Electronic Journal of Probability

Electron. J. Probab. **26** (2021), article no. 8, 1–31.
ISSN: 1083-6489 https://doi.org/10.1214/20-EJP578

Nonlinear matrix concentration via semigroup methods

De Huang* Joel A. Tropp†

Abstract

Matrix concentration inequalities provide information about the probability that a random matrix is close to its expectation with respect to the ℓ_2 operator norm. This paper uses semigroup methods to derive sharp nonlinear matrix inequalities. The main result is that the classical Bakry–Émery curvature criterion implies subgaussian concentration for "matrix Lipschitz" functions. This argument circumvents the need to develop a matrix version of the log-Sobolev inequality, a technical obstacle that has blocked previous attempts to derive matrix concentration inequalities in this setting. The approach unifies and extends much of the previous work on matrix concentration. When applied to a product measure, the theory reproduces the matrix Efron–Stein inequalities due to Paulin et al. It also handles matrix-valued functions on a Riemannian manifold with uniformly positive Ricci curvature.

Keywords: Bakry–Émery criterion; concentration inequality; functional inequality; Markov process; matrix concentration; local Poincaré inequality; semigroup.
MSC2020 subject classifications: Primary 60B20; 46N30, Secondary 60J25; 46L53.
Submitted to EJP on June 30, 2020, final version accepted on December 25, 2020.

1 Motivation

Matrix concentration inequalities describe the probability that a random matrix is close to its expectation, with deviations measured in the ℓ_2 operator norm. The basic models—sums of independent random matrices and matrix-valued martingales—have been studied extensively, and they admit a wide spectrum of applications [Tro15]. Nevertheless, we lack a complete understanding of more general random matrix models. The purpose of this paper is to develop a systematic approach for deriving "nonlinear" matrix concentration inequalities.

In the scalar setting, functional inequalities offer a powerful framework for studying nonlinear concentration. For example, consider a real-valued Lipschitz function $f(Z)$ of a real random variable Z with distribution μ. If the measure μ satisfies a Poincaré

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

inequality, then the variance of $f(Z)$ is controlled by the total energy of f, which is further controlled by the squared Lipschitz constant. If the measure satisfies a log-Sobolev inequality, then $f(Z)$ enjoys subgaussian concentration on the scale of the Lipschitz constant.

Now, suppose that we can construct a semigroup, acting on real-valued functions, with stationary distribution μ. Functional inequalities for the measure μ are intimately related to the convergence of the semigroup. In particular, the measure admits a Poincaré inequality if and only if the semigroup rapidly tends to equilibrium (in the sense that the variance is exponentially ergodic). Meanwhile, log-Sobolev inequalities are associated with finer types of ergodicity.

In recent years, researchers have attempted to use functional inequalities and semigroup tools to prove matrix concentration results. So far, these arguments have met some success, but they are not strong enough to reproduce the results that are already available for the simplest random matrix models. The main obstacle has been the lack of a suitable extension of the log-Sobolev inequality to the matrix setting. See Section 3.5 for an account of prior work.

The purpose of this paper is to advance the theory of semigroups acting on matrix-valued functions and to apply these methods to obtain matrix concentration inequalities for nonlinear random matrix models. To do so, we argue that the classical Bakry–Émery curvature criterion for a semigroup acting on real-valued functions ensures that an associated matrix semigroup also satisfies a curvature condition. This property further implies local ergodicity of the matrix semigroup, which we can use to prove strong bounds on the trace moments of nonlinear random matrix models.

The power of this approach is that the Bakry–Émery condition has already been verified for a large number of semigroups. We can exploit these results to identify many new settings where matrix concentration is in force. This program entirely evades the question about the proper way to extend log-Sobolev inequalities to matrices.

We begin with a treatment of Markov semigroups acting on matrix-valued functions (Section 2). Afterward, in Section 3, we present our main results on matrix concentration for nonlinear functions. Section 4 explains how these abstract results lead to many concrete matrix concentration. For instance, our approach easily reproduces the matrix Efron–Stein inequalities [PMT16]. Among other new results, we obtain subgaussian concentration for a matrix-valued "Lipschitz" function on a positively curved Riemannian manifold, such as the sphere or the special orthogonal group. The remaining sections of the paper contain complete proofs.

2 Matrix Markov semigroups: foundations

To start, we develop some basic facts about an important class of Markov semigroups that acts on matrix-valued functions. Given a Markov process, we define the associated matrix Markov semigroup and its infinitesimal generator. Then we construct the matrix carré du champ operator and the Dirichlet form. Afterward, we outline the connection between convergence properties of the semigroup and Poincaré inequalities. Parts of our treatment are adapted from [CHT17, ABY20], but some elements appear to be new.

2.1 Notation

Let $\mathbb{M}_d$ be the algebra of all $d \times d$ complex matrices. The real-linear subspace $\mathbb{H}_d$ contains all Hermitian matrices, and $\mathbb{H}_d^+$ is the cone of all positive-semidefinite matrices. Matrices are written in boldface. In particular, $\mathbf{I}_d$ is the d-dimensional identity matrix, while $\boldsymbol{f}$, $\boldsymbol{g}$ and $\boldsymbol{h}$ refer to matrix-valued functions. We use the symbol $\preccurlyeq$ for the semidefinite partial order on Hermitian matrices: for matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{H}_d$, the inequality $\boldsymbol{A} \preccurlyeq \boldsymbol{B}$ means that $\boldsymbol{B} - \boldsymbol{A} \in \mathbb{H}_d^+$.

For a matrix $\boldsymbol{A} \in \mathbb{M}_d$, we write $\|\boldsymbol{A}\|$ for the ℓ_2 operator norm, $\|\boldsymbol{A}\|_{\mathrm{HS}}$ for the Hilbert–Schmidt norm, and $\operatorname{tr} \boldsymbol{A}$ for the trace. The normalized trace is defined as $\bar{\operatorname{tr}}\, \boldsymbol{A} := d^{-1} \operatorname{tr} \boldsymbol{A}$. Nonlinear functions bind before the trace. Given a scalar function $\varphi : \mathbb{R} \to \mathbb{R}$, we construct the *standard matrix function* $\varphi : \mathbb{H}_d \to \mathbb{H}_d$ using the eigenvalue decomposition:

$$\varphi(\boldsymbol{A}) := \sum_{i=1}^{d} \varphi(\lambda_i)\, \boldsymbol{u}_i \boldsymbol{u}_i^* \quad \text{where} \quad \boldsymbol{A} = \sum_{i=1}^{d} \lambda_i\, \boldsymbol{u}_i \boldsymbol{u}_i^*.$$

We constantly use basic tools from matrix theory; see [Bha97, Car10].

Let Ω be a Polish space equipped with a Borel probability measure μ. Define $\mathbb{E}_\mu$ and Var_μ to be the expectation and variance of a real-valued function with respect to the measure μ. When applied to a random matrix, $\mathbb{E}_\mu$ computes the entrywise expectation. Nonlinear functions bind before the expectation.

2.2 Markov semigroups acting on matrices

This paper focuses on a special class of Markov semigroups acting on matrices. In this model, a classical Markov process drives the evolution of a matrix-valued function. Remark 2.1 mentions some generalizations.

Suppose that $(Z_t)_{t \ge 0} \subset \Omega$ is a time-homogeneous Markov process on the state space Ω with stationary measure μ. For each matrix dimension $d \in \mathbb{N}$, we can construct a Markov semigroup $(P_t)_{t\ge 0}$ that acts on a (bounded) measurable matrix-valued function $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ according to

$$(P_t \boldsymbol{f})(z) := \mathbb{E}[\boldsymbol{f}(Z_t) \,|\, Z_0 = z] \quad \text{for all } t \ge 0 \text{ and all } z \in \Omega. \tag{2.1}$$

The semigroup property $P_{t+s} = P_t P_s = P_s P_t$ holds for all $s, t \ge 0$ because $(Z_t)_{t\ge 0}$ is a homogeneous Markov process.

Note that the operator P_0 is the identity map: $P_0 \boldsymbol{f} = \boldsymbol{f}$. For a fixed $\boldsymbol{A} \in \mathbb{H}_d$, regarded as a constant function on Ω, the semigroup also acts as the identity: $P_t \boldsymbol{A} = \boldsymbol{A}$ for all $t \ge 0$. Furthermore, $\mathbb{E}_\mu[P_t \boldsymbol{f}] = \mathbb{E}_\mu[\boldsymbol{f}]$ because $Z_0 \sim \mu$ implies that $Z_t \sim \mu$ for all $t \ge 0$. We use these facts without comment.

Although (2.1) defines a family of semigroups indexed by the matrix dimension d, we will abuse terminology and speak of this collection as if it were as single semigroup. A major theme of this paper is that facts about the action of the semigroup (2.1) on real-valued functions ($d = 1$) imply parallel facts about the action on matrix-valued functions (all $d \in \mathbb{N}$).

Remark 2.1 (Noncommutative semigroups)**.** There is a very general class of noncommutative semigroups acting on a von Neumann algebra where the action is determined by a family of completely positive unital maps [JZ15]. This framework includes (2.1) as a special case; it covers quantum semigroups [CHT17] acting on $\mathbb{H}_d$ with a fixed matrix dimension d; it also includes more exotic examples. We will not study these models, but we will discuss the relationship between our results and prior work.

2.3 Ergodicity and reversibility

We say that the semigroup $(P_t)_{t\ge 0}$ defined in (2.1) is *ergodic* if

$$P_t f \to \mathbb{E}_\mu f \quad \text{in } L_2(\mu) \quad \text{as} \quad t \to +\infty \quad \text{for all } f : \Omega \to \mathbb{R}.$$

Furthermore, $(P_t)_{t\ge 0}$ is *reversible* if each operator P_t is a *symmetric* operator on $L_2(\mu)$. That is,

$$\mathbb{E}_\mu[(P_t f)\, g] = \mathbb{E}_\mu[f\, (P_t g)] \quad \text{for all } t \ge 0 \text{ and all } f, g : \Omega \to \mathbb{R}. \tag{2.2}$$

Note that these definitions involve only real-valued functions ($d = 1$).

In parallel, we say that the Markov process $(Z_t)_{t\geq 0}$ is reversible (resp. ergodic) if the associated Markov semigroup $(P_t)_{t\geq 0}$ is reversible (resp. ergodic). The reversibility of the process $(Z_t)_{t\geq 0}$ implies that, when $Z_0 \sim \mu$, the pair (Z_t, Z_0) is *exchangeable* for all $t \geq 0$. That is, (Z_t, Z_0) and (Z_0, Z_t) follow the same distribution for all $t \geq 0$.

Our matrix concentration results require ergodicity and reversibility of the semigroup action on matrix-valued functions. These properties are actually a consequence of the analogous properties for real-valued functions. Evidently, the ergodicity of $(P_t)_{t\geq 0}$ is equivalent with the statement

$$P_t \boldsymbol{f} \to \mathbb{E}_\mu \boldsymbol{f} \quad \text{in } L_2(\mu) \quad \text{as} \quad t \to +\infty \quad \text{for all } \boldsymbol{f} : \Omega \to \mathbb{H}_d \text{ and each } d \in \mathbb{N}. \tag{2.3}$$

Note that the $L_2(\mu)$ convergence in the matrix setting means $\lim_{t\to\infty} \mathbb{E}_\mu (P_t \boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f})^2 = \mathbf{0}$, which is readily implied by the $L_2(\mu)$ convergence of all entries of $P_t \boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}$. As for reversibility, we have the following result.

Proposition 2.2 (Reversibility)**.** *Let $(P_t)_{t\geq 0}$ be the family of semigroups defined in* (2.1)*. The following are equivalent.*

1. *The semigroup acting on real-valued functions is symmetric, as in* (2.2)*.*
2. *The semigroup acting on matrix-valued functions is symmetric. That is, for each $d \in \mathbb{N}$,*

$$\mathbb{E}_\mu[(P_t \boldsymbol{f})\, \boldsymbol{g}] = \mathbb{E}_\mu[\boldsymbol{f}\, (P_t \boldsymbol{g})] \quad \textit{for all } t \geq 0 \textit{ and all } \boldsymbol{f}, \boldsymbol{g} : \Omega \to \mathbb{H}_d. \tag{2.4}$$

Let us emphasize that (2.4) now involves matrix products. The proof of Proposition 2.2 appears below in Section 5.2. It is based on the obsevation (5.1) that the product of two matrices can be represented by products of scalars. We are grateful to Ramon van Handel for this insight.

Owing to the same observation, many properties of matrix Markov semigroups follows directly from their counterparts in the scalar case. Section 5.2 contains a detailed discussion.

2.4 Convexity

Given a convex function $\Phi : \mathbb{H}_d \to \mathbb{R}$ that is bounded below, the semigroup satisfies a Jensen inequality of the form

$$\Phi(P_t \boldsymbol{f}(z)) = \Phi(\mathbb{E}[\boldsymbol{f}(Z_t) \,|\, Z_0 = z]) \leq \mathbb{E}[\Phi(\boldsymbol{f}(Z_t)) \,|\, Z_0 = z] \quad \text{for all } z \in \Omega.$$

This is an easy consequence of the definition (2.1). In particular,

$$\mathbb{E}_\mu \Phi(P_t \boldsymbol{f}) \leq \mathbb{E}_{Z\sim\mu} \mathbb{E}[\Phi(\boldsymbol{f}(Z_t)) \,|\, Z_0 = Z] = \mathbb{E}_{Z_0\sim\mu}[\Phi(\boldsymbol{f}(Z_t))] = \mathbb{E}_\mu \Phi(\boldsymbol{f}). \tag{2.5}$$

A typical choice of Φ is the trace function $\operatorname{tr}\varphi$, where $\varphi : \mathbb{H}_d \to \mathbb{H}_d$ is a standard matrix function.

2.5 Infinitesimal generator

The *infinitesimal generator* $\mathscr{L}$ of the semigroup (2.1) acts on a (nice) measurable function $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ via the formula

$$(\mathscr{L}\boldsymbol{f})(z) := \lim_{t\downarrow 0} \frac{(P_t \boldsymbol{f})(z) - \boldsymbol{f}(z)}{t} \quad \text{for all } z \in \Omega. \tag{2.6}$$

Because $(P_t)_{t\geq 0}$ is a semigroup, it follows immediately that

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t = \mathscr{L} P_t = P_t \mathscr{L} \quad \text{for all } t \geq 0.$$

The null space of $\mathscr{L}$ contains all constant functions: $\mathscr{L}\boldsymbol{A} = \mathbf{0}$ for each fixed $\boldsymbol{A} \in \mathbb{H}_d$. Moreover,

$$\mathbb{E}_\mu[\mathscr{L}\boldsymbol{f}] = \mathbf{0} \quad \text{for all } \boldsymbol{f} : \Omega \to \mathbb{H}_d. \tag{2.7}$$

That is, the infinitesimal generator converts an arbitrary function into a zero-mean function.

We say that the infinitesimal generator $\mathscr{L}$ is symmetric on $L_2(\mu)$ when its action on matrix-valued functions is symmetric:

$$\mathbb{E}_\mu[(\mathscr{L}\boldsymbol{f})\,\boldsymbol{g}] = \mathbb{E}_\mu[\boldsymbol{f}\,(\mathscr{L}\boldsymbol{g})] \quad \text{for all } \boldsymbol{f}, \boldsymbol{g} : \Omega \to \mathbb{H}_d. \tag{2.8}$$

The generator $\mathscr{L}$ is symmetric if and only if the semigroup $(P_t)_{t\geq 0}$ is symmetric (i.e., reversible). Owing to Proposition 2.2, symmetry of the infinitesimal generator for real-valued functions ($d = 1$) is equivalent with symmetry for matrix-valued functions (all $d \in \mathbb{N}$).

As we have alluded, the limit in (2.6) need not exist for all functions. The set of functions $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ for which $\mathscr{L}\boldsymbol{f}$ is defined μ-almost everywhere is called the *domain* of the generator. It is highly technical, but usually unimportant, to characterize the domain of the generator and related operators.

For our purposes, we may restrict attention to an unspecified algebra of *suitable* functions (say, smooth and compactly supported) where all operations involving limits, derivatives, and integrals are justified. By approximation, we can extend the main results to the entire class of functions where the statements make sense. We refer the reader to the monograph [BGL13] for an extensive discussion about how to make these arguments airtight.

2.6 Carré du champ operator and Dirichlet form

For each $d \in \mathbb{N}$, given the infinitesimal generator $\mathscr{L}$, the matrix *carré du champ operator* is the bilinear form

$$\Gamma(\boldsymbol{f}, \boldsymbol{g}) := \frac{1}{2}\left[\mathscr{L}(\boldsymbol{f}\boldsymbol{g}) - \boldsymbol{f}\mathscr{L}(\boldsymbol{g}) - \mathscr{L}(\boldsymbol{f})\boldsymbol{g}\right] \in \mathbb{M}_d \quad \text{for all suitable } \boldsymbol{f}, \boldsymbol{g} : \Omega \to \mathbb{H}_d. \tag{2.9}$$

The matrix *Dirichlet form* is the bilinear form obtained by integrating the carré du champ:

$$\mathscr{E}(\boldsymbol{f}, \boldsymbol{g}) := \mathbb{E}_\mu\, \Gamma(\boldsymbol{f}, \boldsymbol{g}) \in \mathbb{M}_d \quad \text{for all suitable } \boldsymbol{f}, \boldsymbol{g} : \Omega \to \mathbb{H}_d. \tag{2.10}$$

We abbreviate the associated quadratic forms as $\Gamma(\boldsymbol{f}) := \Gamma(\boldsymbol{f}, \boldsymbol{f})$ and $\mathscr{E}(\boldsymbol{f}) := \mathscr{E}(\boldsymbol{f}, \boldsymbol{f})$. Proposition 5.1 states that both these quadratic forms are positive operators in the sense that they take values in the cone of positive-semidefinite Hermitian matrices. In many instances, the carré du champ $\Gamma(\boldsymbol{f})$ has a natural interpretation as the squared magnitude of the derivative of $\boldsymbol{f}$, while the Dirichlet form $\mathscr{E}(\boldsymbol{f})$ reflects the total energy of the function $\boldsymbol{f}$.

Using (2.7), we can rewrite the Dirichlet form as

$$\mathscr{E}(\boldsymbol{f}, \boldsymbol{g}) = \mathbb{E}_\mu\, \Gamma(\boldsymbol{f}, \boldsymbol{g}) = -\frac{1}{2}\,\mathbb{E}_\mu\left[\boldsymbol{f}\mathscr{L}(\boldsymbol{g}) + \mathscr{L}(\boldsymbol{f})\boldsymbol{g}\right] \tag{2.11}$$

When the semigroup $(P_t)_{t\geq 0}$ is reversible, then (2.8) and (2.11) indicate that

$$\mathscr{E}(\boldsymbol{f}, \boldsymbol{g}) = -\,\mathbb{E}_\mu[\boldsymbol{f}\mathscr{L}(\boldsymbol{g})] = -\,\mathbb{E}_\mu[\mathscr{L}(\boldsymbol{f})\boldsymbol{g}]. \tag{2.12}$$

These alternative expressions are very useful for calculations.

2.7 The matrix Poincaré inequality

For each function $\boldsymbol{f} : \Omega \to \mathbb{H}_d$, the *matrix variance* with respect to the distribution μ is defined as

$$\mathbf{Var}_\mu[\boldsymbol{f}] := \mathbb{E}_\mu \left[(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f})^2 \right] = \mathbb{E}_\mu[\boldsymbol{f}^2] - (\mathbb{E}_\mu \boldsymbol{f})^2 \in \mathbb{H}_d^+.$$

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

$$\mathbf{Var}_\mu(\boldsymbol{f}) \preccurlyeq \alpha \cdot \mathscr{E}(\boldsymbol{f}) \quad \text{for all suitable } \boldsymbol{f} : \Omega \to \mathbb{H}_d. \tag{2.13}$$

This definition seems to be due to Chen et al. [CHT17]; see also Aoun et al. [ABY20].

When the matrix dimension $d = 1$, the inequality (2.13) reduces to the usual scalar Poincaré inequality for the semigroup. For the semigroup (2.1), the scalar Poincaré inequality ($d = 1$) already implies the matrix Poincaré inequality (for all $d \in \mathbb{N}$). Therefore, to check the validity of (2.13), it suffices to consider real-valued functions.

Proposition 2.3 (Poincaré inequalities: Equivalence)**.** *For each $d \in \mathbb{N}$, let $(P_t)_{t\geq 0}$ be the semigroup defined in* (2.1)*. The following are equivalent:*

1. ***Scalar Poincaré inequality.*** $\mathrm{Var}_\mu[f] \leq \alpha \cdot \mathscr{E}(f)$ *for all suitable* $f : \Omega \to \mathbb{R}$.
2. ***Matrix Poincaré inequality.*** $\mathbf{Var}_\mu[\boldsymbol{f}] \preccurlyeq \alpha \cdot \mathscr{E}(\boldsymbol{f})$ *for all suitable* $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ *and all* $d \in \mathbb{N}$.

The proof of Proposition 2.3 appears in Section 5.2. This result also appears in the companion paper [HT20a], and it has also been obtained in the independent work [GKS20] of Garg et al.

2.8 Poincaré inequalities and ergodicity

As in the scalar case, the matrix Poincaré inequality (2.13) is a powerful tool for understanding the action of a semigroup on matrix-valued functions. Assuming ergodicity, the Poincaré inequality is equivalent with the exponential convergence of the Markov semigroup $(P_t)_{t\geq 0}$ to the expectation operator $\mathbb{E}_\mu$. The constant α determines the rate of convergence. The following result makes this principle precise.

Proposition 2.4 (Poincaré inequality: Consequences)**.** *Consider a Markov semigroup $(P_t)_{t\geq 0}$ with stationary measure μ acting on suitable functions $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ for a fixed $d \in \mathbb{N}$, as defined in* (2.1)*. The following are equivalent:*

1. ***Poincaré inequality.*** $\mathbf{Var}_\mu[\boldsymbol{f}] \preccurlyeq \alpha \cdot \mathscr{E}(\boldsymbol{f})$ *for all suitable* $\boldsymbol{f} : \Omega \to \mathbb{H}_d$.
2. ***Exponential ergodicity of variance.*** $\mathbf{Var}_\mu[P_t\boldsymbol{f}] \preccurlyeq \mathrm{e}^{-2t/\alpha} \cdot \mathbf{Var}_\mu[\boldsymbol{f}]$ *for all* $t \geq 0$ *and for all suitable* $\boldsymbol{f} : \Omega \to \mathbb{H}_d$.

Moreover, if the semigroup $(P_t)_{t\geq 0}$ is reversible and ergodic, then the statements above are also equivalent with the following:

3. ***Exponential ergodicity of energy.*** $\mathscr{E}(P_t\boldsymbol{f}) \preccurlyeq \mathrm{e}^{-2t/\alpha} \cdot \mathscr{E}(\boldsymbol{f})$ *for all* $t \geq 0$ *and for all suitable* $\boldsymbol{f} : \Omega \to \mathbb{H}_d$.

Proposition 2.4 is essentially the same as in the scalar case [vH16, Theorem 2.18]. We elaborate in Section 5.2.

2.9 Iterated carré du champ operator

To better understand how quickly a Markov semigroup converges to equilibrium, it is valuable to consider the *iterated carré du champ operator*. In the matrix setting, this operator is defined as

$$\Gamma_2(\boldsymbol{f}, \boldsymbol{g}) := \frac{1}{2}\left[\mathscr{L}\Gamma(\boldsymbol{f}, \boldsymbol{g}) - \Gamma(\boldsymbol{f}, \mathscr{L}(\boldsymbol{g})) - \Gamma(\mathscr{L}(\boldsymbol{f}), \boldsymbol{g})\right] \in \mathbb{M}_d \quad \text{for all suitable } \boldsymbol{f}, \boldsymbol{g} : \Omega \to \mathbb{H}_d. \tag{2.14}$$

As with the carré du champ, we abbreviate the quadratic form $\Gamma_2(\boldsymbol{f}) := \Gamma_2(\boldsymbol{f}, \boldsymbol{f})$. We remark that this quadratic form is not necessarily a positive operator. Rather, $\Gamma_2(\boldsymbol{f})$ reflects the "magnitude" of the squared Hessian of $\boldsymbol{f}$ plus a correction factor that reflects the "curvature" of the matrix semigroup.

When the underlying Markov semigroup $(P_t)_{t\geq 0}$ is reversible, it holds that

$$\mathbb{E}_\mu \Gamma_2(\boldsymbol{f}, \boldsymbol{g}) = \mathbb{E}_\mu \left[\mathscr{L}(\boldsymbol{f})\,\mathscr{L}(\boldsymbol{g})\right] \quad \text{for all suitable } \boldsymbol{f}, \boldsymbol{g} : \Omega \to \mathbb{H}_d.$$

Thus, for a reversible semigroup, the *average* value $\mathbb{E}_\mu \Gamma_2(\boldsymbol{f})$ is a positive-semidefinite matrix.

2.10 Bakry–Émery criterion

When the iterated carré du champ is comparable with the carré du champ, we can obtain more information about the convergence of the Markov semigroup. We say the semigroup satisfies the *matrix Bakry–Émery criterion* with constant $c > 0$ if

$$\Gamma(\boldsymbol{f}) \preccurlyeq c \cdot \Gamma_2(\boldsymbol{f}) \quad \text{for all suitable } \boldsymbol{f} : \Omega \to \mathbb{H}_d. \tag{2.15}$$

Since $\Gamma(\boldsymbol{f})$ and $\Gamma_2(\boldsymbol{f})$ are functions, one interprets this condition as a pointwise inequality that holds μ-almost everywhere in Ω. It reflects uniform positive curvature of the semigroup.

When the matrix dimension $d = 1$, the condition (2.15) reduces to the classical Bakry–Émery criterion [BGL13, Sec. 1.16]. For a semigroup of the form (2.1), the scalar result actually implies the matrix result for all $d \in \mathbb{N}$.

Proposition 2.5 (Bakry–Émery: Equivalence)**.** *Let $(P_t)_{t\geq 0}$ be the family of semigroups defined in* (2.1)*. The following statements are equivalent:*

1. ***Scalar Bakry–Émery criterion.*** $\Gamma(f) \leq c \cdot \Gamma_2(f)$ *for all suitable* $f : \Omega \to \mathbb{R}$.
2. ***Matrix Bakry–Émery criterion.*** $\Gamma(\boldsymbol{f}) \preccurlyeq c \cdot \Gamma_2(\boldsymbol{f})$ *for all suitable* $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ *and all* $d \in \mathbb{N}$.

See Section 5.2 for the proof of Proposition 2.5.

Proposition 2.5 is a very powerful tool, and it is a key part of our method. Indeed, it is already known [BGL13] that many kinds of Markov processes satisfy the scalar Bakry–Émery criterion (1). When contemplating novel settings, we only need to check the scalar criterion, rather than worrying about matrix-valued functions. In all these cases, we obtain the matrix extension for free.

Remark 2.6 (Curvature)**.** The scalar Bakry–Émery criterion, Proposition 2.5(1), is also known as the curvature condition $CD(\rho, \infty)$ with $\rho = c^{-1}$. In the scenario where the infinitesimal generator $\mathscr{L}$ is the Laplace–Beltrami operator $\Delta_{\mathfrak{g}}$ on a Riemannian manifold $(M, \mathfrak{g})$ with co-metric $\mathfrak{g}$, the Bakry–Émery criterion holds if and only if the Ricci curvature tensor is everywhere positive definite, with eigenvalues bounded from below by $\rho > 0$. See [BGL13, Section 1.16] for a discussion. We return to this example in Section 4.3.

2.11 Bakry–Émery and ergodicity

The scalar Bakry–Émery criterion, Proposition 2.5(1), is equivalent with a local Poincaré inequality, which is strictly stronger than the scalar Poincaré inequality, Proposition 2.3(1). It is also equivalent with a powerful local ergodicity property [vH16, Theorem 2.35]. The next result states that the matrix Bakry–Émery criterion (2.15) implies counterparts of these facts.

Proposition 2.7 (Bakry–Émery: Consequences)**.** *Let $(P_t)_{t\geq 0}$ be a Markov semigroup acting on suitable functions $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ for fixed $d \in \mathbb{N}$, as defined in* (2.1). *The following are equivalent:*

1. ***Bakry–Émery criterion.*** $\Gamma(\boldsymbol{f}) \preccurlyeq c \cdot \Gamma_2(\boldsymbol{f})$ *for all suitable $\boldsymbol{f} : \Omega \to \mathbb{H}_d$.*
2. ***Local ergodicity.*** $\Gamma(P_t\boldsymbol{f}) \preccurlyeq \mathrm{e}^{-2t/c} \cdot P_t\Gamma(\boldsymbol{f})$ *for all $t \geq 0$ and for all suitable $\boldsymbol{f} : \Omega \to \mathbb{H}_d$.*
3. ***Local Poincaré inequality.*** $P_t(\boldsymbol{f}^2) - (P_t\boldsymbol{f})^2 \preccurlyeq c\,(1 - \mathrm{e}^{-2t/c}) \cdot P_t\Gamma(\boldsymbol{f})$ *for all $t \geq 0$ and for all suitable $\boldsymbol{f} : \Omega \to \mathbb{H}_d$.*

The proof Proposition 2.7 follows along the same lines as the scalar result [vH16, Theorem 2.36]. See Section 5.2 for a discussion.

Proposition 2.7 plays a central role in this paper. With the aid of Proposition 2.5, we can verify the Bakry–Émery criterion (1) for many particular Markov semigroups. Meanwhile, the local ergodicity property (2) supports short derivations of trace moment inequalities for random matrices.

The results in Proposition 2.7 refine the statements in Proposition 2.4. Indeed, the carré du champ operator $\Gamma(\boldsymbol{f})$ measures the local fluctuation of a function $\boldsymbol{f}$, so the local ergodicity condition (2) means that the fluctuation of $P_t\boldsymbol{f}$ at every point $z \in \Omega$ is decreasing exponentially fast. By applying $\mathbb{E}_\mu$ to both sides of the local ergodicity inequality, we obtain the ergodicity of energy, Proposition 2.4(3).

If $(P_t)_{t\geq 0}$ is ergodic, applying the expectation $\mathbb{E}_\mu$ to the local Poincaré inequality (3) and then taking $t \to +\infty$ yields the matrix Poincaré inequality, Proposition 2.4(1) with constant $\alpha = c$. In fact, a standard method for establishing a Poincaré inequality is to check the Bakry–Émery criterion.

3 Nonlinear matrix concentration: main results

The matrix Poincaré inequality (2.13) has been associated with subexponential concentration inequalities for random matrices [ABY20, HT20a]. The central purpose of this paper is to establish that the (scalar) Bakry–Émery criterion leads to matrix concentration inequalities via a straightforward semigroup argument. This section outlines our main results; the proofs appear in Sections 6 and 7. The key technical ingredient is a novel chain rule inequality for the carré du champ operator.

3.1 Markov processes and random matrices

Let Z be a random variable, taking values in the state space Ω, with the distribution μ. For a matrix-valued function $\boldsymbol{f} : \Omega \to \mathbb{H}_d$, we can define the random matrix $\boldsymbol{f}(Z)$, whose distribution is the push-forward of μ by the function $\boldsymbol{f}$. Our goal is to understand how well the random matrix $\boldsymbol{f}(Z)$ concentrates around its expectation $\mathbb{E}\,\boldsymbol{f}(Z) = \mathbb{E}_\mu\,\boldsymbol{f}$.

To do so, suppose that we can construct a reversible, ergodic Markov process $(Z_t)_{t\geq 0} \subset \Omega$ whose stationary distribution is μ. We have the intuition that the faster that the process $(Z_t)_{t\geq 0}$ converges to equilibrium, the more sharply the random matrix $\boldsymbol{f}(Z)$ concentrates around its expectation.

To quantify the rate of convergence of the matrix Markov process, we use the Bakry–Émery criterion (2.15) to obtain local ergodicity of the semigroup. This property allows us to prove strong bounds on the trace moments of the random matrix. Using standard arguments (e.g., see [PMT16, Section 3]), these moment bounds imply nonlinear matrix concentration inequalities.

3.2 Polynomial concentration

We first introduce a general estimate on the polynomial trace moments of a random matrix under a scalar Bakry–Émery criterion.

Theorem 3.1 (Polynomial moments)**.** *Let Ω be a Polish space equipped with a Borel probability measure μ. Consider a reversible, ergodic Markov semigroup* (2.1) *with stationary measure μ that acts on (suitable) functions $\boldsymbol{f} : \Omega \to \mathbb{H}_d$. Assume that the semigroup satisfies the scalar Bakry–Émery criterion* (2.15) *with a constant $c > 0$ and with $d = 1$. Then, for $q = 1$ and $q \geq 1.5$,*

$$\left[\mathbb{E}_\mu \operatorname{tr} |\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}|^{2q}\right]^{1/(2q)} \leq \sqrt{c\,(2q-1)}\left[\mathbb{E}_\mu \operatorname{tr} \Gamma(\boldsymbol{f})^q\right]^{1/(2q)}. \tag{3.1}$$

In addition, if the variance proxy $v_{\boldsymbol{f}} := \| \|\Gamma(\boldsymbol{f})\| \|_{L_\infty(\mu)} < +\infty$, then

$$\left[\mathbb{E}_\mu \operatorname{tr} |\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}|^{2q}\right]^{1/(2q)} \leq d^{1/(2q)} \sqrt{c\,(2q-1)\, v_{\boldsymbol{f}}}. \tag{3.2}$$

We establish this theorem in Section 7.

With the polynomial moment bounds in Theorem 3.1, we can derive bounds for the expectation and tails of $\|\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}\|$ using the matrix Chebyshev inequality; for example, see [MJC$^+$14, Proposition 6.2]. In particular, when $v_{\boldsymbol{f}} < +\infty$, we obtain subgaussian concentration.

For noncommutative diffusion semigroups, Junge & Zeng [JZ15] have developed polynomial moment bounds similar to Theorem 3.1, but they only obtain moment growth of $O(q)$ in the inequality (3.1). We can trace this discrepancy to the fact that they use a martingale argument based on the noncommutative Burkholder–Davis–Gundy inequality. At present, our proof only applies to the classical semigroup (2.1), but it seems plausible that our approach can be generalized.

3.3 Exponential concentration

As a consequence of the Bakry–Émery criterion (2.15), we can also derive exponential matrix concentration inequalities. In principle, polynomial moment inequalities are stronger, but the exponential inequalities often lead to better constants and more detailed information about tail decay.

Theorem 3.2 (Exponential concentration)**.** *Let Ω be a Polish space equipped with a Borel probability measure μ. Consider a reversible, ergodic Markov semigroup* (2.1) *with stationary measure μ that acts on (suitable) functions $\boldsymbol{f} : \Omega \to \mathbb{H}_d$. Assume that the semigroup satisfies the scalar Bakry–Émery criterion* (2.15) *with a constant $c > 0$ and with $d = 1$. Then*

$$\mathbb{P}_\mu \left\{\lambda_{\max}(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}) \geq t\right\} \leq d \cdot \inf_{\beta > 0} \exp\left(\frac{-t^2}{2c r_{\boldsymbol{f}}(\beta) + 2t\sqrt{c/\beta}}\right) \quad \text{for all } t \geq 0.$$

The function $r_{\boldsymbol{f}}$ computes an exponential mean of the carré du champ:

$$r_{\boldsymbol{f}}(\beta) := \frac{1}{\beta} \log \mathbb{E}_\mu \operatorname{\bar{tr}} \mathrm{e}^{\beta \Gamma(\boldsymbol{f})} \quad \text{for } \beta > 0.$$

In addition, if the variance proxy $v_{\boldsymbol{f}} := \| \|\Gamma(\boldsymbol{f})\| \|_{L_\infty(\mu)} < +\infty$*, then*

$$\mathbb{P}_\mu \left\{ \lambda_{\max}(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}) \geq t \right\} \leq d \cdot \exp\left(\frac{-t^2}{2cv_{\boldsymbol{f}}} \right) \quad \textit{for all } t \geq 0.$$

Furthermore,

$$\mathbb{E}_\mu \lambda_{\max}(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}) \leq \sqrt{2cv_{\boldsymbol{f}} \log d}.$$

Parallel inequalities hold for the minimum eigenvalue $\lambda_{\min}$.

We establish Theorem 3.2 in Section 7.4 as a consequence of an exponential moment inequality, Theorem 7.2, for random matrices.

A partial version of Theorem 3.2 with slightly worse constants appears in [JZ15, Corollary 4.13]. When comparing these results, note that probability measure in [JZ15] is normalized to absorb the dimensional factor d.

3.4 Extension to general rectangular matrices

By a standard formal argument, we can extend the results in this section to a function $\boldsymbol{h} : \Omega \to \mathbb{M}^{d_1 \times d_2}$ that takes rectangular matrix values. To do so, we simply apply the theorems to the self-adjoint dilation

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

See [Tro15] for many examples of this methodology.

3.5 History

Matrix concentration inequalities are noncommutative extensions of their scalar counterparts. They have been studied extensively, and they have had a profound impact on a wide range of areas in computational mathematics and statistics. The models for which the most complete results are available include a sum of independent random matrices [LP86, Rud99, Oli10, Tro12, Hua19] and a matrix-valued martingale sequence [PX97, Oli09, Tro11, JZ15, HRMS18]. We refer to the monograph [Tro15] for an introduction and an extensive bibliography. Very recently, some concentration results for products of random matrices have also been established [HW20, HNWTW20].

In recent years, many authors have sought concentration results for more general random matrix models. One natural idea is to develop matrix versions of scalar concentration techniques based on functional inequalities or based on Markov processes.

In the scalar setting, the subadditivity of the entropy plays a basic role in obtaining modified log-Sobolev inequalities for product spaces, a core ingredient in proving subgaussian concentration results. Chen and Tropp [CT14] established the subadditivity of matrix trace entropy quantities. Unfortunately, the approach in [CT14] requires awkward additional assumptions to derive matrix concentration from modified log-Sobolev inequalities. Cheng et al. [CH16, CHT17, CH19] have extended this line of research.

Mackey et al. [MJC+14, PMT16] observed that the method of exchangeable pairs [Ste72, Ste86, Cha05] leads to more satisfactory matrix concentration inequalities, including matrix generalizations of the Efron–Stein–Steele inequality. The argument in [PMT16] can be viewed as a discrete version of the semigroup approach that we use in this paper; see Section 7.5 for more discussion.

Very recently, Aoun et al. [ABY20] showed how to derive exponential matrix concentration inequalities from the matrix Poincaré inequality (2.13). Their approach is based on the classical iterative argument, due to Aida & Stroock [AS94], that operates in the scalar setting. For matrices, it takes serious effort to implement this technique. In our companion paper [HT20a], we have shown that a trace Poincaré inequality leads to stronger exponential concentration results via an easier argument.

Another appealing contribution of the paper [ABY20] is to establish the validity of a matrix Poincaré inequality for particular matrix-valued Markov processes. Unfortunately, Poincaré inequalities are apparently not strong enough to capture subgaussian concentration. In the scalar case, log-Sobolev inequalities lead to subgaussian concentration inequalities. At present, it is not clear how to extend the theory of log-Sobolev inequalities to matrices, and this obstacle has delayed progress on studying matrix concentration via functional inequalities.

In the scalar setting, one common technique for establishing a log-Sobolev inequality is to prove that the Bakry–Émery criterion holds [vH16, Problem 3.19]. Inspired by this observation, we have chosen to investigate the implications of the Bakry–Émery criterion (2.15) for Markov semigroups acting on matrix-valued functions. Our work demonstrates that this type of curvature condition allows us to establish matrix moment bounds directly, without the intermediation of a log-Sobolev inequality. As a consequence, we can obtain subgaussian and subgamma concentration for nonlinear random matrix models.

After establishing the results in this paper, we discovered that Junge & Zeng [JZ15] have also derived subgaussian matrix concentration inequalities from the (noncommutative) Bakry–Émery criterion. Their approach is based on a noncommutative version of the Burkholder–Davis–Gundy inequality and a martingale argument that applies to a wider class of noncommutative diffusion semigroups acting on von Neumann algebras. As a consequence, their results apply to a larger family of examples, but the moment growth bounds are somewhat worse.

In contrast, our paper develops a direct argument for the classical semigroup (2.1) that does not require any sophisticated tools from operator theory or noncommutative probability. Instead, we establish a new trace inequality (Lemma 6.1) that mimics the chain rule for a scalar diffusion semigroup.

4 Examples

Before we prove the main theorems, we first motivate the reader with some immediate applications of our results. This section contains some examples of Markov semigroups that satisfy the Bakry–Émery criterion (2.15). We will use these semigroups to derive matrix concentration results for several random matrix models.

4.1 Product measures

Consider a product space $\Omega = \Omega_1 \otimes \Omega_2 \otimes \cdots \otimes \Omega_n$ equipped with a product measure $\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$. We can construct a Markov process $(Z_t)_{t\geq 0} = (Z_t^1, Z_t^2, \dots, Z_t^n)_{t\geq 0}$ on Ω whose stationary measure is μ. Let $(N_t^i)_{i=1}^n$ be an independent family of standard Poisson processes on the positive real line. At each time t where N_t^i increases for some i, we replace the value of Z_t^i in Z_t by an independent sample from μ_i while keeping the remaining coordinates fixed. It is known that this Markov process is ergodic and reversible. We refer the reader to [vH16, ABY20] for a comprehensive introduction to this Markov process.

The infinitesimal generator $\mathscr{L}$ of the associated Markov semigroup admits the explicit form

$$\mathscr{L}\boldsymbol{f} = -\sum_{i=1}^{n} \delta_i \boldsymbol{f}. \tag{4.1}$$

The difference operator δ_i is given by

$$\delta_i \boldsymbol{f}(z) := \boldsymbol{f}(z) - \mathbb{E}_Z\, \boldsymbol{f}\big((z;Z)_i\big) \quad \text{for all } z \in \Omega,$$

where $(z;Z)_i = (z^1, \dots, z^{i-1}, Z^i, z^{i+1}, \dots, z^n)$. The superscript stands for the index of the coordinate. This infinitesimal generator $\mathscr{L}$ is well defined for all integrable functions, so the class of suitable functions contains $L_1(\mu)$.

Using the expression (4.1) for the infinitesimal generator $\mathscr{L}$, we can compute the (quadratic) matrix carré du champ operator Γ and the (quadratic) iterated matrix carré du champ operator Γ_2 of the semigroup as follows:

$$\Gamma(\boldsymbol{f})(z) = \boldsymbol{V}(\boldsymbol{f})(z) := \frac{1}{2}\sum_{i=1}^n \mathbb{E}_Z\left[\big(\boldsymbol{f}(z) - \boldsymbol{f}((z;Z)_i)\big)^2\right] \tag{4.2}$$

and

$$\begin{aligned}\Gamma_2(\boldsymbol{f})(z) = \frac{1}{4}\sum_{i=1}^n \Big\{ &\mathbb{E}_Z\left[\big(\boldsymbol{f}(z) - \boldsymbol{f}((z;Z)_i)\big)^2\right] + \mathbb{E}_{\tilde{Z}}\,\mathbb{E}_Z\left[\big(\boldsymbol{f}((z;\tilde{Z})_i) - \boldsymbol{f}((z;Z)_i)\big)^2\right]\Big\} \\ &+ \frac{1}{4}\sum_{i\neq j} \mathbb{E}_{\tilde{Z}}\,\mathbb{E}_Z\left[\big(\boldsymbol{f}(z) - \boldsymbol{f}((z;\tilde{Z})_i) - \boldsymbol{f}((z;Z)_j) + \boldsymbol{f}(((z;\tilde{Z})_i;Z)_j)\big)^2\right]. \end{aligned} \tag{4.3}$$

Here $\mathbb{E}_Z := \mathbb{E}_{Z\sim\mu}$, and the random variables Z and $\tilde{Z}$ are independent draws from the measure μ. These expressions are valid for all suitable $\boldsymbol{f}, \boldsymbol{g} : \Omega \to \mathbb{H}_d$ and all $z \in \Omega$.

It is clear that the formula (4.2) for Γ appears within the formula (4.3) for Γ_2. We immediately conclude that the Bakry–Émery criterion holds.

Theorem 4.1 (Product measure: Bakry–Émery)**.** *For the semigroup characterized by the infinitesimal generator* (4.1)*, the scalar Bakry–Émery criterion* (2.15) *holds with* $c = 2$*. That is, for any suitable function* $f : \Omega \to \mathbb{R}$*,*

$$\Gamma(f) \le 2\,\Gamma_2(f).$$

Owing to Proposition 2.5, the semigroup also satisfies the matrix Bakry–Émery criterion for all $d \in \mathbb{N}$*.*

After completing this paper, we learned that Theorem 4.1 appears in [JZ15, Example 6.6] with a different style of proof.

Remark 4.2 (Matrix Poincaré inequality: Constants)**.** Following the discussion in Section 2.10, Theorem 4.1 implies the matrix Poincaré inequality (2.13) with $\alpha = 2$. However, Aoun et al. [ABY20] proved that the Markov process characterized by the infinitesimal generator (4.1) actually satisfies the matrix Poincaré inequality with $\alpha = 1$; see also [CH16, Theorem 5.1]. This gap is not surprising because the averaging operation that is missing in the local Poincaré inequality contributes to the global convergence of the Markov semigroup.

Theorem 4.1 shows that there is a reversible ergodic Markov semigroup whose stationary measure is $\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$ and which satisfies the Bakry–Émery criterion (2.15) with constant $c = 2$. Therefore, we can apply Theorem 3.1 with $c = 2$ to obtain polynomial moment bounds for product measures.

Corollary 4.3 (Product measure: Polynomial moments)**.** *Let* $\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$ *be a product measure on a product space* $\Omega = \Omega_1 \otimes \Omega_2 \otimes \cdots \otimes \Omega_n$*. Let* $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ *be a suitable function. Then, for* $q = 1$ *and* $q \ge 1.5$*,*

$$\left[\mathbb{E}_\mu \operatorname{tr} |\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}|^{2q}\right]^{1/(2q)} \le \sqrt{2(2q-1)}\left[\mathbb{E}_\mu \operatorname{tr} \boldsymbol{V}(\boldsymbol{f})^q\right]^{1/(2q)}.$$

The matrix variance proxy $\boldsymbol{V}(\boldsymbol{f})$ *is defined in* (4.2)*.*

Corollary 4.3 exactly reproduces the matrix polynomial Efron–Stein inequalities established by Paulin et al. [PMT16, Theorem 4.2].

We can also reproduce the matrix exponential Efron–Stein inequalities of Paulin et al. [PMT16, Theorem 4.3] by applying Theorem 3.2 to a product measure. For instance, we obtain the following subgaussian inequality.

Corollary 4.4 (Product measure: Subgaussian concentration)**.** *Let $\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$ be a product measure on a product space $\Omega = \Omega_1 \otimes \Omega_2 \otimes \cdots \otimes \Omega_n$. Let $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ be a suitable function. Define the variance proxy $v_{\boldsymbol{f}} := \| \| \boldsymbol{V}(\boldsymbol{f}) \| \|_{L_\infty(\mu)}$, where $\boldsymbol{V}(\boldsymbol{f})$ is given by* (4.2)*. Then*

$$\mathbb{P}\left\{\lambda_{\max}(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}) \geq t\right\} \leq d \cdot \exp\left(-\frac{t^2}{4v_{\boldsymbol{f}}}\right) \quad \textit{for all } t \geq 0.$$

Furthermore,

$$\mathbb{E}_\mu \lambda_{\max}(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}) \leq 2\sqrt{v_{\boldsymbol{f}} \log d}.$$

Parallel results hold for the minimum eigenvalue $\lambda_{\min}$.

4.2 Log-concave measures

Log-concave distributions [Pré73, ASZ09, SW14] are a fundamental class of probability measures on $\Omega = \mathbb{R}^n$ that are closely related to diffusion processes; the most important example in this class is the standard Gaussian measure. A log-concave measure takes the form $\mathrm{d}\mu = \rho^\infty(z)\,\mathrm{d}z \propto \mathrm{e}^{-W(z)}\,\mathrm{d}z$ where the potential $W : \mathbb{R}^n \to \mathbb{R}$ is a smooth convex function, so it captures a form of negative dependence. The associated diffusion process naturally induces a reversible semigroup whose carré du champ operator takes the form of the squared “magnitude” of the gradient:

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

As usual, $\partial_i := \partial/\partial z_i$ for $i = 1, \ldots, n$.

Many interesting results follow from the condition that the potential W is uniformly strongly convex on $\mathbb{R}^n$. In other words, for a constant $\eta > 0$, we assume that the Hessian matrix satisfies

$$(\operatorname{Hess} W)(z) := \left[\partial_{ij} W(z)\right]_{i,j=1}^n \succcurlyeq \eta \cdot \mathbf{I}_n \quad \text{for all } z \in \mathbb{R}^n. \tag{4.4}$$

In fact, the uniform strong convexity of W implies the ergodicity of the associated Markov semigroup in the sense of (2.3). One can find more detailed ergodicity results in [Hai16, JSY19].

Moreover, it is a standard result [BGL13, Sec. 4.8] that the strong convexity condition (4.4) implies the scalar Bakry–Émery criterion with constant $c = \eta^{-1}$. Therefore, according to Proposition 2.5, the matrix Bakry–Émery criterion (2.15) is valid for every $d \in \mathbb{N}$.

This discussion implies that we can apply our main results to a log-concave probability measure $\mathrm{d}\mu \propto \mathrm{e}^{-W(z)}\,\mathrm{d}z$ on $\mathbb{R}^n$ whose potential W satisfies the strong convexity condition (4.4). For instance, we can apply Theorem 3.1 with $c = \eta^{-1}$ to obtain polynomial moment bounds for μ.

Corollary 4.5 (Log-concave measure: Polynomial moments)**.** *Let $\mathrm{d}\mu \propto \mathrm{e}^{-W(z)}\,\mathrm{d}z$ be a log-concave measure on $\mathbb{R}^n$ whose potential $W : \mathbb{R}^n \to \mathbb{R}$ satisfies a uniform strong convexity condition: $\operatorname{Hess} W \succcurlyeq \eta \cdot \mathbf{I}_n$ with constant $\eta > 0$. Let $\boldsymbol{f} \in \mathbb{R}^n \mapsto \mathbb{H}_d$ be a suitable function. Then, for $q = 1$ and $q \geq 1.5$,*

$$\left[\mathbb{E}_\mu \operatorname{tr} |\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}|^{2q}\right]^{1/(2q)} \leq \sqrt{\frac{2q-1}{\eta}} \left[\mathbb{E}_\mu \operatorname{tr} \left(\sum_{i=1}^n (\partial_i \boldsymbol{f})^2\right)^q\right]^{1/(2q)}.$$

Similarly, we can apply Theorem 3.2 with $c = \eta^{-1}$ to obtain subgaussian concentration inequalities.

Corollary 4.6 (Log-concave measure: Subgaussian concentration). *Let* $\mathrm{d}\mu \propto \mathrm{e}^{-W(z)}\,\mathrm{d}z$ *be a log-concave probability measure on* $\mathbb{R}^n$ *whose potential* $W : \mathbb{R}^n \to \mathbb{R}$ *satisfies a uniform strong convexity condition:* $\operatorname{Hess} W \succcurlyeq \eta \cdot \mathbf{I}_n$ *where* $\eta > 0$*. Let* $\boldsymbol{f} \in \mathbb{R}^n \mapsto \mathbb{H}_d$ *be a suitable function, and define the variance proxy*

$$v_{\boldsymbol{f}} := \sup\nolimits_{z \in \mathbb{R}^n} \left\| \sum_{i=1}^n (\partial_i \boldsymbol{f}(z))^2 \right\|.$$

Then

$$\mathbb{P}_\mu \left\{ \lambda_{\max}(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}) \geq t \right\} \leq d \cdot \exp\left(\frac{-\eta t^2}{2 v_{\boldsymbol{f}}} \right) \quad \textit{for all } t \geq 0.$$

Furthermore,

$$\mathbb{E}_\mu \lambda_{\max}(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f}) \leq \sqrt{2\eta^{-1} v_{\boldsymbol{f}} \log d}.$$

Parallel results hold for the minimum eigenvalue $\lambda_{\min}$*.*

4.2.1 Standard normal distribution

One of the core examples of a log-concave measure is the standard Gaussian measure γ_n on $\mathbb{R}^n$. Its potential, $W(z) = z^{\mathsf{T}} z/2$, is uniformly strongly convex with parameter $\eta = 1$. The associated diffusion process induces the Ornstein–Uhlenbeck semigroup, which satisfies the Bakry–Émery criterion (2.15) with constant $c = 1$. Therefore, Corollary 4.6 gives subgaussian concentration for matrix-valued functions of a Gaussian random vector. To make a comparison with familiar results, we present one of the basic examples.

Example 4.7 (Matrix Gaussian series). Consider the matrix Gaussian series

$$\boldsymbol{f}(z) = \sum_{i=1}^n Z_i \boldsymbol{A}_i \quad \text{where } z = (Z_1, \dots, Z_n) \sim \gamma_n \text{ and } \boldsymbol{A}_i \in \mathbb{H}_d \text{ are fixed.}$$

In this case, the carré du champ is simply

$$\Gamma(\boldsymbol{f})(z) = \sum_{i=1}^n \boldsymbol{A}_i^2.$$

Thus, the expectation bound states that

$$\mathbb{E}_{\gamma_n} \lambda_{\max}(\boldsymbol{f}(z)) \leq \sqrt{2 v_{\boldsymbol{f}} \log d} \quad \text{where} \quad v_{\boldsymbol{f}} = \left\| \sum_{i=1}^n \boldsymbol{A}_i^2 \right\|.$$

Up to and including the constants, this matches the sharp bound that follows from "linear" matrix concentration techniques [Tro15, Chapter 4].

Van Handel (private communication) has outlined out an alternative proof of Corollary 4.6 with slightly worse constants. His approach uses Pisier's method [Pis86, Thm. 2.2] and the noncommutative Khintchine inequality [Buc01] to obtain the statement for the standard normal measure. Then Caffarelli's contraction theorem [Caf00] implies that the same bound holds for every log-concave measure whose potential is uniformly strongly convex with $\eta \geq 1$. This approach is short and conceptual, but it is more limited in scope.

4.3 Measures on Riemannian manifolds

The theory of diffusion processes on Euclidean spaces can be generalized to the setting of Riemannian manifolds. Although this exercise may seem abstract, it allows us to treat some new examples in a unified way. We refer to [BGL13] for a comprehensive treatment of the subject, and we instate their conventions. For an introduction to calculus on Riemannian manifolds, references include [Pet16, Lee18].

Consider an n-dimensional compact Riemannian manifold $(M, \mathfrak{g})$. Let $\mathfrak{g}(x) = (g^{ij}(x) : 1 \le i, j \le n)$ be the matrix representation of the co-metric tensor $\mathfrak{g}$ in local coordinates, which is a symmetric and positive-definite matrix defined for every $x \in M$. The manifold is equipped with a canonical Riemannian probability measure $\mu_{\mathfrak{g}}$ that has local density $\mathrm{d}\mu_{\mathfrak{g}} \propto \det(\mathfrak{g}(x))^{-1/2}\,\mathrm{d}x$ with respect to the Lebesgue measure in local coordinates. This measure $\mu_{\mathfrak{g}}$ is the stationary measure of the diffusion process on M whose infinitesimal generator $\mathscr{L}$ is the Laplace–Beltrami operator $\Delta_{\mathfrak{g}}$. This diffusion process is called the *Riemannian Brownian motion*.[1] Note that the Laplace–Beltrami operator $\Delta_{\mathfrak{g}}$ is symmetric with respect to the measure $\mu_{\mathfrak{g}}$, and hence the diffusion process is reversible.

The associated matrix carré du champ operator coincides with the squared "magnitude" of the differential:

$$\Gamma(\boldsymbol{f})(x) = \sum_{i,j=1}^{n} g^{ij}(x)\,\partial_i \boldsymbol{f}(x)\,\partial_j \boldsymbol{f}(x) \quad \text{for suitable } \boldsymbol{f} : M \to \mathbb{H}_d. \tag{4.5}$$

Here, ∂_i for $i = 1, \dots, n$ are the components of the differential, computed in local coordinates. We emphasize that the matrix carré du champ operator is intrinsic; expressions for the carré du champ resulting from different choices of local coordinates are equivalent under change of variables. In particular, if we compute the partial derivatives ∂_i in local geodesic/normal coordinates, the corresponding co-metric tensor $\mathfrak{g}$ is the identity matrix, and thus the carré du champ operator has the simple form $\Gamma(\boldsymbol{f}) = \sum_{i=1}^{n}(\partial_i \boldsymbol{f})^2$.

As mentioned in Remark 2.6, the scalar Bakry–Émery criterion holds with $c = \rho^{-1}$ if and only if the Ricci curvature tensor of $(M, \mathfrak{g})$ is everywhere positive, with eigenvalues bounded from below by $\rho > 0$. In other words, for Brownian motion on a manifold, the Bakry–Émery criterion is equivalent with the uniform positive curvature of the manifold. Proposition 2.5 ensures that the matrix Bakry–Émery criterion (2.15) holds with $c = \rho^{-1}$ under precisely the same circumstances.

We remark that the uniform positiveness of the Ricci curvature tensor also leads to a Poincaré inequality for the diffusion process on the manifold; see [BGL13, Section 4.8]. Therefore, Proposition 2.4 implies that the associated Markov semigroup is ergodic in the sense of (2.3).

Many examples of positively curved Riemannian manifolds are discussed in [Led01, Gro07, CE08, BGL13]. We highlight two particularly interesting cases.

4.3.1 The sphere

Consider the n-dimensional unit sphere $\mathbb{S}^n \subset \mathbb{R}^{n+1}$, equipped with the Riemannian manifold structure induced by $\mathbb{R}^{n+1}$. The canonical Riemannian measure on the sphere is simply the uniform probability measure. The sphere has a constant Ricci curvature tensor, whose eigenvalues all equal $n-1$; see [BGL13, Section 2.2]. Therefore, the Brownian motion on $\mathbb{S}^n$ satisfies a Bakry–Émery criterion (2.15) with $c = (n-1)^{-1}$ for

[1] Many authors use an alternative convention that Riemmanian Brownian motion has infinitesimal generator $\frac{1}{2}\Delta_{\mathfrak{g}}$.

$n \geq 2$. Theorem 3.1 then implies that, for any suitable function $\boldsymbol{f} : \mathbb{S}^n \to \mathbb{H}_d$,

$$\left[\mathbb{E}_{\sigma_n} \operatorname{tr} |\boldsymbol{f} - \mathbb{E}_{\sigma_n} \boldsymbol{f}|^{2q}\right]^{1/(2q)} \leq \sqrt{\frac{2q-1}{n-1}} \left[\mathbb{E}_{\sigma_n} \operatorname{tr} \Gamma(\boldsymbol{f})^q\right]^{1/(2q)},$$

where the carré du champ $\Gamma(\boldsymbol{f})$ is defined by (4.5). We can also obtain subgaussian tail bounds in terms of the variance proxy $v_{\boldsymbol{f}} := \| \|\Gamma(\boldsymbol{f})\| \|_{L_\infty(\sigma_n)}$. Indeed, Theorem 3.2 yields the bound

$$\mathbb{P}_{\sigma_n} \left\{\lambda_{\max}(\boldsymbol{f} - \mathbb{E}_{\sigma_n} \boldsymbol{f}) \geq t\right\} \leq d \cdot \exp\left(\frac{-(n-1)t^2}{2v_{\boldsymbol{f}}}\right) \quad \text{for all } t \geq 0.$$

To use these concentration inequalities, we need to compute the carré du champ $\Gamma(\boldsymbol{f})$ and bound the variance proxy $v_{\boldsymbol{f}}$ for particular functions $\boldsymbol{f}$.

We give two illustrations where we compute the carré du champ by viewing the sphere $\mathbb{S}^n$ as a submanifold of the Euclidean space $\mathbb{R}^{n+1}$. From this perspective, the carré du champ is simply the squared Euclidean length of the tangential gradient on the sphere; we refer the reader to the arXiv version [HT20b] of this paper for detailed calculations. In each case, let $x = (x_1, \ldots, x_{n+1}) \in \mathbb{S}^n$ be a random vector drawn from the uniform probability measure σ_n. Suppose that $(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_{n+1}) \subset \mathbb{H}_d$ is a list of deterministic Hermitian matrices.

Example 4.8 (Sphere I)**.** Consider the random matrix $\boldsymbol{f}(x) = \sum_{i=1}^{n+1} x_i \boldsymbol{A}_i$. We can compute the carré du champ as

$$\Gamma(\boldsymbol{f})(x) = \sum_{i=1}^{n+1} \boldsymbol{A}_i^2 - \left(\sum_{i=1}^{n+1} x_i \boldsymbol{A}_i\right)^2 \succcurlyeq \mathbf{0}.$$

It is obvious that $\Gamma(\boldsymbol{f})(x) \preccurlyeq \sum_{i=1}^{n+1} \boldsymbol{A}_i^2$ for all $x \in \mathbb{S}^n$, so the variance proxy $v_{\boldsymbol{f}} \leq \left\|\sum_{i=1}^{n+1} \boldsymbol{A}_i^2\right\|$.

Compare this result with Example 4.7, where the coefficients follow the standard normal distribution. For the sphere, the carré du champ operator is smaller because a finite-dimensional sphere has slightly more curvature than the Gauss space.

Example 4.9 (Sphere II)**.** Consider the random matrix $\boldsymbol{f}(x) = \sum_{i=1}^{n+1} x_i^2 \boldsymbol{A}_i$. The carré du champ admits the expression

$$\Gamma(\boldsymbol{f})(x) = 2 \sum_{i,j=1}^{n+1} x_i^2 x_j^2 (\boldsymbol{A}_i - \boldsymbol{A}_j)^2.$$

A simple bound shows that the variance proxy $v_{\boldsymbol{f}} \leq 2 \max_{i,j} \|\boldsymbol{A}_i - \boldsymbol{A}_j\|$.

4.3.2 The special orthogonal group

The Riemannian manifold framework also encompasses matrix-valued functions of random orthogonal matrices. Consider the special orthogonal group $\mathrm{SO}(d) \subset \mathbb{R}^{d \times d}$ as a Riemannian submanifold of $\mathbb{R}^{d \times d}$. The induced Riemannian metric is the Haar probability measure on $\mathrm{SO}(d)$. For this manifold, it is known that the eigenvalues of the Ricci tensor are uniformly bounded below by $\rho = (d-1)/4$; see [Led01, p. 27]. Therefore, the Brownian motion on the special orthogonal group $\mathrm{SO}(d)$ satisfies the Bakry–Émery criterion (2.15) with $c = 4/(d-1)$.

As a direction extension, the Brownian motion on the product space $\mathrm{SO}(d)^{\otimes n}$ also satisfies the Bakry–Émery criterion (2.15) with the same constant $c = 4/(d-1)$. Theorem 3.2 then implies that, for any suitable function $\boldsymbol{f} : \mathrm{SO}(d)^{\otimes n} \to \mathbb{H}_d$,

$$\mathbb{P}_{\mu^{\otimes n}} \left\{ \lambda_{\max}(\boldsymbol{f} - \mathbb{E}_{\mu^{\otimes n}} \boldsymbol{f}) \geq t \right\} \leq d \cdot \exp\left(\frac{-(d-1)t^2}{8v_{\boldsymbol{f}}} \right) \quad \text{for all } t \geq 0.$$

Here is a particular example where we can bound the variance proxy.

Example 4.10 (Special orthogonal group)**.** Let $(\boldsymbol{A}_1, \dots, \boldsymbol{A}_n) \subset \mathbb{H}_d(\mathbb{R})$ be a fixed list of real, symmetric matrices. Suppose that $\boldsymbol{O}_1, \dots, \boldsymbol{O}_n \in \mathrm{SO}(d)$ are drawn independently and uniformly from the Haar probability measure μ on $\mathrm{SO}(d)$. Consider the random matrix $\boldsymbol{f}(\boldsymbol{O}_1, \dots, \boldsymbol{O}_n) = \sum_{i=1}^n \boldsymbol{O}_i \boldsymbol{A}_i \boldsymbol{O}_i^\intercal$. The carré du champ is

$$\Gamma(\boldsymbol{f})(\boldsymbol{O}_1, \dots, \boldsymbol{O}_n) = \frac{1}{2} \sum_{i=1}^n \boldsymbol{O}_i \left[\left(\mathrm{tr}[\boldsymbol{A}_i^2] - d^{-1}\, \mathrm{tr}[\boldsymbol{A}_i]^2 \right) \cdot \mathbf{I}_d + d \left(\boldsymbol{A}_i - d^{-1}\, \mathrm{tr}[\boldsymbol{A}_i] \cdot \mathbf{I}_d \right)^2 \right] \boldsymbol{O}_i^\intercal.$$

Each matrix $\boldsymbol{O}_i$ is orthogonal, so the variance proxy satisfies

$$v_{\boldsymbol{f}} \leq \frac{1}{2} \sum_{i=1}^n \left[\mathrm{tr}[\boldsymbol{A}_i^2] - d^{-1}\, \mathrm{tr}[\boldsymbol{A}_i]^2 + d \cdot \left\| \boldsymbol{A}_i - d^{-1}\, \mathrm{tr}[\boldsymbol{A}_i] \cdot \mathbf{I}_d \right\|^2 \right].$$

We remark that the carré du champ has been obtained using the Lie group structure of $\mathrm{SO}(d)^{\otimes n}$, so the derivatives can be conveniently calculated under the geodesic frame of a tangent space. See the arXiv version [HT20b] for detailed calculations.

5 Matrix Markov semigroups: properties and proofs

This section presents some other fundamental facts about matrix Markov semigroups. We also provide proofs of the propositions from Section 2. In particular, we argue that many properties of matrix Markov semigroups follow directly from their analogs in the scalar setting.

5.1 Properties of the carré du champ operator

Our first proposition gives the matrix extension of some classical facts about the carré du champ operator Γ. Parts of this result are adapted from [ABY20, Prop. 2.2].

Proposition 5.1 (Matrix carré du champ)**.** *Let $(Z_t)_{t\geq 0}$ be a Markov process. The associated matrix bilinear form Γ has the following properties:*

1. *For all suitable $\boldsymbol{f}, \boldsymbol{g} : \Omega \to \mathbb{H}_d$ and all $z \in \Omega$,*

$$\Gamma(\boldsymbol{f}, \boldsymbol{g})(z) = \lim_{t \downarrow 0} \frac{1}{2t}\, \mathbb{E}\left[\big(\boldsymbol{f}(Z_t) - \boldsymbol{f}(Z_0)\big)\big(\boldsymbol{g}(Z_t) - \boldsymbol{g}(Z_0)\big) \,\big|\, Z_0 = z \right].$$

2. *In particular, the quadratic form $\boldsymbol{f} \mapsto \Gamma(\boldsymbol{f})$ is positive: $\Gamma(\boldsymbol{f}) \succcurlyeq \mathbf{0}$.*

3. *For all suitable $\boldsymbol{f}, \boldsymbol{g} : \Omega \to \mathbb{H}_d$ and all $s > 0$,*

$$\Gamma(\boldsymbol{f}, \boldsymbol{g}) + \Gamma(\boldsymbol{g}, \boldsymbol{f}) \preccurlyeq s\, \Gamma(\boldsymbol{f}) + s^{-1}\, \Gamma(\boldsymbol{g}).$$

4. *The quadratic form induced by Γ is operator convex:*

$$\Gamma\big(\tau \boldsymbol{f} + (1-\tau)\boldsymbol{g}\big) \preccurlyeq \tau\, \Gamma(\boldsymbol{f}) + (1-\tau)\, \Gamma(\boldsymbol{g}) \quad \textit{for each } \tau \in [0,1].$$

Similar results hold for the matrix Dirichlet form, owing to the definition (2.10)*.*

Proof. The limit form of the carré du champ in (1) can be verified with a short calculation:

$$\begin{aligned}\Gamma(\boldsymbol{f},\boldsymbol{g})(z) &= \lim_{t\downarrow 0}\frac{1}{2t}\big[\,\mathbb{E}[\boldsymbol{f}(Z_t)\boldsymbol{g}(Z_t)\,|\,Z_0=z]-\boldsymbol{f}(z)\boldsymbol{g}(z)\big]\\ &\qquad -\lim_{t\downarrow 0}\frac{1}{2t}\big[\boldsymbol{f}(z)\big(\,\mathbb{E}[\boldsymbol{g}(Z_t)\,|\,Z_0=z]-\boldsymbol{g}(z)\big)\big]\\ &\qquad -\lim_{t\downarrow 0}\frac{1}{2t}\big[\big(\,\mathbb{E}[\boldsymbol{f}(Z_t)\,|\,Z_0=z]-\boldsymbol{f}(z)\big)\boldsymbol{g}(z)\big]\\ &= \lim_{t\downarrow 0}\frac{1}{2t}\,\mathbb{E}\left[\boldsymbol{f}(Z_t)\boldsymbol{g}(Z_t)-\boldsymbol{f}(z)\boldsymbol{g}(Z_t)-\boldsymbol{f}(Z_t)\boldsymbol{g}(z)+\boldsymbol{f}(z)\boldsymbol{g}(z)\,|\,Z_0=z\right]\\ &= \lim_{t\downarrow 0}\frac{1}{2t}\,\mathbb{E}\left[(\boldsymbol{f}(Z_t)-\boldsymbol{f}(Z_0))(\boldsymbol{g}(Z_t)-\boldsymbol{g}(Z_0))\,|\,Z_0=z\right].\end{aligned}$$

The first relation depends on the definition (2.9) of Γ and the definition (2.6) of $\mathscr{L}$. Statement (2) is a direct consequence of (1) with $\boldsymbol{f}=\boldsymbol{g}$, because the square of a matrix is positive-semidefinite and the expectation preserves positivity. Statements (3) and (4) are direct consequences of (2). □

The next lemma is an extension of Proposition 5.1(1). We use this result to establish the all-important chain rule inequality in Section 7.

Lemma 5.2 (Triple product)**.** *Let $(Z_t)_{t\geq 0}$ be a reversible Markov process with a stationary measure μ and infinitesimal generator $\mathscr{L}$. For all suitable $\boldsymbol{f},\boldsymbol{g},\boldsymbol{h}:\Omega\to\mathbb{H}_d$ and all $z\in\Omega$,*

$$\begin{aligned}&\lim_{t\downarrow 0}\frac{1}{t}\operatorname{tr}\mathbb{E}\left[\big(\boldsymbol{f}(Z_t)-\boldsymbol{f}(Z_0)\big)\big(\boldsymbol{g}(Z_t)-\boldsymbol{g}(Z_0)\big)\big(\boldsymbol{h}(Z_t)-\boldsymbol{h}(Z_0)\big)\big]\,\big|\,Z_0=z\right]\\ &\quad=\operatorname{tr}\left[\mathscr{L}(\boldsymbol{fgh})-\mathscr{L}(\boldsymbol{fg})\boldsymbol{h}-\mathscr{L}(\boldsymbol{hf})\boldsymbol{g}-\mathscr{L}(\boldsymbol{gh})\boldsymbol{f}+\mathscr{L}(\boldsymbol{f})\boldsymbol{gh}+\mathscr{L}(\boldsymbol{g})\boldsymbol{hf}+\mathscr{L}(\boldsymbol{h})\boldsymbol{fg}\right](z).\end{aligned}$$

In particular,

$$\mathbb{E}_{Z\sim\mu}\lim_{t\downarrow 0}\frac{1}{t}\operatorname{tr}\mathbb{E}\left[\big(\boldsymbol{f}(Z_t)-\boldsymbol{f}(Z_0)\big)\big(\boldsymbol{g}(Z_t)-\boldsymbol{g}(Z_0)\big)\big(\boldsymbol{h}(Z_t)-\boldsymbol{h}(Z_0)\big)\big]\,\big|\,Z_0=Z\right]=0.$$

Proof. For simplicity, we abbreviate

$$\boldsymbol{f}_t=\boldsymbol{f}(Z_t),\quad \boldsymbol{g}_t=\boldsymbol{g}(Z_t),\quad \boldsymbol{h}_t=\boldsymbol{h}(Z_t)\quad\text{and}\quad \boldsymbol{f}_0=\boldsymbol{f}(Z_0),\quad \boldsymbol{g}_0=\boldsymbol{g}(Z_0),\quad \boldsymbol{h}_0=\boldsymbol{h}(Z_0).$$

Direct calculation gives

$$\begin{aligned}&\lim_{t\downarrow 0}\frac{1}{t}\operatorname{tr}\mathbb{E}\left[\big(\boldsymbol{f}(Z_t)-\boldsymbol{f}(Z_0)\big)\big(\boldsymbol{g}(Z_t)-\boldsymbol{g}(Z_0)\big)\big(\boldsymbol{h}(Z_t)-\boldsymbol{h}(Z_0)\big)\,\big|\,Z_0=z\right]\\ &\quad=\lim_{t\downarrow 0}\frac{1}{t}\operatorname{tr}\mathbb{E}\,[\boldsymbol{f}_t\boldsymbol{g}_t\boldsymbol{h}_t-\boldsymbol{f}_t\boldsymbol{g}_t\boldsymbol{h}_0-\boldsymbol{f}_t\boldsymbol{g}_0\boldsymbol{h}_t+\boldsymbol{f}_t\boldsymbol{g}_0\boldsymbol{h}_0\\ &\qquad\qquad -\boldsymbol{f}_0\boldsymbol{g}_t\boldsymbol{h}_t+\boldsymbol{f}_0\boldsymbol{g}_t\boldsymbol{h}_0+\boldsymbol{f}_0\boldsymbol{g}_0\boldsymbol{h}_t-\boldsymbol{f}_0\boldsymbol{g}_0\boldsymbol{h}_0\,\big|\,Z_0=z\big]\\ &\quad=\lim_{t\downarrow 0}\frac{1}{t}\operatorname{tr}\mathbb{E}\,\big[\big(\boldsymbol{f}_t\boldsymbol{g}_t\boldsymbol{h}_t-\boldsymbol{f}_0\boldsymbol{g}_0\boldsymbol{h}_0\big)-\big((\boldsymbol{f}_t\boldsymbol{g}_t-\boldsymbol{f}_0\boldsymbol{g}_0)\boldsymbol{h}_0\big)\\ &\qquad\qquad -\ \big((\boldsymbol{h}_t\boldsymbol{f}_t-\boldsymbol{h}_0\boldsymbol{f}_0)\boldsymbol{g}_0\big)+\big((\boldsymbol{f}_t-\boldsymbol{f}_0)\boldsymbol{g}_0\boldsymbol{h}_0\big)\\ &\qquad\qquad -\ \big((\boldsymbol{g}_t\boldsymbol{h}_t-\boldsymbol{g}_0\boldsymbol{h}_0)\boldsymbol{f}_0\big)+\big((\boldsymbol{g}_t-\boldsymbol{g}_0)\boldsymbol{h}_0\boldsymbol{f}_0\big)+\big((\boldsymbol{h}_t-\boldsymbol{h}_0)\boldsymbol{f}_0\boldsymbol{g}_0\big)\,\big|\,Z_0=z\big]\\ &\quad=\operatorname{tr}\big[\mathscr{L}(\boldsymbol{fgh})(z)-\mathscr{L}(\boldsymbol{fg})(z)\boldsymbol{h}(z)-\mathscr{L}(\boldsymbol{hf})(z)\boldsymbol{g}(z)-\mathscr{L}(\boldsymbol{gh})(z)\boldsymbol{f}(z)\\ &\qquad\qquad +\ \mathscr{L}(\boldsymbol{f})(z)\boldsymbol{g}(z)\boldsymbol{h}(z)+\mathscr{L}(\boldsymbol{g})(z)\boldsymbol{h}(z)\boldsymbol{f}(z)+\mathscr{L}(\boldsymbol{h})(z)\boldsymbol{f}(z)\boldsymbol{g}(z)\big].\end{aligned}$$

We have applied the cyclic property of the trace. Using the reversibility (2.8) of the

Markov process and the zero-mean property (2.7) of the infinitesimal generator, we have

$$\begin{aligned}
&\mathbb{E}_\mu \operatorname{tr}[\mathscr{L}(\boldsymbol{fgh}) - \mathscr{L}(\boldsymbol{fg})\boldsymbol{h} - \mathscr{L}(\boldsymbol{hf})\boldsymbol{g} - \mathscr{L}(\boldsymbol{gh})\boldsymbol{f} + \mathscr{L}(\boldsymbol{f})\boldsymbol{gh} + \mathscr{L}(\boldsymbol{g})\boldsymbol{hf} + \mathscr{L}(\boldsymbol{h})\boldsymbol{fg}] \\
&\quad = \operatorname{tr}[\mathbb{E}_\mu[\mathscr{L}(\boldsymbol{fgh})] - \mathbb{E}_\mu[\mathscr{L}(\boldsymbol{fg})\boldsymbol{h} - \boldsymbol{fg}\mathscr{L}(\boldsymbol{h})] \\
&\qquad\qquad - \mathbb{E}_\mu[\mathscr{L}(\boldsymbol{hf})\boldsymbol{g} - \boldsymbol{hf}\mathscr{L}(\boldsymbol{g})] - \mathbb{E}_\mu[\mathscr{L}(\boldsymbol{gh})\boldsymbol{f} - \boldsymbol{gh}\mathscr{L}(\boldsymbol{f})]] \\
&\quad = 0.
\end{aligned}$$

This concludes the second part of the lemma. □

5.2 Dimension reduction

In this section, we demonstrate that many properties of matrix Markov semigroups introduced in Section 2 are equivalent with their analogs in the scalar setting ($d = 1$). The pattern of argument was suggested to us by Ramon van Handel.

The technique is based on the simple observation that, for all vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^d$ and all matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{H}_d$,

$$\begin{aligned}
\boldsymbol{u}^*(\boldsymbol{AB})\boldsymbol{v} &= \sum_{j=1}^d (\boldsymbol{u}^*\boldsymbol{A}\mathbf{e}_j)(\mathbf{e}_j^*\boldsymbol{B}\boldsymbol{v}) =: \sum_{j=1}^d a_j \bar{b}_j \\
&= \sum_{j=1}^d [\operatorname{Re}(a_j)\operatorname{Re}(b_j) + \operatorname{Im}(a_j)\operatorname{Im}(b_j) - \mathrm{i}\operatorname{Re}(a_j)\operatorname{Im}(b_j) + \mathrm{i}\operatorname{Im}(a_j)\operatorname{Re}(b_j)]. \quad (5.1)
\end{aligned}$$

We have defined $a_j := \boldsymbol{u}^*\boldsymbol{A}\mathbf{e}_j$ and $b_j := \boldsymbol{v}^*\boldsymbol{A}\mathbf{e}_j$ for each $j = 1, \dots, d$. As usual, $(\mathbf{e}_j : 1 \leq j \leq d)$ is the standard basis for $\mathbb{C}^d$. The real and imaginary parts of a complex number $w \in \mathbb{C}$ are given by $\operatorname{Re}(w) := \frac{1}{2}(w + w^*) \in \mathbb{R}$ and $\operatorname{Im}(w) := \frac{1}{2\mathrm{i}}(w - w^*) \in \mathbb{R}$, where * denotes the complex conjugate.

As an example, we show how to use this observation to prove Proposition 2.2, which states that reversibility of the semigroup (2.1) on real-valued functions is equivalent with the reversibility of the semigroup on matrix-valued functions.

Proof of Proposition 2.2. The implication that matrix reversibility (2.4) for all $d \in \mathbb{N}$ implies scalar reversibility is obvious: just take $d = 1$. To check the converse, we consider two matrix-valued functions $\boldsymbol{f}, \boldsymbol{g} : \Omega \to \mathbb{H}_d$ and introduce the scalar functions $f_j := \boldsymbol{u}^*\boldsymbol{f}\mathbf{e}_j$ and $g_j := \boldsymbol{v}^*\boldsymbol{g}\mathbf{e}_j$ for each $j = 1, \dots, d$. The definition (2.1) of the semigroup $(P_t)_{t\geq 0}$ as an expectation ensures that

$$\boldsymbol{u}^*(P_t\boldsymbol{f})\mathbf{e}_j = P_t f_j = P_t(\operatorname{Re}(f_j)) + \mathrm{i}\, P_t(\operatorname{Im}(f_j)) = \operatorname{Re}(P_t f_j) + \mathrm{i}\operatorname{Im}(P_t f_j).$$

The parallel statement holds for $\boldsymbol{v}^*(P_t\boldsymbol{g})\mathbf{e}_j$. Therefore, we can use formula (5.1) to compute that

$$\begin{aligned}
&\boldsymbol{u}^*\,\mathbb{E}_\mu[(P_t\boldsymbol{f})\,\boldsymbol{g}]\boldsymbol{v} \\
&= \sum_{j=1}^d \mathbb{E}_\mu[\boldsymbol{u}^*(P_t\boldsymbol{f})\mathbf{e}_j\mathbf{e}_j^*\boldsymbol{g}\boldsymbol{v}] = \sum_{j=1}^d \mathbb{E}_\mu[(P_t f_j)\,\bar{g}_j] \\
&= \sum_{j=1}^d \mathbb{E}_\mu\,[(P_t\operatorname{Re}(f_j))\operatorname{Re}(g_j) + (P_t\operatorname{Im}(f_j))\operatorname{Im}(g_j) - \mathrm{i}(P_t\operatorname{Re}(f_j))\operatorname{Im}(g_j) + \mathrm{i}(P_t\operatorname{Im}(f_j))\operatorname{Re}(g_j)] \\
&= \sum_{j=1}^d \mathbb{E}_\mu\,[\operatorname{Re}(f_j)(P_t\operatorname{Re}(g_j)) + \operatorname{Im}(f_j)(P_t\operatorname{Im}(g_j)) - \mathrm{i}\operatorname{Re}(f_j)(P_t\operatorname{Im}(g_j)) + \mathrm{i}\operatorname{Im}(f_j)(P_t\operatorname{Re}(g_j))] \\
&= \sum_{j=1}^d \mathbb{E}_\mu[f_j\,(P_t\bar{g}_j)] = \sum_{j=1}^d \mathbb{E}_\mu[\boldsymbol{u}^*\boldsymbol{f}\mathbf{e}_j\mathbf{e}_j^*(P_t\boldsymbol{g})\boldsymbol{v}] = \boldsymbol{u}^*\,\mathbb{E}_\mu[\boldsymbol{f}\,(P_t\boldsymbol{g})]\boldsymbol{v}.
\end{aligned}$$

The matrix identity (2.4) follows immediately because $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^d$ are arbitrary. □

By applying formula (5.1) to the matrix products in the definitions (2.9) and (2.14) of the carré du champ operators and using the linearity of the infinitesimal generator (2.6), we immediately obtain the following result. It explains how to relate the carré du champ operator of a matrix-valued function to the carré du champ operators of some scalar functions.

Lemma 5.3 (Dimension reduction of carré du champ)**.** *Let $(P_t)_{t\geq 0}$ be the semigroup defined in* (2.1)*. The carré du champ operator Γ and the iterated carré du champ operator Γ_2 satisfy*

$$\boldsymbol{u}^*\Gamma(\boldsymbol{f})\boldsymbol{u} = \sum_{j=1}^{d} \big(\Gamma\big(\operatorname{Re}(\boldsymbol{u}^*\boldsymbol{f}\mathbf{e}_j)\big) + \Gamma\big(\operatorname{Im}(\boldsymbol{u}^*\boldsymbol{f}\mathbf{e}_j)\big)\big)\,;$$
$$\boldsymbol{u}^*\Gamma_2(\boldsymbol{f})\boldsymbol{u} = \sum_{j=1}^{d} \big(\Gamma_2\big(\operatorname{Re}(\boldsymbol{u}^*\boldsymbol{f}\mathbf{e}_j)\big) + \Gamma_2\big(\operatorname{Im}(\boldsymbol{u}^*\boldsymbol{f}\mathbf{e}_j)\big)\big)\,.$$

These formulae hold for all $d \in \mathbb{N}$, for all suitable functions $\boldsymbol{f} : \Omega \to \mathbb{H}_d$, and for all vectors $\boldsymbol{u} \in \mathbb{C}^d$.

The proof involves calculations similar to those in the proof of Proposition 2.2, so we omit the repetitive details. Lemma 5.3 helps us transform the scalar Poincaré inequality and the scalar Bakry–Émery criterion to their matrix equivalents. For instance, if the semigroup $(P_t)_{t\geq 0}$ satisfies the scalar Bakry-Émery criterion in Proposition 2.5(1), then for any vector $\boldsymbol{u} \in \mathbb{C}^d$ and any suitable function $\boldsymbol{f} : \Omega \mapsto \mathbb{H}_d$,

$$\begin{aligned}
\boldsymbol{u}^*\Gamma(\boldsymbol{f})\boldsymbol{u} &= \sum_{j=1}^{d} \big(\Gamma\big(\operatorname{Re}(\boldsymbol{u}^*\boldsymbol{f}\mathbf{e}_j)\big) + \Gamma\big(\operatorname{Im}(\boldsymbol{u}^*\boldsymbol{f}\mathbf{e}_j)\big)\big) \\
&\leq c\sum_{j=1}^{d} \big(\Gamma_2\big(\operatorname{Re}(\boldsymbol{u}^*\boldsymbol{f}\mathbf{e}_j)\big) + \Gamma_2\big(\operatorname{Im}(\boldsymbol{u}^*\boldsymbol{f}\mathbf{e}_j)\big)\big) \\
&= c \cdot \boldsymbol{u}^*\Gamma_2(\boldsymbol{f})\boldsymbol{u}.
\end{aligned}$$

Since $\boldsymbol{u} \in \mathbb{C}^d$ is arbitrary, we immediately obtain the matrix Bakry–Émery criterion (2.15). This proves Proposition 2.5. Similarly, Proposition 2.3 follows from the same type of calculations.

Moreover, Lemma 5.3 implies that Proposition 2.4 and Proposition 2.7 are equivalent with their counterparts in the scalar case [vH16, Theorem 2.18 & Theorem 2.36]. This releases us from proving these results all over again in the matrix setting.

6 A chain rule inequality

In this section, we state and prove the key new technical result that we need to establish polynomial and exponential matrix concentration inequalities (Theorems 3.1 and 3.2).

Lemma 6.1 (Chain rule inequality)**.** *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a scalar function for which $\psi := |\varphi'|$ is convex. For all suitable $\boldsymbol{f}, \boldsymbol{g} : \Omega \to \mathbb{H}_d$,*

$$\mathbb{E}_\mu \operatorname{tr} \Gamma(\boldsymbol{g}, \varphi(\boldsymbol{f})) \leq \Big(\mathbb{E}_\mu \operatorname{tr} [\Gamma(\boldsymbol{f})\, \psi(\boldsymbol{f})] \cdot \mathbb{E}_\mu \operatorname{tr} [\Gamma(\boldsymbol{g})\, \psi(\boldsymbol{f})] \Big)^{1/2}. \tag{6.1}$$

We use Lemma 6.1 to control the trace of the carré du champ within a standard semigroup argument. The rest of this section is devoted to the proof of the lemma.

To give some context, we remark that the carré du champ operator Γ of a (scalar-valued, reversible) diffusion semigroup satisfies a chain rule [BGL13, Sec. 1.11]:

$$\Gamma(g, \varphi(f)) = \Gamma(g)\, \varphi'(f) \quad \text{for smooth } f, g, \varphi : \mathbb{R} \to \mathbb{R}.$$

The formula (6.1) provides a substitute for this relation for an arbitrary reversible Markov semigroup acting on matrices. In the scalar setting, the lemma is related to the Stroock–Varopoulos inequality [Str84, Var85].

6.1 Mean value trace inequality

Lemma 6.1 relies on a new matrix trace inequality.

Lemma 6.2 (Mean value trace inequality). *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a function such that $\psi := |\varphi'|$ is convex. For all $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{H}_d$,*

$$\operatorname{tr}\left[\boldsymbol{C}\left(\varphi(\boldsymbol{A}) - \varphi(\boldsymbol{B})\right)\right] \leq \inf_{s>0} \frac{1}{4} \operatorname{tr}\left[\left(s\,(\boldsymbol{A}-\boldsymbol{B})^2 + s^{-1}\,\boldsymbol{C}^2\right)\left(\psi(\boldsymbol{A}) + \psi(\boldsymbol{B})\right)\right].$$

Lemma 6.2 is a common generalization of [PMT16, Lemmas 9.2 and 12.2]. The proof is similar in spirit, but it uses some additional ingredients from matrix analysis.

The key idea is to use tensorization to lift a pair of noncommuting matrices to a pair of commuting tensors. This step gives us access to tools that are not available for general matrices. For any two Hermitian matrices $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{H}_d$, define a linear operator $\boldsymbol{X} \otimes \boldsymbol{Y} : \mathbb{M}_d \to \mathbb{M}_d$ whose action is given by

$$(\boldsymbol{X} \otimes \boldsymbol{Y})(\boldsymbol{Z}) = \boldsymbol{X}\boldsymbol{Z}\boldsymbol{Y} \quad \text{for all } \boldsymbol{Z} \in \mathbb{M}_d.$$

The linear operator $\boldsymbol{X} \otimes \boldsymbol{Y}$ is self-adjoint with respect to the standard inner product on $\mathbb{M}_d$:

$$\langle \boldsymbol{Z}_1, (\boldsymbol{X} \otimes \boldsymbol{Y})(\boldsymbol{Z}_2)\rangle_{\mathbb{M}_d} = \operatorname{tr}[\boldsymbol{Z}_1^* \boldsymbol{X} \boldsymbol{Z}_2 \boldsymbol{Y}] = \operatorname{tr}[\boldsymbol{Y} \boldsymbol{Z}_1^* \boldsymbol{X} \boldsymbol{Z}_2] = \langle (\boldsymbol{X} \otimes \boldsymbol{Y})(\boldsymbol{Z}_1), \boldsymbol{Z}_2\rangle_{\mathbb{M}_d}$$
$$\text{for all } \boldsymbol{Z}_1, \boldsymbol{Z}_2 \in \mathbb{M}_d.$$

Therefore, for any function $\varphi : \mathbb{R} \to \mathbb{R}$, we can define the tensor function $\varphi(\boldsymbol{X} \otimes \boldsymbol{Y})$ using the spectral resolution of $\boldsymbol{X} \otimes \boldsymbol{Y}$. It is not hard to check that

$$\varphi(\boldsymbol{X} \otimes \mathbf{I}) = \varphi(\boldsymbol{X}) \otimes \mathbf{I} \quad \text{and} \quad \varphi(\mathbf{I} \otimes \boldsymbol{Y}) = \mathbf{I} \otimes \varphi(\boldsymbol{Y}).$$

Note that the tensors $\boldsymbol{X} \otimes \mathbf{I}$ and $\mathbf{I} \otimes \boldsymbol{Y}$ commute with each other, regardless of whether $\boldsymbol{X}$ and $\boldsymbol{Y}$ commute.

Proof of Lemma 6.2. We can write

$$\begin{aligned}\varphi(\boldsymbol{A}) - \varphi(\boldsymbol{B}) &= \big(\varphi(\boldsymbol{A}) \otimes \mathbf{I} - \mathbf{I} \otimes \varphi(\boldsymbol{B})\big)(\mathbf{I}) \\ &= \big(\varphi(\boldsymbol{A} \otimes \mathbf{I}) - \varphi(\mathbf{I} \otimes \boldsymbol{B})\big)(\mathbf{I}) = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\tau} \varphi\big(\tau \boldsymbol{A} \otimes \mathbf{I} + (1-\tau)\mathbf{I} \otimes \boldsymbol{B}\big)(\mathbf{I})\, \mathrm{d}\tau.\end{aligned}$$

Since $\boldsymbol{A} \otimes \mathbf{I}$ commutes with $\mathbf{I} \otimes \boldsymbol{B}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}\tau} \varphi\big(\tau \boldsymbol{A} \otimes \mathbf{I} + (1-\tau)\mathbf{I} \otimes \boldsymbol{B}\big) = \varphi'\big(\tau \boldsymbol{A} \otimes \mathbf{I} + (1-\tau)\mathbf{I} \otimes \boldsymbol{B}\big)(\boldsymbol{A} \otimes \mathbf{I} - \mathbf{I} \otimes \boldsymbol{B}).$$

As a consequence,

$$\begin{aligned}\varphi(\boldsymbol{A}) - \varphi(\boldsymbol{B}) &= \int_0^1 \varphi'\big(\tau \boldsymbol{A} \otimes \mathbf{I} + (1-\tau)\mathbf{I} \otimes \boldsymbol{B}\big)(\boldsymbol{A} \otimes \mathbf{I} - \mathbf{I} \otimes \boldsymbol{B})(\mathbf{I})\, \mathrm{d}\tau \\ &= \int_0^1 \varphi'\big(\tau \boldsymbol{A} \otimes \mathbf{I} + (1-\tau)\mathbf{I} \otimes \boldsymbol{B}\big)(\boldsymbol{A} - \boldsymbol{B})\, \mathrm{d}\tau =: \int_0^1 \mathscr{M}_\tau(\boldsymbol{A}, \boldsymbol{B})(\boldsymbol{A} - \boldsymbol{B})\, \mathrm{d}\tau.\end{aligned}$$

Since $\mathscr{M}_\tau(\boldsymbol{A},\boldsymbol{B})$ is a self-adjoint linear operator on the Hilbert space $\mathbb{M}_d$, we can apply the operator Cauchy–Schwarz inequality [PMT16, Lemma A.2]. For any $s>0$,

$$\begin{aligned}\operatorname{tr}\left[\boldsymbol{C}\left(\varphi(\boldsymbol{A})-\varphi(\boldsymbol{B})\right)\right] &= \langle \boldsymbol{C},\ \varphi(\boldsymbol{A})-\varphi(\boldsymbol{B})\rangle_{\mathbb{M}_d} = \int_0^1 \langle \boldsymbol{C},\ \mathscr{M}_\tau(\boldsymbol{A},\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})\rangle_{\mathbb{M}_d}\,\mathrm{d}\tau \\ &\le \int_0^1 \left[\frac{s}{2}\langle \boldsymbol{A}-\boldsymbol{B},\ \left|\mathscr{M}_\tau(\boldsymbol{A},\boldsymbol{B})\right|(\boldsymbol{A}-\boldsymbol{B})\rangle_{\mathbb{M}_d} + \frac{s^{-1}}{2}\langle \boldsymbol{C},\ \left|\mathscr{M}_\tau(\boldsymbol{A},\boldsymbol{B})\right|(\boldsymbol{C})\rangle_{\mathbb{M}_d}\right]\mathrm{d}\tau. \quad (6.2)\end{aligned}$$

By assumption, $\psi := |\varphi'|$ is convex. Thus, for all $\tau\in[0,1]$,

$$\begin{aligned}\left|\mathscr{M}_\tau(\boldsymbol{A},\boldsymbol{B})\right| &= \left|\varphi'\big(\tau\boldsymbol{A}\otimes\mathbf{I}+(1-\tau)\mathbf{I}\otimes\boldsymbol{B}\big)\right| = \psi\big(\tau\boldsymbol{A}\otimes\mathbf{I}+(1-\tau)\mathbf{I}\otimes\boldsymbol{B}\big) \\ &\preccurlyeq \tau\cdot\psi\left(\boldsymbol{A}\otimes\mathbf{I}\right)+(1-\tau)\cdot\psi\left(\mathbf{I}\otimes\boldsymbol{B}\right) = \tau\cdot\psi(\boldsymbol{A})\otimes\mathbf{I}+(1-\tau)\cdot\mathbf{I}\otimes\psi(\boldsymbol{B}).\end{aligned}$$

The argument above exploits the commutativity of $\boldsymbol{A}\otimes\mathbf{I}$ and $\mathbf{I}\otimes\boldsymbol{B}$, so we do not need ψ to be operator convex. Hence, for any $\boldsymbol{Z}\in\mathbb{M}_d$,

$$\begin{aligned}&\int_0^1 \langle \boldsymbol{Z},\ \left|\mathscr{M}_\tau(\boldsymbol{A},\boldsymbol{B})\right|(\boldsymbol{Z})\rangle_{\mathbb{M}_d}\,\mathrm{d}\tau \le \int_0^1 \big\langle \boldsymbol{Z},\ \big(\tau\cdot\psi(\boldsymbol{A})\otimes\mathbf{I}+(1-\tau)\cdot\mathbf{I}\otimes\psi(\boldsymbol{B})\big)(\boldsymbol{Z})\big\rangle_{\mathbb{M}_d}\,\mathrm{d}\tau \\ &= \frac{1}{2}\big(\langle \boldsymbol{Z},\ \psi(\boldsymbol{A})\,\boldsymbol{Z}\rangle_{\mathbb{M}_d} + \langle \boldsymbol{Z},\ \boldsymbol{Z}\,\psi(\boldsymbol{B})\rangle_{\mathbb{M}_d}\big) = \frac{1}{2}\big(\operatorname{tr}\left[\boldsymbol{Z}\boldsymbol{Z}^*\,\psi(\boldsymbol{A})\right]+\operatorname{tr}\left[\boldsymbol{Z}^*\boldsymbol{Z}\,\psi(\boldsymbol{B})\right]\big). \quad (6.3)\end{aligned}$$

Applying (6.3) to (6.2), substituting $\boldsymbol{A}-\boldsymbol{B}$ and $\boldsymbol{C}$ for $\boldsymbol{Z}$, we arrive at

$$\operatorname{tr}\left[\boldsymbol{C}\left(\varphi(\boldsymbol{A})-\varphi(\boldsymbol{B})\right)\right] \le \frac{1}{4}\operatorname{tr}\left[\left(s\left(\boldsymbol{A}-\boldsymbol{B}\right)^2+s^{-1}\,\boldsymbol{C}^2\right)\left(\psi(\boldsymbol{A})+\psi(\boldsymbol{B})\right)\right].$$

Optimize over $s>0$ to achieve the stated result. □

6.2 Proof of chain rule inequality

We are now ready to prove Lemma 6.1 from Lemma 6.2.

Proof of Lemma 6.1. For simplicity, we abbreviate

$$\boldsymbol{f}_t=\boldsymbol{f}(Z_t),\quad \boldsymbol{g}_t=\boldsymbol{g}(Z_t)\quad\text{and}\quad \boldsymbol{f}_0=\boldsymbol{f}(Z_0),\quad \boldsymbol{g}_0=\boldsymbol{g}(Z_0).$$

By Proposition 5.1(1),

$$\begin{aligned}\mathbb{E}_\mu \operatorname{tr}\Gamma(\boldsymbol{g},\varphi(\boldsymbol{f})) &= \mathbb{E}_{Z\sim\mu}\operatorname{tr}\lim_{t\downarrow 0}\frac{1}{2t}\,\mathbb{E}\left[(\boldsymbol{g}_t-\boldsymbol{g}_0)\left(\varphi(\boldsymbol{f}_t)-\varphi(\boldsymbol{f}_0)\right)\,\middle|\, Z_0=Z\right] \\ &= \mathbb{E}_{Z\sim\mu}\lim_{t\downarrow 0}\frac{1}{2t}\,\mathbb{E}\left[\operatorname{tr}\left[(\boldsymbol{g}_t-\boldsymbol{g}_0)\left(\varphi(\boldsymbol{f}_t)-\varphi(\boldsymbol{f}_0)\right)\right]\,\middle|\, Z_0=Z\right].\end{aligned} \quad (6.4)$$

Fix a parameter $s>0$. For each $t>0$, the mean value trace inequality, Lemma 6.2, yields

$$\begin{aligned}\operatorname{tr}\left[(\boldsymbol{g}_t-\boldsymbol{g}_0)\left(\varphi(\boldsymbol{f}_t)-\varphi(\boldsymbol{f}_0)\right)\right] &\le \frac{1}{4}\operatorname{tr}\left[\left(s\left(\boldsymbol{f}_t-\boldsymbol{f}_0\right)^2+s^{-1}\left(\boldsymbol{g}_t-\boldsymbol{g}_0\right)^2\right)\left(\psi(\boldsymbol{f}_t)+\psi(\boldsymbol{f}_0)\right)\right] \\ &= \frac{1}{2}\operatorname{tr}\left[\left(s\left(\boldsymbol{f}_t-\boldsymbol{f}_0\right)^2+s^{-1}\left(\boldsymbol{g}_t-\boldsymbol{g}_0\right)^2\right)\psi(\boldsymbol{f}_0)\right] \\ &\quad + \frac{1}{4}\operatorname{tr}\left[\left(s(\boldsymbol{f}_t-\boldsymbol{f}_0)^2+s^{-1}\left(\boldsymbol{g}_t-\boldsymbol{g}_0\right)^2\right)\left(\psi(\boldsymbol{f}_t)-\psi(\boldsymbol{f}_0)\right)\right].\end{aligned} \quad (6.5)$$

It follows from the triple product result, Lemma 5.2, that the second term satisfies

$$\mathbb{E}_{Z\sim\mu}\lim_{t\downarrow 0}\frac{1}{t}\operatorname{tr}\mathbb{E}\left[\left(s\left(\boldsymbol{f}_t-\boldsymbol{f}_0\right)^2+s^{-1}\left(\boldsymbol{g}_t-\boldsymbol{g}_0\right)^2\right)\left(\psi(\boldsymbol{f}_t)-\psi(\boldsymbol{f}_0)\right)\,\middle|\, Z_0=Z\right]=0. \quad (6.6)$$

Sequence the displays (6.4), (6.5) and (6.6) to reach

$$\begin{aligned}\mathbb{E}_\mu \operatorname{tr}\Gamma(\boldsymbol{g},\varphi(\boldsymbol{f})) &\le \frac{1}{2}\,\mathbb{E}_{Z\sim\mu}\lim_{t\downarrow 0}\frac{1}{2t}\operatorname{tr}\mathbb{E}\left[\left(s\,(\boldsymbol{f}_t-\boldsymbol{f}_0)^2+s^{-1}\,(\boldsymbol{g}_t-\boldsymbol{g}_0)^2\right)\psi(\boldsymbol{f}_0)\,\middle|\,Z_0=Z\right]\\
&= \frac{1}{2}\,\mathbb{E}_{Z\sim\mu}\operatorname{tr}\Big[\Big(s\lim_{t\downarrow 0}\frac{1}{2t}\,\mathbb{E}[(\boldsymbol{f}_t-\boldsymbol{f}_0)^2\,|\,Z_0=Z]\\
&\qquad\qquad + s^{-1}\lim_{t\downarrow 0}\frac{1}{2t}\,\mathbb{E}[(\boldsymbol{g}_t-\boldsymbol{g}_0)^2\,|\,Z_0=Z]\Big)\psi(\boldsymbol{f}(Z))\Big]\\
&= \frac{1}{2}\,\mathbb{E}_\mu\operatorname{tr}\left[\left(s\,\Gamma(\boldsymbol{f})+s^{-1}\,\Gamma(\boldsymbol{g})\right)\,\psi(\boldsymbol{f})\right].\end{aligned}$$

The last relation is Proposition 5.1(1). Minimize the right-hand side over $s\in(0,\infty)$ to arrive at

$$\mathbb{E}_\mu \operatorname{tr}\Gamma(\boldsymbol{g},\varphi(\boldsymbol{f})) \le \left(\,\mathbb{E}_\mu\operatorname{tr}\left[\Gamma(\boldsymbol{f})\,\psi(\boldsymbol{f})\right]\right)^{1/2}\cdot\left(\,\mathbb{E}_\mu\operatorname{tr}\left[\Gamma(\boldsymbol{g})\,\psi(\boldsymbol{f})\right]\right)^{1/2}.$$

This completes the proof of Lemma 6.1. □

7 From curvature conditions to matrix moment inequalities

The main results of this paper, Theorems 3.1 and 3.2, demonstrate that the Bakry–Émery criterion (2.15) leads to trace moment inequalities for random matrices. This section is dedicated to the proofs of these theorems. These arguments appear to be new, even in the scalar setting, but see [Led92, Sch99] for some precedents.

7.1 A Markov semigroup argument

We first explain the common strategy that we will use to prove the main theorems. It is the most essential part in our method.

Let us recall the hypotheses. Suppose that Ω is a Polish space equipped with a Borel probability measure μ. We consider a reversible, ergodic semigroup $(P_t)_{t\ge 0}$ with stationary measure μ that acts on suitable functions $\boldsymbol{f}:\Omega\to\mathbb{H}_d$. Without loss of generality, we can fix a dimension $d\in\mathbb{N}$, and we select an $\mathbb{H}_d$-valued function $\boldsymbol{f}$ with zero mean: $\mathbb{E}_\mu\,\boldsymbol{f}=\mathbf{0}$.

Our core assumption is that the semigroup satisfies the scalar Bakry–Émery criterion (2.15) for a constant $c>0$. Proposition 2.5 states that the semigroup also satisfies the matrix Bakry–Émery criterion for dimension d. Therefore, Proposition 2.7 guarantees that the action of the semigroup on $\mathbb{H}_d$-valued functions is locally ergodic.

Let $\varphi:\mathbb{R}\to\mathbb{R}$ be a scalar function. The core idea is to estimate a trace moment of the form $\mathbb{E}_\mu\operatorname{tr}[\boldsymbol{f}\,\varphi(\boldsymbol{f})]$ via a classical semigroup argument:

$$\mathbb{E}_\mu\operatorname{tr}[\boldsymbol{f}\,\varphi(\boldsymbol{f})]=\mathbb{E}_\mu\operatorname{tr}[P_0(\boldsymbol{f})\,\varphi(\boldsymbol{f})]=\lim_{t\to\infty}\mathbb{E}_\mu\operatorname{tr}[P_t(\boldsymbol{f})\,\varphi(\boldsymbol{f})]-\int_0^\infty\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}_\mu\operatorname{tr}[P_t(\boldsymbol{f})\,\varphi(\boldsymbol{f})]\,\mathrm{d}t.$$

By ergodicity (2.3), we confirm that $\lim_{t\to\infty}\mathbb{E}_\mu\operatorname{tr}[P_t(\boldsymbol{f})\,\varphi(\boldsymbol{f})]=\mathbb{E}_\mu\operatorname{tr}[(\mathbb{E}_\mu\,\boldsymbol{f})\,\varphi(\boldsymbol{f})]=0$. In the second term on the right-hand side, the time derivative places the infinitesimal generator $\mathscr{L}$ in the integrand, which then yields

$$\mathbb{E}_\mu\operatorname{tr}[\boldsymbol{f}\,\varphi(\boldsymbol{f})]=-\int_0^\infty\mathbb{E}_\mu\operatorname{tr}[\mathscr{L}(P_t\boldsymbol{f})\,\varphi(\boldsymbol{f})]\,\mathrm{d}t=\int_0^\infty\mathbb{E}_\mu\operatorname{tr}\Gamma(P_t\boldsymbol{f},\varphi(\boldsymbol{f}))\,\mathrm{d}t. \tag{7.1}$$

The second equality above follows from the formula (2.12) and is the starting point for our method.

Now, suppose that $\psi=|\varphi'|$ is convex. Then we can employ the chain rule inequality, Lemma 6.1, to control the trace of the carré du champ:

$$\mathbb{E}_\mu\operatorname{tr}\Gamma(P_t\boldsymbol{f},\varphi(\boldsymbol{f}))\le\Big(\,\mathbb{E}_\mu\operatorname{tr}\left[\Gamma(\boldsymbol{f})\,\psi(\boldsymbol{f})\right]\cdot\mathbb{E}_\mu\operatorname{tr}\left[\Gamma(P_t\boldsymbol{f})\,\psi(\boldsymbol{f})\right]\Big)^{1/2}. \tag{7.2}$$

To proceed, we invoke the local ergodicity property, Proposition 2.7(2), to estimate $\Gamma(P_t \boldsymbol{f})$. That is, $\Gamma(P_t \boldsymbol{f}) \preccurlyeq \mathrm{e}^{-2t/c} P_t \Gamma(\boldsymbol{f})$. Note that $\psi(\boldsymbol{f}) = |\varphi'|(\boldsymbol{f}) \succcurlyeq \mathbf{0}$. Thus, we have

$$\mathbb{E}_\mu \operatorname{tr} [\Gamma(P_t \boldsymbol{f}) \, \psi(\boldsymbol{f})] \leq \mathrm{e}^{-2t/c} \, \mathbb{E}_\mu \operatorname{tr} [(P_t \Gamma(\boldsymbol{f})) \, \psi(\boldsymbol{f})] \, . \tag{7.3}$$

Sequencing the displays (7.1), (7.2) and (7.3) yields

$$\mathbb{E}_\mu \operatorname{tr}[\boldsymbol{f} \, \varphi(\boldsymbol{f})] \leq \Big(\mathbb{E}_\mu \operatorname{tr} [\Gamma(\boldsymbol{f}) \, \psi(\boldsymbol{f})] \Big)^{1/2} \int_0^\infty \mathrm{e}^{-t/c} \Big(\mathbb{E}_\mu \operatorname{tr} [(P_t \Gamma(\boldsymbol{f})) \, \psi(\boldsymbol{f})] \Big)^{1/2} \, \mathrm{d}t. \tag{7.4}$$

Last, we apply the matrix decoupling techniques, based on Hölder and Young trace inequalities, to bound $\mathbb{E} \operatorname{tr} [\Gamma(\boldsymbol{f}) \, \psi(\boldsymbol{f})]$ and $\mathbb{E} \operatorname{tr} [(P_t \Gamma(\boldsymbol{f})) \, \psi(\boldsymbol{f})]$ in terms of the original quantity of interest $\mathbb{E}_\mu \operatorname{tr}[\boldsymbol{f} \, \varphi(\boldsymbol{f})]$. The subsequent sections supply full details in case φ is a polynomial or exponential function.

Our approach incorporates some techniques and ideas from [PMT16, Theorems 4.2 and 4.3], but the argument is distinct. Section 7.5 gives more details about the connection.

7.2 Polynomial moments

This section is dedicated to the proof of Theorem 3.1, which states that the Bakry–Émery criterion implies matrix polynomial moment bounds.

7.2.1 Setup

Fix a suitable function $\boldsymbol{f} : \Omega \to \mathbb{H}_d$. Proposition 5.1(1) implies that the carré du champ is shift invariant. In particular, $\Gamma(\boldsymbol{f}) = \Gamma(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f})$. Therefore, we may assume that $\mathbb{E}_\mu \boldsymbol{f} = \mathbf{0}$.

The quantity of interest is

$$\mathbb{E}_\mu \operatorname{tr} |\boldsymbol{f}|^{2q} = \mathbb{E}_\mu \operatorname{tr} \left[\boldsymbol{f} \cdot \operatorname{sgn}(\boldsymbol{f}) \cdot |\boldsymbol{f}|^{2q-1} \right] =: \mathbb{E}_\mu \operatorname{tr} [\boldsymbol{f} \, \varphi(\boldsymbol{f})] \, .$$

We have introduced the signed moment function $\varphi : x \mapsto \operatorname{sgn}(x) \cdot |x|^{2q-1}$ for $x \in \mathbb{R}$. Note that the absolute derivative $\psi(x) := |\varphi'(x)| = (2q-1) \, |x|^{2q-2}$ is convex when $q = 1$ or when $q \geq 1.5$.

Remark 7.1 (Missing powers)**.** A similar argument holds when $q \in (1, 1.5)$. It requires a variant of Lemma 6.1 that holds for monotone ψ, but has an extra factor of 2 on the right-hand side.

7.2.2 Decoupling

We can invoke the semigroup argument from Section 7.1 with the function $\varphi(x) = \operatorname{sgn}(x) \cdot |x|^{2q-1}$. In this case, the inequality (7.4) reads

$$\mathbb{E}_\mu \operatorname{tr} |\boldsymbol{f}|^{2q} \leq (2q-1) \, \left(\mathbb{E}_\mu \operatorname{tr} \left[\Gamma(\boldsymbol{f}) \, |\boldsymbol{f}|^{2q-2} \right] \right)^{1/2} \int_0^\infty \mathrm{e}^{-t/c} \left(\mathbb{E}_\mu \operatorname{tr} \left[(P_t \Gamma(\boldsymbol{f})) \, |\boldsymbol{f}|^{2q-2} \right] \right)^{1/2} \mathrm{d}t. \tag{7.5}$$

Apply Hölder's inequality for the trace followed by Hölder's inequality for the expectation to obtain

$$\begin{aligned} \mathbb{E}_\mu \operatorname{tr} \left[\Gamma(\boldsymbol{f}) \, |\boldsymbol{f}|^{2q-2} \right] &\leq (\mathbb{E}_\mu \operatorname{tr} \Gamma(\boldsymbol{f})^q)^{1/q} \cdot \left(\mathbb{E}_\mu \operatorname{tr} |\boldsymbol{f}|^{2q} \right)^{(q-1)/q} ; \\ \mathbb{E}_\mu \operatorname{tr} \left[(P_t \Gamma(\boldsymbol{f})) \, |\boldsymbol{f}|^{2q-2} \right] &\leq (\mathbb{E}_\mu \operatorname{tr} (P_t \Gamma(\boldsymbol{f}))^q)^{1/q} \cdot \left(\mathbb{E}_\mu \operatorname{tr} |\boldsymbol{f}|^{2q} \right)^{(q-1)/q} . \end{aligned} \tag{7.6}$$

Introduce (7.6) into (7.5). Rearrange the expression to reach

$$\left(\mathbb{E}_\mu \operatorname{tr} |\boldsymbol{f}|^{2q} \right)^{1/q} \leq (2q-1) \, (\mathbb{E}_\mu \operatorname{tr} \Gamma(\boldsymbol{f})^q)^{1/(2q)} \int_0^\infty \mathrm{e}^{-t/c} \, (\mathbb{E}_\mu \operatorname{tr} (P_t \Gamma(\boldsymbol{f}))^q)^{1/(2q)} \, \mathrm{d}t. \tag{7.7}$$

It remains to remove the semigroup from the integral.

7.2.3 Endgame

The trace power $\operatorname{tr}[(\cdot)^q]$ is convex on $\mathbb{H}_d$ for $q \geq 1$; see [Car10, Theorem 2.10]. Therefore, the Jensen inequality (2.5) for the semigroup implies that

$$\mathbb{E}_\mu \operatorname{tr} (P_t \Gamma(\boldsymbol{f}))^q \leq \mathbb{E}_\mu \operatorname{tr} \Gamma(\boldsymbol{f})^q. \tag{7.8}$$

Substituting (7.8) into (7.7) yields

$$\left(\mathbb{E}_\mu \operatorname{tr} |\boldsymbol{f}|^{2q} \right)^{1/q} \leq (2q-1) \left(\mathbb{E}_\mu \operatorname{tr} \Gamma(\boldsymbol{f})^q \right)^{1/q} \int_0^\infty \mathrm{e}^{-t/c} \, \mathrm{d}t = c\,(2q-1) \left(\mathbb{E}_\mu \operatorname{tr} \Gamma(\boldsymbol{f})^q \right)^{1/q}.$$

This establishes (3.1).

Define the uniform bound $v_{\boldsymbol{f}} := \| \|\Gamma(\boldsymbol{f})\| \|_{L_\infty(\mu)}$. We have the further estimate

$$\left(\mathbb{E}_\mu \operatorname{tr} \left[\Gamma(\boldsymbol{f})^q \right] \right)^{1/(2q)} \leq d^{1/(2q)} \sqrt{v_{\boldsymbol{f}}}.$$

The statement (3.2) now follows from (3.1). This step completes the proof of Theorem 3.1.

7.3 Exponential moments

In this section, we establish Theorem 3.2, the exponential matrix concentration inequality. The main technical ingredient is a bound on exponential moments:

Theorem 7.2 (Exponential moments)**.** *Instate the hypotheses of Theorem 3.2. For all* $\theta \in (-\sqrt{\beta/c}, \sqrt{\beta/c})$,

$$\log \mathbb{E}_\mu \operatorname{\bar{tr}} \mathrm{e}^{\theta(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f})} \leq \frac{c\theta^2 r_{\boldsymbol{f}}(\beta)}{2(1 - c\theta^2/\beta)}. \tag{7.9}$$

Moreover, if $v_{\boldsymbol{f}} < +\infty$*, then*

$$\log \mathbb{E}_\mu \operatorname{\bar{tr}} \mathrm{e}^{\theta(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f})} \leq \frac{c v_{\boldsymbol{f}} \theta^2}{2} \quad \text{for all } \theta \in \mathbb{R}. \tag{7.10}$$

The proof of Theorem 7.2 occupies the rest of this subsection. Afterward, in Section 7.4, we derive Theorem 3.2.

7.3.1 Setup

Fix a suitable function $\boldsymbol{f} : \Omega \to \mathbb{H}_d$. Again, we may assume that $\mathbb{E}_\mu \boldsymbol{f} = \mathbf{0}$. Furthermore, we only need to consider the case $\theta \geq 0$. The results for $\theta < 0$ follow formally under the change of variables $\theta \mapsto -\theta$ and $\boldsymbol{f} \mapsto -\boldsymbol{f}$.

The quantity of interest is the normalized trace moment generating function (mgf):

$$m(\theta) := \mathbb{E}_\mu \operatorname{\bar{tr}} \mathrm{e}^{\theta \boldsymbol{f}} \quad \text{for } \theta \geq 0.$$

We will bound the derivative of this function:

$$m'(\theta) = \mathbb{E}_\mu \operatorname{\bar{tr}} \left[\boldsymbol{f} \, \mathrm{e}^{\theta \boldsymbol{f}} \right] =: \mathbb{E}_\mu \operatorname{\bar{tr}} [\boldsymbol{f} \, \varphi(\boldsymbol{f})].$$

We have introduced the function $\varphi : x \mapsto \mathrm{e}^{\theta x}$ for $x \in \mathbb{R}$. Note that its absolute derivative $\psi(x) := |\varphi'(x)| = \theta \mathrm{e}^{\theta x}$ is a convex function, since $\theta \geq 0$.

7.3.2 Decoupling

We invoke the semigroup argument from Section 7.1 with the function $\varphi(x) = \mathrm{e}^{\theta x}$. The inequality (7.4) now states that

$$m'(\theta) \le \theta \left(\mathbb{E}_\mu \,\bar{\mathrm{tr}}\left[\Gamma(\boldsymbol{f})\,\mathrm{e}^{\theta \boldsymbol{f}}\right]\right)^{1/2} \int_0^\infty \mathrm{e}^{-t/c} \left(\mathbb{E}_\mu \,\bar{\mathrm{tr}}\left[(P_t\Gamma(\boldsymbol{f}))\,\mathrm{e}^{\theta \boldsymbol{f}}\right]\right)^{1/2} \mathrm{d}t. \tag{7.11}$$

To decouple the carré du champ operator in (7.11) from the matrix exponential, we need to use an entropy inequality.

Fact 7.3. *(Young's inequality for matrix entropy) Let $\boldsymbol{X}$ be a random matrix in $\mathbb{H}_d$, and let $\boldsymbol{Y}$ be a random matrix in $\mathbb{H}_d^+$ such that $\mathbb{E}\,\bar{\mathrm{tr}}\,\boldsymbol{Y} = 1$. Then*

$$\mathbb{E}\,\bar{\mathrm{tr}}\,[\boldsymbol{XY}] \le \log \mathbb{E}\,\bar{\mathrm{tr}}\,\mathrm{e}^{\boldsymbol{X}} + \mathbb{E}\,\bar{\mathrm{tr}}\,[\boldsymbol{Y}\log \boldsymbol{Y}]\,.$$

This result appears as [MJC$^+$14, Proposition A.3]; see also [Car10, Theorem 2.13].

Apply Fact 7.3 to see that, for any $\beta > 0$,

$$\begin{aligned}
\mathbb{E}_\mu \,\bar{\mathrm{tr}}\left[\Gamma(\boldsymbol{f})\,\mathrm{e}^{\theta \boldsymbol{f}}\right] &= \frac{m(\theta)}{\beta}\,\mathbb{E}_\mu \,\bar{\mathrm{tr}}\left[\beta\Gamma(\boldsymbol{f})\frac{\mathrm{e}^{\theta \boldsymbol{f}}}{m(\theta)}\right] \\
&\le \frac{m(\theta)}{\beta}\left(\log \mathbb{E}_\mu \,\bar{\mathrm{tr}} \exp\left(\beta\Gamma(\boldsymbol{f})\right) + \mathbb{E}_\mu \,\bar{\mathrm{tr}}\left[\frac{\mathrm{e}^{\theta \boldsymbol{f}}}{m(\theta)} \log \frac{\mathrm{e}^{\theta \boldsymbol{f}}}{m(\theta)}\right]\right) \\
&= m(\theta)\, r(\beta) + \frac{1}{\beta}\,\mathbb{E}_\mu \,\bar{\mathrm{tr}}\left[\mathrm{e}^{\theta \boldsymbol{f}} \log \frac{\mathrm{e}^{\theta \boldsymbol{f}}}{m(\theta)}\right].
\end{aligned} \tag{7.12}$$

We have identified the exponential mean $r(\beta) := \beta^{-1} \log \mathbb{E}_\mu \,\bar{\mathrm{tr}} \exp\left(\beta\Gamma(\boldsymbol{f})\right)$.

Likewise,

$$\mathbb{E}_\mu \,\bar{\mathrm{tr}}\left[(P_t\Gamma(\boldsymbol{f}))\,\mathrm{e}^{\theta \boldsymbol{f}}\right] \le \frac{m(\theta)}{\beta} \log \mathbb{E}_\mu \,\bar{\mathrm{tr}} \exp\left(\beta P_t \Gamma(\boldsymbol{f})\right) + \frac{1}{\beta}\,\mathbb{E}_\mu \,\bar{\mathrm{tr}}\left[\mathrm{e}^{\theta \boldsymbol{f}} \log \frac{\mathrm{e}^{\theta \boldsymbol{f}}}{m(\theta)}\right].$$

The trace exponential $\bar{\mathrm{tr}} \exp(\cdot)$ is operator convex; see [Car10, Theorem 2.10]. The Jensen inequality (2.5) for the semigroup implies that

$$\mathbb{E}_\mu \,\bar{\mathrm{tr}} \exp\left(\beta P_t \Gamma(\boldsymbol{f})\right) \le \mathbb{E}_\mu \,\bar{\mathrm{tr}} \exp\left(\beta\Gamma(\boldsymbol{f})\right) = \exp\left(\beta r(\beta)\right).$$

Combine the last two displays to obtain

$$\mathbb{E}_\mu \,\bar{\mathrm{tr}}\left[(P_t\Gamma(\boldsymbol{f}))\,\mathrm{e}^{\theta \boldsymbol{f}}\right] \le m(\theta)\, r(\beta) + \frac{1}{\beta}\,\mathbb{E}_\mu \,\bar{\mathrm{tr}}\left[\mathrm{e}^{\theta \boldsymbol{f}} \log \frac{\mathrm{e}^{\theta \boldsymbol{f}}}{m(\theta)}\right]. \tag{7.13}$$

Thus, the two terms on the right-hand side of (7.11) have matching bounds.

Sequence the displays (7.11), (7.12), and (7.13) to reach

$$m'(\theta) \le c\,\theta \left(m(\theta)\, r(\beta) + \frac{1}{\beta}\,\mathbb{E}_\mu \,\bar{\mathrm{tr}}\left[\mathrm{e}^{\theta \boldsymbol{f}} \log \frac{\mathrm{e}^{\theta \boldsymbol{f}}}{m(\theta)}\right]\right). \tag{7.14}$$

Next, we simplify this expression to arrive at a differential inequality.

7.3.3 A differential inequality

Since $\mathrm{e}^{\theta \boldsymbol{f}} \succcurlyeq \boldsymbol{I}$, we have $\log m(\theta) \ge \log \bar{\mathrm{tr}}\,\boldsymbol{I} = 0$ and hence

$$\log \frac{\mathrm{e}^{\theta \boldsymbol{f}}}{m(\theta)} = \theta \boldsymbol{f} - \log m(\theta) \cdot \mathbf{I} \preccurlyeq \theta \boldsymbol{f}.$$

It follows that

$$\mathbb{E}_\mu \bar{\operatorname{tr}}\left[\mathrm{e}^{\theta \boldsymbol{f}} \log \frac{\mathrm{e}^{\theta \boldsymbol{f}}}{m(\theta)}\right] \leq \theta\, \mathbb{E}_\mu \bar{\operatorname{tr}}\left[\boldsymbol{f}\, \mathrm{e}^{\theta \boldsymbol{f}}\right] = \theta\, m'(\theta). \tag{7.15}$$

Combine (7.14) and (7.15) to arrive at the differential inequality

$$m'(\theta) \leq c\theta\, m(\theta)\, r(\beta) + \frac{c\theta^2}{\beta} m'(\theta) \quad \text{for } \theta \geq 0. \tag{7.16}$$

Finally, we need to solve for the trace mgf.

7.3.4 Solving the differential inequality

Fix parameters θ and β where $0 \leq \theta < \sqrt{\beta/c}$. By rearranging the expression (7.16), we find that

$$\frac{\mathrm{d}}{\mathrm{d}\zeta} \log m(\zeta) \leq \frac{c\zeta\, r(\beta)}{1 - c\zeta^2/\beta} \leq \frac{c\zeta\, r(\beta)}{1 - c\theta^2/\beta} \quad \text{for } \zeta \in (0, \theta].$$

Since $\log m(0) = \log \bar{\operatorname{tr}}\, \boldsymbol{I} = 0$, we can integrate this bound over $[0, \theta]$ to obtain

$$\log m(\theta) \leq \frac{c\theta^2 r(\beta)}{2(1 - c\theta^2/\beta)}.$$

This is the first claim (7.9).

Moreover, it is easy to check that $r(\beta) \leq v_{\boldsymbol{f}}$. Since this bound is independent of β, we can take $\beta \to +\infty$ in (7.9) to achieve (7.10). This completes the proof of Theorem 7.2.

7.4 Exponential matrix concentration

We are now ready to prove Theorem 3.2, the exponential matrix concentration inequality, as a consequence of the moment bounds of Theorem 7.2. To do so, we use the standard matrix Laplace transform method. For example, see [MJC+14, Proposition 3.3]. The following proposition states a special case of this method.

Proposition 7.4. *Let $\boldsymbol{X} \in \mathbb{H}_d$ be a random matrix with normalized trace mgf $m(\theta) := \mathbb{E}\, \bar{\operatorname{tr}}\, \mathrm{e}^{\theta \boldsymbol{X}}$. Assume that there are constants $c_1, c_2 \geq 0$ for which*

$$\log m(\theta) \leq \frac{c_1 \theta^2}{2(1 - c_2\theta)} \quad \textit{when } 0 \leq \theta < \frac{1}{c_2}.$$

Then for all $t \geq 0$,

$$\mathbb{P}\left\{\lambda_{\max}(\boldsymbol{X} \geq t)\right\} \leq d \cdot \exp\left(\frac{-t^2}{2c_1 + 2c_2 t}\right).$$

Furthermore,

$$\mathbb{E}\, \lambda_{\max}(\boldsymbol{X}) \leq \sqrt{2c_1 \log d} + c_2 \log d.$$

This result converts estimates on the trace mgf into bounds on the largest eigenvalue of a random matrix. See [MJC+14, Section 4.2.4] for a proof.

Proof of Theorem 3.2 from Theorem 7.2. To obtain inequalities for the maximum eigenvalue $\lambda_{\max}$, we apply Proposition 7.4 to the random matrix $\boldsymbol{X} = \boldsymbol{f}(Z) - \mathbb{E}_\mu \boldsymbol{f}$ where $Z \sim \mu$. To do so, we first need to weaken the moment bound (7.9):

$$\log \mathbb{E}_\mu \bar{\operatorname{tr}}\, \mathrm{e}^{\theta(\boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f})} \leq \frac{c\theta^2 r(\beta)}{2(1 - c\theta^2/\beta)} \leq \frac{c\theta^2 r(\beta)}{2(1 - \theta\sqrt{c/\beta})} \quad \text{for } 0 \leq \theta < \sqrt{\beta/c}.$$

Then substitute $c_1 = cr(\beta)$ and $c_2 = \sqrt{c/\beta}$ into Proposition 7.4 to achieve the results stated in Theorem 3.2.

To obtain bounds for the minimum eigenvalue $\lambda_{\min}$, we apply Proposition 7.4 instead to the random matrix $\boldsymbol{X} = -(\boldsymbol{f}(Z) - \mathbb{E}_\mu \boldsymbol{f})$ where $Z \sim \mu$. □

7.5 Connection with Stein's method

There is an established approach to proving matrix concentration inequalities using the method of exchangeable pairs; see [Cha05] for the scalar setting and [MJC+14, PMT16] for matrix extensions. The approach in our paper can be viewed as a continuous version of this method. Let us detail the connection, glossing over some technical details.

First, we give an alternative account of the semigroup argument in Section 7.1. Let $\boldsymbol{f} : \Omega \to \mathbb{H}_d$ be a suitable function. Suppose that we solve the Poisson problem for the semigroup:

$$\boldsymbol{g}_0 := (-\mathscr{L})^{-1}\boldsymbol{f} = \int_0^\infty P_t \boldsymbol{f} \,\mathrm{d}t.$$

We can justify the convergence of this integral when the semigroup satisfies a Poincaré inequality (so is exponentially ergodic):

$$\boldsymbol{f} = \boldsymbol{f} - \mathbb{E}_\mu \boldsymbol{f} = P_0\boldsymbol{f} - P_\infty \boldsymbol{f} = -\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t} P_t \boldsymbol{f} \,\mathrm{d}t = -\mathscr{L} \int_0^\infty P_t \boldsymbol{f} \,\mathrm{d}t = -\mathscr{L}\boldsymbol{g}_0 \quad \text{in } L_2(\mu).$$

Consequently, we have

$$\mathbb{E}_\mu[\boldsymbol{f}\,\varphi(\boldsymbol{f})] = -\mathbb{E}_\mu\left[\mathscr{L}(\boldsymbol{g}_0)\,\varphi(\boldsymbol{f})\right] = \mathbb{E}_\mu \Gamma(\boldsymbol{g}_0, \varphi(\boldsymbol{f})). \tag{7.17}$$

Using the integral expression for $\boldsymbol{g}_0$ and the linearity of the carré du champ form, we can obtain

$$\mathbb{E}_\mu[\boldsymbol{f}\,\varphi(\boldsymbol{f})] = \int_0^\infty \mathbb{E}_\mu \Gamma(P_t\boldsymbol{f}, \varphi(\boldsymbol{f}))\,\mathrm{d}t.$$

This formula coincides with the identity (7.1).

Paulin et al. [PMT16] use a discrete version of this argument to obtain the matrix Efron–Stein inequalities. Given a reversible, exponentially ergodic Markov process $(Z_t)_{t\geq 0}$ with a stationary measure μ, one can construct an exchangeable pair as follows. Fix a time $t > 0$. Let Z be drawn from the measure μ, and let $\tilde{Z} = Z_t$ where $Z_0 = Z$. By reversibility, it is easy to check that $(Z, \tilde{Z})$ is an exchangeable pair; that is, $(Z, \tilde{Z})$ has the same distribution as $(\tilde{Z}, Z)$.

For a zero-mean function $\boldsymbol{f} : \Omega \to \mathbb{H}_d$, define the function $\boldsymbol{g}_t : \Omega \to \mathbb{H}_d$ by

$$\boldsymbol{g}_t = \left(\frac{P_0 - P_t}{t}\right)^{-1} \boldsymbol{f} = t\sum_{k=0}^\infty P_{kt}\boldsymbol{f}.$$

To justify the convergence of this series (in the product space case), Paulin et al. use a standard coupling argument. We remark that the function $\boldsymbol{g}_t$ is a solution to the Poisson problem for the Markov chain with transition kernel P_t. The function $\boldsymbol{g}_0$ defined above is simply the limit of $\boldsymbol{g}_t$ as $t \downarrow 0$.

The pair $(\boldsymbol{f}(Z), \boldsymbol{f}(\tilde{Z}))$ is called a *kernel Stein pair* associated with the kernel

$$\boldsymbol{K}_t(z, \tilde{z}) := \frac{\boldsymbol{g}_t(z) - \boldsymbol{g}_t(\tilde{z})}{t} \quad \text{for all } z, \tilde{z} \in \Omega.$$

By construction, for all $z, \tilde{z} \in \Omega$,

$$\begin{aligned} \boldsymbol{K}_t(z, \tilde{z}) &= -\boldsymbol{K}_t(\tilde{z}, z); \\ \mathbb{E}\left[\boldsymbol{K}_t(Z, \tilde{Z}) \,|\, Z = z\right] &= \boldsymbol{f}(z). \end{aligned} \tag{7.18}$$

This construction is inspired by Stein's work [Ste86]; see Chatterjee's PhD thesis [Cha05, Section 4.1]. The property (7.18) yields the identity

$$\mathbb{E}[\boldsymbol{f}(Z)\,\varphi(\boldsymbol{f}(Z))] = \mathbb{E}[\boldsymbol{K}_t(Z, \tilde{Z})\,\varphi(\boldsymbol{f}(Z))] = \frac{1}{2}\,\mathbb{E}\left[\boldsymbol{K}_t(Z, \tilde{Z})\left(\varphi(\boldsymbol{f}(Z)) - \varphi(\boldsymbol{f}(\tilde{Z}))\right)\right], \tag{7.19}$$

This is roughly the analog of the continuous formula (7.17). In fact, the limit formula, Proposition 5.1(1), implies that taking the limit $t \downarrow 0$ on the right hand side of (7.19) yields exactly the right hand side of (7.17).

Paulin et al. [PMT16] use a variant of (7.19) to establish matrix Efron–Stein inequalities for product measures, much in the same way that we derive Theorem 3.1 and Theorem 7.2. Instead of constructing $\boldsymbol{K}_t(Z, Z')$ using a continuous-time Markov process, they construct the associated Stein kernel $\boldsymbol{K}(Z, Z')$ in a similar way based on a reversible, exponentially ergodic Markov chain. This Markov chain can be seen as a discrete version of the Markov process introduced in Section 4.1. Correspondingly, they obtain the kernel relation (7.19) with $\boldsymbol{K}_t(Z, Z')$ replaced by $\boldsymbol{K}(Z, Z')$. Their argument then consists of three major steps that are parallel with our method.

First, they use more specific versions ([PMT16, Lemmas 9.2 and 12.2]) of the mean value trace inequality, Lemma 6.2, to bound the trace of the right hand side of (7.19): for any $s > 0$,

$$\begin{aligned}
&\frac{1}{2}\,\mathbb{E}\,\mathrm{tr}\left[\boldsymbol{K}(Z,\tilde{Z})\left(\varphi(\boldsymbol{f}(Z)) - \varphi(\boldsymbol{f}(\tilde{Z}))\right)\right] \\
&\le \frac{1}{8}\,\mathbb{E}\,\mathrm{tr}\left[\left(s\big(\boldsymbol{f}(Z) - \boldsymbol{f}(Z')\big)^2 + s^{-1}\boldsymbol{K}(Z,Z')^2\right)\left(\psi(\boldsymbol{f}(Z)) + \psi(\boldsymbol{f}(Z'))\right)\right] \\
&= \frac{1}{4}\,\mathbb{E}\,\mathrm{tr}\left[\left(s\,\mathbb{E}\left[\big(\boldsymbol{f}(Z) - \boldsymbol{f}(Z')\big)^2 \,\middle|\, Z\right] + s^{-1}\,\mathbb{E}\left[\boldsymbol{K}(Z,Z')^2 \,\middle|\, Z\right]\right)\psi(\boldsymbol{f}(Z))\right].
\end{aligned}$$

The equality follows from the fact that (Z, Z') is an exchangeable pair. This step represents a discrete version of the chain rule inequality (6.1). Second, they use the exponential ergodicity of the constructed Markov chain to control the kernel $\boldsymbol{K}(Z, Z')$ in terms of the matrix variance proxy $\boldsymbol{V}(\boldsymbol{f})$ that is defined in (4.2). This step implicitly relies on a discrete version of the local ergodicity condition, Proposition 2.7(2). Last, they use the same decoupling techniques as in our arguments to obtain prior bounds of the desired trace moments on the size of the kernel.

From this discussion, one can see that the Stein's method for establishing concentration inequalities is essentially an implementation of the semigroup approach in the discrete setting. So far, this method only applies to product measures. For comparison, our semigroup method is more general and technically simpler because it can directly apply the ergodicity theories of many well-studied Markov processes. Nevertheless, our work is strongly inspired by the tools and techniques developed by Paulin et al. [PMT16].

References

[ABY20] R. Aoun, M. Banna, and P. Youssef. Matrix Poincaré inequalities and concentration. *Adv. Math.*, 371:107251, 2020. MR-4108222

[AS94] S. Aida and D. Stroock. Moment estimates derived from Poincaré and logarithmic Sobolev inequalities. *Math. Res. Lett.*, 1(1):75–86, 1994. MR-1258492

[ASZ09] L. Ambrosio, G. Savaré, and L. Zambotti. Existence and stability for Fokker–Planck equations with log-concave reference measure. *Probab. Theory Related Fields*, 145(3-4):517–564, 2009. MR-2529438

[BGL13] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*. Springer, 2013. MR-3155209

[Bha97] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997. MR-1477662

[Buc01] A. Buchholz. Operator Khintchine inequality in non-commutative probability. *Math. Ann.*, 319(1):1–16, 2001. MR-1812816

[Caf00] L. A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.*, 214(3):547–563, 2000. MR-1800860

[Car10] E. Carlen. Trace inequalities and quantum entropy: an introductory course. In *Entropy and the quantum*, volume 529 of *Contemp. Math.*, pages 73–140. Amer. Math. Soc., Providence, RI, 2010. MR-2681769

[CE08] J. Cheeger and D. G. Ebin. *Comparison theorems in Riemannian geometry*. AMS Chelsea Publishing, Providence, RI, 2008. Revised reprint of the 1975 original. MR-2394158

[CH16] H.-C. Cheng and M.-H. Hsieh. Characterizations of matrix and operator-valued Φ-entropies, and operator Efron–Stein inequalities. *Proc. Royal Soc. A*, 472(2187):20150563, 2016. MR-3488705

[CH19] H.-C. Cheng and M.-H. Hsieh. Matrix Poincaré, Φ-Sobolev inequalities, and quantum ensembles. *J. Math. Phys.*, 60(3):032201, 2019. MR-3922774

[Cha05] S. Chatterjee. *Concentration inequalities with exchangeable pairs*. ProQuest LLC, Ann Arbor, MI, 2005. Thesis (Ph.D.)–Stanford University. MR-2707160

[CHT17] H.-C. Cheng, M.-H. Hsieh, and M. Tomamichel. Exponential decay of matrix Φ-entropies on Markov semigroups with applications to dynamical evolutions of quantum ensembles. *J. Math. Phys.*, 58(9):092202, 2017. MR-3702668

[CT14] R. Y. Chen and J. A. Tropp. Subadditivity of matrix φ-entropy and concentration of random matrices. *Electron. J. Probab*, 19(27):1–30, 2014. MR-3174839

[GKS20] A. Garg, T. Kathuria, and N. Srivastava. Scalar Poincaré implies matrix Poincaré, 2020.

[Gro07] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Modern Birkhäuser Classics. Birkhäuser Boston, Inc., Boston, MA, english edition, 2007. Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates. MR-2307192

[Hai16] M. Hairer. Convergence of Markov processes, 2016. Lecture notes.

[HNWTW20] D. Huang, J. Niles-Weed, J. A. Tropp, and R. Ward. Matrix concentration for products, 2020.

[HRMS18] S. R. Howard, A. Ramdas, J. McAuliffe, and J. Sekhon. Exponential line-crossing inequalities, 2018.

[HT20a] D. Huang and J. A. Tropp. From Poincaré inequalities to nonlinear matrix concentration, March 2020.

[HT20b] D. Huang and J. A. Tropp. Nonlinear matrix concentration via semigroup methods, 2020. arXiv:2006.16562.

[Hua19] D. Huang. A generalized Lieb's theorem and its applications to spectrum estimates for a sum of random matrices. *Linear Algebra Appl.*, 579:419–448, 2019. MR-3968118

[HW20] A. Henriksen and R. Ward. Concentration inequalities for random matrix products. *Linear Algebra Appl.*, 594:81–94, 2020. MR-4067998

[JSY19] M. Ji, Z. Shen, and Y. Yi. Convergence to equilibrium in Fokker–Planck equations. *J. Dynam. Differential Equations*, 31(3):1591–1615, 2019. MR-3992083

[JZ15] M. Junge and Q. Zeng. Noncommutative martingale deviation and Poincaré type inequalities with applications. *Probab. Theory Related Fields*, 161(3-4):449–507, 2015. MR-3334274

[Led92] M. Ledoux. A heat semigroup approach to concentration on the sphere and on a compact Riemannian manifold. *Geom. Funct. Anal.*, 2(2):221–224, 1992. MR-1159831

[Led01] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001. MR-1849347

[Lee18] J. M. Lee. *Introduction to Riemannian manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, Cham, 2018. MR-3887684

[LP86] F. Lust-Piquard. Inégalités de Khintchine dans C_p $(1 < p < \infty)$. *C. R. Math. Acad. Sci. Paris*, 303:289–292, 1986. MR-0859804

[MJC+14] L. Mackey, M. I. Jordan, R. Y. Chen, B. Farrell, and J. A. Tropp. Matrix concentration inequalities via the method of exchangeable pairs. *Ann. Probab.*, 42(3):906–945, 05 2014. MR-3189061

[Oli09] R. I. Oliveira. Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges, 2009.

[Oli10] R. I. Oliveira. Sums of random Hermitian matrices and an inequality by Rudelson. *Electron. Commun. Probab.*, 15:203–212, 2010. MR-2653725

[Pet16] P. Petersen. *Riemannian geometry*, volume 171 of *Graduate Texts in Mathematics*. Springer, Cham, 3rd edition, 2016. MR-3469435

[Pis86] G. Pisier. Probabilistic methods in the geometry of Banach spaces. In *Probability and analysis (Varenna, 1985)*, volume 1206 of *Lecture Notes in Math.*, pages 167–241. Springer, Berlin, 1986. MR-0864714

[PMT16] D. Paulin, L. Mackey, and J. A. Tropp. Efron–Stein inequalities for random matrices. *Ann. Probab.*, 44(5):3431–3473, 2016. MR-3551202

[Pré73] A. Prékopa. On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)*, 34:335–343, 1973. MR-0404557

[PX97] G. Pisier and Q. Xu. Non-commutative martingale inequalities. *Comm. Math. Phys.*, 189(3):667–698, 1997. MR-1482934

[Rud99] M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164(1):60–72, 1999. MR-1694526

[Sch99] M. Schmuckenschläger. Curvature of nonlocal Markov generators. In *Convex geometric analysis (Berkeley, CA, 1996)*, volume 34 of *Math. Sci. Res. Inst. Publ.*, pages 189–197. Cambridge Univ. Press, Cambridge, 1999. MR-1665591

[Ste72] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, pages 583–602, Berkeley, Calif., 1972. Univ. California Press. MR-0402873

[Ste86] C. Stein. Approximate computation of expectations. *Lecture Notes-Monograph Series*, 7:i–164, 1986. MR-0882007

[Str84] D. W. Stroock. *An introduction to the theory of large deviations*. Universitext. Springer-Verlag, New York, 1984. MR-0755154

[SW14] A. Saumard and J. A. Wellner. Log-concavity and strong log-concavity: a review. *Stat. Surv.*, 8:45, 2014. MR-3290441

[Tro11] J. A. Tropp. Freedman's inequality for matrix martingales. *Electron. Commun. Probab.*, 16:262–270, 2011. MR-2802042

[Tro12] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012. MR-2946459

[Tro15] J. A. Tropp. An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.*, 8(1-2):1–230, 2015. MR-3311439

[Var85] N. T. Varopoulos. Hardy–Littlewood theory for semigroups. *J. Funct. Anal.*, 63(2):240–260, 1985. MR-0803094

[vH16] R. van Handel. Probability in high dimension. APC 550 lecture notes, Princeton Univ., 2016.

Acknowledgments. We thank Ramon van Handel for his feedback on an early version of this manuscript. He is responsible for the observation and proof that matrix Poincaré inequalities are equivalent with scalar Poincaré inequalities, and we are grateful to him for allowing us to incorporate these ideas. We appreciate the thoughtful feedback from the anonymous referees, which has helped us streamline this paper.

DH was funded by NSF grants DMS-1907977 and DMS-1912654. JAT gratefully acknowledges funding from ONR awards N00014-17-12146 and N00014-18-12363, and he would like to thank his family for their support in these difficult times.