Matrix Hoeffding and Bernstein Bounds with Sharp Constants for Markov Chains

Zhao Song*

August 18, 2026

Abstract

Matrix concentration for Markov chains was initiated in the expander-walk setting by Garg, Lee, Song, and Srivastava'18 [GLSS18]. However, the constant obtained in [GLSS18] is quite loose, and it is natural to ask whether a tighter proof can yield the same constant as in the independent matrix concentration setting. In this paper, we provide a positive answer to this question. Our Hoeffding exponent is sharp, as shown by a scalar obstruction. Our Chernoff and Bernstein constants improve upon those in [GLSS18] and Neeman, Shi, and Ward'24 [NSW24], respectively.

1 Introduction

Concentration inequalities provide quantitative control over the deviations of random sums from their expectations and play a central role in probability theory and computer science. For sums of independent bounded random variables, the classical inequalities of Chernoff [Che52], Hoeffding [Hoe63], and Bernstein [Ber24] give exponentially decaying tail bounds, with Hoeffding's bound depending on the ranges of the summands and Bernstein's bound additionally exploiting their variances.

A foundational matrix Chernoff inequality for independent matrix-valued summands was established by Ahlswede and Winter [AW02]. Tropp later developed a systematic collection of Chernoff-, Hoeffding-, and Bernstein-type inequalities for independent random matrices [Tro12]. In parallel, scalar concentration for Markov chains developed from Gillman's expander-walk bound and Healy's generalization [Gil98, Hea08]. Lezaud and León–Perron obtained spectral Chernoff–Hoeffding bounds for finite-state chains, while Chung, Lam, Liu, and Mitzenmacher treated general nonreversible finite-state chains [Lez98, LP04, CLLM12]. Paulin established concentration inequalities using coupling and spectral methods, including Bernstein-type bounds [Pau15]. Subsequent results include general-state-space Bernstein and Hoeffding inequalities [JSF18, FJS21], time-dependent Hoeffding bounds for reversible finite-state chains [Rao19], and finite-sample Chernoff bounds attaining the optimal large-deviation rate [MA19].

Wigderson and Xiao initiated the study of randomness-efficient sampling for matrix-valued functions along expander walks [WX05, WX08]. Wigderson and Xiao conjectured a matrix Chernoff bound for expander walks [WX05, WX08]. Garg, Lee, Song, and Srivastava later proved the conjectured matrix Chernoff bound for expander walks [GLSS18]. Qiu, Wang, Liao, Peng, and Tang then established a matrix Chernoff bound for regular finite Markov chains and applied it to co-occurrence matrices [QWL+20].

*magic.linuxkde@gmail.com.

Neeman, Shi, and Ward extended this line of work to time-dependent matrix-valued observables of stationary, possibly nonreversible Markov chains on general state spaces, obtaining Hoeffding- and Bernstein-type inequalities [NSW24]. More recent directions include spectral concentration under ψ-mixing [WS26], Banach-space concentration for reversible finite chains [Rao26], matrix Hoeffding and Bernstein inequalities for ergodic chains [PXZ25], and matrix Chernoff inequalities for inhomogeneous chains [Zan26]. Sen also proved trace-distance concentration for quantum soft covering along expander walks [Sen25].

A parallel line of work concerns multivariate trace inequalities. For Hermitian matrices A and B, the classical Golden–Thompson inequality, proved independently by Golden and Thompson, states that $\mathrm{tr}[\exp(A+B)] \le \mathrm{tr}[\exp(A)\exp(B)]$ [Gol65, Tho65]. Sutter, Berta, and Tomamichel established a multivariate Golden–Thompson inequality whose integral representation ranges over the entire real line and uses complex phases [SBT17]. Garg, Lee, Song, and Srivastava established a variant with integration over a bounded interval in their analysis of expander walks [GLSS18].

1.1 Our Results

The constant obtained in [GLSS18] is quite loose, and it is natural to ask whether a tighter proof can yield the same constant as in the independent matrix concentration setting [Son19]. In this paper, we provide a positive answer to this question.

Theorem 1.1 (Markov matrix Hoeffding bound, informal version of Theorem 4.1)**.** *Let $s_1, \dots, s_n$ be a stationary trajectory of a Markov chain with L^2 contraction $\lambda < 1$. Suppose that the Hermitian matrices $F_j(s_j)$ are centered and satisfy $a_j I \preceq F_j(x) \preceq b_j I$. We define $\alpha(\lambda) := (1+\lambda)/(1-\lambda)$. Then, for every $t \ge 0$,*

$$\Pr_{\mu}[\lambda_{\max}(\sum_{j=1}^{n} F_j(s_j)) \ge t] \le d \exp(-\frac{2t^2}{\alpha(\lambda)\sum_{j=1}^{n}(b_j-a_j)^2}).$$

When $\lambda = 0$, we have $P = \Pi$, so the trajectory consists of independent samples from μ. The theorem then gives the independent matrix Hoeffding bound obtained from the interval-width matrix moment generating function estimate and the matrix Laplace-transform method [Tro12, Corollary 3.7].

Theorem 1.2 (Markov matrix Bernstein bound, informal version of Theorem 5.1)**.** *Under the same Markov-chain assumptions, suppose that the centered Hermitian matrices satisfy $\|F_j(x)\| \le \mathcal{M}$ and $\|\mathbb{E}_{\mu}[F_j^2]\| \le \mathcal{V}_j$. Set $\sigma^2 := \sum_{j=1}^{n} \mathcal{V}_j$ and define $\alpha(\lambda) := (1+\lambda)/(1-\lambda)$. If $0 < \lambda < 1$, then, for every $t \ge 0$,*

$$\Pr_{\mu}[\lambda_{\max}(\sum_{j=1}^{n} F_j(s_j)) \ge t] \le d \exp(-\frac{t^2/2}{\alpha(\lambda)\sigma^2 + 2\mathcal{M}t/(1-\lambda)}).$$

For $\lambda = 0$, the denominator in the exponent is $\sigma^2 + \mathcal{M}t/3$.

When $\lambda = 0$, the trajectory consists of independent samples from μ, and the bound recovers the independent matrix Bernstein inequality [Tro12, Theorem 1.4]. The coefficient $\alpha(\lambda)$ multiplying the variance proxy is optimal in the small-deviation regime.

Corollary 1.3 (Expander-Chernoff normalization, informal version of Corollary 4.2)**.** *Under the same Markov-chain assumptions, suppose that $-I \preceq F_j(x) \preceq I$ for every j and x. Then, for every $\epsilon \ge 0$,*

$$\Pr_{\mu}[\lambda_{\max}(\frac{1}{n}\sum_{j=1}^{n} F_j(s_j)) \ge \epsilon] \le d \exp(-\frac{\epsilon^2(1-\lambda)n}{2(1+\lambda)}).$$

When $\lambda = 0$, the trajectory consists of independent samples from μ. Applying the independent matrix Chernoff inequality [Tro12, Theorem 5.1] to $X_j := (F_j + I)/2$, followed by the standard quadratic lower bound for binary relative entropy, recovers this estimate.

Remark 1.4 (Comparison with [GLSS18] and [NSW24])**.** In the expander-walk setting, [GLSS18, Theorem 4.1] has prefactor $d^{2-\pi/4}$ and exponent $-\epsilon^2(1-\lambda)n/80$. Since every probability is at most one, [GLSS18, Remark 4.2] converts this to a bound with prefactor d and exponent

$$-\frac{\epsilon^2(1-\lambda)n}{80(2-\pi/4)}.$$

At the common prefactor d, Corollary 1.3 therefore improves the exponent by the factor $40(2-\pi/4)/(1+\lambda)$.

For real symmetric matrices, [NSW24, Theorem 2.5] has prefactor $d^{2-\pi/4}$ and exponent

$$-\frac{\pi^2 t^2}{8\alpha(\lambda)\sum_{j=1}^n (b_j - a_j)^2}.$$

The same prefactor–exponent tradeoff gives numerator constant $\pi^2/(8(2-\pi/4))$ at prefactor d. Thus, at the common prefactor d, Theorem 1.1 improves the exponent by the factor $16(2-\pi/4)/\pi^2$. Specifically, after Eq. (6.12), [NSW24] replaces the phase factor $\cos^2(\phi)$ by 1 before applying the bounded multivariate trace inequality; this is the source of the conversion loss. For complex Hermitian matrices, [NSW24, Corollary 2.7] has an additional factor 2, whereas Theorem 1.1 retains prefactor d.

2 Preliminaries

For matrices, $\|\cdot\|$ denotes the operator norm, while $\|\cdot\|_2$ denotes the Frobenius norm, equivalently the Schatten 2-norm. We use i to denote $\sqrt{-1}$. We use $\otimes$ to denote the tensor product. We use $\mathrm{tr}[]$ to denote the trace of a matrix.

Definition 2.1 (Markov matrix setup)**.** *Let P be a Markov kernel on a state space $\mathcal{X}$ with stationary probability measure μ. Let $\Pi f := \mathbb{E}_\mu[f]$, and assume*

$$\|P - \Pi\|_{L^2(\mu)\to L^2(\mu)} \le \lambda < 1.$$

Let $s_1, \dots, s_n$ be a stationary trajectory of P. Throughout the paper, $F_j : \mathcal{X} \to \mathbb{H}_d$ is measurable and centered:

$$\mathbb{E}_\mu[F_j] = 0.$$

We use the abbreviations

$$S_F := \sum_{j=1}^n F_j(s_j), \qquad \alpha(\lambda) := \frac{1+\lambda}{1-\lambda}.$$

We state a tool from previous work.

Lemma 2.2 (Sutter–Berta–Tomamichel, Corollary 3.3 of [SBT17])**.** *Let $H_1, \dots, H_n$ be Hermitian matrices. There is a probability density β_0 on $\mathbb{R}$ such that*

$$\log \|\exp(\sum_{j=1}^n H_j)\|_2 \le \int_{\mathbb{R}} \beta_0(u) \log \|\prod_{j=1}^n \exp((1+iu)H_j)\|_2 \mathrm{d}u.$$

We also use the realification map for a complex Hermitian matrix $Z = X + iY$, where X is real symmetric and Y is real skew-symmetric:

$$\mathcal{C}(Z) := \begin{pmatrix} X & -Y \\ Y & X \end{pmatrix}. \tag{1}$$

The map $\mathcal{C}$ preserves sums, products, exponentials, operator norms, and Loewner-order interval bounds. Every eigenvalue of Z occurs twice as an eigenvalue of $\mathcal{C}(Z)$, and hence

$$\mathrm{tr}_{\mathbb{R}}[\exp(\mathcal{C}(Z))] = 2\,\mathrm{tr}_{\mathbb{C}}[\exp(Z)]. \tag{2}$$

3 Shared trace-MGF conversion

Lemma 3.1 (SBT-to-trace-MGF conversion)**.** *Under Definition 2.1, suppose that each $F_j(x)$ is real symmetric and that there are finite constants $M_1, \ldots, M_n$ such that $\|F_j(x)\| \le M_j$ for every j and x. Let $\mathcal{J} \subseteq (0, \infty)$, let $D > 0$, and let $\Psi : \mathcal{J} \to \mathbb{R}$. Suppose that, whenever $r > 0$ and $\phi \in (-\pi/2, \pi/2)$ satisfy $r\cos(\phi) \in \mathcal{J}$,*

$$\mathbb{E}_\mu[\| \prod_{j=1}^{n} \exp(\frac{re^{i\phi}}{2} F_j(s_j))\|_2^2] \le D \exp(\Psi(r\cos(\phi))). \tag{3}$$

Then, for every $\theta \in \mathcal{J}$,

$$\mathbb{E}_\mu[\mathrm{tr}[\exp(\theta S_F)]] \le D \exp(\Psi(\theta)). \tag{4}$$

Proof. Fix $\theta \in \mathcal{J}$ and apply Lemma 2.2 with $H_j := \theta F_j(s_j)/2$. For every $u \in \mathbb{R}$, define

$$Q_u := \| \prod_{j=1}^{n} \exp(\frac{\theta(1+iu)}{2} F_j(s_j))\|_2^2.$$

Then we can show,

$$\begin{aligned} \log \mathrm{tr}[\exp(\theta S_F)] &= 2 \log \| \exp(\frac{\theta S_F}{2})\|_2 \\ &\le 2 \int_{\mathbb{R}} \beta_0(u) \log \| \prod_{j=1}^{n} \exp(\frac{\theta(1+iu)}{2} F_j(s_j))\|_2 \mathrm{d}u \\ &= \int_{\mathbb{R}} \beta_0(u) \log Q_u \mathrm{d}u. \end{aligned} \tag{5}$$

where the first step follows from $\| \exp(\theta S_F/2)\|_2^2 = \mathrm{tr}[\exp(\theta S_F)]$, the second step follows from Lemma 2.2, and the third step follows from the definition of Q_u. Moreover,

$$Q_u \le d \exp(\theta \sum_{j=1}^{n} M_j),$$

so Q_u is uniformly bounded in u for each trajectory. Since β_0 is a probability density, we can show,

$$\begin{aligned} \mathrm{tr}[\exp(\theta S_F)] &\le \exp(\int_{\mathbb{R}} \beta_0(u) \log Q_u \mathrm{d}u) \\ &\le \int_{\mathbb{R}} \beta_0(u) \exp(\log Q_u) \mathrm{d}u \end{aligned}$$

$$= \int_{\mathbb{R}} \beta_0(u) Q_u \mathrm{d}u. \tag{6}$$

where the first step follows from exponentiating Eq. (5), the second step follows from Jensen's inequality for the convex function exp, and the third step follows from $\exp(\log Q_u) = Q_u$.

Next, we can show

$$\mathbb{E}_\mu[\mathrm{tr}[\exp(\theta S_F)]] \le \mathbb{E}_\mu[\int_{\mathbb{R}} \beta_0(u) Q_u \mathrm{d}u] = \int_{\mathbb{R}} \beta_0(u) \mathbb{E}_\mu[Q_u] \mathrm{d}u. \tag{7}$$

where the first step follows from taking expectations in Eq. (6), and the second step follows from Tonelli's theorem [Fol99].

For every $u \in \mathbb{R}$, set

$$r := \theta\sqrt{1+u^2}, \qquad \phi := \arctan(u).$$

Then $re^{i\phi} = \theta(1+iu)$ and $r\cos(\phi) = \theta \in \mathcal{J}$. Hence Eq. (3) gives

$$\mathbb{E}_\mu[Q_u] \le D \exp(\Psi(\theta)).$$

Combining this estimate with Eq. (7) and using $\int_{\mathbb{R}} \beta_0(u)\mathrm{d}u = 1$ proves Eq. (4). □

4 Sharp Hoeffding bound

4.1 Markov matrix Hoeffding bound

Theorem 4.1 (Markov matrix Hoeffding bound, formal version of Theorem 1.1)**.** *Under Definition 2.1, suppose that*

$$a_j I \preceq F_j(x) \preceq b_j I$$

for every j and x. Set

$$\mathcal{R} := \sum_{j=1}^n (b_j - a_j)^2.$$

Then, for every $t \ge 0$,

$$\Pr_\mu[\lambda_{\max}(S_F) \ge t] \le d \exp(-\frac{2t^2}{\alpha(\lambda)\mathcal{R}}).$$

Equivalently,

$$\Pr_\mu[\lambda_{\max}(S_F) \ge t] \le d \exp(-\frac{2(1-\lambda)t^2}{(1+\lambda)\sum_{j=1}^n (b_j - a_j)^2}).$$

Proof. We first assume that the matrices are real symmetric. The interval bounds imply $\|F_j(x)\| \le M_j$ for $M_j := \max\{|a_j|, |b_j|\}$. Apply Lemma 3.1 with

$$\mathcal{J} := (0, \infty), \qquad D := d, \qquad \Psi(\eta) := \frac{\alpha(\lambda)\eta^2}{8}\mathcal{R}.$$

The required product estimate follows from Lemma 4.3, because

$$d \exp(\frac{\alpha(\lambda) r^2 \cos^2(\phi)}{8}\mathcal{R}) = d \exp(\Psi(r\cos(\phi))).$$

Therefore, for every $\theta > 0$,

$$\mathbb{E}_\mu[\mathrm{tr}[\exp(\theta S_F)]] \le d \exp(\frac{\alpha(\lambda)\theta^2}{8}\mathcal{R}). \tag{8}$$

The Laplace transform argument yields

$$\Pr_{\mu}[\lambda_{\max}(S_F) \geq t] \leq \inf_{\theta>0} d\exp(-\theta t + \frac{\alpha(\lambda)\theta^2}{8}\mathcal{R}).$$

The minimizing value is $\theta_\star = 4t/(\alpha(\lambda)\mathcal{R})$, which gives the claimed bound.

For complex Hermitian matrices, apply Eq. (8) to the $2d$-dimensional realifications $\mathcal{C}(F_j)$. The right-hand side initially has prefactor $2d$, while Eq. (2) shows that the trace on the left is twice the complex trace. Dividing by two recovers Eq. (8) with prefactor d and completes the proof. □

4.2 Expander-Chernoff normalization

Corollary 4.2 (Expander-Chernoff normalization, formal version of Corollary 1.3)**.** *Under Definition 2.1, suppose that $-I \preceq F_j(x) \preceq I$ for every j and x. Then, for every $\epsilon \geq 0$,*

$$\Pr_{\mu}[\lambda_{\max}(\frac{1}{n}S_F) \geq \epsilon] \leq d\exp(-\frac{\epsilon^2(1-\lambda)n}{2(1+\lambda)}).$$

Consequently, a bound of the form $d\exp(-c\epsilon^2(1-\lambda)n)$ holds with $c(\lambda) = 1/(2(1+\lambda))$, and uniformly over $\lambda \in [0,1)$ with $c = 1/4$.

Proof. Here $\mathcal{R} = 4n$. Applying Theorem 4.1 with $t = n\epsilon$ gives

$$\frac{2t^2}{\alpha(\lambda)\mathcal{R}} = \frac{\epsilon^2 n}{2\alpha(\lambda)} = \frac{\epsilon^2(1-\lambda)n}{2(1+\lambda)}.$$

□

4.3 Phase-sensitive product estimate

Lemma 4.3 (Phase sensitive hoeffding product estimate)**.** *Under the assumptions of Theorem 4.1, assume in addition that $F_j(x)$ is real symmetric for every j and x. Then, for every $r > 0$ and $\phi \in [-\pi/2, \pi/2]$,*

$$\mathbb{E}_{\mu}[\|\prod_{j=1}^{n}\exp(\frac{re^{i\phi}}{2}F_j(s_j))\|_2^2] \leq d\exp(\frac{\alpha(\lambda)r^2\cos^2(\phi)}{8}\mathcal{R}).$$

Proof. This is exactly the product estimate stated in Eq. (6.12) of [NSW24]. □

4.4 Scalar obstruction

Proposition 4.4 (Scalar obstruction)**.** *No inequality valid under the hypotheses of Corollary 4.2 can replace $1/(2(1+\lambda))$ by a larger coefficient depending only on λ. Consequently, no universal coefficient larger than $1/4$ is possible.*

Proof. Consider the stationary two-state chain $X_j \in \{-1, 1\}$ with transition matrix

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

and take the scalar observable $F(X_j) = X_j$. Set $p := (1+\lambda)/2$ and $q := (1-\lambda)/2$. The tilted transition matrix is similar to

$$M_\theta = \begin{pmatrix} pe^{\theta} & q \\ q & pe^{-\theta} \end{pmatrix},$$

whose Perron eigenvalue is

$$\rho(\theta) = p\cosh(\theta) + \sqrt{q^2 + p^2\sinh^2(\theta)}.$$

Therefore the limiting cumulant-generating function satisfies

$$\Lambda(\theta) = \log\rho(\theta) = \frac{1+\lambda}{2(1-\lambda)}\theta^2 + O(\theta^4).$$

The corresponding Legendre-transform rate function satisfies

$$I(\epsilon) = \frac{\epsilon^2(1-\lambda)}{2(1+\lambda)} + O(\epsilon^4).$$

Thus a uniform quadratic upper-tail inequality with a strictly larger coefficient would contradict the exact scalar large-deviation rate for sufficiently small ϵ. Finally,

$$\inf_{0\le\lambda<1} \frac{1}{2(1+\lambda)} = \frac{1}{4}.$$

□

5 Sharp Bernstein bound

5.1 Markov matrix Bernstein bound

Theorem 5.1 (Markov matrix Bernstein bound, formal version of Theorem 1.2)**.** *Under Definition 2.1, suppose that*

$$\|\mathbb{E}_\mu[F_j^2]\| \le \mathcal{V}_j, \qquad \|F_j(x)\| \le \mathcal{M}$$

for every j and x, where $\mathcal{M} > 0$. Set

$$\sigma^2 := \sum_{j=1}^n \mathcal{V}_j$$

and define

$$\mathcal{I}_\lambda := \begin{cases} (0,\infty), & \lambda = 0, \\ (0, \log(1/\lambda)/\mathcal{M}), & 0 < \lambda < 1. \end{cases}$$

For $\theta \in \mathcal{I}_\lambda$, let

$$G_\lambda(\theta) := \frac{\sigma^2}{\mathcal{M}^2}(e^{\mathcal{M}\theta} - \mathcal{M}\theta - 1 + \frac{\lambda(e^{\mathcal{M}\theta}-1)^2}{1-\lambda e^{\mathcal{M}\theta}}).$$

Then, for every $t \ge 0$,

$$\Pr_\mu[\lambda_{\max}(S_F) \ge t] \le d \inf_{\theta\in\mathcal{I}_\lambda} \exp(-\theta t + G_\lambda(\theta)).$$

In particular, if $0 < \lambda < 1$, then

$$\Pr_\mu[\lambda_{\max}(S_F) \ge t] \le d\exp(-\frac{t^2/2}{\alpha(\lambda)\sigma^2 + 2\mathcal{M}t/(1-\lambda)}).$$

If $\lambda = 0$, then

$$\Pr_\mu[\lambda_{\max}(S_F) \ge t] \le d\exp(-\frac{t^2/2}{\sigma^2 + \mathcal{M}t/3}).$$

Proof. We first assume that the matrices are real symmetric. Apply Lemma 3.1 with

$$\mathcal{J} := \mathcal{I}_\lambda, \qquad D := d, \qquad \Psi(\eta) := G_\lambda(\eta).$$

The norm bound gives $\|F_j(x)\| \le \mathcal{M}$, and the required product estimate is exactly Lemma 5.3. Therefore, for every $\theta \in \mathcal{I}_\lambda$,

$$\mathbb{E}_\mu[\mathrm{tr}[\exp(\theta S_F)]] \le d \exp(G_\lambda(\theta)). \tag{9}$$

The Laplace transform argument now yields

$$\Pr_\mu[\lambda_{\max}(S_F) \ge t] \le d \inf_{\theta \in \mathcal{I}_\lambda} \exp(-\theta t + G_\lambda(\theta)),$$

which proves the exact bound.

It remains to derive the displayed Bernstein simplifications. Normalize by setting

$$\widetilde{F}_j := \frac{F_j}{\mathcal{M}}, \qquad \widetilde{t} := \frac{t}{\mathcal{M}}, \qquad \widetilde{\sigma}^2 := \frac{\sigma^2}{\mathcal{M}^2}.$$

For $0 < \lambda < 1$, define

$$g(\theta) := \widetilde{\sigma}^2(e^\theta - \theta - 1) + \frac{\widetilde{\sigma}^2 \lambda \theta^2}{1 - \lambda - 2\theta}$$

for $0 \le \theta < (1-\lambda)/2$. Eqs. (6.27)–(6.28) of [NSW24], used with scaling parameter $L = 1$, give $\widetilde{G}_\lambda(\theta) \le g(\theta)$ on this interval. Lemma D.2 of [NSW24], also with $L = 1$, implies

$$g^*(\widetilde{t}) \ge \frac{\widetilde{t}^2/2}{\alpha(\lambda)\widetilde{\sigma}^2 + 2\widetilde{t}/(1-\lambda)}.$$

Since $\widetilde{G}_\lambda \le g$, their convex conjugates satisfy $\widetilde{G}^*_\lambda \ge g^*$. Restoring $\mathcal{M}$ gives

$$G^*_\lambda(t) \ge \frac{t^2/2}{\alpha(\lambda)\sigma^2 + 2\mathcal{M}t/(1-\lambda)}.$$

For $\lambda = 0$, the second part of the same lemma gives

$$G^*_0(t) \ge \frac{t^2/2}{\sigma^2 + \mathcal{M}t/3}.$$

Substitution into the exact bound proves both simplified inequalities.

Finally, for complex Hermitian matrices apply Eq. (9) to the realifications $\mathcal{C}(F_j)$. Realification preserves the variance and norm bounds because

$$\mathcal{C}(F_j)^2 = \mathcal{C}(F_j^2), \qquad \|\mathbb{E}_\mu[\mathcal{C}(F_j)^2]\| = \|\mathbb{E}_\mu[F_j^2]\|.$$

The real trace and dimension are both doubled. Dividing Eq. (9) by two therefore recovers the same estimate with prefactor d, completing the proof. □

Remark 5.2 (Comparison with Neeman–Shi–Ward)**.** Theorem 2.6 of [NSW24] has prefactor $d^{2-\pi/4}$ and numerator constant $\pi^2/32$ in the Bernstein exponent. Using the same prefactor–exponent tradeoff as [GLSS18, Remark 4.2], this gives numerator constant $\pi^2/(32(2-\pi/4))$ at prefactor d. Hence, at the common prefactor d, Theorem 5.1 improves the small-deviation exponent by the factor $16(2-\pi/4)/\pi^2$. The exact infimum in Theorem 5.1 is stronger than the displayed simplification. This theorem improves the conversion from the NSW product estimate to a trace moment-generating function; it does not improve the NSW product estimate or certify a sharper variance proxy.

5.2 Phase-sensitive product estimate

Lemma 5.3 (Phase-sensitive Bernstein product estimate)**.** *Under the assumptions of Theorem 5.1, assume in addition that $F_j(x)$ is real symmetric for every j and x. Let $r > 0$ and $\phi \in [-\pi/2, \pi/2]$. If $\lambda = 0$, or if*

$$0 \le r\cos(\phi) < \frac{\log(1/\lambda)}{\mathcal{M}},$$

then

$$\mathbb{E}_\mu[\| \prod_{j=1}^{n} \exp(\frac{re^{i\phi}}{2} F_j(s_j))\|_2^2] \le d\exp(G_\lambda(r\cos(\phi))).$$

Proof. Eq. (6.19) of Neeman, Shi, and Ward [NSW24] reduces the product moment on the left-hand side to norms of multiplication operators generated by

$$T_j(x) = \frac{\cos(\phi)}{2}(F_j(x) \otimes I + I \otimes F_j(x)).$$

In Section 6.3 of [NSW24], immediately before [NSW24, Lemma 6.7], set $\phi = 0$ only to simplify the notation; they state that the scalar factor $\cos(\phi)$ can be carried through the argument. Consequently, [NSW24, Lemma 6.7 and Claims 6.8–6.9] apply here after replacing their parameter θ by

$$\eta = r\cos(\phi).$$

Applying [NSW24, Eqs. (6.20)–(6.23)] with this substitution gives

$$\mathbb{E}_\mu[\| \prod_{j=1}^{n} \exp(\frac{re^{i\phi}}{2} F_j(s_j))\|_2^2] \le d\exp(\frac{\sigma^2}{\mathcal{M}^2}(e^{\mathcal{M}\eta} - \mathcal{M}\eta - 1 + \frac{\lambda(e^{\mathcal{M}\eta}-1)^2}{1-\lambda e^{\mathcal{M}\eta}})).$$

The exponent on the right is $G_\lambda(\eta)$ by definition. If $\lambda > 0$, the hypothesis $\eta < \log(1/\lambda)/\mathcal{M}$ is equivalent to $\lambda e^{\mathcal{M}\eta} < 1$, which is the condition used in [NSW24, Claim 6.8]. If $\lambda = 0$, the denominator equals one and no upper restriction on η is needed. This proves the lemma. □

5.3 Sharpness of the variance term

Proposition 5.4 (Scalar obstruction for the Bernstein variance term)**.** *Fix $\lambda \in [0, 1)$. Under the assumptions of Theorem 5.1, the coefficient $\alpha(\lambda)$ multiplying σ^2 in the simplified Bernstein denominator is optimal in the small-deviation regime. More precisely, there do not exist constants $0 < c_\lambda < \alpha(\lambda)$ and $0 \le C_\lambda < \infty$ such that*

$$\Pr_\mu[\lambda_{\max}(S_F) \ge t] \le d\exp(-\frac{t^2/2}{c_\lambda\sigma^2 + C_\lambda\mathcal{M}t})$$

holds for every n, every admissible stationary chain and matrix-valued observable, and every $t > 0$.

Proof. Consider the stationary two-state chain and scalar observable from Proposition 4.4. In this example, $d = 1$, $\mathcal{M} = 1$, and $\sigma^2 = n$. Moreover, for $t = n\epsilon$, the exact scalar large-deviation rate satisfies $I(\epsilon) = \frac{\epsilon^2}{2\alpha(\lambda)} + O(\epsilon^4)$. If the displayed inequality held with $c_\lambda < \alpha(\lambda)$, then

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr_\mu[S_F \ge n\epsilon] \le -\frac{\epsilon^2}{2(c_\lambda + C_\lambda\epsilon)}.$$

For all sufficiently small fixed $\epsilon > 0$, the decay rate on the right is strictly larger than $I(\epsilon)$, contradicting the exact scalar large-deviation rate. Hence the coefficient $\alpha(\lambda)$ cannot be decreased. □

References

[AW02] Rudolf Ahlswede and Andreas Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

[Ber24] Sergei N. Bernstein. On a modification of chebyshev's inequality and of the error formula of laplace. *Uchenye Zapiski Nauchno-Issledovatel'skikh Kafedr Ukrainy, Otdelenie Matematiki*, 1(1):38–49, 1924. In Russian.

[Che52] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952.

[CLLM12] Kai-Min Chung, Henry Lam, Zhenming Liu, and Michael Mitzenmacher. Chernoff–hoeffding bounds for markov chains: Generalized and simplified. In *29th International Symposium on Theoretical Aspects of Computer Science*, volume 14 of *Leibniz International Proceedings in Informatics*, pages 124–135, 2012.

[FJS21] Jianqing Fan, Bai Jiang, and Qiang Sun. Hoeffding's inequality for general markov chains and its applications to statistical learning. *Journal of Machine Learning Research*, 22(139):1–35, 2021.

[Fol99] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, 2 edition, 1999.

[Gil98] David Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

[GLSS18] Ankit Garg, Yin Tat Lee, Zhao Song, and Nikhil Srivastava. A matrix expander chernoff bound. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1102–1114. ACM, 2018.

[Gol65] Sidney Golden. Lower bounds for the helmholtz function. *Physical Review*, 137(4B):B1127–B1128, 1965.

[Hea08] Alexander Healy. Randomness-efficient sampling within NC^1. *Computational Complexity*, 17(1):3–37, 2008.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[JSF18] Bai Jiang, Qiang Sun, and Jianqing Fan. Bernstein's inequalities for general markov chains, 2018. arXiv:1805.10721.

[Lez98] Pascal Lezaud. Chernoff-type bound for finite markov chains. *The Annals of Applied Probability*, 8(3):849–867, 1998.

[LP04] Carlos A. León and François Perron. Optimal hoeffding bounds for discrete reversible markov chains. *The Annals of Applied Probability*, 14(2):958–970, 2004.

[MA19] Vrettos Moulos and Venkat Anantharam. Optimal chernoff and hoeffding bounds for finite state markov chains, 2019. arXiv:1907.04467.

[NSW24] Joe Neeman, Bobby Shi, and Rachel Ward. Concentration inequalities for sums of markov-dependent random matrices. *Information and Inference: A Journal of the IMA*, 13(4):iaae032, 2024.

[Pau15] Daniel Paulin. Concentration inequalities for markov chains by marton couplings and spectral methods. *Electronic Journal of Probability*, 20:1–32, 2015.

[PXZ25] Yang Peng, Yuchen Xin, and Zhihua Zhang. Matrix moment and concentration inequalities for martingales and ergodic markov chains with applications in statistical learning, 2025. arXiv:2508.04327.

[QWL+20] Jiezhong Qiu, Chi Wang, Ben Liao, Richard Peng, and Jie Tang. A matrix chernoff bound for markov chains and its application to co-occurrence matrices. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

[Rao19] Shravas Rao. A hoeffding inequality for markov chains. *Electronic Communications in Probability*, 24(14):1–11, 2019.

[Rao26] Shravas Rao. Towards a banach space chernoff bound for markov chains via chaining arguments. *Proceedings of the American Mathematical Society*, 154:4037–4048, 2026.

[SBT17] David Sutter, Mario Berta, and Marco Tomamichel. Multivariate trace inequalities. *Communications in Mathematical Physics*, 352(1):37–58, 2017.

[Sen25] Pranab Sen. Matrix chernoff concentration bounds for multipartite soft covering and expander walks, 2025. arXiv:2504.04067.

[Son19] Zhao Song. *Matrix Theory: Optimization, Concentration and Algorithms*. PhD thesis, The University of Texas at Austin, August 2019.

[Tho65] Colin J. Thompson. Inequality with applications in statistical mechanics. *Journal of Mathematical Physics*, 6(11):1812–1813, 1965.

[Tro12] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

[WS26] Alexander Van Werde and Jaron Sanders. Sharp concentration for sums of matrices with markovian dependence through universality. *Probability Theory and Related Fields*, 2026.

[WX05] Avi Wigderson and David Xiao. A randomness-efficient sampler for matrix-valued functions and applications. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 397–406. IEEE, 2005.

[WX08] Avi Wigderson and David Xiao. Derandomizing the ahlswede–winter matrix-valued chernoff bound using pessimistic estimators, and applications. *Theory of Computing*, 4(3):53–76, 2008.

[Zan26] Luca Zanetti. Matrix concentration inequalities for time-inhomogeneous markov chains, 2026. arXiv:2605.24445.

Acknowledgments

The author used Gemini Pro 3.1, Codex 5.6 Sol and Claude Code Fable 5 to assist with language editing and grammar checking.