Bernstein type inequality for a class of dependent random matrices

Marwa Banna[a], Florence Merlevède[b], Pierre Youssef[c]

[a b] Université Paris-Est, UPEM, LAMA, CNRS, UPEC, 5 Boulevard Descartes, 77 454 Marne La Vallée, France.

E-mail: marwa.banna@u-pem.fr; florence.merlevede@u-pem.fr

[c] Department of Mathematical and Statistical sciences. University of Alberta, Canada.

Email: pyoussef@ualberta.ca

Key words: Random matrices, Bernstein inequality, Deviation inequality, Absolute regularity, β-mixing coefficients.

Mathematics Subject Classification (2010): 60B20, 60F10.

Abstract.

In this paper we obtain a Bernstein type inequality for the sum of self-adjoint centered and geometrically absolutely regular random matrices with bounded largest eigenvalue. This inequality can be viewed as an extension to the matrix setting of the Bernstein-type inequality obtained by Merlevède et al. [11] in the context of real-valued bounded random variables that are geometrically absolutely regular. The proofs rely on decoupling the Laplace transform of a sum on a Cantor-like set of random matrices.

1 Introduction

The analysis of the spectrum of large matrices has known significant development recently due to its important role in several domains. One of the questions is to study the fluctuations of a Hermitian matrix $\mathbb{X}$ from its expectation measured by its largest eigenvalue. Matrix concentration inequalities give probabilistic bounds for such fluctuations and provide effective methods for studying several models. The Laplace transform method, which is due to Bernstein in the scalar case, was generalized to the sum of independent Hermitian random matrices by Ahlswede and Winter in [2]. The starting point is that the usual Chernoff bound when we deal with the partial sums associated to real-valued random variables has the following counterpart in the matrix setting:

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{i=1}^n \mathbb{X}_i\big) \geq x\Big) \leq \inf_{t>0}\Big\{\mathrm{e}^{-tx}\cdot \mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\sum_{i=1}^n \mathbb{X}_i}\big)\Big\} \tag{1}$$

(see [2]). Here and all along the paper, $(\mathbb{X}_i)_{i\geq 1}$ is a family of $d\times d$ self-adjoint random matrices. The main problem is then to give a suitable bound for $L_n(t) := \mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\sum_{i=1}^n \mathbb{X}_i}\big)$. In the independent case, starting from the Golden-Thompson inequality stating that if $\mathbb{A}$ and $\mathbb{B}$ are two self-adjoint matrices,

$$\mathrm{Tr}\big(\mathrm{e}^{\mathbb{A}+\mathbb{B}}\big) \leq \mathrm{Tr}\big(\mathrm{e}^{\mathbb{A}}\mathrm{e}^{\mathbb{B}}\big)\,,$$

Ahlswede and Winter have observed that

$$\mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\sum_{i=1}^n \mathbb{X}_i}\big) \leq \lambda_{\max}(\mathbb{E}(\mathrm{e}^{t\mathbb{X}_n}))\cdot \mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\sum_{i=1}^{n-1} \mathbb{X}_i}\big) \tag{2}$$

and gave a bound for $L_n(t)$ by iterating the procedure above. In [17], Tropp used Lieb's concavity theorem (see [9]) to improve the bound on $L_n(t)$ stated in [2] and obtained Lemma 4 of Section

4.1. This lemma then allows to extend to the matrix setting the usual Bernstein inequality for the partial sum associated with independent real-valued random variables.

Let us mention that an extension of the so-called Hoeffding-Azuma inequality for matrix martingales and of the so-called McDiarmid bounded difference inequality for matrix-valued functions of independent random variables is also stated in [17].

Taking another direction, Mackey et al. [10] extended to the matrix setting Chatterjee's technique for developing scalar concentration inequalities via Stein's method of exchangeable pairs (see [4] and [5]), and established Bernstein and Hoeffding inequalities as well as concentration inequalities. Following this approach, Paulin et al. [14] established a so-called McDiarmid inequality for matrix-valued functions of dependent random variables under conditions on the associated Dobrushin interdependence matrix.

The aim of this paper is to give an extension of the Bernstein deviation inequality when we consider the largest eigenvalue of the partial sums associated with self-adjoint, centered and geometrically absolutely regular random matrices with bounded largest eigenvalue. A family $(\mathbb{X}_i)_{i\geq 1}$ of $d\times d$ matrices will be said to be absolutely regular if $\lim_{n\to\infty}\beta_n = 0$ where the coefficients of absolute regularity $(\beta_n)_{n\geq 0}$ are defined at the beginning of Section 2, and geometrically absolutely regular if the rate of convergence is exponentially fast. These coefficients aim at quantifying the strength of dependence between the sigma-algebra generated by $(\mathbb{X}_i)_{1\leq i\leq k}$ and the one generated by $(\mathbb{X}_i)_{i\geq k+n}$ for all $k\in\mathbb{N}^*$. So asking β_n to converge to zero means that the sigma-algebra generated by $(\mathbb{X}_i)_{i\geq k+n}$ is less and less dependent from $\sigma((\mathbb{X}_i)_{1\leq i\leq k})$ the larger n is. This kind of dependence cannot be compared to the dependence structure imposed in [10] or in [14].

Note that for dependent random matrices, the first step given by (2) of the iterative procedure in [2] fails as well as the concave trace function method used in [17]. Therefore additional transformations on the Laplace transform have to be made. Even in the scalar dependent case, obtaining sharp Bernstein-type inequalities is a challenging problem and a dependence structure of the underlying process has obviously to be specified. For instance, Adamczak [1] proved a Bernstein-type inequality for the partial sum associated with bounded functions of a geometrically ergodic Harris recurrent Markov chain. He showed that even in this context where it is possible to go back to the independent setting by creating random iid cycles, a logarithmic extra factor (compared to the independent case) cannot be avoided (see Theorem 6 and Section 3.3 in [1]).

In [11] and [12], Merlevède et al. considered more general dependence structures than Harris recurrent Markov chains and proved Bernstein-type inequalities for the partial sums associated with bounded real-valued random variables whose strong mixing coefficients (or τ-dependent coefficients) decrease geometrically or sub-geometrically. Note that in [12], the case of real-valued random variables that are not necessarily bounded is also treated. The method used in both papers mentioned consists of partitioning the n random variables in blocks indexed by Cantor-type sets plus a remainder. The idea is then to control the log-Laplace transform of each partial sum on the Cantor-type sets. The log-Laplace transform of the total partial sum is then handled with the help of a general result which provides bounds for the log-Laplace transform of any sum of real-valued random variables (see our Lemma 5 in the context of random matrices). Obviously, the main step is to obtain a suitable upper bound of the log-Laplace transform of the partial sum on each of the Cantor-type set. The dependence structure assumed in [11] or [12] allow the following control: for any index sets Q and Q' of natural numbers such that $Q\subset[1,p]$ and $Q'\subset[n+p,\infty)$ and any $t>0$,

$$\mathbb{E}\big(\mathrm{e}^{t\sum_{i\in Q}X_i}\mathrm{e}^{t\sum_{i\in Q'}X_i}\big)\leq\mathbb{E}\big(\mathrm{e}^{t\sum_{i\in Q}X_i}\big)\mathbb{E}\big(\mathrm{e}^{t\sum_{i\in Q'}X_i}\big)+\varepsilon(n)\|\mathrm{e}^{t\sum_{i\in Q}X_i}\|_\infty\|\mathrm{e}^{t\sum_{i\in Q'}X_i}\|_\infty\,,\quad(3)$$

where $\varepsilon(n)$ is a sequence of positive real numbers depending on the dependent coefficients. The binary tree structure of the Cantor-type sets allows to iterate the decorrelation procedure above-mentioned allowing then to suitably handle the log-Laplace transform of the partial sum on each of the Cantor-type set.

In the random matrix setting, iterating a procedure as (3) cannot lead to suitable exponential inequalities essentially due to the fact that the extension of the Golden-Thompson inequality to three or more Hermitian matrices fails, and then the extension of the exponential inequalities stated in [11] and [12] to the matrix setting is not straightforward. To benefit from the ideas developed in [2] or in [17], we shall rather bound the log-Laplace transform of the partial sum indexed by a Cantor-type set, say K, by the log-Laplace transform of the sum of 2^ℓ independent and self-adjoint random matrices plus a small error term (here ℓ depends on the cardinality of K). Lemma 8 is in this direction and can be viewed as a decoupling lemma for the Laplace transform in the matrix setting. As we shall see, a well-adapted dependence structure allowing such a procedure is the absolute regularity structure. Indeed, the Berbee's coupling lemma (see Lemma 6 below) allows a "good coupling" in terms of absolute regular coefficients (see the definition (4)) even when the underlying random variables take values in a high dimensional space (working with $d \times d$ random matrices can be viewed as working with random vectors of dimension d^2). The decoupling lemma 8 associated with additional coupling arguments will then allow us to prove our key Proposition 7 giving a bound for the Laplace transform of the partial sum indexed by Cantor-type set of self-adjoint random matrices. As we shall see, our method allows to extend the scalar Bernstein type inequality given in [11] to the matrix setting.

Our paper is organized as follows. In Section 2, we introduce some notations and definitions and state our Bernstein-type inequality for the class of random matrices we consider (see Theorem 1). Section 3 is devoted to some examples of matrix models where this Bernstein-type inequality applies. The proof of the main result is given in Section 4.

2 Main Result

For any $d \times d$ matrix $\mathbb{X} = [(\mathbb{X})_{i,j}]_{i,j=1}^d$ whose entries belong to $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, we associate its corresponding vector $\mathbf{X}$ in $\mathbb{K}^{d^2}$ whose coordinates are the entries of $\mathbb{X}$ i.e.

$$\mathbf{X} = \big((\mathbb{X})_{i,j}\,, 1 \leq i \leq d\big)_{1 \leq j \leq d}\,.$$

Therefore $\mathbf{X} = \big(X_i\,, 1 \leq i \leq d^2\big)$ where

$$X_i = (\mathbb{X})_{i-(j-1)d,j} \quad \text{for } (j-1)d + 1 \leq i \leq jd\,,$$

and $\mathbf{X}$ will be called the vector associated with $\mathbb{X}$. Reciprocally, given $\mathbf{X} = \big(X_\ell\,, 1 \leq \ell \leq d^2\big)$ in $\mathbb{K}^{d^2}$ we shall associate a $d \times d$ matrix $\mathbb{X}$ by setting

$$\mathbb{X} = \big[(\mathbb{X})_{i,j}\big]_{i,j=1}^n \text{ where } (\mathbb{X})_{i,j} = X_{i+(j-1)d}\,.$$

The matrix $\mathbb{X}$ will be referred to as the matrix associated with $\mathbf{X}$.

In all the paper we consider a family $(\mathbb{X}_i)_{i\geqslant 1}$ of $d \times d$ self-adjoint random matrices whose entries are defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and with values in $\mathbb{K}$, and that are geometrically absolutely regular in the following sense. Let

$$\beta_0 = 1 \quad \text{and} \quad \beta_k = \sup_{j\geqslant 1} \beta(\sigma(\mathbf{X}_i\,,\, i \leq j), \sigma(\mathbf{X}_i\,,\, i \geqslant j+k))\,, \text{ for any } k \geq 1, \tag{4}$$

where

$$\beta(\mathcal{A}, \mathcal{B}) = \frac{1}{2} \sup \Big\{ \sum_{i\in I} \sum_{j\in J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| \Big\}\,,$$

the maximum being taken over all finite partitions $(A_i)_{i\in I}$ and $(B_i)_{i\in J}$ of Ω respectively with elements in $\mathcal{A}$ and $\mathcal{B}$.

The $(\beta_k)_{k\geqslant 0}$ are usually called the coefficients of absolute regularity of the sequence of vectors $(\mathbf{X}_i)_{i\geqslant 1}$ and we shall assume in this paper that they decrease geometrically in the sense that there exists $c>0$ such that for any integer $k\geqslant 1$,

$$\beta_k=\sup_{j\geqslant 1}\beta(\sigma(\mathbf{X}_i\,,\,i\le j),\sigma(\mathbf{X}_i\,,\,i\geqslant j+k))\le \mathrm{e}^{-c(k-1)}\,. \tag{5}$$

Note that the β_k coefficients have been introduced by Kolmogorov and Rozanov [8] and even if they are more restrictive than the so-called Rosenblatt strong mixing coefficients α_k they can be computed in many situations. For instance, we refer to the work by Doob [6] for sufficient conditions on Markov chains to be geometrically absolutely regular or by Mokkadem [13] for mild conditions ensuring vector ARMA processes to be also geometrically β-mixing.

In all the paper, we will assume that the underlying probability space $(\Omega,\mathcal{A},\mathbb{P})$ is rich enough to contain a sequence $(\epsilon_i)_{i\in\mathbb{Z}}=(\delta_i,\eta_i)_{i\in\mathbb{Z}}$ of iid random variables with uniform distribution over $[0,1]^2$, independent of $(\mathbf{X}_i)_{i\geqslant 0}$. In addition, the following notations will be used : $\log x:=\ln x$, $\log_2 x=\frac{\log x}{\log 2}$, we write $\mathbf{0}$ for the zero matrix and $\mathbb{I}_d$ for the $d\times d$ identity matrix, we use the curly inequalities to denote the semidefinite ordering i.e. $\mathbf{0}\preceq\mathbb{X}$ means that $\mathbb{X}$ is positive semidefinite.

Theorem 1 *Let $(\mathbb{X}_i)_{i\geqslant 1}$ be a family of self-adjoint random matrices of size d. Assume that* (5) *holds and that there exists a positive constant M such that for any $i\ge 1$,*

$$\mathbb{E}(\mathbb{X}_i)=\mathbf{0}\quad\textit{and}\quad\lambda_{\max}(\mathbb{X}_i)\le M\quad\textit{almost surely.} \tag{6}$$

Then there exists a universal positive constant C such that for any $x>0$ and any integer $n\geqslant 2$,

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{i=1}^n\mathbb{X}_i\big)\ge x\Big)\le d\exp\Big(-\frac{Cx^2}{v^2n+c^{-1}M^2+xM\gamma(c,n)}\Big)\,,$$

where

$$v^2=\sup_{K\subseteq\{1,\dots,n\}}\frac{1}{\mathrm{Card}K}\lambda_{\max}\Big(\mathbb{E}\big(\sum_{i\in K}\mathbb{X}_i\big)^2\Big) \tag{7}$$

and

$$\gamma(c,n)=\frac{\log n}{\log 2}\max\Big(2,\frac{32\log n}{c\log 2}\Big)\,. \tag{8}$$

In the definition of v^2 above, the maximum is taken over all nonempty subsets $K\subseteq\{1,\dots,n\}$.

To prove the deviation inequality stated in Theorem 1, we shall use the matrix Chernoff bound (1). The theorem will then follow from the following control of the matrix log-Laplace transform that is proved in Section 4.3: Under the conditions of Theorem 1, for any positive t such that $tM<1/\gamma(c,n)$, we have

$$\log\mathbb{E}\mathrm{Tr}\Big(\exp\big(t\sum_{i=1}^n\mathbb{X}_i\big)\Big)\le\log d+\frac{t^2n\big(15v+2M/(cn)^{1/2}\big)^2}{1-tM\gamma(c,n)}\,.$$

As proved in Section 4.2.4 of [10], this inequality together with Jensen's inequality leads to the following upper bound for the expectation of the largest eigenvalue of $\sum_{i=1}^n\mathbb{X}_i$: Under the conditions of Theorem 1,

$$\mathbb{E}\lambda_{\max}\Big(\sum_{i=1}^n\mathbb{X}_i\Big)\le 30v\sqrt{n\log d}+4Mc^{-1/2}\sqrt{\log d}+M\gamma(c,n)\log d\,.$$

3 Applications

Let $(\tau_k)_k$ be a stationary sequence of real-valued random variables such that $\|\tau_1\|_\infty \le 1$ a.s. Consider a family $(\mathbb{Y}_k)_k$ of independent real and symmetric $d \times d$ random matrices which is independent of $(\tau_k)_k$. For any $i = 1, \dots, n$, let $\mathbb{X}_i = \tau_i \mathbb{Y}_i$ and note that in this case

$$\beta_k = \beta(\sigma(\tau_i\,,\, i \le 0), \sigma(\tau_i\,,\, i \ge k))\,.$$

Corollary 2 *Assume that there exists a positive constant c such that $\beta_k \le e^{-c(k-1)}$ for any $k \ge 1$ and suppose that each random matrix $\mathbb{Y}_k$ satisfies*

$$\mathbb{E}\mathbb{Y}_k = \mathbf{0} \quad , \quad \lambda_{\max}(\mathbb{Y}_k) \le M \quad \textit{and} \quad \lambda_{\min}(\mathbb{Y}_k) \ge -M \quad \textit{almost surely.}$$

Then for any $t > 0$ and any integer $n \ge 2$,

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{k=1}^{n} \tau_k \mathbb{Y}_k\big) \ge t\Big) \le d \exp\Big(-\frac{Ct^2}{nM^2\mathbb{E}(\tau_0^2) + M^2 + tM(\log n)^2}\Big)\,,$$

where C is a positive constant depending only on c.

Proof. The above corollary follows by noting that for any $K \subseteq \{1, \dots, n\}$

$$\Sigma_K := \mathbb{E}\big(\sum_{k \in K} \tau_k \mathbb{Y}_k\big)^2 = \sum_{k \in K} \mathbb{E}(\tau_k^2)\mathbb{E}(\mathbb{Y}_k^2) = \mathbb{E}(\tau_0^2) \sum_{k \in K} \mathbb{E}(\mathbb{Y}_k^2),$$

which, by Weyl's inequality, implies that $\lambda_{\max}(\Sigma_K) \le M^2 \mathrm{Card}(K)\mathbb{E}(\tau_0^2)$. Therefore, we infer that $v^2 \le M^2\mathbb{E}(\tau_0^2)$. $\square$

We consider now another model for which Theorem 1 can be applied. Let $(X_k)_{k \in \mathbb{Z}}$ be a geometrically absolutely regular sequence of real-valued centered random variables. That is, there exists a positive constant c_0 such that for any $k \ge 1$,

$$\sup_{\ell \in \mathbb{Z}} \beta\big(\sigma(X_i\,,\, i \le \ell), \sigma(X_i\,,\, i \ge k + \ell)\big) \le e^{-c_0(k-1)}\,. \tag{9}$$

For any $i = 1, \dots, n$, let $\mathbb{X}_i$ be the $d \times d$ random matrix defined by $\mathbb{X}_i = \mathbf{C}_i \mathbf{C}_i^T - \mathbb{E}(\mathbf{C}_i \mathbf{C}_i^T)$ where $\mathbf{C}_i = (X_{(i-1)d+1}, \dots, X_{id})^T$. Note that in this case,

$$\beta_k = \sup_{\ell \in \mathbb{Z}} \beta\big(\sigma(\mathbf{C}_i, i \le \ell), \sigma(\mathbf{C}_i\,,\, i \ge \ell + k)\big) \le e^{-c_0 d(k-1)}\,.$$

for any $k \ge 1$.

Corollary 3 *Assume that $(X_k)_k$ satisfies (9). Suppose in addition that there exists a positive constant M satisfying $\sup_k \|X_k\|_\infty \le M$ a.s. Then, for any $x > 0$ and any integer $n \ge 2$*

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{i=1}^{n} \mathbb{X}_i\big) \ge x\Big) \le d \exp\Big(-\frac{Cx^2}{ndM^4 + dM^4 + xM^2(d\log n + \log^2 n)}\Big)\,,$$

where C is a positive constant depending only on c_0.

Proof. For any $i \in \{1, \dots, n\}$, note that $\lambda_{\max}(\mathbb{X}_i) \le \lambda_{\max}(\mathbf{C}_i \mathbf{C}_i^T)$ implying that $\lambda_{\max}(\mathbb{X}_i) \le dM^2$ a.s. To get the desired result, it remains to control v^2. We have for any $K \subseteq \{1, \dots, N\}$,

$$\Sigma_K := \mathbb{E}\big(\sum_{i \in K} \mathbb{X}_i\big)^2 = \sum_{i,j \in K} \mathrm{Cov}(\mathbf{C}_i \mathbf{C}_i^T, \mathbf{C}_j \mathbf{C}_j^T)$$

and we note that the $(k,\ell)^{th}$ component of Σ_K is

$$(\Sigma_K)_{k,\ell} = \Big[\mathbb{E}\big(\sum_{i\in K}\mathbb{X}_i\big)^2\Big]_{k,\ell} = \sum_{i,j\in K}\sum_{s=1}^{d}\mathrm{Cov}\big(X_{(i-1)d+k}\,X_{(i-1)d+s}\,,X_{(j-1)d+s}\,X_{(j-1)d+\ell}\big)\,.$$

Therefore we infer by Gerschgorin's theorem that

$$\begin{aligned}&\big|\lambda_{\max}\big(\Sigma_K\big)\big| \\ &\qquad \leqslant \sup_k\sum_{\ell=1}^{d}|(\Sigma_K)_{k,\ell}| \leqslant \sup_k\sum_{i,j\in K}\sum_{\ell=1}^{d}\sum_{s=1}^{d}\big|\mathrm{Cov}\big(X_{(i-1)d+k}\,X_{(i-1)d+s}\,,X_{(j-1)d+s}\,X_{(j-1)d+\ell}\big)\big|\,.\end{aligned}$$

After tedious computations involving Ibragimov's covariance inequality (see [7]), we infer that $v^2 \leq c_1 d M^4$ where c_1 is a positive constant depending only on c_0. Applying Theorem 1 with these upper bounds ends the proof. $\square$

4 Proof of Theorem 1

The proof of Theorem 1 being very technical, it is divided into several steps. In Section 4.1, we first collect some technical preliminary lemmas that will be necessary all along the proof. In Section 4.2, we give the main ingredient to prove our Bernstein-type inequality, namely: a bound for the Laplace transform of the partial sum, indexed by a suitable Cantor-type set, of the self-adjoint random matrices under consideration (see Proposition 7 and Section 4.2.1 for the construction of this suitable Cantor-set). As quoted in the introduction, this key result is based on a decoupling lemma which is stated in Section 4.2.2. The proof of Theorem 1 is completed in Section 4.3.

4.1 Preliminary materials

The following lemma is due to Tropp [17]. Under the form stated below, it is a combination of his Lemmas 3.4 and 6.7 together with the proof of his Corollary 3.7.

Lemma 4 *Let K be a finite subset of positive integers. Consider a family $(\mathbb{U}_k)_{k\in K}$ of $d\times d$ self-adjoint random matrices that are mutually independent. Assume that for any $k\in K$,*

$$\mathbb{E}(\mathbb{U}_k) = \mathbf{0} \quad \textit{and} \quad \lambda_{\max}(\mathbb{U}_k) \leq B \ \textit{a.s.}$$

where B is a positive constant. Then for any $t>0$,

$$\mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\sum_{k\in K}\mathbb{U}_k}\big) \leq d\exp\Big(t^2 g(tB)\lambda_{\max}\Big(\sum_{k\in K}\mathbb{E}(\mathbb{U}_k^2)\Big)\Big)\,, \tag{10}$$

where $g(x) = x^{-2}(\mathrm{e}^x - x - 1)$.

The next lemma is an adaptation of Lemma 3 in [12] to the case of the log-Laplace transform of any sum of $d\times d$ self-adjoint random matrices.

Lemma 5 *Let $\mathbb{U}_0, \mathbb{U}_1, \dots$ be a sequence of $d\times d$ self-adjoint random matrices. Assume that there exists positive constants $\sigma_0, \sigma_1, \dots$ and $\kappa_0, \kappa_1, \dots$ such that, for any $i\geq 0$ and any t in $[0, 1/\kappa_i[$,*

$$\log\mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\,\mathbb{U}_i}\big) \leq C_d + (\sigma_i t)^2/(1-\kappa_i t)\,,$$

where C_d is a positive constant depending only on d. Then, for any positive n and any t in $[0, 1/(\kappa_0 + \kappa_1 + \cdots + \kappa_n)[$,

$$\log \mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\sum_{k=0}^{n} \mathbb{U}_k}\big) \le C_d + (\sigma t)^2/(1 - \kappa t),$$

where $\sigma = \sigma_0 + \sigma_1 + \cdots + \sigma_n$ and $\kappa = \kappa_0 + \kappa_1 + \cdots + \kappa_n$.

Proof. Lemma 5 follows from the case $n = 1$ by induction on n. For any $t \ge 0$, let

$$L(t) = \log \mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t(\mathbb{U}_0+\mathbb{U}_1)}\big)$$

and notice that by the Golden-Thompson inequality,

$$L(t) \le \log \mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\,\mathbb{U}_0}\mathrm{e}^{t\,\mathbb{U}_1}\big)\,. \tag{11}$$

Define the functions γ_i by

$$\gamma_i(t) = (\sigma_i t)^2/(1 - \kappa_i t) \ \text{ for } t \in [0, 1/\kappa_i[\ \text{ and } \gamma_i(t) = +\infty \ \text{ for } t \ge 1/\gamma_i\,,$$

and recall the non-commutative Hölder inequality (see for instance exercise 1.3.9 in [16]): if A and B are $d \times d$ self-adjoint random matrices then, for any $1 \le p, q \le \infty$ with $p^{-1} + q^{-1} = 1$,

$$|\mathrm{Tr}(AB)| \le \|A\|_{\mathcal{S}^p} \|B\|_{\mathcal{S}^q}\,, \tag{12}$$

where $\|A\|_{\mathcal{S}^p} = \|(\lambda_i(A))_{i=1}^d\|_{\ell_n^p} = \Big(\sum_{i=1}^d |\lambda_i(A)|^p\Big)^{1/p}$ (resp. $\|B\|_{\mathcal{S}^q}$) is the p-Schatten norm of A (resp the q-Schatten norm of B).

Starting from (11) and applying (12) with $A = \mathrm{e}^{t\,\mathbb{U}_0}$ and $B = \mathrm{e}^{t\,\mathbb{U}_0}$, we derive that for any $t > 0$ and any $p \in]1, \infty[$

$$L(t) \le \log \mathbb{E}\Big(\|\mathrm{e}^{t\,\mathbb{U}_0}\|_{\mathcal{S}^p}\|\mathrm{e}^{t\,\mathbb{U}_1}\|_{\mathcal{S}^q}\Big)\,,$$

which gives by applying Hölder's inequality

$$L(t) \le p^{-1} \log \mathbb{E}\|\mathrm{e}^{t\,\mathbb{U}_0}\|_{\mathcal{S}^p}^p + q^{-1} \log \mathbb{E}\|\mathrm{e}^{t\,\mathbb{U}_1}\|_{\mathcal{S}^q}^q\,.$$

Observe now that since $\mathbb{U}_0$ is self-adjoint

$$\|\mathrm{e}^{t\,\mathbb{U}_0}\|_{\mathcal{S}^p}^p = \sum_{i=1}^d |\lambda_i(\mathrm{e}^{t\,\mathbb{U}_0})|^p = \sum_{i=1}^d \lambda_i(\mathrm{e}^{tp\,\mathbb{U}_0}) = \mathrm{Tr}\big(\mathrm{e}^{tp\,\mathbb{U}_0}\big)\,,$$

and similarly $\|\mathrm{e}^{t\,\mathbb{U}_1}\|_{\mathcal{S}^q}^q = \mathrm{Tr}\big(\mathrm{e}^{tq\,\mathbb{U}_1}\big)$. So, overall,

$$L(t) \le p^{-1} \log \mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{tp\,\mathbb{U}_0}\big) + q^{-1} \log \mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{tq\,\mathbb{U}_1}\big)\,. \tag{13}$$

For any t in $[0, 1/\kappa[$, take $u_t = (\sigma_0/\sigma)(1 - \kappa t) + \kappa_0 t$ (here $\kappa = \kappa_0 + \kappa_1$ and $\sigma = \sigma_0 + \sigma_1$). With this choice $1 - u_t = (\sigma_1/\sigma)(1 - \kappa t) + \kappa_1 t$, so that u_t belongs to $]0, 1[$. Applying Inequality (13) with $p = 1/u_t$, we get that for any t in $[0, 1/\kappa[$,

$$L(t) \le u_t \gamma_0(t/u_t) + (1 - u_t)\gamma_1(t/(1 - u_t)) = (\sigma t)^2/(1 - \kappa t)\,,$$

which completes the proof of Lemma 5. □

Next lemma allows coupling and is due to Berbee [3].

Lemma 6 *Let X and Y be two random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking their values in Borel spaces B_1 and B_2 respectively. Assume that $(\Omega, \mathcal{A}, \mathbb{P})$ is rich enough to contain a random variable δ with uniform distribution over $[0,1]$ independent of (X, Y). Then there exists a random variable $Y^* = f(X, Y, \delta)$ where f is a measurable function from $B_1 \times B_2 \times [0,1]$ into B_2 such that Y^* is independent of X, has the same distribution as Y and*

$$\mathbb{P}(Y \neq Y^*) = \beta(\sigma(X), \sigma(Y)) .$$

Let us note that the β-mixing coefficient $\beta(\sigma(X), \sigma(Y))$ has the following equivalent definition:

$$\beta(\sigma(X), \sigma(Y)) = \frac{1}{2} \| P_{X,Y} - P_X \otimes P_Y \| , \tag{14}$$

where $P_{X,Y}$ is the joint distribution of (X, Y) and P_X and P_Y are respectively the distributions of X and Y and, for two positive measures μ and ν, the notation $\|\mu - \nu\|$ denotes the total variation of $\mu - \nu$.

4.2 A key result

The next proposition is the main ingredient to prove Theorem 1. It is based on a suitable construction of a subset K_A of $\{1, \dots, A\}$ for which it is possible to give a good upper bound for the Laplace transform of $\sum_{i \in K_A} \mathbb{X}_i$. Its proof is based on the decoupling Lemma 8 below that allows to compare $\mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t \sum_{i \in K_A} \mathbb{X}_i}\big)$ with the same quantity but replacing $\sum_{i \in K_A} \mathbb{X}_i$ with a sum of independent blocks.

Proposition 7 *Let $(\mathbb{X}_i)_{i \geqslant 1}$ be as in Theorem 1. Let A be a positive integer larger than 2. Then there exists a subset K_A of $\{1, \dots, A\}$ with $\mathrm{Card}(K_A) \geqslant A/2$, such that for any positive t such that $tM \leq \min\big(\frac{1}{2}, \frac{c \log 2}{32 \log A}\big)$,*

$$\log \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t \sum_{i \in K_A} \mathbb{X}_i}\Big) \leq \log d + 4 \times 3.1 t^2 A v^2 + \frac{9(tM)^2}{c} \mathrm{e}^{-3c/(32tM)} , \tag{15}$$

where v^2 is defined in (7).

The proof of this proposition is divided into several steps.

4.2.1 Construction of a Cantor-like subset K_A

As in [11] and [12], the set K_A will be a finite union of 2^ℓ disjoint sets of consecutive integers with same cardinality spaced according to a recursive 'Cantor"-like construction. Let

$$\delta = \frac{\log 2}{2 \log A} \quad \text{and} \quad \ell := \ell_A = \sup\{k \in \mathbb{N}^* \, : \, \frac{A\delta(1-\delta)^{k-1}}{2^k} \geqslant 2\} .$$

Note that $\ell \leq \log A / \log 2$ and $\delta \leq 1/2$ (since $A \geqslant 2$). Let $n_0 = A$ and for any $j \in \{1, \dots, \ell\}$,

$$n_j = \lceil \frac{A(1-\delta)^j}{2^j} \rceil \quad \text{and} \quad d_{j-1} = n_{j-1} - 2n_j . \tag{16}$$

For any nonnegative x, the notation $\lceil x \rceil$ means the smallest integer which is larger or equal to x. Note that for any $j \in \{0, \dots, \ell - 1\}$,

$$d_j \geqslant \frac{A\delta(1-\delta)^j}{2^j} - 2 \geqslant \frac{A\delta(1-\delta)^j}{2^{j+1}} , \tag{17}$$

where the last inequality comes from the definition of ℓ. Moreover,

$$n_\ell \leq \frac{A(1-\delta)^\ell}{2^\ell} + 1 \leq \frac{A(1-\delta)^\ell}{2^{\ell-1}}, \tag{18}$$

where the last inequality comes from the fact that $\dfrac{A\delta(1-\delta)^{\ell-1}}{2^\ell} \times \dfrac{1-\delta}{\delta} \geqslant 2$ by the definition of ℓ and the fact that $\delta \leq 1/2$

To construct K_A we proceed as follows. At the first step, we divide the set $\{1, \ldots, A\}$ into three disjoint subsets of consecutive integers: $I_{1,1}$, $I^*_{0,1}$ and $I_{1,2}$. These subsets are such that $\mathrm{Card}(I_{1,1}) = \mathrm{Card}(I_{1,2}) = n_1$ and $\mathrm{Card}(I^*_{0,1}) = d_0$. At the second step, each of the sets of integers $I_{1,i}$, $i = 1, 2$, is divided into three disjoint subsets of consecutive integers as follows: for any $i = 1, 2$, $I_{1,i} = I_{2,2i-1} \cup I^*_{1,i} \cup I_{2,2i}$ where $\mathrm{Card}(I_{2,2i-1}) = \mathrm{Card}(I_{2,2i}) = n_2$ and $\mathrm{Card}(I^*_{1,i}) = d_1$. Iterating this procedure we have constructed after j steps ($1 \leq j \leq \ell_A$), 2^j sets of consecutive integers, $I_{j,i}$, $i = 1, \ldots, 2^j$, each of cardinality n_j such that $a_{j,2k} - b_{j,2k-1} - 1 = d_{j-1}$ for any $k = 1, \ldots, 2^{j-1}$, where $a_{j,i} = \min\{k \in I_{j,i}\}$ and $b_{j,i} = \max\{k \in I_{j,i}\}$. Moreover if, for any $i = 1, \ldots, 2^{j-1}$, we set $I^*_{j-1,i} = \{b_{j,2i-1} + 1, \ldots, a_{j,2i} - 1\}$, then $I_{j-1,i} = I_{j,2i-1} \cup I^*_{j-1,i} \cup I_{j,2i}$. After ℓ steps we then have constructed 2^ℓ sets of consecutive integers, $I_{\ell,i}$, $i = 1, \ldots, 2^\ell$, each of cardinality n_ℓ such that $I_{\ell,2i-1}$ and $I_{\ell,2i}$ are spaced by $d_{\ell-1}$ integers. The set of consecutive integers K_A is then defined by

$$K_A = \bigcup_{k=1}^{2^\ell} I_{\ell,k}\,.$$

Note that

$$\{1, \ldots, A\} = K_A \cup (\cup_{j=0}^{\ell-1} \cup_{i=1}^{2^j} I^*_{j,i})$$

Therefore

$$\mathrm{Card}(\{1, \ldots, A\} \setminus K_A) = \sum_{j=0}^{\ell-1} \sum_{i=1}^{2^j} \mathrm{Card}(I^*_{j,i}) = \sum_{j=0}^{\ell-1} 2^j d_j = A - 2^\ell n_\ell\,.$$

But

$$A - 2^\ell n_\ell \leq A\big(1 - (1-\delta)^\ell\big) = A\delta \sum_{j=0}^{\ell-1} (1-\delta)^j \leq A\delta\ell \leq \frac{A}{2}\,. \tag{19}$$

Therefore $A \geqslant \mathrm{Card}(K_A) \geqslant A/2$.

In the rest of the proof, the following notation will be also useful: for any $k \in \{0, 1, \ldots, \ell\}$ and any $j \in \{1, \ldots, 2^k\}$, let

$$K_{k,j} := K_{A,k,j} = \bigcup_{i=(j-1)2^{\ell-k}+1}^{j2^{\ell-k}} I_{\ell,i} \tag{20}$$

Therefore $K_{0,1} = K_A$ and, for any $j \in \{1, \ldots, 2^\ell\}$, $K_{\ell,j} = I_{\ell,j}$. Moreover, for any $k \in \{1, \ldots, \ell\}$ and any $j \in \{1, \ldots, 2^{k-1}\}$, there are exactly d_{k-1} integers between $K_{k,2j-1}$ and $K_{k,2j}$.

4.2.2 A fundamental decoupling lemma

We start by introducing some notations, then we state the decoupling Lemma 8 below that is fundamental to prove Proposition 7. Let K_A be defined as in Step 1. In the rest of the proof, we will adopt the following notation. For any integer $m \in \{0, \ldots, \ell\}$, $(\mathbf{V}_j^{(m)})_{1 \leq j \leq 2^m}$ will denote a family of 2^m mutually independent random vectors defined on $(\Omega, \mathcal{A}, \mathbb{P})$, each of dimension $s_{d,\ell,m} := d^2 \mathrm{Card}(K_{m,j}) = d^2 2^{\ell-m} n_\ell$ and such that

$$\mathbf{V}_j^{(m)} =^{\mathcal{D}} (\mathbf{X}_i\,, i \in K_{m,j})\,. \tag{21}$$

The existence of such a family is ensured by the Skorohod lemma (see [15]). Indeed since $(\Omega, \mathcal{A}, \mathbb{P})$ is assumed to be large enough to contain a sequence $(\delta_i)_{i\in\mathbb{Z}}$ of iid random variables uniformly distributed on $[0,1]$ and independent of the sequence $(\mathbf{X}_i)_{i\geqslant 0}$, there exist measurable functions f_j such that the vectors $\mathbf{V}_j^{(m)} = f_j\big((\mathbf{X}_i\,, i \in K_{m,k})_{k=1,\dots,j}, \delta_j\big)$, $j = 1, \dots, 2^m$, are independent and satisfy (21).

Let $\pi_i^{(m)}$ be the i-th canonical projection from $\mathbb{K}^{s_{d,\ell,m}}$ onto $\mathbb{K}^{d^2}$, namely: for any vector $\mathbf{x} = (\mathbf{x}_i\,, i \in K_{m,j})$ of $\mathbb{K}^{s_{d,\ell,m}}$, $\pi_i^{(m)}(\mathbf{x}) = \mathbf{x}_i$. For any $i \in K_{m,j}$, let

$$\mathbf{X}_j^{(m)}(i) = \pi_i^{(m)}(\mathbf{V}_j^{(m)}) \ \text{ and } \ \mathbb{S}_j^{(m)} = \sum_{i\in K_{m,j}} \mathbb{X}_j^{(m)}(i)\,, \tag{22}$$

where $\mathbb{X}_j^{(m)}(i)$ is the $d \times d$ random matrix associated with $\mathbf{X}_j^{(m)}(i)$ (recall that this means that the (k,ℓ)-th entry of $\mathbb{X}_j^{(m)}(i)$ is the $((\ell-1)d+k)$-th coordinate of the vector $\mathbf{X}_j^{(m)}(i)$).

With the above notations, we have

$$\mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\sum_{i\in K_A}\mathbb{X}_i}\big) = \mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\,\mathbb{S}_1^{(0)}}\big)\,. \tag{23}$$

We are now in position to state the following lemma which will be a key step in the proof of Proposition 7 and allows decoupling when we deal with the Laplace transform of a sum of self adjoint random matrices.

Lemma 8 *Assume that* (6) *holds. Then for any $t > 0$ and any $k \in \{0, \dots, \ell-1\}$,*

$$\mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}}\Big) \le \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\sum_{j=1}^{2^{k+1}}\mathbb{S}_j^{(k+1)}}\Big)\Big(1 + \beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\Big)^{2^k}\,,$$

where $(\mathbb{S}_j^{(k)})_{j=1,\dots,2^k}$ is the family of mutually independent random matrices defined in (22).

Proof. Note that for any $k \in \{0, \dots, \ell-1\}$ and any $j \in \{1, \dots, 2^k\}$,

$$K_{k,j} = K_{k+1,2j-1} \cup K_{k+1,2j}$$

where the union is disjoint. Therefore

$$\mathbb{S}_j^{(k)} = \mathbb{S}_{j,1}^{(k)} + \mathbb{S}_{j,2}^{(k)} \ \text{ and } \ \mathbf{V}_j^{(k)} = (\mathbf{V}_{j,1}^{(k)}, \mathbf{V}_{j,2}^{(k)})\,,$$

where $\mathbb{S}_{j,1}^{(k)} := \sum_{i\in K_{k+1,2j-1}} \mathbb{X}_j^{(k)}(i)$, $\mathbb{S}_{j,2}^{(k)} := \sum_{i\in K_{k+1,2j}} \mathbb{X}_j^{(k)}(i)$, $\mathbf{V}_{j,1}^{(k)} := \big(\mathbf{X}_j^{(k)}(i)\,, i \in K_{k+1,2j-1}\big)$ and $\mathbf{V}_{j,2}^{(k)} := \big(\mathbf{X}_j^{(k)}(i)\,, i \in K_{k+1,2j}\big)$. Note that there are exactly d_k integers between $K_{k+1,2j-1}$ and $K_{k+1,2j}$ and that for any $i \in \{1, \dots, 2^{k+1}\}$,

$$\mathrm{Card}(K_{k+1,i}) = \mathrm{Card}(K_{k+1,1}) = 2^{\ell-(k+1)}n_\ell\,.$$

Recall that the probability space is assumed to be large enough to contain a sequence $(\delta_i, \eta_i)_{i\in\mathbb{Z}}$ of iid random variables uniformly distributed on $[0,1]^2$ independent of the sequence $(\mathbf{X}_i)_{i\geqslant 0}$. Therefore according to the remark on the existence of the family $(\mathbf{V}_j^{(m)})_{1\le j\le 2^m}$ made at the beginning of Section 4.2.2, the sequence $(\eta_i)_{i\in\mathbb{Z}}$ is independent of $(\mathbf{V}_j^{(m)})_{1\le j\le 2^m}$. According to Lemma 6 there exists a random vector $\widetilde{\mathbf{V}}_{1,2}^{(k)}$ of size $d^2\mathrm{Card}(K_{k+1,2})$ with the same law as $\mathbf{V}_{1,2}^{(k)}$ that is measurable with respect to $\sigma(\eta_1) \vee \sigma(\mathbf{V}_{1,1}^{(k)}) \vee \sigma(\mathbf{V}_{1,2}^{(k)})$, independent of $\sigma(\mathbf{V}_{1,1}^{(k)})$ and such that

$$\mathbb{P}(\widetilde{\mathbf{V}}_{1,2}^{(k)} \neq \mathbf{V}_{1,2}^{(k)}) = \beta\big(\sigma(\mathbf{V}_{1,1}^{(k)}), \sigma(\mathbf{V}_{1,2}^{(k)})\big) \le \beta_{d_k+1}\,,$$

where the inequality comes from the fact that, by relation (14), the quantity $\beta\big(\sigma(\mathbf{V}_{1,1}^{(k)}),\sigma(\mathbf{V}_{1,2}^{(k)})\big)$ depends only on the joint distribution of $(\mathbf{V}_{1,1}^{(k)},\mathbf{V}_{1,2}^{(k)})$ and therefore, by (21),

$$\beta\big(\sigma(\mathbf{V}_{1,1}^{(k)}),\sigma(\mathbf{V}_{1,2}^{(k)})\big)=\beta\big(\sigma(\mathbf{X}_i\,,i\in K_{k+1,1}),\ \sigma(\mathbf{X}_i\,,i\in K_{k+1,2})\big)\le\beta_{d_k+1}\,.$$

Note that by construction, $\widetilde{\mathbf{V}}_{1,2}^{(k)}$ is independent of $\sigma\big(\mathbf{V}_{1,1}^{(k)},(\mathbf{V}_j^{(k)})_{j=2,\dots,2^k}\big)$.
For any $i\in K_{k+1,2}$, let

$$\widetilde{\mathbf{X}}_{1,2}^{(k)}(i)=\pi_i^{(k+1)}(\widetilde{\mathbf{V}}_{1,2}^{(k)})\ \text{ and }\ \widetilde{\mathbb{S}}_{1,2}^{(k)}=\sum_{i\in K_{k+1,2}}\widetilde{\mathbb{X}}_{1,2}^{(k)}(i)\,,$$

where $\widetilde{\mathbb{X}}_{1,2}^{(k)}(i)$ is the $d\times d$ random matrix associated with the random vector $\widetilde{\mathbf{X}}_{1,2}^{(k)}(i)$.
With the notations above, we have

$$\begin{aligned}\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big)&=\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)}=\mathbf{V}_{1,2}^{(k)}}\mathrm{Tr}\exp\big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\big)\Big)+\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)}\neq\mathbf{V}_{1,2}^{(k)}}\mathrm{Tr}\exp\big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\big)\Big)\\&\le\mathbb{E}\mathrm{Tr}\exp\Big(t\mathbb{S}_{1,1}^{(k)}+t\widetilde{\mathbb{S}}_{1,2}^{(k)}+t\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}\Big)+\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)}\neq\mathbf{V}_{1,2}^{(k)}}\mathrm{Tr}\exp\big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\big)\Big)\,.\end{aligned}\tag{24}$$

(With usual convention, $\sum_{j=\ell}^{2^k}\mathbb{S}_j^{(k)}$ is the null vector if $\ell>2^k$). By Golden-Thompson inequality, we have

$$\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big)\le\mathrm{Tr}\Big(\mathrm{e}^{t\mathbb{S}_1^{(k)}}\cdot\mathrm{e}^{t\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}}\Big).$$

Hence, since $\sigma\big(\mathbf{V}_j^{(k)}\,,\,j=2,\dots,2^k\big)$ is independent of $\sigma\big(\mathbf{V}_{1,1}^{(k)},\mathbf{V}_{1,2}^{(k)},\widetilde{\mathbf{V}}_{1,2}^{(k)}\big)$, we get

$$\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)}\neq\mathbf{V}_{1,2}^{(k)}}\mathrm{Tr}\exp\big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\big)\Big)\le\mathrm{Tr}\Big(\mathbb{E}\big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)}\neq\mathbf{V}_{1,2}^{(k)}}\mathrm{e}^{t\,\mathbb{S}_1^{(k)}}\big)\cdot\mathbb{E}\exp\big(t\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}\big)\Big)\,.$$

Note now the following fact: if $\mathbb{U}$ is a $d\times d$ self-adjoint random matrix with entries defined on $(\Omega,\mathcal{A},\mathbb{P})$ and such that $\lambda_{\max}(\mathbb{U})\le b$ a.s., then for any $\Gamma\in\mathcal{A}$,

$$\mathbf{1}_\Gamma\mathrm{e}^{\mathbb{U}}\preceq\mathrm{e}^b\mathbb{I}_d\mathbf{1}_\Gamma\ \text{ a.s. and so }\ \lambda_{\max}\mathbb{E}\big(\mathbf{1}_\Gamma\mathrm{e}^{\mathbb{U}}\big)\le\mathrm{e}^b\mathbb{P}(\Gamma)\,.$$

Therefore if we consider $\mathbb{V}$ a $d\times d$ self-adjoint random matrix with entries defined on $(\Omega,\mathcal{A},\mathbb{P})$, the following inequality is valid:

$$\mathrm{Tr}\big(\mathbb{E}(\mathbf{1}_\Gamma\mathrm{e}^{\mathbb{U}})\mathbb{E}(\mathrm{e}^{\mathbb{V}})\big)\le\mathrm{e}^b\mathbb{P}(\Gamma)\cdot\mathbb{E}\mathrm{Tr}(\mathrm{e}^{\mathbb{V}})\,.\tag{25}$$

Notice now that $(\mathbf{X}_1^{(k)}(i)\,,i\in K_{k,1})$ has the same distribution as $(\mathbf{X}_i\,,i\in K_{k,1})$. Therefore $\lambda_{\max}(\mathbf{X}_1^{(k)}(i))\le M$ a.s. for any i, implying by Weyl's inequality that

$$\lambda_{\max}(t\,\mathbb{S}_1^{(k)})\le tM\mathrm{Card}(K_{k,1})=tM2^{\ell-k}n_\ell\ \text{ a.s.}$$

Hence, applying (25) with $b=tM2^{\ell-k}n_\ell$, $\Gamma=\{\widetilde{\mathbf{V}}_{1,2}^{(k)}\neq\mathbf{V}_{1,2}^{(k)}\}$ and $\mathbb{V}=t\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}$, and taking into account that $\mathbb{P}(\Gamma)\le\beta_{d_k+1}$, we obtain

$$\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)}\neq\mathbf{V}_{1,2}^{(k)}}\mathrm{Tr}\exp\big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\big)\Big)\le\beta_{d_k+1}\mathrm{e}^{tn_\ell2^{\ell-k}M}\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}\Big)\,.\tag{26}$$

Note now that if $\mathbb{V}$ and $\mathbb{W}$ are two independent random matrices with entries defined on $(\Omega, \mathcal{A}, \mathbb{P})$ and such that $\mathbb{E}(\mathbb{W}) = \mathbf{0}$ then

$$\mathbb{E}\mathrm{Tr}\exp(\mathbb{V}) = \mathbb{E}\mathrm{Tr}\exp\big(\mathbb{E}\big(\mathbb{V}+\mathbb{W}|\sigma(\mathbb{V})\big)\big)\,.$$

Since $\mathrm{Tr}\circ\exp$ is convex, it follows from Jensen's inequality applied to the conditional expectation that

$$\mathbb{E}\mathrm{Tr}\exp(\mathbb{V}) \leq \mathbb{E}\big(\mathbb{E}\big(\mathrm{Tr}\,\mathrm{e}^{\mathbb{V}+\mathbb{W}}|\sigma(\mathbb{V})\big)\big) = \mathbb{E}\big(\mathrm{Tr}\,\mathrm{e}^{\mathbb{V}+\mathbb{W}}\big)\,. \tag{27}$$

Since $\mathbb{E}(\mathbb{X}_1^{(k)}(i)) = \mathbb{E}(\mathbb{X}_i) = \mathbf{0}$ for any $i \in K_{k,1}$ and $\sigma(\mathbb{S}_{1,1}^{(k)}, \widetilde{\mathbb{S}}_{1,2}^{(k)})$ is independent of $\sigma(\mathbb{S}_j^{(k)}, j = 2, \ldots, 2^k)$, we can apply the inequality above with $\mathbb{W} = t(\mathbb{S}_{1,1}^{(k)} + \widetilde{\mathbb{S}}_{1,2}^{(k)})$ and $\mathbb{V} = t\sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}$. Therefore, starting from (26) and using (27), we get

$$\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}_{1,2}^{(k)} \neq \mathbf{V}_{1,2}^{(k)}}\mathrm{Tr}\exp\big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\big)\Big) \leq \beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\,\mathbb{E}\mathrm{Tr}\exp\Big(t\big(\mathbb{S}_{1,1}^{(k)} + \widetilde{\mathbb{S}}_{1,2}^{(k)} + \sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}\big)\Big)\,. \tag{28}$$

Starting from (24) and considering (28), it follows that

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big) \leq (1 + \beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M})\,\mathbb{E}\mathrm{Tr}\exp\Big(t\big(\mathbb{S}_{1,1}^{(k)} + \widetilde{\mathbb{S}}_{1,2}^{(k)} + \sum_{j=2}^{2^k}\mathbb{S}_j^{(k)}\big)\Big)\,. \tag{29}$$

The proof of Lemma 8 will then be achieved after having iterated this procedure $2^k - 1$ times more. For the sake of clarity, let us explain the way to go from the j-th step to the $(j+1)$-th step.

At the end of the j-th step, assume that we have constructed with the help of the coupling Lemma 6, j random vectors $\widetilde{\mathbf{V}}_{i,2}^{(k)}$, $i = 1, \ldots, j$, each of dimension $d^2\mathrm{Card}(K_{k+1,1})$ and satisfying the following properties: for any i in $\{1, \ldots, j\}$, $\widetilde{\mathbf{V}}_{i,2}^{(k)}$ is a measurable function of $(\mathbf{V}_{i,1}^{(k)}, \mathbf{V}_{i,2}^{(k)}, \eta_i)$, it has the same distribution as $\mathbf{V}_{i,2}^{(k)}$, is such that $\mathbb{P}(\widetilde{\mathbf{V}}_{i,2}^{(k)} \neq \mathbf{V}_{i,2}^{(k)}) \leq \beta_{d_k+1}$, is independent of $\mathbf{V}_{i,1}^{(k)}$ and it satisfies

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}\Big) \leq \big(1 + \beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\big)^j \cdot \mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i=1}^{j}(\mathbb{S}_{i,1}^{(k)} + \widetilde{\mathbb{S}}_{i,2}^{(k)}) + t\sum_{i=j+1}^{2^k}\mathbb{S}_i^{(k)}\Big)\,, \tag{30}$$

where we have implemented the following notation:

$$\widetilde{\mathbb{S}}_{i,2}^{(k)} = \sum_{r \in K_{k+1,2i}}\widetilde{\mathbb{X}}_{i,2}^{(k)}(r)\,. \tag{31}$$

In the notation above, $\widetilde{\mathbb{X}}_{i,2}^{(k)}(r)$ is the $d \times d$ random matrix associated with the random vector $\widetilde{\mathbf{X}}_{i,2}^{(k)}(r)$ of $\mathbb{K}^{d^2}$ defined by

$$\widetilde{\mathbf{X}}_{i,2}^{(k)}(r) = \pi_r^{(k+1)}(\widetilde{\mathbf{V}}_{i,2}^{(k)}) \text{ for any } r \in K_{k+1,2i}\,.$$

Note that the induction assumption above has been proven at the beginning of the proof for $j = 1$. Moreover, note that since, for any $m \in \{1, \ldots, \ell\}$, $(\mathbf{V}_j^{(m)})_{1 \leq j \leq 2^m}$ is a family of independent random vectors, the random vectors $\widetilde{\mathbf{V}}_{i,2}^{(k)}$, $i = 1, \ldots, j$, defined above are also such that, for any $i \in \{1, \ldots, j\}$, $\widetilde{\mathbf{V}}_{i,2}^{(k)}$ is independent of $\sigma\big((\mathbf{V}_{\ell,1}^{(k)})_{\ell=1,\ldots,i}, (\widetilde{\mathbf{V}}_{\ell,2}^{(k)})_{\ell=1,\ldots,i-1}, (\mathbf{V}_\ell^{(k)})_{\ell=i+1,\ldots,2^k}\big)$.

Now to show that the induction hypothesis also holds at step $j+1$, we proceed as follows. By Lemma 6, there exists a random vector $\widetilde{\mathbf{V}}_{j+1,2}^{(k)}$ of size $d^2\mathrm{Card}(K_{k+1,1})$ with the same law as

$\mathbf{V}^{(k)}_{j+1,2}$, measurable with respect to $\sigma(\eta_{j+1}) \vee \sigma(\mathbf{V}^{(k)}_{j+1,1}) \vee \sigma(\mathbf{V}^{(k)}_{j+1,2})$, independent of $\sigma(\mathbf{V}^{(k)}_{j+1,1})$ and such that

$$\mathbb{P}(\widetilde{\mathbf{V}}^{(k)}_{j+1,2} \neq \mathbf{V}^{(k)}_{j+1,2}) \leq \beta_{d_k+1} \,. \tag{32}$$

(The inequality above comes again from (21) and the equivalent definition (14) of the β-coefficients). Note that by construction, $\sigma\big((\mathbf{V}^{(k)}_{i,1})_{i=1,\dots,j+1}, (\widetilde{\mathbf{V}}^{(k)}_{i,2})_{i=1,\dots,j}, (\mathbf{V}^{(k)}_{i})_{i=j+2,\dots,2^k}\big)$ and $\sigma(\widetilde{\mathbf{V}}^{(k)}_{j+1,2})$ are independent. With the notation (31), we have the following decomposition:

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2}) + t\sum_{i=j+1}^{2^k}\mathbb{S}^{(k)}_{i}\Big) \leq \mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i=1}^{j+1}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2}) + t\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}\Big)$$
$$+\,\mathbb{E}\Big(\mathbf{1}_{\widetilde{\mathbf{V}}^{(k)}_{j+1,2}\neq\mathbf{V}^{(k)}_{j+1,2}}\mathrm{Tr}\exp\big(t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2}) + t\sum_{i=j+1}^{2^k}\mathbb{S}^{(k)}_{i}\big)\Big)\,. \tag{33}$$

Using Golden-Thompson inequality, we have

$$\mathrm{Tr}\exp\Big(t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2}) + t\sum_{i=j+1}^{2^k}\mathbb{S}^{(k)}_{i}\Big)$$
$$\leq \mathrm{Tr}\Big(\exp\big(t\,\mathbb{S}^{(k)}_{j+1}\big)\cdot\exp\big(t\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2}) + t\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}\big)\Big)\,.$$

Hence, since the sigma algebra generated by $\big((\mathbf{V}^{(k)}_{i,1})_{i=1,\dots,j}, (\widetilde{\mathbf{V}}^{(k)}_{i,2})_{i=1,\dots,j}, (\mathbf{V}^{(k)}_{i})_{i=j+2,\dots,2^k}\big)$ is independent of that generated by $\big(\mathbf{V}^{(k)}_{j+1,1}, \mathbf{V}^{(k)}_{j+1,2}, \widetilde{\mathbf{V}}^{(k)}_{j+1,2}\big)$, we get

$$\mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\big(\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+\sum_{i=j+1}^{2^k}\mathbb{S}^{(k)}_{i}\big)}\mathbf{1}_{\widetilde{\mathbf{V}}^{(k)}_{j+1,2}\neq\mathbf{V}^{(k)}_{j+1,2}}\Big)$$
$$\leq \mathrm{Tr}\Big(\mathbb{E}\mathrm{e}^{t\big(\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}\big)}\cdot\mathbb{E}\big(\mathrm{e}^{t\,\mathbb{S}^{(k)}_{j+1}}\mathbf{1}_{\widetilde{\mathbf{V}}^{(k)}_{j+1,2}\neq\mathbf{V}^{(k)}_{j+1,2}}\big)\Big)\,. \tag{34}$$

By Weyl's inequality,

$$\lambda_{\max}\big(t\,\mathbb{S}^{(k)}_{j+1}\big) \leq t\sum_{r\in K_{k,j+1}}\lambda_{\max}(\mathbb{X}^{(k)}_{j+1}(r)) \text{ a.s.}$$

Using that $\mathbf{V}^{(k)}_{j+1} =^{\mathcal{D}} (\mathbf{X}_i\,, i\in K_{k,j+1})$ and that $\lambda_{\max}(\mathbb{X}_i)\leq M$ a.s. for any i, it follows that

$$\lambda_{\max}\big(t\,\mathbb{S}^{(k)}_{j+1}\big) \leq tM\mathrm{Card}(K_{k,j+1}) = tM2^{\ell-k}n_k \text{ a.s.}$$

In addition, we notice that $\mathbb{S}^{(k)}_{j+1,1}+\widetilde{\mathbb{S}}^{(k)}_{j+1,2}$ is independent of $\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}$ and, since $\widetilde{\mathbf{V}}^{(k)}_{j+1,2} =^{\mathcal{D}} \mathbf{V}^{(k)}_{j+1,2}$ and $\mathbf{V}^{(k)}_{j+1} =^{\mathcal{D}} (\mathbf{X}_i\,, i\in K_{k,j+1})$, $\mathbb{E}(\mathbb{S}^{(k)}_{j+1,1}+\widetilde{\mathbb{S}}^{(k)}_{j+1,2}) = \mathbf{0}$. Therefore, starting from (34) and taking into account (32), an application of inequality (25) with $b = tM2^{\ell-k}n_\ell$, $\Gamma = \{\widetilde{\mathbf{V}}^{(k)}_{j+1,2}\neq\mathbf{V}^{(k)}_{j+1,2}\}$ and $\mathbb{V} = t\big(\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}\big)$, followed by an application of inequality (27) with $\mathbb{W} = t(\mathbb{S}^{(k)}_{j+1,1}+\widetilde{\mathbb{S}}^{(k)}_{j+1,2})$, gives

$$\mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\big(\sum_{i=1}^{j}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+\sum_{i=j+1}^{2^k}\mathbb{S}^{(k)}_{i}\big)}\mathbf{1}_{\widetilde{\mathbf{V}}^{(k)}_{j+1,2}\neq\mathbf{V}^{(k)}_{j+1,2}}\Big)$$
$$\leq \beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\times\mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\big(\sum_{i=1}^{j+1}(\mathbb{S}^{(k)}_{i,1}+\widetilde{\mathbb{S}}^{(k)}_{i,2})+\sum_{i=j+2}^{2^k}\mathbb{S}^{(k)}_{i}\big)}\Big)\,. \tag{35}$$

Therefore, starting from (33) and using (35), we get

$$\mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\big(\sum_{i=1}^{j}(\mathbb{S}_{i,1}^{(k)}+\widetilde{\mathbb{S}}_{i,2}^{(k)})+\sum_{i=j+1}^{2^k}\mathbb{S}_i^{(k)}\big)}\Big) \\ \leq \big(1+\beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\big)\times \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\big(\sum_{i=1}^{j+1}(\mathbb{S}_{i,1}^{(k)}+\widetilde{\mathbb{S}}_{i,2}^{(k)})+\sum_{i=j+2}^{2^k}\mathbb{S}_i^{(k)}\big)}\Big),$$

which combined with (30) implies that

$$\mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}}\Big) \leq \big(1+\beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\big)^{j+1}\times \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\big(\sum_{i=1}^{j+1}(\mathbb{S}_{i,1}^{(k)}+\widetilde{\mathbb{S}}_{i,2}^{(k)})+\sum_{i=j+2}^{2^k}\mathbb{S}_i^{(k)}\big)}\Big),$$

proving the induction hypothesis for the step $j+1$. Finally 2^k steps of the procedure lead to

$$\mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\sum_{j=1}^{2^k}\mathbb{S}_j^{(k)}}\Big) \leq \big(1+\beta_{d_k+1}\mathrm{e}^{tn_\ell 2^{\ell-k}M}\big)^{2^k}\times \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\sum_{i=1}^{2^k}(\mathbb{S}_{i,1}^{(k)}+\widetilde{\mathbb{S}}_{i,2}^{(k)})}\Big). \tag{36}$$

To end the proof of the lemma it suffices to notice the following facts: the random vectors $\mathbf{V}_{i,1}^{(k)}, \widetilde{\mathbf{V}}_{i,2}^{(k)}$, $i=1,\dots,2^k$, are mutually independent and such that $\mathbf{V}_{i,1}^{(k)} =^{\mathcal{D}} \mathbf{V}_{2i-1}^{(k+1)}$ and $\mathbf{V}_{i,2}^{(k)} =^{\mathcal{D}} \mathbf{V}_{2i}^{(k+1)}$. In addition, the random vectors $\mathbf{V}_i^{(k+1)}$, $i=1,\dots,2^{k+1}$, are mutually independent. This obviously implies that

$$\mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\sum_{i=1}^{2^k}(\mathbb{S}_{i,1}^{(k)}+\widetilde{\mathbb{S}}_{i,2}^{(k)})}\Big) = \mathbb{E}\mathrm{Tr}\Big(\mathrm{e}^{t\sum_{i=1}^{2^{k+1}}\mathbb{S}_i^{(k+1)}}\Big),$$

which ends the proof of the lemma. □

4.2.3 Proof of Proposition 7

We shall prove Inequality (15) with K_A defined in Section 4.2.1.

Let us prove it first in the case where $0 < tM \leq 4/A$. Since by Weyl's inequality,

$$\lambda_{\max}\big(\sum_{i\in K_A}\mathbb{X}_i\big) \leq \sum_{i\in K_A}\lambda_{\max}\big(\mathbb{X}_i\big) \leq AM,$$

and $\mathbb{E}(\mathbb{X}_i) = \mathbf{0}$ for any $i \in K_A$, it follows by using Lemma 4 applied with $K = \{1\}$ and $\mathbb{U}_1 = \sum_{i\in K_A}\mathbb{X}_i$ that, for any $t>0$,

$$\mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\sum_{i\in K_A}\mathbb{X}_i}\big) \leq d\exp\Big(t^2 g(tAM)\lambda_{\max}\Big(\mathbb{E}\big(\sum_{i\in K_A}\mathbb{X}_i\big)^2\Big)\Big).$$

Therefore by the definition of v^2, since g is increasing, $tAM < 4$ and $g(4) \leq 3.1$, we get

$$\mathbb{E}\mathrm{Tr}\big(\mathrm{e}^{t\sum_{i\in K_A}\mathbb{X}_i}\big) \leq d\exp(3.1\times At^2v^2),$$

proving then (15).

We prove now Inequality (15) in the case where $4/A < tM \leq \min\big(\frac{1}{2}, \frac{c\log 2}{32\log A}\big)$. Let

$$\kappa = \frac{c}{8} \ \text{ and } \ k(t) = \inf\Big\{k\in\mathbb{Z} : A((1-\delta)/2)^k \leq \min\big(\frac{\kappa}{(tM)^2}, A\big)\Big\}. \tag{37}$$

Note that if $t^2M^2 \leq \kappa/A$ then $k(t)=0$ whereas $k(t)\geq 1$ if $t^2M^2 > \kappa/A$. In addition by the selection of ℓ_A, $A((1-\delta)/2)^\ell < 4/\delta$. Therefore $k(t) \leq \ell_A$ since $(tM)^2 \leq c\delta/32$. Then, starting from (23), considering the selection of $k(t)$ and using Lemma 8, we get by induction that

$$\mathbb{E}\mathrm{Tr}\exp\big(t\sum_{i\in K_A}\mathbb{X}_i\big) \leq \prod_{k=0}^{k(t)-1}\Big(1+\beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\Big)^{2^k}\,\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^{k(t)}}\mathbb{S}_j^{(k(t))}\Big), \tag{38}$$

with the usual convention that $\prod_{k=0}^{-1} a_k = 1$. Note that in the inequality above, $(\mathbb{S}_j^{(k(t))})_{j=1,\dots,2^{k(t)}}$ is a family of mutually independent random matrices defined in (22). They are then constructed from a family $(\mathbf{V}_j^{(k(t))})_{1\le j\le 2^{k(t)}}$ of $2^{k(t)}$ mutually independent random vectors that satisfy (21). Therefore we have that, for any $j \in 1,\dots,2^{k(t)}$, $\mathbb{S}_j^{(k(t))} =^{\mathcal{D}} \sum_{i\in K_{k(t),j}} \mathbb{X}_i$. Moreover, according to the remark on the existence of the family $(\mathbf{V}_j^{(k(t))})_{1\le j\le 2^{k(t)}}$ made at the beginning of Section 4.2.2, the entries of each random matrix $\mathbb{S}_j^{(k(t))}$ are measurable functions of $(\mathbf{X}_i, \delta_i)_{i\in\mathbb{Z}}$.

The rest of the proof consists of giving a suitable upper bound for $\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^{k(t)}} \mathbb{S}_j^{(k(t))}\Big)$. With this aim, let p be a positive integer to be chosen later such that

$$2p \le \mathrm{Card}(K_{k(t),j}) := q\,. \tag{39}$$

Note that $q = 2^{\ell-k(t)} n_\ell$ and by (19)

$$q \ge \frac{A}{2^{k(t)+1}}\,.$$

Therefore if $k(t)=0$ then $q\ge A/2$ implying that $q\ge 2$ (since we have $4/A < tM \le 1$). Now if $k(t)\ge 1$ and therefore if $t^2M^2 > \kappa/A$, by the definition of $k(t)$, we have $q \ge \frac{\kappa}{(tM)^2}$ and then $q\ge 2$ since $(tM)^2 \le \kappa/2$. Hence in all cases, $q\ge 2$ implying that the selection of a positive integer p satisfying (39) is always possible.

Let $m_{q,p} = [q/(2p)]$. For any $j\in\{1,\dots,2^{k(t)}\}$, we divide $K_{k(t),j}$ into $2m_{q,p}$ consecutive intervals $(J_{j,i}^{(k(t))}$, $1\le i\le 2m_{q,p})$ each containing p consecutive integers plus a remainder interval $J_{j,2m_{q,p}+1}^{(k(t))}$ containing $r = q-2pm_{q,p}$ consecutive integers. Note that this last interval contains at most $2p-1$ integers. Let $\mathbb{X}_j^{(k(t))}(k)$ be the $d\times d$ random matrix associated with the random vector $\mathbf{X}_j^{(k(t))}(k)$ defined in (22) and define

$$\mathbb{Z}_{j,i}^{(k(t))} = \sum_{k\in K_{k(t),j}\cap J_{j,i}^{(k(t))}} \mathbb{X}_j^{(k(t))}(k)\,. \tag{40}$$

With this notation

$$\mathbb{S}_j^{(k(t))} = \sum_{i=1}^{m_{q,p}+1} \mathbb{Z}_{j,2i-1}^{(k(t))} + \sum_{i=1}^{m_{q,p}} \mathbb{Z}_{j,2i}^{(k(t))}\,.$$

Since $\mathrm{Tr}\circ\exp$ is a convex function, we get

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{j=1}^{2^{k(t)}} \mathbb{S}_j^{(k(t))}\Big) \le \frac{1}{2}\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}+1} \mathbb{Z}_{j,2i-1}^{(k(t))}\Big) + \frac{1}{2}\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}} \mathbb{Z}_{j,2i}^{(k(t))}\Big)\,. \tag{41}$$

We start by giving an upper bound for $\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}} \mathbb{Z}_{j,2i}^{(k(t))}\Big)$. With this aim, let us define the following vectors

$$\mathbf{U}_{j,i}^{(k(t))} = \big(\mathbf{X}_j^{(k(t))}(k)\,,\ k\in K_{k(t),j}\cap J_{j,i}^{(k(t))}\big) \ \text{ and }\ \mathbf{W}_j^{(k(t))} = \big(\mathbf{U}_{j,i}^{(k(t))}\,,\ i\in\{1,\dots,2m_{q,p}+1\}\big)\,. \tag{42}$$

Proceeding by induction and using the coupling lemma 6, one can construct random vectors $\mathbf{U}_{j,2i}^{*(k(t))}$, $j=1,\dots,2^{k(t)}$, $i=1,\dots,m_{q,p}$, that satisfy the following properties:

(i) $(\mathbf{U}_{j,2i}^{*(k(t))}, (j,i)\in\{1,\dots,2^{k(t)}\}\times\{1,\dots,m_{q,p}\})$ is a family of mutually independent random vectors,

(ii) $\mathbf{U}_{j,2i}^{*(k(t))}$ has the same distribution as $\mathbf{U}_{j,2i}^{(k(t))}$,

(iii) $\mathbb{P}(\mathbf{U}_{j,2i}^{*(k(t))} \neq \mathbf{U}_{j,2i}^{(k(t))}) \le \beta_{p+1}$.

Let us explain the construction. Recall first that $(\Omega, \mathcal{A}, \mathbb{P})$ is assumed to be rich enough to contain a sequence $(\eta_i)_{i\in\mathbb{Z}}$ of iid random variables with uniform distribution over $[0,1]$ independent of $(\mathbf{X}_i, \delta_i)_{i\in\mathbb{Z}}$ (the sequence $(\delta_i)_{i\in\mathbb{Z}}$ has been used to construct the independent random matrices $\mathbb{S}_j^{(k(t))}$, $j = 1, \ldots, k(t)$, involved in inequality (38)). For any $j \in \{1, \ldots, 2^{k(t)}\}$, let $\mathbf{U}_{j,2}^{*(k(t))} = \mathbf{U}_{j,2}^{(k(t))}$, and construct the random vectors $\mathbf{U}_{j,2i}^{*(k(t))}$, $i = 2, \ldots, m_{q,p}$, recursively from $(\mathbf{U}_{j,2\ell}^{*(k(t))}, 1 \le \ell \le i-1)$ as follows. According to Lemma 6, there exists a random vector $\mathbf{U}_{j,2i}^{*(k(t))}$ such that

$$\mathbf{U}_{j,2i}^{*(k(t))} = f_{i,j}\big((\mathbf{U}_{j,2\ell}^{*(k(t))})_{1\le\ell\le i-1}, \mathbf{U}_{j,2i}^{(k(t))}, \eta_{i+(j-1)2^{k(t)}}\big) \tag{43}$$

where $f_{i,j}$ is a measurable function, $\mathbf{U}_{j,2i}^{*(k(t))}$ has the same law as $\mathbf{U}_{j,2i}^{(k(t))}$, is independent of $\sigma\big(\mathbf{U}_{j,2\ell}^{*(k(t))}, 1 \le \ell \le i-1\big)$ and

$$\mathbb{P}(\mathbf{U}_{j,2i}^{*(k(t))} \neq \mathbf{U}_{j,2i}^{(k(t))}) = \beta\big(\sigma\big(\mathbf{U}_{j,2\ell}^{*(k(t))}, 1 \le \ell \le i-1\big), \sigma(\mathbf{U}_{j,2i}^{(k(t))})\big) \le \beta_{p+1} .$$

By construction, for any fixed $j \in \{1, \ldots, 2^{k(t)}\}$, the random vectors $\mathbf{U}_{j,2i}^{*(k(t))}$, $i = 1, \ldots, m_{q,p}$, are mutually independent. In addition, by (43) and the fact that $(\mathbf{W}_j^{(k(t))}, j = 1, \ldots, 2^{k(t)})$ is a family of mutually independent random vectors, we note that $(\mathbf{U}_{j,2i}^{*(k(t))}, (i,j) \in \{1, \ldots, m_{q,p}\} \times \{1, \ldots, 2^{k(t)}\})$ is also so. Therefore the constructed random vectors $\mathbf{U}_{j,2i}^{*(k(t))}$ $i = 1, \ldots, m_{q,p}$, $j = 1, \ldots, 2^{k(t)}$, satisfy Items (i) and (ii) above. Moreover, by (43), we have

$$\sigma\big(\mathbf{U}_{j,2\ell}^{*(k(t))}, 1 \le \ell \le i-1\big) \subseteq \sigma\big(\mathbf{U}_{j,2\ell}^{(k(t))}, 1 \le \ell \le i-1\big) \vee \sigma\big(\eta_{\ell+(j-1)2^{k(t)}}, 1 \le \ell \le i-1\big) .$$

Since $(\eta_i)_{i\in\mathbb{Z}}$ is independent of $(\mathbf{X}_i, \delta_i)_{i\in\mathbb{Z}}$, we have

$$\beta\big(\sigma\big(\mathbf{U}_{j,2\ell}^{*(k(t))}, 1 \le \ell \le i-1\big), \sigma(\mathbf{U}_{j,2i}^{(k(t))})\big) \le \beta\big(\sigma\big(\mathbf{U}_{j,2\ell}^{(k(t))}, 1 \le \ell \le i-1\big), \sigma(\mathbf{U}_{j,2i}^{(k(t))})\big) .$$

By relation (14), the quantity $\beta\big(\sigma\big(\mathbf{U}_{j,2\ell}^{(k(t))}, 1 \le \ell \le i-1\big), \sigma(\mathbf{U}_{j,2i}^{(k(t))})\big)$ depends only on the joint distribution of $\big((\mathbf{U}_{j,2\ell}^{(k(t))})_{1\le\ell\le i-1}, \mathbf{U}_{j,2i}^{(k(t))}\big)$. By the definition (42) of the $\mathbf{U}_{j,\ell}^{(k(t))}$'s, the definition (22) of the $\mathbf{X}_j^{(k(t))}(k)$'s and (21), we infer that

$$\begin{aligned}&\beta\big(\sigma\big(\mathbf{U}_{j,2\ell}^{(k(t))}, 1 \le \ell \le i-1\big), \sigma(\mathbf{U}_{j,2i}^{(k(t))})\big)\\ &\qquad = \beta\big(\sigma\big(\mathbf{X}_k\,,\, k \in \cup_{\ell=1}^{i-1} K_{k(t),j} \cap J_{j,2\ell}^{(k(t))}\big), \sigma\big(\mathbf{X}_k\,,\, k \in K_{k(t),j} \cap J_{j,2i}^{(k(t))}\big)\big) \le \beta_{p+1} .\end{aligned}$$

So, overall, the constructed random vectors $\mathbf{U}_{j,2i}^{*(k(t))}$ $i = 1, \ldots, m_{q,p}$, $j = 1, \ldots, 2^{k(t)}$, satisfy also Item (iii) above.

Denote now

$$\mathbf{X}_{j,2i}^{*(k(t))}(\ell) = \pi_\ell(\mathbf{U}_{j,2i}^{*(k(t))})$$

where $\pi_i^{(m)}$ is the ℓ-th canonical projection from $\mathbb{K}^{pd^2}$ onto $\mathbb{K}^{d^2}$, namely: for any vector $\mathbf{x} = (\mathbf{x}_i\,, i \in \{1, \ldots, p\})$ of $\mathbb{K}^{pd^2}$, $\pi_\ell(\mathbf{x}) = \mathbf{x}_\ell$. Let $\mathbb{X}_{j,2i}^{*(k(t))}(\ell)$ be the $d \times d$ random matrix associated with $\mathbf{X}_{j,2i}^{*(k(t))}(\ell)$ and define, for any $i = 1 \ldots, m_{q,p}$,

$$\mathbb{Z}_{j,2i}^{*(k(t))} = \sum_{\ell \in K_{k(t),j} \cap J_{j,2i}^{(k(t))}} \mathbb{X}_{j,2i}^{*(k(t))}(\ell) .$$

Observe that by Item (ii) above, $\mathbb{Z}^{*(k(t))}_{j,2i} =^{\mathcal{D}} \mathbb{Z}^{(k(t))}_{j,2i}$ (where we recall that $\mathbb{Z}^{(k(t))}_{j,2i}$ is defined by (40)) and by Item (i), the random matrices $\mathbb{Z}^{*(k(t))}_{j,2i}$, $i = 1, \dots, m_{q,p}$, $j = 1, \dots, 2^{k(t)}$, are mutually independent. The aim now is to prove that the following inequality is valid:

$$\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{(k(t))}_{j,2i}\Big) \le \Big(1+(m_{q,p}-1)\mathrm{e}^{qtM}\beta_{p+1}\Big)^{2^{k(t)}}\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{*(k(t))}_{j,2i}\Big). \quad (44)$$

Obviously, this can be done by induction if we can show that, for any ℓ in $\{1, \dots, 2^{k(t)}\}$,

$$\begin{aligned}&\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{\ell-1}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{*(k(t))}_{j,2i} + 2t\sum_{j=\ell}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{(k(t))}_{j,2i}\Big)\\ &\qquad \le \Big(1+(m_{q,p}-1)\mathrm{e}^{qtM}\ \beta_{p+1}\Big)\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{\ell}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{*(k(t))}_{j,2i} + 2t\sum_{j=\ell+1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{(k(t))}_{j,2i}\Big). \quad (45)\end{aligned}$$

To prove the inequality above, we set

$$\mathbb{C}_{\ell-1,\ell}(t) = 2t\sum_{j=1}^{\ell-1}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{*(k(t))}_{j,2i} + 2t\sum_{j=\ell}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{(k(t))}_{j,2i}$$

and we write

$$\begin{aligned}\mathbb{E}\mathrm{Tr}\exp\big(\mathbb{C}_{\ell-1,\ell}(t)\big) &= \mathbb{E}\Big(\prod_{i=2}^{m_{q,p}}\mathbf{1}_{\mathbf{U}^{(k(t))}_{\ell,2i}=\mathbf{U}^{*(k(t))}_{\ell,2i}}\mathrm{Tr}\exp\big(\mathbb{C}_{\ell-1,\ell}(t)\big)\Big)\\ &\qquad + \mathbb{E}\Big(\mathbf{1}_{\exists i\in\{2,\dots,m_{q,p}\}\,:\,\mathbf{U}^{(k(t))}_{\ell,2i}\neq\mathbf{U}^{*(k(t))}_{\ell,2i}}\mathrm{Tr}\exp\big(\mathbb{C}_{\ell-1,\ell}(t)\big)\Big)\\ &\le \mathbb{E}\mathrm{Tr}\exp\big(\mathbb{C}_{\ell,\ell+1}(t)\big) + \mathbb{E}\Big(\mathbf{1}_{\exists i\in\{2,\dots,m_{q,p}\}\,:\,\mathbf{U}^{(k(t))}_{\ell,2i}\neq\mathbf{U}^{*(k(t))}_{\ell,2i}}\mathrm{Tr}\exp\big(\mathbb{C}_{\ell-1,\ell}(t)\big)\Big). \quad (46)\end{aligned}$$

Note that the sigma algebra generated by the random vectors $(\mathbf{U}^{*(k)}_{j,2i})_{i\in\{1,\dots,m_{q,p}\},j\in\{1,\dots,\ell-1\}}$ and $(\mathbf{U}^{(k)}_{j,2i})_{i\in\{1,\dots,m_{q,p}\},\,j\in\{\ell+1,\dots,2^{k(t)}\}}$ is independent of $\sigma\big((\mathbf{U}^{(k)}_{\ell,2i},\mathbf{U}^{*(k)}_{\ell,2i})_{i\in\{1,\dots,m_{q,p}\}}\big)$. This fact together with the Golden-Thomson inequality give

$$\begin{aligned}&\mathbb{E}\Big(\mathbf{1}_{\exists i\in\{2,\dots,m_{q,p}\}\,:\,\mathbf{U}^{(k(t))}_{\ell,2i}\neq\mathbf{U}^{*(k(t))}_{\ell,2i}}\mathrm{Tr}\exp\big(\mathbb{C}_{\ell-1,\ell}(t)\big)\Big)\\ &\quad\le \mathrm{Tr}\Big(\mathbb{E}\Big(\exp\big(2t\sum_{j=1}^{\ell-1}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{*(k(t))}_{j,2i} + 2t\sum_{j=\ell+1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{(k(t))}_{j,2i}\big)\Big)\\ &\qquad\qquad\times\ \mathbb{E}\Big(\mathbf{1}_{\exists i\in\{2,\dots,m_{q,p}\}\,:\,\mathbf{U}^{(k(t))}_{\ell,2i}\neq\mathbf{U}^{*(k(t))}_{\ell,2i}}\exp\big(2t\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{(k(t))}_{\ell,2i}\big)\Big)\Big).\end{aligned}$$

By Weyl's inequality and (21), we infer that, almost surely,

$$\lambda_{\max}\big(2t\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{(k(t))}_{\ell,2i}\big) \le 2t\sum_{i=1}^{m_{q,p}}\sum_{k\in K_{k(t),\ell}\cap J^{(k(t))}_{\ell,2i}}\lambda_{\max}\big(\mathbb{X}_k\big) \le 2tm_{q,p}pM \le tqM. \quad (47)$$

Therefore, applying (25) with $b = tqM$, $\Gamma = \{\exists i \in \{2, \dots, m_{q,p}\} : \mathbf{U}^{(k(t))}_{\ell,2i} \neq \mathbf{U}^{*(k(t))}_{\ell,2i}\}$ and $\mathbb{V} = 2t\sum_{j=1}^{\ell-1}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{*(k(t))}_{j,2i} + 2t\sum_{j=\ell+1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{(k(t))}_{j,2i}$ and taking into account that

$$\mathbb{P}(\Gamma) \le \sum_{i=2}^{m_{q,p}}\mathbb{P}(\mathbf{U}^{(k(t))}_{\ell,2i} \neq \mathbf{U}^{*(k(t))}_{\ell,2i}) \le (m_{q,p}-1)\beta_{p+1},$$

we get

$$\mathbb{E}\Big(\mathbf{1}_{\exists i\in\{2,\dots,m_{q,p}\}\,:\,\mathbf{U}^{(k(t))}_{\ell,2i}\neq\mathbf{U}^{*(k(t))}_{\ell,2i}}\operatorname{Tr}\exp\big(\mathbb{C}_{\ell-1,\ell}(t)\big)\Big)$$
$$\leq (m_{q,p}-1)\beta_{p+1}\mathrm{e}^{qtM}\,\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{\ell-1}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{*(k(t))}_{j,2i}+2t\sum_{j=\ell+1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{(k(t))}_{j,2i}\Big)\,.$$

Using that the sigma algebra generated by the random vectors $(\mathbf{U}^{*(k)}_{j,2i})_{i\in\{1,\dots,m_{q,p}\},\,j\in\{1,\dots,\ell-1\}}$ and $(\mathbf{U}^{(k)}_{j,2i})_{i\in\{1,\dots,m_{q,p}\},j\in\{\ell+1,\dots,2^{k(t)}\}}$ is independent of $\sigma\big((\mathbf{U}^{*(k)}_{\ell,2i})_{i\in\{1,\dots,m_{q,p}\}}\big)$, and noticing that by construction, $\mathbb{E}(\mathbb{Z}^{*(k(t))}_{\ell,2i})=\mathbb{E}(\mathbb{Z}^{(k(t))}_{\ell,2i})=\mathbf{0}$, an application of inequality (27) then gives

$$\mathbb{E}\Big(\mathbf{1}_{\exists i\in\{2,\dots,m_{q,p}\}\,:\,\mathbf{U}^{(k(t))}_{\ell,2i}\neq\mathbf{U}^{*(k(t))}_{\ell,2i}}\operatorname{Tr}\exp\big(\mathbb{C}_{\ell-1,\ell}(t)\big)\Big)$$
$$\leq \beta_{p+1}(m_{q,p}-1)\mathrm{e}^{qtM}\,\mathbb{E}\mathrm{Tr}\exp\big(\mathbb{C}_{\ell,\ell+1}(t)\big)\,. \quad (48)$$

Starting from (46) and taking into account (48), inequality (45) follows and so does inequality (44).

With the same arguments as above and with obvious notations, we infer that

$$\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p+1}}\mathbb{Z}^{(k(t))}_{j,2i-1}\Big)$$
$$\leq\Big(1+m_{q,p}\mathrm{e}^{2qtM}\beta_{p+1}\Big)^{2^{k(t)}}\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p+1}}\mathbb{Z}^{*(k(t))}_{j,2i-1}\Big)\,. \quad (49)$$

Note that to get the above inequality, we used instead of (47) that, almost surely,

$$\lambda_{\max}\big(2t\sum_{i=1}^{m_{q,p+1}}\mathbb{Z}^{(k(t))}_{\ell,2i-1}\big)\leq 2t\sum_{i=1}^{m_{q,p+1}}\sum_{k\in K_{k(t),\ell}\cap J^{(k(t))}_{\ell,2i-1}}\lambda_{\max}\big(\mathbb{X}_k\big)$$
$$\leq 2Mt(m_{q,p}p+q-2pm_{q,p})=2Mt(q-pm_{q,p})\leq Mt(q+2p)\leq 2tqM\,.$$

Starting from (38) and taking into account (41), (44) and (49), we then derive

$$\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i\in K_A}\mathbb{X}_i\Big)\leq\Big(1+m_{q,p}\mathrm{e}^{2qtM}\beta_{p+1}\Big)^{2^{k(t)}}\prod_{k=0}^{k(t)-1}\Big(1+\beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\Big)^{2^k}$$
$$\times\Big(\frac{1}{2}\mathbb{E}\mathrm{Tr}\exp\big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{*(k(t))}_{j,2i}\big)+\frac{1}{2}\mathbb{E}\mathrm{Tr}\exp\big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}+1}\mathbb{Z}^{*(k(t))}_{j,2i-1}\big)\Big)\,. \quad (50)$$

Now we choose

$$p=\Big[\frac{2}{tM}\Big]\wedge\Big[\frac{q}{2}\Big]\,.$$

Note that the random vectors $(\mathbb{Z}^{*(k(t))}_{j,2i-1})_{i,j}$ are mutually independent and centered. Moreover,

$$2\lambda_{\max}(\mathbb{Z}^{*(k(t))}_{j,2i-1})\leq 2Mp\leq\frac{4}{t}\quad\text{a.s.}$$

Therefore by using Lemma 4 together with the definition of v^2 and the fact that $2^{k(t)}(m_{q,p}+1)p \leq 2^{k(t)}q \leq A$, we get

$$\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}+1}\mathbb{Z}^{*(k(t))}_{j,2i-1}\Big) \leq d\exp(4\times 3.1\times At^2v^2)\,. \tag{51}$$

Similarly, we obtain that

$$\mathbb{E}\mathrm{Tr}\exp\Big(2t\sum_{j=1}^{2^{k(t)}}\sum_{i=1}^{m_{q,p}}\mathbb{Z}^{*(k(t))}_{j,2i}\Big) \leq d\exp(4\times 3.1\times At^2v^2)\,. \tag{52}$$

Next, by using Condition (5) and (19), we get

$$\log\Big(1+m_{q,p}\mathrm{e}^{2tqM}\beta_{p+1}\Big)^{2^{k(t)}} \leq 2^{k(t)}m_{q,p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp} \leq \frac{A}{2p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp}\,. \tag{53}$$

Several situations can occur. Either $(tM)^2 \leq \kappa/A$ and in this case $k(t)=0$ implying that $A/2 \leq q \leq A \leq \kappa/(tM)^2$. If in addition $q \geq 4/(tM)$ then $p=[2/(tM)] \geq 1/tM$ (since $tM \leq 1$) and

$$\frac{A}{2p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp} \leq \frac{AtM}{2}\mathrm{e}^{2\kappa/(tM)}\mathrm{e}^{-c/(tM)} \leq \frac{AtM}{2}\mathrm{e}^{-3c/(4tM)} \leq \frac{(tM)^2}{c}\mathrm{e}^{-3c/(16tM)}\,,$$

where we have used that $\log_2 A \leq \frac{c}{32tM}$, $A \geq 2$, and $\mathrm{e}^{-3c/(8tM)} \leq \frac{8tM}{3c}$ for the last inequality. If otherwise $q < 4/(tM)$ then $p=[q/2] \geq q/4$. Hence, since $2tM \leq c/16$ (since $\log A \geq \log 2$) and $tM > 4/A$,

$$\frac{A}{2p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp} \leq \frac{2A}{q}\mathrm{e}^{-3cq/16} \leq 4\mathrm{e}^{-3cA/32} \leq AtM\mathrm{e}^{-3c/(8tM)} \leq \frac{(tM)^2}{c}\mathrm{e}^{-3c/(32tM)}\,,$$

where we have used that $A/2 \leq q$ for the second inequality, and that $\log_2 A \leq \frac{c}{32tM}$, $A \geq 2$ and $\mathrm{e}^{-3c/(16tM)} \leq \frac{16tM}{3c}$ for the last one.

Either $(tM)^2 > \kappa/A$ and in this case $k(t) \geq 1$ and by using (19) and the definition of $k(t)$, we have

$$q \geq \frac{A}{2^{k(t)+1}} \geq \frac{\kappa}{4(tM)^2}\,. \tag{54}$$

If in addition $q \geq 4/(tM)$ then $p=[2/(tM)] \geq 1/tM$, and by (18) and the definition of $k(t)$,

$$q \leq 2A\frac{(1-\delta)^\ell}{2^{k(t)}} \leq \frac{2\kappa}{(tM)^2}\,.$$

It follows that

$$\frac{A}{2p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp} \leq \frac{AtM}{2}\mathrm{e}^{4\kappa/(tM)}\mathrm{e}^{-c/(tM)} \leq \frac{AtM}{2}\mathrm{e}^{-c/(2tM)} \leq \frac{(tM)^2}{c}\mathrm{e}^{-c/(8tM)}\,,$$

where we have used that $\log_2 A \leq \frac{c}{32tM}$, $A \geq 2$ and $\mathrm{e}^{-c/(4tM)} \leq \frac{4tM}{c}$ for the last inequality. Now if $q < 4/(tM)$ then $p=[q/2] \geq q/4$. Hence, using again the fact that $2tM \leq c/16$ combined with (54), we get

$$\frac{A}{2p}\mathrm{e}^{2tqM}\mathrm{e}^{-cp} \leq \frac{8A(tM)^2}{\kappa}\mathrm{e}^{-3cq/16} \leq \frac{8^2A(tM)^2}{c}\mathrm{e}^{-\frac{3c^2}{16\times 4\times 8(tM)^2}} \leq \frac{8(tM)^2}{c}\mathrm{e}^{-3c/(32tM)}\,,$$

where we have used that $\log_2 A \leq \frac{c^2}{(32tM)^2}$ and $A \geq 2$ for the last inequality.

So, overall, starting from (53), we get

$$\log\Big(1+m_{q,p}\mathrm{e}^{2tqM}\beta_{p+1}\Big)^{2^{k(t)}} \le \frac{8(tM)^2}{c}\mathrm{e}^{-3c/(32tM)}\,. \tag{55}$$

We handle now the term $\prod_{k=0}^{k(t)-1}\Big(1+\beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\Big)^{2^k}$ only in the case where $(\kappa/A)^{1/2}<tM$, otherwise this term is equal to one. By taking into account (5), (17), (18) and the fact that $tM\le c\delta/8$, we have

$$\begin{aligned}\log\prod_{k=0}^{k(t)-1}\Big(1+\beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\Big)^{2^k} &\le \sum_{k=0}^{k(t)-1}2^k\exp\Big(-c\frac{A\delta(1-\delta)^k}{2^{k+1}}+2tM\frac{A(1-\delta)^\ell}{2^k}\Big)\\ &\le \sum_{k=0}^{k(t)-1}2^k\exp\Big(-c\frac{A\delta(1-\delta)^k}{2^{k+2}}\Big)\\ &\le 2^{k(t)}\exp\Big(-\frac{Ac\delta(1-\delta)^{k(t)-1}}{2^{k(t)+1}}\Big)\,.\end{aligned}$$

By the definition of $k(t)$, we have $A\dfrac{(1-\delta)^{k(t)-1}}{2^{k(t)-1}}>\dfrac{\kappa}{(tM)^2}$. Therefore $2^{k(t)}\le 2A\frac{(tM)^2}{\kappa}$. Moreover

$$Ac\delta\frac{(1-\delta)^{k(t)-1}}{2^{k(t)+1}}>\frac{c\kappa\delta}{4(tM)^2}\geqslant\frac{2\kappa}{tM}\,,$$

since $tM\le c\delta/8$. It follows that

$$\log\prod_{k=0}^{k(t)-1}\Big(1+\beta_{d_k+1}\mathrm{e}^{tMn_\ell 2^{\ell-k}}\Big)^{2^k}\le 2A\frac{(tM)^2}{\kappa}\exp\big(-2\kappa/(tM)\big)\le\frac{(tM)^2}{c}\mathrm{e}^{-3c/(32tM)}\,, \tag{56}$$

where we have used the fact that $\log_2 A\le\frac{c}{32tM}$. So, overall, starting from (50) and considering the upper bounds (51), (52), (55) and (56), we get

$$\log\mathbb{E}\mathrm{Tr}\exp\Big(t\sum_{i\in K_A}\mathbb{X}_i\Big)\le\log d+4\times 3.1At^2v^2+\frac{9(tM)^2}{c}\mathrm{e}^{-3c/(32tM)}\,.$$

Therefore Inequality (15) also holds in the case where $4/A<tM\le\min\big(\frac12,\frac{c\log 2}{32\log A}\big)$. This ends the proof of the proposition. □

4.3 Proof of Theorem 1

Let $A_0=A=n$ and $\mathbb{Y}^{(0)}(k)=\mathbb{X}_k$ for any $k=1,\dots,A_0$. Let K_{A_0} be the discrete Cantor type set as defined from $\{1,\dots,A\}$ in Section 4.2.1. Let $A_1=A_0-\mathrm{Card}(K_{A_0})$ and define for any $k=1,\dots,A_1$,

$$\mathbb{Y}^{(1)}(k)=\mathbb{X}_{i_k}\text{ where }\{i_1,\dots,i_{A_1}\}=\{1,\dots,A\}\setminus K_A\,.$$

Now for $i\ge 1$, let K_{A_i} be defined from $\{1,\dots,A_i\}$ exactly as K_A is defined from $\{1,\dots,A\}$. Set $A_{i+1}=A_i-\mathrm{Card}(K_{A_i})$ and $\{j_1,\dots,j_{A_{i+1}}\}=\{1,\dots,A_i\}\setminus K_{A_i}$. Define now

$$\mathbb{Y}^{(i+1)}(k)=\mathbb{Y}^{(i)}(j_k)\text{ for }k=1,\dots,A_{i+1}\,,$$

and set

$$L=L_n=\inf\{j\in\mathbb{N}^*\,,\ A_j\le 2\}\,.$$

Note that, for any $i \in \{0, \ldots, L-1\}$, $A_i > 2$ and $\mathrm{Card}(K_{A_i}) \geqslant A_i/2$. Moreover $A_i \leq n2^{-i}$.

The following decomposition clearly holds

$$\sum_{k=1}^{n} \mathbb{X}_k = \sum_{i=0}^{L-1} \sum_{k \in K_{A_i}} \mathbb{Y}^{(i)}(k) + \sum_{k=1}^{A_L} \mathbb{Y}^{(L)}(k) . \tag{57}$$

Let

$$\mathbb{U}_i = \sum_{k \in K_{A_i}} \mathbb{Y}^{(i)}(k) \text{ for } 0 \leq i \leq L-1 \text{ and } \mathbb{U}_L = \sum_{k=1}^{A_L} \mathbb{Y}^{(L)}(k) ,$$

For any positive x, let

$$h(c, x) = \min\Big(\frac{1}{2}, \frac{c \log 2}{32 \log x}\Big) .$$

For any $i \in \{0, \ldots, L-1\}$, noticing that the self-adjoint random matrices $(Y^{(i)}(k))_k$ satisfy the condition (5) with the same constant c, we can apply Proposition 7 and get that for any positive t satisfying $tM < h(c, n/2^i)$,

$$\log \mathbb{E}\mathrm{Tr}\big(\exp(t\mathbb{U}_i)\big) \leq \log d + \frac{4t^2 n 2^{-i} (2v + \sqrt{3} \times 2^{5i/2} M/(n^{5/2}\sqrt{c}))^2}{1 - tM/h(c, n2^{-i})} . \tag{58}$$

On the other hand, by Weyl's inequality,

$$\lambda_{\max}\big(\mathbb{U}_L\big) \leq M A_L \leq 2M .$$

Therefore by using Lemma 4, for any positive t,

$$\mathbb{E}\mathrm{Tr}\big(\exp(t\mathbb{U}_L)\big) \leq d \exp\Big(t^2 g(2tM) \lambda_{\max}\big(\mathbb{E}\big(\mathbb{U}_L^2\big)\Big) .$$

Hence by the definition of v^2, for any positive t such that $tM < 1$, we get

$$\log \mathbb{E}\mathrm{Tr}\big(\exp(t\mathbb{U}_L)\big) \leq \log d + 2t^2 v^2 \leq \log d + \frac{2t^2 v^2}{1 - tM} . \tag{59}$$

Let

$$\kappa_i = \frac{M}{h(c, n/2^i)} \text{ for } 0 \leq i \leq L-1 \text{ and } \kappa_L = M$$

and

$$\sigma_i = 2\frac{\sqrt{n}}{2^{i/2}}\Big(2v + \sqrt{3} \times \frac{2^i M}{n\sqrt{c}}\Big) \text{ for } 0 \leq i \leq L-1 \text{ and } \sigma_L = v\sqrt{2} .$$

Since

$$L \leq \Big\lceil \frac{\log n - \log 2}{\log 2} \Big\rceil + 1 ,$$

we get

$$\sum_{i=0}^{L} \kappa_i \leq M\Big(\sum_{i=0}^{L-1} \frac{1}{h(c, n/2^i)} + 1\Big) \leq M \frac{\log n}{\log 2} \max\Big(2, \frac{32 \log n}{c \log 2}\Big) = M\gamma(c, n) .$$

Moreover

$$\begin{aligned}
\sum_{i=0}^{L} \sigma_i &= 2\sqrt{n} \sum_{i=0}^{L-1} 2^{-i/2}\Big(2v + \sqrt{3} \times \frac{2^{5i/2} M}{n^{5/2}\sqrt{c}}\Big) \; + \; v\sqrt{2} \\
&\leq 14\sqrt{n}v + 2c^{-1/2} n^{-2} M 2^{2L} + v\sqrt{2} \\
&\leq 15\sqrt{n}v + 2c^{-1/2} M .
\end{aligned}$$

Taking into account (58) and (59), we get overall by Lemma 5, that for any positive t such that $tM < 1/\gamma(c,n)$,

$$\log \mathbb{E}\mathrm{Tr}\Big(\exp\big(t \sum_{i=1}^{n} \mathbb{X}_i\big)\Big) \leq \log d + \frac{t^2 n \big(15v + 2M/(cn)^{1/2}\big)^2}{1 - tM\gamma(c,n)} := \gamma_n(t)\,. \tag{60}$$

To end the proof of the theorem, it suffices to notice that for any positive x

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{i=1}^{n} \mathbb{X}_i\big) \geqslant x\Big) \leq \inf_{t>0\,:\,tM\leq 1/\gamma(c,n)} \exp\big(-tx + \gamma_n(t)\big)\,,$$

where $\gamma_n(t)$ is defined in (60). □

Acknowledgements. This work was initiated when the first named author was visiting the third named author at the University of Alberta. She would like to thank Nicole Tomczak-Jaegermann and Alexander Litvak for their hospitality and the University of Alberta for the excellent working conditions.

References

[1] Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. Journal Probab.* **13**, 1000-1034.

[2] Ahlswede, R. and Winter, A. (2002). Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory*, **48**, no. 3, 569-579.

[3] Berbee H.C.P. (1979). Random walks with stationary increments and renewal theory. *Mathematical Centre Tracts*, **112**, Mathematisch Centrum, Amsterdam, 223 pp.

[4] Chatterjee, S. (2006). Concentration inequalities with exchangeable pairs. Ph.D. thesis, *Stanford Univ., Palo Alto..*

[5] Chatterjee, S. (2007). Steins method for concentration inequalities. *Probability theory and related fields*, **138**, no. 1, 305-321.

[6] Doob, J. L. (1953) *Stochastic processes.* Wiley, New York.

[7] Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Teor. Verojatnost. i Primenen.* **7**, 361-392.

[8] Kolmogorov, A.N., Rozanov, Y.A. (1960). On the strong mixing conditions for stationary gaussian sequences. *Theor. Probab. Appl.*, **5**, 204-207.

[9] Lieb, E. H. (1973). Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Adv. Math.* **11**, 267-288.

[10] Mackey, L., Jordan, M. I., Chen, R. Y., Farrell, B., Tropp, J. A. (2014). Matrix concentration inequalities via the method of exchangeable pairs. *Ann. Probab.*, **42**, no. 3, 906-945.

[11] Merlevède F., Peligrad M. and Rio E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. *High dimensional probability V: the Luminy volume, 273292, Inst. Math. Stat. Collect., 5, Inst. Math. Statist., Beachwood, OH*

[12] Merlevède F., Peligrad M. and Rio E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Related Fields* **151**, no. 3-4, 435-474.

[13] Mokkadem, A. (1990). Propriété de mélange des processus autorégressifs polynomiaux. *Ann. Instit Henri Poincaré.* **26**, 133-141.

[14] Paulin, D., Mackey, L. and Tropp, J. A. (2014). Efron-Stein Inequalities for Random Matrices. *arXiv:1408.3470.*

[15] Skorohod, A. V. (1976). On a representation of random variables. *Theory Probab. Appl.*, **21**, 628-632.

[16] Tao, T. (2012). *Topics in random matrix theory.* Graduate Studies in Mathematics, 132. American Mathematical Society, Providence, RI, 282 pp.

[17] Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, **12**, no. 4, 389-434.