

A Hoeffding inequality for Markov chains

Shravas Rao*

Abstract

We prove deviation bounds for the random variable $\sum_{i=1}^n f_i(Y_i)$ in which $\{Y_i\}_{i=1}^\infty$ is a Markov chain with stationary distribution and state space $[N]$, and $f_i : [N] \rightarrow [-a_i, a_i]$. Our bound improves upon previously known bounds in that the dependence is on $\sqrt{a_1^2 + \cdots + a_n^2}$ rather than $\max_i \{a_i\} \sqrt{n}$. We also prove deviation bounds for certain types of sums of vector-valued random variables obtained from a Markov chain in a similar manner. One application includes bounding the expected value of the Schatten ∞ -norm of a random matrix whose entries are obtained from a Markov chain.

Keywords: Hoeffding bound; Markov chain; generic chaining; random matrices.

AMS MSC 2010: 60F10.

Submitted to ECP on September 13, 2018, final version accepted on February 17, 2019.

Supersedes arXiv:1806.11519.

1 Introduction

Consider a Markov chain $\{Y_i\}_{i=1}^\infty$ with state space $[N]$, transition matrix A , and stationary distribution π such that Y_1 is distributed as π . Let E_π be the associated averaging operator defined by $(E_\pi)_{ij} = \pi_j$, so that for $v \in \mathbb{R}^N$ $E_\pi v = \mathbb{E}_\pi[v] \mathbf{1}$ where $\mathbf{1}$ is the vector whose entries are all 1.

In the case that the Y_i are independent, that is $A = E_\pi$, then it is well known (see [10]) that for functions $f_1, \dots, f_n : [N] \rightarrow [-1, 1]$ with $\mathbb{E}[f_i(Y_i)] = 0$ for all i and $u \geq 0$, that

$$\Pr[|f_1(Y_1) + \cdots + f_n(Y_n)| \geq u\sqrt{n}] \leq 2 \exp(-u^2/2). \quad (1.1)$$

Gillman generalized Eq. (1.1) to all Markov chains with a stationary distribution, in terms of the quantity $\lambda = \|A - E_\pi\|_{L_2(\pi) \rightarrow L_2(\pi)}$ in the case $f_1 = \cdots = f_n$ [7]. These bounds were refined in a long series of work including [4, 11, 13, 22, 12, 9, 3, 8, 16, 15, 17]. We state the following version due to Healy [9], which handles the case in which the f_i are not necessarily equal.

$$\Pr[|f_1(Y_1) + \cdots + f_n(Y_n)| \geq u\sqrt{n}] \leq 2 \exp\left(\frac{-u^2(1-\lambda)}{4}\right). \quad (1.2)$$

*New York University, United States of America. E-mail: rao@cs.nyu.edu, cs.nyu.edu/~rao

†This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1342536.

Back in the case of independent random variables, Hoeffding generalized Eq. (1.1) to the case when the function f_i has range $[-a_i, a_i]$, obtaining the following bound [10].

$$\Pr \left[|f_1(Y_1) + \dots + f_n(Y_n)| \geq u \left(\sum_{i=1}^n a_i^2 \right)^{1/2} \right] \leq 2 \exp(-u^2/2). \quad (1.3)$$

In this work, we generalize Eq. (1.3) to Markov chains with a stationary distribution. In particular, we prove the following.

Theorem 1.1. *Let $\{Y_i\}_{i=1}^\infty$ be a stationary Markov chain with state space $[N]$, transition matrix A , stationary probability measure π , and averaging operator E_π , so that Y_1 is distributed according to π . Let $\lambda = \|A - E_\pi\|_{L_2(\pi) \rightarrow L_2(\pi)}$ and let $f_1, \dots, f_n : [N] \rightarrow \mathbb{R}$ 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 $u \geq 0$,*

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

One interpretation of Theorem 1.1 is that for a Markov chain $\{Y_i\}_{i=1}^\infty$ and functions $f_1, \dots, f_n : [N] \rightarrow [-1, 1]$, the random vector $(f_1(Y_1), \dots, f_n(Y_n))$ is sub-gaussian.

We remark that the dependence on λ in both Eq. (1.2) and Theorem 1.1 is optimal, as shown in [12] which considered the case that the f_i are equal. 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 $f_1(Y_1) + \dots + f_n(Y_n)$ is similar to the sum of $n(1 - \lambda)$ random variables that are close to $1/(1 - \lambda)$ or close to $-1/(1 - \lambda)$, both with equal probability.

We also remark that Theorem 1.1 holds even for non-reversible Markov chains, continuing the work of [3] who were the first to consider this setting. It is possible, if the Markov chain is not reversible, for $\|A - E_\pi\|_{L_2(\pi) \rightarrow L_2(\pi)}$ to be greater than 1, and thus the bound in Theorem 1.1 is trivial.

1.1 Extension to vector-valued random variables

Recently, much attention has been paid to tail bounds for sums of vector-valued random variables. Naor [14] obtained tail bounds for sums of random variables from a Banach space satisfying certain properties. Before stating the corresponding theorem, we define a quantity called the modulus of uniform smoothness.

Definition 1.2. *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 $f_i : [N] \rightarrow \{x \in X : \|x\| \leq a_i\}$ and such that $\mathbb{E}[f_i(Y_i)] = 0$, it was shown that

$$\Pr \left[\|f_1(Y_1) + \dots + f_n(Y_n)\| \geq u \left(\sum_{i=1}^n a_i^2 \right)^{1/2} \right] \leq \exp(s + 2 - cu^2) \quad (1.4)$$

for some universal constant c .

We extend Theorem 1.1 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. (1.4). 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 1.3. *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 Markov chain with state space $[N]$, transition matrix A , stationary probability measure π , and averaging operator E_π , so that Y_1 is distributed according to π . Let $\lambda = \|A - E_\pi\|_{L_2(\pi) \rightarrow L_2(\pi)}$, and let $f_1, \dots, f_n : [N] \rightarrow [-1, 1]$ 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 $u \geq 0$,*

$$\Pr[\|f_1(Y_1)X_1 + \dots + f_n(Y_n)X_n\| \geq uC\mathbb{E}[\|g_1X_1 + \dots + g_nX_n\|]] \leq L \exp(-Cu^2(1 - \lambda))$$

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

Note that Eq. (1.4) implies that $\mathbb{E}[\|g_1X_1 + \dots + g_nX_n\|] \leq C\sqrt{s(\|X_1\|^2 + \dots + \|X_n\|^2)}$ 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

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

1.1.1 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 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 [2] that

$$\mathbb{E}[\|X\|_{S_\infty}] \leq \min \left\{ C(\sigma + \sigma_*\sqrt{\log d}), \|B\|_{S_\infty} \right\} \tag{1.5}$$

for some absolute constant C , where

$$\sigma = \max_i \sqrt{\sum_j b_{i,j}^2} \text{ and } \sigma_* = \max_{i,j} |b_{i,j}|. \tag{1.6}$$

We generalize Eq. (1.5) to Markov chains with a stationary distribution. In particular, we obtain a similar bound in terms of $\lambda = \|A - E_\pi\|_{L_2(\pi) \rightarrow L_2(\pi)}$ 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 $A = E_\pi$ corresponds to choosing the entries of the diagonal and upper triangular part of the matrix independently, as in [2].

Corollary 1.4. *Let $\{Y_i\}_{i=1}^\infty$ be a stationary Markov chain with state space $[N]$, transition matrix A , stationary probability measure π , and averaging operator E_π , so that Y_1 is distributed according to π . Let $\lambda = \|A - E_\pi\|_{L_2(\pi) \rightarrow L_2(\pi)}$, let $f : V \rightarrow [-1, 1]$ 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 $\omega : \mathcal{I} \rightarrow \{1, 2, \dots, (d^2 + d)/2\}$, 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,

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

for some absolute constant C , where σ and σ_* are defined as in Eq. (1.6).

1.2 Related Work

In recent independent work by Fan, Jiang, and Sun [5], a Hoeffding bound for general Markov chains was also given. Their bound is sharper, and in particular the constant $64e$ can be replaced by 2 after replacing $1 - \lambda$ by $(1 - \lambda)/(1 + \lambda)$. However, our proof is arguably somewhat simpler.

In work by Garg, Lee, Song and Srivastava [6], a version of Eq. (1.4) was proved for Markov chains when the Banach space is the set of $d \times d$ matrices under the Schatten ∞ -norm, generalizing a result first shown by Ahlswede and Winter [1] (see also the monograph by Tropp [21]). Note that the Schatten ∞ -norm of a $d \times d$ matrix is up to constant factors equal to the Schatten $\log(d)$ -norm of that matrix, and the modulus of uniform smoothness of the set of $d \times d$ matrices under the Schatten p -norm is $O(p\tau^2)$. Thus in this case the left-hand side of Eq. (1.4) is bounded above by $d \exp(c - cu^2)$. Garg, Lee, Song and Srivastava showed that for a Markov chain $\{Y_i\}_{i=1}^\infty$ and functions $f_i : [N] \rightarrow \{x \in \mathbb{R}^{d \times d} : \|x\|_{S_\infty} \leq 1\}$ such that $\mathbb{E}[f_i(Y_i)] = 0$,

$$\Pr[\|f_1(Y_1) + \dots + f_n(Y_n)\|_{S_\infty} \leq u\sqrt{n}] \leq 2d \exp(-c(1 - \lambda)u^2)$$

2 Preliminaries

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

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

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

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

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

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

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 $\text{diag}(v)$ be the diagonal matrix where $\text{diag}(v)_{i,i} = v_i$.

Let A be a stochastic matrix, and let π be a stationary distribution for A . We let $(E_\pi)_{ij} = \pi_j$ be the averaging operator on $L_\infty(\pi) \rightarrow L_\infty(\pi)$. Note that E_π is also stochastic, and that $E_\pi A = AE_\pi = E_\pi^2 = E_\pi$.

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

Claim 2.1. For any matrix T ,

$$\|T\|_{L_2(\pi) \rightarrow L_2(\pi)}^2 \leq \|T\|_{L_1(\pi) \rightarrow L_1(\pi)} \|T\|_{L_\infty(\pi) \rightarrow L_\infty(\pi)}.$$

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

$$\begin{aligned} \|Tx\|_{L_2(\pi) \rightarrow L_2(\pi)}^2 &= \sum_{i=1}^n \pi_i \left(\sum_{j=1}^n T_{ij} x_j \right)^2 \leq \sum_{i=1}^n \pi_i \left(\sum_{j=1}^n |T_{ij}| \right) \left(\sum_{j=1}^n |T_{ij}| x_j^2 \right) \\ &\leq \|T\|_{L_\infty(\pi) \rightarrow L_\infty(\pi)} \|T(x \circ x)\|_{L_1(\pi) \rightarrow L_1(\pi)} \leq \|T\|_{L_\infty(\pi) \rightarrow L_\infty(\pi)} \|T\|_{L_1(\pi) \rightarrow L_1(\pi)} \|x\|_{L_2(\pi)}^2 \end{aligned}$$

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

3 Proof of Theorem 1.1

To prove Theorem 1.1, we follow the strategy of bounding the q th moment for some even integer q , and using Markov's inequality to obtain a tail bound. We start by expanding $(f_1(Y_1) + \dots + f_n(Y_n))^q$ into a sum of monomials.

The following lemma bounds the expectation of monomials in the $f_i(Y_i)$. The statement is similar to Lemma 3.3 in [17]. Most of the proof is the same and is deferred to the appendix. Let $S_{q-1} \subset \{0, 1\}^{q-1}$ be 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 3.1. Let $\{Y_i\}_{i=1}^\infty$ be a stationary Markov chain with state space $[N]$, transition matrix A , stationary probability measure π , and averaging operator E_π , so that Y_1 is distributed according to π . Let $\lambda = \|A - E_\pi\|_{L_2(\pi) \rightarrow L_2(\pi)}$ and let $f_1, \dots, f_n : [N] \rightarrow \mathbb{R}$ 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 $w_1 \leq w_2 \leq \dots \leq w_q$

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

Proof. We apply Lemma A.3, letting $k = q - 1$, $u_i(v) = f_{w_i}(v)$ for all $v \in [N]$, and $T_i = A^{w_{i+1} - w_i} - E_\pi$. Note that for all $k \geq 0$,

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

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

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

Theorem 3.2. Let $\{Y_i\}_{i=1}^\infty$ be a stationary Markov chain with state space $[N]$, transition matrix A , stationary probability measure π , and averaging operator E_π , so that Y_1 is distributed according to π . Let $\lambda = \|A - E_\pi\|_{L_2(\pi) \rightarrow L_2(\pi)}$ be less than 1, and let $f_1, \dots, f_n : [N] \rightarrow \mathbb{R}$ 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 ,

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

A Hoeffding inequality for Markov chains

Proof. Let $\sigma : [n]^q \rightarrow [n]^q$ be the function where $\sigma(w)$ is the sorted list of coordinates of w in non-decreasing order. Then by Lemma 3.1,

$$\begin{aligned} \mathbb{E}[(f_1(Y_1) + \dots + f_n(Y_n))^q] &= \sum_{w \in [n]^q} \mathbb{E}[f_{w_1}(Y_{w_1})f_{w_2}(Y_{w_2}) \cdots f_{w_q}(Y_{w_q})] \\ &\leq \sum_{w \in [n]^q} a_{w_1}a_{w_2} \cdots a_{w_q} \sum_{s \in S_{q-1}} \left(\prod_{i:s_i=1} \lambda^{\sigma(w)_{i+1} - \sigma(w)_i} \right). \end{aligned} \quad (3.1)$$

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 $j \in [n]$,

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

i.e. the multi-set $\bigcup_{i \in \mathcal{I}} \{w_i\}$ is equal to the multi-set $\{\sigma(w)_1, \sigma(w)_3, \sigma(w)_5, \dots, \sigma(w)_{q-1}\}$. Let $w_{\mathcal{I}}, w_{[q] \setminus \mathcal{I}} \in [n]^{q/2}$ be the restriction of w to the coordinates in \mathcal{I} and $[q] \setminus \mathcal{I}$ 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 $w_{[q] \setminus \mathcal{I}}$ th column of $T_{\mathcal{I},s}$ is

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

Because

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

Eq. (3.1) can be bounded above by

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

where $a^{\otimes q/2} \in \mathbb{R}^{n^{q/2}}$ is the vector such that $a_{i_1, \dots, i_{q/2}}^{\otimes q/2} = a_{i_1}a_{i_2} \cdots a_{i_{q/2}}$ 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 2.1, it is enough to show that

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

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

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

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

as desired. The first inequality follows from the fact that $w \in W_{\mathcal{I}}$ and $w_1, w_3, w_5, \dots, w_{q-1}$ determine $\sigma(w)_1, \sigma(w)_3, \sigma(w)_5, \dots, \sigma(w)_{q-1}$ exactly, and that there are at most $(q/2)!$

possible orderings of w_2, w_4, \dots, 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. \square

Finally, Theorem 1.1 follows by considering the moment generating function and applying Markov's inequality.

Proof of Theorem 1.1. 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 $\theta = (1-\lambda)u/(32(a_1^2 + \dots + a_n^2)^{1/2})$. By Theorem 3.2 and keeping in mind that by Jensen's inequality, odd moments are bounded above by even moments,

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

By Markov's inequality,

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

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 $2 \exp(-u^2(1-\lambda)/(64e)) \geq 1$. \square

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 Extension to vector-valued random variables

To prove Theorem 1.3 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 4.1. 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 1.1 essentially states that for a Markov chain $\{Y_i\}_{i=1}^\infty$ and functions $f_1, \dots, f_n : [N] \rightarrow [-1, 1]$ with $\mathbb{E}[f_i(Y_i)] = 0$, the process $(Z_t)_{t \in T}$ defined by $Z_t = (f_1(Y_1)t_1, \dots, f_n(Y_n)t_n)$ 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 4.2.

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

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

The majorizing measures theorem, due to Talagrand [19] (see also Theorem 2.4.1 in [20]), 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 4.3 (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 [20].

Theorem 4.4. *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 1.3. 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 [18]). For fixed $X_1, \dots, X_n \in X$, let $T \subset \mathbb{R}^n$ be the set of points,

$$T = \{(\langle x^*, X_1 \rangle, \langle x^*, X_2 \rangle, \dots, \langle x^*, X_n \rangle) : x^* \in B^*\}. \tag{4.1}$$

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

$$\|f_1 X_1 + \dots + f_n X_n\| = \sup_{t \in T} \langle f, t \rangle. \tag{4.2}$$

Finally, we prove Theorem 1.3.

Proof of Theorem 1.3. Consider the metric space (T, d) where T is as constructed in Eq. (4.1) and $d(s, t) = \sqrt{32e/(1-\lambda)} \|s - t\|_{\ell_2}$. Then by Theorem 1.1, the process $(Z_t)_{t \in T}$ defined by $Z_t = (f_1(Y_1)t_1, \dots, f_n(Y_n)t_n)$ satisfies the increment condition.

Additionally, consider the Gaussian process $(Z'_t)_{t \in T}$ on the metric space (T, d') , so that $Z_t = g_1 t_1 + \dots + g_n t_n$ for independent standard Gaussian variables g_1, \dots, g_n and $d' = \mathbb{E}[(Z_s - Z_t)^2]^{1/2}$. Then by Theorem 4.3,

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

The theorem then follows from Theorem 4.4 the observation that $\sup_{s,t} |Z_s - Z_t| = 2 \sup_t Z_t$ as T is symmetric, and Eq. (4.2). \square

4.1 Comparison to matrices with independent entries

We prove Corollary 1.4, which follows from a straightforward application of Theorem 1.3.

In order to apply Theorem 1.3, 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. (1.5)). This setting was also discussed in [2] in which it was shown that

$$\mathbb{E}[\|X'\|_{S_\infty}] \leq C(\sigma + \sigma_* \sqrt{\log d}), \tag{4.3}$$

where σ and σ_* are defined as in Eq. (1.6).

Proof of Corollary 1.4. Let X' be the random matrix defined above. Then by Theorem 1.3 and Eq. (4.3),

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

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. \square

A Appendix

In this section, we give the tools needed to prove Lemma 3.1. They are either taken directly from [17] (which is based on techniques used in [15]), or are straightforward adaptations.

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

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

Claim A.2. For all $k \geq 1$, vectors $u_1, \dots, u_k \in \mathbb{R}^N$, $U_i = \text{diag}(u_i)$ for all i , distributions π over $[N]$ and matrices $T_1, \dots, T_{k-1} \in \mathbb{R}^{N \times N}$,

$$\|U_1 T_1 U_2 T_2 \dots T_{k-1} U_k \mathbf{1}\|_{L_1(\pi)} \leq \|u_k\|_{L_\infty(\pi)} \prod_{i=1}^{k-1} \|u_i\|_{L_\infty(\pi)} \|T_i\|_{L_2(\pi)}.$$

Proof. By Jensen's inequality, the right-hand side is bounded above by

$$\|U_1 T_1 U_2 T_2 \dots T_{k-1} U_k \mathbf{1}\|_{L_2(\pi)}$$

and the claim follows by the definition of operator norm, and the fact that $\|U_i\|_{L_2(\pi) \rightarrow L_2(\pi)} = \|u_i\|_{L_\infty(\pi)}$. \square

Lemma A.3. Let $k \geq 1$ be an integer. Let $S_k \subset \{0, 1\}^k$ be the subset of $\{0, 1\}^k$ of vectors s with no two consecutive 0s and so that $s_1, s_k = 1$. Let π be a distribution over $[N]$, let $u_1, \dots, u_{k+1} \in \mathbb{R}^N$ be N -dimensional vectors such that $u_i^t \pi = 0$ for all i , and let $U_i = \text{diag}(u_i)$ for all i . Finally, let $T_1, \dots, T_k \in \mathbb{R}^{N \times N}$. Then,

$$\left| \langle \mathbf{1}, U_1(T_1 + E_\pi)U_2(T_2 + E_\pi)U_3 \cdots U_k(T_k + E_\pi)U_{k+1}\mathbf{1} \rangle_{L_2(\pi)} \right| \leq \|u_1\|_{L_\infty(\pi)} \|u_2\|_{L_\infty(\pi)} \cdots \|u_{k+1}\|_{L_\infty(\pi)} \sum_{s \in S_k} \prod_{j: s_j=1} \|T_j\|_{L_2(\pi) \rightarrow L_2(\pi)}. \quad (\text{A.1})$$

Proof. For $j = 1, \dots, k$, let $T_{j,0} = E_\pi$ and $T_{j,1} = T_j$. Then using the triangle inequality, the left-hand side of (A.1) is at most

$$\sum_{s \in \{0,1\}^k} \left| \left\langle \mathbf{1}, \left(\prod_{j=1}^k U_j T_{j,s_j} \right) U_{k+1} \mathbf{1} \right\rangle_{L_2(\pi)} \right| = \sum_{s \in S_k} \left| \left\langle \mathbf{1}, \left(\prod_{j=1}^k U_j T_{j,s_j} \right) U_{k+1} \mathbf{1} \right\rangle_{L_2(\pi)} \right|, \quad (\text{A.2})$$

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 $E_\pi U_j E_\pi = 0$ (or $E_\pi U_{k+1} \mathbf{1} = 0$) appears. Additionally, terms corresponding to vectors s with $s_1 = 0$ are equal to 0, as $\langle \mathbf{1}, U_1 E_\pi v \rangle_{L_2(\pi)} = 0$ for all $v \in \mathbb{R}^N$.

Fix an $s \in S_k$, and let r_1, r_2, \dots, r_ℓ be the indices of s that are 0. By Claim A.1, the term corresponding to s in Eq. (A.2) is at most

$$\|U_1 T_1 U T_2 \cdots T_{r_1-1} U_{r_1} \mathbf{1}\|_{L_1(\pi)} \cdot \|U_{r_1+1} T_{r_1+1} U_{r_1+2} T_{r_1+2} \cdots T_{r_2-1} U_{r_2} \mathbf{1}\|_{L_1(\pi)} \cdots \|U_{r_\ell+1} T_{r_\ell+1} U_{r_\ell+2} T_{r_\ell+2} \cdots T_k U_{k+1} \mathbf{1}\|_{L_1(\pi)}.$$

The claim now follows by applying Claim A.2. □

References

- [1] Rudolf Ahlswede and Andreas Winter, *Strong converse for identification via quantum channels*, IEEE Trans. Inform. Theory **48** (2002), no. 3, 569–579. MR-1889969
- [2] Afonso S. Bandeira and Ramon van Handel, *Sharp nonasymptotic bounds on the norm of random matrices with independent entries*, Ann. Probab. **44** (2016), no. 4, 2479–2506. MR-3531673
- [3] Kai-Min Chung, Henry Lam, Zhenming Liu, and Michael Mitzenmacher, *Chernoff-Hoeffding bounds for Markov chains: Generalized and simplified*, STACS, 2012, arXiv:1201.0559, pp. 124–135. MR-2909308
- [4] I. H. Dinwoodie, *A probability inequality for the occupation measure of a reversible Markov chain*, Ann. Appl. Probab. **5** (1995), no. 1, 37–43. MR-1325039
- [5] Jianqing Fan, Bai Jiang, and Qiang Sun, *Hoeffding’s lemma for Markov chains and its applications to statistical learning*, 2018.
- [6] Ankit Garg, Yin Tat Lee, Zhao Song, and Nikhil Srivastava, *A matrix expander Chernoff bound*, 2017.
- [7] David Gillman, *A Chernoff bound for random walks on expander graphs*, SIAM J. Comput. **27** (1998), no. 4, 1203–1220. MR-1621958
- [8] Jan Hazla and Thomas Holenstein, *Upper tail estimates with combinatorial proofs*, STACS, 2015, arXiv:1405.2349, pp. 392–405. MR-3356428
- [9] Alexander D. Healy, *Randomness-efficient sampling within NC¹*, Comput. Complexity **17** (2008), no. 1, 3–37. MR-2402893
- [10] Wassily Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30. MR-0144363

A Hoeffding inequality for Markov chains

- [11] Nabil Kahale, *Large deviation bounds for Markov chains*, *Combin. Probab. Comput.* **6** (1997), no. 4, 465–474. MR-1483429
- [12] Carlos A. León and François Perron, *Optimal Hoeffding bounds for discrete reversible Markov chains*, *Ann. Appl. Probab.* **14** (2004), no. 2, 958–970. MR-2052909
- [13] Pascal Lezaud, *Chernoff-type bound for finite Markov chains*, *Ann. Appl. Probab.* **8** (1998), no. 3, 849–867. MR-1627795
- [14] Assaf Naor, *On the Banach-space-valued Azuma inequality and small-set isoperimetry of Alon-Roichman graphs*, *Combin. Probab. Comput.* **21** (2012), no. 4, 623–634. MR-2942733
- [15] Assaf Naor, Shramas Rao, and Oded Regev, *On the rate of convergence of the vector-valued ergodic theorem for Markov chains with a spectral gap*, 2017, In preparation.
- [16] Daniel Paulin, *Concentration inequalities for Markov chains by Marton couplings and spectral methods*, *Electron. J. Probab.* **20** (2015), 32 pp. MR-3383563
- [17] Shramas Rao and Oded Regev, *A sharp tail bound for the expander random sampler*, 2017.
- [18] Walter Rudin, *Functional analysis*, second ed., International Series in Pure and Applied Mathematics, McGraw-Hill, Inc., New York, 1991. MR-1157815
- [19] Michel Talagrand, *Regularity of Gaussian processes*, *Acta Math.* **159** (1987), no. 1-2, 99–149. MR-0906527
- [20] Michel Talagrand, *Upper and lower bounds for stochastic processes*, *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]*, vol. 60, Springer, Heidelberg, 2014, Modern methods and classical problems.
- [21] Joel A. Tropp, *An introduction to matrix concentration inequalities*, *Found. Trends Mach. Learn.* **8** (2015), no. 1-2, 1–230.
- [22] Roy Wagner, *Tail estimates for sums of variables sampled by a random walk*, *Comb. Probab. Comput.* **17** (2008), no. 2, 307–316, arXiv:math/0608740. MR-2396356

Acknowledgments. I would like to thank Oded Regev, Noah Stephens-Davidowitz, and the anonymous referees for their valuable comments. I would also like to thank the anonymous referees for pointing out that the proof also applies to non-reversible Markov chains.

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS¹)
- Easy interface (EJMS²)

Economical model of EJP-ECP

- Non profit, sponsored by IMS³, BS⁴, ProjectEuclid⁵
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

²EJMS: Electronic Journal Management System <http://www.vtex.lt/en/ejms.html>

³IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

⁴BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁵Project Euclid: <https://projecteuclid.org/>

⁶IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>