s,j)|)} = \frac{1}{1-\lambda_0^{1-|1-2/p(s,j)|}} \leq \frac{1}{(1-|1-2/p(s,j)|)(1-\lambda_0)}\ ,$$

where we used the convexity of λ_0^α in α. Therefore, (3.6) is at most

$$\frac{(2n)^{q_0+1}}{(q_0+1)!}\Big(\frac{1}{1-\lambda_0}\Big)^{q_1}\prod_{j:s_j=1}(1-|1-2/p(s,j)|)^{-1}\,. \tag{3.7}$$

Recalling the definition of $p(s,j)$, notice that the contribution to the product in (3.7) coming from a run of consecutive 1s of length r in s is

$$\prod_{i=1}^{r}\Big(1-\Big|\frac{r+1-2i}{q}\Big|\Big)^{-1}=\prod_{i=1}^{\lfloor r/2\rfloor}\Big(\frac{q}{q-r-1+2i}\Big)^2\,.$$

Assuming for simplicity that r is even (a similar calculation holds for odd r), this is

$$\begin{aligned}2^{-r}q^r\cdot\Big(\Big(\frac{q-r-1}{2}\Big)!\Big/\Big(\frac{q-1}{2}\Big)!\Big)^2 &\le 2^{-r}q^r\cdot e^2\cdot\Big(\frac{q-r-1}{2e}\Big)^{q-r-1}\Big(\frac{q-1}{2e}\Big)^{-(q-1)}\\ &\le 2^{-r}q^r\cdot e^2\cdot\Big(\frac{q-1}{2e}\Big)^{-r}\\ &\le 2^{-r}\cdot e\cdot e^2\cdot(2e)^r\\ &\le e^{4r}\,,\end{aligned}$$

where we used Stirling's approximation, $\sqrt{2\mu}\le n!/((n/e)^n n^{1/2})\le\sqrt{2\mu}e$. Summarizing, the product in (3.7) (over all j such that $s_j=1$) is at most $e^{4q_1}\le e^{4q}$, and therefore (3.7) is at most

$$e^{4q}\cdot\frac{(2n)^{q_0+1}}{(q_0+1)!}\Big(\frac{1}{1-\lambda_0}\Big)^{q_1}\le e^{4q}\cdot\frac{(2n)^{q/2}}{q^{q_1-q/2}(q_0+1)!}\Big(\frac{1}{1-\lambda_0}\Big)^{q/2}\le e^{4q}\cdot\frac{(2n)^{q/2}}{(q/2)!}\Big(\frac{1}{1-\lambda}\Big)^{q/2} \tag{3.8}$$

The first inequality in Eq. (3.8) follows by noting that $q_1\ge q/2$ and that $\lambda_0\le 1-q/(2n)$ and thus $1/(1-\lambda_0)\le 2n/q$. The second follows by noting that because $q_1\ge q/2$, and $q_0+q_1=q-1$, we have $q^{q_1-q/2}(q_0+1)!\ge(q/2)!$, and that $(1-\lambda_0)^{-1}\le(1-\lambda)^{-1}$.

We complete the proof by plugging this back into (3.4) and noting that there are at most

2^{q-1} elements in S_{q-1}. □

Finally, we obtain the moment bound described in the introduction.

Proof of Theorem 4. If $q > n$, we have

$$\begin{aligned}\mathbb{E}[|f(Y_1) + \cdots + f(Y_n)|^q]^{1/q} &= n \cdot \mathbb{E}\left[\left|\frac{f(Y_1) + \cdots + f(Y_n)}{n}\right|^q\right]^{1/q} \\ &\leq n \cdot \mathbb{E}\left[\frac{|f(Y_1)|^q + \cdots + |f(Y_n)|^q}{n}\right]^{1/q} \\ &\leq n \\ &\leq \sqrt{qn},\end{aligned}$$

as desired, where the first inequality follows from Jensen's inequality.

When $2 \leq q \leq n$ is an even integer we combine Lemma 3 from Chapter 2 and Lemma 5 to obtain the following inequality.

$$\mathbb{E}[|f(Y_1) + \cdots + f(Y_n)|^q] \leq C^q(1-\lambda)^{-q/2}\|u\|^q_{L_q(\mu)}\frac{q!n^{q/2}}{(q/2)!}.$$

The theorem follows in this case by noting that $q!/(q/2)! \leq q^{q/2}$.

We now treat the general case when q is not an even integer. Let p be the largest even integer smaller than q. Let T_n be the linear transformation from functions on $[N]$ to functions on $[N]^n$ defined as

$$T_n f(W) = \sum_{i=1}^{n} f(Y_i)$$

for $W \in [N]^n$. Consider the operator $I_N - E_\mu$, where I_N is the identity operator. First, by Claim 4 (or directly by the triangle inequality), $\|I_N - E_\mu\|_{L_p(\mu)\to L_p(\mu)} \leq 2$. Moreover, for all functions f on $[N]$, $\mathbb{E}[(I_N - E_\mu)f(W)] = 0$, i.e, the image under $I_N - E_\mu$ of any function f is a centered function. Therefore, by the even integer case of the theorem, we have that

$$\|T_n \circ (I_N - E_\mu)\|_{L_p(\mu)\to L_p(\mu_n)} \lesssim \sqrt{\frac{pn}{1-\lambda}}\,, \tag{3.9}$$

where μ_n is the law of $(Y_1, \ldots, Y_n)$. Similarly, we can replace p with $p+2$ in Eq. (3.9). By Theorem 6, we obtain that there exists some $0 \le \theta \le 1$ such that

$$\begin{aligned}\|T_n \circ (I_N - E_\mu)\|_{L_q(\mu)\to L_q(\mu_n)} &\le \|T_n \circ (I_N - E_\mu)\|^{1-\theta}_{L_p(\mu)\to L_p(\mu_n)} \|T_n \circ (I_N - E_\mu)\|^{\theta}_{L_{p+2}(\mu)\to L_{p+2}(\mu_n)} \\ &\lesssim \sqrt{\frac{(p+2)n}{1-\lambda}} \\ &\lesssim \sqrt{\frac{qn}{1-\lambda}},\end{aligned}$$

which implies the desired conclusion of the theorem. □

3.3 Calculating moment bounds in the ℓ_p case

In this section we prove Theorem 5. We consider functions of the form $f : [N] \to \mathbb{R}^m$ with the property that $\mathbb{E}[f(Y_1)] = 0$. Then we will bound $\mathbb{E}[\|S_n\|^q_{\ell_p}]^{1/q}$ where $S_n = f(Y_1) + \cdots + f(Y_n)$. The proof proceeds in three steps. In the first step, we handle the case $p = q$, which follows almost immediately from Theorem 4. We then derive from this the case $p = 2$ by using Dvoretzky's theorem (Dvo61), which provides an embedding of ℓ_2 into ℓ_q. Finally, we derive the case $2 \le p \le q$ by interpolation, and the case $p > q$ follows from the case $p = q$.

3.3.1 A bound on the qth moment for ℓ_q

Bounding the qth moment for ℓ_q only requires a bound on the sum of qth moments of many scalar random variables, which we already have from Theorem 4. We carry out these computations in the following lemma.

Lemma 6. *Let $\{Y_s\}_{s=1}^{\infty}$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A and stationary probability measure $r(s) \leq$, so that Y_1 is distributed according to μ. Let $\lambda = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$ and let $f : [N] \to \mathbb{R}^m$ so that $\mathbb{E}[f(Y_1)] = 0$ and $\mathbb{E}[\|f(Y_1)\|_{\ell_q}^q] \leq 1$. Let $S_n = f(Y_1) + \cdots + f(Y_n)$. Then for $q \geq 2$,*

$$\mathbb{E}[\|S_n\|_{\ell_q}^q]^{1/q} \lesssim \sqrt{\frac{qn}{1-\lambda}}.$$

Proof. Note that

$$\mathbb{E}[\|S_n\|_{\ell_q}^q] = \mathbb{E}\left[|(S_n)_1|^q + |(S_n)_2|^q + \cdots + |(S_n)_m|^q\right].$$

Applying Theorem 4 to each $|(S_n)_i|^q$, we find that this is bounded above by

$$(Cqn/(1-\lambda))^{q/2}(\mathbb{E}[|f_1|^q] + \cdots + \mathbb{E}[|f_m|^q]) \leq (Cqn/(1-\lambda))^{q/2},$$

for some universal constant C as desired. □

3.3.2 A bound for ℓ_2

Dvoretzky's theorem, which we state below, shows that ℓ_2 can be embedded into ℓ_q for any $q \geq 1$ (Dvo61).

Theorem 7. *For all $1 \leq q < \infty$, $\epsilon > 0$, and integer m, there exists $N_q(m, \epsilon)$ and a linear function $g : \mathbb{R}^m \to \mathbb{R}^{N_q(m,\epsilon)}$ so that for all $x \in \mathbb{R}^m$,*

$$\|x\|_{\ell_2} \leq \|g(x)\|_{\ell_q} \leq (1+\epsilon)\|x\|_{\ell_2}.$$

Lemma 7. *Let $\{Y_s\}_{s=1}^{\infty}$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A and stationary probability measure μ, so that Y_1 is distributed according to μ. Let*

$\lambda = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$ and let $f : [N] \to \mathbb{R}^m$ be so that $\mathbb{E}[f(Y_1)] = 0$ and $\mathbb{E}[\|f(Y_1)\|_{\ell_2}^q] = 1$. Let $S_n = f(Y_1) + \cdots + f(Y_n)$. Then for $q \geq 2$,

$$\mathbb{E}[\|S_n\|_{\ell_2}^q]^{1/q} \lesssim \sqrt{\frac{qn}{1-\lambda}}.$$

Proof. Let g be the function in Theorem 7 for $\epsilon = 1$ and m. Then,

$$\begin{aligned}\mathbb{E}[\|S_n\|_{\ell_2}^q]^{1/q} &\leq \mathbb{E}[\|g(S_n)\|_{\ell_q}^q]^{1/q} \\ &= \mathbb{E}[\|g(f(Y_1)) + \cdots + g(f(Y_n))\|_{\ell_q}^q]^{1/q} \\ &\lesssim \mathbb{E}[\|g(f(Y_1))\|_{\ell_q}^q]^{1/q}\sqrt{qn/(1-\lambda)} \\ &\leq 2\mathbb{E}[\|f(Y_1)\|_{\ell_2}^q]^{1/q}\sqrt{qn/(1-\lambda)}\ ,\end{aligned}$$

where the first inequality is by Theorem 7, the second is by Lemma 6 (applied to the function $g \circ f$), and the last is again by Theorem 7. □

3.3.3 A bound for ℓ_p for all $p \geq 2$

We need the following generalization of the Riesz-Thorin interpolation theorem, which can be obtained from Theorems 4.1.2 and 5.1.2 in (BL76).

Theorem 8. *Let $1 \leq p_0 \leq p_1 \leq \infty$ and $1 \leq q \leq \infty$, and μ and μ' be distributions. Let T be a linear operator such that $\|T\|_{L_q(\mu;\ell_{p_0})\to L_q(\mu';\ell_{p_0})}$ and $\|T\|_{L_q(\mu;\ell_{p_1})\to L_q(\mu';\ell_{p_1})}$ are finite. Let $1/p_\theta = (1-\theta)/p_0 + \theta/p_1$. Then*

$$\|T\|_{L_q(\mu;\ell_{p_\theta})\to L_q(\mu';\ell_{p_\theta})} \leq \|T\|_{L_q(\mu;\ell_{p_0})\to L_q(\mu';\ell_{p_0})}^{1-\theta}\|T\|_{L_q(\mu;\ell_{p_1})\to L_q(\mu';\ell_{p_1})}^{\theta}.$$

The following is the upper bound on $\mathbb{E}[\|S_n\|_{\ell_p}^q]^{1/q}$.

Proof of Theorem 5. For the case $p \leq q$, as in Theorem 4, let T_n be a linear transformation from functions f on $[N]$ to functions on $[N]^n$, so that

$$T_n f(W) = \sum_{i=1}^{n} f(Y_i)$$

for $W \in [N]^n$. As in the proof of Theorem 4, let I_N be the identity operator, and note that by the triangle inequality $\|I_N - E_\mu\|_{L_q(\mu;\ell_p)\to L_q(\mu;\ell_p)} \leq 2$, for $p = 2, q$. Then by Lemma 7 and by Lemma 6,

$$\|T_n \circ (I_n - \mathbb{E}_\mu)\|_{L_q(\mu;\ell_2)\to L_q(\mu_n;\ell_2)}, \|T_n \circ (I_n - \mathbb{E}_\mu)\|_{L_q(\mu;\ell_q)\to L_q(\mu_n;\ell_q)} \lesssim \sqrt{\frac{qn}{1-\lambda}},$$

where μ_n is the law of $(Y_1, \dots, Y_n)$. The statement of the theorem follows from Theorem 8 and setting $p_0 = 2$, $p_1 = q$, and $p_\theta = p$.

If $p \geq q$, then by Jensen's inequality, $\mathbb{E}[\|S_n\|_{\ell_p}^q]^{1/q} \leq \mathbb{E}[\|S_n\|_{\ell_p}^p]^{1/p}$, and the theorem follows from Lemma 6. □

3.4 Tail bounds

We now prove the two tail bounds mentioned in the introduction. Both proofs follow easily from Markov's inequality, and are nearly identical.

Proof of Corollary 1. The first tail bound follows by noting that

$$\Pr[|f(Y_1) + \cdots + f(Y_n)| \geq u\sqrt{n}] = \Pr[|f(Y_1) + \cdots + f(Y_n)|^q \geq (u^2 n)^{q/2}]$$

and using Markov's inequality and Theorem 4.

For the second tail bound, note that if $u \leq \sqrt{2Ce/(1-\lambda)}$, then the right-hand side is greater than 1 and the inequality holds trivially. Otherwise, the bound follows by taking $q = u^2(1-\lambda)/(Ce)$, which is greater than 2. □

Proof of Corollary 2. The first tail bound follows by noting that

$$\Pr[\|f(Y_1) + \cdots + f(Y_n)\|_{\ell_p} \geq u\sqrt{n}] = \Pr[\|f(Y_1) + \cdots + f(Y_n)\|_{\ell_p}^q \geq (u^2 n)^{q/2}]$$

and using Markov's inequality and Theorem 5.

For the second tail bound, note that if $u \leq \sqrt{pCe/(1-\lambda)}$, then the right-hand side is at least 1 and the inequality holds trivially. Otherwise the bound follows by taking $q = u^2(1-\lambda)/(Ce)$ which is greater than p, and thus $\max\{p, q\} \leq 2q$. □

Chapter 4

A Hoeffding Inequality for Markov Chains

In this chapter, we prove a Markov chain version of Hoeffding's inequality. We also use this to prove tail bounds for sums of vector-valued random variables, and in pariticular point out an application to norms of random matrices whose entries are obtained from a Markov chain.

Let $Y_1, \ldots, Y_n$ be independent random variables uniform over $[N]$ and let the functions $f_1, \ldots, f_n$ be on $[N]$ so that f_i has range $[-a_i, a_i]$. Recall from the introduction that Hoeffding obtained the following tail bound (Hoe63).

$$\Pr\left[|f_1(Y_1) + \cdots + f_n(Y_n)| \geq u \left(\sum_{i=1}^{n} a_i^2\right)^{1/2}\right] \leq 2\exp(-u^2/2). \tag{4.1}$$

In this work, we generalize Eq. (4.1) to reversible Markov chains with a stationary distribution. In particular, we prove the following.

Theorem 9. *Let $\{Y_i\}_{i=1}^{\infty}$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A, stationary probability measure μ, and averaging operator E_μ, so that Y_1*

is distributed according to μ. Let $\lambda = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$ and let $f_1, \dots, f_n : [N] \to \mathbb{R}$ so that $\mathbb{E}[f_i(Y_i)] = 0$ for all i and $|f_i(v)| \le a_i$ for all $v \in [N]$ and all i. Then for $u \ge 0$,

$$\Pr\left[|f_1(Y_1) + \cdots + f_n(Y_n)| \ge u\left(\sum_{i=1}^{n} a_i^2\right)^{1/2}\right] \le 2\exp(-u^2(1-\lambda)/(64e)).$$

One interpretation of Theorem 9 is that for a Markov chain $\{Y_i\}_{i=1}^{\infty}$ and functions $f_1, \dots, f_n : [N] \to [-1, 1]$, the random vector $(f_1(Y_1), \dots, f_n(Y_n))$ is sub–gaussian.

4.0.1 Extension to vector–valued random variables

We extend Theorem 9 to random variables from a fixed Banach space as follows. We stress that the setting in the following theorem is more limited than that of Eq. (1.8). In particular we only allow random variables of the form $f(Y_i)X_i$ in which $f(Y_i)$ is a random scalar and X_i is a fixed element from the Banach space.

Theorem 10. *Let $(X, \|\cdot\|)$ be a Banach space, and let $X_1, \dots, X_n \in X$. Let $\{Y_i\}_{i=1}^{\infty}$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A, stationary probability measure μ, and averaging operator E_μ, so that Y_1 is distributed according to μ. Let $\lambda = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$, and let $f_1, \dots, f_n : [N] \to [-1, 1]$ be such that $\mathbb{E}[f_i(Y_i)] = 0$ for all i. Then there exist universal constants C and L, such that for any $u \ge 0$,*

$$\Pr\left[\|f_1(Y_1)X_1 + \cdots + f_n(Y_n)X_n\| \ge uC\mathbb{E}[\|g_1X_1 + \cdots + g_nX_n\|]\right] \le L\exp(-Cu^2(1-\lambda))$$

where $g_1, \dots, g_n \sim \mathcal{N}(0, 1)$ are independent standard Gaussian random variables.

Note that Eq. (1.8) implies that $\mathbb{E}[\|g_1X_1 + \cdots + g_nX_n\|] \le C\sqrt{s(\|X_1\|^2 + \cdots + \|X_n\|^2)}$ for some constant C. This follows from the fact that the distribution of the normalized sum of independent Rademacher random variables approaches that of a Gaussian, in the limit.

Thus for Banach spaces that satisfy $\rho_X(\tau) \leq s\tau^2$, we also have the bound

$$\Pr\left[\|f_1(Y_1)X_1 + \cdots + f_n(Y_n)X_n\| \geq uC\sqrt{s(\|X_1\|^2 + \cdots + \|X_n\|^2)}\right] \leq L\exp(-Cu^2(1-\lambda))$$

4.0.1.1 Bounds on the Schatten ∞-norm of a random matrix

As an application, we are able to generalize bounds on the Schatten ∞-norm of a matrix with independent entries to matrices whose entries are obtained from a reversible Markov chain with stationary distribution.

Let $\mathcal{I} \subseteq [d] \times [d]$ be the set of pairs (i,j) such that $i \leq j$, and let $B = (b_{i,j}) \in \mathbb{R}^{d\times d}$ be a symmetric matrix with positive entries. Let $X \in \mathbb{R}^{d\times d}$ be the random symmetric matrix whose entries are

$$X_{i,j} = \begin{cases} \varepsilon_{i,j}b_{i,j} & \text{if } (i,j) \in \mathcal{I} \\ \varepsilon_{j,i}b_{i,j} & \text{otherwise} \end{cases}$$

where $\varepsilon_{i,j}$ are independent Rademacher random variables. Then it was shown in (BvH16) that

$$\mathbb{E}[\|X\|_{S_\infty}] \leq \min\left\{C(\sigma + \sigma_*\sqrt{\log d}), \|B\|_{S_\infty}\right\} \tag{4.2}$$

for some absolute constant C, where

$$\sigma = \max_i \sqrt{\sum_j b_{i,j}^2} \text{ and } \sigma_* = \max_{i,j} |b_{i,j}|. \tag{4.3}$$

We generalize Eq. (4.2) to reversible Markov chains with a stationary distribution. In particular, we obtain a similar bound in terms of $\lambda = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$ on the Schatten ∞-norm of a matrix whose entries are chosen in the following manner. We start by choosing an arbitrary permutation of the entries in the diagonal and upper triangular part of the matrix. Then we fill in the entries according to the order given by the permutation, using the values given by the Markov chain. Finally we fill in the entries in the lower triangular part of the matrix, so that the matrix is symmetric. The case that the transition matrix is $A = E_\mu$ corresponds to choosing the entries of the diagonal and upper triangular part of the matrix independently, as in (BvH16).

Corollary 3. *Let $\{Y_i\}_{i=1}^\infty$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A, stationary probability measure μ, and averaging operator E_μ, so that Y_1 is distributed according to μ. Let $\lambda = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$, let $f : V \to [-1, 1]$ be such that $\mathbb{E}[f(Y_i)] = 0$, and let $B \in \mathbb{R}^{d\times d}$ be a symmetric $d \times d$ matrix with positive entries. For any injective function $\omega : \mathcal{I} \to \{1, 2, \ldots, (d^2 + d)/2\}$, let X be the symmetric matrix defined by*

$$X_{i,j} = \begin{cases} f(Y_{\omega(i,j)})b_{i,j} & \text{if } (i,j) \in \mathcal{I} \\ f(Y_{\omega(j,i)})b_{j,i} & \text{otherwise} \end{cases}$$

Then,

$$\mathbb{E}[\|X\|_{S_\infty}] \le \min\left\{\frac{C}{\sqrt{1-\lambda}}(\sigma + \sigma_*\sqrt{\log d}), \|B\|_{S_\infty}\right\},$$

for some absolute constant C, where σ and σ_* are defined as in Eq. (4.3).

4.1 Preliminaries

The following preliminaries are specific to this chapter.

The following simple claim bounds $\|T\|_{L_2(\mu)\to L_2(\mu)}$ for a matrix T in terms of $\|T\|_{L_1(\mu)\to L_1(\mu)}$ and $\|T\|_{L_\infty(\mu)\to L_\infty(\mu)}$. This can be viewed as a special case of interpolation of matrix norms.

Claim 5. *For any matrix T,*

$$\|T\|^2_{L_2(\mu)\to L_2(\mu)} \le \|T\|_{L_1(\mu)\to L_1(\mu)}\|T\|_{L_\infty(\mu)\to L_\infty(\mu)}.$$

Proof. For all $x, \mu \in \mathbb{R}^n$ so that μ has positive entries,

$$\|Tx\|^2_{L_2(\mu)\to L_2(\mu)} = \sum_{i=1}^n \mu_i \left(\sum_{j=1}^n T_{ij}x_j\right)^2 \leq \sum_{i=1}^n \mu_i \left(\sum_{j=1}^n |T_{ij}|\right)\left(\sum_{j=1}^n |T_{ij}|x_j^2\right)$$
$$\leq \|T\|_{L_\infty(\mu)\to L_\infty(\mu)} \|T(x\circ x)\|_{L_1(\mu)\to L_1(\mu)} \leq \|T\|_{L_\infty(\mu)\to L_\infty(\mu)}\|T\|_{L_1(\mu)\to L_1(\mu)}\|x\|^2_{L_2(\mu)}$$

where the first inequality follows by Cauchy-Schwarz, and $\circ$ denotes entrywise product. □

4.2 Proof of Theorem 9

To prove Theorem 9, we follow the strategy of bounding the qth moment for some even integer q, and using Markov's inequality to obtain a tail bound. We start by expanding $(f_1(Y_1) + \cdots + f_n(Y_n))^q$ into a sum of monomials.

The following lemma bounds the expectation of monomials in the $f_i(Y_i)$, and can be derived from Lemma 2 in the introduction. Recall that $S_{q-1} \subset \{0,1\}^{q-1}$ is the set of strings with no consecutive 0's and so that $s_1, s_{q-1} = 1$ for all $s \in S_{q-1}$.

Lemma 8. *Let $\{Y_i\}_{i=1}^\infty$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A, stationary probability measure μ, and averaging operator E_μ, so that Y_1 is distributed according to μ. Let $\lambda = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$ and let $f_1, \ldots, f_n : [N] \to \mathbb{R}$ so that $\mathbb{E}[f_i(Y_i)] = 0$ for all i and $|f_i(v)| \leq a_i$ for all $v \in [N]$ and all i. For all q, and $w \in [n]^q$ such that $w_1 \leq w_2 \leq \cdots \leq w_q$*

$$\mathbb{E}[f_{w_1}(Y_{w_1})f_{w_2}(Y_{w_2})\cdots f_{w_q}(Y_{w_q})] \leq a_{w_1}a_{w_2}\cdots a_{w_q} \sum_{s\in S_{q-1}} \left(\prod_{i:s_i=1} \lambda^{w_{i+1}-w_i}\right).$$

Proof. We apply Eq. (2.5) from Lemma 2 in Chapter 2, letting $k = q-1$, $u_i(v) = f_{w_i}(v)$ for all $v \in [N]$, and $T_i = A^{w_{i+1}-w_i} - E_\mu$. Note that for all $k \geq 0$,

$$A^k - E_\mu = A^k - A^{k-1}E_\mu - E_\mu A + E_\mu^2 = (A^{k-1} - E_\mu)(A - E_\mu) = (A - E_\mu)^k.$$

The lemma follows by noting that $\|u_i\|_{L_\infty(\mu)} \leq a_{w_i}$ and $\|T_i\|_{L_2(\mu)\to L_2(\mu)} \leq \lambda^{w_{i+1}-w_i}$ □

We obtain the following bound on the moments of $f_1(Y_1) + \cdots + f_n(Y_n)$.

Theorem 11. *Let $\{Y_i\}_{i=1}^\infty$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A, stationary probability measure μ, and averaging operator E_μ, so that Y_1 is distributed according to μ. Let $\lambda = \|A - E_\mu\|_{L_2(\mu)\to L_2(\mu)}$ be less than 1, and let $f_1, \ldots, f_n : [N] \to \mathbb{R}$ so that $\mathbb{E}[f_i(Y_i)] = 0$ for all i and $|f_i(v)| \leq a_i$ for all $v \in [N]$ and all i. Then for even q,*

$$\mathbb{E}[(f_1(Y_1) + \cdots + f_n(Y_n))^q] \leq 4^q (q/2)! \left(\frac{1}{1-\lambda}\right)^{q/2} \left(\sum_{i=1}^n a_i^2\right)^{q/2}.$$

Proof. Let $\sigma : [n]^q \to [n]^q$ be the function where $\sigma(w)$ is the sorted list of coordinates of w in non-decreasing order. Then by Lemma 9,

$$\begin{aligned}\mathbb{E}[(f_1(Y_1) + \cdots + f_n(Y_n))^q] &= \sum_{w\in[n]^q} \mathbb{E}[f_{w_1}(Y_{w_1}) f_{w_2}(Y_{w_2}) \cdots f_{w_q}(Y_{w_q})] \\ &\leq \sum_{w\in[n]^q} a_{w_1} a_{w_2} \cdots a_{w_q} \sum_{s\in S_{q-1}} \left(\prod_{i:s_i=1} \lambda^{\sigma(w)_{i+1}-\sigma(w)_i}\right). \qquad (4.4)\end{aligned}$$

Let $\binom{[q]}{q/2}$ denote the collection of subsets of $[q]$ of size exactly $q/2$. For each subset $\mathcal{I} \in \binom{[q]}{q/2}$, let $W_\mathcal{I} \subset [n]^q$ be the set of all vectors w such that for each $j \in [n]$,

$$|\{i : i \in \mathcal{I} \text{ and } w_i = j\}| = |\{i : i \in \{1, 3, 5, \ldots, q-1\} \text{ and } \sigma(w)_i = j\}|,$$

i.e. the multi-set $\bigcup_{i\in\mathcal{I}}\{w_i\}$ is equal to the multi-set $\{\sigma(w)_1, \sigma(w)_3, \sigma(w)_5, \ldots, \sigma(w)_{q-1}\}$. Let $w_{\mathcal{I}}, w_{[q]\backslash\mathcal{I}} \in [n]^{q/2}$ be the restriction of w to the coordinates in $\mathcal{I}$ and $[q]\backslash\mathcal{I}$ respectively. Additionally, for each $\mathcal{I} \in \binom{[q]}{q/2}$ and $s \in S_{q-1}$, let $T_{\mathcal{I},s}$ be the $n^{q/2} \times n^{q/2}$ matrix defined as follows. For each $w \in [n]^q$, the entry in the $w_{\mathcal{I}}$th row and $w_{[q]\backslash\mathcal{I}}$th column of $T_{\mathcal{I},s}$ is

$$T_{\mathcal{I},s}(w_{\mathcal{I}}, w_{[q]\backslash\mathcal{I}}) = \begin{cases} \prod_{i:s_i=1} \lambda^{\sigma(w)_{i+1}-\sigma(w)_i} & \text{if } w \in W_{\mathcal{I}} \\ 0 & \text{otherwise.} \end{cases}$$

Because

$$\bigcup_{\mathcal{I}\in\binom{[q]}{q/2}} W_{\mathcal{I}} = [n]^q,$$

Eq. (4.4) can be bounded above by

$$\begin{aligned} \sum_{s\in S_{q-1}} \sum_{\mathcal{I}\in\binom{[q]}{q/2}} \sum_{w\in W_{\mathcal{I}}} a_{w_1}a_{w_2}\cdots a_{w_q}\left(\prod_{i:s_i=1}\lambda^{\sigma(w)_{i+1}-\sigma(w)_i}\right) &= \sum_{s\in S_{q-1}} \sum_{\mathcal{I}\in\binom{[q]}{q/2}} \left\langle a^{\otimes q/2}, T_{\mathcal{I},s}a^{\otimes q/2}\right\rangle_{\ell_2} \\ &\leq |S_{q-1}|\binom{q}{q/2} \max_{s\in S_{q-1},\mathcal{I}\in\binom{[q]}{q/2}} \|T_{\mathcal{I},s}\|_{\ell_2\to\ell_2}\|a\|_{\ell_2}^q, \end{aligned}$$

where $a^{\otimes q/2} \in \mathbb{R}^{n^{q/2}}$ is the vector such that $a^{\otimes q/2}_{i_1,\ldots,i_{q/2}} = a_{i_1}a_{i_2}\cdots a_{i_{q/2}}$ for $i \in [n]^{q/2}$ and thus $\|a^{\otimes q/2}\|_{\ell_2} = \|a\|_{\ell_2}^{q/2}$. Both $|S_{q-1}|$ and $\binom{q}{q/2}$ are each bounded above by 2^q. Thus by Claim 5, it is enough to show that

$$\|T_{\mathcal{I},s}\|_{\ell_1\to\ell_1}, \|T_{\mathcal{I},s}\|_{\ell_\infty\to\ell_\infty} \leq (q/2)!\left(\frac{1}{1-\lambda}\right)^{q/2}.$$

We show this for $\|T_{\mathcal{I},s}\|_{\ell_\infty\to\ell_\infty}$; the proof for $\|T_{\mathcal{I},s}\|_{\ell_1\to\ell_1}$ is similar.

Because the entries of T are positive, $\|T_{\mathcal{I},s}\|_{\ell_\infty\to\ell_\infty}$ is just the largest row sum of $T_{\mathcal{I},s}$. Without loss of generality, assume that $\mathcal{I} = \{1, 3, 5, \ldots, q-1\}$. Then the sum of the entries of the row corresponding to $w_{\mathcal{I}} = (w_1, w_3, w_5, \ldots, w_{q-1})$ is

$$\sum_{w_2, w_4, \ldots, w_q \colon w \in W_{\mathcal{I}}} T_{\mathcal{I},s}(w_{\mathcal{I}}, w_{[q]\setminus\mathcal{I}}) \leq (q/2)! \sum_{w_2=\sigma(w)_1}^{\sigma(w)_3} \sum_{w_4=\sigma(w)_3}^{\sigma(w)_5} \cdots \sum_{w_q=\sigma(w)_{q-1}}^{n} \prod_{i:s_i=1} \lambda^{\sigma(w)_{i+1}-\sigma(w)_i}$$
$$\leq (q/2)! \left(\frac{1}{1-\lambda}\right)^{q/2},$$

as desired. The first inequality follows from the fact that $w \in W_{\mathcal{I}}$ and $w_1, w_3, w_5, \ldots, w_{q-1}$ determine $\sigma(w)_1, \sigma(w)_3, \sigma(w)_5, \ldots, \sigma(w)_{q-1}$ exactly, and that there are at most $(q/2)!$ possible orderings of $w_2, w_4, \ldots, w_q$. The second inequality follows from the definition of S_{q-1}, which implies that for every positive even integer $k \leq q$, either $s_{k-1} = 1$ or $s_k = 1$, along with the formula for the sum of an infinite geometric series. □

Finally, Theorem 9 follows by considering the moment generating function and applying Markov's inequality.

Proof of Theorem 9. If $\lambda \geq 1$ or if $u \leq 8/\sqrt{1-\lambda}$, the theorem holds trivially as the right-hand side is greater than 1.

Otherwise, we start by bounding the moment generating function. Let $\theta = (1-\lambda)u/(32(a_1^2 + \cdots + a_n^2)^{1/2})$ By Theorem 11 and keeping in mind that by Jensen's inequality, odd moments are bounded above by even moments,

$$\mathbb{E}\left[\exp(\theta(f_1(Y_1) + \cdots + f_n(Y_n)))\right] = \sum_{q=0}^{\infty} \frac{\mathbb{E}[\theta(f_1(Y_1) + \cdots + f_n(Y_n))^q]}{q!}$$
$$\leq 1 + \sum_{q=1}^{\infty} \frac{(1-\lambda)^{(2q-1)/2} u^{2q-1} q!}{8^{2q-1}(2q-1)!} + \frac{(1-\lambda)^q u^{2q} q!}{8^{2q}(2q)!}$$
$$\leq 2 \sum_{q=0}^{\infty} \frac{(1-\lambda)^q u^{2q}}{8^{2q} q!}$$
$$= 2 \exp\left(u^2(1-\lambda)/64\right).$$

By Markov's inequality,

$$\begin{aligned}
&\Pr\left[f_1(Y_1) + \cdots + f_n(Y_n) \geq u\left(\sum_{i=1}^{n} a_i^2\right)^{1/2}\right] \\
&\qquad\qquad = \Pr\left[\exp(\theta(f_1(Y_1) + \cdots + f_n(Y_n))) \geq \exp\left(\theta u\left(\sum_{i=1}^{n} a_i^2\right)^{1/2}\right)\right] \\
&\qquad\qquad \leq \frac{\mathbb{E}\left[\exp(\theta(f_1(Y_1) + \cdots + f_n(Y_n)))\right]}{\exp\left(\theta u\left(\sum_{i=1}^{n} a_i^2\right)^{1/2}\right)} \\
&\qquad\qquad \leq 2\exp\left(u^2(1-\lambda)/64 - u^2(1-\lambda)/32\right) \\
&\qquad\qquad = 2\exp\left(-u^2(1-\lambda)/64\right)
\end{aligned}$$

The final bound follows by doing the same for the left tail, and noting that if $u \geq 8/\sqrt{1-\lambda}$, either $4\exp(-u^2(1-\lambda)/64) \leq 2\exp(-u^2(1-\lambda)/(64e))$, or $2\exp(-u^2(1-\lambda)/(64e) \geq 1$.

□

We note that it is possible to obtain stronger tail bounds that improve on the constant factor by optimizing some of the calculations above, but we will not do so here.

4.3 Extension to vector–valued random variables

To prove Theorem 10 we use the techniques of Talagrand's generic chaining. These techniques apply to random variables that satisfy the "increment condition," which we define below.

Definition 2. *A metric space (T, d) and process $(Z_t)_{t \in T}$ satisfies the increment condition if for all u and all $s, t \in T$,*

$$\Pr[|Z_s - Z_t| \geq u] \leq 2\exp\left(-\frac{u^2}{2d(s,t)^2}\right).$$

When $(Z_t)_{t \in T}$ is a gaussian process, that is Z_t is gaussian for all $t \in T$, we can equip T with the canonical distance, $d(s,t) = \mathbb{E}[(Z_s - Z_t)^2]^{1/2}$.

Theorem 9 essentially states that for a a Markov chain $\{Y_i\}_{i=1}^{\infty}$ and functions $f_1, \ldots, f_n : [N] \to [-1,1]$ with $\mathbb{E}[f_i(Y_i)] = 0$, the process $(Z_t)_{t \in T}$ defined by $Z_t = (f_1(Y_1)t_1, \ldots, f_n(Y_n)t_n)$ for $T = \mathbb{R}^n$ satisfies the increment condition if the associated distance is $\sqrt{32e/(1-\lambda)}$ times the Euclidean distance.

We also define the γ_2 functional.

Definition 3.

$$\gamma_2(T,d) = \inf \sup_{t \in T} \sum_{i=0}^{\infty} 2^{i/2} \min_{t' \in T_i} d(t,t'),$$

where the infimum is taken over all sequences of subsets $T_0 \subseteq T_1 \subseteq \cdots \subseteq T$ such that $|T_0| = 1$ and $|T_i| \leq 2^{2^i}$ for $i \geq 1$.

The majorizing measures theorem, due to Talagrand (Tal87) (see also Theorem 2.4.1 in (Tal14)), gives bounds on the expected value of $\sup_{t \in T} Z_t$, where $(Z_t)_{t \in T}$ is a gaussian process, in terms of $\gamma_2(T,d)$ where d is the canonical distance. We state the theorem below.

Theorem 12 (Talagrand's majorizing measures theorem)**.** *For some universal constant C, and for every gaussian process $(Z_t)_{t \in T}$,*

$$\frac{1}{C}\gamma_2(T,d) \leq \mathbb{E}\left[\sup_{t \in T} Z_t\right] \leq C\gamma_2(T,d),$$

where $d(s,t) = \mathbb{E}[(Z_s - Z_t)^2]^{1/2}$.

We also use the following tail bound for any process that satisfies the increment condition, which is given as Theorem 2.2.27 in (Tal14).

Theorem 13. *If the process (Z_t) satisfies the increment condition, then for $u > 0$, Then,*

$$\Pr\left[\sup_{s,t \in T} |X_s - X_t| \geq L\gamma_2(T,d) + uL \sup_{t_1,t_2 \in T} d(t_1,t_2)\right] \leq L\exp(-u^2).$$

We now describe how to select T to apply the above tools to the setting of Theorem 10. Let $(X, \|\cdot\|)$ be a Banach space, and let $(X^*, \|\cdot\|_*)$ be the dual space of X with closed unit ball B^*. Recall that for $x \in X$,

$$\|x\| = \sup_{x^* \in B^*} |\langle x^*, x\rangle|.$$

(see for instance, Theorem 4.3 in (Rud91)). For fixed $X_1, \ldots, X_n \in X$, let $T \subset \mathbb{R}^n$ be the set of points,

$$T = \{(\langle x^*, X_1\rangle, \langle x^*, X_2\rangle, \ldots, \langle x^*, X_n\rangle) : x^* \in B^*\}. \tag{4.5}$$

Note that T is symmetric, as for every $x^* \in B^*$, we also have $-x^* \in B^*$. It follows that

$$\|f_1X_1 + \cdots + f_nX_n\| = \sup_{t \in T}\langle f, t\rangle. \tag{4.6}$$

Finally, we prove Theorem 10.

Proof of Theorem 10. Consider the metric space (T, d) where T is as constructed in Eq. (4.5) and $d(s,t) = \sqrt{32e/(1-\lambda)}\|s - t\|_{\ell_2}$. Then by Theorem 9, the process $(Z_t)_{t\in T}$ defined by $Z_t = (f_1(Y_1)t_1, \ldots, f_n(Y_n)t_n)$ satisfies the increment condition.

Additionally, consider the Gaussian process $(Z'_t)_{t\in T}$ on the metric space (T, d'), so that $Z_t = g_1t_1 + \cdots + g_nt_n$ for independent standard Gaussian variables $g_1, \ldots, g_n$ and $d' = \mathbb{E}[(Z_s - Z_t)^2]^{1/2}$. Then by Theorem 12,

$$\gamma_2(T, d) = \sqrt{\frac{32e}{1-\lambda}}\gamma_2(T, d') \leq \frac{C}{1-\lambda}\mathbb{E}\left[\sup_{t\in T} Z'_t\right]$$

The theorem then follows from Theorem 13 the observation that $\sup_{s,t}|Z_s - Z_t| = 2\sup_t Z_t$ as T is symmetric, and Eq. (4.6). $\square$

4.3.1 Comparison to matrices with independent entries

We prove Corollary 3, which follows from a straightforward application of Theorem 10.

In order to apply Theorem 10, we need a bound on $\mathbb{E}[\|X'\|_{S_\infty}]$ when X' is the random symmetric matrix whose entries are

$$X'_{i,j} = \begin{cases} g_{i,j} b_{i,j} & \text{if } (i,j) \in \mathcal{I} \\ g_{j,i} b_{i,j} & \text{otherwise} \end{cases}$$

where $g_{i,j} \sim \mathcal{N}(0,1)$ are independent standard Gaussian random variables (rather than Rademacher random variables, as in Eq. (4.2)). This setting was also discussed in (BvH16) in which it was shown that

$$\mathbb{E}[\|X'\|_{S_\infty}] \leq C(\sigma + \sigma_* \sqrt{\log d}), \tag{4.7}$$

where σ and σ_* are defined as in Eq. (4.3).

Proof of Corollary 3. Let X' be the random matrix defined above. Then by Theorem 10 and Eq. (4.7),

$$\mathbb{E}[\|X\|_{S_\infty}] \leq \frac{C}{\sqrt{1-\lambda}} \mathbb{E}[\|X'\|_{S_\infty}] \leq \frac{C'}{\sqrt{1-\lambda}} (\sigma + \sigma_* \sqrt{\log d})$$

Finally, because $|f(v)| \leq 1$ for all $v \in [N]$ and B has positive entries, it follows that $\|X\|_{S_\infty} \leq \|B\|_{S_\infty}$, always. □

Chapter 5

Poisson Approximations for Markov Chains

In this chapter, we investigate to what extent sums of random variables obtained from a Markov chain share properties with Poisson random variables. As an application, we construct explicit resilient functions matching the nonconstructive versions shown to exist due to Ajtai and Linial (AL93).

In particular, let $\{Y_i\}_{i=1}^{\infty}$ be a reversible Markov chain with state space $[N]$ and stationary distribution μ, and let $f_1, \ldots, f_n : V \to [0, 1]$ (where often $f_1 = \cdots = f_n = f$) be so that $\mathbb{E}[f_1(Y_1) + \cdots + f_n(Y_n)] = \Phi$. When the Y_i are completely independent of each other, a simple calculation shows that the moment generating function is bounded above by

$$\mathbb{E}[\alpha^{f_1(Y_1)+\cdots+f_n(Y_n)}] \leq \exp((\alpha - 1) \cdot \Phi) \ .$$

The above bound easily imply tail bounds on $f_1(Y_1) + \cdots + f_n(Y_n)$ by Markov's inequality.

In this chapter we prove an analogous bound for the case of Markov chains

Theorem 14. $\{Y_i\}_{i=1}^{\infty}$ *be a reversible Markov chain with state space* $[N]$ *and stationary* *distribution μ, and let $\lambda = \lambda(Y)$. Let $f_1, \ldots, f_n : V \to [0,1]$, and let $\sigma_i = \mathbb{E}[f_i(v)]$. Let $\Phi := \mathbb{E}[f_1(Y_1) + \cdots + f_n(Y_n)]$. Then for $1 < \alpha < 1/\lambda$,*

$$\mathbb{E}[\alpha^{f_1(Y_1)+\cdots+f_n(Y_n)}] \leq \exp\left((\alpha - 1) \cdot \Phi \cdot \left(\frac{1-\lambda}{1-\alpha\lambda}\right)\right). \tag{5.1}$$

The proof strategy we use for Theorem 14 is to first bound expressions of the form

$$\mathbb{E}[Z_{w_1} Z_{w_2} \cdots Z_{w_k}]$$

where $Z_i = f(Y_i)$ for all i. The will then allow us to bound expressions of the form

$$\mathbb{E}\left[(f_1(Y_1) + \cdots + f_n(Y_n))^q\right]$$

for integers q, which will finally allow us to obtain a bound on the moment generating function.

As stated previously, bounds on the moment generating function immediately imply tail bounds. In particular, by plugging in $\alpha = \lambda^{-1} - (1-\lambda)/(t^{1/2}\lambda^{3/2})$ in Theorem 14 we obtain the following.

Corollary 4. *In the setting of Theorem 14, for all $t > 1/\lambda$,*

$$\Pr[S_n \geq t\Phi] \leq \left(\frac{1}{\lambda} - \frac{1-\lambda}{t^{1/2}\lambda^{3/2}}\right)^{-t\Phi} \exp(\Phi \cdot (1-\lambda)(\sqrt{t\lambda} - 1)/\lambda) . \tag{5.2}$$

For instance, for $\lambda = 1/2$, we obtain

$$\Pr[S_n \geq t\Phi] \leq \left(2 - \sqrt{\frac{2}{t}}\right)^{-t\Phi} \exp(\Phi \cdot (\sqrt{t/2} - 1)) .$$

We also note that for large t, the bound in (5.2) is roughly $\lambda^{t\Phi}$, which is again close to tight by the example in Section 5.0.1.

5.0.1 Sharpness

We now sketch an argument showing that Theorem 14 is sharp in the following sense. Fix arbitrary $\lambda \in [0,1)$ and $\Phi > 0$, and consider the Markov chain with transition matrix $A = \lambda I + (1-\lambda)J$ of dimensions $N \times N$ where J is the matrix whose entries are all $1/N$. This corresponds to the walk where at each step we either stay in place with probability λ, or choose a uniform vertex with probability $1-\lambda$. Let $f_1 = \cdots = f_n = f$ be the function that assigns 1 to a $\mu = \Phi/n$ fraction of "marked" states and 0 to the remaining states (where we assume for simplicity that $\Phi|V|/n$ is integer). Equivalently, one can consider a Markov chain with two states, one marked and one unmarked; a step in the chain stays in the same state with probability λ and otherwise chooses from the stationary distribution, which assigns mass μ to the marked state and $1-\mu$ to the unmarked state.

Then we claim that as n goes to infinity, the left-hand side of Eq. (5.1) converges to the right-hand side. To see that, we say that a step of the walk is a "hit" if (1) the walk chooses a uniform state (which happens with probability $1-\lambda$), and (2) that chosen state is marked. Then observe that the random variable counting the number of hits during the walk converges to a Poisson distribution with expectation $(1-\lambda)\Phi$ (since it is the sum of n independent Bernoulli random variables, each with probability $(1-\lambda)\mu$ of being 1). Moreover, each time a hit occurs, we stay in that vertex a number of steps that is distributed like a geometric distribution with success probability $1-\lambda$. (We are ignoring here lower order effects, such as reaching the end of the walk.) Therefore, using the probability mass function of the Poisson distribution and the moment generating function of the geometric distribution, we see that for any $\alpha < 1/\lambda$, as n goes to infinity, $\mathbb{E}[\alpha^{S_n}]$ converges to

$$\sum_{k=0}^{\infty} \frac{\exp(-(1-\lambda)\Phi)((1-\lambda)\Phi)^k}{k!} \Big(\frac{(1-\lambda)\alpha}{1-\lambda\alpha}\Big)^k = \exp\Big(-(1-\lambda)\Phi + (1-\lambda)\Phi\frac{(1-\lambda)\alpha}{1-\lambda\alpha}\Big)$$
$$= \exp\left(\Phi \cdot \left(\frac{(1-\lambda)(\alpha-1)}{1-\alpha\lambda}\right)\right),$$

as desired.

5.1 Bounding monomials

In this section we prove Lemma 9, bounding the expectation of monomials in the $f_i(Y_i)$.

Lemma 9. *Let $\{Y_i\}_{i=1}^{\infty}$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A and stationary distribution μ so that Y_1 is distributed as μ, and let $\lambda = \lambda(Y)$. Let $f_1, \ldots, f_n : V \to [0,1]$, and let $\sigma_i = \mathbb{E}[f_i(Y_i)]$, and let $Z_i = f_i(Y_i)$ for all i. Then for all $k \geq 1$ and $w \in [n]^k$ such that $w_1 \leq w_2 \leq \cdots \leq w_k$,*

$$\mathbb{E}[Z_{w_1} Z_{w_2} \cdots Z_{w_k}] \leq \sum_{s \in \{0,1\}^{k-1}} \sqrt{\sigma_{w_1}\sigma_{w_k}} \left(\prod_{i: s_i = 0} \sqrt{\sigma_{w_i}\sigma_{w_{i+1}}}\right)\left(\prod_{i: s_i=1} \lambda^{w_{i+1}-w_i}\right).$$

Proof. Let $d_i = w_{i+1} - w_i$ for all i. Let u_i be given by $(u_i)_v = f_{w_i}(v)$ and let $U_i = \operatorname{diag}(u_i)$. Then

$$\mathbb{E}[Z_{w_1} Z_{w_2} \cdots Z_{w_k}] = \|U_1 A^{d_1} U_2 A^{d_2} \cdots A^{d_{k-1}} U_k \mathbf{1}\|_{L_1(\mu)}. \tag{5.3}$$

Let $T_{i,0} = E_\mu$ and $T_{i,1} = A^{d_i} - E_\mu$ and notice that $\|T_{i,1}\|_{L_2(\mu)\to L_2(\mu)} \leq \lambda^{d_i}$. Using the triangle inequality, we can bound the right-hand side of Eq. (5.3) from above by

$$\sum_{s \in \{0,1\}^{k-1}} \|U_1 T_{1,s_1} U_2 T_{2,s_2} \cdots T_{k-1,s_{k-1}} U_k \mathbf{1}\|_{L_1(\mu)}.$$

By Eq. (2.6) in Lemma 2 in Chapter 2, for each s, the term corresponding to s satisfies

$$\|U_1 T_{1,s_1} U_2 T_{2,s_2} \cdots T_{k-1,s_{k-1}} U_k \mathbf{1}\|_{L_1(\mu)} \leq \sqrt{\sigma_{w_1}\sigma_{w_k}} \left(\prod_{i:s_i=0} \sqrt{\sigma_{w_i}\sigma_{w_{i+1}}} \right) \left(\prod_{i:s_i=1} \lambda^{d_i} \right) . \tag{5.4}$$

The lemma follows by summing over $s \in \{0,1\}^{k-1}$. □

It is interesting to note that we can also let $T_{i,0} = (1-\lambda^{d_i})E_\mu$ in the proof above to obtain a bound of

$$\begin{aligned} \mathbb{E}[Z_{w_1} Z_{w_2} \cdots Z_{w_k}] &\leq \sum_{s \in \{0,1\}^{k-1}} \sqrt{\sigma_{w_1}\sigma_{w_k}} \left(\prod_{i:s_i=0} (1-\lambda^{w_{i+1}-w_i}) \sqrt{\sigma_{w_i}\sigma_{w_{i+1}}} \right) \left(\prod_{i:s_i=1} \lambda^{w_{i+1}-w_i} \right) \\ &= \sqrt{\sigma_{w_1}\sigma_{w_k}} \prod_{i=1}^{k-1} ((1-\lambda^{w_{i+1}-w_i}) \sqrt{\sigma_{w_i}\sigma_{w_{i+1}}} + \lambda^{w_{i+1}-w_i}). \end{aligned} \tag{5.5}$$

When $\sigma_1 = \cdots = \sigma_n = \mu$, this bound simplifies to

$$\mathbb{E}[Z_{w_1} Z_{w_2} \cdots Z_{w_k}] \leq \mu \prod_{i=1}^{k-1} ((1-\lambda^{w_{i+1}-w_i})\mu + \lambda^{w_{i+1}-w_i}).$$

Observe that for the two-state Markov chain described in Section 5.0.1, for every $n \geq 1$ and every $1 \leq w_1 \leq \cdots \leq w_k \leq n$, this inequality is actually an equality. Indeed, the left-hand side is the probability that we are in the marked state at all the steps $w_1, \ldots, w_k$. The probability of being in the marked state at step w_1 is μ (as we are in the stationary distribution); and the probability of being in the marked state at step w_{i+1} conditioned on being there at step w_i is $(1-\lambda^{w_{i+1}-w_i})\mu + \lambda^{w_{i+1}-w_i}$.

This observation implies that the moment generating function $\mathbb{E}[\alpha^{S_n}]$ of an arbitrary graph and arbitrary $f_1, \ldots, f_n$ with all $\mathbb{E}[f_i]$ equal can be bounded by the moment generating function of the corresponding two-state Markov chain (as can be seen from the Taylor expansion; see Eq. (5.8) below). This can be used to give an alternative (and perhaps more intuitive) proof of Theorem 14. We do not include this proof here since it is not clear how to extend it to the case of general σ_i.

5.2 Proof of Theorem 14

In this section we complete the proof of the main theorem using the bound in Lemma 9. We start with the following easy corollary of Cauchy-Schwarz.

Claim 6. *Let $P \in \mathbb{R}[X_1, \ldots, X_n]$ be a multivariate polynomial with non-negative coefficients. Then for $x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{R}$,*

$$P(x_1y_1, x_2y_2, \ldots, x_ny_n) \leq \max\{P(x_1^2, x_2^2, \ldots, x_n^2), P(y_1^2, y_2^2, \ldots, y_n^2)\}.$$

Proof. Let

$$P(X_1, \ldots, X_n) = \sum_{m \in \mathbb{N}^n} a_m X_1^{m_1} X_2^{m_2} \cdots X_n^{m_n}$$

for some $a_m \geq 0$. Then

$$\begin{aligned} P(x_1y_1, x_2y_2, \ldots, x_ny_n) &= \sum_{m \in \mathbb{N}^n} a_m (x_1y_1)^{m_1} (x_2y_2)^{m_2} \cdots (x_ny_n)^{m_n} \\ &= \sum_{m \in \mathbb{N}^n} (\sqrt{a_m} x_1^{m_1} x_2^{m_2} \cdots x_n^{m_n})(\sqrt{a_m} y_1^{m_1} y_2^{m_2} \cdots y_n^{m_n}) \\ &\leq \left(\sum_{m \in \mathbb{N}^n} a_m x_1^{2m_1} x_2^{2m_2} \cdots x_n^{2m_n} \right)^{1/2} \left(\sum_{m \in \mathbb{N}^n} a_m y_1^{2m_1} y_2^{2m_2} \cdots y_n^{2m_n} \right)^{1/2} \\ &\leq \max\{P(x_1^2, x_2^2, \ldots, x_n^2), P(y_1^2, y_2^2, \ldots, y_n^2)\}, \end{aligned}$$

where the first inequality follows from Cauchy-Schwarz. □

Lemma 10. *Let $\{Y_i\}_{i=1}^{\infty}$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A and stationary distribution μ so that Y_1 is distributed as μ, and let $\lambda = \lambda(Y)$.*

Let $f_1, \ldots, f_n : [N] \to [0,1]$, let $Z_i = f_i(Y_i)$ for all i, and let $\sigma_i = \mathbb{E}[f_i(v)]$, $\Phi = \sigma_1 + \cdots + \sigma_n$. For all $k \in [n]$, let $W_k \subseteq [n]^k$ be the set of all w such that $w_1 < w_2 < \cdots < w_k$. Then

$$\mathbb{E}\left[\sum_{w \in W_k} Z_{w_1} Z_{w_2} \cdots Z_{w_k}\right] \le \sum_{i=0}^{k-1} \binom{k-1}{i} \frac{\Phi^{i+1} \lambda^{k-i-1}}{(i+1)!(1-\lambda)^{k-i-1}} .$$

Proof. By Lemma 9 and Claim 6,

$$\begin{aligned}
\mathbb{E}\left[\sum_{w \in W_k} Z_{w_1} Z_{w_2} \cdots Z_{w_k}\right] &\le \sum_{w \in W_k} \sum_{s \in \{0,1\}^{k-1}} \sqrt{\sigma_{w_1}\sigma_{w_k}} \left(\prod_{i:s_i=0} \sqrt{\sigma_{w_i}\sigma_{w_{i+1}}}\right)\left(\prod_{i:s_i=1} \lambda^{w_{i+1}-w_i}\right) \\
&\le \max\Bigg\{ \sum_{w \in W_k} \sum_{s \in \{0,1\}^{k-1}} \sigma_{w_1} \prod_{i:s_i=0} \sigma_{w_{i+1}} \prod_{i:s_i=1} \lambda^{w_{i+1}-w_i}, \\
&\qquad \sum_{w \in W_k} \sum_{s \in \{0,1\}^{k-1}} \sigma_{w_k} \prod_{i:s_i=0} \sigma_{w_i} \prod_{i:s_i=1} \lambda^{w_{i+1}-w_i} \Bigg\}.
\end{aligned}$$

We assume for the remainder of the proof that the second term is the maximum. A similar proof holds under the assumption that the first term is the maximum.

We will show that for each $s \in \{0,1\}^{k-1}$,

$$\sum_{w \in W_k} \sigma_{w_k} \left(\prod_{i:s_i=0} \sigma_{w_i}\right)\left(\prod_{i:s_i=1} \lambda^{w_{i+1}-w_i}\right) \le \frac{(\sigma_1 + \cdots + \sigma_n)^{k-|s|}\lambda^{|s|}}{(k-|s|)!(1-\lambda)^{|s|}}, \tag{5.6}$$

where $|s|$ is the number of coordinates of s equal to 1. This proves the lemma, as there are $\binom{k-1}{k-j-1}$ vectors $s \in \{0,1\}^{k-1}$ such that $|s| = j$. From this point we fix s.

Let $w_{\bar{s}}$ be w restricted to the coordinates i such that $s_i = 0$ along with the kth coordinate.

Then

$$\sum_{w\in W_k}\sigma_{w_k}\left(\prod_{i:s_i=0}\sigma_{w_i}\right)\left(\prod_{i:s_i=1}\lambda^{w_{i+1}-w_i}\right) = \sum_{v\in W_{k-|s|}}\left(\prod_{j=1}^{k-|s|}\sigma_{v_j}\right)\left(\sum_{w:w_{\bar{s}}=v}\prod_{i:s_i=1}\lambda^{w_{i+1}-w_i}\right)$$
$$\leq \sum_{v\in W_{k-|s|}}\left(\prod_{j=1}^{k-|s|}\sigma_{v_j}\right)\left(\frac{\lambda}{1-\lambda}\right)^{|s|}.$$

where the last inequality uses the observation that the function that maps any $w \in W_{k-|s|}$ with $w_{\bar{s}} = v$ to the sequence of positive values $(w_{i+1}-w_i)_{i:s_i=1}$ is an injective function. Finally, Eq. (5.6) follows by noting that

$$\sum_{v\in\binom{[n]}{k-|s|}}\left(\prod_{j=1}^{k-|s|}\sigma_{v_j}\right) \leq \frac{(\sigma_1+\cdots+\sigma_n)^{k-|s|}}{(k-|s|)!}.$$

□

The following lemma gives an upper bound on the moments of S_n. We denote by $\left\{ n \atop k \right\}$ the Stirling number of the second kind. This counts the number of ways to partition a set of n labelled objects into k nonempty unlabelled subsets

Lemma 11. *Let $\{Y_i\}_{i=1}^{\infty}$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A and stationary distribution μ so that Y_1 is distributed as μ, and let $\lambda = \lambda(Y)$. Let $f_1, \ldots, f_n : V \to [0,1]$, let $Z_i = f_i(Y_i)$ for all i, and let $S_n = Z_1 + \cdots + Z_n$, and let $\sigma_i = \mathbb{E}[Z_i]$, $\Phi = \sigma_1 + \cdots + \sigma_n$. Then for all positive integers q,*

$$\mathbb{E}[S_n^q] \leq \sum_{k=1}^{q}\left\{ q \atop k \right\} k! \sum_{i=0}^{k-1}\binom{k-1}{i}\frac{\Phi^{i+1}\lambda^{k-i-1}}{(i+1)!(1-\lambda)^{k-i-1}}.$$

Proof. Consider the subset $D_k \subseteq [n]^q$ of vectors with exactly k distinct coordinates. Note that

$$\mathbb{E}[S_n^q] = \sum_{w \in [n]^q} \mathbb{E}\left[\prod_{j=1}^{q} Z_{w_j}\right] = \sum_{k=1}^{q} \sum_{w \in D_k} \mathbb{E}\left[\prod_{j=1}^{q} Z_{w_j}\right].$$

We will upper bound each term on the right-hand side separately.

Fix a k, and let $W_k \subseteq [n]^k$ be the set of vectors w so that $w_1 < w_2 < \cdots < w_k$. Let $\psi : D_k \to W_k$ be the function mapping each $w \in D_k$ to the vector whose coordinates are exactly those in w in sorted order and without repetition. Then because $Z_i \in [0, 1]$ for all i,

$$\sum_{w \in D_k} \mathbb{E}\left[\prod_{j=1}^{q} Z_{w_j}\right] \leq \sum_{w \in D_k} \mathbb{E}\left[\prod_{j=1}^{k} Z_{\psi(w)_j}\right]. \tag{5.7}$$

Moreover, for all $w \in W_k$ we have $|\psi^{-1}(w)| = \left\{ {q \atop k} \right\} k!$ (as this is the number of ways to partition q labeled balls into k nonempty labeled boxes), and thus Eq. (5.7) is equal to

$$\left\{ {q \atop k} \right\} k! \sum_{w \in W_k} \mathbb{E}\left[\prod_{j=1}^{k} Z_{w_j}\right].$$

The lemma then follows from Lemma 10. □

Finally we can insert the upper bounds from Lemma 11 in the Taylor expansion of α^{S_n} to prove Theorem 14.

Proof of Theorem 14. By Lemma 11,

$$\begin{aligned} \mathbb{E}[\alpha^{S_n}] &= \sum_{q=0}^{\infty} \frac{\log(\alpha)^q \mathbb{E}[S_n^q]}{q!} \\ &\leq 1 + \sum_{q=1}^{\infty} \frac{\log(\alpha)^q}{q!} \sum_{k=1}^{q} \left\{ {q \atop k} \right\} k! \sum_{i=1}^{k} \binom{k-1}{i-1} \frac{\Phi^i \lambda^{k-i}}{i!(1-\lambda)^{k-i}}. \end{aligned} \tag{5.8}$$

Rearranging the sums yields

$$1+\sum_{i=1}^{\infty}\frac{\Phi^i}{i!}\sum_{k=i}^{\infty}\binom{k-1}{i-1}\frac{\lambda^{k-i}}{(1-\lambda)^{k-i}}\sum_{q=k}^{\infty}\left\{ {q \atop k} \right\}\frac{\log(\alpha)^q k!}{q!}. \tag{5.9}$$

Using the following identity (Sta12, Eq. 1.94(b)),

$$\sum_{q=k}^{\infty}\left\{ {q \atop k} \right\}\frac{\log(\alpha)^q}{q!}=\frac{(\alpha-1)^k}{k!},$$

(which can be seen by writing $\alpha-1=e^{\log(\alpha)}-1$ as $\log(\alpha)+\frac{1}{2!}\log(\alpha)^2+\frac{1}{3!}\log(\alpha)^3+\cdots$) we can rewrite Eq. (5.9) as

$$1+\sum_{i=1}^{\infty}\frac{\Phi^i}{i!}(\alpha-1)^i\sum_{k=i}^{\infty}\binom{k-1}{i-1}\frac{\lambda^{k-i}}{(1-\lambda)^{k-i}}(\alpha-1)^{k-i}\ . \tag{5.10}$$

Using the following identity for $0\leq x<1$,

$$\sum_{j=i}^{\infty}\binom{j-1}{i-1}x^{j-i}=(1-x)^{-i},$$

(which follows from differentiating $\sum_{j=0}^{\infty}x^j=(1-x)^{-1}$ a total of $i-1$ times) we can rewrite Eq. (5.10) as

$$1+\sum_{i=1}^{\infty}\frac{\Phi^i(\alpha-1)^i}{i!}\left(1-\frac{\lambda(\alpha-1)}{1-\lambda}\right)^{-i}=\exp\left(\Phi\cdot\left(\frac{(1-\lambda)(\alpha-1)}{1-\alpha\lambda}\right)\right)\ .$$

□

5.3 Explicit constructions of resilient functions

We apply our expander sampler to construct resilient functions. Informally, these are Boolean functions for which any not too large subset of the inputs cannot affect the output of the function when the other inputs are set randomly. We define resiliency formally.

Definition 4. *Given a function $f : \{0,1\}^\ell \to \{0,1\}$, and $Q \subseteq [\ell]$, let $I_Q(f)$ be the probability that a random assignment of the variables in coordinates $[\ell]\backslash Q$ does not determine the value of f. Let $I_q(f) = \max_{Q\subseteq[\ell],|Q|\leq q} I_Q(f)$. We say that f is τ-strongly resilient if $I_q(f) \leq \tau \cdot q$ for all $q \leq \ell$.*

Note that a function that is a constant, or close to a constant, will by definition be resilient, and thus we restrict our attention to functions that are almost balanced, i.e., are 0 on about half the inputs, and 1 on the other half. Almost-balanced resilient functions were shown to exist by Ajtai and Linial (AL93). An explicit construction was shown by Meka (Mek17), who proved the following.

Theorem 15. *For some universal constants $c_1, c_2 \geq 1$, the following holds. For infinitely many ℓ, there exists an efficiently computable function $f : \{0,1\}^\ell \to \{0,1\}$ such that*

- *f is almost balanced, that is $\Pr_{x\in\{0,1\}}[f(x) = 1] = 1/2 \pm 1/10$.*
- *f is $(c_1(\log^2 \ell)/\ell)$-strongly resilient.*
- *f has a depth 3 monotone circuit of size at most ℓ^{c_2}.*

The proof in (Mek17) uses a sampler based on the extractors constructed in (Zuc97). We will show how to instead use expander samplers, leading to a somewhat simpler proof of the same theorem.

We start by outlining the general strategy used to prove Theorem 15. The function is constructed as follows. Let $\mathcal{P} = \{P^1, P^2, \ldots, P^u\}$ be a set of partitions of $[\ell]$ such that each P^α

is a collection of v sets, $P_1^\alpha, P_2^\alpha, \ldots, P_v^\alpha$, so that $|P_j^\alpha| = \ell/v$ for all j and $P_1^\alpha \cup P_2^\alpha \cup \cdots \cup P_N^\alpha = [\ell]$. The construction is the function $f_\mathcal{P} : \{0,1\}^\ell \to \{0,1\}$ defined as

$$f_\mathcal{P}(x) = \bigwedge_{\alpha \in [u]} \bigvee_{i \in [v]} \left(\bigwedge_{k \in P_i^\alpha} x_k \right). \tag{5.11}$$

To prove that $f_\mathcal{P}$ is almost balanced, Meka starts by showing that his choice of partitions is a design, which we define below.

Definition 5. *Let $\mathcal{P} = \{P^1, \ldots, P^u\}$ be a set of partitions of $[\ell]$ so that every partition has v parts, each of size w, and thus $\ell = v \cdot w$. For $d \leq w$, $\mathcal{P}$ is a d-design if for all $\alpha \neq \beta \in [u]$ and $i, j \in [v]$, $|P_i^\alpha \cap P_j^\beta| \leq w - d$. For $d \leq k < w$ and $\delta \in (0,1)$, we say that $\mathcal{P}$ is a (d,k,δ)-design if it is a d-design and for all $\alpha \in [u]$ and $i, j \in [v]$,*

$$\Pr_{\beta \sim [n]}[|P_i^\alpha \cap P_j^\beta| \leq w - k] \geq 1 - \delta.$$

Now, define the function

$$\mathrm{bias}(u,v,w) := (1 - (1 - 2^{-w})^v)^u$$

which can be interpreted as the probability that $f(x) = 1$ if the leaves of the formula in (5.11) (i.e., all occurrences of the x_k) are also independent and uniform over $\{0,1\}$. Meka proceeds by showing that if a set of partitions is a (d,k,δ)-design, and $\mathrm{bias}(u,v,w)$ is close to $1/2$, then so is the bias of the function $f_\mathcal{P}$. In particular, he shows the following.

Theorem 16. *Let $\mathcal{P} = \{P^1, \ldots, P^u\}$ be a set of partitions of $[\ell]$ so that every partition has v parts, each of size w, that is a (d,k,δ)-design. Assume $v = \theta(1)w2^w$ and $1/3 \leq$*

$\mathrm{bias}(u, v, w) \leq 2/3$. *Then,*

$$\left| \Pr_{x \sim \{0,1\}^\ell}[f_{\mathcal{P}} = 1] - \mathrm{bias}(u, v, w) \right| \leq C\bigg(w \exp(-\Omega(k)) + \exp(-\Omega(d)) + 2^w \delta \bigg).$$

To prove that $f_{\mathcal{P}}$ is resilient, Meka shows that his choice of partitions is load-balancing, which we define below.

Definition 6. *Let $\mathcal{P} = \{P^1, \ldots, P^u\}$ be a set of partitions of $[\ell]$ so that every partition has v parts, each of size w, and thus $\ell = v \cdot w$. $\mathcal{P}$ is (q, t)-load balancing if for all $Q \subseteq [\ell]$ with $|Q| \leq q$ and $j \in [v]$,*

$$\mathbb{E}_{\alpha \sim [u]} \left[\nVdash (Q \cap P_j^\alpha \neq \emptyset) 2^{|Q \cap P_j^\alpha|} \right] \leq \frac{tq}{v}.$$

Note that $\mathbb{E}_{\alpha \sim [u]}[|Q \cap P_j^\alpha|] = q/v$. In Theorem 2.5 in (Mek17), it was shown that if a set of partitions is load-balancing, then there is a bound on $I_q(f_{\mathcal{P}})$. In particular, we have the following.

Theorem 17. *Let $\mathcal{P} = \{P^1, \ldots, P^u\}$ be a set of partitions of $[\ell]$ so that every partition has v parts, each of size w, and is (q, t)-load balancing. Then $f_{\mathcal{P}}$ is $(u(1 - 2^{-w})^{v-q}) \cdot (t2^{-w})$-strongly resilient.*

5.3.1 Our construction

Our construction differs from Meka's only in the choice of partitions $\mathcal{P}$, and thus it is enough to prove that our partitions are also a design and are load balancing. Let c_1, c_2, λ and d be constants to be chosen below. Let $G = (\mathbb{Z}_N, E)$ be a d-regular undirected graph with vertex set $\mathbb{Z}_N$ for some prime N to be chosen later. Let $\mathcal{T} \subset \mathbb{Z}_N^n$ be the set of paths of length n on G. Define the function $r : \mathbb{Z}_N \to \mathbb{Z}_N^{c_1}$ by

$$r(v) = (v \bmod N, 2v \bmod N, \ldots, c_1 v \bmod N),$$

and let $\mathcal{U} = \{(r(t_1), r(t_2), \ldots, r(t_n)) : t \in \mathcal{T}\}$ be a subset of vectors of $\mathbb{Z}_N^{nc_1}$, obtained by concatenating the result of r on the vertices of paths in $\mathcal{T}$.

For $i \in \mathbb{Z}_N$ and $\alpha \in \mathcal{U}$, let

$$P_i^\alpha = \{(k-1)N + ((i - \alpha_k) \bmod N) : k \in [nc_1]\} \subseteq \mathbb{Z}_{Nnc_1},$$

so that P^α is a partition of the set $\mathbb{Z}_{Nnc_1}$, and thus our function will be on the set $\{0,1\}^{Nnc_1}$. We will refer to each set P_i^α as a part of the partition P^α. We let $\mathcal{P}(G) = \{P^\alpha : \alpha \in \mathcal{U}\}$ be the collection of partitions constructed from a graph G.

Parameters

Our construction depends on the constants c_1, c_2, λ and d, and also n and N. Below, we describe how to choose c_1, c_2, λ and d. For the rest of the section, we assume that these conditions hold.

- c_1 is a universal constant chosen in the proof of Corollary 7.
- $\lambda = \dfrac{1}{8^{2c_1}}$.
- $d = O(1/\lambda^2)$ is the degree of expander graphs with spectral gap λ
- c_2 is the constant from Claim 7 below, and does not depend on anything.

It is known that there exist d-regular expander graphs G of any large enough size so that $d = O(1/\lambda^2)$ for all λ (for example, the expander graphs constructed in (LPS88). Thus, there exist constant factors so that the first three dependencies hold.

We allow n to be any integer, and we choose N according to the following claim.

Claim 7. *There exists a constant $c_2 < 1$ such that for all n and all d, the following holds. There exists a prime number N such that for $u = Nd^{n-1}$,*

$$0 \leq N - 2^{c_1 n}(\ln(u/\ln 2)) \leq 2^{c_2 c_1 n}.$$

Proof. Consider the function

$$\phi(x) = x - 2^{c_1 n}(\ln x + (n-1)\ln d + \ln\ln 2).$$

Because $\phi(x)$ is continuous, there exists an $x^* \geq 1$ so that $\phi(x^*) = 0$, and we let N be the next largest prime. Note that $x^* \leq 2(c_1 + \ln d)n2^{c_1 n}$ as $\phi(x^*)$ is positive for this x^*. By results on gaps between primes (see for example, (BHP01)), there exists a prime number in the range $[x^*, x^* + 2^{c_2 c_1 n}]$ for some universal constant c_2 less than 1. Thus if we assign to N this prime number, $\phi(x^*) = 0 \leq N - 2^{c_1 n}(\ln(u/\ln 2)) = \phi(N) \leq N - x^* \leq 2^{c_2 c_1 n}$. □

Claim 7 can be used to apply the following fact, which will be used both to prove that $f_{\mathcal{P}(G)}$ is almost-balanced and strongly resilient.

Fact 1. *Let $n \leq N \leq u$, $B \geq 1$ and $0 \leq N - 2^n(\ln(u/\ln 2)) \leq B$ and $\theta = (1 - 2^{-n})^N$. Then $(1+\theta)^u = O(1)$ and $(1-\theta)^u = 1/2 \pm O(B \ln u)2^{-n}$.*

5.3.2 Strongly resilient

To prove that $f_{\mathcal{P}(G)}$ is resilient, we will use the same general strategy as in (Mek17) but use the result of Theorem 14 in place of (Mek17, Theorem 1.8). In particular, we will show that $\mathcal{P}(G)$ is load balancing, and then apply Theorem 17.

Lemma 12. *Let $G = (\mathbb{Z}_N, E)$ be a graph, let $\lambda = \lambda(G)$. Then $\mathcal{P}(G)$ is (q,t)-load balancing* *for $q \leq N$ and*

$$t = \exp\left((2^{c_1} - 1)\left(\frac{1-\lambda}{1-2^{c_1}\lambda}\right)\right).$$

Proof. Fix a $Q \subseteq \mathbb{Z}_{Nnc_1}$ so that $|Q| = q$, and fix $j \in \mathbb{Z}_N$. Let α be a uniformly random element of $\mathcal{U}$, and define the random variable

$$g_i = \big|\{k : j - \alpha_k \in Q \cap \{(k-1)N, (k-1)N+1, \ldots, kN-1\} \text{ and } k \in [(i-1)c_1 + 1, ic_1]\}\big|.$$

Then,

$$S_n := |Q \cap P_j^\alpha| = g_1 + \cdots + g_n.$$

and

$$\begin{aligned}
\mathbb{E}\big(\nVdash(Q \cap P_j^\alpha \neq \emptyset)2^{|Q \cap P_j^\alpha|}\big) &\leq \mathbb{E}[2^{S_n}] - 1 + \mathbb{E}[S_n] \\
&\leq \exp\left((2^{c_1} - 1)\frac{|Q|}{N}\left(\frac{1-\lambda}{1-2^{c_1}\lambda}\right)\right) - 1 + \frac{|Q|}{N} \\
&\leq \frac{|Q|}{N}\exp\left((2^{c_1} - 1)\left(\frac{1-\lambda}{1-2^{c_1}\lambda}\right)\right)
\end{aligned}$$

as desired, where the second inequality follows from Theorem 14 and noting that g_i is bounded above by c_1, and the third inequality follows because $|Q|/N \leq 1$. Note that because $\lambda < 2^{-c_1}$, the conditions of Theorem 14 hold. □

We can now conclude that $f_{\mathcal{P}(G)}$ is strongly-resilient.

Corollary 5. *Let n be any integer, and let N be chosen according to Claim 7. There exists a graph $G = (\mathbb{Z}_N, E)$ such that the function $f_{\mathcal{P}(G)} : \{0,1\}^{nNc_1} \to \{0,1\}$ is $O((\log^2 \ell)/\ell)$-strongly resilient.*

Proof. By Lemma 12 and Theorem 17, $f_{\mathcal{P}(G)}$ is $(u(1-2^{-nc_1})^{N-q}) \cdot (t2^{-nc_1})$-strongly resilient for some constant t. For $q = O(2^{nc_1})$, it holds that $u(1-2^{-nc_1})^{N-q} = O(1)$, which follows from applying Fact 1 and Claim 7. The corollary follows in this case by noting that $2^{nc_1} = O(\ell/(\log^2 \ell))$. When $q \geq 2^{nc_1}$, the corollary holds trivially. □

5.3.3 Almost balanced

To prove that $f_{\mathcal{P}(G)}$ is almost balanced, we will also use the same general strategy as in (Mek17). In particular, we will show that $\mathcal{P}(G)$ is a design, and then apply Theorem 16. To aid in the proof that the set of partitions $\mathcal{P}(G)$ described is a design, we also define the following distance $d_H(x, x')$.

Definition 7. *For two sequences $x, x' \in \mathbb{Z}_N^\ell$, let $d_H(x, x') = \min_{a \in \mathbb{Z}_N} |\{i \in [d] : x_i - x'_i \neq a \bmod N\}|$.*

We now prove that $\mathcal{P}$ is in fact a design.

Lemma 13. *Let $G = (\mathbb{Z}_N, E)$ be a graph such that $\lambda(G) = \lambda$. Then if $\lambda \leq 1/4$ and $n \leq \sqrt{N}$, then $\mathcal{P}(G)$ is a $(c_1 - 1, n(c_1 - 1)/2, e^{1/\lambda}(2\lambda)^{n/2})$-design.*

Proof. As in (Mek17), we note that for $\alpha, \beta \in \mathbb{Z}_N^n$ and $i, j \in \mathbb{Z}_N$, it holds that

$$|P_i^\alpha \cap P_j^\beta| = |\{k \in [n] : \beta_k - \alpha_k = (j - i) \bmod N\}| \leq w - d_H(\alpha, \beta).$$

Note that if $v, u \in \mathbb{Z}_N$ are distinct, then $d_H(r(v), r(u)) \geq c_1 - 1$, and thus $d_H(\alpha, \beta) \geq c_1 - 1$ for all distinct $\alpha, \beta \in \mathcal{U}$.

Now fix a path $t \in \mathcal{T}$. By Corollary 4, the probability that a random walk $(Y_1, \ldots, Y_n) \in \mathcal{T}$ agrees with t in at least $n/2$ coordinates is bounded above by

$$\Big(\frac{1}{\lambda} - \frac{1-\lambda}{(N/2)^{1/2}\lambda^{3/2}}\Big)^{-n/2} \exp\Big(\frac{n}{N} \cdot (\sqrt{N\lambda/2} - 1)\frac{1-\lambda}{\lambda}\Big) \leq e^{1/\lambda}(2\lambda)^{n/2},$$

where the inequality follows from the conditions set on λ and n. To see this, let the function

g_i be the indicator function for t_i, that is $g_i(Y_i) = 1$ if $Y_i = t_i$, and is 0 otherwise. Then, $S_n = g_1(Y_1) + \cdots + f_n(Y_n)$ is the number of vertices that agree with t, and $\mathbb{E}[S_n] = n/N$. If $S_n \leq n/2$, then $d_H((r(t_1), \ldots, r(t_n)), ((r(s_1), \ldots, r(s_n)) \geq n(c_1 - 1)/2$. □

Now we can apply Theorem 16 to get the following.

Corollary 6. *For any n and d, let N be as in Claim 7, and $u = |\mathcal{U}|$, and assume that $1/3 \leq \text{bias}(u, N, nc_1) \leq 2/3$. Let $G = (\mathbb{Z}_N, E)$ be a d-regular graph such that $\lambda(G) = \lambda$ for $\lambda \leq 1/4$. Then*

$$\left|\Pr_{x \sim \{0,1\}^{Nn}}[f_{\mathcal{P}(G)} = 1] - \text{bias}(u, N, nc_1)\right| \leq$$
$$C\Bigg(nc_1 \exp\Bigg(-\Omega\Bigg(\frac{n(c_1-1)}{4}\Bigg)\Bigg) + \exp(-\Omega(c_1 - 1)) + e^{1/\lambda}(4^{c_1}\lambda^{1/2})^n\Bigg).$$

Proof. Apply Lemma 13 and Theorem 16. □

Finally, we prove that $f_{\mathcal{P}(G)}$ is almost balanced, using a bound on $\text{bias}(u, N, nc_1)$ that follows from Corollary 6 and Claim 7.

Corollary 7. *Let n be any integer, and let N be chosen according to Claim 7. There exists a graph $G = (\mathbb{Z}_N, E)$ such that the function $f_{\mathcal{P}(G)} : \{0,1\}^{nNc_1} \to \{0,1\}$ has the property that*

$$\Pr_{x \in \{0,1\}}[f_{\mathcal{P}(G)}(x) = 1] = 1/2 \pm c.$$

for some constant $c < 1/10$.

Proof. If we let $\theta = (1 - 2^{-nc_1})^N$, then

$$\text{bias}(u, N, nc_1) = (1 - \theta)^u = 1/2 \pm o(1)$$

as desired by Fact 1 and Claim 7. The corollary follows from this, Corollary 6, and the fact that c_1 can be made arbitrarily large. □

Chapter 6

The Littlewood-Offord Problem for Markov Chains

Let $v_1, \ldots, v_n \in \mathbb{R}^d$ be fixed vectors of Euclidean length at least 1, and let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher random variables, so that $\Pr[\varepsilon_i = 1] = \Pr[\varepsilon_i = -1] = 1/2$ for all i. The celebrated Littlewood-Offord problem (LO43) asks for an upper bound on the probability,

$$\Pr[\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n \in B] \tag{6.1}$$

for an open Euclidean ball B with radius 1. This question was first investigated by Littlewood and Offord for the case $d = 1$ and $d = 2$ (LO43). A tight bound of $\binom{n}{n/2}/2^n = \Theta(1/\sqrt{n})$ when n is even, with the worst case being when the vectors are equal, was found by Erdős for the case $d = 1$ using a clever combinatorial argument (Erd45). Such bounds can be contrasted with concentration inequalities like the Hoeffding inequality in the scalar case and the Khintchine-Kahane inequality in the vector case, both of which give an upper bound on the probability $\Pr[\|\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n\| \geq k\sqrt{n}]$ for positive k. In contrast, an upper bound on Eq. (6.1) can be considered a form of anti-concentration, that is showing that the random sum is unlikely to be in B.

In the case that the v_i are d-dimensional vectors, a tight bound up to constant factors of $C/\sqrt{n}$ was found by Kleitman (Kle70), and was improved by series of work (Sal83; Sal85; FF88; TV12). In the scalar case, under the restriction that $v_1, \dots, v_n$ are distinct integers, an upper bound of $n^{-3/2}$ was found by Sárközy and Szemeredi (SS65).

In this chapter, we investigate the case in which $\varepsilon_1, \dots, \varepsilon_n$ are not independent, but are obtained from a stationary reversible Markov chain $\{Y_i\}_{i=1}^{\infty}$ with state space $[N]$ and transition matrix A, and functions $f_1, \dots, f_n : [N] \to \{-1, 1\}$, using $\varepsilon_i = f_i(Y_i)$.

Let μ be the stationary distribution for the Markov chain, and let E_μ be the associated averaging operator defined by $(E_\mu)_{ij} = \mu_j$, so that for $v \in \mathbb{R}^N$, $E_\mu v = \mathbb{E}_\mu[v]\mathbf{1}$ where $\mathbf{1}$ is the vector whose entries are all 1. As in the rest of this thesis, our generalizations will be in terms of the quantity

$$\lambda = \|A - E_\mu\|_{L_2(\mu) \to L_2(\mu)}.$$

We show that the Littlewood-Offord problem can also be generalized to Markov chains with an extra dependence on λ, for all dimensions. We additionally consider the one-dimensional case when the scalars are distinct integers. In all cases, the proof is based off a Fourier-analytic argument due to Halász (Hal77).

The random variables in all cases are defined in the same way, which we state below.

Setting 1. *Let $\{Y_i\}_{i=1}^{\infty}$ be a stationary reversible Markov chain with state space $[N]$, transition matrix A, stationary probability measure μ, and averaging operator E_μ so that Y_1 is distributed according to μ. Let $\lambda = \|A - E_\mu\|_{L_2(\mu) \to L_2(\mu)}$, and let $f_1, \dots, f_n : [N] \to \{-1, 1\}$ be such that $\mathbb{E}[f_i(Y_i)] = 0$ for every i. Then consider the random variables $f_1(Y_1), f_2(Y_2), \dots, f_n(Y_n)$.*

We obtain the following theorem that upper bounds the probability that the random sum is concentrated on any unit ball. In the case that the v_i are one-dimensional, the bound is tight up to a factor of $\sqrt{(1-\lambda)/(1+\lambda)}$ in λ. Note that the bound depends on the dimension, while in the independent case, there is no dependence on the dimension.

Theorem 18. *Assume the setting of 1. Let $x_0 \in \mathbb{R}^d$ and $R \geq \frac{1}{C\sqrt{d}}$ for some universal constant C'. For every set of vectors $v_1, \ldots, v_n \in \mathbb{R}^d$ of Euclidean length at least* 1*,*

$$\Pr[\|f_1(Y_1)v_1 + f_2(Y_2)v_2 + \cdots + f_n(Y_n)v_n - x_0\|_{\ell_2} \leq R] \leq \frac{C \cdot R\sqrt{d}}{(1-\lambda)\sqrt{n}}.$$

for some universal constant C.

In the one-dimensional case, we also consider the restriction that $v_1, \ldots, v_n$ are *distinct* integers.

Theorem 19. *Assume the setting of 1. Then for every set of distinct integers $v_1, \ldots, v_n \geq 1$ and $b \in \mathbb{Z}$,*

$$\Pr[f_1(Y_1)v_1 + f_2(Y_2)v_2 + \cdots + f_n(Y_n)v_n = b] \leq \frac{C}{(1-\lambda)^3 n^{3/2}}$$

for some universal constant C.

Finally, we consider a different setting, where rather than choosing $\varepsilon_1, \ldots, \varepsilon_n$ independently, we choose these uniformly at random from a subset D of $\{-1, 1\}^n$ that we can construct explicitly.

Theorem 20. *For every n, there exists an explicit set $D \subseteq \{-1, 1\}^n$ of cardinality at most $2^{C_1\sqrt{n}}$ for some universal constant C_1 such that the following holds. For every $v_1, \ldots, v_n \geq 1$ and $b \in \mathbb{R}$ and ε chosen uniformly at random from D*

$$\Pr[|\varepsilon_1 v_1 + \varepsilon_2 v_2 + \cdots + \varepsilon_n v_n - b| \leq 1] \leq \frac{C}{\sqrt{n}}.$$

for some universal constant C independent of n.

One interpretation of Theorem 20 is that one can obtain similar results as in the Littlewood-Offord problem in one dimension using much less randomness, and in particular, using $C_1\sqrt{n}$ bits of randomness rather than n.

This setting was also considered in (KKL17), in which they were able to construct an explicit set of cardinality $n2^{n^c}$, from which a random sample satisfies

$$\Pr[f_1(Y_1)v_1 + f_2(Y_2)v_2 + \cdots + f_n(Y_n)v_n = b] \leq \frac{\log(n)^{C_1/c}}{\sqrt{n}}.$$

for any constant c bounded above by 1. Sampling from the set in Theorem 20 guarantees a stronger bound on the probability that the sum lands in any interval, while requiring more randomness when $c < 1/2$.

6.1 The Littlewood-Offord problem for independent random variables

As warm up, we present the bound in the independent case for 1-dimensional vectors, or scalars. These calculations will be used later in the proofs of Theorems 18, 19, and 20,. This bound was first proved by Erdős (Erd45) who used a clever combinatorial argument that applies Sperner's theorem. The proof we present is in spirit, due to Halász (Hal77) and is based on techniques from Fourier analysis.

We start by presenting the following concentration inequality due to Esséen (Ess66), which will allow us to upper-bound probabilities. This inequality is in the spirit of Fourier inversion, but written in a way that can be more readily applied for our purposes.

Theorem 21 (Esséen concentration inequality). *Let $X \in \mathbb{R}^d$ be a random variable taking a* *finite number of values. For $R, \varepsilon > 0$,*

$$\sup_{x_0 \in \mathbb{R}^d} \Pr\left[\|X - x_0\|_{\ell_2} \leq R\right] = O\left(\frac{R}{\sqrt{d}} + \frac{\sqrt{d}}{\varepsilon}\right)^d \int_{\xi \in \mathbb{R}^d \colon \|\xi\|_{\ell_2} \leq \varepsilon} |\mathbb{E}[\exp(2\pi i \langle \xi, X \rangle)]| \, d\xi.$$

The following bound is implicit in the proof of Proposition 7.18 in (TV06) and will be used to further bound the quantities obtained from Theorem 21

Claim 8. *Let $v_1, \ldots, v_k \in \mathbb{R}$ be such that $|v_j| \geq 1$ for all j. Then*

$$\int_{-1}^{1} \left(\prod_{j \in k} |\cos(2\pi \xi v_j)| \right) d\xi \leq \frac{C}{\sqrt{|k|}},$$

for some constant C.

We now prove the bound in the independent case.

Theorem 22. *Let $v_1, \ldots, v_n \in \mathbb{R}$ be non-zero, and let $\varepsilon_1, \ldots, \varepsilon_n$ be independent random variables uniform over the set $\{-1, 1\}$. Then for all $x_0 \in \mathbb{R}$,*

$$\Pr[|\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n - x_0| \leq 1] \leq \frac{C}{\sqrt{n}}.$$

for some constant C independent of n.

Proof. By Theorem 21, the left-hand side can be bounded above by

$$\begin{aligned} C_1 \int_{-1}^{1} |\mathbb{E}[\exp(2\pi i \xi (\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n))]| \, d\xi &= C_1 \int_{-1}^{1} \prod_{j=1}^{n} |\mathbb{E}[\exp(2\pi i \xi \varepsilon_j v_j)]| \, d\xi \\ &= C_1 \int_{-1}^{1} \prod_{j=1}^{n} |\cos(2\pi \xi v_j)| \, d\xi \qquad (6.2) \\ &\leq \frac{C_2}{\sqrt{n}} \end{aligned}$$

for some constants C_1 and C_2. The first equality follows from the independence of the ε_j, the next equality follows from the fact that ε_j is uniform over $\{-1, 1\}$ for all j, and the subsequent inequality follows from Claim 8. □

6.2 The Littlewood-Offord Problem for Random Variables from a Markov chain

Now we consider the case that $\varepsilon_1, \dots, \varepsilon_n$ are obtained from a Markov chain. The proof follows very closely the proof for independent random variables in Proposition 7.18 in (TV06) which itself is due to Halász (Hal77).

Before proving Theorem 18, we first prove the following that will allow us to upper-bound negative moments of binomial random variables.

Claim 9. *Let $x = B(n, p)$ be a binomial random variable with n trials, each with success probability $p > 0$. Then for all positive integers d,*

$$\mathbb{E}\left[\frac{1}{(x+1)^d}\right] \leq \frac{d^d}{n^d p^d}.$$

Proof. Note that because $d(i+1) \geq i+d$ for all non-negative i, the right-hand side is bounded above by $d^d \mathbb{E}\left[\frac{x!}{(x+d)!}\right]$, where the term inside the expected value can be written as

$$\begin{aligned}
\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \frac{i!}{(i+d)!} &= \sum_{i=0}^{n} \frac{n!}{(n-i)!(i+d)!} p^i (1-p)^{n-i} \\
&= \sum_{i=0}^{n} \binom{n+d}{i+d} p^{i+d} (1-p)^{n-i} \frac{n!}{(n+d)! p^d} \\
&\leq \frac{n!}{(n+d)! p^d}.
\end{aligned}$$

The claim follows by noting that $n \leq n+i$ for $1 \leq i \leq d$. □

We start by considering the case of 1-dimensional vectors, or scalars. We allow in this case for at most one-half of the v_i to have length less than 1. This will allow us to generalize to higher dimensions. We note that in the case of independent random variables the corresponding statement follows from the usual Littlewood-Offord problem, by conditioning on the ε_i such that $|v_i| < 1$, for just an increase in the constant factor in the bound.

Lemma 14. *Assume the setting of 1. Then for every $v_1, \ldots, v_n \in \mathbb{R}$ such that $|\{i : |v_i| \geq 1\}| \geq n/2$ and $x_0 \in \mathbb{R}$,*

$$\Pr[|f_1(Y_1)v_1 + f_2(Y_2)v_2 + \cdots + f_n(Y_n)v_n - x_0| \leq 1] \leq \frac{C}{(1-\lambda)\sqrt{n}}.$$

for some universal constant C.

Proof. By Theorem 21,

$$\Pr[|f_1(Y_1)v_1 + \cdots + f_n(Y_n)v_n - x_0| \leq 1] \leq \\ C_1 \int_{-1}^{1} |\mathbb{E}[\exp(2\pi i\xi(f_1(Y_1)v_1 + \cdots + f_n(Y_n)v_n))]| \, d\xi \quad (6.3)$$

for some constant C_1. Note that

$$\mathbb{E}[\exp(2\pi i\xi(f_1(Y_1)v_1 + \cdots + f_n(Y_n)v_n))] = \mathbb{E}\left[\prod_{j=1}^{n} \exp(2\pi i\xi f_j(Y_j)v_i)\right]. \quad (6.4)$$

Let $T_j = A - (1-\lambda)E_\mu$, let u_j be the vector defined by $u_j(y) = \exp(2\pi i\xi f_j(y)v_j)$ for $y \in [N]$, and let $U_j = \mathrm{diag}(u_j)$. For $s \in \{0,1\}^{n-1}$, let $t(s)$ be the set of indices j such that $s_{j-1} = s_j = 0$, and also includes 1 if $s_1 = 0$ and includes n if $s_{n-1} = 0$. Then the right-hand side of Eq. (6.4) is bounded above by

$$\|U_1(T_1+(1-\lambda)E_\mu)U_2(T_2+(1-\lambda)E_\mu)U_3\cdots U_{n-1}(T_{n-1}+(1-\lambda)E_\mu)U_n\mathbf{1}\|_{L_1(\mu)} \leq \sum_{s\in\{0,1\}^{n-1}}\left(\prod_{j:s_j=1}\lambda\right)\left(\prod_{j:s_j=0}(1-\lambda)\right)\left(\prod_{j\in t(s)}|\cos(2\pi\xi v_j)|\right),$$

where the inequality follows by Eq. (2.7) in Lemma 2 in Chapter 2 and evaluating $|\langle\mu,u\rangle_{L_2(\mu)}|$.

Let $t'(s)$ be the set of indices $j\in t(s)$ such that $|v_j|$ is greater than 1. When $|t'(s)|=0$, the corresponding product disappears. When $|t'(s)|>0$, we can apply Claim 8. Thus, the right-hand side of Eq. (6.3) can be bounded above by

$$C_1\sum_{s\in\{0,1\}^{n-1}}\left(\prod_{j:s_j=1}\lambda\right)\left(\prod_{j:s_j=0}(1-\lambda)\right)\frac{C_2}{\sqrt{|t'(s)|+1}}. \tag{6.5}$$

Let $r:\{0,1\}^{n-1}\to[n-1]$ be defined as

$$r=\left|\{j:s_j=s_{j+1}=0\text{ and }|v_j|\geq 1\}\right|,$$

so that $r(s)\leq|t'(s)|$ for all $s\in\{0,1\}^{n-1}$. Let $\mathbf{s}$ be a random vector from $\{0,1\}^{n-1}$ so that for each $s\in\{0,1\}^{n-1}$

$$\Pr[\mathbf{s}=s]=\left(\prod_{j:s_j=1}\lambda\right)\left(\prod_{j:s_j=0}(1-\lambda)\right).$$

By the definition of r and $\mathbf{s}$, the right-hand side of Eq. (6.5) is bounded above by,

$$C_1\mathbb{E}\left[\frac{C_2}{\sqrt{r(\mathbf{s})+1}}\right].$$

We conclude with the following argument. Let $r'=B(\lfloor n/4\rfloor-1,(1-\lambda)^2)+1$ where

$B(n,p)$ denotes a binomial random variable with n trials, each with success probability p. It follows that r' is dominated by $r(\mathbf{s})+1$, and thus

$$\mathbb{E}\left[\frac{C}{\sqrt{r(\mathbf{s})+1}}\right] \leq \mathbb{E}\left[\frac{C}{\sqrt{r'}}\right] \leq \left(\mathbb{E}\left[\frac{C^2}{r'}\right]\right)^{1/2}, \tag{6.6}$$

where the second inequality follows by Jensen's inequality. Finally, by Claim 9, the right-hand side of Eq. (6.6) is bounded above by $C\left((1-\lambda)\sqrt{\lfloor n/4 \rfloor}\right)^{-1}$ as desired. □

Before proving Theorem 18, we prove the following bound on random unit vectors.

Claim 10. *Let $v \in \mathbb{R}^d$ be a random unit vector uniform over the $d-1$-dimensional sphere. Then there exists a constant C such that*

$$\Pr\left[|v_1| \geq \frac{1}{C\sqrt{d}}\right] \geq \frac{1}{2}$$

Proof. We start by noting that the probability density function of v_1 at t is proportional to $(1-t^2)^{(d-3)/2}$, which is also the probability density of the beta distribution, shifted so that the domain is $[-1,1]$. The probability density function at all points is bounded above by

$$\frac{1}{2^{d-3}} \cdot \frac{\Gamma(d-1)}{\Gamma((d-1)/2)^2} \leq \frac{1}{2^{d-3}} \cdot \frac{C_1(d-1)^{d-3/2}e^{-d+2}}{C_1^2((d-1)/2)^{d-2}e^{-d+1}} \leq C_2\sqrt{d-1}$$

for some constants C_1 and C_2, where the inequality follows from Stirling's approximation (see (Jam15)). The claim follows by letting $C = C_2/4$. □

We now use Lemma 14 to prove Theorem 18 as follows.

Proof of Theorem 18. Let $A \in \mathcal{SO}(d)$ be a random rotation uniform over the Haar measure of the special orthogonal group. Then it is enough to consider the random variable $\|Af_1(Y_1)v_1 + \cdots + Af_n(Y_n)v_n - Ax_0\|_{\ell_2}$. Additionally, the left-hand side in the statement of the theorem is bounded above by

$$\Pr[|(Af_1(Y_1)v_1 + \cdots + Af_n(Y_n)v_n - Ax_0)_1| \leq R]. \tag{6.7}$$

This is because if the absolute value of the first coordinate of the random vector is greater than R, so is the Euclidean norm.

By Claim 10, for any fixed d, it holds that $|f_i(Y_i)v_i| \geq 1/(C'\sqrt{d})$ for at least half of the i for some constant C'. By Lemma 14, we have that Eq. (6.7) is bounded above by

$$C' \cdot R\sqrt{d} \sup_{x_0 \in \mathbb{R}} \Pr\left[|(Af_1(Y_1)v_1 + \cdots + Af_n(Y_n)v_n - x_0)_1| \leq \frac{1}{C'\sqrt{d}}\right] \leq \frac{C \cdot R\sqrt{d}}{(1-\lambda)\sqrt{n}}$$

as desired. □

Remark 1. *For scalars, Theorem 18 is tight up to a factor of* $\sqrt{(1-\lambda)/(1+\lambda)}$. *To see this, consider the transition matrix on two states defined by*

$$A = \begin{pmatrix} \frac{1-\lambda}{2} & \frac{1+\lambda}{2} \\ \frac{1+\lambda}{2} & \frac{1-\lambda}{2} \end{pmatrix}$$

with $f(1) = 1$ *and* $f(2) = -1$, *and stationary distribution uniform over both states. Such a Markov chain can be interpreted as first choosing a state at random, and then at each subsequent step choosing a new state uniformly at random with probability* $1 - \lambda$, *or switching states with probability* λ. *We can associate with this walk a sequence of numbers,* $(X_1, X_2, \ldots)$ *obtained as follows. Whenever a state is chosen at random, we add a new entry in the sequence starting at* 1, *and increase this entry every time the state is switched. Then conditioned on this sequence,* $f(Y_1) + f(Y_2) + \cdots + f(Y_n)$ *is distributed as* $\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{\mathbf{n}}$ *where* $\mathbf{n}$ *is the* *number of entries in the sequence that are odd. Thus, if* $\mathbf{n}$ *is considered as a random variable,*

$$\Pr[f(Y_1) + f(Y_2) + \cdots + f(Y_n) = 0] \leq \mathbb{E}\left[\frac{C}{\sqrt{\mathbf{n}}}\right]$$

If we assume that n *is large, then the probability that any given step in the walk is the start of a entry that will eventually be of odd length is approximately* $1/(1+\lambda)$*, and thus,* $\mathbf{n}$ *is approximately distributed like* $B(n, (1-\lambda)/(1+\lambda))$*, and thus*

$$\mathbb{E}\left[\frac{C}{\sqrt{\mathbf{n}}}\right] \leq \frac{C}{\sqrt{(1-\lambda)n/(1+\lambda)}}$$

6.3 Extension to distinct v_i's

Theorem 22, the bound obtained in the independent case, is tight when $v_1 = \cdots = v_n = 1$. It is reasonable to ask if one can obtain better bounds on the probability $\Pr[\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n \in B]$ under certain restrictions of $v_1, \ldots, v_n$. In particular, when the v_i are distinct integers, Sárközy and Szemeredi (SS65) showed that for all x_0 and for some constant C

$$\Pr[\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n = b] \leq \frac{C}{n^{3/2}}, \tag{6.8}$$

which is a factor n smaller than Theorem 22.

Like Erdős's proof of Theorem 22, the proof of the above by Sárközy and Szemeredi uses a clever combinatorial argument. However, Halász's Fourier-analytic argument can also be used to prove the above. A similar bound can be achieved in the case of Markov Chains, as in Theorem 18.

Our proof is based on the techniques used in (TV06) for the same problem, in which the Fourier-analytic argument is over the group $\mathbb{Z}_p$ for some large enough p, rather than over the integers or over the real numbers. The following claim is implicit in Corollary 7.16 in (TV06) and will be used in our computation.

Claim 11. *If $v_1, \ldots, v_n$ are distinct positive integers, then there exists a prime p such that $p \geq v_i$ for all i, and*

$$\frac{1}{p} \sum_{\xi \in \mathbb{Z}_p} \left[\prod_{i=1}^{n} |\cos(2\pi\xi \cdot v_i)| \right] \leq \frac{C}{n^{3/2}}.$$

We use Claim 11 to prove Theorem 19 which is a Markov chain version of Eq. (6.8).

Proof of Theorem 19. Let p be the prime in Claim 11. Note that by Fourier inversion,

$$\begin{aligned}
&\Pr[f_1(Y_1)v_1 + f_2(Y_2)v_2 + \cdots + f_n(Y_n)v_n = x_0] \\
&\quad \leq \Pr[f_1(Y_1)v_1 + f_2(Y_2)v_2 + \cdots + f_n(Y_n)v_n \equiv x_0 \bmod p] \\
&\quad = \frac{1}{p} \sum_{\xi \in \mathbb{Z}_p} \left| \exp\left(-\frac{2\pi i}{N}\xi \cdot x_0\right) \mathbb{E}\left[\exp\left(\frac{2\pi i}{N}\xi \cdot (f(Y_1)v_1 + f(Y_2)v_2 + \cdots + f(Y_n)v_n)\right)\right] \right|.
\end{aligned} \tag{6.9}$$

Let $T_j = A - (1-\lambda)E_\mu$ for all j, and let u_i be the vector defined by $u_j(y) = \exp(2\pi i(\xi \cdot f_j(y)v_j)/N)$. Then the absolute value of the expectation inside the right-hand side of Eq. (6.9) is bounded above by

$$\|U_1(T_1 + (1-\lambda)E_\mu)U_2(T_2 + (1-\lambda)E_\mu)U_3 \cdots U_{n-1}(T_{n-1} + (1-\lambda)E_\mu)U_n\mathbf{1}\|_{L_1(\mu)} \leq \sum_{s \in \{0,1\}^{n-1}} \left(\prod_{j:s_j=1} \lambda\right) \left(\prod_{j:s_j=0} (1-\lambda)\right) \left(\prod_{j \in t(s)} |\cos(2\pi\xi \cdot v_j)|\right),$$

by Eq. (2.7) in Lemma 2 in Chapter 2, where for each $s \in \{0,1\}^{n-1}$, we define $t(s)$ to be the set of indices j such that $s_{j-1} = s_j = 0$, or $s_j = 0$ if $j = 1$ or $s_{j-1} = 0$ if $j = k+1$. Thus by

Claim 11, we can upper bound on the right-hand side of Eq. (6.9) by

$$\frac{1}{2\pi}\sum_{s\in\{0,1\}^{n-1}}\left(\prod_{j:s_j=1}\lambda\right)\left(\prod_{j:s_j=0}(1-\lambda)\right)\frac{C}{(|t(s)|+1)^{3/2}},$$

where the inequality also holds in the case that $|t(s)|=0$.

As in the proof of Theorem 18, let $r:\{0,1\}^{n-1}\to[n-1]$ be defined as

$$r=\left|\{j:s_j=s_{j+1}=0\}\right|,$$

so that $r(s)\le|t(s)|$ for all $s\in\{0,1\}^{n-1}$, and let $\mathbf{s}$ be a random vector from $\{0,1\}^{n-1}$ so that for each $s\in\{0,1\}^{n-1}$

$$\Pr[\mathbf{s}=s]=\left(\prod_{j:s_j=1}\lambda\right)\left(\prod_{j:s_j=0}(1-\lambda)\right).$$

By the definition of $r(\mathbf{s})$, we have

$$\Pr[f(Y_1)v_1+f(Y_2)v_2+\cdots+f(Y_n)v_n=0]\le\frac{1}{2\pi}\mathbb{E}\left[\frac{C}{(r(\mathbf{s})+1)^{3/2}}\right]$$

As before, let $r'=B(\lfloor(n/2\rfloor-1,(1-\lambda)^2)+1$. Then because r' is dominated by $r(s)$,

$$\mathbb{E}\left[\frac{C}{(r(\mathbf{s})+1)^{3/2}}\right]\le\mathbb{E}\left[\frac{C}{r'^{3/2}}\right]\le\left(\mathbb{E}\left[\frac{C^{4/3}}{r'^2}\right]\right)^{3/4}, \qquad (6.10)$$

where again the second inequality follows by Jensen's inequality. Finally, Claim 9 can be used to upper-bound the right-hand side of Eq. (6.10). □

6.4 A Pseudorandom Generator for the Littlewood-Offord Problem

In this section we prove Theorem 20. As stated in the introduction, this theorem can be interpreted as proving the existence of a pseudorandom generator for the Littlewood-Offord problem.

We start by describing the construction of D. Our construction will be based on expander graphs which we define as follows. Given a d-regular graph $G = (V, E)$, let A be the normalized adjacency matrix of G and let J be the matrix whose entries are all $1/|V|$. We say that a family of d-regular graphs $\mathcal{G}$ is a family of expanders if for all graphs G in the family,

$$\|A - J\|_{L_2(\mu)\to L_2(\mu)} \le \lambda$$

for some constant λ bounded away from 1, where μ is the vector whose entries are all $1/|V|$. In particular $1 - \|A - J\|_{L_2(\mu)\to L_2(\mu)}$ is also the spectral gap of the Markov chain that is a simple random walk on G. When $G = (V, E)$ is d-regular, the stationary distribution is μ, and the averaging operator is J. It is well known that there exist infinite families of expander graphs of constant degree d (see for example, (LPS88) and (Mar88)).

Let $G = (\{-1, 1\}^k, E)$ be a d-regular graph from such a family so that $\|A-J\|_{L_2(\mu)\to L_2(\mu)} \le \lambda$ for some constant λ independent of k. We let our set D be the set of concatenations of the labels of walks of length n/k on G, and thus D has cardinality $2^{k+C_1 n/k}$ for some constant C_1 independent of n and k.

Proof of Theorem 20. Let μ be the uniform measure on $\{-1, 1\}^k$ and let D be as defined above. Then by Theorem 21,

$$\sup_{x_0\in\mathbb{R}} \Pr_{\varepsilon\sim D}[|\varepsilon_1 v_1 + \varepsilon_2 v_2 + \cdots + \varepsilon_n v_n - x_0| \leq 1] \leq C\int_{-1}^{1} |\mathbb{E}[\exp(2\pi i\xi(\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n))]|d\xi. \tag{6.11}$$

For each $j \in [n/k]$, let $T_j = A - (1-\lambda)J$ and let $u_j \in \mathbb{R}^{\{-1,1\}^k}$ be the vector defined by

$$u_j(w) = \exp(2\pi i\omega(w_{(j-1)k+1}v_{(j-1)k+1} + \cdots + w_{jk}v_{jk}))S$$

and let $U_j = \mathrm{diag}(u_j)$. Then $|\mathbb{E}[\exp(2\pi i\xi(\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n))]|$ is bounded above by,

$$\left\|U_1(T_1 + (1-\lambda)J)U_2(T_2 + (1-\lambda)J)U_3 \cdots U_{n/k-1}(T_{n/k-1} + (1-\lambda)J)U_{n/k}\mathbf{1}\right\|_{L_1(\mu)} \leq \sum_{s\in\{0,1\}^k} \left(\prod_{j:s_j=1} \lambda\right)\left(\prod_{j:s_j=0} (1-\lambda)\right)\left(\prod_{j\in t(s)} \left|\langle u_j, \mu\rangle_{L_2(\mu)}\right|\right),$$

where the inequality follows by Eq. (2.7) in Lemma 2 in Chapter 2, and for each $s \in \{0,1\}^{n/k-1}$, we define $t(s)$ to be the set of indices j such that $s_{j-1} = s_j = 0$, or $s_j = 0$ if $j = 1$ or $s_{j-1} = 0$ if $j = n/k$.

Note that $\langle u_j, \mu\rangle_{L_2(\mu)}$ is the Fourier transform at ξ of the random variable $w_{(j-1)k+1}v_{(j-1)k+1} + \cdots + w_{jk}v_{jk}$ where each coordinate of w is uniformly random over the set $\{-1,1\}$. This brings us back to the original setting of completely independent random variables, and by Eq. (6.2), it follows that

$$\langle u_j, \mu\rangle_{L_2(\mu)} = \prod_{\ell=1}^{k} \cos(2\pi v_{(j-1)k+\ell}\xi).$$

$$\frac{1}{2\pi}\sum_{s\in\{0,1\}^{n/k-1}}\int_{-1}^{1}\left(\prod_{j:s_j=1}\lambda\right)\left(\prod_{j:s_j=0}(1-\lambda)\right)\left(\prod_{j\in t(s)}\prod_{\ell=1}^{k}\left|\cos(2\pi v_{(j-1)k+\ell}\xi)\right|\right)d\xi \leq$$

$$\frac{1}{2\pi}\sum_{s\in\{0,1\}^{n/k-1}}\left(\prod_{j:s_j=1}\lambda\right)\left(\prod_{j:s_j=0}(1-\lambda)\right)\frac{C}{\sqrt{k(|t(\mathbf{s})|+1)}},$$

where the inequality follows from Claim 8, We proceed by using the same argument as in Lemma 14 starting from Eq. (6.5), which gives an upper bound of $C/\sqrt{k\cdot(n/k)} = C/\sqrt{n}$ as desired. Finally, we obtain a construction of the desired size by letting $k=\sqrt{n}$. □

Bibliography

[AKK99] S. Arora, D. Karger, and M. Karpinski. Polynomial time approximation schemes for dense instances of NP-hard problems. *J. Comput. System Sci.*, 58(1):193–210, 1999. ISSN 0022-0000. doi:10.1006/jcss.1998.1605.

[AL93] M. Ajtai and N. Linial. The influence of large coalitions. *Combinatorica*, 13(2):129–145, 1993. ISSN 0209-9683. doi:10.1007/BF01303199.

[AS99] N. Alon and B. Sudakov. On two segmentation problems. *J. Algorithms*, 33(1):173–184, 1999. ISSN 0196-6774. doi:10.1006/jagm.1999.1024.

[AW02] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory*, 48(3):569–579, 2002. ISSN 0018-9448. doi:10.1109/18.985947.

[BHP01] R. C. Baker, G. Harman, and J. Pintz. The difference between consecutive primes. II. *Proc. London Math. Soc. (3)*, 83(3):532–562, 2001. ISSN 0024-6115. doi:10.1112/plms/83.3.532.

[BL76] J. Bergh and J. Löfström. *Interpolation spaces. An introduction.* Springer-Verlag, Berlin-New York, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223.

[BvH16] A. S. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.*, 44(4):2479–2506, 2016. ISSN 0091-1798.

[CLLM12] K. Chung, H. Lam, Z. Liu, and M. Mitzenmacher. Chernoff-Hoeffding bounds for Markov chains: Generalized and simplified. In *STACS*, pages 124–135. 2012. ArXiv:1201.0559.

[Din95] I. H. Dinwoodie. A probability inequality for the occupation measure of a reversible Markov chain. *Ann. Appl. Probab.*, 5(1):37–43, 1995. ISSN 1050-5164.

[Dvo61] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, pages 123–160. Jerusalem Academic Press, Jerusalem; Pergamon, Oxford, 1961.

[Erd45] P. Erdös. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.*, 51:898–902, 1945. ISSN 0002-9904.

[Ess66] C. G. Esseen. On the Kolmogorov-Rogozin inequality for the concentration function. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 5:210–216, 1966. doi:10.1007/BF00533057.

[FF88] P. Frankl and Z. Füredi. Solution of the Littlewood-Offord problem in high dimensions. *Ann. of Math. (2)*, 128(2):259–270, 1988. ISSN 0003-486X.

[FJS18] J. Fan, B. Jiang, and Q. Sun. Hoeffding's lemma for Markov chains and its applications to statistical learning, 2018.

[Gil98] D. Gillman. A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.*, 27(4):1203–1220, 1998. ISSN 0097-5397. doi:10.1137/S0097539794268765.

[GLSS17] A. Garg, Y. T. Lee, Z. Song, and N. Srivastava. A matrix expander Chernoff bound, 2017.

[Hal77] G. Halász. Estimates for the concentration function of combinatorial number theory and probability. *Period. Math. Hungar.*, 8(3-4):197–211, 1977. ISSN 0031-5303.

[Hea08] A. D. Healy. Randomness-efficient sampling within NC^1. *Comput. Complexity*, 17(1):3–37, 2008. ISSN 1016-3328. doi:10.1007/s00037-007-0238-5.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *BULL. AMER. MATH. SOC.*, 43(4):439–561, 2006.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963. ISSN 0162-1459.

[HOW15] B. Hemenway, R. Ostrovsky, and M. Wootters. Local correctability of expander codes. *Inform. and Comput.*, 243:178–190, 2015. ISSN 0890-5401. doi:10.1016/j.ic.2014.12.013.

[Jam15] G. J. O. Jameson. A simple proof of Stirling's formula for the gamma function. *Math. Gaz.*, 99(544):68–74, 2015. ISSN 0025-5572. doi:10.1017/mag.2014.9.

[JSF18] B. Jiang, Q. Sun, and J. Fan. Bernstein's inequality for general markov chains, 2018.

[Kah97] N. Kahale. Large deviation bounds for Markov chains. *Combin. Probab. Comput.*, 6(4):465–474, 1997. ISSN 0963-5483. doi:10.1017/S0963548397003209.

[Kar07] V. Kargin. A large deviation inequality for vector functions on finite reversible Markov chains. *Ann. Appl. Probab.*, 17(4):1202–1221, 2007. ISSN 1050-5164. doi:10.1214/105051607000000078.

[KKL17] V. Kabanets, D. M. Kane, and Z. Lu. A polynomial restriction lemma with applications. In *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 615–628. ACM, New York, 2017.

[Kle70] D. J. Kleitman. On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors. *Advances in Math.*, 5:155–157 (1970), 1970. ISSN 0001-8708.

[Lez98] P. Lezaud. Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.*, 8(3):849–867, 1998. ISSN 1050-5164. doi:10.1214/aoap/1028903453.

[LO43] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation. III. *Rec. Math. [Mat. Sbornik] N.S.*, 12(54):277–286, 1943.

[LP04] C. A. León and F. Perron. Optimal Hoeffding bounds for discrete reversible Markov chains. *Ann. Appl. Probab.*, 14(2):958–970, 2004. ISSN 1050-5164. doi:10.1214/105051604000000170.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988. ISSN 1439-6912. doi:10.1007/BF02126799.

[Mar88] G. A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii*, 24(1):51–60, 1988. ISSN 0555-2923.

[Mek17] R. Meka. Explicit resilient functions matching Ajtai-Linial. In *SODA*. 2017.

[MT06] R. Montenegro and P. Tetali. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.*, 1(3):x+121, 2006. ISSN 1551-305X. doi:10.1561/0400000003.

[MZ37] J. Marcinkiewicz and A. Zygmund. Sur les fonctions indépendantes. *Fundamenta Mathematicae*, 29(1):60–90, 1937.

[Nao12] A. Naor. On the Banach-space-valued Azuma inequality and small-set isoperimetry of Alon-Roichman graphs. *Combin. Probab. Comput.*, 21(4):623–634, 2012. ISSN 0963-5483. doi:10.1017/S0963548311000757.

[NRR17] A. Naor, S. Rao, and O. Regev. On the rate of convergence of the vector-valued ergodic theorem for Markov chains with a spectral gap, 2017. In preparation.

[Pau15] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.*, 20:no. 79, 32, 2015. ISSN 1083-6489. doi:10.1214/EJP.v20-4039.

[Rao19a] S. Rao. A Hoeffding inequality for Markov chains. *Electron. Commun. Probab.*, 24:11 pp., 2019. doi:10.1214/19-ECP219.

[Rao19b] S. Rao. The littlewood-offord problem for markov chains, 2019.

[Rie27] M. Riesz. Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires. *Acta Math.*, 49(3-4):465–497, 1927. ISSN 0001-5962. doi:10.1007/BF02564121.

[RR17] S. Rao and O. Regev. A sharp tail bound for the expander random sampler, 2017.

[Rud91] W. Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991. ISBN 0-07-054236-8.

[Sal83] A. Sali. Stronger form of an M-part Sperner theorem. *European J. Combin.*, 4(2):179–183, 1983. ISSN 0195-6698.

[Sal85] A. Sali. A Sperner-type theorem. *Order*, 2(2):123–127, 1985. ISSN 0167-8094.

[SS65] A. Sárközi and E. Szemerédi. über ein Problem von Erdös und Moser. *Acta Arith.*, 11:205–208, 1965. ISSN 0065-1036.

[Sta12] R. P. Stanley. *Enumerative combinatorics. Volume 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2012. ISBN 978-1-107-60262-5.

[Tal87] M. Talagrand. Regularity of Gaussian processes. *Acta Math.*, 159(1-2):99–149, 1987. ISSN 0001-5962.

[Tal14] M. Talagrand. *Upper and lower bounds for stochastic processes*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer, Heidelberg, 2014. ISBN 978-3-642-54074-5; 978-3-642-54075-2. doi:10.1007/978-3-642-54075-2. Modern methods and classical problems.

[Tho48] G. O. Thorin. Convexity theorems generalizing those of M. Riesz and Hadamard with some applications. *Comm. Sem. Math. Univ. Lund [Medd. Lunds Univ. Mat. Sem.]*, 9:1–58, 1948.

[Tro15] J. A. Tropp. An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.*, 8(1-2):1–230, May 2015. ISSN 1935-8237. doi:10.1561/2200000048.

[TV06] T. Tao and V. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006. ISBN 978-0-521-85386-6; 0-521-85386-9. doi:10.1017/CBO9780511755149.

[TV12] T. Tao and V. Vu. The Littlewood-Offord problem in high dimensions and a conjecture of Frankl and Füredi. *Combinatorica*, 32(3):363–372, 2012. ISSN 0209-9683.

[Wag08] R. Wagner. Tail estimates for sums of variables sampled by a random walk. *Comb. Probab. Comput.*, 17(2):307–316, March 2008. ArXiv:math/0608740.

[Zuc97] D. Zuckerman. Randomness-optimal oblivious sampling. In *Proceedings of the Workshop on Randomized Algorithms and Computation (Berkeley, CA, 1995)*, volume 11, pages 345–367. 1997. ISSN 1042-9832. doi:10.1002/(SICI)1098-2418(199712)11:4⟨345::AID-RSA4⟩3.3.CO;2-7.

[Zuc07] D. Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory Comput.*, 3:103–128, 2007. ISSN 1557-2862. doi:10.4086/toc.2007.v003a006.