Electronic Colloquium on Computational Complexity, Report No. 216 (2026)

Matrix Hoeffding and Bernstein Bounds with Sharp Constants for Markov Chains

Zhao Song*

September 25, 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

*magic.linuxkde@gmail.com. The first version of this paper was completed on August 18, 2026; this is version 2. We provide a detailed comparison between the two versions in the Acknowledgments section.

ISSN 1433-8092

Tang then established a matrix Chernoff bound for regular finite Markov chains and applied it to co-occurrence matrices [QWL+20].

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. This question was raised in [Son19]. In this paper, we provide a positive answer to this question. Throughout the paper, in tail-bound exponents we use the conventions $0/0 = 0$ and $a/0 = +\infty$ for $a > 0$.

Theorem 1.1 (Markov matrix Hoeffding bound, informal version of Theorem 4.1)**.** *Let $s_1, \ldots, s_n$ 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 $\sigma^2 := \sum_{j=1}^{n} \mathcal{V}_j$, $\alpha(\lambda) := (1+\lambda)/(1-\lambda)$, and $\beta(\lambda) := (1+4\lambda+\lambda^2)/(3(1-\lambda^2))$. Then, for every $0 \le \lambda < 1$ and every $t \ge 0$,*

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

When $\lambda = 0$, we have $\alpha(0) = 1$ and $\beta(0) = 1/3$, and the trajectory consists of independent samples from μ. In this case, the exponent has numerator $-t^2/2$ and denominator $\sigma^2 + \mathcal{M}t/3$, with prefactor d. Thus our bound recovers exactly the one-sided Hermitian form of Recht's non-commutative Bernstein inequality [Rec11, Theorem 4 and Appendix A], obtained by omitting the rectangular dilation in its proof. The coefficient $\alpha(\lambda)$ multiplying the variance proxy is optimal in the small-deviation regime.

We also obtain a complementary Bernstein bound whose leading term uses the norm of the summed second moments. Its dependence correction vanishes in the independent case.

Theorem 1.3 (Markov matrix Bernstein bound with summed variance, simplified form of Theorem 6.1)**.** *Under the same Markov-chain assumptions, suppose that the centered Hermitian matrices satisfy* $\|F_j(x)\| \le \mathcal{M}$*, where* $\mathcal{M} > 0$*. Define* $v := \|\sum_{j=1}^n \mathbb{E}_\mu[F_j^2]\|$ *and* $w := \sum_{j=1}^n \|\mathbb{E}_\mu[F_j^2]\|$*, and let* $\alpha(\lambda)$ *and* $\beta(\lambda)$ *be as in Theorem 1.2. For* $p > 1$*, set* $q := p/(p-1)$ *and* $V_p(\lambda) := pv + \alpha(\lambda)\lambda^2 qw$*. Define* $C_p(0) := p/3$ *and* $C_p(\lambda) := p + \beta(\lambda)q$ *for* $0 < \lambda < 1$*. Then, for every* $0 \le \lambda < 1$*, every* $p > 1$*, and every* $t \ge 0$*,*

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

When $\lambda = 0$, we have $V_p(0) = pv$ and $C_p(0) = p/3$. Letting $p \downarrow 1$, the exponent has numerator $-t^2/2$ and denominator $v + \mathcal{M}t/3$, with prefactor d. Thus the theorem recovers Tropp's independent matrix Bernstein inequality [Tro12, Theorem 1.4], with the same variance proxy and constants.

Corollary 1.4 (Expander-Chernoff normalization, informal version of Corollary 4.2)**.** *Under the same Markov-chain assumptions, suppose that* $\mathbb{E}_\mu[F_j] = 0$ *and* $-I \preceq F_j(x) \preceq I$ *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 1.5 (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 1.4 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.

Roadmap. The remainder of the paper is organized as follows. Section 2 defines notation and introduces the multivariate trace inequality. Section 3 proves the common conversion from phase-sensitive product estimates to trace moment-generating-function bounds. Section 4 shows the sharp Markov matrix Hoeffding inequality, its expander-Chernoff normalization, the needed product estimate, and the scalar sharpness obstruction. Section 5 establishes the exact and simplified Bernstein bounds, verifies the phase-sensitive product estimate, and proves optimality of the variance coefficient. Section 6 derives a complementary variational Bernstein bound and recovers Tropp's independent matrix Bernstein inequality.

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 $s_1, \ldots, s_n$ 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 $H_1, \ldots, H_n$ be Hermitian matrices. For the fixed probability density $\beta_0(u) := \pi/(2(\cosh(\pi u)+1))$, independent of $H_1, \ldots, H_n$, we have*

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

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

The product estimates used later control a Schatten-norm expression rather than the trace moment-generating function needed by the Laplace method. The following lemma supplies the conversion by combining the multivariate trace inequality with Jensen's inequality and the phase-sensitive hypothesis. It will serve as the common interface for both the Hoeffding and Bernstein arguments.

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 $M_1, \dots, M_n$ 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,

$$\begin{aligned} \mathrm{tr}[\exp(\theta S_F)] &\le \exp(\int_{\mathbb{R}} \beta_0(u) \log Q_u \mathrm{d}u) \\ &\le \int_{\mathbb{R}} \beta_0(u) \exp(\log Q_u) \mathrm{d}u \\ &= \int_{\mathbb{R}} \beta_0(u) Q_u \mathrm{d}u. \end{aligned} \tag{6}$$

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

Section 4.1 proves the sharp Markov matrix Hoeffding bound by converting the phase-sensitive product estimate into a trace moment-generating-function bound. Section 4.2 specializes this result to normalized observables and records the resulting uniform expander-Chernoff constant. Section 4.3 states the phase-sensitive product estimate that supplies the probabilistic input to the proof. Section 4.4 uses a stationary two-state chain to show that the dependence on λ is optimal.

4.1 Markov matrix Hoeffding bound

We first state the main Hoeffding inequality in the notation of Definition 2.1. The proof separates the real symmetric case, where the phase-sensitive product estimate applies directly, from the complex Hermitian case, which follows by realification. The resulting exponent matches the independent matrix Hoeffding constant when $\lambda = 0$.

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. The case $t = 0$ is trivial. If $\mathcal{R} = 0$, centering implies $S_F = 0$ almost surely, so the claim holds under the stated conventions; hence assume $t > 0$ and $\mathcal{R} > 0$. 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

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

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

The preceding theorem becomes especially transparent for observables normalized to lie between $-I$ and I. Substituting interval width two and scaling the threshold by n yields the following expander-walk form, together with a uniform constant over all $\lambda \in [0, 1)$.

Corollary 4.2 (Expander-Chernoff normalization, formal version of Corollary 1.4)**.** *Under Definition 2.1, suppose that $-I \preceq F_j(x) \preceq I$ for every j and x. Then, for every $\epsilon \ge 0$,*

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

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

The only model-specific input in the Hoeffding proof is a product-moment estimate that retains the phase factor arising from the multivariate trace inequality. We isolate that estimate here so that its role in the trace-MGF conversion is explicit.

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 $\phi \in [-\pi/2, \pi/2]$,*

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

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

4.4 Scalar obstruction

The dependence on λ in Corollary 4.2 is not an artifact of the proof. A two-state reversible chain already attains the corresponding quadratic large-deviation rate, which yields the following 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,

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

$\square$

5 Sharp Bernstein bound

Section 5.1 proves the exact variational Bernstein bound and its simplified variance-sensitive forms. Section 5.2 isolates the phase-sensitive product estimate needed for the trace moment-generating-function conversion. Section 5.3 uses the two-state example to prove that the variance coefficient $\alpha(\lambda)$ is optimal in the small-deviation regime.

5.1 Markov matrix Bernstein bound

We next turn to variance-sensitive concentration. In addition to the exact variational bound, the theorem records a Bernstein-form simplification that separates the variance and range contributions. As in the Hoeffding case, the argument first treats real symmetric matrices and then passes to complex Hermitian matrices by realification.

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

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

Set $G_\lambda(0) := 0$ and, for $\theta \in \mathcal{I}_\lambda$, let

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

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, define $\beta(\lambda) := (1+4\lambda+\lambda^2)/(3(1-\lambda^2))$. For every $0 \le \lambda < 1$ and every $t \ge 0$,

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

Since $\alpha(0) = 1$ and $\beta(0) = 1/3$, setting $\lambda = 0$ gives the denominator $\sigma^2 + \mathcal{M}t/3$.

Proof. The case $t = 0$ is trivial under the stated conventions. If $\sigma^2 = 0$, then $S_F = 0$ almost surely and all the bounds hold; hence assume $t > 0$ and $\sigma^2 > 0$. 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 Bernstein simplification. We prove a scalar estimate that keeps the dependence on λ explicit. Write $r := \lambda/(1-\lambda)$, $a := \alpha(\lambda) = 1 + 2r$, $\delta := 1/(6a)$, and $b := \beta(\lambda) = a/2 - \delta$. Since $a \ge 1$, we have $\delta \le 1/6$ and $b - r = 1/2 - \delta \ge 1/3$.

For $u \geq 0$ with $\lambda e^u < 1$, define

$$h_\lambda(u) := e^u - u - 1 + \frac{\lambda(e^u-1)^2}{1-\lambda e^u}.$$

We claim that every $0 \leq u < 1/b$ satisfies $\lambda e^u < 1$ and

$$h_\lambda(u) \leq \frac{au^2}{2(1-bu)}. \tag{10}$$

Fix $0 \leq u < 1/b$ and set

$$Q(u) := 1-(b-r)u + r\delta u^2, \qquad f(u) := u + \delta u^2 - Q(u)(e^u - 1).$$

Direct differentiation gives

$$f(0) = f'(0) = f''(0) = 0, \qquad f'''(u) = \delta e^u u(2-ru) \geq 0.$$

The inequality follows from $0 \leq ru \leq r/b < 1$, since $b > r$. Integrating the nonnegative third derivative three times, using the three initial conditions, gives $f(u) \geq 0$.

Moreover,

$$Q(u) = 1 - bu + ru(1+\delta u) > 0,$$

since $1-bu > 0$ and $r, u, \delta \geq 0$. Writing $z := e^u - 1$, the inequality $f(u) \geq 0$ therefore implies

$$0 \leq z \leq \frac{u(1+\delta u)}{Q(u)}.$$

Consequently,

$$rz \leq \frac{ru(1+\delta u)}{Q(u)} < 1.$$

The strict inequality follows from $Q(u) - ru(1+\delta u) = 1 - bu > 0$. Since $1-\lambda e^u = (1-\lambda)(1-rz) > 0$, this also verifies the claimed domain condition.

We can now show

$$\begin{aligned} h_\lambda(u) &= \frac{z}{1-rz} - u \\ &\leq \frac{u(1+\delta u)}{1-bu} - u \\ &= \frac{au^2}{2(1-bu)}. \end{aligned}$$

The first step uses $r = \lambda/(1-\lambda)$ and $z = e^u - 1$. The second step uses the preceding bound on z and monotonicity of $z/(1-rz)$ on $0 \leq rz < 1$. The final step uses $b + \delta = a/2$. This proves Eq. (10).

Apply Eq. (10) with $u = \mathcal{M}\theta$. For every $0 < \theta < 1/(\beta(\lambda)\mathcal{M})$, the domain verification above gives $\theta \in \mathcal{I}_\lambda$, and

$$G_\lambda(\theta) = \frac{\sigma^2}{\mathcal{M}^2} h_\lambda(\mathcal{M}\theta) \leq \frac{\alpha(\lambda)\sigma^2\theta^2}{2(1-\beta(\lambda)\mathcal{M}\theta)}.$$

Combining this estimate with Eq. (9) and the Laplace transform bound gives

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

Choose

$$\theta := \frac{t}{\alpha(\lambda)\sigma^2 + \beta(\lambda)\mathcal{M}t}.$$

This lies in the stated interval because $t > 0$ and $\sigma^2 > 0$. Substitution gives

$$-\theta t + \frac{\alpha(\lambda)\sigma^2\theta^2}{2(1 - \beta(\lambda)\mathcal{M}\theta)} = -\frac{t^2/2}{\alpha(\lambda)\sigma^2 + \beta(\lambda)\mathcal{M}t},$$

which proves the claimed Bernstein bound for every $0 \le \lambda < 1$.

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

To invoke Lemma 3.1 in the Bernstein setting, we need a product estimate whose exponent depends only on the real part of the complex tilting parameter. The following lemma extracts precisely this phase-sensitive form from the Neeman–Shi–Ward argument, with a domain restriction that keeps the denominator positive.

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. If $\cos(\phi) = 0$, all factors are unitary, so both sides equal d. Otherwise, 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

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

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. $\square$

5.3 Sharpness of the variance term

The variance coefficient in the simplified Bernstein denominator is forced by the same low-dimensional example as the Hoeffding sharpness result. Examining small deviations of the two-state chain shows that no uniformly smaller multiplier of σ^2 can be compensated for by a finite linear 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 $0 \leq C_\lambda < \infty$ such that*

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

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

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

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. $\square$

6 Variational Bernstein bound

The variance proxy in Section 5 takes an operator norm before summing over time. We next give a complementary bound that retains the norm of the summed second moments in its leading term and separates the remaining dependence correction. Section 6.1 proves the variational bound. Section 6.2 establishes the conditional trace estimate and the contraction estimate used in its proof.

6.1 Variational Markov matrix Bernstein bound

The following theorem combines a conditional trace argument with the Bernstein trace-MGF estimate from Section 5. Its first term uses the variance proxy from Tropp's independent matrix Bernstein inequality, while its second term vanishes when the samples are independent. For positive dependence, this estimate complements rather than uniformly improves Theorem 5.1.

Theorem 6.1 (Variational Markov matrix Bernstein bound)**.** *Under Definition 2.1, suppose that* $\|F_j(x)\| \le \mathcal{M}$ *for every* j *and* x*, where* $\mathcal{M} > 0$*. Define*

$$v := \|\sum_{j=1}^{n} \mathbb{E}_\mu[F_j^2]\|, \qquad w := \sum_{j=1}^{n} \|\mathbb{E}_\mu[F_j^2]\|.$$

Let

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

and, for $u \in \mathcal{J}_\lambda$*, define*

$$h_\lambda(u) := e^u - u - 1 + \frac{\lambda(e^u-1)^2}{1-\lambda e^u}.$$

For $p > 1$ *and* $\theta > 0$*, set*

$$q := \frac{p}{p-1}, \qquad \psi_p(\theta) := \frac{e^{p\mathcal{M}\theta}-1}{p-1}.$$

Whenever $\psi_p(\theta) \in \mathcal{J}_\lambda$*, define*

$$K_{\lambda,p}(\theta) := \frac{v}{p\mathcal{M}^2} h_0(p\mathcal{M}\theta) + \frac{\lambda^2 w}{q\mathcal{M}^2} h_\lambda(\psi_p(\theta)). \tag{11}$$

Then

$$\mathbb{E}_\mu[\mathrm{tr}[\exp(\theta S_F)]] \le d\exp(K_{\lambda,p}(\theta)). \tag{12}$$

Consequently, for every $t \ge 0$*,*

$$\Pr_\mu[\lambda_{\max}(S_F) \ge t] \le d \inf_{\substack{p>1,\ \theta>0 \\ \psi_p(\theta)\in\mathcal{J}_\lambda}} \exp(-\theta t + K_{\lambda,p}(\theta)). \tag{13}$$

When $\lambda = 0$*, the second term in Eq. (11) vanishes, and letting* $p \downarrow 1$ *yields*

$$\Pr_\mu[\lambda_{\max}(S_F) \ge t] \le d\exp(-\frac{t^2/2}{v + \mathcal{M}t/3}). \tag{14}$$

Thus the independent case recovers [Tro12, Theorem 1.4], with the same variance proxy and constants.

Proof. Fix an admissible pair (p,θ) and put

$$c := \frac{h_0(p\mathcal{M}\theta)}{p\mathcal{M}^2}.$$

Let $\mathcal{F}_0$ be the trivial sigma-algebra and let $\mathcal{F}_j := \sigma(s_1, \ldots, s_j)$. Define

$$B := \sum_{j=2}^{n} PF_j(s_{j-1}),$$
$$A := \mathbb{E}_\mu[F_1^2] + \sum_{j=2}^{n} P(F_j^2)(s_{j-1}),$$
$$H := A - \sum_{j=1}^{n} \mathbb{E}_\mu[F_j^2].$$

The Markov property identifies B and A as the sums of the conditional means and conditional second moments of $F_j(s_j)$. Applying Lemma 6.2 with $z = p\theta$ gives

$$\mathbb{E}_\mu[\mathrm{tr}[\exp(U)]] \le d, \qquad U := p\theta S_F - p\theta B - pcA. \tag{15}$$

For Hermitian matrices U and W, the Golden–Thompson inequality followed by Schatten Hölder gives

$$\mathrm{tr}[\exp(U/p + W/q)] \le (\mathrm{tr}[\exp(U)])^{1/p}(\mathrm{tr}[\exp(W)])^{1/q}.$$

Apply this inequality with $W := q\theta B + qcA$. Since $1/p + 1/q = 1$ and $U/p + W/q = \theta S_F$, we can show

$$\begin{aligned}\mathbb{E}_\mu[\mathrm{tr}[\exp(\theta S_F)]] &\le (\mathbb{E}_\mu[\mathrm{tr}[\exp(U)]])^{1/p}(\mathbb{E}_\mu[\mathrm{tr}[\exp(W)]])^{1/q} \\ &\le d^{1/p} e^{cv}(\mathbb{E}_\mu[\mathrm{tr}[\exp(q\theta B + qcH)]])^{1/q}.\end{aligned} \tag{16}$$

The first step follows from the displayed trace inequality and scalar Hölder. The second step follows from Eq. (15), $A = \sum_{j=1}^{n} \mathbb{E}_\mu[F_j^2] + H$, and $\sum_{j=1}^{n} \mathbb{E}_\mu[F_j^2] \preceq vI$, using monotonicity of the trace exponential in the Loewner order.

For $j = 2, \ldots, n$, define the matrix-valued function

$$Z_j := P(\theta F_j + c(F_j^2 - \mathbb{E}_\mu[F_j^2])).$$

Then

$$\theta B + cH = \sum_{j=2}^{n} Z_j(s_{j-1}).$$

By Lemma 6.3, these functions are centered and satisfy

$$\|Z_j(x)\| \le L, \qquad \|\mathbb{E}_\mu[Z_j^2]\| \le \lambda^2(\theta + c\mathcal{M})^2 \|\mathbb{E}_\mu[F_j^2]\|,$$

where

$$L := \mathcal{M}(\theta + c\mathcal{M}) = \frac{e^{p\mathcal{M}\theta} - 1}{p}. \tag{17}$$

In particular,

$$\sum_{j=2}^{n} \|\mathbb{E}_\mu[Z_j^2]\| \le \frac{\lambda^2 L^2}{\mathcal{M}^2} w.$$

Apply Eq. (9) to the functions $Z_2, \ldots, Z_n$ along the stationary trajectory $s_1, \ldots, s_{n-1}$, with range bound L and tilting parameter q. Its domain condition is $qL \in \mathcal{J}_\lambda$, which holds because $qL = \psi_p(\theta)$. We obtain

$$\mathbb{E}_\mu[\mathrm{tr}[\exp(q\theta B + qcH)]] \le d \exp(\frac{\lambda^2 w}{\mathcal{M}^2} h_\lambda(\psi_p(\theta))). \tag{18}$$

If $n = 1$, the correction sum is empty, and this inequality follows directly from $\mathrm{tr}[I] = d$ and $h_\lambda \geq 0$.

Substituting Eq. (18) into Eq. (16) gives

$$\mathbb{E}_\mu[\mathrm{tr}[\exp(\theta S_F)]] \leq d \exp(cv + \frac{\lambda^2 w}{q\mathcal{M}^2} h_\lambda(\psi_p(\theta))).$$

The exponent is $K_{\lambda,p}(\theta)$ by the definition of c, proving Eq. (12). Finally, the matrix Laplace transform bound gives

$$\Pr_\mu[\lambda_{\max}(S_F) \geq t] \leq e^{-\theta t} \mathbb{E}_\mu[\mathrm{tr}[\exp(\theta S_F)]].$$

Taking the infimum over all admissible pairs proves Eq. (13). The argument applies directly to complex Hermitian matrices, since Eq. (9) already includes that case.

For $\lambda = 0$, every $p > 1$ and $\theta > 0$ is admissible, and the second term in Eq. (11) is zero. Fix $\theta > 0$ and let $p \downarrow 1$ in Eq. (12). Continuity gives

$$\mathbb{E}_\mu[\mathrm{tr}[\exp(\theta S_F)]] \leq d \exp(\frac{v}{\mathcal{M}^2} h_0(\mathcal{M}\theta)). \tag{19}$$

For $0 \leq x < 3$, comparison of Taylor coefficients using $k! \geq 2 \cdot 3^{k-2}$ for $k \geq 2$ gives

$$h_0(x) = \sum_{k=2}^{\infty} \frac{x^k}{k!} \leq \frac{x^2}{2(1 - x/3)}.$$

Hence the Laplace transform bound and Eq. (19) imply

$$\Pr_\mu[\lambda_{\max}(S_F) \geq t] \leq d \exp(-\theta t + \frac{v\theta^2}{2(1 - \mathcal{M}\theta/3)})$$

for $0 < \theta < 3/\mathcal{M}$. If $t > 0$ and $v > 0$, choose $\theta = t/(v + \mathcal{M}t/3)$ to obtain Eq. (14). The case $t = 0$ is trivial, and if $v = 0$, then all the summands vanish almost surely.

The independent conclusion also holds under Tropp's original one-sided assumption $F_j \preceq \mathcal{M}I$, with finite second moments. Indeed, the scalar exponential estimate in the proof of Lemma 6.2 holds for every eigenvalue at most $\mathcal{M}$. Independence and centering then give

$$\log \mathbb{E}_\mu[e^{\theta F_j}] \preceq g_{\mathcal{M}}(\theta) \mathbb{E}_\mu[F_j^2].$$

Applying [Tro12, Lemma 3.4] to these deterministic bounds and using $\sum_{j=1}^n \mathbb{E}_\mu[F_j^2] \preceq vI$ yields the same trace-MGF estimate and the same tail bound. Thus the independent recovery does not require a lower spectral bound.

We finally derive the simplified bound stated in Theorem 1.3. The scalar estimate from Section 5 controls the dependence correction while also verifying the admissible range of the tilting parameter.

Fix $p > 1$ and set $q := p/(p-1)$. The case $t = 0$ is trivial. If $v = 0$, then $\sum_j \mathbb{E}_\mu[F_j^2] = 0$ and each $\mathbb{E}_\mu[F_j^2]$ is positive semidefinite, so every $F_j(s_j)$ vanishes almost surely. Hence assume $t > 0$ and $v > 0$.

First suppose that $0 < \lambda < 1$. Write $a := \alpha(\lambda)$, $b := \beta(\lambda)$, $C := p + bq = C_p(\lambda)$, and $x := \mathcal{M}\theta$. Fix $0 < x < 1/C$. For $0 \leq y < 2$, comparison of Taylor coefficients using $k! \geq 2^{k-1}$ for $k \geq 1$ gives

$$e^y - 1 \leq \frac{y}{1 - y/2}.$$

Since $px < 1$, applying this estimate with $y = px$ yields

$$\psi_p(\theta) \leq \frac{qx}{1 - px/2} < \frac{1}{b}.$$

The strict inequality follows from $(p/2+bq)x < Cx < 1$. By the domain verification accompanying Eq. (10), we have $\lambda e^{\psi_p(\theta)} < 1$. Thus (p,θ) satisfies the required domain condition.

We can now show

$$\begin{aligned} h_\lambda(\psi_p(\theta)) &\le \frac{a\psi_p(\theta)^2}{2(1-b\psi_p(\theta))} \\ &\le \frac{aq^2x^2}{2(1-px/2)(1-(p/2+bq)x)} \\ &\le \frac{aq^2x^2}{2(1-Cx)}. \end{aligned}$$

The first step follows from Eq. (10). The second step uses the preceding bound on $\psi_p(\theta)$ and monotonicity of $u^2/(1-bu)$ on $0 \le u < 1/b$. The third step follows from

$$(1-px/2)(1-(p/2+bq)x) = 1 - Cx + \frac{p}{2}(\frac{p}{2}+bq)x^2 \ge 1 - Cx > 0.$$

Moreover,

$$h_0(px) \le \frac{p^2x^2}{2(1-px/3)} \le \frac{p^2x^2}{2(1-Cx)}.$$

The first inequality is the Taylor-coefficient estimate above, and the second uses $C \ge p/3$. Substituting these two estimates into Eq. (11), and using $x = \mathcal{M}\theta$, gives

$$K_{\lambda,p}(\theta) \le \frac{(pv+\alpha(\lambda)\lambda^2 qw)\theta^2}{2(1-C_p(\lambda)\mathcal{M}\theta)} = \frac{V_p(\lambda)\theta^2}{2(1-C_p(\lambda)\mathcal{M}\theta)}. \tag{20}$$

This holds for every $0 < \theta < 1/(C_p(\lambda)\mathcal{M})$.

When $\lambda = 0$, the correction term in Eq. (11) vanishes. Applying the same estimate for $h_0(p\mathcal{M}\theta)$ directly gives Eq. (20) with $V_p(0) = pv$ and $C_p(0) = p/3$, for every $0 < \theta < 3/(p\mathcal{M})$.

Finally, Eq. (12) and the matrix Laplace transform bound imply

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

Choose $\theta := t/(V_p(\lambda) + C_p(\lambda)\mathcal{M}t)$. Since $V_p(\lambda) \ge pv > 0$, this lies in the stated interval. Substitution gives exponent $-(t^2/2)/(V_p(\lambda)+C_p(\lambda)\mathcal{M}t)$, proving Theorem 1.3 for every $0 \le \lambda < 1$. □

6.2 Conditional trace estimates

We first record the conditional trace estimate used above. This is the exponential-supermartingale argument obtained from Lieb's concavity theorem, in the form of [Tro12, Corollary 3.3]. We keep the conditional mean explicitly, since the Markov summands need not form a martingale difference sequence.

Lemma 6.2 (Conditional trace-MGF estimate)**.** *Let $X_1, \ldots, X_n$ be Hermitian random matrices adapted to a filtration $(\mathcal{F}_j)_{j=0}^n$, with $\|X_j\| \le \mathcal{M}$ almost surely, where $\mathcal{M} > 0$. Write $\mathbb{E}_{j-1}[\cdot] := \mathbb{E}[\cdot \mid \mathcal{F}_{j-1}]$ and define*

$$S_k := \sum_{j=1}^k X_j, \qquad B_k := \sum_{j=1}^k \mathbb{E}_{j-1}[X_j], \qquad A_k := \sum_{j=1}^k \mathbb{E}_{j-1}[X_j^2].$$

For $z > 0$, let

$$g_{\mathcal{M}}(z) := \frac{e^{\mathcal{M}z} - \mathcal{M}z - 1}{\mathcal{M}^2}.$$

Then

$$\mathbb{E}[\mathrm{tr}[\exp(zS_n - zB_n - g_{\mathcal{M}}(z)A_n)]] \leq d. \tag{21}$$

Proof. For every real $x \leq \mathcal{M}$, Taylor's formula with integral remainder gives

$$e^{zx} - 1 - zx = z^2x^2 \int_0^1 (1-r)e^{rzx}\mathrm{d}r \leq g_{\mathcal{M}}(z)x^2.$$

The inequality follows from $e^{rzx} \leq e^{rz\mathcal{M}}$ for $r \in [0,1]$. Functional calculus and conditional expectation therefore imply

$$\mathbb{E}_{j-1}[e^{zX_j}] \preceq I + z\mathbb{E}_{j-1}[X_j] + g_{\mathcal{M}}(z)\mathbb{E}_{j-1}[X_j^2].$$

Set

$$C_j := z\mathbb{E}_{j-1}[X_j] + g_{\mathcal{M}}(z)\mathbb{E}_{j-1}[X_j^2].$$

The matrix $I + C_j$ is positive definite, since it dominates $\mathbb{E}_{j-1}[e^{zX_j}]$. Thus

$$\log \mathbb{E}_{j-1}[e^{zX_j}] \preceq \log(I + C_j) \preceq C_j. \tag{22}$$

The first step uses operator monotonicity of the logarithm. The second step follows by applying $\log(1+x) \leq x$ to the eigenvalues of C_j.

Define

$$R_k := zS_k - \sum_{j=1}^{k} C_j, \qquad R_0 := 0.$$

Since $R_{k-1} - C_k$ is $\mathcal{F}_{k-1}$-measurable, we can show

$$\begin{aligned}
\mathbb{E}_{k-1}[\mathrm{tr}[\exp(R_k)]] &= \mathbb{E}_{k-1}[\mathrm{tr}[\exp(R_{k-1} - C_k + zX_k)]] \\
&\leq \mathrm{tr}[\exp(R_{k-1} - C_k + \log \mathbb{E}_{k-1}[e^{zX_k}])] \\
&\leq \mathrm{tr}[\exp(R_{k-1})].
\end{aligned} \tag{23}$$

The first step follows from the definition of R_k. The second step is the conditional form of [Tro12, Corollary 3.3]. The third step follows from Eq. (22) and monotonicity of the trace exponential. Taking expectations and iterating gives $\mathbb{E}[\mathrm{tr}[\exp(R_n)]] \leq \mathrm{tr}[I] = d$. Since $R_n = zS_n - zB_n - g_{\mathcal{M}}(z)A_n$, this proves Eq. (21). □

The next lemma controls the centered correction after one application of the Markov kernel. The L^2 contraction is applied to vector-valued functions before taking an operator norm. This yields a factor λ^2 while preserving the individual variance matrix in the Loewner order.

Lemma 6.3 (Contraction of the centered correction)**.** *Under Definition 2.1, let $F : \mathcal{X} \to \mathbb{H}_d$ satisfy $\mathbb{E}_\mu[F] = 0$ and $\|F(x)\| \leq \mathcal{M}$. For $\theta, c \geq 0$, define*

$$\Sigma := \mathbb{E}_\mu[F^2], \qquad Z := P(\theta F + c(F^2 - \Sigma)).$$

Then

$$\mathbb{E}_\mu[Z] = 0, \qquad \|Z(x)\| \leq \mathcal{M}(\theta + c\mathcal{M}),$$

and

$$\mathbb{E}_\mu[Z^2] \preceq \lambda^2(\theta + c\mathcal{M})^2\Sigma. \tag{24}$$

Proof. Stationarity gives $\mathbb{E}_\mu[Z] = 0$. Since $0 \preceq F(x)^2 \preceq \mathcal{M}^2 I$ and $0 \preceq \Sigma \preceq \mathcal{M}^2 I$, we have $\|F(x)^2 - \Sigma\| \le \mathcal{M}^2$. Taking averages under the Markov kernel and using the triangle inequality proves the norm bound.

Fix $u \in \mathbb{C}^d$, with the Euclidean norm used for vectors. Expanding the square gives

$$\begin{aligned}\mathbb{E}_\mu[(F^2 - \Sigma)^2] &= \mathbb{E}_\mu[F^4] - \Sigma^2 \\ &\preceq \mathbb{E}_\mu[F^4] \\ &\preceq \mathcal{M}^2 \Sigma. \end{aligned} \tag{25}$$

The first step uses $\mathbb{E}_\mu[F^2] = \Sigma$. The second step uses $\Sigma^2 \succeq 0$, and the third step follows from $F^4 \preceq \mathcal{M}^2 F^2$ by functional calculus.

Put $T := \theta F + c(F^2 - \Sigma)$. The triangle inequality in $L^2(\mu; \mathbb{C}^d)$ gives

$$\begin{aligned}(\mathbb{E}_\mu[\|Tu\|^2])^{1/2} &\le \theta(\mathbb{E}_\mu[\|Fu\|^2])^{1/2} + c(\mathbb{E}_\mu[\|(F^2 - \Sigma)u\|^2])^{1/2} \\ &\le (\theta + c\mathcal{M})(u^* \Sigma u)^{1/2}.\end{aligned}$$

The second step follows from Eq. (25) and $\mathbb{E}_\mu[\|Fu\|^2] = u^* \Sigma u$.

Since $\mathbb{E}_\mu[T] = 0$, the L^2 contraction of $P - \Pi$, applied coordinatewise to Tu, implies

$$\begin{aligned}u^* \mathbb{E}_\mu[Z^2] u &= \mathbb{E}_\mu[\|P(Tu)\|^2] \\ &\le \lambda^2 \mathbb{E}_\mu[\|Tu\|^2] \\ &\le \lambda^2 (\theta + c\mathcal{M})^2 u^* \Sigma u.\end{aligned}$$

The same contraction holds for complex coordinates by applying the real L^2 estimate to real and imaginary parts. Since the displayed inequality holds for every u, it proves Eq. (24). $\square$

Acknowlegements and AI disclosure

The first version of this paper was completed on August 18, 2026. In preparing the first version, the author used Gemini Pro 3.1, Codex 5.6 Sol, and Claude Code Fable 5 to assist with language editing and grammar checking. The second version of this paper includes a new result, Theorem 1.3, with its proof in Section 6; this result does not appear in the first version. To obtain Theorem 1.3, the author used ChatGPT 6 to come up with proof ideas. All of the proofs have been substantially verified and rewritten by the author. The author takes full responsibility for the correctness of the proofs in this paper.

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.

[Rec11] Benjamin Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12(104):3413–3430, 2011.

[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.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il