Matrix Poincaré inequalities and concentration

Richard Aoun, Marwa Banna, Pierre Youssef

MATRIX POINCARÉ INEQUALITIES AND CONCENTRATION

RICHARD AOUN, MARWA BANNA, AND PIERRE YOUSSEF

Abstract. We show that any probability measure satisfying a Matrix Poincaré inequality with respect to some reversible Markov generator satisfies an exponential matrix concentration inequality depending on the associated matrix carré du champ operator. This extends to the matrix setting a classical phenomenon in the scalar case. Moreover, the proof gives rise to new matrix trace inequalities which could be of independent interest. We then apply this general fact by establishing matrix Poincaré inequalities to derive matrix concentration inequalities for Gaussian measures, product measures and for Strong Rayleigh measures. The latter represents the first instance of matrix concentration for general matrix functions of negatively dependent random variables.

1. Introduction and main results

Concentration inequalities are versatile tools which found use in several pure and applied mathematical problems. While in their essence, these inequalities are just a quantification of the law of large numbers, they represent further illustrations of deep high dimensional phenomena in areas across Mathematics. On a conceptual level, they assert that a random variable measurable with respect to a large number of independent (or weakly dependent) random variables and "depending little" on each individually, is almost constant with high probability. While many methods were developed to prove concentration inequalities, perhaps the most insightful one is the one based on functional inequalities such as Poincaré and log-Sobolev inequalities. Indeed, these functional inequalities serve as a further illustration of the conceptual description we mentioned above. For instance, in its classical form, Poincaré inequality relates the variance of a function of a random variable to the average length of the gradient of the function. One then readily sees that if a function varies little locally (in the usual sense of variations), then with high probability it also varies little when evaluated at a random point. More generally, these inequalities relate statistics of the measure, such as the variance and entropy of any function, to the derivative along a semi-group associated with a Markov process generating the measure. Beside the concentration phenomenon, such functional inequalities provide a further understanding of the measure as they are intimately connected to the convergence rate of the corresponding Markov process generating it. The interconnection between concentration and functional inequalities is by now very well understood, and the use of such inequalities to derive concentration proved to be very powerful due to its flexibility in dealing with any measure, provided one can architect a suitable Markov process which generates it. We refer to [5] and [20] for more on classical concentration inequalities.

Matrix concentration inequalities are noncommutative extensions of their scalar counterpart and have been extensively developed in the last decade [1, 23, 25, 28, 29]. A big effort was made to transfer our understanding of the scalar case to the matrix one. In this direction, many papers were devoted to extending scalar methods for deriving concentration inequalities in the matrix setting. We refer to the book of Tropp [30] for a detailed introduction to the subject and an extensive list of references. As mentioned in the previous paragraph, the approach leading to concentration inequalities based on functional inequalities has been successful in the scalar case: it allowed to establish concentration inequalities in dependent settings and treat general functions beyond the example of sums of random variables. In light of this, several efforts were made to

extend this theory to the matrix setting. Several papers were devoted to properly defining the matrix entropy and establishing its basic properties such as the subadditivity [6, 7, 8, 9, 10]. In [7], Chen and Tropp aimed at extending ϕ-Sobolev inequalities to the matrix setting. In [8, 9], Cheng and Hsieh further investigated the notion of Φ-entropy for operator valued functions and established several matrix functional inequalities such as a matrix analogue of the Efron-Stein inequality. The subadditivity of matrix entropy was used in [7] with the aim of developing a matrix version of the entropy method. However, as was noted in [7], the attempt to adapt the Herbst's argument had some shortcomings requiring additional assumptions to be successfully implemented. In addition to [7], we should mention the paper [25] where a family of exponential matrix Efron-Stein inequalities are established and turned into matrix concentration. However, the literature lacks a unified general framework for relating matrix functional inequalities to matrix concentration.

In this paper, we focus on Poincaré inequalities and aim to implement a general procedure turning a *matrix* Poincaré inequality into a concentration inequality. In the scalar case, such a procedure was first introduced by Gromov and Milman [14] and alternative arguments were later developed. We will adapt here the approach of Aida and Stroock [2] (see also [19, Section 2.5]). One of our contributions is to elaborate such a procedure, then establish matrix Poincaré inequalities and use this to derive new matrix concentration inequalities for Gaussian measures and negatively dependent measures.

Matrix Poincaré implies matrix concentration. Let us denote by $\mathcal{H}_d$ (resp. $\mathcal{H}_d^+$) the set of $d \times d$ Hermitian (resp. positive semi-definite) matrices. Given a probability measure μ on some Polish space Ω and $f : \Omega \to \mathcal{H}_d$ whose matrix coefficients belong to $\mathrm{L}^2(\mu)$, the variance of f is given by

$$\mathrm{Var}_\mu(f) = \mathbb{E}_\mu[f^2] - (\mathbb{E}_\mu f)^2,$$

where $\mathbb{E}_\mu := \int f d\mu$. It can be easily checked that $\mathrm{Var}_\mu(f) \succeq 0$, where $\succeq$ refers to the positive semi-definite ordering. We will say that μ satisfies a *matrix Poincaré inequality* with constant α and *matrix Markov generator* $\mathcal{L}$ with Dirichlet domain $\mathcal{D}(\mathcal{L})$ if for any $f \in \mathcal{D}(\mathcal{L})$ we have

$$\mathrm{Var}_\mu(f) \preceq \alpha \mathcal{E}(f),$$

where $\mathcal{E}(f) = -\mathbb{E}_\mu[f\mathcal{L}f]$ is the *matrix Dirichlet form* associated with $\mathcal{L}$. As is verified in Proposition 2.2, this definition makes sense as $\mathcal{E}(f) \in \mathcal{H}_d^+$. The notion of matrix Poincaré inequality (with respect to the positive semi-definite ordering) appears in the works of Chen and Hsieh [8, 9] although not expressed in the language of semigroups. Together with Tomamichel [10], the aforementioned authors later developed the basic definitions and properties of semigroups acting on matrix functions, as well as the matrix Dirichlet form and matrix carré du champ operator, which we use here. These notions will be recalled in Section 2 for completeness, and can be thought of at this stage as natural extensions of their scalar counterpart by considering the action of the Markov generator on each entry of the matrix valued function. Let Γ be the matrix carré du champ operator associated to the matrix generator $\mathcal{L}$ defined on an algebra $\mathcal{A}$ of $\mathcal{D}(\mathcal{L})$. Our first main result states as follows.

Theorem 1.1. *Let μ be a probability measure on some Polish space Ω. Suppose that μ satisfies a matrix Poincaré inequality with constant α and matrix Markov generator $\mathcal{L}$ reversible with respect to μ. Then for any $f \in \mathcal{A}$ and $t \geq 0$, we have*

$$\mu\Big(\lambda_{max}\big(f - \mathbb{E}_\mu f\big) \geq t\Big) \leq d \exp\left(- \frac{t^2}{2\alpha v_f + t\sqrt{2\alpha v_f}} \right),$$

where $v_f = \big\| \|\Gamma(f)\| \big\|_{L_\infty}$.

The above theorem provides a general machinery turning a matrix Poincaré inequality into a corresponding matrix concentration inequality. In view of such general phenomenon paralleling

its scalar counterpart, establishing a matrix concentration inequality is reduced to proving a matrix Poincaré inequality. To this aim, for a given probability measure, the main task lies in designing the appropriate Markov generator and calculating the corresponding matrix carré du champ operator.

The proof of Theorem 1.1 hides many challenging obstacles arising because of noncommutativity. As is customary, obtaining a concentration inequality follows by combining a Chernoff bound with an estimate on the Laplace transform. The Poincaré inequality is then used to obtain a recursive relation involving the Laplace transform, which when properly arranged produces the desired bound on the Laplace transform. This simple looking procedure carries a great amount of difficulties when one attempts to extend it to the matrix setting. For instance, given $g : \Omega \to \mathcal{H}_d$, one starts applying the matrix Poincaré inequality to e^g to get

$$\mathbb{E}_\mu[e^{2g}] \preceq (\mathbb{E}_\mu e^g)^2 + \alpha \mathcal{E}(e^g).$$

In the scalar case, $\mathcal{E}(e^g)$ can be easily related to the Laplace transform which would automatically translates the above relation into a recursive formula on the Laplace transform. Such a relation is far from trivial in the matrix setting and requires the development of new matrix trace inequalities which could be of independent interest. Such a relation is established in Section 3 (see Theorem 3.1) where, in particular, a new matrix trace inequality is elaborated (see Theorem 3.3). While obtaining a recursive formula is the end of the story in the scalar case, such a relation cannot be directly iterated in the matrix setting. Indeed, since the square function is not operator monotone, one cannot reapply the same procedure to bound $(\mathbb{E}e^g)^2$ in the above formula. To overcome this issue, we exploit the operator monotonicity of the trace of such functions and combine it with special convexity arguments to implement the iterative procedure.

Matrix Poincaré and Concentration for product measures. We derive a matrix Poincaré inequality for the standard Gaussian measure and use the mechanism in Theorem 1.1 to deduce corresponding concentration. To this aim, we consider the Ornstein-Uhlenbeck Markov process whose matrix Dirichlet form is precisely the expectation of the sum of the squared partial derivatives matrices. In this case, we obtain the following matrix Gaussian Poincaré inequality.

Theorem 1.2. *Let $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ be the standard Gaussian measure on $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathcal{H}_d$ be such that all of its matrix coefficients, together with their partial derivatives, are smooth and in $\mathrm{L}^2(\mu)$. Then*

$$\mathrm{Var}_\mu(f) \preceq \int \sum_{i=1}^n (\partial_i f)^2 d\mu,$$

where $\partial_i f(x_1, \ldots, x_n) := \frac{\partial f}{\partial x_i}(x_1, \ldots, x_n)$ is the matrix whose entries are the i-th partial derivatives of the corresponding entries of f.

The above inequality appears to be new and extends to the matrix setting the scalar Gaussian Poincaré inequality. Indeed, when f is a scalar function, the right-hand side is precisely the integral of the Euclidean norm squared of the gradient of f. A related Poincaré inequality for the Gaussian unitary ensemble was obtained in [9, Theorem 4] with the trace applied to both sides of the inequality. Combined with Theorem 1.1, the above statement implies the following concentration inequality.

Theorem 1.3. *Let $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ be the standard Gaussian measure on $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathcal{H}_d$ be such that all of its matrix coefficients are smooth and in $\mathrm{L}^2(\mu)$. Then for any $t \in \mathbb{R}_+$,*

$$\mu\Big(\lambda_{max}\big(f - \mathbb{E}_\mu f\big) \geq t\Big) \leq d \exp\bigg(-\frac{t^2}{2v_f + t\sqrt{2v_f}}\bigg),$$

where $v_f = \sup_{(x_1,\ldots,x_n)\in\mathbb{R}^n} \big\| \sum_{i=1}^n (\partial_i f)^2(x_1, \cdots, x_n) \big\|$ with $\partial_i f(x_1, \ldots, x_n) = \frac{\partial f}{\partial x_i}(x_1, \ldots, x_n)$.

The interesting feature in the above theorem is that it captures concentration in terms of the variations of the matrix function, in the usual sense of variations. We couldn't locate a comparable result in the literature, as previous matrix concentration inequalities designed specifically for the Gaussian measure dealt with matrix Gaussian series; i.e. $f(x_1, \dots, x_n) = \sum_{i=1}^n x_i A_i$ for some deterministic $A_1, \dots, A_n \in \mathcal{H}_d$. We should note that for this particular example, the literature contains sub-Gaussian bounds on matrix concentration (see [30, Chapter 4]).

In Section 5, we further illustrate this procedure by investigating general product measures. To this aim, we prove a corresponding matrix Poincaré inequality (Theorem 5.3) and derive an exponential matrix concentration inequality (Theorem 5.1). In this setting, the matrix Poincaré inequality is equivalent to the matrix Efron-Stein inequality [8, Theorem 5.1]. We provide an alternative proof of this by building an appropriate Markov process.

We should note that, as in the scalar case, the approach based on Poincaré inequalities cannot lead to sub-Gaussian bounds on concentration. A possible approach to deriving sub-Gaussian bounds would be the elaboration of matrix log-Sobolev inequalities and of a general procedure turning these into corresponding matrix concentration inequalities. As of this writing, such procedure remains a challenging task and it is not clear how it could be implemented.

Matrix Poincaré and concentration for SCP measures. Concentration inequalities become increasingly more challenging without the independence structure. The matrix setting adds another layer of difficulty to the problem. In view of this, it is not surprising that there are few matrix concentration inequalities in the dependent case with [3, 22] only dealing with sums of random matrices, while in the works [24, 25] a matrix bounded difference inequality was established under a form of weak dependence.

The interesting feature in Theorem 1.1 is its ability to deal with any probability measure μ, the main remaining task lies in the construction of a suitable Markov process having μ as its stationary measure. With this perspective in mind, we are able to establish a matrix concentration inequality for functions of negatively dependent random variables. More precisely, we prove in Theorem 1.4 a matrix Poincare inequality for any homogeneous probability measure on the n-dimensional unit cube satisfying a form of negative dependence known as the *stochastic covering property* (SCP). Combined with Theorem 1.1, this implies a corresponding matrix exponential concentration inequality. In the scalar case, sub-Gaussian concentration bounds were obtained by Pemantle–Peres [26] and Hermon–Salez [16] who also established a modified log-Sobolev inequality. The proof of Theorem 1.4 relies on the approach of [16] and extends it to the matrix setting.

The stochastic covering property was put forward in [26] as a form of negative dependence. Indeed, it was shown in [26] that the *strong Rayleigh property* implies SCP. The class of strong Rayleigh measures was introduced by Borcea–Brändén–Liggett in [4] with the aim of building a theory of negative dependence. One of the main features of this class is its stability under many natural operation such as conditioning and projecting. Due to this, the strong Rayleigh property, which implies negative association, is more commonly used. Moreover, the class of strong Rayleigh measures (and thus the ones satisfying SCP) contains several interesting examples, such as: determinantal measures and point processes, independent Bernoullis conditioned on their sum, measures obtained by running the exclusion dynamics from a deterministic state.

We begin by recalling the definition of the stochastic covering property. Let $n \in \mathbb{N}$. We equip the n-dimensional discrete unit cube $\{0,1\}^n$ with the partial order $\preceq$ defined by

$$x \preceq y \iff x = y \quad \text{or} \quad \exists!\ i \leq n; y_i = x_i + 1.$$

We extend this order to the set of probability measures on $\{0,1\}^n$ in the following way. If μ_1 and μ_2 are two probability measures on $\{0,1\}^n$, we write $\mu_1 \preceq \mu_2$ if there exists a coupling κ on $\{0,1\}^n \times \{0,1\}^n$ such that $\text{support}(\kappa) \subseteq \{(x,y) \in \{0,1\}^n \times \{0,1\}^n : x \preceq y\}$.

Fix now $k \in \{1, \cdots, n\}$. Let μ be a probability measure on $\{0,1\}^n$ and ξ be a random variable on Ω with distribution μ. We say that μ is *k-homogeneous* if $\mathbb{P}(\sum_{i=1}^n \xi_i = k) = 1$, that is, μ is a probability measure on $\Omega := \{x \in \{0,1\}^n : \sum_{i=1}^n x_i = k\}$. We say that μ has the *stochastic covering property* if for any subset S of $\{1, \cdots, n\}$, and any $x_S, y_S \in \{0,1\}^S$,

$$x_S \preceq y_S \Longrightarrow \mu\left(\cdot \mid \xi_S = y_S\right) \preceq \mu\left(\cdot \mid \xi_S = x_S\right),$$

where ξ_S denotes the restriction of ξ to the coordinates indexed by S. We should note that $\mu\left(\cdot \mid \xi_S = y_S\right)$ is a probability measure on $\{0,1\}^{S^c}$ supported on vectors z_{S^c} satisfying $\|z_{S^c}\|_1 = k - \|y_S\|_1$, where $\|\cdot\|_1$ stands for the ℓ_1-norm (here equal to the number of coordinates equal to one). We are now able to state the matrix Poincaré inequality for SCP measures.

Theorem 1.4. *Let μ be a k-homogeneous probability measure on $\Omega := \{0,1\}^n$ with the SCP property and denote by $\widetilde{\Omega} = \{(x,y) \in \Omega^2 : x$ and y differ on exactly 2 coordinates$\}$. Then there exists a Markov generator Q supported on $\widetilde{\Omega}$ and satisfying $\max\{-Q(x,x) : x \in \Omega\} \leq 1$, such that for any $f : \Omega \to \mathcal{H}_d$, we have*

$$\mathrm{Var}_\mu(f) \preceq 2k\mathcal{E}(f),$$

where $\mathcal{E}$ is the Dirichlet form associated with Q.

We refer to Section 6 where an explicit expression of the Markov generator is given. The result above states that any probability measure with the SCP property satisfies a matrix Poincaré inequality with normalized Markov generator and constant $2k$. The analogous result in the scalar case was recently established in [16]. While in [16] a scalar Poincaré inequality (and a modified log-Sobolev inequality) is derived by means of an induction method introduced in [21], we extract from this inductive procedure the explicit Markov generator and prove the matrix Poincaré inequality directly using operator convexity arguments. Combined with Theorem 1.1, this implies the following matrix concentration inequality.

Theorem 1.5. *Let μ be a k-homogeneous probability measure on $\Omega = \{0,1\}^n$ with the SCP property. Let $f : \Omega \to \mathcal{H}_d$ be a 1-Lipschitz matrix function in the sense that for any $x, y \in \Omega$,*

$$\|f(x) - f(y)\| \leq \|x - y\|_1,$$

where $\|\cdot\|$ stands for the operator norm. Then, for any $t \in \mathbb{R}_+$,

$$\mu\Big(\lambda_{max}\big(f - \mathbb{E}_\mu f\big) \geq t\Big) \leq d \exp\Big(-\frac{t^2}{8k + 2t\sqrt{2k}}\Big).$$

A special case when $f(x_1, \ldots, x_n) = \sum_{i=1}^n x_i A_i$ for some $A_1, \ldots, A_n \in \mathcal{H}_d$, was recently investigated in [18] where, up to a logarithmic term, a sub-Gaussian concentration bound is derived. The method developed in [18] extends to the matrix setting the martingale approach elaborated in [26]. Once again, while our matrix Poincaré inequality cannot yield sub-Gaussian concentration, it provides a matrix concentration valid for any matrix Lipschitz function while the results in [18] are only concerned with the specific example provided by $f(x_1, \ldots, x_n) = \sum_{i=1}^n x_i A_i$ for some $A_1, \ldots, A_n \in \mathcal{H}_d$.

The paper is organized as follows. In Section 2, we recall the notion of matrix Markov generator, its associated matrix Dirichlet form and carré du champ operator and their properties. In section 3, we establish the relation between the matrix Dirichlet form and the Laplace transform. Section 4 is devoted to the proof of Theorem 1.1. In Section 5, we investigate the matrix Poincaré inequality for product measures and in particular for the standard Gaussian measure. Finally, Section 6 is devoted to the study of the matrix Poincaré inequality for SCP measures.

Aknowledgement: The authors are grateful to the anonymous referee whose remarks and generous suggestions greatly improved the presentation. In particular, the current proofs of Theorem 3.3 and Lemma 3.4 were communicated by the referee and replace previously lengthy arguments. The authors would also like to thank De Huang and Lester Mackey for helpful comments. This work was initiated during visits of the authors to Université Paris Diderot and the American University of Beirut. We would like to thank these institutions for the excellent working conditions. P.Y. was supported by grant ANR-16-CE40-0024-01.

2. Matrix markov semi-group and generator

In [10, Section 3], the authors developed a framework of Markov semigroups on matrix-valued functions and defined a non-commutative version of the carré du champ operator and Dirichlet form. In this preliminary section, we recall these notions in our context for completeness and state the necessary properties which we will use. We refer to [10] for more on this topic.

Let Ω be a Polish space and $(X_t)_{t\geq 0}$ be a Markov process with stationary measure μ. Let $\mathrm{L}^2(\mu)$ be the Hilbert space of square integrable functions with respect to μ. The Markov semi-group $(P_t)_{t\geq 0}$ associated to $(X_t)_{t\geq 0}$ defines an operator on $\mathrm{L}^2(\mu)$ through the formula

$$P_t f(x) = \mathbb{E}[f(X_t) \mid X_0 = x],$$

for any $x \in \Omega$. Recall that the Markov process is said to be reversible if for every $f, g \in \mathrm{L}^2(\mu)$,

$$\langle f, P_t g\rangle_\mu = \langle P_t f, g\rangle_\mu,$$

where $\langle \cdot, \cdot \rangle_\mu$ is the usual inner product of $\mathrm{L}^2(\mu)$. It is said to be ergodic if $P_t f \underset{t\to+\infty}{\longrightarrow} \mathbb{E}_\mu(f)$ in $\mathrm{L}^2(\mu)$ for every $f \in \mathrm{L}^2(\mu)$.

To the Markov semi-group $(P_t)_{t\geq 0}$ is associated its infinitesimal generator $\mathcal{L}$ defined by

$$\mathcal{L} f = \lim_{t\to 0} \frac{P_t f - f}{t},$$

for every function $f \in \mathcal{D}(\mathcal{L})$, where $\mathcal{D}(\mathcal{L})$ is the $\mathrm{L}^2(\mu)$-domain of $\mathcal{L}$. In this paper, we are mostly interested in (Hermitian) matrix valued functions. The action of the Markov semi-group and that of the infinitesimal generator can be naturally extended to matrix valued functions by considering the action of the semi-group on each entry of the matrix valued function. Therefore, given a function $f : \Omega \to \mathcal{H}_d$ whose matrix coefficients belong to $\mathrm{L}^2(\mu)$ (or to $\mathcal{D}(\mathcal{L})$), we set similarly

$$P_t f(x) = \mathbb{E}[f(X_t) \mid X_0 = x],$$

for any $x \in \Omega$ and $\mathcal{L} f = \lim_{t\to 0} \frac{P_t f - f}{t}$. We will refer to P_t as *matrix Markov semi-group* and $\mathcal{L}$ as *matrix Markov generator* to emphasize that we will be interested in their action on matrix valued functions. By abuse of notation, the space of functions $f : \Omega \to \mathcal{H}_d$ whose matrix coefficients belong to $\mathcal{D}(\mathcal{L})$ will be still denoted by $\mathcal{D}(\mathcal{L})$. Below, we collect some of the basic properties paralleling their scalar counterpart and refer to [10] for their proofs.

Proposition 2.1. *Let Ω be a Polish space and $(X_t)_{t\geq 0}$ be a Markov process with stationary measure μ. The matrix Markov semi-group $(P_t)_{t\geq 0}$ and matrix Markov generator $\mathcal{L}$ satisfy the following elementary properties for every matrix valued functions $f, g \in \mathcal{D}(\mathcal{L})$:*

1. *P_t and $\mathcal{L}$ commute.*
2. *If $(X_t)_{t\in\mathbb{R}_+}$ is reversible then $\mathbb{E}_\mu[f\mathcal{L}(g)] = \mathbb{E}_\mu[\mathcal{L}(f)g]$.*
3. *$\mathbb{E}_\mu[\mathcal{L} f] = 0$.*
4. *$P_t f$ takes values in $\mathcal{H}_d^+$.*
5. *$(P_t f)^2 \preceq P_t f^2$.*
6. *If $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function, then*

$$\mathrm{Tr}\big(\phi(P_t f)\big) \leq \mathrm{Tr}\big(P_t \phi(f)\big).$$

We introduce the *matrix Dirichlet form* given by

$$\mathcal{E}(f) = -\mathbb{E}_\mu[f\mathcal{L}(f)],$$

for any $f \in \mathcal{D}(\mathcal{L})$. Given an algebra $\mathcal{A} \subseteq \mathcal{D}(\mathcal{L})$, define the *matrix carré du champ* operator by

$$\Gamma(f) := \frac{1}{2}\big(\mathcal{L}(f^2) - f\mathcal{L}(f) - \mathcal{L}(f)f\big),$$

for $f \in \mathcal{A}$, Let us note that already in the above definitions, we see the subtlety of the non-commutative nature of the objects manipulated. For instance, while in the scalar counterpart f and $\mathcal{L}(f)$ commute, this is no longer the case here and one needs to take into account this when dealing with the above notions.

It is not clear at first glance if the usual properties of the Dirichlet form and the carré du champ operator extend to their matrix counterparts. This is the case when the underlying Markov process is reversible as we verify in the next proposition.

Proposition 2.2. *Let Ω be a Polish space and $(X_t)_{t\geq 0}$ be a reversible Markov process with stationary measure μ. Then the matrix Dirichlet form $\mathcal{E}$ and the carré du champ operator Γ satisfy the following properties for every matrix valued function $f \in \mathcal{A}$:*

(1) *For any $a \in \mathbb{R}$, we have $\Gamma(af) = a^2\,\Gamma(f)$.*
(2) *$\mathcal{E}(f) = \mathbb{E}_\mu[\Gamma(f)]$.*
(3) *We have*

$$\Gamma(f)(x) = \lim_{t\to 0} \frac{\mathbb{E}\big[\big(f(X_t) - f(X_0)\big)^2 \mid X_0 = x\big]}{2t}$$

and

$$\mathcal{E}(f) = \lim_{t\to 0} \frac{\mathbb{E}\big[\big(f(X_t) - f(X_0)\big)^2\big]}{2t}, \quad X_0 \sim \mu.$$

In particular, $\Gamma(f) : \Omega \to \mathcal{H}_d^+$ and $\mathcal{E}(f) \in \mathcal{H}_d^+$.
(4) *If Ω is finite, then for any $x \in \Omega$ we have*

$$\Gamma(f)(x) = \frac{1}{2}\sum_{y\in\Omega} Q(x,y)\big(f(y) - f(x)\big)^2,$$

and

$$\mathcal{E}(f) = \frac{1}{2}\sum_{x,y\in\Omega} \mu(x)Q(x,y)\big(f(y) - f(x)\big)^2,$$

where Q is the $|\Omega| \times |\Omega|$ matrix representing the Markov generator.

Proof. The first property follows easily from the linearity of $\mathcal{L}$. To establish the second property, note that

$$\mathbb{E}_\mu[\Gamma(f)] = \frac{1}{2}\big(\mathbb{E}_\mu[\mathcal{L}(f^2)] - \mathbb{E}_\mu[f\mathcal{L}(f)] - \mathbb{E}_\mu[\mathcal{L}(f)f]\big).$$

Using that $\mathbb{E}_\mu[\mathcal{L}(f^2)] = 0$ and that $\mathbb{E}_\mu[f\mathcal{L}(f)] = \mathbb{E}_\mu[\mathcal{L}(f)f]$ by reversibility, we get the claim. To establish the third, we use the definition of $\mathcal{L}$ to explicitly write

$$\begin{aligned}\Gamma(f)(x) &= \lim_{t\to 0} \frac{\mathbb{E}[f^2(X_t) \mid X_0 = x] - f^2(x)}{2t} - \lim_{t\to 0} \frac{f(x)\big(\mathbb{E}[f(X_t) \mid X_0 = x] - f(x)\big)}{2t} \\ &\quad - \lim_{t\to 0} \frac{\big(\mathbb{E}[f(X_t) \mid X_0 = x] - f(x)\big)f(x)}{2t} \\ &= \lim_{t\to 0} \frac{\mathbb{E}[f^2(X_t) \mid X_0 = x] + f^2(x) - f(x)\mathbb{E}[f(X_t) \mid X_0 = x] - \mathbb{E}[f(X_t) \mid X_0 = x]f(x)}{2t} \\ &= \lim_{t\to 0} \frac{\mathbb{E}\big[\big(f(X_t) - f(X_0)\big)^2 \mid X_0 = x\big]}{2t},\end{aligned}$$

which establishes the claim for $\Gamma(f)$. Since $\mathcal{E}(f) = \mathbb{E}_\mu[\Gamma(f)]$, the expression of $\mathcal{E}(f)$ readily follows. From these representations, it is clear that $\Gamma(f)$ and $\mathcal{E}$ take values in $\mathcal{H}_d^+$.

Finally, when Ω is finite, using the above and defining $h_x : \Omega \to \mathcal{H}_d$ by $h_x(y) = \big(f(y) - f(x)\big)^2$, we get

$$\Gamma(f)(x) = \lim_{t\to 0} \frac{(P_t h_x)(x)}{2t} = \frac{1}{2}(Qh_x)(x) = \frac{1}{2}\sum_{y\in\Omega} Q(x,y)h_x(y),$$

which proves the expression of $\Gamma(f)$ in last property of the proposition. It remains to use that $\mathcal{E}(f) = \mathbb{E}_\mu[\Gamma(f)]$ to derive the expression of $\mathcal{E}(f)$ and finish the proof. □

We collect furthermore some useful identities connecting the variance and the Dirichlet form. Unlike the scalar case, the following identity requires reversibility of the Markov process.

Lemma 2.3. *Let Ω be a Polish space and $(X_t)_{t\geq 0}$ be a* reversible *Markov process with stationary measure μ. Then for any $f \in \mathcal{D}(\mathcal{L})$, we have*

$$\frac{d}{dt}\mathrm{Var}_\mu(P_t f) = -2\mathcal{E}(P_t f).$$

Moreover, if the Markov process is ergodic then

$$\mathrm{Var}_\mu(f) = 2\int_0^\infty \mathcal{E}(P_t f)\, dt.$$

Proof. We start by proving the first identity. As $\mathbb{E}_\mu[P_t f] = \mathbb{E}_\mu[f]$ then

$$\frac{d}{dt}\mathrm{Var}_\mu[P_t f] = \frac{d}{dt}\mathbb{E}_\mu[(P_t f)^2] = \mathbb{E}_\mu\Big[\Big(\frac{d}{dt}P_t f\Big)P_t f\Big] + \mathbb{E}_\mu\Big[P_t f\Big(\frac{d}{dt}P_t f\Big)\Big].$$

Noting that $\frac{d}{dt}P_t f = \mathcal{L}(P_t f)$ and using reversibility, we get

$$\frac{d}{dt}\mathrm{Var}_\mu[P_t f] = \mathbb{E}_\mu\Big[\mathcal{L}(P_t f)P_t f\Big] + \mathbb{E}_\mu\Big[P_t f\mathcal{L}(P_t f)\Big] = 2\mathbb{E}_\mu\Big[P_t f\mathcal{L}(P_t f)\Big] = -2\mathcal{E}(P_t f).$$

In order to prove the second identity, we shall first prove that $\mathrm{Var}_\mu[P_t f]$ converges to zero as t goes to infinity. In fact, the ergodicity of the Markov process implies that, for any $f : \Omega \to \mathcal{H}_d$ and any $i, j \in \{1, \dots, d\}$, $(P_t f)_{ij}$ converges to $(\mathbb{E}_\mu[f])_{ij}$ in $L^2(\mu)$. Then for any $i, j = 1, \dots d$,

using Cauchy-Schwarz inequality

$$\begin{aligned}\big|(\mathrm{Var}_\mu[P_tf])_{ij}\big| &= \Big|\sum_{k=1}^d \mathbb{E}_\mu[(P_tf-\mathbb{E}_\mu[f])_{ik}(P_tf-\mathbb{E}_\mu[f])_{kj}]\Big| \\ &\le \sum_{k=1}^d \big(\mathbb{E}_\mu[(P_tf-\mathbb{E}_\mu[f])_{ik}^2]\big)^{\frac12}\big(\mathbb{E}_\mu[(P_tf-\mathbb{E}_\mu[f])_{kj}^2]\big)^{\frac12} \xrightarrow[t\to\infty]{} 0\end{aligned}$$

and hence $\mathrm{Var}_\mu[P_tf] \xrightarrow[t\to\infty]{} 0$. Therefore, we get by the fundamental theorem of calculus

$$\mathrm{Var}_\mu[f] = \mathrm{Var}_\mu[P_0f] - \lim_{t\to\infty}\mathrm{Var}_\mu[P_tf] = -\int_0^\infty \frac{d}{dt}\mathrm{Var}_\mu[P_tf]dt = 2\int_0^\infty \mathcal{E}(P_tf,P_tf)\,dt.$$

□

3. Some matrix trace inequalities

The goal of this section is to establish the following trace inequality relating the matrix Dirichlet form to the Laplace transform.

Theorem 3.1. *Let Ω be a Polish space and $(X_t)_{t\ge0}$ be a reversible Markov process with stationary measure μ. Then, for any $g:\Omega\to\mathcal{H}_d$ belonging to the domain of Γ and any $p\ge1$,*

$$\mathrm{Tr}\Big[\big(\mathcal{E}(e^g)\big)^p\Big] \le \Big\|\|\Gamma(g)\|^p\Big\|_{L_\infty}\mathrm{Tr}\big(\mathbb{E}_\mu[e^{2pg}]\big),$$

where $\mathcal{E}$ and Γ are respectively the matrix Dirichlet form and matrix carré du champ operator associated with $(X_t)_{t\ge0}$.

The above theorem is the cornerstone of the argument relating the matrix poincaré inequality to matrix concentration. In order to prove this statement, we establish some matrix trace inequalities which we believe might be of independent interest. Before stating and proving these inequalities, let us mention a key ingredient which we will rely on. The following statement is a particular case of the results in [17], and provides an integral representation of the matrix logarithmic mean in terms of the matrix arithmetic mean.

Theorem 3.2. [17, Corollary 2.4] *There exists a probability measure ν on $\mathbb{R}$ such that the following holds. Let H,K be two $d\times d$ positive definite matrices and let $X\in\mathcal{H}_d$. Then*

$$\int_0^1 H^\tau X K^{1-\tau}\,d\tau = \int_{\mathbb{R}} H^{is}\frac{HX+XK}{2}K^{-is}\,d\nu(s).$$

The probability measure ν in the above statement can be made explicit (see equation 2.7 in [17]), however this won't be needed for our purposes. Let us mention that the above is only one of several integral representations established by Hiai and Kosaki in [17].

Let us recall that given a function $f:\mathbb{R}\to\mathbb{R}$, it can be extended to a function on Hermitian matrices by applying it to each eigenvalue of the matrix. More precisely, if $A=\sum_{i=1}^d\lambda_iv_iv_i^*$ is the spectral decomposition of $A\in\mathcal{H}_d$, then one defines $f(A)=\sum_{i=1}^d f(\lambda_i)v_iv_i^*$. The main trace inequality used to prove Theorem 3.1 is the following.

Theorem 3.3. *Let A be a $d\times d$ Hermitian (deterministic) matrix and let B be a $d\times d$ Hermitian random matrix. Then for every increasing, convex function $f:\mathbb{R}_+\to\mathbb{R}_+$,*

$$\mathrm{Tr}\Big[f\Big(\mathbb{E}[(e^A-e^B)^2]\Big)\Big] \le \frac12\mathrm{Tr}\Big[f\big(\mathbb{E}[(A-B)e^{2B}(A-B)]\big)\Big] + \frac12\mathrm{Tr}\Big[f\big(e^A\mathbb{E}[(A-B)^2]e^A\big)\Big].$$

In particular, the above holds for $f:t\to t^p$ when $p\ge1$.

Proof. First, note that $\frac{d}{d\tau}e^{\tau A}e^{(1-\tau)B} = e^{\tau A}(A-B)e^{(1-\tau)B}$. Therefore, we can write

$$e^A - e^B = \int_0^1 \frac{d}{d\tau} e^{\tau A} e^{(1-\tau)B}\, d\tau = \int_0^1 e^{\tau A}(A-B)e^{(1-\tau)B}\, d\tau.$$

Denoting $C := A - B$ and using Theorem 3.2, we have

$$e^A - e^B = \int_{\mathbb{R}} e^{isA}\frac{e^A C + Ce^B}{2} e^{-isB}\, d\nu(s) := \int_{\mathbb{R}} \Gamma(s)\, d\nu(s).$$

Using [11, Corollary 2.8], we have that

$$\int_{\mathbb{R}} \Gamma(s)\, d\nu(s) \cdot \int_{\mathbb{R}} \Gamma(s)^*\, d\nu(s) \preceq \int_{\mathbb{R}} \Gamma(s)\Gamma(s)^*\, d\nu(s).$$

Putting the above together, we get

$$\begin{aligned}(e^A - e^B)^2 &\preceq \int_{\mathbb{R}} e^{isA}\Big(\frac{e^A C + Ce^B}{2}\Big)\cdot\Big(\frac{e^A C + Ce^B}{2}\Big)^* e^{-isA}\, d\nu(s)\\ &\preceq \int_{\mathbb{R}} e^{isA}\frac{e^A C^2 e^A + Ce^{2B}C}{2} e^{-isA}\, d\nu(s),\end{aligned}$$

where the last inequality follows after using that $(e^A C + Ce^B)(e^A C + Ce^B)^* \preceq 2(e^A C)(e^A C)^* + 2(Ce^B)(Ce^B)^*$.

Since f is continuous, convex monotone increasing, then so is $\mathrm{Tr}\circ f$ (see for instance [6, Theorem 2.10]). Therefore, recalling that A is deterministic, we have

$$\begin{aligned}\mathrm{Tr}\Big[f\Big(\mathbb{E}[(e^A - e^B)^2]\Big)\Big] &\leq \mathrm{Tr}\Big[f\Big(\int_{\mathbb{R}} e^{isA}\mathbb{E}\Big[\frac{e^A C^2 e^A + Ce^{2B}C}{2}\Big]e^{-isA}\, d\nu(s)\Big)\Big]\\ &\leq \int_{\mathbb{R}} \mathrm{Tr}\Big[f\Big(e^{isA}\mathbb{E}\Big[\frac{e^A C^2 e^A + Ce^{2B}C}{2}\Big]e^{-isA}\Big)\Big]\, d\nu(s),\end{aligned}$$

where the first inequality uses the monotonicity of $\mathrm{Tr}\circ f$, and the second its convexity. Now note that e^{isA} is unitary, therefore using that $\mathrm{Tr}\circ f$ is unitary invariant, we get

$$\begin{aligned}\mathrm{Tr}\Big[f\Big(\mathbb{E}[(e^A - e^B)^2]\Big)\Big] &\leq \mathrm{Tr}\Big[f\Big(\mathbb{E}\Big[\frac{e^A C^2 e^A + Ce^{2B}C}{2}\Big]\Big)\Big]\\ &\leq \frac{1}{2}\mathrm{Tr}\Big[f\big(\mathbb{E}[(A-B)e^{2B}(A-B)]\big)\Big] + \frac{1}{2}\mathrm{Tr}\Big[f\big(e^A\mathbb{E}[(A-B)^2]e^A\big)\Big],\end{aligned}$$

where the last inequality follows by convexity of $\mathrm{Tr}\circ f$. □

The above theorem seems new even when both A and B are deterministic. For instance, we get that for any $p \in \mathbb{N}$

$$\mathrm{Tr}\Big[(e^A - e^B)^{2p}\Big] \leq \frac{1}{2}\mathrm{Tr}\Big[\big((A-B)^2 e^{2B}\big)^p\Big] + \frac{1}{2}\mathrm{Tr}\Big[\big(e^{2A}(A-B)^2\big)^p\Big].$$

The case $p = 1$ is related to some of the results in [25] even though it is incomparable to them and cannot be derived from them.

To deduce Theorem 3.1 from Theorem 3.3, we will need the following lemma which relies on an operator convexity inequality from [15].

Lemma 3.4. *Let $K \in \mathcal{H}_d$ and $Z \in \mathcal{H}_d^+$ be random matrices, not necessarily independent, and assume that $\mathbb{E}[K^2] \preceq Id$. Then for any $p \geq 1$, we have*

$$\mathbb{E}[KZK] \preceq \big(\mathbb{E}[KZ^pK]\big)^{\frac{1}{p}}.$$

In particular,

$$\mathrm{Tr}\Big[\big(\mathbb{E}[KZK]\big)^p\Big] \leq \mathbb{E}\mathrm{Tr}[KZ^pK].$$

Proof. By a truncation argument, we may suppose without loss of generality that Z is uniformly bounded. Let $\mathcal{A}$ be the C^*-algebra of uniformly bounded random Hermitian $d \times d$ matrices and let $K \in \mathcal{H}_d$ be such that $\mathbb{E}[K^2] \preceq Id$. Consider $\Phi : \mathcal{A} \to \mathcal{H}_d$ defined by

$$\Phi(Z) = \mathbb{E}[KZK].$$

Clearly, Φ is linear and positive i.e. $\Phi(Z) \succeq 0$ if $Z \succeq 0$, and $\Phi(0) = 0$. Moreover, if Z is uniformly bounded by 1, then using that $\mathbb{E}[K^2] \preceq Id$ we see that the operator norm of $\Phi(Z)$ is bounded by 1. Thus, Φ is a contraction and we can apply [15, Corollary 2.2] to deduce that for any $p \geq 1$ and any $Z \in \mathcal{A}$

$$\Phi(Z^{\frac{1}{p}}) \preceq \Phi(Z)^{\frac{1}{p}},$$

where we have used that the function $t \to t^{\frac{1}{p}}$ is operator concave for $p \geq 1$. The first claim then follows after a change of variables, while the second uses that $\mathrm{Tr} \circ (\cdot)^p$ is operator monotone. □

With the help of Lemma 3.4, we are now ready to show how Theorem 3.3 implies Theorem 3.1.

Proof of Theorem 3.1. Using Proposition 2.2, we start writing

$$\mathrm{Tr}\Big[\big(\mathcal{E}(e^g)\big)^p\Big] = \lim_{t\to 0} \frac{1}{(2t)^p} \mathrm{Tr}\Big[\big(\mathbb{E}\big(e^{g(X_t)} - e^{g(X_0)}\big)^2\big)^p\Big].$$

Let $(A, B) := \big(g(X_0), g(X_t)\big)$ and note that it follows from the reversibility of $(X_t)_{t\geq 0}$ that (A, B) is an exchangeable pair. Since $\mathrm{Tr}[(\cdot)^p]$ is operator convex, then Jensen's inequality implies that

$$\mathrm{Tr}\Big[\big(\mathcal{E}(e^g)\big)^p\Big] \leq \lim_{t\to 0} \frac{1}{(2t)^p} \mathbb{E}\mathrm{Tr}\Big[\big(\mathbb{E}\big[\big(e^A - e^B\big)^2 \mid A\big]\big)^p\Big]. \tag{1}$$

Denote $\alpha = \Big\|\mathbb{E}[(A-B)^2 \mid A]\Big\|_{L^\infty(A)}$. Applying Theorem 3.3 conditionally on A, we get

$$\mathbb{E}\mathrm{Tr}\Big[\big(\mathbb{E}[(e^A - e^B)^2 \mid A]\big)^p\Big] \leq \frac{1}{2}\mathbb{E}\mathrm{Tr}\Big[\big(\mathbb{E}[(A-B)e^{2B}(A-B) \mid A]\big)^p\Big] + \frac{1}{2}\mathrm{Tr}\Big[\big(e^A \mathbb{E}[(A-B)^2 \mid A] e^A\big)^p\Big].$$

Now note that $e^A \mathbb{E}[(A-B)^2 \mid A] e^A \preceq \|\mathbb{E}[(A-B)^2 \mid A]\|\, e^{2A}$ for any realization of A. Therefore, using the monotonicity of $\mathrm{Tr}[(\cdot)^p]$, we get that

$$\mathrm{Tr}\Big[\big(e^A \mathbb{E}[(A-B)^2 \mid A] e^A\big)^p\Big] \leq \alpha^p \,\mathrm{Tr}(e^{2pA}),$$

for all realizations of A. On the other hand, using Lemma 3.4 conditionally on A with $K = \frac{A-B}{\sqrt{\alpha}}$ and $Z = e^{2B}$, we have

$$\mathrm{Tr}\Big[\big(\mathbb{E}[(A-B)e^{2B}(A-B) \mid A]\big)^p\Big] \leq \|\mathbb{E}[(A-B)^2 \mid A]\|^{p-1} \mathrm{Tr}\Big[\mathbb{E}[(A-B)e^{2pB}(A-B) \mid A]\Big],$$

for all realizations of A. Combining these estimates, we deduce that

$$\begin{aligned}\mathbb{E}\mathrm{Tr}\Big[\big(\mathbb{E}[(e^A - e^B)^2 \mid A]\big)^p\Big] &\leq \frac{1}{2}\alpha^{p-1}\mathbb{E}\mathrm{Tr}\Big[\mathbb{E}[(A-B)e^{2pB}(A-B) \mid A]\Big] + \frac{1}{2}\alpha^p \,\mathrm{Tr}(e^{2pA}) \\ &= \frac{1}{2}\alpha^{p-1}\mathbb{E}\mathrm{Tr}\Big[(A-B)^2 e^{2pB}\Big] + \frac{1}{2}\alpha^p \,\mathbb{E}\mathrm{Tr}(e^{2pA}).\end{aligned}$$

Using the exchangeability of A and B, we have

$$\mathbb{E}\mathrm{Tr}\Big[(A-B)^2 e^{2pB}\Big] \leq \alpha\, \mathbb{E}\mathrm{Tr}(e^{2pA}).$$

Combining the two previous estimates, we deduce that

$$\mathbb{E}\mathrm{Tr}\Big[\big(\mathbb{E}[(e^A - e^B)^2 \mid A]\big)^p\Big] \leq \Big\|\mathbb{E}[(A-B)^2 \mid A]\Big\|_{L^\infty}^p \,\mathbb{E}\mathrm{Tr}(e^{2pA}).$$

Replacing A and B by their values, and plugging back the above inequality in (1), we get

$$\mathrm{Tr}\Big[\big(\mathcal{E}(e^g)\big)^p\Big] \leq \Big\| \lim_{t\to 0} \frac{1}{2t}\mathbb{E}[\big(g(X_t) - g(X_0)\big)^2 \mid X_0]\Big\|_{L^\infty}^p \,\mathbb{E}_\mu \mathrm{Tr}(e^{2pg}).$$

Using Proposition 2.2, we have

$$\lim_{t\to 0}\frac{1}{2t}\mathbb{E}[\big(g(X_t)-g(X_0)\big)^2 \mid X_0]=\Gamma(g),$$

which after replacement in the previous inequality, finishes the proof. □

4. From matrix Poincaré inequality to matrix concentration

The goal of this section is to prove Theorem 1.1. Like usual, providing a bound on the Laplace transform will be sufficient to derive the corresponding concentration inequality through the use of a Chernoff bound. More precisely, we prove the following.

Theorem 4.1. *Let μ be a probability measure on some Polish space Ω. Suppose that μ satisfies a matrix Poincaré inequality with constant α and matrix Markov generator $\mathcal{L}$ reversible with respect to μ. Given $f:\Omega\to\mathcal{H}_d$ belonging to the domain of Γ, denote*

$$v_f=\sup\big\|\Gamma(f)\big\|.$$

Suppose that $e^{\lambda f}$ has matrix coefficients in $\mathrm{L}^1(\mu)$ for any $\lambda>0$. Then, for any $\delta<\sqrt{\frac{2}{\alpha v_f}}$, we have

$$\mathbb{E}_\mu\mathrm{Tr}(e^{\delta(f-\mathbb{E}_\mu f)})\leq\frac{2d}{2-\alpha v_f\delta^2}.$$

With this bound in hand, we can easily derive the concentration inequality promised in Theorem 1.1.

Proof of Theorem 1.1. The proof follows the standard method initiated by Ahlswede-Winter [1]. Using Chernoff bound, we write

$$\mu\Big(\lambda_{max}\big(f-\mathbb{E}_\mu f\big)\geq t\Big)\leq\inf_{\delta\in\mathbb{R}_+}e^{-\delta t}\mathbb{E}_\mu\mathrm{Tr}(e^{\delta(f-\mathbb{E}_\mu f)}).$$

Using Theorem 4.1, we get

$$\mu\Big(\lambda_{max}\big(f-\mathbb{E}_\mu f\big)\geq t\Big)\leq\inf_{\delta<\sqrt{\frac{2}{\alpha v_f}}}\frac{2de^{-\delta t}}{2-\alpha v_f\delta^2}\leq\inf_{\delta<\sqrt{\frac{2}{\alpha v_f}}}d\cdot e^{-\delta t+\frac{\alpha v_f\delta^2}{2-\sqrt{2\alpha v_f}\delta}}.$$

Choosing $\delta=\frac{t}{\alpha v_f+t\sqrt{\alpha v_f/2}}$, we get the result. □

The rest of this section is devoted to the proof of Theorem 4.1. As explained in the introduction, the key is to establish some recursive relation involving the Laplace transform. To this aim, the result of the previous section relating the matrix Dirichlet form to the Laplace transform will play a crucial role in the derivation of such recursive formula. We start with the following elementary lemma.

Lemma 4.2. *Let A and B be two Hermitian matrices and $\gamma>1$. Then, for any $p\in\mathbb{N}$, we have*

$$\mathrm{Tr}\big[(A+B)^p\big]\leq\Big(\frac{\gamma}{\gamma-1}\Big)^{p-1}\mathrm{Tr}(A^p)+\gamma^{p-1}\mathrm{Tr}(B^p).$$

Proof. Let $\widetilde{A}=\frac{\gamma}{\gamma-1}A$ and $\widetilde{B}=\gamma B$. With these notations, we have

$$\mathrm{Tr}\big[(A+B)^p\big]=\mathrm{Tr}\Big[\big(\frac{\gamma-1}{\gamma}\widetilde{A}+\frac{1}{\gamma}\widetilde{B}\big)^p\Big].$$

Using the convexity of $\mathrm{Tr}[(\cdot)^p]$, we get

$$\mathrm{Tr}\big[(A+B)^p\big]\leq\frac{\gamma-1}{\gamma}\mathrm{Tr}(\widetilde{A}^p)+\frac{1}{\gamma}\mathrm{Tr}(\widetilde{B}^p).$$

Replacing $\widetilde{A}$ and $\widetilde{B}$ by their expressions, we finish the proof. □

The next lemma will help us implement an induction argument to prove Theorem 4.1.

Lemma 4.3. *Let μ be a probability measure on some Polish space Ω. Suppose that μ satisfies a matrix Poincaré inequality with constant α and matrix Markov generator $\mathcal{L}$ reversible with respect to μ. Given $g : \Omega \to \mathcal{H}_d$ with $\mathbb{E}_\mu[g] = 0$, let*

$$v_g = \sup \big\|\Gamma(g)\big\|.$$

If $\alpha v_g < 1$, then for any $p \in \mathbb{N}$ we have

$$\mathrm{Tr}\Big[\big(\mathbb{E}_\mu[e^{2g}]\big)^p\Big] \leq \frac{1}{(1-\alpha v_g)^{p-1}}\mathrm{Tr}\Big[\big(\mathbb{E}_\mu[e^{g}]\big)^{2p}\Big] + \alpha v_g \mathrm{Tr}\Big[\mathbb{E}_\mu[e^{2pg}]\Big].$$

Proof. Let $g : \Omega \to \mathcal{H}_d$. Since μ satisfies a matrix Poincaré inequality, then

$$\mathrm{Var}_\mu(e^g) \preceq \alpha \mathcal{E}(e^g),$$

which amounts to

$$\mathbb{E}_\mu[e^{2g}] \preceq \big(\mathbb{E}_\mu[e^g]\big)^2 + \alpha\mathcal{E}(e^g).$$

Using that $\mathrm{Tr}[(\cdot)^p]$ is operator monotone, the above inequality implies that

$$\mathrm{Tr}\Big[\big(\mathbb{E}_\mu[e^{2g}]\big)^p\Big] \leq \mathrm{Tr}\Big[\Big(\big(\mathbb{E}_\mu[e^{g}]\big)^2 + \alpha\mathcal{E}(e^g)\Big)^p\Big].$$

Now using Lemma 4.2, we get

$$\mathrm{Tr}\Big[\big(\mathbb{E}_\mu[e^{2g}]\big)^p\Big] \leq \Big(\frac{\gamma}{\gamma-1}\Big)^{p-1}\mathrm{Tr}\Big[\big(\mathbb{E}_\mu[e^{g}]\big)^{2p}\Big] + \gamma^{p-1}\alpha^p \mathrm{Tr}\Big[\Big(\mathcal{E}(e^g)\Big)^p\Big],$$

where $\gamma > 1$ will be chosen in the sequel. Using Theorem 3.1, we deduce that

$$\mathrm{Tr}\Big[\big(\mathbb{E}_\mu[e^{2g}]\big)^p\Big] \leq \Big(\frac{\gamma}{\gamma-1}\Big)^{p-1}\mathrm{Tr}\Big[\big(\mathbb{E}_\mu[e^{g}]\big)^{2p}\Big] + \gamma^{p-1}\alpha^p v_g^p\, \mathbb{E}_\mu \mathrm{Tr}(e^{2p\, g}).$$

It remains to choose $\gamma = (\alpha v_g)^{-1}$ to finish the proof. □

Proof of Theorem 4.1. Without loss of generality, we assume that $\mathbb{E}_\mu[f] = 0$. We will implement an induction procedure based on the previous lemma. We start by applying Lemma 4.3 with $g_1 = \frac{\delta f}{2}$ and $p = 1$ to get

$$\mathrm{Tr}\Big[\mathbb{E}_\mu[e^{\delta f}]\Big] \leq \frac{1}{1-\alpha v_f(\delta/2)^2}\mathrm{Tr}\Big[\Big(\mathbb{E}_\mu[e^{\frac{\delta f}{2}}]\Big)^2\Big],$$

where we used that $v_{g_1} = (\delta/2)^2 v_f$. Now, apply again Lemma 4.3 with $g_2 = \delta f/2^2$ and $p = 2$ to get

$$\mathrm{Tr}\Big[\mathbb{E}_\mu[e^{\delta f}]\Big] \leq \frac{1}{\big(1-\alpha v_f(\delta/2)^2\big)\big(1-\alpha v_f(\delta/2^2)^2\big)}\mathrm{Tr}\Big[\Big(\mathbb{E}_\mu[e^{\frac{\delta f}{2^2}}]\Big)^{2^2}\Big] + \frac{\alpha v_f(\delta/2^2)^2}{\big(1-\alpha v_f(\delta/2)^2\big)}\mathrm{Tr}\Big[\mathbb{E}_\mu[e^{\delta f}]\Big],$$

which after rearrangement leads to

$$\begin{aligned}
\mathrm{Tr}\Big[\mathbb{E}_\mu[e^{\delta f}]\Big] &\leq \frac{1}{\big(1-\alpha v_f(\delta/2)^2 - \alpha v_f(\delta/2^2)^2\big)\big(1-\alpha v_f(\delta/2^2)^2\big)}\mathrm{Tr}\Big[\Big(\mathbb{E}_\mu[e^{\frac{\delta f}{2^2}}]\Big)^{2^2}\Big] \\
&\leq \frac{1}{\big(1-\alpha v_f(\delta/2)^2 - 2\alpha v_f(\delta/2^2)^2\big)}\mathrm{Tr}\Big[\Big(\mathbb{E}_\mu[e^{\frac{\delta f}{2^2}}]\Big)^{2^2}\Big] \\
&\leq \frac{1}{\big(1-\alpha v_f \delta^2(1-2^{-2})/2\big)}\mathrm{Tr}\Big[\Big(\mathbb{E}_\mu[e^{\frac{\delta f}{2^2}}]\Big)^{2^2}\Big].
\end{aligned}$$

We will prove by induction on s that

$$\mathrm{Tr}\Big[\mathbb{E}_\mu[e^{\delta f}]\Big] \leq \frac{1}{\big(1-\alpha v_f \delta^2(1-2^{-s})/2\big)}\mathrm{Tr}\Big[\Big(\mathbb{E}_\mu[e^{\frac{\delta f}{2^s}}]\Big)^{2^s}\Big].$$

We verified the above inequality for $s = 1$ and $s = 2$. Suppose it is true for s, and apply Lemma 4.3 with $g = \delta f/2^{s+1}$ and $p = 2^s$ to get

$$\mathrm{Tr}\Big[\Big(\mathbb{E}_\mu[e^{\frac{\delta f}{2^s}}]\Big)^{2^s}\Big] \leq \frac{1}{\big(1-\alpha v_f(\delta/2^{s+1})^2\big)^{2^s-1}}\mathrm{Tr}\Big[\big(\mathbb{E}_\mu[e^{\delta f/2^{s+1}}]\big)^{2^{s+1}}\Big] + \alpha v_f(\delta/2^{s+1})^2\mathrm{Tr}\Big[\mathbb{E}_\mu[e^{\delta f}]\Big].$$

Combining the above with the induction hypothesis, we get that

$$\mathrm{Tr}\Big[\mathbb{E}_\mu[e^{\delta f}]\Big] \leq \frac{1}{\big(1-2^{-1}\alpha v_f\delta^2(1-2^{-s}) - \alpha v_f(\delta/2^{s+1})^2\big)\big(1-\alpha v_f(\delta/2^{s+1})^2\big)^{2^s-1}}\mathrm{Tr}\Big[\big(\mathbb{E}_\mu[e^{\delta f/2^{s+1}}]\big)^{2^{s+1}}\Big].$$

Now using that $(1-x)^n \geq 1-nx$ when $x \leq 1$, we deduce from the above that

$$\mathrm{Tr}\Big[\mathbb{E}_\mu[e^{\delta f}]\Big] \leq \frac{1}{1-2^{-1}\alpha v_f\delta^2(1-2^{-s}) - \alpha v_f(\delta/2^{s+1})^2 - (2^s-1)\alpha v_f(\delta/2^{s+1})^2}\mathrm{Tr}\Big[\big(\mathbb{E}_\mu[e^{\delta f/2^{s+1}}]\big)^{2^{s+1}}\Big],$$

which after a short calculation finishes the induction. To finish the proof, take the limit as $s \to \infty$ and notice that

$$\mathrm{Tr}\Big[\Big(\mathbb{E}_\mu[e^{\frac{\delta f}{2^s}}]\Big)^{2^s}\Big] \underset{s\to\infty}{\longrightarrow} \mathrm{Tr}\Big[e^{\delta\mathbb{E}_\mu[f]}\Big] = d,$$

to finish the proof. □

5. Matrix Poincaré inequality for product measures

The aim of this section is to prove Theorems 1.2 and 1.3. Before doing so, we will investigate general product measures. We will first show that an arbitrary product measure $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ satisfies a suitable matrix Poincaré inequality, then will compute the associated carré du champ operator to deduce the following matrix concentration inequality.

Theorem 5.1. *Let $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ be any product measure on some Polish space Ω^n. Let $f : \Omega^n \to \mathcal{H}_d$ be such that*

$$v_f := \sup_{(x_1,\ldots,x_n)\in\Omega^n} \Big\| \sum_{i=1}^n \int \big(f(x_1,\ldots,x_n) - f(x_1,\ldots,x_{i-1},z,x_{i+1},\ldots,x_n)\big)^2 d\mu_i(z)\Big\|$$

is finite. Then, for any $t \geq 0$, we have

$$\mu\Big(\lambda_{max}\big(f-\mathbb{E}_\mu f\big) \geq t\Big) \leq d\exp\Big(-\frac{t^2}{v_f + t\sqrt{v_f}}\Big).$$

The proof of Theorem 5.1 simply consists of constructing a Markov process with μ as stationary measure and having a suitable Markov generator $\mathcal{L}$ for which we prove a matrix Poincaré inequality. In this case, the matrix carré du champ operator consists of the sum of the squared variation in each coordinate of the matrix function. One of the simplest and most natural smoothness assumptions on a matrix function f is the so-called bounded difference condition; i.e. for any $i = 1,\ldots,n$ there exists a deterministic matrix $A_i \in \mathcal{H}_d$ such that

$$\big(f(x_1,\ldots,x_n) - f(x_1,\ldots,x_{i-1},x',x_{i+1},\ldots,x_n)\big)^2 \preceq A_i^2$$

for any $x', x_1, \ldots, x_n \in \Omega$. In this case, we instantly get that $v_f \leq \|\sum_{i=1}^n A_i^2\|$ and hence the inequality

$$\mu\Big(\lambda_{max}\big(f-\mathbb{E}_\mu f\big) \geq t\Big) \leq d\exp\Big(-\frac{t^2}{\sigma^2 + t\sigma}\Big), \tag{2}$$

where $\sigma^2 := \|\sum_{i=1}^n A_i^2\|$. This is a weak form of the matrix bounded difference inequality, as Poincaré inequality cannot capture sub-Gaussian concentration. The matrix bounded difference inequality with sub-Gaussian tail bounds has been established as a consequence of matrix Azuma inequality [29, Section 7], and later recovered with improved constant factors as a consequence of a matrix exponential Efron-Stein inequality [25, Section 5] (see also [24] where the inequality is

derived with an optimal exponent). While the approach based on the matrix Poincaré inequality is unable to compete with such refined results, it provides a unifying framework for several exponential concentration inequalities allowing to cover a wide range of examples and deriving a variety of concentration inequalities. While Theorem 5.1 is stated for any product measure, it could not be used for the example $f(x_1,\dots,x_n)=\sum_{i=1}^n x_i A_i$ with the standard Gaussian measure as the bounded difference condition is violated in this case. As a remedy, Theorem 1.2 provides a refined matrix Poincaré inequality yielding the concentration given in Theorem 1.3, thus recovering the same bound as in (2) for matrix Gaussian series.

Theorem 5.1 follows by combining Theorem 1.1 with Theorem 5.3 below, and using Proposition 5.2 which provides the expression of the matrix carré du champs operator. We start by introducing a Markov process $X_t=(X_t^1,\dots,X_t^n)_{t\in\mathbb{R}_+}$ having μ as stationary measure and through which we obtain a matrix Poincaré inequality with respect to a suitable Dirichlet form. Such a construction is known, see for instance [31, Chapter 2].

For each coordinate $i=1,\dots,n$, we associate an independent Poisson process $N^i=(N_t^i)_{t\in\mathbb{R}_+}$ with rate 1 and construct X_t as follows: we draw X_0 according to μ independently of the Poisson process. Then, whenever N_t^i jumps for some i, we replace the value of X_t^i by an independent sample from μ_i while keeping the remaining coordinates fixed.

Proposition 5.2. *Let $\mu=\mu_1\otimes\ldots\otimes\mu_n$ be any product measure on some Polish space Ω^n. The process $(X_t)_{t\in\mathbb{R}_+}$ defined above is a reversible Markov process with μ as stationary measure and semi-group given by*

$$P_tf(x)=\sum_{I\subseteq\{1,\dots,n\}}(1-e^{-t})^{|I|}e^{-t(n-|I|)}\int f(x_1,\dots,x_n)\prod_{i\in I}d\mu_i(x_i),$$

for any $x=(x_1,\dots,x_n)\in\Omega^n$ and any $f:\Omega^n\to\mathcal{H}_d$ whose matrix coefficients belong to $\mathrm{L}^2(\mu)$. Moreover, the carré du champ and Dirichlet form are respectively given by

$$\Gamma(f)(x)=\frac{1}{2}\sum_{i=1}^n\int\big(f(x_1,\dots,x_n)-f(x_1,\dots,x_{i-1},z,x_{i+1},\dots,x_n)\big)^2d\mu_i(z)$$

and

$$\mathcal{E}(f)=\sum_{i=1}^n\int\left(f(x_1,\dots,x_n)-\int f(x_1,\dots,x_{i-1},z,x_{i+1},\dots,x_n)d\mu_i(z)\right)^2d\mu(x).$$

Proof. It is easy to verify that X_t is a Markov process with μ as stationary measure and that X_t is reversible with respect to μ. Let $f:\Omega^n\to\mathcal{H}_d$ and $x=(x_1,\dots,x_n)\in\Omega^n$. By construction, the Markov semi-group is computed explicitly

$$\begin{aligned}P_tf(x)&=\sum_{I\subseteq\{1,\dots,n\}}\mathbb{P}\big[N_t^i>0,\ \text{for } i\in I, N_t^i=0 \text{ for } i\notin I\big]\int f(x_1,\dots,x_n)\prod_{i\in I}d\mu_i(x_i)\\&=\sum_{I\subseteq\{1,\dots,n\}}(1-e^{-t})^{|I|}e^{-t(n-|I|)}\int f(x_1,\dots,x_n)\prod_{i\in I}d\mu_i(x_i).\end{aligned}$$

Moreover as $\lim_{t\to\infty}(1-e^{-t})^{|I|}e^{-t(n-|I|)}=0$ whenever $|I|<n$, one can readily see that the process is ergodic. In light of Proposition 2.2, the carré du champ operator is given by

$$\begin{aligned}\Gamma(f)(x)&=\lim_{t\to0}\frac{\mathbb{E}\big[\big(f(X_t)-f(X_0)\big)^2\mid X_0=x\big]}{2t}\\&=\lim_{t\to0}\frac{\mathbb{E}\big[\big(f(X_t)-f(x)\big)^2\mid X_0=x\big]}{2t}=\lim_{t\to0}\frac{P_th_x(x)}{2t}\end{aligned}$$

where $h_x : \Omega \to \mathcal{H}_d$ is the function defined by $h_x(y) = (f(x) - f(y))^2$. Now noting that $\lim_{t\to 0} t^{-1}(1-e^{-t})^{|I|}e^{-t(n-|I|)} = 0$ whenever $|I| \geq 2$ and that $h_x(x) = 0$, the explicit expression of the Markov semigroup then yields that

$$\lim_{t\to 0} \frac{P_t h_x(x)}{t} = \sum_{i=1}^{n} \int \big(f(x_1,\dots,x_n) - f(x_1,\dots,x_{i-1},z,x_{i+1},\dots,x_n)\big)^2 d\mu_i(z).$$

Finally, recalling that $\mathcal{E}(f) = \mathbb{E}_\mu[\Gamma(f)]$ and using that

$$\begin{aligned}\int \big(f(x)-f(x_1,\dots,x_{i-1},z,x_{i+1},\dots,x_n)\big)^2 d\mu_i(z)d\mu_i(x_i)&\\ = 2\int \left(f(x) - \int f(x_1,\dots,x_{i-1},z,x_{i+1},\dots,x_n)d\mu_i(z)\right)^2 d\mu_i(x_i),&\end{aligned}$$

we get the expression of $\mathcal{E}(f)$. □

We are now ready to prove that μ satisfies a matrix Poincaré inequality with constant 1 with respect to the above Dirichlet form.

Theorem 5.3. *Let $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ be any product measure on some Polish space Ω^n. Then for any $f : \Omega^n \to \mathcal{H}_d$ whose matrix coefficients belong to $\mathrm{L}^2(\mu)$,*

$$\mathrm{Var}_\mu(f) \preceq \mathcal{E}(f).$$

Proof. Let $f : \Omega^n \to \mathcal{H}_d$. Define $\delta_i f$ by

$$\delta_i f(x) := f(x) - \int f(x_1,\dots,x_{i-1},z,x_{i+1},\dots,x_n)d\mu_i(z)$$

and note that

$$\mathcal{E}(f) = \sum_{i=1}^{n} \int \big(\delta_i f(x)\big)^2 d\mu(x).$$

Since $(X_t)_{t\in\mathbb{R}_+}$ is ergodic and reversible, we apply Lemma 2.3 to write

$$\mathrm{Var}_\mu(f) = 2\int_0^\infty \mathcal{E}(P_t f)\,dt = 2\sum_{i=1}^{n}\int_0^\infty \int (\delta_i P_t f(x))^2 d\mu(x)\,dt.$$

Computing $\delta_i P_t f(x)$ explicitly, we get

$$\delta_i P_t f(x) = e^{-t} \sum_{\substack{I\subseteq\{1,\dots,n\}\\ i\notin I}} (1-e^{-t})^{|I|}e^{-t(n-1-|I|)} \int \delta_i f(x_1,\dots,x_n) \prod_{i\in I} d\mu_i(x_i)\,.$$

Since $\sum_{I\not\ni i}(1-e^{-t})^{|I|}e^{-t(n-1-|I|)} = 1$ and the square is operator convex, then by convexity and Jensen's inequality we obtain

$$\begin{aligned}(\delta_i P_t f(x))^2 &\preceq e^{-2t} \sum_{\substack{I\subseteq\{1,\dots,n\}\\ i\notin I}} (1-e^{-t})^{|I|}e^{-t(n-1-|I|)} \left(\int \delta_i f(x_1,\dots,x_n) \prod_{i\in I} d\mu_i(x_i)\right)^2\\ &\preceq e^{-2t} \sum_{\substack{I\subseteq\{1,\dots,n\}\\ i\notin I}} (1-e^{-t})^{|I|}e^{-t(n-1-|I|)} \int \Big(\delta_i f(x_1,\dots,x_n)\Big)^2 \prod_{i\in I} d\mu_i(x_i)\,.\end{aligned}$$

Taking the expectation we get,

$$\int (\delta_i P_t f(x))^2 d\mu(x) \preceq e^{-2t}\int (\delta_i f(x))^2 d\mu(x)$$

and hence

$$\mathrm{Var}_\mu(f) \preceq \left(2\int_0^\infty e^{-2t}\,dt\right)\sum_{i=1}^n \int (\delta_i f(x))^2 d\mu(x) = \mathcal{E}(f).$$

□

Remark 5.4. In view of the expression of the Dirichlet form, the above matrix Poincaré inequality implies the subadditivity of the variance

$$\mathrm{Var}_\mu(f) \preceq \mathcal{E}(f) = \sum_{i=1}^n \int \mathrm{Var}_{\mu_i}(f) d\mu,$$

and hence the matrix Efron-Stein inequality for product measures. This shows that the latter is a particular case of matrix Poincaré inequalities. We refer to [8, Theorem 5.1] for a direct proof of the Matrix Efron-Stein inequality.

Matrix Poincaré for the standard Gaussian measure. The matrix Poincaré inequality established above applies for any product measure. However, when given a specific product measure, it is possible to architect a suitable Markov generator and prove other matrix Poincaré inequalities which could result in better concentration inequalities. In the remaining part of this section, we investigate the case of the n-dimensional standard Gaussian measure and prove Theorem 1.2. With this in hand, Theorem 1.3 will then follow by using Theorem 1.1 together with the expession of the matrix carré du champ operator given in Proposition 5.5 below.

As we have seen in Remark 5.4, the matrix Poincaré inequality we established can be interpreted as a matrix Efron Stein inequality. In view of this, it is enough to investigate the matrix Poincaré inequality for the one dimensional standard Gaussian measure and then extend it by tensorization to the n-dimensional case. To this aim, let us consider the Ornstein-Uhlenbeck semi-group acting on matrix valued functions in the obvious way, by considering the action entrywise. More precisely, the Ornstein-Uhlenbeck semi-group is defined by

$$P_t f(x) = \mathbb{E}\big[f\big(e^{-t}x + \sqrt{1-e^{-2t}}\xi\big)\big], \qquad \xi \sim N(0,1),$$

for any $f : \mathbb{R} \to \mathcal{H}_d$. The Ornstein-Uhlenbeck process is a reversible ergodic Markov process with stationary measure the standard Gaussian measure. Moreover, the associated Markov generator is given by

$$\mathcal{L}(f)(x) = -xf'(x) + f''(x),$$

where $f'(x)$ (resp. $f''(x)$) denotes the matrix whose entries are the derivatives (resp. second derivatives) of the corresponding entries of $f(x)$.

Proposition 5.5. *The matrix Dirichlet form and matrix carré du champ operator associated with the Ornstein-Uhlenbeck process and standard Gaussian measure γ are given by*

$$\Gamma(f)(x) = (f'(x))^2 \quad \textit{and} \quad \mathcal{E}(f) = \mathbb{E}_\gamma[(f')^2],$$

for any smooth function $f : \mathbb{R} \to \mathcal{H}_d$ whose matrix coefficients and their derivatives belong to $\mathrm{L}^2(\mu)$.

Proof. To compute the matrix carré du champ operator, we start writing

$$\begin{aligned}\Gamma(f)(x) &= \frac{1}{2}\big(\mathcal{L}(f^2)(x) - f(x)\mathcal{L}(f)(x) - \mathcal{L}(f)(x)f(x)\big)\\ &= \frac{1}{2}\Big(-x(f^2)'(x) + (f^2)''(x) - f(x)\big(-xf'(x)+f''(x)\big) - \big(-xf'(x)+f''(x)\big)f(x)\Big)\\ &= (f'(x))^2.\end{aligned}$$

Finally, we finish the proof by recalling that the Dirichlet form is the expectation of the carré du champ operator. □

We are ready now to prove the Gaussian matrix Poincaré inequality.

Proof of Theorem 1.2. We first prove the one dimensional version of the theorem. Recalling the expression of the semi-group, we note that $(P_t f)'(x) = e^{-t} P_t f'(x)$. Using this together with Property (5) of Proposition 2.1, we have

$$\begin{aligned}\mathcal{E}(P_t f) &= \mathbb{E}_\gamma\big[\big((P_t f)'\big)^2\big] = e^{-2t}\mathbb{E}_\gamma\big[(P_t f')^2\big] \\ &\preceq e^{-2t}\mathbb{E}_\gamma\big[P_t(f')^2\big] = e^{-2t}\mathbb{E}_\gamma\big[(f')^2\big] = e^{-2t}\mathcal{E}(f).\end{aligned}$$

Integrating over $\mathbb{R}_+$ and using Lemma 2.3, we deduce the desired matrix Poincaré inequality in the one dimensional case. To derive the inequality for the n-dimenstional standard Gaussian measure $\mu = \mu_1 \otimes \ldots \otimes \mu_n$, we use Remark 5.4 to write

$$\mathrm{Var}_\mu(f) \preceq \int \sum_{i=1}^n \mathrm{Var}_{\mu_i}(f) d\mu,$$

then use the established one dimensional matrix Poincaré to get

$$\mathrm{Var}_\mu(f) \preceq \int \sum_{i=1}^n \Big(\frac{\partial}{\partial x_i} f(x_1, \ldots, x_n)\Big)^2 d\mu,$$

and finish the proof. □

6. Matrix Poincaré inequality for SCP measures

The goal of this section is to prove Theorem 1.4 from which the concentration inequality in Theorem 1.5 follows. In the sequel, μ denotes a probability measure on $\Omega := \{x \in \{0,1\}^n : \sum_{i=1}^n x_i = k\}$ with the SCP property and ξ a random vector on Ω distributed according to μ. We will start by introducing the (normalized) Markov generator for which μ satisfies a Poincaré inequality with constant $2k$. To this aim, given $x, y \in \Omega$, we denote $x \sim y$ if x and y coincide on all but exactly 2 coordinates.

Given $x \sim y$, we denote by s_{xy} (resp. s_{yx}) the unique coordinate $i \in \{1, \ldots, n\}$ such that $x_i = 0$ and $y_i = 1$ (resp. $x_i = 1$ and $y_i = 0$). Note that for any two vectors x, y in Ω, it is possible to construct a sequence of intermediate vectors $(z_i)_{0 \le i \le \ell}$ such that $z_0 = x$, $z_\ell = y$ and $z_i \sim z_{i+1}$ for any $i = 0, \ldots, \ell - 1$. Indeed, the intermediate sequence can be derived by swapping zeros and ones (step by step) on the coordinates where x and y differ. This motivates us to build the generator on vectors differing exactly by one such swap.

Before providing the explicit expression of the generator, let us describe briefly the intuition behind it. Given $x \sim y$, to transition from x to y, a swap has to be made between the coordinates s_{xy} and s_{yx}, and the transition probability is governed by μ. We will uncover the coordinates of x and y in a random order until reaching the coordinate where the two differ (which could be s_{xy} or s_{yx}), in which case we exhibit a "swapping" probability of this coordinate. The uncovered coordinates will be indexed by an ordered subset $S = (s_1, \ldots, s_\ell) \subset [n]$. We will say that (S, x, y) is *admissible* if $x \sim y$ and $x_S = y_S$, that is, the restriction of x and y to the coordinates in the ordered set S coincide. Note that this automatically implies that S does not contain s_{xy} and s_{yx}. Now given an admissible triple (S, x, y) and $s \notin S$, since μ satisfies the SCP property, then there exists a coupling κ_S^s of the measures $\mu(\cdot \mid \xi_S = x_S, \xi_s = 0)$ and $\mu(\cdot \mid \xi_S = x_S, \xi_s = 1)$ which is supported on $\{(x_{\bar{S}}, y_{\bar{S}}) \in \{0,1\}^{\bar{S}} \times \{0,1\}^{\bar{S}} : x_{\bar{S}} \succeq y_{\bar{S}} \text{ and } \|x_{\bar{S}}\|_1 = k - \|x_S\|_1\}$, where we denoted by $\bar{S}$ the unordered set $\bar{S} = (S \cup \{s\})^c$.

We are now ready to introduce the Markov generator Q defined for every $x \sim y$ by

$$Q(x,y) := \frac{1}{2k} \sum_{\ell=0}^{n-2} \frac{(n-1-\ell)!}{n!} \sum_{\substack{S:\,|S|=\ell \\ (S,x,y) \text{ admissible}}} \frac{H_S^{s_{xy}}(x,y) + H_S^{s_{yx}}(y,x)}{\mu(x \mid \xi_S = x_S)}, \tag{3}$$

where

$$H_S^s(x,y) := \kappa_S^s(x,y)\mathbb{P}(\xi_s = 0 \mid \xi_S = x_S)\mathbb{P}(\xi_s = 1 \mid \xi_S = x_S).$$

We set $Q(x,x) = -\sum_{y\sim x} Q(x,y)$ to complete the construction. The above expression puts in place the informal description provided previously. Indeed, it is obtained by averaging over all possible ways of uncovering the coordinates of x and y. This can be seen by noting that the factor $\frac{(n-1-\ell)!}{n!}$ represents the probability of uncovering the coordinates in some fixed order $(s_1, \dots, s_\ell, s_{xy})$. Finally, after uncovering the coordinates, we exhibit the transition probability on the differing coordinate using the corresponding coupling between the measures obtained by conditioning on the uncovered coordinates. We should note that the above Markov generator is the one implicitly used in [16]. Indeed the above expression can be recovered by carefully following the iterative procedure implemented there.

Clearly, Q is reversible by construction. Moreover, Q is normalized as we check in the next proposition.

Proposition 6.1. *With above notations, we have*

$$\max_{x\in\Omega}\{-Q(x,x)\} \le 1.$$

Proof. Let $x \in \Omega$ and denote by $\operatorname{Supp} x := \{i \in \{1,\dots,n\} : x_i = 1\}$ its support. We start writing

$$-Q(x,x) = \sum_{y\sim x} Q(x,y) = \frac{1}{2k}\sum_{\ell=0}^{n-2} \frac{(n-1-\ell)!}{n!}(\alpha_\ell + \beta_\ell),$$

where

$$\alpha_\ell := \sum_{y\sim x} \sum_{\substack{S:\,|S|=\ell \\ (S,x,y) \text{ admissible}}} \frac{H_S^{s_{xy}}(x,y)}{\mu(x \mid \xi_S = x_S)},$$

and

$$\beta_\ell := \sum_{y\sim x} \sum_{\substack{S:\,|S|=\ell \\ (S,x,y) \text{ admissible}}} \frac{H_S^{s_{yx}}(y,x)}{\mu(x \mid \xi_S = x_S)}.$$

We will estimate α_ℓ and β_ℓ separately.

Note that, for x given, the collection of all admissible triples (S,x,y) is in a one to one correspondence with admissible triples $(s,S,y_{\bar S})$ where $s \notin \operatorname{Supp} x$, $S \subset [n]\setminus\{s\}$ ordered set, and $y_{\bar S}$ a 0/1 vector on $\bar S = (S\cup\{s\})^c$ satisfying $\|y_{\bar S}\|_1 = \|x_{\bar S}\|_1 - 1$. To see this, given $y_{\bar S}$, note that one can uniquely define $y \sim x$ by concatenating x_S, $y_{\bar S}$ and setting $y_s = 1$. In view of this, we can write

$$\alpha_\ell = \sum_{s\notin \operatorname{Supp} x} \sum_{\substack{S\subset[n]\setminus\{s\} \\ S \text{ ordered},\,|S|=\ell}} \sum_{y_{\bar S}} \frac{\kappa_S^s(x,y)\mathbb{P}(\xi_s = 0 \mid \xi_S = x_S)\mathbb{P}(\xi_s = 1 \mid \xi_S = x_S)}{\mu(x \mid \xi_S = x_S)}.$$

Recalling that

$$\sum_{y_{\bar S}} \kappa_S^s(x,y) = \mu(x \mid \xi_S = x_S, \xi_s = 0),$$

we get

$$\begin{aligned}\alpha_\ell &= \sum_{s\notin \operatorname{Supp} x} \sum_{\substack{S\subset[n]\setminus\{s\}\\ S \text{ ordered},\, |S|=\ell}} \mathbb{P}(\xi_s = 1 \mid \xi_S = x_S)\\ &\leq \sum_{\substack{S\subset[n]\\ S \text{ ordered},\, |S|=\ell}} \sum_{s\in S^c} \mathbb{E}[\xi_s \mid \xi_S = x_S]\\ &= \sum_{\substack{S\subset[n]\\ S \text{ ordered},\, |S|=\ell}} \sum_{i\notin S} x_i\\ &= \sum_{i=1}^n \sum_{\substack{S\subset[n]\setminus\{i\}\\ S \text{ ordered},\, |S|=\ell}} x_i = k\frac{(n-1)!}{(n-1-\ell)!}.\end{aligned}$$

where we used that $\mathbb{E}[\sum_{s\in S^c}\xi_s \mid \xi_S = x_S] = \sum_{i\notin S} x_i$ (due to the homogeneity of μ).

To estimate β_ℓ, note that, given x, the collection of all admissible triples (S, x, y) is in a one to one correspondence with all triples $(s, S, y_{\bar S})$ where $s \in \operatorname{Supp} x$, $S \subset [n]\setminus\{s\}$ ordered set, and $y_{\bar S}$ a 0/1 vector on $\bar S = (S\cup\{s\})^c$ satisfying $\|y_{\bar S}\|_1 = \|x_{\bar S}\|_1$. To see this, given $y_{\bar S}$, note that one can uniquely define $y \sim x$ by concatenating x_S, $y_{\bar S}$ and setting $y_s = 0$. Therefore, we have

$$\beta_\ell = \sum_{s\in\operatorname{Supp} x} \sum_{\substack{S\subset[n]\setminus\{s\}\\ S \text{ ordered},\, |S|=\ell}} \sum_{y_{\bar S}} \frac{\kappa_S^s(y,x)\mathbb{P}(\xi_s = 0 \mid \xi_S = x_S)\mathbb{P}(\xi_s = 1 \mid \xi_S = x_S)}{\mu(x \mid \xi_S = x_S)}.$$

Recalling that

$$\sum_{y_{\bar S}} \kappa_S^s(y,x) = \mu(x \mid \xi_S = x_S, \xi_s = 1),$$

we get

$$\beta_\ell = \sum_{s\in\operatorname{Supp} x} \sum_{\substack{S\subset[n]\setminus\{s\}\\ S \text{ ordered},\, |S|=\ell}} \mathbb{P}(\xi_s = 0 \mid \xi_S = x_S) \leq k\frac{(n-1)!}{(n-1-\ell)!},$$

where we used the trivial bound $\mathbb{P}(\xi_s = 0 \mid \xi_S = x_S) \leq 1$.

Putting together the estimates on α_ℓ and β_ℓ, we deduce that

$$-Q(x,x) \leq \frac{n-1}{n} \leq 1,$$

and finish the proof. □

Before we proceed with the proof of the matrix Poincaré inequality promised in the introduction, we need the following identity which we interpret as a two state matrix Poincaré.

Lemma 6.2. *Let π be a probability measure on $\{0,1\}$ and $\widetilde{Q}$ be a reversible Markov generator. Then for any $f : \{0,1\} \to \mathcal{H}_d$, we have*

$$\operatorname{Var}_\pi(f) = \frac{1}{\widetilde{Q}(0,1) + \widetilde{Q}(1,0)}\mathcal{E}(f,f).$$

Proof. First note that

$$\operatorname{Var}_\pi(f) = \pi(0)\big(f(0) - \mathbb{E}_\pi[f]\big)^2 + \pi(1)\big(f(1) - \mathbb{E}_p i[f]\big)^2 = \pi(0)\pi(1)\big(f(0) - f(1)\big)^2.$$

On the other hand, using the reversibility of $\widetilde{Q}$, we can write

$$\mathcal{E}(f,f) = \pi(0)\widetilde{Q}(0,1)\big(f(0) - f(1)\big)^2.$$

Also, by reversibility of $\widetilde{Q}$, it is easy to see that $\pi(1) = \frac{\widetilde{Q}(0,1)}{\widetilde{Q}(0,1)+\widetilde{Q}(1,0)}$. This finishes the proof. $\square$

Proof of Theorem 1.4. We will show that μ satisfies a matrix Poincaré inequality with constant $2k$ with respect to the Markov generator defined in (3). We showed in Proposition 6.1 that Q is normalized, and it remains to show that for any $f : \Omega \to \mathcal{H}_d$, we have

$$\mathrm{Var}_\mu(f) \preceq 2k\mathcal{E}(f,f).$$

First, by Proposition 2.2,

$$2k\mathcal{E}(f,f) = \sum_{\ell=0}^{n-2} \frac{(n-1-\ell)!}{n!} \sum_{x\sim y} \sum_{\substack{S:\,|S|=\ell \\ (S,x,y)\text{ admissible}}} \mathbb{P}(\xi_S = x_S)\, H_S^{s_{xy}}(x,y)\big(f(x)-f(y)\big)^2,$$

where we have used the reversibility of Q to simplify the expression. Note that the collection of all admissible triples (S,x,y) is in one to one correspondence with all quintuples $(s,S,x_S,x_{\bar S},y_{\bar S})$, where $\bar S := (S\cup\{s\})^c$, $s\notin S$, $x_{\bar S} \succeq y_{\bar S}$ and $\|x_{\bar S}\|_1 = k - \|x_S\|_1$. To see this, note that if (S,x,y) is admissible then $x_S = y_S$. Moreover, given $(s,S,x_S,x_{\bar S},y_{\bar S})$, it is possible to uniquely reconstruct x (resp. y) by concatenating x_S and $x_{\bar S}$ (resp. x_S and $y_{\bar S}$) and setting $x_s = 0$ (resp. $y_s = 1$). In the sequel, given an admissible quintuple $(s,S,x_S,x_{\bar S},y_{\bar S})$, x and y refer to the vectors constructed as we just described. In view of this, for any $\ell = 0,\dots,n-2$, we can write

$$\begin{aligned}\gamma_\ell &:= \sum_{x\sim y} \sum_{\substack{S:\,|S|=\ell \\ (S,x,y)\text{ admissible}}} \mathbb{P}(\xi_S = x_S)\, H_S^{s_{xy}}(x,y)\big(f(x)-f(y)\big)^2 \\ &= \sum_{s=1}^n \sum_{\substack{S\subset[n]\setminus\{s\} \\ S\text{ ordered},\,|S|=\ell}} \sum_{x_S} \mathbb{P}(\xi_s = 0 \mid \xi_S = x_S)\,\mathbb{P}(\xi_s = 1,\,\xi_S = x_S) \sum_{(x_{\bar S},y_{\bar S})} \kappa_S^s(x,y)\big(f(x)-f(y)\big)^2.\end{aligned}$$

Since the square is operator convex and κ_S^s is a probability measure on $\bar S\times\bar S$, then using Jensen's inequality we get

$$\gamma_\ell \succeq \sum_{s=1}^n \sum_{\substack{S\subset[n]\setminus\{s\} \\ S\text{ ordered},\,|S|=\ell}} \sum_{x_S} \mathbb{P}(\xi_s = 0 \mid \xi_S = x_S)\,\mathbb{P}(\xi_s = 1,\,\xi_S = x_S)\big(f_{(s,S,x_S)}(0) - f_{(s,S,x_S)}(1)\big)^2,$$

where $f_{(s,S,x_S)} : \{0,1\} \to \mathcal{H}_d$ is defined by

$$f_{(s,S,x_S)}(0) = \sum_{(x_{\bar S},y_{\bar S})} \kappa_S^s(x,y)\, f(x) = \sum_{x_{\bar S}} \mu(x \mid \xi_s = 0, \xi_S = x_S) f(x) = \mathbb{E}[f(\xi) \mid \xi_s = 0, \xi_S = x_S],$$

and

$$f_{(s,S,x_S)}(1) = \sum_{(x_{\bar S},y_{\bar S})} \kappa_S^s(x,y)\, f(y) = \sum_{y_{\bar S}} \mu(y \mid \xi_s = 1, \xi_S = x_S) f(x) = \mathbb{E}[f(\xi) \mid \xi_s = 1, \xi_S = x_S].$$

Now for a given triple (s,S,x_S), define a probability measure π on $\{0,1\}$ by $\pi(0) = \mathbb{P}(\xi_s = 0 \mid \xi_S = x_S)$ and $\pi(1) = \mathbb{P}(\xi_s = 1 \mid \xi_S = x_S)$. Moreover, define a reversible Markov generator $\widetilde{Q}$ by $\widetilde{Q}(0,1) = \mathbb{P}(\xi_s = 1,\,\xi_S = x_S)$. On the one hand,

$$\begin{aligned}\mathcal{E}(f_{(s,S,x_S)}, f_{(s,S,x_S)}) &= \pi(0)\pi(1)\widetilde{Q}(0,1)\big(f_{(s,S,x_S)}(0) - f_{(s,S,x_S)}(1)\big)^2 \\ &= \mathbb{P}(\xi_s = 0 \mid \xi_S = x_S)\mathbb{P}(\xi_s = 1,\,\xi_S = x_S)\big(f_{(s,S,x_S)}(0) - f_{(s,S,x_S)}(1)\big)^2.\end{aligned}$$

On the other hand, by the two-state matrix Poincaré inequality (Lemma 6.2),

$$\mathcal{E}(f_{(s,S,x_S)}, f_{(s,S,x_S)}) = \big(\widetilde{Q}(0,1) + \widetilde{Q}(1,0)\big)\mathrm{Var}_\pi(f_{(s,S,x_S)}) = \mathbb{P}(\xi_S = x_S)\mathrm{Var}_\pi(f_{(s,S,x_S)}).$$

Thus we get

$$\mathbb{P}(\xi_s = 0 \mid \xi_S = x_S)\mathbb{P}(\xi_s = 1,\, \xi_S = x_S)\big(f_{(s,S,x_S)}(0) - f_{(s,S,x_S)}(1)\big)^2 = \mathbb{P}(\xi_S = x_S)\mathrm{Var}_\pi(f_{(s,S,x_S)}),$$

which when replaced in the expression of γ_ℓ yields to

$$\gamma_\ell \succeq \sum_{s=1}^{n} \sum_{\substack{S\subset[n]\setminus\{s\} \\ S \text{ ordered},\, |S|=\ell}} \sum_{x_S} \mathbb{P}(\xi_S = x_S)\mathrm{Var}_\pi(f_{(s,S,x_S)}).$$

Now note that

$$\mathbb{E}_\pi[f_{(s,S,x_S)}] = \mathbb{E}[f(\xi) \mid \xi_S = x_S],$$

and

$$\begin{aligned}\sum_{x_S} \mathbb{P}(\xi_S = x_S)\mathbb{E}_\pi[f^2_{(s,S,x_S)}] &= \sum_{x_S} \mathbb{P}(\xi_s = 0, \xi_S = x_S)\big(\mathbb{E}[f(\xi) \mid \xi_s = 0, \xi_S = x_S]\big)^2 \\ &\quad + \sum_{x_S} \mathbb{P}(\xi_s = 1, \xi_S = x_S)\big(\mathbb{E}[f(\xi) \mid \xi_s = 1, \xi_S = x_S]\big)^2 \\ &= \sum_{x_{S\cup\{s\}}} \mathbb{P}(\xi_{S\cup\{s\}} = x_{S\cup\{s\}})\big(\mathbb{E}[f(\xi) \mid \xi_{S\cup\{s\}} = x_{S\cup\{s\}}]\big)^2.\end{aligned}$$

Putting together these identities, we get

$$\begin{aligned}\gamma_\ell &\succeq \sum_{s=1}^{n} \sum_{\substack{S\subset[n]\setminus\{s\} \\ S \text{ ordered},\, |S|=\ell}} \sum_{x_{S\cup\{s\}}} \mathbb{P}(\xi_{S\cup\{s\}} = x_{S\cup\{s\}})\big(\mathbb{E}[f(\xi) \mid \xi_{S\cup\{s\}} = x_{S\cup\{s\}}]\big)^2 \\ &\quad - \sum_{s=1}^{n} \sum_{\substack{S\subset[n]\setminus\{s\} \\ S \text{ ordered},\, |S|=\ell}} \sum_{x_S} \mathbb{P}(\xi_S = x_S)\big(\mathbb{E}[f(\xi) \mid \xi_S = x_S]\big)^2 \\ &= \sum_{\substack{S\subset[n] \\ S \text{ ordered},\, |S|=\ell+1}} \sum_{x_S} \mathbb{P}(\xi_S = x_S)\big(\mathbb{E}[f(\xi) \mid \xi_S = x_S]\big)^2 \\ &\quad - (n-\ell) \sum_{\substack{S\subset[n] \\ S \text{ ordered},\, |S|=\ell}} \sum_{x_S} \mathbb{P}(\xi_S = x_S)\big(\mathbb{E}[f(\xi) \mid \xi_S = x_S]\big)^2.\end{aligned}$$

In view of this, the sum involving γ_ℓ is a telescopic sum, yielding to

$$\begin{aligned}2k\mathcal{E}(f,f) &= \sum_{\ell=0}^{n-2} \frac{(n-1-\ell)!}{n!}\gamma_\ell \\ &\succeq \frac{1}{n!} \sum_{\substack{S\subset[n] \\ S \text{ ordered},\, |S|=n-1}} \sum_{x_S} \mathbb{P}(\xi_S = x_S)\big(\mathbb{E}[f(\xi) \mid \xi_S = x_S]\big)^2 - \big(\mathbb{E}[f(\xi)]\big)^2.\end{aligned}$$

Finally, note that by homogeneity, fixing $n-1$ coordinates automatically determines the remaining coordinate. Therefore, for any ordered set S of size $n-1$, we have

$$\sum_{x_S} \mathbb{P}(\xi_S = x_S)\big(\mathbb{E}[f(\xi) \mid \xi_S = x_S]\big)^2 = \mathbb{E}[f^2(\xi)],$$

which when replaced in the previous inequality finishes the proof. □

Finally, we end the section by showing how to derive the concentration inequality stated in Theorem 1.5.

Proof of Theorem 1.5. The proof will follow by combining Theorem 1.1 and Theorem 1.4. First, using Proposition 2.2, we have for any $f : \Omega \to \mathcal{H}_d$ and any $x \in \Omega$

$$\Gamma(f)(x) = \frac{1}{2}\sum_{y\in\Omega} Q(x,y)\big(f(x) - f(y)\big)^2,$$

where Q is the Markov generator defined in (3). Note that if f is 1-Lipschitz in the sense of Theorem 1.5, then $\|f(x) - f(y)\| \leq 2$ for any $x \sim y$. Using this together with the triangular inequality, we deduce that for any 1-Lipschitz matrix function f, we have

$$\|\Gamma(f)(x)\| \leq 2,$$

for any $x \in \Omega$. Replacing this estimate in Theorem 1.1 together with the value of the matrix Poincaré constant from Theorem 1.4, we finish the proof. □

References

[1] R. Ahlswede and A. Winter. Addendum to: Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory*, 49(1):346, 2003.

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

[3] M. Banna, F. Merlevède, and P. Youssef. Bernstein-type inequality for a class of dependent random matrices. *Random Matrices Theory Appl.*, 5(2):1650006, 28, 2016.

[4] J. Borcea, P. Brändén, and T. M. Liggett. Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.*, 22(2):521–567, 2009.

[5] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

[6] 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.

[7] R. Y. Chen and J. A. Tropp. Subadditivity of matrix ϕ-entropy and concentration of random matrices. *Electron. J. Probab.*, 19:no. 27, 30, 2014.

[8] H.-C. Cheng and M.-H. Hsieh. Characterizations of matrix and operator-valued Φ-entropies, and operator Efron-Stein inequalities. *Proc. R. Soc. A*, 472(2187):20150563, 20, 2016.

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

[10] 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, 24, 2017.

[11] M.D. Choi. A Schwarz inequality for positive linear maps on C^*-algebras. *Illinois J. Math.*, 18 (1974), pp. 565–574.

[12] P. J. Forrester and C. J. Thompson. The Golden-Thompson inequality: historical aspects and random matrix applications. *J. Math. Phys.*, 55(2):023503, 12, 2014.

[13] S. Golden. Lower bounds for the Helmholtz function. *Phys. Rev. (2)*, 137:B1127–B1128, 1965.

[14] M. Gromov and V. D. Milman. A topological application of the isoperimetric inequality. *Amer. J. Math.*, 105(4):843–854, 1983.

[15] F. Hansen and G. K. Pedersen. Jensen's inequality for operators and Löwners theorem. *Mathematische Annalen*, 258(1982), 229–241.

[16] J. Hermon and J. Salez. Modified log-sobolev inequalities for strong-rayleigh measures. *arXiv preprint arXiv:1902.02775*, 2019.

[17] F. Hiai and H. Kosaki. Means for matrices and comparison of their norms. *Indiana Univ. Math. J.*, 48 (1999), 899–936.

[18] R. Kyng and Z. Song. A matrix chernoff bound for strongly rayleigh distributions and spectral sparsifiers from a fewrandom spanning trees. In 2018 IEEE 59th Annual Symposium on Foundations of Computer Science(FOCS). IEEE, 2018, pp. 373–384.

[19] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 120–216. Springer, Berlin, 1999.

[20] M. Ledoux. The concentration of measure phenomenon. Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence, RI(2001).

[21] S. L. Lu and H.-T. Yau. Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics. *Comm. Math. Phys.*, 156(2):399–433, 1993.

[22] 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, 2014.
[23] R. I. Oliveira. Sums of random Hermitian matrices and an inequality by Rudelson. *Electron. Commun. Probab.*, 15:203–212, 2010.
[24] D. Paulin, L. Mackey, and J. A. Tropp. Deriving matrix concentration inequalities from kernel couplings. Available at arXiv:1305.0612, 2014.
[25] D. Paulin, L. Mackey, and J. A. Tropp. Efron-Stein inequalities for random matrices. *Ann. Probab.*, 44(5):3431–3473, 2016.
[26] R. Pemantle and Y. Peres. Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures. *Combin. Probab. Comput.*, 23(1):140–160, 2014.
[27] C. J. Thompson. Inequality with applications in statistical mechanics. *J. Mathematical Phys.*, 6:1812–1813, 1965.
[28] J. A. Tropp. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16: 262–270, 2011.
[29] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012.
[30] J. A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.
[31] R. van Handel. Probability in high dimension. ORF 570 Lecture Notes, Princeton University, June 2014.

Richard Aoun,
American University of Beirut, Department of Mathematics, Faculty of Arts and Sciences, P.O. Box 11-0236 Riad El Solh, Beirut 1107 2020, Lebanon,
E-mail: ra279@aub.edu.lb

Marwa Banna,
Saarland University, Fachbereich Mathematik, 66041 Saarbrücken, Germany
E-mail: banna@math.uni-sb.de

Pierre Youssef,
Laboratoire de Probabilités, Statistique et Modélisation, Université Paris Diderot, France
And
Mathematics, Division of Science, New York University Abu Dhabi, UAE
E-mail: youssef@lpsm.paris