From Poincaré inequalities to nonlinear matrix concentration

DE HUANG[*] and JOEL A. TROPP[†]

California Institute of Technology, Pasadena, USA. E-mail: [*]dhuang@caltech.edu; [†]jtropp@cms.caltech.edu

This paper deduces exponential matrix concentration from a Poincaré inequality via a short, conceptual argument. Among other examples, this theory applies to matrix-valued functions of a uniformly log-concave random vector. The proof relies on the subadditivity of Poincaré inequalities and a chain rule inequality for the trace of the matrix Dirichlet form. It also uses a symmetrization technique to avoid difficulties associated with a direct extension of the classic scalar argument.

Keywords: Concentration inequality; functional inequality; Markov process; matrix concentration; Poincaré inequality; semigroup

1. Introduction

Matrix concentration inequalities describe the probability that a random matrix is close to its expected value, with deviations measured by the ℓ_2 operator norm. These results have had a profound impact on a wide range of areas in computational mathematics and statistics. See the monograph [46] for an introduction to the subject and its applications.

As did the field of scalar concentration, matrix concentration theory began with simple models, such as sums of independent random matrices [1,29,42,45] and matrix-valued martingale sequences [33,40, 44]. In recent years, researchers have sought to develop results that hold for a wider class of random matrix models. Some initial successful efforts were based on exchangeable pairs techniques [30,36], but these methods do not address all cases of interest.

Researchers have also tried to extend scalar concentration techniques based on functional inequalities. An early attempt, by Chen and Tropp [11], demonstrates that (traces of) matrix variance and entropy quantities are subadditive, which leads to some Poincaré and modified log-Sobolev inequalities. A number of papers, including [12–14], have pursued this line of work. Unfortunately, these approaches have not been sufficient to reproduce all the results that have been established in the simpler models.

Very recently, Aoun et al. [3] have shown that a matrix form of the Poincaré inequality implies subexponential concentration of a random matrix with respect to the ℓ_2 operator norm. We believe that this is the first instance where a matrix functional inequality leads unconditionally to a matrix concentration result (with respect to the operator norm). Nevertheless, it remains an open question to deduce a full spectrum of matrix concentration results from matrix functional inequalities.

In this paper, we improve on [3] by demonstrating that the ordinary scalar Poincaré inequality also leads to subexponential concentration with respect to the operator norm. Our argument has some elements in common with the work in [3], but we have found a route to avoid most of the technical difficulty of their approach.

The basic idea is to bound the trace of an odd function (for example, the hyperbolic sine) of the random matrix using a symmetrization argument. The variance appears, and the Poincaré inequality yields a bound on the variance in terms of the Dirichlet form. Last, we apply a new matrix chain rule

inequality for the Dirichlet form to obtain a moment comparison. After this paper was written, we learned that Bobkov and Ledoux proposed a similar argument in the scalar case [5], Section 4.

As in the scalar setting, Poincaré inequalities may not capture the strongest concentration properties that are possible. In a companion paper [22], we demonstrate that *local* Poincaré inequalities lead to the optimal subgaussian concentration results. The analysis in [22] captures most of the previous results on matrix concentration, but it involves more technical machinery. We also refer the reader to work of Junge and Zeng [23] that contains similar results in the fully noncommutative setting.

An expert reader may still wonder about the role of (modified) log-Sobolev inequalities in establishing matrix concentration. At the time of writing, it is not clear how to obtain a matrix analog of the log-Sobolev inequality that would imply subgaussian matrix concentration inequalities in the same spirit as the ones in the related works [22,23].

2. Main result

This section summarizes our notation and the setting for our problem. It highlights our primary result on matrix concentration, and it gives a number of examples. In the next section, we comment further on the relationship with previous work.

2.1. Notation

Let $\mathbb{H}_d$ be the real linear space of $d \times d$ self-adjoint complex matrices, equipped with the ℓ_2 operator norm $\|\cdot\|$. We work with both the ordinary trace, tr, and the normalized trace, $\bar{\mathrm{tr}} := d^{-1}\,\mathrm{tr}$ on the space $\mathbb{H}_d$. Matrices, and occasionally vectors, are written in boldface italic. In particular, $\boldsymbol{f}$ and $\boldsymbol{g}$ refer to functions taking values in $\mathbb{H}_d$. The cone $\mathbb{H}_d^+$ contains the positive semidefinite matrices, and the symbol $\preccurlyeq$ refers to the semidefinite order.

Given a function $\varphi : \mathbb{R} \to \mathbb{R}$ taking real values, we extend it to a function $\varphi : \mathbb{H}_d \to \mathbb{H}_d$ on self-adjoint matrices by means of the spectral resolution:

$$\boldsymbol{A} = \sum_{\lambda \in \mathrm{spec}(\boldsymbol{A})} \lambda \boldsymbol{P}_\lambda \in \mathbb{H}_d \quad \text{implies} \quad \varphi(\boldsymbol{A}) = \sum_{\lambda \in \mathrm{spec}(\boldsymbol{A})} \varphi(\lambda) \boldsymbol{P}_\lambda \in \mathbb{H}_d.$$

Whenever we apply a scalar function, such as a power or a hyperbolic function, to a matrix, we are referring to the standard matrix function. Nonlinear functions bind before the trace.

We use familiar notation from probability. The operator $\mathbb{E}$ returns the expectation, and $\mathbb{P}\{\cdot\}$ is the probability of an event. The symbol $\sim$ means "has the distribution." Nonlinear functions bind before the expectation.

2.2. Random matrices

Let Ω be a Polish space, equipped with a probability measure μ, and write $\mathbb{E}_\mu$ for the integral with respect to the measure μ. Consider a μ-integrable matrix-valued function $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ on the state space Ω. By drawing a random variable $Z \sim \mu$, we can construct a random matrix $\boldsymbol{f}(Z)$. Our goal is to understand the concentration of $\boldsymbol{f}(Z)$ around its mean $\mathbb{E}_\mu \boldsymbol{f}$.

Example 2.1 (Gaussians). Consider the Gaussian space $(\mathbb{R}^n, \gamma_n)$ of n-dimensional real vectors equipped with the standard normal measure γ_n. Suppose we are interested in a matrix-valued function $\boldsymbol{f}(X)$ of a standard normal random vector $X \sim \gamma_n$. A familiar example [46], Chapter 5, is the

matrix Gaussian series

$$\boldsymbol{f}(\boldsymbol{X}) = \sum_{i=1}^{n} X_i \boldsymbol{A}_i \quad \text{where } \boldsymbol{X} \sim \gamma_n \text{ and } \boldsymbol{A}_1, \dots, \boldsymbol{A}_n \in \mathbb{H}_d \text{ are fixed.} \tag{2.1}$$

We will use the Gaussian case as a running example to illustrate the concepts that we introduce.

2.3. Markov processes

Suppose that we can identify an ergodic, reversible, time-homogeneous Markov process $(Z_t : t \geq 0) \subset \Omega$ with initial value Z_0 and stationary measure μ. This induces a matrix-valued Markov process $(\boldsymbol{f}(Z_t) : t \geq 0) \subset \mathbb{H}_d$.

By ergodicity, for any point $z \in \Omega$, we have the limit $\mathbb{E}[\boldsymbol{f}(Z_t) \mid Z_0 = z] \to \mathbb{E}_\mu \boldsymbol{f}$ as $t \to \infty$. The results in this paper build on the intuition that a random matrix $\boldsymbol{f}(Z)$ with $Z \sim \mu$ concentrates sharply about its mean when the matrix-valued Markov process tends quickly to equilibrium.

Example 2.2 (Gaussians). We can construct a reversible Markov process $(\boldsymbol{X}_t : t \geq 0) \subset \mathbb{R}^n$, called the *Ornstein–Uhlenbeck (OU) process*, by means of the stochastic differential equation

$$\mathrm{d}\boldsymbol{X}_t = -\boldsymbol{X}_t \, \mathrm{d}t + \sqrt{2} \, \mathrm{d}\boldsymbol{B}_t \quad \text{with initial value } \boldsymbol{X}_0 \in \mathbb{R}^n,$$

where $(\boldsymbol{B}_t : t \geq 0) \subset \mathbb{R}^n$ is random vector whose coordinates are independent Brownian motions. The stationary measure of the OU process is the standard normal distribution γ_n.

2.4. Derivatives and energy

To understand how quickly a matrix Markov process $\boldsymbol{f}(Z_t)$ converges to stationarity, we introduce notions of the "squared derivative" and the "energy" of the function $\boldsymbol{f}$.

Inspired by [3], we define the matrix *carré du champ operator* by the formula

$$\boldsymbol{\Gamma}(\boldsymbol{f})(z) := \lim_{t \downarrow 0} \frac{1}{2t} \mathbb{E}\big[\big(\boldsymbol{f}(Z_t) - \boldsymbol{f}(Z_0)\big)^2 \mid Z_0 = z\big] \in \mathbb{H}_d^+ \quad \text{for } z \in \Omega. \tag{2.2}$$

In many instances, the carré du champ $\boldsymbol{\Gamma}(\boldsymbol{f})$ has a natural interpretation as a squared derivative of $\boldsymbol{f}$. The expectation of the carré du champ is called the *matrix Dirichlet form*:

$$\mathcal{E}(\boldsymbol{f}) := \lim_{t \downarrow 0} \frac{1}{2t} \mathbb{E}_{Z \sim \mu}\big[\big(\boldsymbol{f}(Z_t) - \boldsymbol{f}(Z_0)\big)^2 \mid Z_0 = Z\big] \in \mathbb{H}_d^+. \tag{2.3}$$

The Dirichlet form $\mathcal{E}(\boldsymbol{f})$ reflects the total energy of the function $\boldsymbol{f}$.

In a general setting, it requires some care to make sense of the definitions (2.2) and (2.3). Without further comment, we restrict our attention to a "nice" class of functions where the limit in (2.2) exists pointwise and in $L_1(\mu)$ and where calculus operations are justified. By approximation, our main results on concentration hold for a wider class of functions.

Example 2.3 (Gaussians). According to [3], Proposition 5.5, the matrix carré du champ operator and matrix Dirichlet form of the OU process are determined by

$$\boldsymbol{\Gamma}(\boldsymbol{f})(\boldsymbol{x}) = \sum_{i=1}^{n}\big(\partial_i \boldsymbol{f}(\boldsymbol{x})\big)^2 \quad \text{for } \boldsymbol{x} \in \mathbb{R}^n \quad \text{and} \quad \mathcal{E}(\boldsymbol{f}) = \sum_{i=1}^{n} \mathbb{E}_{\gamma_n}(\partial_i \boldsymbol{f})^2.$$

The interpretations as a squared derivative and an energy are evident, and it is easy to check when the carré du champ is defined.

The matrix Gaussian series (2.1) provides an illustration:

$$\mathbf{\Gamma}(f) = \mathcal{E}(f) = \sum_{i=1}^{n} A_i^2.$$

This quantity is familiar from work on matrix concentration for Gaussian series [46], Chapter 4.

2.5. Trace Poincaré inequalities

The *matrix variance* of a function $f : \Omega \to \mathbb{H}_d$ with respect to the distribution μ is defined as

$$\mathbf{Var}_\mu[f] := \mathbb{E}_\mu\big[(f - \mathbb{E}_\mu\, f)^2\big] = \mathbb{E}_\mu\big[f^2\big] - (\mathbb{E}_\mu\, f)^2 \in \mathbb{H}_d^+. \tag{2.4}$$

As in the scalar case, the variance reflects fluctuations of the random matrix $f(Z)$ about its mean, where the random variable $Z \sim \mu$.

We say that the Markov process satisfies a *trace Poincaré inequality* with constant $\alpha > 0$ if

$$\mathrm{tr}\,\mathbf{Var}_\mu[f] \leq \alpha \cdot \mathrm{tr}\,\mathcal{E}(f) \quad \text{for all } f : \Omega \to \mathbb{H}_d. \tag{2.5}$$

In other words, the trace variance of $f(Z)$ is controlled by the energy in the function f. The inequality (2.5) provides a way to quantify the ergodicity of the Markov process.

As it happens, the trace Poincaré inequality is equivalent to an ordinary Poincaré inequality. We are grateful to Ramon Van Handel for this observation. The same result has recently appeared in the independent work of Garg et al. [16].

Proposition 2.4 (Equivalence of Poincaré inequalities). *Consider a Markov process* $(Z_t : t \geq 0) \subset \Omega$ *with stationary measure* μ. *The following are equivalent*:

1. *Scalar Poincaré. For all* $f : \Omega \to \mathbb{R}$, *we have* $\mathbf{Var}_\mu[f] \leq \alpha \cdot \mathcal{E}(f)$.
2. *Trace Poincaré. For all* $d \in \mathbb{N}$ *and all* $f : \Omega \to \mathbb{H}_d$, *we have* $\mathrm{tr}\,\mathbf{Var}_\mu[f] \leq \alpha \cdot \mathrm{tr}\,\mathcal{E}(f)$.
3. *Matrix Poincaré. For all* $d \in \mathbb{N}$ *and all* $f : \Omega \to \mathbb{H}_d$, *we have* $\mathbf{Var}_\mu[f] \preccurlyeq \alpha \cdot \mathcal{E}(f)$.

The Poincaré constant $\alpha \geq 0$ *is the same in all three cases.*

Proof. It is evident that (3) $\implies$ (2) $\implies$ (1). Our task is to prove that (1) $\implies$ (3).

Consider a real matrix-valued function $f : \Omega \to \mathbb{H}_d(\mathbb{R})$ with zero mean. For vectors $u, v \in \mathbb{R}^d$, define the scalar function $g(z) = \langle u, f(z)v \rangle \in \mathbb{R}$. Apply the scalar Poincaré inequality to g and invoke the definition (2.3) of the Dirichlet form. Thus,

$$\mathbb{E}_\mu \langle u, f(z)v \rangle^2 \leq \alpha \cdot \lim_{t \downarrow 0} \frac{1}{2t} \mathbb{E}_{Z \sim \mu}\big[\langle u, (f(Z_t) - f(Z_0))v \rangle^2 \mid Z_0 = Z\big].$$

Instate this inequality with $v = e_i$ for each $i = 1, \ldots, d$ and sum over i to arrive at

$$\langle u, \mathbf{Var}_\mu[f]u \rangle \leq \langle u, \mathcal{E}(f)u \rangle.$$

Since $u \in \mathbb{R}^d$ is arbitrary, this is the matrix Poincaré inequality (3).

To extend this argument to complex matrices, apply the same approach to the real and imaginary parts of the inner product.   □

The main result of this paper is that concentration properties of the random matrix $\boldsymbol{f}(Z)$ follow from the trace Poincaré inequality (2.5) or, equivalently, the scalar Poincaré inequality.

Example 2.5 (Gaussians). It is well known that the OU process satisfies the Poincaré inequality with constant $\alpha = 1$. Thus, it satisfies the trace Poincaré inequality (2.5) with $\alpha = 1$. For an alternative proof, see [3], Theorem 1.2.

2.6. Subexponential concentration and expectation bounds

We are now prepared to present our main result. It demands several hypotheses, which will be enforced throughout the paper.

Assumption 2.6 (Conditions). We assume that

1. The Markov process $(Z_t : t \geq 0) \subset \Omega$ is reversible and homogeneous, with initial value Z_0 and stationary measure μ.
2. The process admits a trace Poincaré inequality (2.5) with constant α. Equivalently, the process admits a scalar Poincaré inequality with the same constant α.
3. The class of valid functions is suitably restricted so that manipulations of expectations, limits, and derivatives are justified.

Under Assumption 2.6, we will deduce subexponential concentration of the random matrix $\boldsymbol{f}(Z)$ around its mean $\mathbb{E}_\mu \boldsymbol{f}$, where we measure the size of deviations with the ℓ_2 operator norm $\|\cdot\|$.

Theorem 2.7 (Subexponential Concentration). *Enforce Assumption 2.6. Let $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ be a matrix-valued function, and define the variance proxy*

$$v_f := \big\| \big\| \boldsymbol{\Gamma}(\boldsymbol{f})(z) \big\| \big\|_{L_\infty(\mu)}.$$

For all $\lambda > 0$,

$$\mathbb{P}_\mu \big\{ \| \boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f} \| \geq \sqrt{\alpha v_f} \cdot \lambda \big\} \leq 6d \cdot \mathrm{e}^{-\lambda}. \tag{2.6}$$

In particular,

$$\mathbb{E}_\mu \| \boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f} \| \leq \log(6\mathrm{e}d) \cdot \sqrt{\alpha v_f}. \tag{2.7}$$

The proof of Theorem 2.7 appears in Section 6 after we present some more background on matrix-valued Markov processes. Note that we have made no effort to refine constants.

The main point of the tail bound (2.6) is that the random matrix $\boldsymbol{f}(Z)$ exhibits exponential concentration on the scale $\sqrt{\alpha v_f}$. The variance proxy v_f is analogous to a global bound on the Lipschitz constant of $\boldsymbol{f}$. Be aware that we cannot achieve tail decay faster than exponential under the sole assumption of a Poincaré inequality, so this approach may not capture the strongest possible concentration. The leading constant in (2.6) reflects the ambient dimension d of the matrix; this feature is typical of matrix concentration bounds.

The expectation bound (2.7) shows that the average value of $\| \boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f} \|$ is proportional to the square root $\sqrt{v_f}$ of the variance proxy and to the logarithm of the ambient dimension. For many

examples, the optimal bound contains the *square-root* of the logarithm, but the result (2.7) is nontrivial and informative.

Example 2.8 (Gaussian Series). Theorem 2.7 applies to a matrix-valued function of a standard normal vector. For instance, according to (2.7), a matrix Gaussian series satisfies the expectation bound

$$\mathbb{E}\left\|\sum_{i=1}^{n} X_i A_i\right\| \leq \log(6\mathrm{e}d) \cdot \left\|\sum_{i=1}^{n} A_i^2\right\|^{1/2}.$$

Modulo the constant and the power on the logarithm, this bound is qualitatively correct for worst-case examples. On the other hand, the subexponential tail bound (2.6) does not reproduce the actual sub-Gaussian behavior. See [46], Chapter 5, for discussion. Example 7.3 describes an application to Gaussian chaos that requires tools more delicate than Theorem 2.7.

Remark 2.9 (Extensions). The bounds in Theorem 2.7 can be refined in several ways. We can replace the variance proxy v_f with less stringent measures of the size of the carré du champ $\boldsymbol{\Gamma}(\boldsymbol{f})$, such as the expected Schatten q-norm with $q \approx \log d$. It is also possible to replace the ambient dimension d with a measure of the intrinsic dimension of the random matrix $\boldsymbol{f}(Z)$. See Section 7.

Remark 2.10 (Rectangular case). By a standard formal argument, we can extend all the results here to a function $\boldsymbol{h} : \Omega \to \mathbb{M}_{d_1 \times d_2}$ that takes values in the $d_1 \times d_2$ complex matrices. To do so, we simply apply our results to the self-adjoint function

$$\boldsymbol{f}(z) = \begin{bmatrix} \mathbf{0} & \boldsymbol{h}(z) \\ \boldsymbol{h}(z)^* & \mathbf{0} \end{bmatrix} \in \mathbb{H}_{d_1+d_2} \quad \text{for } z \in \Omega.$$

See [46], Section 2.1.17, for details.

2.7. Examples

To indicate the scope of Theorem 2.7, let us present some more examples. Most of these examples are actually known to exhibit sub-Gaussian matrix concentration, but there is at least one case (Section 2.7.4) where the results here are currently the best available.

2.7.1. *Log-concave measures*

The Gaussian case is a particular example of a more general result for log-concave measures. Suppose that $J : \mathbb{R}^n \to \mathbb{R}$ is a strongly convex function that satisfies $\operatorname{Hess} J \succcurlyeq \eta \mathbf{I}$ uniformly for $\eta > 0$. Construct the probability measure μ on $\mathbb{R}^n$ whose density is proportional to e^{-J}. In this example, we briefly discuss concentration of matrix-valued functions $\boldsymbol{f}(X)$ where $X \sim \mu$. This model is interesting because it captures a type of negative dependence.

The appropriate Markov process $(X_t : t \geq 0)$ evolves with the stochastic differential equation

$$\mathrm{d}X_t = -\nabla J(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\boldsymbol{B}_t \quad \text{with initial value } X_0 \in \mathbb{R}^n,$$

where $(\boldsymbol{B}_t : t \geq 0) \subset \mathbb{R}^n$ is Brownian motion. The stationary distribution is μ, and the matrix carré du champ is

$$\boldsymbol{\Gamma}(\boldsymbol{f})(\boldsymbol{x}) = \sum_{i=1}^{n} \big(\partial_i \boldsymbol{f}(\boldsymbol{x})\big)^2 \quad \text{for } \boldsymbol{x} \in \mathbb{R}^n.$$

It is well known that these diffusions satisfy a Poincaré inequality with constant $\alpha = 1/\eta$; see [4], Corollary 4.8.2. Therefore, Theorem 2.7 applies.

A fortiori, these log-concave measures also satisfy a local Poincaré inequality, which leads to sub-Gaussian matrix concentration [22], Section 2.12.2.

2.7.2. *Riemannian manifolds with positive curvature*

More generally, let $(M, \mathfrak{g})$ be a compact Riemannian manifold with co-metric $\mathfrak{g}$. The manifold carries a canonical Riemannian probability measure $\mu_{\mathfrak{g}}$. The diffusion whose infinitesimal generator is the Laplace–Beltrami operator $\Delta_{\mathfrak{g}}$ is called the Brownian motion on the manifold. This is a reversible, ergodic Markov process. Its matrix carré du champ takes the form

$$\mathbf{\Gamma}(f)(z) = \sum_{i,j} g^{ij}(z)\big(\partial_i f(z)\big)\big(\partial_j f(z)\big) \quad \text{for } f : \Omega \to \mathbb{H}_d. \tag{2.8}$$

The co-metric g and the partial derivatives $\partial_i f$ are computed with respect to local coordinates. See [4] for an introduction to diffusions on manifolds; the companion paper [22] treats matrix-valued functions on manifolds.

Suppose that the manifold has uniformly positive Ricci curvature, where the Ricci tensor has eigenvalues bounded below by ρ. By now, it is a classic fact that the Brownian motion on this manifold satisfies the scalar Poincaré inequality with constant $\alpha = \rho^{-1}$. See [4], Section 4.8.

Example 2.11 (Sphere). For $n \geq 2$, the unit sphere $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ is a Riemannian submanifold of $\mathbb{R}^{n+1}$. Its canonical measure is the uniform probability distribution, and the carré du champ of the Brownian motion on the sphere is computed using (2.8). The sphere has positive Ricci curvature with $\rho = n - 1$, so it admits a Poincaré inequality with $\alpha = (n - 1)^{-1}$. Thus, matrix-valued functions on the sphere satisfy exponential matrix concentration.

A similar story can be told about every positively curved manifold. In fact, in this setting, we even have sub-Gaussian matrix concentration because of the stronger arguments in [22].

2.7.3. *Products*

Consider a probability space (Ω, μ). It is common to work with multivariate functions defined on the product space $(\Omega^n, \mu^{\otimes n})$. There is a standard construction [48], Section 2.3.2, of a Markov process on the product space. Aoun et al. [3] verify that the (matrix) carré du champ of this process is

$$\mathbf{\Gamma}(f)(z) = \frac{1}{2} \sum_{i=1}^n \mathbb{E}_{Z \sim \mu}\big[\big(f(z_1, \ldots, z_n) - f(z_1, \ldots, z_{i-1}, Z, z_{i+1}, \ldots, z_n)\big)^2\big].$$

This is the sum of squared discrete derivatives, each averaging over perturbations in a single coordinate. The variance proxy v_f takes the form

$$v_f = \operatorname{esssup}_{z \in \Omega^n} \frac{1}{2} \left\| \sum_{i=1}^n \mathbb{E}_{Z \sim \mu}\big[\big(f(z_1, \ldots, z_n) - f(z_1, \ldots, z_{i-1}, Z, z_{i+1}, \ldots, z_n)\big)^2\big] \right\|.$$

The variance proxy coincides with the matrix bounded difference that arises in Paulin et al. [36]. Aoun et al. prove that the Markov process satisfies a trace Poincaré inequality (2.5) with constant $\alpha = 1$. Therefore, Theorem 2.7 yields a suboptimal version of the exponential matrix Efron–Stein inequality [36], Theorem 4.3. See [3,22] for more details.

Remark 2.12 (Bernstein concentration?). In the scalar case, Bobkov and Ledoux [5], Corollary 3.2, have shown that a function $f : \Omega^n \to \mathbb{R}$ on a product space $(\Omega^n, \mu^{\otimes n})$ exhibits Bernstein-type concentration when μ admits a Poincaré inequality with constant α. In detail,

$$\mathbb{P}\{|f - \mathbb{E}\, f| > t\} \leq \exp\left(\frac{-ct^2}{v + Bt}\right) \quad \text{with } v = \left\|\sum_{i=1}^n (\partial_i f)^2\right\|_{L_\infty} \text{ and } B = \max_i \|\partial_i f\|_{L_\infty}.$$

The constant c depends on the constant α in the Poincaré inequality. Our approach does not seem powerful enough to transfer this insight to the matrix setting. Nevertheless, our paper [22] demonstrates that a *local* Poincaré inequality is sufficient to achieve Bernstein concentration.

2.7.4. *Stochastic covering property*

Aoun et al. [3] have considered a model for negatively dependent functions on the hypercube $\{0, 1\}^n$, namely the class of measures with the *stochastic covering property* (SCP). For a k-homogeneous measure μ with the SCP, it is possible to construct a Markov process that satisfies the trace Poincaré inequality (2.5) with constant $2k$. Thus, Theorem 2.7 applies. See [3,21,37] for a more complete discussion of this example.

Remark 2.13 (Subgaussian concentration?). Although the Markov process associated with an SCP measure satisfies a log-Sobolev inequality, we do not know if it satisfies the local Poincaré inequality that we would need to activate the sub-Gaussian concentration inequalities in [22].

2.7.5. *Expander graphs*

Let $G = (\Omega, E)$ be a k-regular, connected, undirected graph. We can construct a Markov process $(Z_t : t \geq 0) \subset \Omega$ by taking a continuous-time random walk on the vertex set Ω. The stationary measure μ of the random walk is the uniform measure on Ω. The carré du champ operator takes the form

$$\boldsymbol{\Gamma}(\boldsymbol{f})(z) = \frac{1}{2k} \sum_{(z',z) \in E} \left(\boldsymbol{f}(z') - \boldsymbol{f}(z)\right)^2.$$

In other words, the "squared gradient" is just the half the average squared difference between the matrix at the current vertex and its k neighbors. Thus,

$$v_f := \max_{z \in \Omega} \left\|\frac{1}{2k} \sum_{(z',z) \in E} \left(\boldsymbol{f}(z') - \boldsymbol{f}(z)\right)^2\right\|.$$

The variance proxy is just the maximum "squared gradient" at any vertex.

Assume that the Markov process satisfies the (scalar) Poincaré inequality, Proposition 2.4(1), with constant α. In this case, we say that G is an α-expander graph. According to Theorem 2.7, a matrix-valued function $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ on an α-expander graph satisfies a subexponential matrix concentration inequality:

$$\mathbb{P}\{\|\boldsymbol{f} - \mathbb{E}_\mu\, \boldsymbol{f}\| \geq \sqrt{\alpha v_f} \cdot \lambda\} \leq 6d \cdot \mathrm{e}^{-\lambda}.$$

Therefore, local control over the fluctuations yields global concentration around the mean. The number of vertices where the function departs from its mean is controlled by the quality α of the expander.

This example is potentially interesting because the Markov process does not satisfy a (modified) log-Sobolev inequality with a constant independent of the number of vertices in the graph [6], Section 5. Indeed, to achieve the functional inequality

$$\text{Ent}_\mu(f) \le \beta \cdot \mathcal{E}(f, \log f) \quad \text{for all } f : \Omega \to \mathbb{R},$$

it is necessary that $\beta \ge \text{const} \cdot \log(\#\Omega)$. See [32], Section 4.3, for related results on curvature-dimension conditions of Bakry–Émery type.

Remark 2.14 (Matrix Expander Chernoff). Although the modified log-Sobolev inequality fails, it is still possible to establish sub-Gaussian concentration for the ergodic averages of a matrix-valued random walk on an expander graph [17].

3. Related work

3.1. Markov processes

Much of the classical research on Markov processes concerns the relationship between the geometry of the state space and the behavior of canonical diffusion processes (e.g., Brownian motion on a Riemannian manifold). For an introduction, we recommend the lecture notes [48]. A more comprehensive source is the treatise [4].

Matrix-valued Markov processes first arose in the mathematical physics literature as a model for the evolution of a quantum system. Some of the foundational works include Davies [15] and Lindblad [28]. Quantum information theory has provided a new impetus for studying matrix-valued Markov processes; see [24] for a discussion and some background references.

Here, we are interested in a mixed classical-quantum setting, where a classical Markov process drives a matrix-valued function. Surprisingly, this model does not seem to have received much attention until the last few years. See Cheng et al. [14] for a more expansive framework that includes this case. Other foundational results appear in [3,22].

3.2. Functional inequalities

In the scalar setting, the connection between functional inequalities, convergence of Markov processes, and concentration is a long-standing topic of research. References include [4,7,26,48].

Functional inequalities for matrices were originally formulated in the mathematical physics literature; for example, see the work of Gross [19]. The application of functional inequalities to the ergodicity of matrix-valued Markov processes dates back at least as far as the papers [31,35].

Functional inequalities in the mixed classical-quantum setting seem to have a more recent vintage. Chen and Tropp [11] formulated subadditivity properties for tracial entropy-like quantities, including the trace variance (2.4). They showed that these properties imply some Sobolev and modified log-Sobolev-type inequalities for random matrices, and they obtained some restricted matrix concentration inequalities. Some of the partial results from [11] were completed in [12,20,41].

Cheng et al. [14] developed a framework for studying Markov processes in the mixed classical-quantum setting (and beyond), and they showed an equivalence between tracial log-Sobolev inequalities and exponential ergodicity of the trace entropy. Further results and implications for concentration appear in [13]. At present, we do not have a full picture of the relationships between matrix functional inequalities and matrix concentration.

Van Handel (personal communication) has pointed out that we can derive nonlinear matrix concentration for functions of a log-concave measure with a strongly convex potential by combining Pisier's method [39], the (sharp) noncommutative Khintchine inequality [8,47], and Caffarelli's contraction theorem [9]. This approach gives subgaussian concentration, which is better than we can obtain via the trace Poincaré inequality for log-concave measures, but it apparently does not extend beyond this setting.

3.3. From Poincaré to concentration

It has been recognized for about 40 years that Poincaré inequalities imply exponential concentration. Gromov and Milman [18], Theorem 4.1, prove such a theorem in the context of Riemannian manifolds. The standard argument, a recursive estimate of the moment generating function, is attributed to Aida and Stroock [2]. For a textbook presentation, see [7], Section 3.6.

Consider a matrix Poincaré inequality of the form

$$\mathbf{Var}_\mu[\boldsymbol{f}] \preccurlyeq \alpha \cdot \boldsymbol{\mathcal{E}}(\boldsymbol{f}). \tag{3.1}$$

Proposition 2.4 shows that this matrix Poincaré inequality (3.1) is also equivalent to a scalar Poincaré inequality with the same constant α. The papers [12–14] demonstrate that (3.1) leads to some inequalities for the matrix variance (2.4) and its trace, but these approaches do not lead to matrix concentration inequalities like Theorem 2.7.

Aoun, Banna, and Youssef [3] have recently shown that the matrix Poincaré inequality (3.1) does imply exponential concentration of a random matrix about its mean with respect to the ℓ_2 operator norm. Modulo constants, their result is equivalent with the tail bound (2.6), but it is weaker than the bounds in Theorem 7.1. The proof in [3] is a direct analog of the argument of Aida and Stroock [2]. But, in the matrix setting, the recursive estimate requires some heavy lifting. Another contribution of the paper [3] is to establish that some particular matrix-valued Markov processes satisfy (3.1). Nevertheless, Proposition 2.4 indicates that no additional effort is required on that end.

Our approach is similar in spirit to the work of Aoun et al. [3], but we use a symmetrization argument to avoid the difficult recursion. For related work in the scalar setting, see [5], Section 4. In a companion paper [22], we show that *local* Poincaré inequalities lead to much stronger ergodicity and concentration properties. The theory in the companion paper is significantly more involved than the development here, so we have chosen to separate them. For results on (local) Poincaré inequalities in noncommutative probability spaces, see Junge and Zeng [23].

4. Subadditivity

To control the variance of a function of several independent variables, it is helpful to understand the influence of each individual variable. As in the scalar setting, the matrix variance can be bounded by a sum of conditional variances. We can control each conditional variance by a conditional application of a trace Poincaré inequality. For simplicity, we focus on the case that is relevant to our proof, but these results hold in greater generality (products of more than two spaces that may carry different measures). Some of the material in this section is drawn from [11,12]. See [25] for classic work on subadditivity and concentration.

4.1. Influence of a coordinate

Consider the product space $(\Omega^2, \mu \otimes \mu)$. We want to study how individual coordinates affect the behavior of a matrix-valued function $\boldsymbol{g} : \Omega^2 \to \mathbb{H}_d$.

First, introduce notation for the expectation of the function with respect to each coordinate:

$$\mathbb{E}_1[\boldsymbol{g}](z_2) := \mathbb{E}_{Z \sim \mu}\big[\boldsymbol{g}(Z, z_2)\big] \in \mathbb{H}_d \quad \text{for all } z_1 \in \Omega;$$

$$\mathbb{E}_2[\boldsymbol{g}](z_1) := \mathbb{E}_{Z \sim \mu}\big[\boldsymbol{g}(z_1, Z)\big] \in \mathbb{H}_d \quad \text{for all } z_2 \in \Omega.$$

The coordinate-wise variance is the positive-semidefinite random matrix

$$\mathbf{Var}_i[\boldsymbol{g}] := \mathbb{E}_i\big[(\boldsymbol{g} - \mathbb{E}_i\,\boldsymbol{g})^2\big] \in \mathbb{H}_d^+ \quad \text{for } i = 1, 2.$$

This matrix reflects the fluctuation in the ith coordinate, with the other coordinate held fixed.

Similarly, we can introduce the coordinate-wise carré du champ operator and Dirichlet form:

$$
\begin{aligned}
\boldsymbol{\Gamma}_1(\boldsymbol{g})(z_1, z_2) &:= \lim_{t \downarrow 0} \frac{1}{2t} \mathbb{E}\big[\big(\boldsymbol{g}(Z_t, z_2) - \boldsymbol{g}(Z_0, z_2)\big)^2 \mid Z_0 = z_1\big] \in \mathbb{H}_d^+; \\
\boldsymbol{\mathcal{E}}_1(\boldsymbol{g})(z_2) &:= \lim_{t \downarrow 0} \frac{1}{2t} \mathbb{E}_{Z \sim \mu}\big[\big(\boldsymbol{g}(Z_t, z_2) - \boldsymbol{g}(Z_0, z_2)\big)^2 \mid Z_0 = Z\big] \in \mathbb{H}_d^+.
\end{aligned}
\tag{4.1}
$$

As usual, $(Z_t : t \geq 0)$ is a realization of the Markov process with initial value Z_0. We make analogous definitions for the second coordinate $i = 2$.

Last, we extend the carré du champ operator and the Dirichlet form to bivariate functions:

$$\boldsymbol{\Gamma}(\boldsymbol{g}) := \boldsymbol{\Gamma}_1(\boldsymbol{g}) + \boldsymbol{\Gamma}_2(\boldsymbol{g}) \in \mathbb{H}_d^+; \tag{4.2}$$

$$\boldsymbol{\mathcal{E}}(\boldsymbol{g}) := \mathbb{E}_{\mu \otimes \mu}\big[\boldsymbol{\mathcal{E}}_1(\boldsymbol{g}) + \boldsymbol{\mathcal{E}}_2(\boldsymbol{g})\big] \in \mathbb{H}_d^+. \tag{4.3}$$

These formulas have a heuristic interpretation: the squared derivative of a bivariate function is the sum of the squared partial derivatives.

4.2. Trace variance is subadditive

Observe that the trace variance is controlled by the sum of the coordinate-wise variances.

Fact 4.1 (Trace variance: Subadditivity). Let $\boldsymbol{g} : \Omega^2 \to \mathbb{H}_d$ be a matrix-valued function on the product space $(\Omega^2, \mu \otimes \mu)$. Then

$$\operatorname{tr} \mathbf{Var}_{\mu \otimes \mu}[\boldsymbol{g}] \leq \mathbb{E}_2 \operatorname{tr} \mathbf{Var}_1[\boldsymbol{g}] + \mathbb{E}_1 \operatorname{tr} \mathbf{Var}_2[\boldsymbol{g}].$$

This result is due to Chen and Tropp [11], who showed that other matrix functions are also subadditive. Later, Cheng and Hsieh [12] noticed that an analogous result holds without the trace. Similar decompositions are also valid for functions on the n-fold product $(\Omega_1 \otimes \Omega_2 \otimes \cdots \otimes \Omega_n, \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n)$. See [48], Section 2.1, for the standard result in the scalar setting.

Proof. The proof is the same as in the scalar setting. Writing $\mathbb{E} = \mathbb{E}_{\mu \otimes \mu}$ for the total expectation,

$$\mathbb{E}\,\mathbf{Var}_{\mu \otimes \mu}[\boldsymbol{g}] := \mathbb{E}\big[(\boldsymbol{g} - \mathbb{E}\,\boldsymbol{g})^2\big] = \mathbb{E}\big[(\boldsymbol{g} - \mathbb{E}_1\,\boldsymbol{g})^2 + (\mathbb{E}_1\,\boldsymbol{g} - \mathbb{E}_1\,\mathbb{E}_2\,\boldsymbol{g})^2\big]$$
$$\preccurlyeq \mathbb{E}\big[(\boldsymbol{g} - \mathbb{E}_1\,\boldsymbol{g})^2\big] + \mathbb{E}\big[(\boldsymbol{g} - \mathbb{E}_2\,\boldsymbol{g})^2\big]$$
$$= \mathbb{E}_2\,\mathbf{Var}_1[\boldsymbol{g}] + \mathbb{E}_1\,\mathbf{Var}_2[\boldsymbol{g}].$$

The first line relies on orthogonality of the conditionally centered random matrices. The second line requires the operator convexity of the square, applied conditionally. Last, take the trace. $\square$

4.3. Trace Poincaré inequalities are subadditive

If the Markov process satisfies a trace Poincaré inequality, then the variance of a bivariate function also satisfies a trace Poincaré inequality.

Proposition 4.2 (Trace Poincaré: Subadditivity). *Suppose that the Markov process satisfies a trace Poincaré inequality* (2.5) *with constant* α. *Let* $\boldsymbol{g} : \Omega^2 \to \mathbb{H}_d$ *be a suitable bivariate matrix-valued function. Then*

$$\mathrm{tr}\,\mathbf{Var}_{\mu \otimes \mu}[\boldsymbol{g}] \leq \alpha \cdot \mathrm{tr}\,\boldsymbol{\mathcal{E}}(\boldsymbol{g}).$$

Proof. Start with Fact 4.1. Apply the trace Poincaré inequality (2.5) coordinate-wise to control each of the two coordinate-wise variances. Last, introduce the definition (4.3) of the Dirichlet form for a bivariate function. $\square$

5. Chain rule inequality for the Dirichlet form

The key new tool in our approach is a simple trace inequality for the matrix Dirichlet form that shows how it interacts with composition.

Proposition 5.1 (Chain rule inequality). *Enforce Assumption* 2.6. *Let* $\varphi : \mathbb{R} \to \mathbb{R}$ *be a scalar function whose squared derivative* $\psi = (\varphi')^2$ *is convex. Then*

$$\mathrm{tr}\,\boldsymbol{\mathcal{E}}\big(\varphi(\boldsymbol{f})\big) = \mathbb{E}_\mu\,\mathrm{tr}\,\boldsymbol{\Gamma}\big(\varphi(\boldsymbol{f})\big) \leq \mathbb{E}_\mu\,\mathrm{tr}\big[\boldsymbol{\Gamma}(\boldsymbol{f})\psi(\boldsymbol{f})\big] \quad \textit{for all suitable } \boldsymbol{f} : \Omega \to \mathbb{H}_d.$$

In particular,

$$\mathrm{tr}\,\boldsymbol{\mathcal{E}}\big(\varphi(\boldsymbol{g})\big) = \mathbb{E}_{\mu \otimes \mu}\,\mathrm{tr}\,\boldsymbol{\Gamma}\big(\varphi(\boldsymbol{g})\big) \leq \mathbb{E}_{\mu \otimes \mu}\,\mathrm{tr}\big[\boldsymbol{\Gamma}(\boldsymbol{g})\psi(\boldsymbol{g})\big] \quad \textit{for all suitable } \boldsymbol{g} : \Omega^2 \to \mathbb{H}_d.$$

The proof of Proposition 5.1 consumes the rest of this section.

For context, recall that the carré du champ operator Γ of a (scalar-valued, reversible) diffusion process satisfies a chain rule [4], Section 1.11:

$$\Gamma\big(\varphi(f)\big) = \Gamma(f)\varphi'(f)^2 \quad \text{for smooth } f \text{ and } \varphi.$$

Proposition 5.1 provides a substitute for this relation for an arbitrary reversible Markov process that takes matrix values. In exchange for the wider applicability, we need some additional averaging (provided by the Dirichlet form); we must restrict our attention to functions φ with a convexity property; and the equality is relaxed to an inequality. In the scalar case, Proposition 5.1 is related to the Stroock–Varopoulos inequality [43,49].

5.1. Mean-value inequality for trace functions

The argument hinges on an elementary trace inequality for deterministic matrices. This result is obtained by lifting a numerical inequality to self-adjoint matrices. A very similar statement [30], Lemma 3.4, animates the exchangeable pairs approach to matrix concentration, which is motivated by work in the scalar setting [10].

Lemma 5.2 (Mean-value trace inequality). *Let $A, B \in \mathbb{H}_d$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a scalar function whose squared derivative $\psi = (\varphi')^2$ is convex. Then*

$$\operatorname{tr}\big[\big(\varphi(A) - \varphi(B)\big)^2\big] \leq \frac{1}{2}\operatorname{tr}\big[(A - B)^2\big(\psi(A) + \psi(B)\big)\big].$$

Proof. Let $a, b \in \mathbb{R}$. The fundamental theorem of calculus and Jensen's inequality together deliver the relations

$$\big(\varphi(a) - \varphi(b)\big)^2 = (a - b)^2 \bigg[\int_0^1 \mathrm{d}\tau\, \varphi'\big(\tau a + (1 - \tau)b\big)\bigg]^2$$

$$\leq (a - b)^2 \int_0^1 \mathrm{d}\tau\, \psi\big(\tau a + (1 - \tau)b\big)$$

$$\leq (a - b)^2 \int_0^1 \mathrm{d}\tau\, \big[\tau \psi(a) + (1 - \tau)\psi(b)\big] = \frac{1}{2}(a - b)^2\big(\psi(a) + \psi(b)\big).$$

The generalized Klein inequality [38], Proposition 3, allows us to lift this numerical fact to a trace inequality for matrices $A, B \in \mathbb{H}_d$. □

5.2. Proof of Proposition 5.1

The result follows from a short calculation. First, we use the definition (2.3) of the Dirichlet form as a limit:

$$\operatorname{tr}\mathcal{E}\big(\varphi(f)\big) = \lim_{t \downarrow 0} \frac{1}{2t}\, \mathbb{E}_{Z \sim \mu}\big[\operatorname{tr}\big[\big(\varphi\big(f(Z_t)\big) - \varphi\big(f(Z_0)\big)\big)^2\big] \mid Z_0 = Z\big]$$

$$\leq \lim_{t \downarrow 0} \frac{1}{4t}\, \mathbb{E}_{Z \sim \mu}\operatorname{tr}\big[\big(f(Z_t) - f(Z_0)\big)^2\big(\psi\big(f(Z_t)\big) + \psi\big(f(Z_0)\big)\big) \mid Z_0 = Z\big]$$

$$= \lim_{t \downarrow 0} \frac{1}{2t}\, \mathbb{E}_{Z \sim \mu}\operatorname{tr}\big[\big(f(Z_t) - f(Z_0)\big)^2 \psi\big(f(Z_0)\big) \mid Z_0 = Z\big]$$

$$= \mathbb{E}_{Z \sim \mu}\operatorname{tr}\bigg[\lim_{t \downarrow 0} \frac{1}{2t}\, \mathbb{E}\big[\big(f(Z_t) - f(Z_0)\big)^2 \mid Z_0 = Z\big]\psi\big(f(Z)\big)\bigg]$$

$$= \mathbb{E}_\mu \operatorname{tr}\big[\Gamma(f)\psi(f)\big].$$

The inequality is Lemma 5.2. To reach the third line, we use the fact that (Z_0, Z_t) is an exchangeable pair for each $t \geq 0$, a consequence of the reversibility of the Markov process $(Z_t : t \geq 0)$ and the fact that $Z_0 \sim \mu$. Last, we condition on the value of Z_0, invoke dominated convergence to pass the expectation through the limit, and we apply the definition (2.2) of the carré du champ operator.

6. Exponential moments

Our main technical result is a bound for the exponential moments of a general matrix-valued function on the state space. In contrast to the usual approach of bounding the moment generating function, we will compute the expectation of a hyperbolic function.

Theorem 6.1 (Exponential moments). *Enforce Assumption* 2.6. *Let $f : \Omega \to \mathbb{H}_d$ be a function with $\mathbb{E}_\mu f = 0$. For $\theta > 0$,*

$$\mathbb{E}_\mu \operatorname{tr} \cosh(\theta f) \le d \cdot \left[1 + \frac{\alpha \theta^2 \, \bar{\operatorname{tr}} \, \mathcal{E}(f)}{(1 - \alpha v_f \theta^2/2)_+} \right]. \tag{6.1}$$

The variance proxy v_f is defined in (2.7), *and $\bar{\operatorname{tr}}$ is the normalized trace.*

The proof of Theorem 6.1 occupies the rest of the section. But first, we use this moment bound to derive our main result, Theorem 2.7.

Proof of Theorem 2.7 from Theorem 6.1. Without loss, assume that $\mathbb{E}_\mu f = 0$. We use the matrix moment method [1,34]:

$$\mathbb{P}_\mu \{ \|f\| \ge \lambda \} \le \inf_{\theta > 0} \frac{1}{\cosh(\theta\lambda)} \cdot \mathbb{E}_\mu \cosh(\theta\|f\|) = \inf_{\theta > 0} \frac{2}{e^{\theta\lambda} + e^{-\theta\lambda}} \cdot \mathbb{E}_\mu \|\cosh(\theta f)\|$$

$$\le \inf_{\theta > 0} \frac{2}{e^{\theta\lambda}} \cdot \mathbb{E}_\mu \operatorname{tr} \cosh(\theta f) \le 2d \cdot \inf_{\theta > 0} e^{-\theta\lambda} \cdot \left[1 + \frac{\alpha \theta^2 \, \bar{\operatorname{tr}} \, \mathcal{E}(f)}{(1 - \alpha v_f \theta^2/2)_+} \right]$$

$$\le 2d \cdot e^{-\lambda/\sqrt{\alpha v_f}} \cdot \left[1 + \frac{2 \, \bar{\operatorname{tr}} \, \mathcal{E}(f)}{v_f} \right] \le 6d \cdot e^{-\lambda/\sqrt{\alpha v_f}}.$$

The first inequality is Markov's. The second relation is the spectral mapping theorem. The ℓ_2 operator norm of a positive-definite matrix is obviously bounded by its trace. Then invoke Theorem 6.1 to control the moment. We have chosen $\theta = (\alpha v_f)^{-1/2}$. Last, we have noted that $\bar{\operatorname{tr}} \, \mathcal{E}(f) \le \|\mathcal{E}(f)\| \le v_f$. To finish the proof of the tail bound (2.6), make the change of variables $\lambda \mapsto \lambda \sqrt{\alpha v_f}$.

The expectation bound (2.7) follows when we integrate the tail bound (2.6), taking into account that the probability is also bounded by one; we omit the details. □

6.1. Proof of Theorem 6.1

Our goal is to develop an exponential moment bound for a function $f : \Omega \to \mathbb{H}_d$ that satisfies $\mathbb{E}_\mu f = 0$. We will need to work with both the hyperbolic sine and cosine, passing between them using simple identities. After writing this paper, we learned that this proof is a matrix analog of an argument proposed by Bobkov and Ledoux [5], Section 4.

6.1.1. *Symmetrization*

The first step is to symmetrize the function. Let $Z, Z' \in \Omega$ be independent random variables, each with distribution μ. Since $f(Z')$ has mean zero, a conditional application of Jensen's inequality yields

$$\mathbb{E} \operatorname{tr} \sinh^2(\theta f(Z)) \le \mathbb{E} \operatorname{tr} \sinh^2(\theta(f(Z) - f(Z'))) =: \mathbb{E} \operatorname{tr} \sinh^2(\theta g(Z, Z')). \tag{6.2}$$

Indeed, since $\sinh^2$ is convex, the function $\operatorname{tr} \sinh^2$ is also convex [38], Proposition 2. We have defined the antisymmetric function $g(z, z') := f(z) - f(z')$ for $z, z' \in \Omega$.

6.1.2. *From moments to variance*

Next, we will write the expectation as a variance. Consider the odd function $\varphi(s) = \sinh(\theta s)$. First, we claim that

$$\mathbb{E}\sinh\big(\theta\boldsymbol{g}(Z, Z')\big) = \boldsymbol{0}.$$

Indeed, using the antisymmetry of $\boldsymbol{g}$ and the oddness of φ,

$$\mathbb{E}\sinh\big(\theta\boldsymbol{g}(Z, Z')\big) = \mathbb{E}\sinh\big(-\theta\boldsymbol{g}(Z', Z)\big) = -\mathbb{E}\sinh\big(\theta\boldsymbol{g}(Z', Z)\big) = -\mathbb{E}\sinh\big(\theta\boldsymbol{g}(Z, Z')\big).$$

The last identity holds because (Z, Z') is exchangeable.

As an immediate consequence,

$$\mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\sinh^2(\theta\boldsymbol{g}) = \mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\sinh^2(\theta\boldsymbol{g}) - \operatorname{tr}\big[\big(\mathbb{E}_{\mu\otimes\mu}\sinh(\theta\boldsymbol{g})\big)^2\big] = \operatorname{tr}\mathbf{Var}_{\mu\otimes\mu}\big[\sinh(\theta\boldsymbol{g})\big].$$

The appearance of the variance gives us access to Poincaré inequalities.

6.1.3. *Poincaré inequality*

To continue, we apply Proposition 4.2, the trace Poincaré inequality for bivariate functions:

$$\mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\sinh^2(\theta\boldsymbol{g}) \leq \alpha \cdot \operatorname{tr}\boldsymbol{\mathcal{E}}\big(\sinh(\theta\boldsymbol{g})\big) \leq \alpha\theta^2 \cdot \mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\big[\boldsymbol{\Gamma}(\boldsymbol{g})\cosh^2(\theta\boldsymbol{g})\big]. \tag{6.3}$$

The second inequality is the chain rule, Proposition 5.1, for the Dirichlet form. To activate it, we note that the squared derivative of $\varphi(s) = \sinh(\theta s)$ is the convex function $\psi(s) = \theta^2 \cosh^2(\theta s)$.

6.1.4. *Moment comparison*

A moment comparison argument allows us to isolate the exponential moment. Define the variance proxy of the bivariate function:

$$v_{\boldsymbol{g}} := \big\|\|\boldsymbol{\Gamma}(\boldsymbol{g})\|\big\|_{L_\infty(\mu\otimes\mu)}. \tag{6.4}$$

Continuing from (6.3), the identity $\cosh^2 = 1 + \sinh^2$ implies that

$$\mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\sinh^2(\theta\boldsymbol{g}) \leq \alpha\theta^2 \cdot \mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\boldsymbol{\Gamma}(\boldsymbol{g}) + \alpha\theta^2 \cdot \mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\big[\boldsymbol{\Gamma}(\boldsymbol{g})\sinh^2(\theta\boldsymbol{g})\big]$$

$$\leq \alpha\theta^2 \cdot \operatorname{tr}\boldsymbol{\mathcal{E}}(\boldsymbol{g}) + \alpha\theta^2 v_{\boldsymbol{g}} \cdot \mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\sinh^2(\theta\boldsymbol{g}).$$

The second inequality is just the usual operator-norm bound for the trace. Rearrange this identity to arrive at

$$\mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\sinh^2(\theta\boldsymbol{g}) \leq \frac{\alpha\theta^2 \operatorname{tr}\boldsymbol{\mathcal{E}}(\boldsymbol{g})}{(1 - \alpha v_{\boldsymbol{g}}\theta^2)_+}. \tag{6.5}$$

It remains to revert to the original function $\boldsymbol{f}$.

6.1.5. *Comparison of carré du champs*

Let us compute the Dirichlet form $\boldsymbol{\mathcal{E}}(\boldsymbol{g})$ and the variance proxy $v_{\boldsymbol{f}}$ in terms of the original function $\boldsymbol{f}$. To that end, observe that the coordinate-wise carré du champ (4.1) satisfies

$$\boldsymbol{\Gamma}_1(\boldsymbol{g})(z, z') = \mathbb{E}\lim_{t\downarrow 0}\frac{1}{2t}\big[\big(\boldsymbol{g}(Z_t, z') - \boldsymbol{g}(Z_0, z')\big)^2 \mid Z_0 = z\big]$$

$$= \mathbb{E}\lim_{t\downarrow 0}\frac{1}{2t}\big[\big(\boldsymbol{f}(Z_t) - \boldsymbol{f}(Z_0)\big)^2 \mid Z_0 = z\big] = \boldsymbol{\Gamma}(\boldsymbol{f})(z).$$

A similar calculation reveals that $\boldsymbol{\Gamma}_2(\boldsymbol{g})(z, z') = \boldsymbol{\Gamma}(\boldsymbol{f})(z')$. Thus, the bivariate carré du champ (4.2) takes the form

$$\boldsymbol{\Gamma}(\boldsymbol{g})(z, z') = \boldsymbol{\Gamma}_1(\boldsymbol{g})(z, z') + \boldsymbol{\Gamma}_2(\boldsymbol{g})(z, z') = \boldsymbol{\Gamma}(\boldsymbol{f})(z) + \boldsymbol{\Gamma}(\boldsymbol{f})(z'). \tag{6.6}$$

As a consequence, the Dirichlet form can be calculated as

$$\boldsymbol{\mathcal{E}}(\boldsymbol{g}) = \mathbb{E}_{\mu \otimes \mu}\, \boldsymbol{\Gamma}(\boldsymbol{g}) = \mathbb{E}_{\mu \otimes \mu}\big[\boldsymbol{\Gamma}(\boldsymbol{f})(z) + \boldsymbol{\Gamma}(\boldsymbol{f})(z')\big] = 2\boldsymbol{\mathcal{E}}(\boldsymbol{f}). \tag{6.7}$$

The variance proxy (6.4) of the bivariate function satisfies

$$v_{\boldsymbol{g}} = \big\|\|\boldsymbol{\Gamma}(\boldsymbol{g})(z, z')\|\big\|_{L_\infty(\mu \otimes \mu)} \le \big\|\|\boldsymbol{\Gamma}(\boldsymbol{f})(z)\|\big\|_{L_\infty(\mu)} + \big\|\|\boldsymbol{\Gamma}(\boldsymbol{f})(z')\|\big\|_{L_\infty(\mu)} = 2v_f. \tag{6.8}$$

The last relation is the definition (2.7) of the variance proxy v_f.

6.1.6. *Endgame*

Combining (6.2) and (6.5), we see that

$$\mathbb{E}_\mu\, \mathrm{tr}\sinh^2(\theta \boldsymbol{f}) \le \frac{\alpha\theta^2\, \mathrm{tr}\,\boldsymbol{\mathcal{E}}(\boldsymbol{g})}{(1 - \alpha v_{\boldsymbol{g}}\theta^2)_+} \le \frac{2\alpha\theta^2\, \mathrm{tr}\,\boldsymbol{\mathcal{E}}(\boldsymbol{f})}{(1 - 2\alpha v_f\theta^2)_+}.$$

We have also used the relations (6.7) and (6.8) from the last section.

To compete the proof of (6.1), invoke the identity $\sinh^2(s) = (\cosh(2s) - 1)/2$ to see that

$$\mathbb{E}_\mu\, \mathrm{tr}\cosh(2\theta \boldsymbol{f}) \le d + \frac{4\alpha\theta^2\, \mathrm{tr}\,\boldsymbol{\mathcal{E}}(\boldsymbol{f})}{(1 - 2\alpha v_f\theta^2)_+}.$$

Finally, introduce the normalized trace, $\bar{\mathrm{tr}}$, and make the change of variables $\theta \mapsto \theta/2$ to arrive at

$$\mathbb{E}_\mu\, \mathrm{tr}\cosh(\theta \boldsymbol{f}) \le d \cdot \left[1 + \frac{\alpha\theta^2\, \bar{\mathrm{tr}}\,\boldsymbol{\mathcal{E}}(\boldsymbol{f})}{(1 - \alpha v_f\theta^2/2)_+}\right].$$

This is the stated result.

7. Polynomial moments

By a simple variation on the proof of Theorem 6.1, we can also obtain bounds for the polynomial moments of a random matrix.

Theorem 7.1 (Polynomial moments). *Enforce Assumption 2.6. Let $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ be a function with $\mathbb{E}_\mu\, \boldsymbol{f} = \boldsymbol{0}$. For $q = 1$ and $q \ge 1.5$,*

$$\big(\mathbb{E}_\mu\, \mathrm{tr}\,|\boldsymbol{f}|^{2q}\big)^{1/(2q)} \le \sqrt{2\alpha q^2} \cdot \big(\mathbb{E}_\mu\, \mathrm{tr}\,\boldsymbol{\Gamma}(\boldsymbol{f})^q\big)^{1/(2q)}. \tag{7.1}$$

By combining Theorem 7.1 with the moment method, we can obtain probability bounds for $\|f\|$. Let us summarize how these results compare with the main result, Theorem 2.7. Observe that Theorem 7.1 gives a bound on the Schatten $2q$-norm of the random matrix $f(Z)$ in terms of the Schatten $2q$-norm of $\Gamma(f)^{1/2}$. We have the relation

$$\left(\mathbb{E}_\mu \operatorname{tr} \Gamma(f)^q\right)^{1/(2q)} \le d^{1/(2q)} \cdot \left\|\,\|\Gamma(f)(z)\|\,\right\|_{L_\infty(\mu)}^{1/2} = d^{1/(2q)} \cdot \sqrt{v_f}. \tag{7.2}$$

Therefore, Theorem 7.1 potentially yields stronger bounds than Theorem 6.1.

In particular, Theorem 7.1 applies even when $\Gamma(f)$ is not uniformly bounded. Example 7.3 illustrates why this flexibility is valuable. In Section 7.2, we show that slightly better polynomial moment bounds are possible when $\Gamma(f)$ is uniformly bounded.

Remark 7.2 (Missing Parameters). Theorem 7.1 also holds for $q \in (1, 1.5)$, with an extra factor of $\sqrt{2}$ on the right-hand side. The proof uses a variant of Proposition 5.1 that only requires the function ψ to be monotone.

Example 7.3 (Gaussian Chaos). Consider the matrix Gaussian chaos

$$f(X) = \sum_{i,j=1}^n X_i X_j A_{ij} \quad \text{where } X \sim \gamma_n \text{ and } A_{ij} = A_{ji} \in \mathbb{H}_d.$$

To bound the trace moments of $f(X)$, observe that the carré du champ takes the form

$$\Gamma(f)(x) = \sum_{i=1}^n \left(\partial_i f(x)\right)^2 = 4 \sum_{i=1}^n \left(\sum_{j=1}^n x_j A_{ij}\right)^2.$$

Evidently, $\Gamma(f)$ is unbounded, so Theorem 2.7 does not apply. But Theorem 7.1 yields

$$\left(\mathbb{E}_{\gamma_n} \operatorname{tr} |f|^{2q}\right)^{1/(2q)} \le \sqrt{8q^2} \cdot \left(\mathbb{E}_{\gamma_n} \operatorname{tr}\left[\sum_{i=1}^n \left(\sum_{j=1}^n X_j A_{ij}\right)^2\right]^q\right)^{1/(2q)}. \tag{7.3}$$

We have used the fact that the Poincaré constant of the OU process is $\alpha = 1$. Further bounds can be obtained by applying Theorem 7.1 repeatedly. In the matrix setting, there are obstacles that prevent us from simplifying (7.3) completely (related to the fact that the partial transpose operator is not completely bounded).

In the scalar case $d = 1$, we can obtain more transparent conclusions. Consider the scalar-valued Gaussian chaos

$$f(X) = \sum_{i,j=1}^n X_i X_j a_{ij}.$$

The most satisfying outcome takes place when $A = [a_{ij}]$ is positive semidefinite. In this case, the result (7.3) implies that

$$\left(\mathbb{E}_{\gamma_n} |f|^{2q}\right)^{1/(2q)} \leq \sqrt{8q^2} \cdot \left(\mathbb{E}_{\gamma_n}\left[\sum_{i=1}^n \left(\sum_{j=1}^n X_j a_{ij}\right)^2\right]^q\right)^{1/(2q)}$$

$$= \sqrt{8q^2} \cdot \left(\mathbb{E}_{\gamma_n}\left[\sum_{j,k=1}^n X_j X_k \left(A^2\right)_{jk}\right]^q\right)^{1/(2q)}$$

$$\leq \sqrt{8q^2 \|A\|} \cdot \left(\mathbb{E}_{\gamma_n} |f|^q\right)^{1/2q} \leq \sqrt{8q^2 \|A\|} \cdot \left(\mathbb{E}_{\gamma_n} |f|^{2q}\right)^{1/4q}.$$

Solving, we obtain the bound

$$\left(\mathbb{E}_{\gamma_n} |f|^{2q}\right)^{1/(2q)} \leq 8q^2 \|A\|.$$

This result gives a suboptimal estimate for large moments of the Gaussian chaos; as $q \to \infty$, the moments should grow in proportion to $q\|A\|$ rather than $q^2\|A\|$. For example, see [27], Corollary 3.9.

7.1. Proof of Theorem 7.1

For a parameter $q = 1$ or $q \geq 1.5$, we wish to estimate the Schatten $2q$-norm of a function $f : \Omega \to \mathbb{H}_d$ that satisfies $\mathbb{E}_\mu f = 0$. The argument has the same structure as Theorem 6.1.

First, we symmetrize. Let $Z, Z' \in \Omega$ be independent random variables, each with distribution μ. Jensen's inequality implies that

$$\mathbb{E}\,\mathrm{tr}\,|f|^{2q} \leq \mathbb{E}\,\mathrm{tr}\,|f(Z) - f(Z')|^{2q} =: \mathbb{E}\,\mathrm{tr}\,|g(Z, Z')|^{2q}. \tag{7.4}$$

Since $|\cdot|^{2q}$ is convex, the function $\mathrm{tr}\,|\cdot|^{2q}$ is also convex [38], Proposition 2.

Define the signed moment function $\varphi(s) := \mathrm{sgn}(s) \cdot |s|^q$, which is odd. Note that its squared derivative $\psi(s) := (\varphi'(s))^2 = q^2 |s|^{2(q-1)}$ is convex. Since $g(z, z') = f(z) - f(z')$ is antisymmetric,

$$\mathbb{E}_{\mu \otimes \mu}\,\mathrm{tr}\,|g|^{2q} = \mathbb{E}_{\mu \otimes \mu}\,\mathrm{tr}\,\varphi(g)^2$$

$$= \mathbb{E}_{\mu \otimes \mu}\,\mathrm{tr}\,\varphi(g)^2 - \mathrm{tr}\left[\left(\mathbb{E}_{\mu \otimes \mu}\,\varphi(g)\right)^2\right] = \mathrm{tr}\,\mathbf{Var}_{\mu \otimes \mu}[\varphi(g)].$$

Apply the bivariate trace Poincaré inequality, Proposition 4.2:

$$\mathrm{tr}\,\mathbf{Var}_{\mu \otimes \mu}[\varphi(g)] \leq \alpha \cdot \mathrm{tr}\,\mathcal{E}(\varphi(g)) \leq \alpha q^2 \cdot \mathbb{E}_{\mu \otimes \mu}\,\mathrm{tr}[\mathbf{\Gamma}(g)\psi(g)].$$

The second bound is the chain rule inequality, Proposition 5.1. In summary,

$$\mathbb{E}_{\mu \otimes \mu}\,\mathrm{tr}\,|g|^{2q} \leq \alpha q^2 \cdot \mathbb{E}_{\mu \otimes \mu}\,\mathrm{tr}\left[\mathbf{\Gamma}(g)|g|^{2(q-1)}\right]. \tag{7.5}$$

This formula allow us to perform a moment comparison.

To isolate the carré du champ $\mathbf{\Gamma}(g)$, invoke Hölder's inequality for the Schatten norms:

$$\mathbb{E}_{\mu \otimes \mu}\,\mathrm{tr}\,|g|^{2q} \leq \alpha q^2 \left(\mathbb{E}_{\mu \otimes \mu}\,\mathrm{tr}\,\mathbf{\Gamma}(g)^q\right)^{1/q} \left(\mathbb{E}_{\mu \otimes \mu}\,\mathrm{tr}\,|g|^{2q}\right)^{(q-1)/q}.$$

Rearrange the last display, and use the initial bound (7.4) to arrive at

$$\left(\mathbb{E}_{\mu\otimes\mu}\operatorname{tr}|\boldsymbol{f}|^{2q}\right)^{1/(2q)} \leq \left(\mathbb{E}_{\mu\otimes\mu}\operatorname{tr}|\boldsymbol{g}|^{2q}\right)^{1/(2q)} \leq \sqrt{\alpha q^2} \cdot \left(\mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\boldsymbol{\Gamma}(\boldsymbol{g})^q\right)^{1/(2q)}.$$

To finish the proof of (7.1), recall the expression (6.6) for the carré du champ $\boldsymbol{\Gamma}(\boldsymbol{g})$. Thus,

$$\left(\mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\boldsymbol{\Gamma}(\boldsymbol{g})^q\right)^{1/q} = \left(\mathbb{E}_{\mu\otimes\mu}\operatorname{tr}\left|\boldsymbol{\Gamma}(\boldsymbol{f})(z)+\boldsymbol{\Gamma}(\boldsymbol{f})(z')\right|^q\right)^{1/q} \leq 2\left(\mathbb{E}_{\mu}\operatorname{tr}\boldsymbol{\Gamma}(\boldsymbol{f})^q\right)^{1/q}.$$

This point follows from the triangle inequality. The argument is complete.

7.2. A variant of the argument

The *intrinsic dimension* of a positive-semidefinite matrix is

$$\operatorname{intdim}(\boldsymbol{A}) := \frac{\operatorname{tr}(\boldsymbol{A})}{\|\boldsymbol{A}\|} \quad \text{for } \boldsymbol{A} \in \mathbb{H}_d^+.$$

We also set $\operatorname{intdim}(\boldsymbol{0}) = 0$. For a nonzero matrix $\boldsymbol{A}$, the intrinsic dimension satisfies $1 \leq \operatorname{intdim}(\boldsymbol{A}) \leq \operatorname{rank}(\boldsymbol{A})$. It can be interpreted as a continuous measure of the rank.

Suppose that q is a natural number. If we use the uniform bound (6.4) for the carré du champ instead of Hölder's inequality, we can apply the bound (7.5) iteratively to obtain

$$\mathbb{E}_{\mu\otimes\mu}\operatorname{tr}|\boldsymbol{g}|^{2q} \leq \operatorname{intdim}\left(\boldsymbol{\mathcal{E}}(\boldsymbol{g})\right) \cdot \alpha^q q! \cdot v_{\boldsymbol{g}}^q.$$

The latter estimate improves over the uniform inequality that follows from Theorem 7.1 and (7.2).

Acknowledgment

Funding

References

[1] Ahlswede, R. and Winter, A. (2002). Strong converse for identification via quantum channels. *IEEE Trans. Inf. Theory* **48** 569–579. MR1889969 https://doi.org/10.1109/18.985947

[2] Aida, S. and Stroock, D. (1994). Moment estimates derived from Poincaré and logarithmic Sobolev inequalities. *Math. Res. Lett.* **1** 75–86. MR1258492 https://doi.org/10.4310/MRL.1994.v1.n1.a9

[3] Aoun, R., Banna, M. and Youssef, P. (2020). Matrix Poincaré inequalities and concentration. *Adv. Math.* **371** 107251, 33. MR4108222 https://doi.org/10.1016/j.aim.2020.107251

[4] Bakry, D., Gentil, I. and Ledoux, M. (2014). *Analysis and Geometry of Markov Diffusion Operators*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **348**. Cham: Springer. MR3155209 https://doi.org/10.1007/978-3-319-00227-9

[5] Bobkov, S. and Ledoux, M. (1997). Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* **107** 383–400. MR1440138 https://doi.org/10.1007/s004400050090

[6] Bobkov, S.G. and Tetali, P. (2006). Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.* **19** 289–336. MR2283379 https://doi.org/10.1007/s10959-006-0016-3

[7] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities*: *A Nonasymptotic Theory of Independence*. Oxford: Oxford Univ. Press. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[8] Buchholz, A. (2001). Operator Khintchine inequality in non-commutative probability. *Math. Ann.* **319** 1–16. MR1812816 https://doi.org/10.1007/PL00004425

[9] Caffarelli, L.A. (2000). Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.* **214** 547–563. MR1800860 https://doi.org/10.1007/s002200000257

[10] Chatterjee, S. (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. MR2288072 https://doi.org/10.1007/s00440-006-0029-y

[11] Chen, R.Y. and Tropp, J.A. (2014). Subadditivity of matrix φ-entropy and concentration of random matrices. *Electron. J. Probab.* **19** no. 27, 30. MR3174839 https://doi.org/10.1214/ejp.v19-2964

[12] Cheng, H.-C. and Hsieh, M.-H. (2016). Characterizations of matrix and operator-valued Φ-entropies, and operator Efron–Stein inequalities. *Proc. R. Soc. A* **472** 20150563, 20. MR3488705 https://doi.org/10.1098/rspa.2015.0563

[13] Cheng, H.-C. and Hsieh, M.-H. (2019). Matrix Poincaré, Φ-Sobolev inequalities, and quantum ensembles. *J. Math. Phys.* **60** 032201, 16. MR3922774 https://doi.org/10.1063/1.5035381

[14] Cheng, H.-C., Hsieh, M.-H. and Tomamichel, M. (2017). Exponential decay of matrix Φ-entropies on Markov semigroups with applications to dynamical evolutions of quantum ensembles. *J. Math. Phys.* **58** 092202, 24. MR3702668 https://doi.org/10.1063/1.5000846

[15] Davies, E.B. (1969). Quantum stochastic processes. *Comm. Math. Phys.* **15** 277–304. MR0266288

[16] Garg, A., Kathuria, T. and Srivastava, N. (2020). Scalar Poincaré implies matrix Poincaré. Preprint. Available at arXiv:2006.09567 [math.PR].

[17] Garg, A., Lee, Y.T., Song, Z. and Srivastava, N. (2018). A matrix expander Chernoff bound. In *STOC'18 – Proceedings of the* 50*th Annual ACM SIGACT Symposium on Theory of Computing* 1102–1114. New York: ACM. MR3826320

[18] Gromov, M. and Milman, V.D. (1983). A topological application of the isoperimetric inequality. *Amer. J. Math.* **105** 843–854. MR0708367 https://doi.org/10.2307/2374298

[19] Gross, L. (1975). Hypercontractivity and logarithmic Sobolev inequalities for the Clifford Dirichlet form. *Duke Math. J.* **42** 383–396. MR0372613

[20] Hansen, F. and Zhang, Z. (2015). Characterisation of matrix entropies. *Lett. Math. Phys.* **105** 1399–1411. MR3395224 https://doi.org/10.1007/s11005-015-0784-8

[21] Hermon, J. and Salez, J. (2019). Modified log-Sobolev inequalities for strong-Rayleigh measures.

[22] Huang, D. and Tropp, J.A. (2021). Nonlinear matrix concentration via semigroup methods. *Electron. J. Probab.* **26** 1–31. https://doi.org/10.1214/20-EJP578

[23] Junge, M. and Zeng, Q. (2015). Noncommutative martingale deviation and Poincaré type inequalities with applications. *Probab. Theory Related Fields* **161** 449–507. MR3334274 https://doi.org/10.1007/s00440-014-0552-1

[24] Kastoryano, M.J. and Temme, K. (2013). Quantum logarithmic Sobolev inequalities and rapid mixing. *J. Math. Phys.* **54** 052202, 30. MR3098923 https://doi.org/10.1063/1.4804995

[25] Ledoux, M. (1995/97). On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Stat.* **1** 63–87. MR1399224 https://doi.org/10.1051/ps:1997103

[26] Ledoux, M. (2001). *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Providence, RI: Amer. Math. Soc. MR1849347 https://doi.org/10.1090/surv/089

[27] Ledoux, M. and Talagrand, M. (2011). *Probability in Banach Spaces*: *Isoperimetry and Processes. Classics in Mathematics*. Berlin: Springer. MR2814399

[28] Lindblad, G. (1976). On the generators of quantum dynamical semigroups. *Comm. Math. Phys.* **48** 119–130. MR0413878

[29] Lust-Piquard, F. (1986). Inégalités de Khintchine dans C_p ($1 < p < \infty$). *C. R. Acad. Sci. Paris Sér. I Math.* **303** 289–292. MR0859804

[30] Mackey, L., Jordan, M.I., Chen, R.Y., Farrell, B. and Tropp, J.A. (2014). Matrix concentration inequalities via the method of exchangeable pairs. *Ann. Probab.* **42** 906–945. MR3189061 https://doi.org/10.1214/13-AOP892

[31] Majewski, A.W., Olkiewicz, R. and Zegarlinski, B. (1998). Dissipative dynamics for quantum spin systems on a lattice. *J. Phys. A* **31** 2045–2056. MR1628657 https://doi.org/10.1088/0305-4470/31/8/015

[32] Münch, F. (2019). Li–Yau inequality under $cd(0, n)$ on graphs. Preprint. Available at arXiv:1909.10242 [math.DG].

[33] Oliveira, R.I. (2009). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges.

[34] Oliveira, R.I. (2010). Sums of random Hermitian matrices and an inequality by Rudelson. *Electron. Commun. Probab.* **15** 203–212. MR2653725 https://doi.org/10.1214/ECP.v15-1544

[35] Olkiewicz, R. and Zegarlinski, B. (1999). Hypercontractivity in noncommutative L_p spaces. *J. Funct. Anal.* **161** 246–285. MR1670230 https://doi.org/10.1006/jfan.1998.3342

[36] Paulin, D., Mackey, L. and Tropp, J.A. (2016). Efron–Stein inequalities for random matrices. *Ann. Probab.* **44** 3431–3473. MR3551202 https://doi.org/10.1214/15-AOP1054

[37] Pemantle, R. and Peres, Y. (2014). Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures. *Combin. Probab. Comput.* **23** 140–160. MR3197973 https://doi.org/10.1017/S0963548313000345

[38] Petz, D. (1994). A survey of certain trace inequalities. In *Functional Analysis and Operator Theory* (*Warsaw*, 1992). *Banach Center Publ.* **30** 287–298. Warsaw: Polish Acad. Sci. Inst. Math. MR1285615

[39] Pisier, G. (1986). Probabilistic methods in the geometry of Banach spaces. In *Probability and Analysis* (*Varenna*, 1985). *Lecture Notes in Math.* **1206** 167–241. Berlin: Springer. MR0864714 https://doi.org/10.1007/BFb0076302

[40] Pisier, G. and Xu, Q. (1997). Non-commutative martingale inequalities. *Comm. Math. Phys.* **189** 667–698. MR1482934 https://doi.org/10.1007/s002200050224

[41] Pitrik, J. and Virosztek, D. (2015). On the joint convexity of the Bregman divergence of matrices. *Lett. Math. Phys.* **105** 675–692. MR3339204 https://doi.org/10.1007/s11005-015-0757-y

[42] Rudelson, M. (1999). Random vectors in the isotropic position. *J. Funct. Anal.* **164** 60–72. MR1694526 https://doi.org/10.1006/jfan.1998.3384

[43] Stroock, D.W. (1984). *An Introduction to the Theory of Large Deviations*. *Universitext*. New York: Springer. MR0755154 https://doi.org/10.1007/978-1-4613-8514-1

[44] Tropp, J.A. (2011). Freedman's inequality for matrix martingales. *Electron. Commun. Probab.* **16** 262–270. MR2802042 https://doi.org/10.1214/ECP.v16-1624

[45] Tropp, J.A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434. MR2946459 https://doi.org/10.1007/s10208-011-9099-z

[46] Tropp, J.A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8** 1–230.

[47] Tropp, J.A. (2016). The expected norm of a sum of independent random matrices: An elementary approach. In *High Dimensional Probability VII. Progress in Probability* **71** 173–202. Cham: Springer. MR3565264 https://doi.org/10.1007/978-3-319-40519-3_8

[48] Van Handel, R. (2016). Probability in High Dimension. APC 550 lecture notes, Princeton Univ.

[49] Varopoulos, N.T. (1985). Hardy–Littlewood theory for semigroups. *J. Funct. Anal.* **63** 240–260. MR0803094 https://doi.org/10.1016/0022-1236(85)90087-4