

# Small-Deviation Inequalities for Sums of Random Matrices

Xianjie Gao, Chao Zhang \* and Hongwei Zhang

School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China; xianjegao@foxmail.com (X.G.); hwzhang@dlut.edu.cn (H.Z.)

\* Correspondence: chao.zhang@dlut.edu.cn

Received: 16 April 2019; Accepted: 4 May 2019; Published: 6 May 2019

**Abstract:** Random matrices have played an important role in many fields including machine learning, quantum information theory, and optimization. One of the main research focuses is on the deviation inequalities for eigenvalues of random matrices. Although there are intensive studies on the large-deviation inequalities for random matrices, only a few works discuss the small-deviation behavior of random matrices. In this paper, we present the small-deviation inequalities for the largest eigenvalues of sums of random matrices. Since the resulting inequalities are independent of the matrix dimension, they are applicable to high-dimensional and even the infinite-dimensional cases.

**Keywords:** large-deviation inequality; small-deviation inequality; random matrix; largest singular value

## 1. Introduction

Random matrices have been widely used in many problems, e.g., compressed sensing [1], high-dimensional data analysis [2], matrix approximation [3,4] and dimension reduction [5]. In the literature, one of main research issues is to study the deviation behavior of eigenvalues (or singular values) of random matrices.

In general, there are two types of deviation results studied in probability theory: one is the large-deviation inequality that describes the behavior of the probability  $\mathbb{P}(|x| > t)$  for large  $t$ ; and the other is the small-deviation (or small-ball) inequality that controls the probability  $\mathbb{P}(|x| < \epsilon)$  for small  $\epsilon$ .

The early large-deviation inequalities for sums of random matrices can be dated back to the work of Ahlswede and Winter [6]. Tropp [7] improved their results and developed a user-friendly framework to obtain the large-deviation inequalities for sums of random matrices. To overcome the limitation of the matrix-dimension dependence, Hsu et al. [8] and Minsker [9] introduced the concepts of intrinsic dimension and effective dimension to tighten large-deviation inequalities, respectively. Moreover, Zhang et al. [10] applied a diagonalization method to obtain the dimension-free large-deviation random for largest singular value of sums of random matrices, while it remains a challenge to select the auxiliary matrices and functions. In the scenario of a single random matrix, Ledoux [11] studied the largest eigenvalues of Gaussian unitary ensemble matrices and Vershynin [12] studied the singular values of the sub-Gaussian and sub-exponential matrices.

Small-deviation problems stemmed from some practical applications, e.g., approximation problems [13], Brownian pursuit problems [14], quantization problem [15], and convex geometry [16]. For more details, we refer to the bibliography maintained by Lifshitz [17]. There have been some works on the small-deviation inequalities for the specific types of random matrices. Aubrun [18] obtained the small-deviation inequalities for the largest eigenvalue of a single Gaussian unitary ensemble matrix. Rudelson and Vershynin [19] presented the small-deviation inequalities for the smallest singular value of the random matrix with independent entries. Volodko [20] estimated the

small-deviation probability of the determinant of the matrix  $\mathbf{B}\mathbf{B}^T$ , where  $\mathbf{B}$  is a  $d \times \infty$  random matrix whose entries obey a centered joint Gaussian distribution. To the best of our knowledge, there are few works on the small-deviation inequalities for sums of random matrices.

### 1.1. Related Works

Let  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\} \subset \mathbb{C}^{d \times d}$  be a finite sequence of independent random Hermitian matrices. It follows from Markov's inequality that

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{X}_k \right) \geq t \right\} \leq \inf_{\theta > 0} \left\{ e^{-\theta t} \cdot \text{tr} \left( \mathbb{E} e^{\theta \sum_k \mathbf{X}_k} \right) \right\},$$

where  $\lambda_{\max}$  denotes the largest eigenvalue and  $\mathbb{E}$  stands for the expectation operation. By using Golden-Thompson inequality, Ahlswede and Winter [6] bounded the trace of the matrix moment generating function (mgf) in the following way:

$$\text{tr}(\mathbb{E} e^{\theta \sum_k \mathbf{X}_k}) \leq \text{tr}(\mathbf{I}) \cdot \left[ \prod_k \lambda_{\max}(\mathbb{E} e^{\theta \mathbf{X}_k}) \right] = d \cdot \exp \left( \sum_k \lambda_{\max}(\log \mathbb{E} e^{\theta \mathbf{X}_k}) \right), \quad (1)$$

where  $\text{tr}(\mathbf{A})$  stands for the trace of the matrix  $\mathbf{A}$ . By applying Lieb's concavity theorem, Tropp [7] achieved a tighter matrix mgf bound than the above one:

$$\text{tr}(\mathbb{E} e^{\theta \sum_k \mathbf{X}_k}) \leq d \cdot \exp \left( \lambda_{\max} \left( \sum_k \log \mathbb{E} e^{\theta \mathbf{X}_k} \right) \right), \quad (2)$$

where "the eigenvalue of sum of matrices" is smaller than "the sum of eigenvalues of matrices" in the right-hand side of (1). However, there remains a shortcoming that the result (2) is dependent with the matrix dimension  $d$ , and its right-hand side will become loose for high-dimensional matrices.

To overcome the shortcoming, Hsu et al. [8] employed the intrinsic dimension  $\frac{\text{tr}(\mathbf{X})}{\lambda_{\max}(\mathbf{X})}$  to replace the ambient dimension  $d$  in the case of real symmetric matrices. Minsker [9] provided a dimension-free version of Bernstein's inequality for sequences of independent random matrices. Zhang et al. [10] introduced a diagonalization method to obtain the tail bounds for the largest singular values of the sum of random matrices. Although their bounds are independent of the matrix dimension and overcame the aforementioned first shortcoming, there remains a challenge to select the appropriate parameters to obtain the tighter bounds.

There are also some small-deviation results on one single random matrix. Edelman [21] presented the small-deviation behavior of the smallest singular value of a Gaussian matrix:

$$\lim_{d \rightarrow \infty} \mathbb{P} \left\{ s_{\min}(\mathbf{A}) \leq \frac{\epsilon}{\sqrt{d}} \right\} = 1 - \exp \left( -\epsilon - \frac{\epsilon^2}{2} \right),$$

where  $\mathbf{A}$  is a  $d \times d$  random matrix whose entries are independent standard normal random variables. Rudelson and Vershynin [22] studied the small-deviation bound of the smallest singular value of a sub-Gaussian matrix:

$$\mathbb{P} \left\{ s_{\min}(\mathbf{B}) \leq \frac{\epsilon}{\sqrt{d}} \right\} \leq C \cdot \epsilon + c^d,$$

where  $C > 0$ ,  $c \in (0, 1)$  only depends on the sub-Gaussian moment of its entries and  $\mathbf{B}$  is a  $d \times d$  random matrix whose entries are i.i.d. sub-Gaussian random variables with zero mean and unit variance. However, to the best of our knowledge, there is little work on the small-deviation inequalities for sums of random matrices.

## 1.2. Overview of Main Results

In this paper, we present the small-deviation inequalities for the largest eigenvalue for sums of independent random Hermitian matrices, i.e., the upper bound of

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{X}_k \right) \leq \epsilon \right\}.$$

In particular, we first present some basic small-deviation results of random matrices. We then obtain several types of small-deviation inequalities for the largest eigenvalue of sums of independent random positive semi-definite (PSD) matrices. In contrast to the large-deviation inequalities for random matrices, the resulting small-deviation inequalities are independent of the matrix dimension  $d$  and thus our findings are applicable to the high-dimensional and even infinite-dimensional cases.

The rest of this paper is organized as follows. In Section 2, we introduce some useful notations and then give some basic results on small-deviation inequalities for random matrices. The small-deviation results for sums of random PSD matrices are presented in Section 3. The last section concludes the paper.

## 2. Basic Small-Deviation Inequalities for Random Matrices

In this section, we first introduce the necessary notations and then present some basic small-deviation results of random matrices.

### 2.1. Necessary Notations

Given a Hermitian matrix  $\mathbf{A}$ , denote  $\lambda_{\max}(\mathbf{A})$  and  $\lambda_{\min}(\mathbf{A})$  as the largest and the smallest eigenvalues of  $\mathbf{A}$ , respectively. Denote  $\text{tr}(\mathbf{A})$  and  $\|\mathbf{A}\|$  as the trace and the spectral norm of  $\mathbf{A}$ , respectively. Let  $\mathbf{I}$  be the identity matrix,  $\mathbf{U}$  be the unitary matrix and  $\mathbf{U}^*$  stand for the Hermitian adjoint of  $\mathbf{U}$ .

By the spectral mapping theorem, given a real-value function  $f: \mathbb{R} \rightarrow \mathbb{R}$ , then

$$f(\mathbf{A}) = \mathbf{U} \cdot f(\Lambda) \cdot \mathbf{U}^*,$$

where  $\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^*$  is a diagonalization of  $\mathbf{A}$ . If  $f(a) \leq g(a)$  for  $a \in I$  when the eigenvalues of  $\mathbf{A}$  lie in  $I$ , then there holds that  $f(\mathbf{A}) \preceq g(\mathbf{A})$ , where the semi-definite partial order  $\preceq$  is defined as follows:

$$\mathbf{A} \preceq \mathbf{H} \Leftrightarrow \mathbf{H} - \mathbf{A} \text{ is positive semi-definite.}$$

### 2.2. Basic Small-Deviation Inequalities for Random Matrices

Subsequently, we produce the small-deviation inequalities for random matrices. First, we consider a small-deviation bound for one single matrix:

**Lemma 1.** Let  $\mathbf{Y}$  be a random Hermitian matrix. Then for any  $\epsilon > 0$ ,

$$\mathbb{P} \{ \lambda_{\max}(\mathbf{Y}) \leq \epsilon \} \leq \inf_{\theta > 0} \left\{ \frac{1}{d} \cdot e^{\theta\epsilon} \cdot \mathbb{E} \text{tr} (e^{-\theta\mathbf{Y}}) \right\}.$$

**Proof.** For any  $\theta > 0$ , we have

$$\begin{aligned}\mathbb{P}\{\lambda_{\max}(\mathbf{Y}) \leq \epsilon\} &= \mathbb{P}\{e^{-\lambda_{\max}(\theta\mathbf{Y})} \geq e^{-\theta\epsilon}\} \\ &\leq \mathbb{E} e^{-\lambda_{\max}(\theta\mathbf{Y})} \cdot e^{\theta\epsilon} \quad [\text{by Markov's inequality}] \\ &= \mathbb{E} e^{\lambda_{\min}(-\theta\mathbf{Y})} \cdot e^{\theta\epsilon} \quad [\text{since } -\lambda_{\max}(\mathbf{A}) = \lambda_{\min}(-\mathbf{A})] \\ &= \mathbb{E} \lambda_{\min}(e^{-\theta\mathbf{Y}}) \cdot e^{\theta\epsilon} \quad [\text{by Spectral mapping theorem}] \\ &\leq \frac{1}{d} \cdot e^{\theta\epsilon} \cdot \mathbb{E} \operatorname{tr}(e^{-\theta\mathbf{Y}}).\end{aligned}$$

The last inequality holds because the minimum eigenvalue of a positive definite (pd) matrix is dominated by the  $\operatorname{tr}(\cdot)/d$ . Since this inequality holds for any  $\theta > 0$ , taking an infimum over  $\theta > 0$  completes the proof.  $\square$

Then, by using the subadditivity of the matrix cumulant generating function (see [7], Lemma 3.4), we obtain the small-deviation bound for sums of random matrices:

**Theorem 1.** Let  $\{X_1, X_2, \dots, X_K\}$  be a finite sequence of independent random Hermitian matrices. Then for any  $\epsilon > 0$ ,

$$\mathbb{P}\left\{\lambda_{\max}\left(\sum_k X_k\right) \leq \epsilon\right\} \leq \inf_{\theta > 0} \left\{e^{\theta\epsilon} \cdot \exp\left(\lambda_{\max}\left(\sum_k \log \mathbb{E} e^{-\theta X_k}\right)\right)\right\}. \quad (3)$$

**Proof.** By combining Lemma 1 and Lemma 3.4 of [7], we have for any  $\theta > 0$ ,

$$\begin{aligned}\mathbb{P}\left\{\lambda_{\max}\left(\sum_k X_k\right) \leq \epsilon\right\} &\leq \frac{1}{d} \cdot e^{\theta\epsilon} \cdot \mathbb{E} \operatorname{tr}(e^{-\theta \sum_k X_k}) \\ &\leq \frac{1}{d} \cdot e^{\theta\epsilon} \cdot \operatorname{tr}\left(\exp\left(\sum_k \log \mathbb{E} e^{-\theta X_k}\right)\right) \\ &\leq \frac{1}{d} \cdot e^{\theta\epsilon} \cdot d \cdot \lambda_{\max}\left(\exp\left(\sum_k \log \mathbb{E} e^{-\theta X_k}\right)\right) \\ &= e^{\theta\epsilon} \cdot \exp\left(\lambda_{\max}\left(\sum_k \log \mathbb{E} e^{-\theta X_k}\right)\right).\end{aligned}$$

Taking the infimum over  $\theta > 0$  completes the proof.  $\square$

Please note that the above small-deviation bound is independent of the matrix dimension  $d$ , and thus it is applicable to the scenarios of high-dimensional and even infinite-dimensional matrices. In addition, we also derive the following small-deviation bounds for sums of random matrices.

**Corollary 1.** Let  $\{X_1, X_2, \dots, X_K\}$  be a sequence of independent random Hermitian matrices. Assume that there is a function  $g(\theta)$  and a sequence  $\{A_k\}$  of fixed Hermitian matrices such that

$$\mathbb{E} e^{-\theta X_k} \preceq e^{g(\theta) \cdot A_k}, \quad \forall \theta > 0. \quad (4)$$

1. Define the scalar parameter

$$\eta_1 := \lambda_{\max}\left(\sum_k A_k\right).$$

If  $g(\theta) > 0$ , then for any  $\epsilon > 0$ ,

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{X}_k \right) \leq \epsilon \right\} \leq \inf_{\theta > 0} \left\{ \exp \left( \theta \epsilon + g(\theta) \cdot \eta_1 \right) \right\}. \quad (5)$$

2. Define the scalar parameter

$$\eta_2 := \lambda_{\min} \left( \sum_k \mathbf{A}_k \right).$$

If  $g(\theta) < 0$ , then for any  $\epsilon > 0$ ,

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{X}_k \right) \leq \epsilon \right\} \leq \inf_{\theta > 0} \left\{ \exp \left( \theta \epsilon + g(\theta) \cdot \eta_2 \right) \right\}. \quad (6)$$

**Proof.** It follows from (4) that

$$\log \mathbb{E} e^{-\theta \mathbf{X}_k} \preceq g(\theta) \cdot \mathbf{A}_k,$$

and substituting it into Theorem 1 leads to the result (5). Then, the fact that  $\lambda_{\max}(-\mathbf{X}) = -\lambda_{\min}(\mathbf{X})$  for any Hermitian matrix  $\mathbf{X}$  leads to the result (6). This completes the proof.  $\square$

By using the logarithm operation, we then obtain another small-deviation bound for sums of random matrices:

**Corollary 2.** Let  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}$  be a sequence of independent random Hermitian matrices. Then for any  $\epsilon > 0$ ,

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{X}_k \right) \leq \epsilon \right\} \leq \inf_{\theta > 0} \exp \left( \theta \epsilon + K \cdot \log \lambda_{\max} \left( \frac{1}{K} \sum_{k=1}^K \mathbb{E} e^{-\theta \mathbf{X}_k} \right) \right).$$

**Proof.** Since the matrix logarithm is operator concave, for each  $\theta > 0$ , we have

$$\sum_{i=1}^K \log \mathbb{E} e^{-\theta \mathbf{X}_k} = K \cdot \frac{1}{K} \sum_{i=1}^K \log \mathbb{E} e^{-\theta \mathbf{X}_k} \preceq K \cdot \log \left( \frac{1}{K} \sum_{i=1}^K \mathbb{E} e^{-\theta \mathbf{X}_k} \right).$$

According to (3), we then arrive at

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{X}_k \right) \leq \epsilon \right\} \leq \frac{1}{d} \cdot e^{\theta \epsilon} \cdot \text{tr} \exp \left( K \cdot \log \left( \frac{1}{K} \sum_{i=1}^K \mathbb{E} e^{-\theta \mathbf{X}_k} \right) \right).$$

Since the trace of a matrix can be bounded by  $d$  times of its maximum eigenvalue, taking the infimum over  $\theta > 0$  completes the proof.  $\square$

The following presents the relationship between one random PSD matrix and a sum of PSD random matrices.

**Lemma 2.** Let  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}$  be a sequence of independent random Hermitian PSD matrices. Then for all  $k$  and any  $\epsilon > 0$ ,

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{X}_k \right) \leq \epsilon \right\} \leq \prod_k \mathbb{P} \left\{ \lambda_{\max}(\mathbf{X}_k) \leq \epsilon \right\} \leq \mathbb{P} \left\{ \lambda_{\max}(\mathbf{X}_k) \leq \epsilon \right\}.$$

**Proof.** Since  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}$  are PSD, we have

$$\begin{aligned}\mathbb{P}\left\{\lambda_{\max}\left(\sum_k \mathbf{X}_k\right) \leq \epsilon\right\} &\leq \mathbb{P}\left\{\max\left(\lambda_{\max}(\mathbf{X}_1), \dots, \lambda_{\max}(\mathbf{X}_K)\right) \leq \epsilon\right\} \\ &= \prod_k \mathbb{P}\left\{\lambda_{\max}(\mathbf{X}_k) \leq \epsilon\right\} \\ &\leq \mathbb{P}\left\{\lambda_{\max}(\mathbf{X}_k) \leq \epsilon\right\}.\end{aligned}$$

The last inequality holds for any  $k = 1, 2, \dots, K$ . This completes the proof.  $\square$

This lemma shows that the small-deviation probability for sums of random matrices can be bounded by using the small-deviation probability for one single matrix. This fact suggests that the small-deviation bound could be independent of the size of matrix sequence, while this phenomenon will not arise in the large-deviation scenario.

### 3. Small-Deviation Inequalities for Positive Semi-Definite Random Matrices

In this section, we present several types of small-deviation inequalities for the largest eigenvalue of sums of independent random PSD matrices. Similar to the scalar version of small-deviation inequalities, there remains a challenge to bound the term  $\mathbb{E}e^{-\theta \mathbf{X}_k}$ . Here, we adapt some methods to handle this issue.

First, we introduce the negative moment estimate for the largest eigenvalue to derive a small-deviation inequality for sums of random matrices:

**Theorem 2.** Let  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}$  be a sequence of independent random Hermitian PSD matrices. Given a  $p > 0$ , if there exists a positive constant  $C_p$  such that

$$\left[\lambda_{\max}\left(\sum_k \mathbb{E}\mathbf{X}_k\right)\right]^{-p} < C_p,$$

then there holds that for any  $\epsilon > 0$ ,

$$\mathbb{P}\left\{\lambda_{\max}\left(\sum_k \mathbf{X}_k\right) \leq \epsilon\right\} \leq C_p \epsilon^p.$$

**Proof.** It follows from Jensen's inequality that

$$\mathbb{E}\left(\lambda_{\max}\left(\sum_k \mathbf{X}_k\right)\right)^{-p} \leq \left(\mathbb{E}\lambda_{\max}\left(\sum_k \mathbf{X}_k\right)\right)^{-p} \leq \left[\lambda_{\max}\left(\sum_k \mathbb{E}\mathbf{X}_k\right)\right]^{-p} < C_p.$$

Then, the Markov's inequality yields

$$\begin{aligned}\mathbb{P}\left\{\lambda_{\max}\left(\sum_k \mathbf{X}_k\right) \leq \epsilon\right\} &= \mathbb{P}\left\{\left[\lambda_{\max}\left(\sum_k \mathbf{X}_k\right)\right]^{-p} \geq \epsilon^{-p}\right\} \\ &\leq \epsilon^p \cdot \mathbb{E}\left(\lambda_{\max}\left(\sum_k \mathbf{X}_k\right)\right)^{-p} \leq C_p \epsilon^p.\end{aligned}$$

This completes the proof.  $\square$

In this theorem, we impose an assumption that the negative moment of  $\lambda_{\max}(\sum_k \mathbf{X}_k)$  is bounded. In general, this assumption is mild, and can be satisfied in most cases. The following small-deviation results are derived under that condition that the eigenvalues of the matrices  $\{\mathbf{X}_k\}$  are bounded:

**Theorem 3.** Let  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}$  be a sequence of independent random Hermitian PSD matrices such that  $\lambda_{\max}(\mathbf{X}_k) \leq L$  ( $\forall k = 1, 2, \dots, K$ ) almost surely. Then for any  $\epsilon > 0$ ,

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{X}_k \right) \leq \epsilon \right\} \leq \left( \frac{\mu}{\epsilon} \right)^{\epsilon/L} \cdot \exp \left( \frac{\epsilon - \mu}{L} \right), \quad (7)$$

where

$$\mu := \lambda_{\min} \left( \sum_k \mathbb{E} \mathbf{X}_k \right).$$

Furthermore, there holds that for any  $\epsilon > 0$ ,

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{X}_k \right) \leq \epsilon \right\} \leq \left( \frac{1}{\epsilon} \right)^{K\epsilon/L} \cdot \left( \prod_{k=1}^K \mu_k \right)^{\epsilon/L} \cdot \exp \left( \frac{K\epsilon - \sum_k \mu_k}{L} \right), \quad (8)$$

where

$$\mu_k = \lambda_{\min}(\mathbb{E} \mathbf{X}_k).$$

**Proof.** For any  $\theta > 0$  and  $x \in [0, L]$ , there holds that

$$e^{-\theta x} \leq 1 + \frac{e^{-\theta L} - 1}{L} \cdot x \leq \exp \left( \frac{e^{-\theta L} - 1}{L} \cdot x \right).$$

According to transfer rule, we have,

$$\log \mathbb{E} e^{-\theta \mathbf{X}_k} \preceq \frac{e^{-\theta L} - 1}{L} \mathbb{E} \mathbf{X}_k. \quad (9)$$

By substituting (9) into Corollary 1, we then have,

$$\begin{aligned} \mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{X}_k \right) \leq \epsilon \right\} &\leq \inf_{\theta > 0} e^{\theta \epsilon} \cdot \exp \left( \lambda_{\max} \left( \sum_k \frac{e^{-\theta L} - 1}{L} \mathbb{E} \mathbf{X}_k \right) \right) \\ &= \inf_{\theta > 0} \exp \left( \theta \epsilon + \lambda_{\max} \left( \frac{e^{-\theta L} - 1}{L} \sum_k \mathbb{E} \mathbf{X}_k \right) \right) \\ &= \inf_{\theta > 0} \exp \left( \theta \epsilon + \frac{e^{-\theta L} - 1}{L} \cdot \lambda_{\min} \left( \sum_k \mathbb{E} \mathbf{X}_k \right) \right) \\ &= \inf_{\theta > 0} \exp \left( \theta \epsilon + \frac{e^{-\theta L} - 1}{L} \cdot \mu \right), \end{aligned}$$

since  $\frac{e^{-\theta L} - 1}{L} < 0$ . The infimum is achieved at  $\theta = \frac{1}{L} \log \left( \frac{\mu}{\epsilon} \right)$ , which leads to the result of (7).

Moreover, the combination of Lemma 1 and (9) leads to

$$\mathbb{P} \{ \lambda_{\max}(\mathbf{X}_k) \leq \epsilon \} \leq \left( \frac{\mu_k}{\epsilon} \right)^{\epsilon/L} \cdot \exp \left( \frac{\epsilon - \mu_k}{L} \right).$$

Then, the result (8) is derived from Lemma 2. This completes the proof.  $\square$

Actually, the above results are derived from the geometric point of view, where the term  $e^{-\theta x}$  is bounded by the linear function  $1 + \frac{e^{-\theta L} - 1}{L} x$  for any  $x \in [0, L]$ . Finally, we study the small-deviation

inequalities for random matrix series  $\sum_k x_k \mathbf{A}_k$ , which is a sum of fixed Hermitian pd matrices  $\mathbf{A}_k$  weighted by random variables  $x_k$ .

**Theorem 4.** Let  $\{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_K\}$  be a sequence of fixed Hermitian pd matrices, and  $\{x_1, x_2, \dots, x_K\}$  be a finite sequence of independent variables. If there exist the constants  $C > 0$  and  $\alpha > 0$  such that for all  $\theta > 0$ ,

$$\mathbb{E} e^{-\theta x_k} \leq C \cdot \theta^{-\alpha}, \quad (10)$$

then there holds that for any  $0 < \epsilon < \frac{K\alpha}{e} \cdot \sqrt[\alpha]{\frac{K}{C\nu}}$ ,

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k x_k \mathbf{A}_k \right) \leq \epsilon \right\} \leq \left( \frac{e\epsilon}{K\alpha} \right)^{\alpha K} \cdot \left( \frac{C\nu}{K} \right)^K, \quad (11)$$

where

$$\nu = \lambda_{\max} \left( \sum_k \mathbf{A}_k^{-\alpha} \right).$$

Furthermore, for any  $0 < \epsilon < (\alpha/e) \cdot C^{-\frac{1}{\alpha}} \cdot \left( \prod_{k=1}^K \nu_k \right)^{-\frac{1}{\alpha K}}$ ,

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k x_k \mathbf{A}_k \right) \leq \epsilon \right\} \leq \left( \prod_{k=1}^K \nu_k \right) \cdot C^K \cdot \left( \frac{e\epsilon}{\alpha} \right)^{K\alpha}, \quad (12)$$

where

$$\nu_k = \lambda_{\max}(\mathbf{A}_k^{-\alpha}).$$

**Proof.** According to transfer rule, we have,

$$\mathbb{E} e^{-\theta x_k \mathbf{A}_k} \leq C \cdot (\theta \mathbf{A}_k)^{-\alpha}. \quad (13)$$

By substituting (13) into Corollary 2, we then have,

$$\begin{aligned} \mathbb{P} \left\{ \lambda_{\max} \left( \sum_k x_k \mathbf{A}_k \right) \leq \epsilon \right\} &\leq \inf_{\theta > 0} \exp \left( \theta \epsilon + K \cdot \log \lambda_{\max} \left( \frac{1}{K} \sum_{k=1}^K C \cdot (\theta \mathbf{A}_k)^{-\alpha} \right) \right) \\ &= \inf_{\theta > 0} \exp \left( \theta \epsilon + K \cdot \log \lambda_{\max} \left( \frac{C\theta^{-\alpha}}{K} \sum_{k=1}^K \mathbf{A}_k^{-\alpha} \right) \right) \\ &= \inf_{\theta > 0} \exp \left( \theta \epsilon + K \cdot \log \frac{C\nu}{K\theta^\alpha} \right). \end{aligned}$$

The infimum will be attained at  $\theta = \frac{\alpha K}{\epsilon}$ , and it leads to the result of (11). Moreover, the combination of Lemma 1 and (13) leads to

$$\mathbb{P} \{ \lambda_{\max}(x_k \mathbf{A}_k) \leq \epsilon \} \leq C \cdot \nu_k \cdot \left( \frac{e\epsilon}{\alpha} \right)^\alpha.$$

Then, the result (12) is resulted from Lemma 2. This completes the proof.  $\square$

The above results hold under the condition (10) that  $\mathbb{E} e^{-\theta x_k}$  has a power-type upper bound of  $C \cdot \theta^{-\alpha}$ . This condition is mild, and we refer to [23] for the details. Moreover, to keep the results (11) and (12) non-trivial, their right-hand sides should be less than one, and thus we arrive at  $\epsilon < \frac{K\alpha}{e} \cdot \sqrt[\alpha]{\frac{K}{C\nu}}$  and  $\epsilon < (\alpha/e) \cdot C^{-\frac{1}{\alpha}} \cdot \left( \prod_{k=1}^K \nu_k \right)^{-\frac{1}{\alpha K}}$ , respectively.



#### 4. Conclusions

In this paper, we present the small-deviation inequalities for the largest eigenvalues of sums of random matrices. In particular, we first give some basic results on small-deviation inequalities for random matrices. We then study the small-deviation inequalities for sums of independent random PSD matrices. In contrast to the large-deviation inequalities for random matrices, our results are independent of the matrix dimension  $d$  and thus can be applicable to the scenarios of high-dimensional and even infinite-dimensional matrices. In addition, by using the Hermitian dilation (see Section 2.6 of [7]), our small-deviation results can also be extended to the scenario of non-Hermitian random matrices.

**Author Contributions:** Conceptualization, X.G. and C.Z.; methodology, X.G. and C.Z.; validation, X.G., C.Z. and H.Z.; resources, C.Z.; writing—original draft preparation, X.G.; writing—review and editing, C.Z.; supervision, C.Z. and H.Z.; project administration, C.Z. and H.Z.; funding acquisition, C.Z.

**Funding:** This research was funded by National Natural Science Foundation of China grant number 61473328 and 11401076.

**Acknowledgments:** We are grateful to the anonymous reviewers and the editors for their valuable comments and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

- Chandrasekaran, V.; Recht, B.; Parrilo, P.A.; Willsky, A.S. The convex geometry of linear inverse problems. *Found. Comput. Math.* **2012**, *12*, 805–849.
- Bühlmann P.; van de Geer, S. *Statistics for High-Dimensional Data: Methods, Theory and Applications*; Springer: Berlin, Germany, 2011.
- Gittens, A.; Mahoney, M.W. Revisiting the nyström method for improved large-scale machine learning. *J. Mach. Learn. Res.* **2016**, *17*, 3977–4041.
- Halko, N.; Martinsson, P.G.; Tropp, J.A. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.* **2011**, *53*, 217–288.
- Clarkson, K.L.; Woodruff, D.P. Low rank approximation and regression in input sparsity time. In Proceedings of the forty-fifth annual ACM symposium on Theory of computing, Palo Alto, CA, USA, 1–4 June 2013; pp. 81–90.
- Ahlswede, R.; Winter, A. Strong converse for identification via quantum channels. *IEEE Trans. Inf. Theory.* **2002**, *48*, 569–579.
- Tropp, J.A. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **2012**, *12*, 389–434.
- Hsu, D.; Kakade, S.M.; Zhang, T. Tail inequalities for sums of random matrices that depend on the intrinsic dimension. *Electron. Commun. Probab.* **2012**, *17*, 1–13.
- Minsker, S. On some extensions of Bernstein’s inequality for self-adjoint operators. *Stat. Probabil. Lett.* **2017**, *127*, 111–119.
- Zhang, C.; Du, L.; Tao, D. Lsv-based tail inequalities for sums of random matrices. *Neural. Comput.* **2017**, *29*, 247–262.
- Ledoux, M. Deviation inequalities on largest eigenvalues. In *Geometric Aspects of Functional Analysis*; Milman, V.D., Schechtman, G., Eds.; Springer: Berlin, Germany, 2007; pp. 167–219.
- Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. *arXiv* **2010**, arXiv:1011.3027.
- Li, W.V.; Linde, W. Approximation, metric entropy and small ball estimates for gaussian measures. *Ann. Probab.* **1999**, *27*, 1556–1578.
- Li, W.V.; Shao, Q.M. Capture time of brownian pursuits. *Probab. Theory. Rel.* **2001**, *121*, 30–48.
- Dereich, S.; Fehrer, F.; Matoussi, A.; Scheutzw, M. On the link between small ball probabilities and the quantization problem for gaussian measures on banach spaces. *J. Theor. Probab.* **2003**, *16*, 249–265.
- Klartag, B.; Vershynin, R. Small ball probability and dvoretzky’s theorem. *ISR J. Math.* **2007**, *157*, 193–207.
- Lifshits, M. Bibliography of Small Deviation Probabilities. Available online: <https://www.lpsm.paris/pageperso/smalldev/biblio.pdf> (accessed on 15 December 2016).

18. Aubrun, G. A sharp small deviation inequality for the largest eigenvalue of a random matrix. In *Séminaire de Probabilités XXXVIII*; Springer: Berlin, Germany, 2005; pp. 320–337.
19. Rudelson, M.; Vershynin, R. Non-asymptotic theory of random matrices: extreme singular values. *arXiv* **2010**, arXiv:1003.2990.
20. Volodko, N.V. Small deviations of the determinants of random matrices with gaussian entries. *Stat Probabil. Lett.* **2014**, *84*, 48–53.
21. Edelman, A. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. A* **1988**, *9*, 543–560.
22. Rudelson, M.; Vershynin, R. The littlewood–offord problem and invertibility of random matrices. *Adv. Math.* **2008**, *218*, 600–633.
23. Li, W.V. Small Value Probabilities in Analysis and Mathematical Physics. Available online: [https://www.math.arizona.edu/~mathphys/school\\_2012//WenboLi.pdf](https://www.math.arizona.edu/~mathphys/school_2012//WenboLi.pdf) (accessed on 15 March 2012).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).