

Concentration of Markov chains with bounded moments

Assaf Naor*

Shravas Rao†

Oded Regev‡

Abstract

Let $\{W_t\}_{t=1}^\infty$ be a finite state stationary Markov chain, and suppose that f is a real-valued function on the state space. If f is bounded, then Gillman's *expander Chernoff bound* (1993) provides concentration estimates for the random variable $f(W_1) + \dots + f(W_n)$ that depend on the spectral gap of the Markov chain and the assumed bound on f . Here we obtain analogous inequalities assuming only that the q 'th moment of f is bounded for some $q \geq 2$. Our proof relies on reasoning that differs substantially from the proofs of Gillman's theorem that are available in the literature, and it generalizes to yield dimension-independent bounds for mappings f that take values in an $L_p(\mu)$ for some $p \geq 2$, thus answering (even in the Hilbertian special case $p = 2$) a question of Kargin (2007).

Résumé: Soit $\{W_t\}_{t=1}^\infty$ une chaîne de Markov stationnaire à états finis, et supposons que f soit une fonction à valeur réelle dans l'espace d'états. Si f est borné, le *inégalité de Chernoff pour graphes expandeurs* (1993) de Gillman fournit des estimations de concentration pour la variable aléatoire $f(W_1) + \dots + f(W_n)$ qui dépendent de l'écart spectral de la chaîne de Markov et le borne sur f . Nous obtenons ici des inégalités analogues en supposant seulement que le q ème moment de f est borné pour un certain $q \geq 2$. Notre démonstration repose sur un raisonnement qui diffère substantiellement des démonstrations du théorème de Gillman disponibles dans la littérature, et elle généralise pour générer des bornes indépendantes de la dimension pour les applications f qui prennent des valeurs dans un $L_p(\mu)$ pour quelque $p \geq 2$, répondant ainsi (même dans le cas spécial hilbertien $p = 2$) à une question de Kargin (2007).

Keywords: Markov chains, concentration bounds, expander graphs, Gillman's theorem

1 Introduction

For $N \in \mathbb{N}$, write $[N] \stackrel{\text{def}}{=} \{1, \dots, N\}$ and let $\Delta^{N-1} \stackrel{\text{def}}{=} \{\pi = (\pi_1, \dots, \pi_N) \in [0, 1]^N : \sum_{i=1}^N \pi_i = 1\}$ be the simplex of probability measures on $[N]$. Given $\pi \in \Delta^{N-1}$, denote by $E_\pi \in M_N(\mathbb{R})$ the N -by- N matrix all of whose rows equal π , i.e., $E_\pi u = (\sum_{j=1}^N \pi_j u_j, \dots, \sum_{j=1}^N \pi_j u_j) \in \mathbb{R}^N$ for every $u = (u_1, \dots, u_N) \in \mathbb{R}^N$.

Given $\pi \in \Delta^{N-1}$, a stochastic matrix $A = (a_{ij}) \in M_N(\mathbb{R})$ is π -stationary if $\pi A = \pi$, i.e., $\pi_i = \sum_{j=1}^N \pi_j a_{ji}$ for all $i \in [N]$. We then define $\lambda_\pi(A)$ to be the norm of $A - E_\pi$ as an operator from $L_2(\pi)$ to $L_2(\pi)$, i.e.,

$$\lambda_\pi(A) \stackrel{\text{def}}{=} \|A - E_\pi\|_{L_2(\pi) \rightarrow L_2(\pi)} = \sup \left\{ \left(\sum_{i=1}^N \pi_i \left(\sum_{j=1}^N a_{ij} u_j - \sum_{k=1}^N \pi_k u_k \right)^2 \right)^{\frac{1}{2}} : u \in \mathbb{R}^N \text{ and } \sum_{k=1}^N \pi_k u_k^2 = 1 \right\}.$$

*Mathematics Department, Princeton University. Supported by the Packard Foundation and the Simons Foundation. The research that is presented here was conducted under the auspices of the Simons Algorithms and Geometry (A&G) Think Tank.

†Courant Institute of Mathematical Sciences, New York University. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1342536.

‡Courant Institute of Mathematical Sciences, New York University. Supported by the Simons Collaboration on Algorithms and Geometry and by the National Science Foundation under Grant No. CCF-1814524. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Note that if A is diagonalizable over the Hilbert space $L_2(\pi)$, then we have $\lambda_\pi(A) = \max\{\lambda_2(A), |\lambda_N(A)|\}$, where $1 = \lambda_1(A) \geq \dots \geq \lambda_N(A) \geq -1$ are the eigenvalues of A . This would occur if A were π -reversible, i.e., $\pi_i a_{ij} = \pi_j a_{ji}$ for all $i, j \in [N]$, in which case A would be a self-adjoint operator on $L_2(\pi)$; the reversible setting is the main case of interest in the ensuing discussion, but reversibility is not needed for our proofs.

Let $\mathbf{W} = \{W_t\}_{t=1}^\infty$ be a Markov chain with state space $[N]$ and transition matrix $A \in M_N(\mathbb{R})$. One says that $\mathbf{W}$ is stationary if A is $\pi_{\mathbf{W}}$ -stationary for $\pi_{\mathbf{W}} = (\Pr[W_1 = 1], \dots, \Pr[W_1 = N]) \in \Delta^{N-1}$. Write $\lambda_{\mathbf{W}} = \lambda_{\pi_{\mathbf{W}}}(A)$.

Theorem 1.1. *Suppose that $\mathbf{W} = \{W_t\}_{t=1}^\infty$ is a stationary Markov chain whose state space is $[N]$ and with $\lambda_{\mathbf{W}} < 1$. Then, every $f : [N] \rightarrow \mathbb{R}$ satisfies the following inequality for every $n \in \mathbb{N}$ and every $q \geq 2$.*

$$\left(\mathbb{E} \left[\left| \frac{f(W_1) + \dots + f(W_n)}{n} - \mathbb{E}[f(W_1)] \right|^q \right] \right)^{\frac{1}{q}} \lesssim \sqrt{\frac{q}{(1 - \lambda_{\mathbf{W}})n}} \cdot \left(\mathbb{E}[|f(W_1)|^q] \right)^{\frac{1}{q}}. \quad (1)$$

The (standard) asymptotic notation $\lesssim$ that appears in (1) (as well as throughout the ensuing discussion) means the following. Given two quantities $\alpha, \beta \in [0, \infty)$, the notation $\alpha \lesssim \beta$ stands for the assertion that there exists a universal constant $C \in (0, \infty)$ for which $\alpha \leq C\beta$; this is also denoted by $\beta \gtrsim \alpha$.

The conclusion (1) of Theorem 1.1 with the random variables $f(W_1), \dots, f(W_n)$ replaced by i.i.d. random variables coincides with the classical Marcinkiewicz–Zygmund inequality [17]. Our contribution here is therefore to generalize this statement to random variables that are (images of) stationary Markov chains with a spectral gap; the i.i.d. setting is the special case $A = E_\pi$ of Theorem 1.1. The bound (1) is optimal; see Remark 4 below. A variant of Theorem 1.1 when $1 \leq q \leq 2$ appears in Remark 3 below.

The precursor (and inspiration) of Theorem 1.1 is the following theorem of Gillman [9, 10].

Theorem 1.2. *Suppose that $\mathbf{W} = \{W_t\}_{t=1}^\infty$ is a stationary Markov chain whose state space is $[N]$ and with $\lambda_{\mathbf{W}} < 1$. Then, every $f : [N] \rightarrow \mathbb{R}$ satisfies the following inequality for every $n \in \mathbb{N}$ and every $q \geq 2$.*

$$\left(\mathbb{E} \left[\left| \frac{f(W_1) + \dots + f(W_n)}{n} - \mathbb{E}[f(W_1)] \right|^q \right] \right)^{\frac{1}{q}} \lesssim \sqrt{\frac{q}{(1 - \lambda_{\mathbf{W}})n}} \cdot \max\{|f(1)|, \dots, |f(N)|\}. \quad (2)$$

Note that Theorem 1.2 is typically stated in the literature as the following concentration inequality, which is commonly called the *expander Chernoff bound*.

$$\forall a > 0, \quad \Pr \left[\left| \frac{f(W_1) + \dots + f(W_n)}{n} - \mathbb{E}[f(W_1)] \right| \geq a \max_{j \in [N]} |f(j)| \right] \lesssim e^{-c(1 - \lambda_{\mathbf{W}})na^2}, \quad (3)$$

where $c > 0$ is a universal constant. The equivalence of (2) and (3) is standard; (2) $\implies$ (3) is checked by applying Markov's inequality and optimizing over q , and (3) $\implies$ (2) follows by straightforward integration (both implications appear in Proposition 2.5.2 of the textbook [24]). The same use of Markov's inequality shows mutatis mutandis that Theorem 1.1 implies the following concentration phenomenon.

Corollary 1.3. *There is a universal constant $c > 0$ with the following property. Suppose that $\mathbf{W} = \{W_t\}_{t=1}^\infty$ is a stationary Markov chain whose state space is $[N]$ and with $\lambda_{\mathbf{W}} < 1$. Then, every $f : [N] \rightarrow \mathbb{R}$ satisfies the following inequality for every $n \in \mathbb{N}$, every $q \geq 2$ and every $0 < a \leq \sqrt{q/((1 - \lambda_{\mathbf{W}})n)}$.*

$$\Pr \left[\left| \frac{f(W_1) + \dots + f(W_n)}{n} - \mathbb{E}[f(W_1)] \right| \geq a \left(\mathbb{E}[|f(W_1)|^q] \right)^{\frac{1}{q}} \right] \lesssim e^{-c(1 - \lambda_{\mathbf{W}})na^2}.$$

Remark 1. Kloeckner investigated in [14] the question of obtaining concentration bounds such as (3) with the L_∞ norm $\max_{j \in [N]} |f(j)|$ replaced by other norms of f . As discussed in [14, Remark 2.2], the results of [14] hold in a setting that imposes structural hypotheses on the aforementioned norm of the “observable” f which notably excludes its $L_q(\pi_{\mathbf{W}})$ norm (which appears in the right-hand side of the bound (1) that we prove here), but it is noted in [14, Remark 2.2] that “classically one only makes moment assumptions on the observable.” Corollary 1.3 addresses this question, though note that [14] also covers settings that are not treated here.

The new bound (1) that we obtain differs from Gillman’s estimate (2) only in the replacement of the worst-case bound on f in the right-hand side of (2) by an average-case bound. Rather than being merely a quantitative enhancement, this improvement has conceptual significance which we achieve through a reasoning that differs substantially from the proof of (3) in [9, 10], as well as the several other proofs of (3) and its variants that appeared in the literature [6, 11, 16, 15, 12, 25, 5, 19, 8, 7, 14] (our approach was recently used in [21, 20]).

Assuming a bound on the q ’th moment of f is the appropriate setting for bounding the q ’th moment of $f(W_1) + \dots + f(W_n)$. This compatibility of the left-hand side of (1) and the right-hand side of (1) allows the resulting inequality to tensorize so as to yield dimension-independent vector-valued statements. Specifically, for any measure space (Ω, μ) , if $f : [N] \rightarrow L_q(\mu)$, then by applying (1) to the real-valued mapping $(i \in [N]) \mapsto f(i)(\omega)$ for each $\omega \in \Omega$, and then integrating the (q ’th power of) the resulting point-wise inequality, we see that (under the assumptions of Theorem 1.1),

$$\left(\mathbb{E} \left[\left\| \frac{f(W_1) + \dots + f(W_n)}{n} - \mathbb{E}[f(W_1)] \right\|_{L_q(\mu)}^q \right] \right)^{\frac{1}{q}} \lesssim \sqrt{\frac{q}{(1 - \lambda_{\mathbf{W}})n}} \cdot \left(\mathbb{E}[\|f(W_1)\|_{L_q(\mu)}^q] \right)^{\frac{1}{q}}. \quad (4)$$

The following Hilbertian statement is a consequence of (4) that deserves to be stated separately.

Corollary 1.4. *Suppose that $\mathbf{W} = \{W_t\}_{t=1}^\infty$ is a stationary Markov chain whose state space is $[N]$ and with $\lambda_{\mathbf{W}} < 1$. Let $(H, \|\cdot\|_H)$ be a Hilbert space. The following bound holds for all $n \in \mathbb{N}$, $q \geq 2$ and $f : [N] \rightarrow H$.*

$$\left(\mathbb{E} \left[\left\| \frac{f(W_1) + \dots + f(W_n)}{n} - \mathbb{E}[f(W_1)] \right\|_H^q \right] \right)^{\frac{1}{q}} \lesssim \sqrt{\frac{q}{(1 - \lambda_{\mathbf{W}})n}} \cdot \left(\mathbb{E}[\|f(W_1)\|_H^q] \right)^{\frac{1}{q}}. \quad (5)$$

Corollary 1.4 is nothing more than (4) applied to an isometric copy of H in $L_q(\mu)$, which is known to exist by [2, Chapter 12] (see also the exposition in, e.g., the textbook [1, Proposition 6.4.12]).

Since $\mathbb{E}[\|f(W_1)\|_H^q] \leq \max_{j \in [N]} \|f(j)\|_H^q$, the following corollary is a consequence Corollary 1.4 through the usual application of Markov’s inequality and then an optimization over q .

Corollary 1.5 (Hilbert space-valued expander Chernoff bound). *There is a universal constant $c > 0$ with the following property. Suppose that $\mathbf{W} = \{W_t\}_{t=1}^\infty$ is a stationary Markov chain whose state space is $[N]$ and with $\lambda_{\mathbf{W}} < 1$. Let $(H, \|\cdot\|_H)$ be a Hilbert space. If $f : [N] \rightarrow H$, then for all $n \in \mathbb{N}$ and $a > 0$ we have*

$$\Pr \left[\left\| \frac{1}{n} \sum_{i=1}^n f(W_i) - \mathbb{E}[f(W_1)] \right\|_H \geq a \max_{j \in [N]} \|f(j)\|_H \right] \lesssim e^{-c(1 - \lambda_{\mathbf{W}})na^2}. \quad (6)$$

Remark 2. Kargin studied [12] the vector-valued setting of Gillman’s theorem for functions that take values in the m -dimensional Euclidean space ℓ_2^m . The statement that is obtained in [12] is the same as that of Corollary 1.5, except that it is dimension-dependent; specifically, with the implicit constant in (6) growing to ∞ exponentially with m . Thus, the main new feature of Corollary 1.5 is that it is dimension-independent. Obtaining such a bound was a main question that [12] left open; see [12, Section 4].

Observe that estimates such as (4) can be interpreted as bounds on the operator norm of a certain linear operator between vector-valued L_q -spaces. Specifically, suppose that $(X, \|\cdot\|_X)$ is a Banach space. Let $\mathbf{W} = \{W_t\}_{t=1}^\infty$ be a stationary Markov chain whose state space is $[N]$ and with $\lambda_{\mathbf{W}} < 1$. Denote (as before) the stationary measure of $\mathbf{W}$ by $\pi_{\mathbf{W}}$ and let the transition matrix of $\mathbf{W}$ be $A = (a_{ij}) \in M_N(\mathbb{R})$. For each $n \in \mathbb{N}$ denote the associated probability measure on the trajectories of length n by $\tau_{\mathbf{W}}^n : [N]^n \rightarrow [0, 1]$. Thus, $\tau_{\mathbf{W}}^n$ is the probability measure on $[N]^n$ that is given by $\tau_{\mathbf{W}}^1 = \pi_{\mathbf{W}}$ if $n = 1$, and for $n \geq 2$,

$$\forall (i_1, \dots, i_n) \in [N]^n, \quad \tau_{\mathbf{W}}^n(i_1, \dots, i_n) \stackrel{\text{def}}{=} \Pr[(W_1, \dots, W_n) = (i_1, \dots, i_n)] = \pi_{\mathbf{W}}(i_1) a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{n-1} i_n}.$$

Define a linear operator $T_X : L_q(\pi_{\mathbf{W}}; X) \rightarrow L_q(\tau_{\mathbf{W}}^n; X)$ by setting for $f : [N] \rightarrow X$,

$$\forall (i_1, \dots, i_n) \in [N]^n, \quad T_X f(i_1, \dots, i_n) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{k=1}^n f(i_k) - \sum_{j=1}^N \pi_{\mathbf{W}}(j) f(j) \in X. \quad (7)$$

Here, and in what follows, we are using standard notation for vector-valued Lebesgue–Bochner spaces, though throughout we will need to consider only finitely supported measures, in which case measurability issues do not need to be discussed. So, if (S, σ) is a probability space with $|S| < \infty$, then the Banach space $L_q(\sigma; X)$ is the vector space of all mapping $\psi : S \rightarrow X$, equipped with the norm

$$\|\psi\|_{L_q(\sigma; X)} = \left(\sum_{s \in S} \sigma(s) \|\psi(s)\|_X^q \right)^{\frac{1}{q}}.$$

The validity of (4) under the assumptions of Theorem 1.1 is the same as the operator norm bound

$$\|T_{L_q(\mu)}\|_{L_q(\pi_{\mathbf{W}}; L_q(\mu)) \rightarrow L_q(\tau_{\mathbf{W}}^n; L_q(\mu))} \lesssim \sqrt{\frac{q}{(1 - \lambda_{\mathbf{W}})n}}. \quad (8)$$

In the same vein, Corollary 1.4 is (under the same assumptions) the same as

$$\|T_{L_2(\mu)}\|_{L_q(\pi_{\mathbf{W}}; L_2(\mu)) \rightarrow L_q(\tau_{\mathbf{W}}^n; L_2(\mu))} \lesssim \sqrt{\frac{q}{(1 - \lambda_{\mathbf{W}})n}}. \quad (9)$$

By Calderón's vector-valued extension [4] of the Riesz–Thorin [22, 23] interpolation theorem (see the monograph [3] for background on complex interpolation; the specific statement that we are using here is a combination of Theorem 4.1.2 and Theorem 5.1.2 in [3]), it follows from (8) and (9) that for every $p \in [2, q]$ we have

$$\|T_{L_p(\mu)}\|_{L_q(\pi_{\mathbf{W}}; L_p(\mu)) \rightarrow L_q(\tau_{\mathbf{W}}^n; L_p(\mu))} \lesssim \sqrt{\frac{q}{(1 - \lambda_{\mathbf{W}})n}}.$$

We record this conclusion as the following generalization of Corollary 1.4 and Corollary 1.5.

Corollary 1.6. *Suppose that $p \geq 2$ and that (Ω, μ) is a measure space. Let $\mathbf{W} = \{W_t\}_{t=1}^\infty$ be a stationary Markov chain whose state space is $[N]$ and with $\lambda_{\mathbf{W}} < 1$. If $f : [N] \rightarrow L_p(\mu)$, then for all $n \in \mathbb{N}$ and $q \geq p$,*

$$\left(\mathbb{E} \left[\left\| \frac{f(W_1) + \cdots + f(W_n)}{n} - \mathbb{E}[f(W_1)] \right\|_{L_p(\mu)}^q \right] \right)^{\frac{1}{q}} \lesssim \sqrt{\frac{q}{(1 - \lambda_{\mathbf{W}})n}} \cdot \left(\mathbb{E}[\|f(W_1)\|_{L_p(\mu)}^q] \right)^{\frac{1}{q}}. \quad (10)$$

Consequently, by the usual combination of (10) with Markov's inequality, followed by optimization over $q \geq p$, there exists a universal constant $c \in (0, \infty)$ such that

$$\forall a > 0, \quad \Pr \left[\left\| \frac{f(W_1) + \cdots + f(W_n)}{n} - \mathbb{E}[f(W_1)] \right\|_{L_p(\mu)} \geq a \max_{j \in [N]} \|f(j)\|_{L_p(\mu)} \right] \lesssim e^{p-c(1-\lambda_{\mathbf{W}})na^2}. \quad (11)$$

Remark 3. By convexity we have $\|T_{\mathbb{R}}\|_{L_1(\pi_{\mathbf{W}}) \rightarrow L_1(\tau_{\mathbf{W}}^n)} \leq 2$, since it is evident from (7) that the operator in question is the difference of two averaging operators. By interpolating this (trivial) estimate with the case $q = 2$ of Theorem 1.1 using the (scalar-valued) Riesz–Thorin interpolation theorem as above, we arrive at the following variant of Theorem 1.1 in the range $1 \leq q \leq 2$, which holds under the same assumptions.

$$\left(\mathbb{E} \left[\left| \frac{f(W_1) + \dots + f(W_n)}{n} - \mathbb{E}[f(W_1)] \right|^q \right] \right)^{\frac{1}{q}} \lesssim \left(\frac{1}{(1 - \lambda_{\mathbf{W}})n} \right)^{1 - \frac{1}{q}} \cdot \left(\mathbb{E}[|f(W_1)|^q] \right)^{\frac{1}{q}}. \quad (12)$$

Observe that when the Markov chain $\mathbf{W}$ is reversible, the case $q = 2$ of (1) is a quadratic inequality that could be directly verified in a straightforward manner by expanding both sides in an orthonormal eigenbasis of the transition matrix of $\mathbf{W}$. The more substantial content of Theorem 1.1 is therefore the case $q > 2$, which does not lend itself to such linear-algebraic reasoning.

Remark 4. Both (1) and (12) are sharp (up to the implicit universal constant factors) for large enough $n \in \mathbb{N}$. This is seen by examining the following family of Markov chains. For every $\varepsilon, \lambda \in (0, 1)$ consider the two-state Markov chain $\mathbf{W}(\lambda, \varepsilon)$ whose transition matrix equals

$$\begin{pmatrix} 1 - (1 - \lambda)(1 - \varepsilon) & (1 - \lambda)(1 - \varepsilon) \\ (1 - \lambda)\varepsilon & 1 - (1 - \lambda)\varepsilon \end{pmatrix} = \lambda I_2 + (1 - \lambda)E_{\pi(\varepsilon)} \in M_2(\mathbb{R}), \quad (13)$$

where I_2 is the 2-by-2 identity matrix and $\pi(\varepsilon) = (\varepsilon, 1 - \varepsilon) \in \Delta^1$. Then $\pi_{\mathbf{W}(\lambda, \varepsilon)} = \pi(\varepsilon)$ and $\lambda_{\mathbf{W}(\lambda, \varepsilon)} = \lambda$.

The optimality of (1) is exhibited by taking $\varepsilon = \frac{1}{2}$ and $f : \{1, 2\} \rightarrow \mathbb{R}$ that is given by $f(1) = 1 = -f(2)$. In this case, it is elementary to check that if $n \geq q/(1 - \lambda)$, then both sides of (1) are within universal constant multiples of each other. Next, the optimality of (12) is exhibited by considering $f : \{1, 2\} \rightarrow \mathbb{R}$ that is given by $f(1) = 1$ and $f(2) = 0$. In this case, it is elementary to check that if $n \geq 1/(1 - \lambda)$, then for small enough $\varepsilon > 0$ both sides of (12) are within universal constant multiples of each other. The routine computations that verify these assertions are omitted.

Remark 5. The above discussion raises the question of understanding what is required from a Banach space $(X, \|\cdot\|_X)$ so that the “Gillman phenomenon” for stationary Markov chains (or variants thereof) would hold for X -valued mappings. The present work obtains the first examples (notably, Hilbert space) of such theorems in infinite dimensions (equivalently, dimension-independent bounds). However, much more remains to be understood here. This matter is pursued in the forthcoming work [18], where it is explained how it relates to central themes in Banach space theory. Further infinite dimensional statements are derived in [18], including a treatment of (10) in the range $2 \leq q < p$ which is not covered in Corollary 1.6, through an approach that is entirely different from our reasoning here.

We end the Introduction by noting that the above results have an equivalent dual formulation that is worthwhile to work out explicitly. Given a Banach space $(X, \|\cdot\|_X)$, the operator T_X that is given in (7) has norm $K > 0$ from $L_q(\pi_{\mathbf{W}}; X)$ to $L_q(\tau_{\mathbf{W}}^n; X)$ if and only if its adjoint T_X^* has norm K from $L_{q^*}(\tau_{\mathbf{W}}^n; X^*)$ to $L_{q^*}(\pi_{\mathbf{W}}; X^*)$, where $q^* = q/(q - 1)$. This leads to the following dual formulation of Corollary 1.6, whose derivation is a mechanical unravelling of the definitions (the straightforward details are omitted).

Corollary 1.7 (adjoint of (10)). *Let $\mathbf{W} = \{W_t\}_{t=1}^{\infty}$ be a stationary Markov chain whose state space is $[N]$ and with $\lambda_{\mathbf{W}} < 1$. Fix $n \in \mathbb{N}$ and $p, q \in (1, 2]$ with $q \leq p$. For every measure space (Ω, μ) and $F : [N]^n \rightarrow L_p(\mu)$,*

$$\begin{aligned} & \left(\mathbb{E} \left[\left\| \frac{1}{n} \sum_{i=1}^n \mathbb{E}[F(W_1, \dots, W_n) | W_i] - \mathbb{E}[F(W_1, \dots, W_n)] \right\|_{L_p(\mu)}^q \right] \right)^{\frac{1}{q}} \\ & \lesssim \frac{1}{\sqrt{(q-1)(1-\lambda_{\mathbf{W}})n}} \cdot \left(\mathbb{E}[\|F(W_1, \dots, W_n)\|_{L_p(\mu)}^q] \right)^{\frac{1}{q}}. \end{aligned}$$

2 Proof of Theorem 1.1

Suppose from now on that we are in the setting of Theorem 1.1. We will write for simplicity $\lambda = \lambda_{\mathbf{W}} < 1$ and $\pi = \pi_{\mathbf{W}} \in \Delta^{N-1}$. We will also let $A = (a_{ij}) \in M_N(\mathbb{R})$ be the transition matrix of $\mathbf{W}$.

It suffices to prove (1) when $f : [N] \rightarrow \mathbb{R}$ satisfies $\mathbb{E}[f(W_1)] = 0$. Indeed, this could be then applied to the centered function $f - \mathbb{E}[f(W_1)]$ to yield the estimate

$$\begin{aligned} \left(\mathbb{E} \left[\left| \frac{f(W_1) + \dots + f(W_n)}{n} - \mathbb{E}[f(W_1)] \right|^q \right] \right)^{\frac{1}{q}} &\lesssim \sqrt{\frac{q}{(1-\lambda_{\mathbf{W}})n}} \cdot \left(\mathbb{E}[|f(W_1) - \mathbb{E}[f(W_1)]|^q] \right)^{\frac{1}{q}} \\ &\leq 2 \sqrt{\frac{q}{(1-\lambda_{\mathbf{W}})n}} \cdot \left(\mathbb{E}[|f(W_1)|^q] \right)^{\frac{1}{q}}, \end{aligned} \quad (14)$$

where the last step is the triangle inequality in $L_q(\pi)$. So, assume from now on that $\mathbb{E}[f(W_1)] = 0$. It will be convenient to define $u \in \mathbb{R}^N$ by setting $u_i = f(i)$ for all $i \in [N]$. The assumption on f becomes $\sum_{i=1}^N \pi_i u_i = 0$. Below, we will denote the diagonal matrix whose diagonal is u by $U \in M_N(\mathbb{R})$, i.e.,

$$U \stackrel{\text{def}}{=} \begin{pmatrix} u_1 & 0 & \dots & 0 \\ 0 & u_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & u_N \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} f(1) & 0 & \dots & 0 \\ 0 & f(2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & f(N) \end{pmatrix}.$$

Lemma 2.1. *For every $m \in \mathbb{N}$ we have*

$$\mathbb{E} \left[(f(W_1) + \dots + f(W_n))^{2m} \right] \leq (2m)! \sum_{\substack{v_0, \dots, v_{2m-1} \in \mathbb{N} \cup \{0\} \\ v_0 + \dots + v_{2m-1} \leq n-1}} \|UA^{v_1} UA^{v_2} \dots UA^{v_{2m-1}} u\|_{L_1(\pi)}.$$

Proof. Let V_{2m} be the set of all those vectors in $w \in [n]^{2m}$ that satisfy $1 \leq w_1 \leq w_2 \leq \dots \leq w_{2m} \leq n$. Observe that by the Markov property and stationarity, for every $w \in V_{2m}$ we have the following identity.

$$\begin{aligned} \mathbb{E} \left[\prod_{i=1}^{2m} f(W_{w_i}) \right] &= \sum_{j \in [N]^{2m}} \pi_{j_1} A_{j_1 j_2}^{w_2 - w_1} A_{j_2 j_3}^{w_3 - w_2} \dots A_{j_{2m-1} j_{2m}}^{w_{2m} - w_{2m-1}} \prod_{k=1}^{2m} u_{j_k} \\ &= \sum_{j \in [N]^{2m}} \pi_{j_1} (UA^{w_2 - w_1})_{j_1 j_2} (UA^{w_3 - w_2})_{j_2 j_3} \dots (UA^{w_{2m} - w_{2m-1}})_{j_{2m-1} j_{2m}} u_{j_{2m}} \\ &= \sum_{i \in [N]} \pi_i (UA^{w_2 - w_1} UA^{w_3 - w_2} \dots UA^{w_{2m} - w_{2m-1}} u)_i. \end{aligned}$$

So, by expanding the $(2m)$ -th power of $f(W_1) + \dots + f(W_n)$ and arranging the indices in increasing order,

$$\begin{aligned} \mathbb{E} \left[(f(W_1) + \dots + f(W_n))^{2m} \right] &\leq (2m)! \sum_{w \in V_{2m}} \left| \mathbb{E} \left[\prod_{i=1}^{2m} f(W_{w_i}) \right] \right| \\ &\leq (2m)! \sum_{w \in V_{2m}} \|UA^{w_2 - w_1} UA^{w_3 - w_2} \dots UA^{w_{2m} - w_{2m-1}} u\|_{L_1(\pi)}. \quad \square \end{aligned}$$

Remark 6. It is worthwhile to note in passing that while the proof of Lemma 2.1 relies on what may seem to be innocuous identities, the crucial step that rearranged the factors so that their indices are increasing is inherently commutative, and this is what obstructs the direct use of the ensuing proof for matrix-valued functions, namely the setting of [26, 8]; alternative routes are taken in [8, 18] but it would be interesting to investigate if a more careful reasoning along the lines of the present work could be used to treat the setting of functions that take values in Schatten–von Neuman trace classes.

Towards bounding from above each of the terms $\|UA^{v_1}UA^{v_2}\cdots UA^{v_{2m-1}}u\|_{L_1(\pi)}$ from Lemma 2.1, we record the following iterative application of Hölder's inequality and the definition of operator norms.

Lemma 2.2. *Fix $k \in \mathbb{N}$ and $q \geq k+1$. Then, for every $T_1, \dots, T_k \in M_N(\mathbb{R})$ we have*

$$\|UT_1UT_2\cdots UT_ku\|_{L_1(\pi)} \leq \|u\|_{L_q(\pi)}^{k+1} \prod_{j=1}^k \|T_j\|_{L_{\frac{2q}{q+k+1-2j}}(\pi) \rightarrow L_{\frac{2q}{q+k+1-2j}}(\pi)}.$$

Proof. Suppose that $\alpha(1), \dots, \alpha(k+1) \geq 1$ satisfy $\frac{1}{\alpha(1)} + \cdots + \frac{1}{\alpha(k+1)} \leq 1$. We claim that

$$\|UT_1UT_2\cdots UT_ku\|_{L_{\beta(0)}(\pi)} \leq \left(\prod_{i=1}^{k+1} \|u\|_{L_{\alpha(i)}(\pi)} \right) \prod_{j=1}^k \|T_j\|_{L_{\beta(j)}(\pi) \rightarrow L_{\beta(j)}(\pi)}, \quad (15)$$

where $\beta(0), \dots, \beta(k) \geq 1$ are defined by $\frac{1}{\beta(j)} = \frac{1}{\alpha(j+1)} + \cdots + \frac{1}{\alpha(k+1)}$. The proof of (15) is by induction on k .

The case $k=0$ is tautological. For the induction step, since $\frac{1}{\beta(0)} = \frac{1}{\alpha(1)} + \frac{1}{\beta(1)}$, by Hölder's inequality,

$$\|UT_1UT_2\cdots UT_ku\|_{L_{\beta(0)}(\pi)} \leq \|u\|_{L_{\alpha(1)}(\pi)} \|T_1UT_2\cdots UT_ku\|_{L_{\beta(1)}(\pi)}. \quad (16)$$

By the definition of the operator norm $\|T_1\|_{L_{\beta(1)}(\pi) \rightarrow L_{\beta(1)}(\pi)}$ we have,

$$\|T_1UT_2\cdots UT_ku\|_{L_{\beta(1)}(\pi)} \leq \|T_1\|_{L_{\beta(1)}(\pi) \rightarrow L_{\beta(1)}(\pi)} \|UT_2\cdots UT_ku\|_{L_{\beta(1)}(\pi)}. \quad (17)$$

Now (15) follows by combining (16) and (17) with the inductive hypothesis.

Choose $\alpha(1) = \alpha(k+1) = \frac{2q}{q-k+1}$ and $\alpha(2) = \cdots = \alpha(k) = q$. So,

$$\forall j \in [k], \quad \beta(j) = \frac{1}{\frac{1}{\alpha(j+1)} + \cdots + \frac{1}{\alpha(k+1)}} = \frac{1}{\frac{k-j}{q} + \frac{q-k+1}{2q}} = \frac{2q}{q+k+1-2j},$$

and $\beta(0) = 1$. Hence, with this specific setting of the parameters the bound (15) becomes

$$\|UT_1UT_2\cdots UT_ku\|_{L_1(\pi)} \leq \|u\|_{L_{\frac{2q}{q-k+1}}(\pi)}^2 \|u\|_{L_q(\pi)}^{k-1} \prod_{j=1}^k \|T_j\|_{L_{\frac{2q}{q+k+1-2j}}(\pi) \rightarrow L_{\frac{2q}{q+k+1-2j}}(\pi)}.$$

It remains to note that since $q \geq k+1$ we have $\frac{2q}{q-k+1} \leq q$, and therefore $\|u\|_{L_{\frac{2q}{q-k+1}}(\pi)} \leq \|u\|_{L_q(\pi)}$. $\square$

Fix $m \in \mathbb{N}$. Throughout what follows, it will be notationally convenient to consider each Boolean vector $s \in \{0, 1\}^{2m-1}$ as an infinite vector in $\{0, 1\}^{\mathbb{Z}}$ whose entries vanish on $\mathbb{Z} \setminus [2m-1]$, namely we use the convention $s_i = s_j = 0$ for $i \leq 0$ and $j \geq 2m$. Let $S_{2m-1} \subseteq \{0, 1\}^{2m-1}$ be all those Boolean vectors of length $2m-1$ with no two consecutive 0s, and with $s_{2m-1} = 1$, i.e.,

$$S_{2m-1} \stackrel{\text{def}}{=} \bigcap_{j=1}^{2m-1} \left\{ s \in \{0, 1\}^{2m-1} : (s_j, s_{j+1}) \neq (0, 0) \right\}.$$

For each $j \in [2m-1]$ and $s \in S_{2m-1}$ that satisfy $s_j = 1$, we define a quantity $p(s, j) \geq 1$ in the following way. Consider the consecutive run of 1s in s to which j belongs, and let $i_1(s, j)$ and $i_2(s, j)$ be the first and last indices of this run, respectively. Formally,

$$i_1(s, j) \stackrel{\text{def}}{=} \max\{i \in \{\dots, j-2, j-1\} : s_i = 0\} + 1 \quad \text{and} \quad i_2(s, j) \stackrel{\text{def}}{=} \min\{i \in \{j+1, j+2, \dots\} : s_i = 0\} - 1. \quad (18)$$

With this notation, write

$$p(s, j) \stackrel{\text{def}}{=} \frac{4m}{2m + i_1(s, j) + i_2(s, j) - 2j}. \quad (19)$$

Lemma 2.3. For every $T_1, \dots, T_{2m-1} \in M_N(\mathbb{R})$,

$$\|U(T_1 + E_\pi)U(T_2 + E_\pi) \cdots U(T_{2m-1} + E_\pi)u\|_{L_1(\pi)} \leq \|u\|_{L_{2m}(\pi)}^{2m} \sum_{s \in S_{2m-1}} \prod_{\substack{j \in [2m-1] \\ s_j=1}} \|T_j\|_{L_{p(s,j)}(\pi) \rightarrow L_{p(s,j)}(\pi)}. \quad (20)$$

Proof. For each $j \in [2m-1]$, write $T_{j,0} = E_\pi$ and $T_{j,1} = T_j$. Observe that

$$\forall s \in \{0, 1\}^{2m-1} \setminus S_{2m-1}, \quad UT_{1,s_1}UT_{2,s_2} \cdots UT_{2m-1,s_{2m-1}}u = 0. \quad (21)$$

Indeed, if $s \in \{0, 1\}^{2m-1} \setminus S_{2m-1}$, then either $s_{2m-1} = 0$, in which case $T_{2m-1,s_{2m-1}}u = E_\pi u = \mathbf{0} \in \mathbb{R}^N$, or $s_j = s_{j+1} = 0$ for some $j \in [2m-2]$, in which case $T_{j,s_j}UT_{j+1,s_{j+1}} = E_\pi UE_\pi = \mathbf{0} \in M_N(\mathbb{R})$, where both identities are equivalent to the assumption $\sum_{i=1}^N \pi_i u_i = 0$. Now,

$$\begin{aligned} \|U(T_1 + E_\pi)U(T_2 + E_\pi) \cdots U(T_{2m-1} + E_\pi)u\|_{L_1(\pi)} &= \left\| \sum_{s \in \{0,1\}^{2m-1}} UT_{1,s_1}UT_{2,s_2} \cdots UT_{2m-1,s_{2m-1}}u \right\|_{L_1(\pi)} \\ &\leq \sum_{s \in \{0,1\}^{2m-1}} \|UT_{1,s_1}UT_{2,s_2} \cdots UT_{2m-1,s_{2m-1}}u\|_{L_1(\pi)} \\ &\stackrel{(21)}{=} \sum_{s \in S_{2m-1}} \|UT_{1,s_1}UT_{2,s_2} \cdots UT_{2m-1,s_{2m-1}}u\|_{L_1(\pi)}. \end{aligned} \quad (22)$$

Fix $s \in S_{2m-1}$ and let $1 \leq r_1 < r_2 < \cdots < r_\ell < 2m-1$ be all of the indices at which s vanishes. Define $R_1, \dots, R_{\ell+1} \in M_N(\mathbb{R})$ by setting

$$R_1 \stackrel{\text{def}}{=} (UT_1)(UT_2) \cdots (UT_{r_1-1}), \quad R_{\ell+1} \stackrel{\text{def}}{=} (UT_{r_\ell+1})(UT_{r_\ell+2}) \cdots (UT_{2m-1}),$$

and

$$R_\kappa \stackrel{\text{def}}{=} (UT_{r_{\kappa-1}+1})(UT_{r_{\kappa-1}+2}) \cdots (UT_{r_\kappa-1})$$

for $\kappa \in \{2, \dots, \ell\}$. Using the fact that $UE_\pi v = (\sum_{i=1}^N \pi_i v_i)u$ for every $v \in \mathbb{R}^N$, we have the following identity.

$$UT_{1,s_1}UT_{2,s_2} \cdots UT_{2m-1,s_{2m-1}}u = R_1(UE_\pi)R_2(UE_\pi)R_3 \cdots (UE_\pi)R_{\ell+1}u = \left(\prod_{\kappa=2}^{\ell+1} \sum_{i=1}^N \pi_i (R_\kappa u)_i \right) R_1 u.$$

Consequently,

$$\|UT_{1,s_1}UT_{2,s_2} \cdots UT_{2m-1,s_{2m-1}}u\|_{L_1(\pi)} = \|R_1 u\|_{L_1(\pi)} \prod_{\kappa=2}^{\ell+1} \left| \sum_{i=1}^N \pi_i (R_\kappa u)_i \right| \leq \prod_{\kappa=1}^{\ell+1} \|R_\kappa u\|_{L_1(\pi)}. \quad (23)$$

Next, by Lemma 2.2 with $q = 2m$ and $k = r_1 - 1$ we have

$$\begin{aligned} \|R_1 u\|_{L_1(\pi)} &= \|(UT_1)(UT_2) \cdots (UT_{r_1-1})u\|_{L_1(\pi)} \\ &\leq \|u\|_{L_{2m}(\pi)}^{r_1} \prod_{j=1}^{r_1-1} \|T_j\|_{L_{\frac{4m}{2m+r_1-2j}}(\pi) \rightarrow L_{\frac{4m}{2m+r_1-2j}}(\pi)} \\ &\stackrel{(19)}{=} \|u\|_{L_{2m}(\pi)}^{r_1} \prod_{j=1}^{r_1-1} \|T_j\|_{L_{p(s,j)}(\pi) \rightarrow L_{p(s,j)}(\pi)}. \end{aligned}$$

In the same vein, for every $k \in \{2, \dots, \ell\}$,

$$\|R_\kappa u\|_{L_1(\pi)} \leq \|u\|_{L_{2m}(\pi)}^{r_\kappa - r_{\kappa-1}} \prod_{j=r_{\kappa-1}+1}^{r_\kappa-1} \|T_j\|_{L_{p(s,j)}(\pi) \rightarrow L_{p(s,j)}(\pi)},$$

and also

$$\|R_{\ell+1}u\|_{L_1(\pi)} \leq \|u\|_{L_{2m}(\pi)}^{2m-r_\ell} \prod_{j=r_\ell+1}^{2m-1} \|T_j\|_{L_{p(s,j)}(\pi) \rightarrow L_{p(s,j)}(\pi)}.$$

We therefore have

$$\prod_{\kappa=1}^{\ell+1} \|R_\kappa u\|_{L_1(\pi)} \leq \|u\|_{L_{2m}(\pi)}^{2m} \prod_{\substack{j \in [2m-1] \\ s_j=1}} \|T_j\|_{L_{p(s,j)}(\pi) \rightarrow L_{p(s,j)}(\pi)}. \quad (24)$$

By substituting (24) into (23) and then substituting the resulting estimate into (22), we arrive at (20). $\square$

In light of Lemma 2.1, the following lemma is highly relevant to our goal of proving Theorem 1.1.

Lemma 2.4. *Suppose that $m \in \mathbb{N}$ satisfies $em \leq n(1-\lambda)$. Then,*

$$\left(\sum_{\substack{v_0, \dots, v_{2m-1} \in \mathbb{N} \cup \{0\} \\ v_0 + \dots + v_{2m-1} \leq n-1}} \|UA^{v_1} UA^{v_2} \dots UA^{v_{2m-1}} u\|_{L_1(\pi)} \right)^{\frac{1}{2m}} \lesssim \frac{\sqrt{n/m}}{\sqrt{1-\lambda}} \|u\|_{L_{2m}(\pi)}. \quad (25)$$

Proof. Fix $v_0, \dots, v_{2m-1} \in \mathbb{N} \cup \{0\}$ and denote $T_j = A^{v_j} - E_\pi$ for every $j \in \{0, \dots, 2m-1\}$. Then,

$$\begin{aligned} \|UA^{v_1} UA^{v_2} \dots UA^{v_{2m-1}} u\|_{L_1(\pi)} &= \|U(T_1 + E_\pi)U(T_2 + E_\pi) \dots U(T_{2m-1} + E_\pi)u\|_{L_1(\pi)} \\ &\leq \|u\|_{L_{2m}(\pi)}^{2m} \sum_{s \in S_{2m-1}} \prod_{\substack{j \in [2m-1] \\ s_j=1}} \|T_j\|_{L_{p(s,j)}(\pi) \rightarrow L_{p(s,j)}(\pi)}, \end{aligned} \quad (26)$$

where the last step of (26) is an application of Lemma 2.3.

Fixing $j \in \{0, \dots, 2m-1\}$, note that $AE_\pi = E_\pi$ since A is stochastic and the columns of E_π are constant, and also $E_\pi A = E_\pi$ since A is π -stationary. Consequently $T_j = A^{v_j} - E_\pi = (A - E_\pi)^{v_j}$. So, for every $p \geq 1$,

$$\|T_j\|_{L_p(\pi) \rightarrow L_p(\pi)} = \|(A - E_\pi)^{v_j}\|_{L_p(\pi) \rightarrow L_p(\pi)} \leq \|A - E_\pi\|_{L_p(\pi) \rightarrow L_p(\pi)}^{v_j}. \quad (27)$$

By definition, $\|A - E_\pi\|_{L_2(\pi) \rightarrow L_2(\pi)} = \lambda$. As A and E_π are averaging operators, by convexity and the triangle inequality $\|A - E_\pi\|_{L_r(\pi) \rightarrow L_r(\pi)} \leq \|A\|_{L_r(\pi) \rightarrow L_r(\pi)} + \|E_\pi\|_{L_r(\pi) \rightarrow L_r(\pi)} = 2$ for all $r \geq 1$. By the Riesz–Thorin interpolation theorem [22, 23] (see e.g. Chapter IV in the textbook [13]), this implies that

$$\|A - E_\pi\|_{L_p(\pi) \rightarrow L_p(\pi)} \leq 2\lambda^{2\min\{\frac{1}{p}, 1-\frac{1}{p}\}}. \quad (28)$$

A substitution of (28) into (27), followed by a substitution of the resulting bound into (26) shows that in order to prove the desired inequality (25) it suffices to establish the following estimate.

$$\left(\sum_{s \in S_{2m-1}} \sum_{\substack{v_0, \dots, v_{2m-1} \in \mathbb{N} \cup \{0\} \\ v_0 + \dots + v_{2m-1} \leq n-1}} \prod_{\substack{j \in [2m-1] \\ s_j=1}} \lambda^{\beta(s,j)v_j} \right)^{\frac{1}{m}} \lesssim \frac{n}{m(1-\lambda)}, \quad (29)$$

where for every $s \in S_{2m-1}$ and $j \in [2m-1]$ such that $s_j = 1$, we denote

$$\beta(s, j) \stackrel{\text{def}}{=} 2\min\left\{\frac{1}{p(s, j)}, 1 - \frac{1}{p(s, j)}\right\} \stackrel{(19)}{=} 1 - \frac{|i_1(s, j) + i_2(s, j) - 2j|}{2m}. \quad (30)$$

Fix some $s \in S_{2m-1}$. Denote $Q_0 = \{j \in [2m-1] : s_j = 0\}$ and $Q_1 = [2m-1] \setminus Q_0$. Thus $|Q_0| + |Q_1| = 2m-1$ and by the definition of S_{2m-1} we have $|Q_1| \geq m$. With this notation, we have the following bound.

$$\begin{aligned}
& \sum_{\substack{v_0, \dots, v_{2m-1} \in \mathbb{N} \cup \{0\} \\ v_0 + \dots + v_{2m-1} \leq n-1}} \prod_{\substack{j \in [2m-1] \\ s_j = 1}} \lambda^{\beta(s, j) v_j} \\
&= \sum_{\substack{(v_i)_{i \in \{0\} \cup Q_0} \in (\mathbb{N} \cup \{0\})^{\{0\} \cup Q_0} \\ \sum_{i \in \{0\} \cup Q_0} v_i \leq n-1}} \sum_{\substack{(v_j)_{j \in Q_1} \in (\mathbb{N} \cup \{0\})^{Q_1} \\ \sum_{j \in Q_1} v_j \leq n-1 - \sum_{i \in \{0\} \cup Q_0} v_i}} \prod_{j \in Q_1} \lambda^{\beta(s, j) v_j} \\
&\leq \left| \left\{ (v_i)_{i \in \{0\} \cup Q_0} \in (\mathbb{N} \cup \{0\})^{\{0\} \cup Q_0} : \sum_{i \in \{0\} \cup Q_0} v_i \leq n-1 \right\} \right| \cdot \sum_{(v_j)_{j \in Q_1} \in (\mathbb{N} \cup \{0\})^{Q_1}} \prod_{j \in Q_1} \lambda^{\beta(s, j) v_j} \\
&= \sum_{\ell=0}^{n-1} \binom{|Q_0| + \ell}{|Q_0|} \prod_{j \in Q_1} \sum_{i=0}^{\infty} \lambda^{\beta(s, j) i} \\
&= \binom{|Q_0| + n}{|Q_0| + 1} \prod_{j \in Q_1} \frac{1}{1 - \lambda^{\beta(s, j)}}.
\end{aligned}$$

By the elementary inequality $1 - \lambda^\beta \geq \beta(1 - \lambda)$, which holds for every $\lambda, \beta \in [0, 1]$, it follows from this that

$$\begin{aligned}
\sum_{\substack{v_0, \dots, v_{2m-1} \in \mathbb{N} \cup \{0\} \\ v_0 + \dots + v_{2m-1} \leq n-1}} \prod_{\substack{j \in [2m-1] \\ s_j = 1}} \lambda^{\beta(s, j) v_j} &\leq \frac{1}{(1 - \lambda)^{|Q_1|}} \binom{|Q_0| + n}{|Q_0| + 1} \prod_{j \in Q_1} \frac{1}{\beta(s, j)} \\
&= \frac{(1 - \lambda)^{|Q_0| + 1}}{(1 - \lambda)^{2m}} \binom{|Q_0| + n}{|Q_0| + 1} \prod_{j \in Q_1} \frac{1}{\beta(s, j)} \\
&\lesssim \frac{e^{O(m)}}{(1 - \lambda)^{2m}} \left(\frac{(1 - \lambda)n}{|Q_0| + 1} \right)^{|Q_0| + 1} \prod_{j \in Q_1} \frac{1}{\beta(s, j)},
\end{aligned} \tag{31}$$

where the last step follows from a straightforward application of Stirling's formula. Consider the function $\psi : [0, \infty) \rightarrow [0, \infty)$ that is given by $\psi(z) = ((1 - \lambda)n/z)^z$. Then $(\log \psi(z))' = \log((1 - \lambda)n/(ez))$. Hence, ψ is increasing on the interval $[0, (1 - \lambda)n/e]$. But $|Q_0| + 1 = 2m - |Q_1| \leq m \leq (1 - \lambda)n/e$, by the assumption on m in the statement of Lemma 2.4. Hence $\psi(|Q_0| + 1) \leq \psi(m)$, and therefore

$$\frac{1}{(1 - \lambda)^{2m}} \left(\frac{(1 - \lambda)n}{|Q_0| + 1} \right)^{|Q_0| + 1} = \frac{\psi(|Q_0| + 1)}{(1 - \lambda)^{2m}} \leq \frac{\psi(m)}{(1 - \lambda)^{2m}} = \frac{(n/m)^m}{(1 - \lambda)^m}. \tag{32}$$

We will show next that

$$\prod_{j \in Q_1} \frac{1}{\beta(s, j)} \stackrel{(30)}{=} \prod_{j \in Q_1} \left(1 - \frac{|i_1(s, j) + i_2(s, j) - 2j|}{2m} \right)^{-1} \leq e^{O(m)}. \tag{33}$$

In combination with (31) and (32), this would imply the desired inequality (29) because $|S_{2m-1}| \leq e^{O(m)}$.

For each $j \in Q_1$ with $i_2(s, j) - i_1(s, j) \leq \frac{3m}{2}$ (i.e., the consecutive run of 1s in s to which j belongs is of length at most $1 + \frac{3m}{2}$), we have $|i_1(s, j) + i_2(s, j) - 2j| \leq \frac{3m}{2}$ and therefore its contribution to the product in (33) is at most 4. So, (33) holds if there are no runs of 1s in s of length greater than $\frac{3m}{2}$. Otherwise, there is exactly one run of 1s in s of length $d > \frac{3m}{2}$, and its contribution to the product in (33) equals

$$\prod_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \left(\frac{2m}{2m - d + 1 + 2i} \right)^2 \leq \prod_{i=0}^{2\lfloor \frac{d-1}{2} \rfloor} \frac{2m}{2m - d + 1 + i} \leq \prod_{k=2}^{2m} \frac{2m}{k} = \frac{(2m)^{2m-1}}{(2m)!} \leq e^{O(m)},$$

where the last step follows from Stirling's formula. This proves our goal (33). $\square$

Completion of the proof of Theorem 1.1. By the triangle inequality in L_q (and stationarity) we have

$$\left(\mathbb{E}[|f(W_1) + \dots + f(W_n)|^q]\right)^{\frac{1}{q}} \leq \left(\mathbb{E}[|f(W_1)|^q]\right)^{\frac{1}{q}} + \dots + \left(\mathbb{E}[|f(W_n)|^q]\right)^{\frac{1}{q}} = n\left(\mathbb{E}[|f(W_1)|^q]\right)^{\frac{1}{q}}.$$

This bound implies the desired estimate (1) when $q \gtrsim (1 - \lambda)n$, so we may assume from now on that $q \leq (1 - \lambda)n/e$. Let $m \in \mathbb{N}$ be the largest integer such that $2m \leq q$. Then, $m, m + 1 \leq q \leq (1 - \lambda)n/e$, so the conclusion of Lemma 2.4 holds for both m and $m + 1$. By Lemma 2.1 (and Stirling's formula), this gives

$$\left(\mathbb{E}\left[(f(W_1) + \dots + f(W_n))^{2m}\right]\right)^{\frac{1}{2m}} \lesssim \sqrt{\frac{nm}{1 - \lambda}} \left(\mathbb{E}[|f(W_1)|^{2m}]\right)^{\frac{1}{2m}} \leq \sqrt{\frac{nq}{1 - \lambda}} \left(\mathbb{E}[|f(W_1)|^{2m}]\right)^{\frac{1}{2m}},$$

and similarly

$$\left(\mathbb{E}\left[(f(W_1) + \dots + f(W_n))^{2(m+1)}\right]\right)^{\frac{1}{2(m+1)}} \lesssim \sqrt{\frac{nq}{1 - \lambda}} \left(\mathbb{E}[|f(W_1)|^{2(m+1)}]\right)^{\frac{1}{2(m+1)}}.$$

As in (14), it follows from these bounds (which we derived under the assumption $\mathbb{E}[f(W_1)] = 0$) that the norm of the operator $T_{\mathbb{R}}$ that is given in (7) is bounded by a universal constant multiple of $\sqrt{q/((1 - \lambda)n)}$ both from $L_{2m}(\pi)$ to $L_{2m}(\pi)$ and from $L_{2(m+1)}(\pi)$ to $L_{2(m+1)}(\pi)$. Since $2m \leq q \leq 2(m + 1)$, another application of the Riesz–Thorin theorem gives that the norm of $T_{\mathbb{R}}$ from $L_q(\pi)$ to $L_q(\pi)$ is also bounded by a universal constant multiple of $\sqrt{q/((1 - \lambda)n)}$. This is precisely the desired bound (1). $\square$

References

- [1] F. Albiac and N. J. Kalton. *Topics in Banach space theory*, volume 233 of *Graduate Texts in Mathematics*. Springer, [Cham], second edition, 2016. With a foreword by Gilles Godefory.
- [2] S. Banach. *Théorie des opérations linéaires.*, volume 1. PWN - Państwowe Wydawnictwo Naukowe, Warszawa, 1932.
- [3] J. Bergh and J. Löfström. *Interpolation spaces. An introduction*. Springer-Verlag, Berlin-New York, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223.
- [4] A.-P. Calderón. Intermediate spaces and interpolation, the complex method. *Studia Math.*, 24:113–190, 1964.
- [5] 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.
- [6] I. H. Dinwoodie. A probability inequality for the occupation measure of a reversible Markov chain. *Ann. Appl. Probab.*, 5(1):37–43, 1995.
- [7] J. Fan, B. Jiang, and Q. Sun. Hoeffding's lemma for Markov chains and its applications to statistical learning, 2018.
- [8] A. Garg, Y. T. Lee, Z. Song, and N. Srivastava. A matrix expander Chernoff bound. In *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1102–1114. ACM, New York, 2018.

- [9] D. Gillman. A Chernoff bound for random walks on expander graphs. In *34th Annual Symposium on Foundations of Computer Science (Palo Alto, CA, 1993)*, pages 680–691. IEEE Comput. Soc. Press, Los Alamitos, CA, 1993.
- [10] D. Gillman. A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.*, 27(4):1203–1220, 1998.
- [11] N. Kahale. Large deviation bounds for Markov chains. *Combin. Probab. Comput.*, 6(4):465–474, 1997.
- [12] V. Kargin. A large deviation inequality for vector functions on finite reversible Markov chains. *Ann. Appl. Probab.*, 17(4):1202–1221, 2007.
- [13] Y. Katznelson. *An introduction to harmonic analysis*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2004.
- [14] B. Kloeckner. Effective Berry-Esseen and concentration bounds for Markov chains with a spectral gap. *Ann. Appl. Probab.*, 29(3):1778–1807, 2019. arXiv:1703.09623.
- [15] C. A. León and F. Perron. Optimal Hoeffding bounds for discrete reversible Markov chains. *Ann. Appl. Probab.*, 14(2):958–970, 2004.
- [16] P. Lezaud. Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.*, 8(3):849–867, 1998.
- [17] J. Marcinkiewicz and A. Zygmund. Sur les fonctions indépendantes. *Fundamenta Mathematicae*, 29(1):60–90, 1937.
- [18] A. Naor. Rates in the vector-valued law of large numbers for Markov chains. Preprint, 2019.
- [19] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.*, 20:32 pp., 2015.
- [20] S. Rao. A Hoeffding inequality for Markov chains. *Electron. Commun. Probab.*, 24(14):1–11, 2019.
- [21] S. Rao and O. Regev. A sharp tail bound for the expander random sampler, 2017. arXiv:1703.10205.
- [22] M. Riesz. Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires. *Acta Math.*, 49(3-4):465–497, 1927.
- [23] 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.
- [24] R. Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018.
- [25] R. Wagner. Tail estimates for sums of variables sampled by a random walk. *Comb. Probab. Comput.*, 17(2):307–316, Mar. 2008. arXiv:math/0608740.
- [26] A. Wigderson and D. Xiao. Derandomizing the Ahlswede-Winter matrix-valued Chernoff bound using pessimistic estimators, and applications. *Theory Comput.*, 4:53–76, 2008.