Journal of Mathematical Research with Applications
Sept., 2020, Vol. 40, No. 5, pp. 543–550
DOI:10.3770/j.issn:2095-2651.2020.05.010
Http://jmre.dlut.edu.cn

A Refined Non-Asymptotic Tail Bound of Sub-Gaussian Matrix

Xianjie GAO, Chao ZHANG*, Hongwei ZHANG
School of Mathematical Sciences, Dalian University of Technology, Liaoning 116024*, P. R. China*

Abstract In this paper, we obtain a refined non-asymptotic tail bound for the largest singular value (the soft edge) of sub-Gaussian matrix. As an application, we use the obtained theorem to compute the tail bound of the Gaussian Toeplitz matrix.

Keywords non-asymptotic theory; largest singular value; tail bound; sub-Gaussian matrix

MR(2010) Subject Classification 60B20; 60F10; 46B09

1. Introduction

Random matrix theory (RMT) has been widely applied in many fields, e.g., multivariate statistics [1], high-dimensional data analysis [2], the matrix approximation [3], the combinatorial optimization [4] and the compressed sensing [5]. One main research concern on RMT is to study the tail behavior of the extreme eigenvalues (or singular values) of random matrices.

In general, there are two types of probabilistic statements on the study of probability theory: asymptotic and non-asymptotic. The former aims to analyze the limit behavior of some probability terms, e.g., the central limit theorem

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \xrightarrow{d} g, \quad n\to\infty \tag{1.1}$$

for Bernoulli random variables $x_1,x_2,\ldots,x_n,\ldots$, where g is a Gaussian random variable. There have been many well-known asymptotic results on RMT:

Wigner's semicircle law [6]: Let $\mathbf{A}_n$ be an $n\times n$ symmetric matrix whose entries are independent Gaussian variables. As dimension $n\to\infty$, the spectrum of the Wigner matrices $\mathbf{W}_n=n^{-1/2}\mathbf{A}_n$ is distributed according to the semicircle law with density:

$$f_{sc}(x)=\frac{1}{2\pi}\sqrt{4-x^2}, \quad x\in[-2,2]. \tag{1.2}$$

Marchenko-Pastur law [7]: Let $\mathbf{A}_{m,n}$ $(m\geq n)$ be an $m\times n$ random Gaussian matrix. As the dimensions $m,n\to\infty$ while the aspect ratio n/m converges to a fix number $y\in(0,1]$, the spectrum of the matrices $\frac{1}{m}\mathbf{A}^*\mathbf{A}$ is distributed according to the Marchenko-Pastur law with

Received June 25, 2019; Accepted May 2, 2020
Supported by the Fundamental Research Funds for the Central Universities (Grant No. DUT20LK38).
* Corresponding author
E-mail address: xianjiegao@foxmail.com (Xianjie GAO); chao.zhang@dlut.edu.cn (Chao ZHANG); hwzhang@dlut.edu.cn (Hongwei ZHANG)

density:

$$f_{mp}(x) = \begin{cases} \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}, & a \le x \le b; \\ 0, & \text{otherwise}, \end{cases} \tag{1.3}$$

where $a = (1-\sqrt{y})^2$, $b = (1+\sqrt{y})^2$ and $\mathbf{A}^*$ stands for the Hermitian adjoint of $\mathbf{A}$.

Bai-Yin's law [8]: Let $\mathbf{A}_{m,n}$ $(m \ge n)$ be an $m \times n$ random matrix whose entries are independent copies of a random variable with zero mean, unit variance, and finite fourth moment. As the dimensions $m, n \to \infty$ with n/m converging to a fix number $y \in (0,1]$, the $s_{\min}(\mathbf{A})$ and $s_{\max}(\mathbf{A})$ are subjected to Bai-Yin's law:

$$\begin{aligned} &s_{\min}(\mathbf{A}) = \sqrt{\mathrm{m}} - \sqrt{\mathrm{n}} + \mathrm{o}(\sqrt{\mathrm{n}}), \\ &s_{\max}(\mathbf{A}) = \sqrt{\mathrm{m}} + \sqrt{\mathrm{n}} + \mathrm{o}(\sqrt{\mathrm{n}}), \ \ \text{almost surely}, \end{aligned} \tag{1.4}$$

where the $s_{\min}(\mathbf{A})$ and $s_{\max}(\mathbf{A})$ represent the smallest and largest singular values of $\mathbf{A}$, respectively.

Although these asymptotic statements can provide a precise limit result when the matrix dimension or sample number goes to the infinity, they cannot describe in what rate these probability terms converge to their limits. To handle this issue, there arises the non-asymptotic viewpoint to study these probability terms. For example, one of the non-asymptotic statement of the central limit theorem is Hoeffding's inequality:

$$\mathbb{P}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i > t\Big) \le 2\mathrm{e}^{-t^2/2}. \tag{1.5}$$

There have been many research works on RMT from the non-asymptotic viewpoint. Vershynin [9] gave non-asymptotic methods about the properties of sub-Gaussian and sub-exponential matrix. Tropp [10] proposed a user-friendly framework to study the tail behavior of sums of random matrices. Moreover, there are also other methods for developing the matrix concentration inequalities, e.g., exchangeable pairs [11] and Markov chain couplings [12]. To eliminate the dimension dependence of these tail results for random matrices, the intrinsic dimension (or effective dimension) was employed to improve them [13,14]. Recently, Zhang et al. [15] applied a diagonalization method to obtain the dimension-free tail inequalities of largest singular value for sums of random matrices.

In this paper, we obtain a refined non-asymptotic tail bound for the largest singular value (the soft edge) of sub-Gaussian matrix. We first give a tail bound for the norm of a sub-Gaussian matrix by transforming a sub-Gaussian matrix into a sub-Gaussian variable. We also obtain a tail bound for the norm of a sub-Gaussian matrix by decomposing a sub-Gaussian matrix into a series of sub-Gaussian matrices. By combining the two resulted tail bounds, we obtain the final tail results. As an application, we use the resulted tail inequalities to study the tail behavior of Gaussian Toeplitz matrix.

The rest of this paper is organized as follows. In the next section, we give some preliminary knowledge on random matrices and sub-Gaussian distributions. In Section 3, we present the main results. Section 4 presents the application of our results in the study of Gaussian Toeplitz matrix, and the last section concludes paper.

2. Notations and preliminaries

In this section, we give some preliminary knowledge on random matrices and sub-Gaussian distributions.

A random matrix is a matrix whose entries are random variables. Its distribution is characterized by the joint distribution of the entries. The expected value of an $m \times n$ random matrix $\mathbf{B}$ is the $m \times n$ matrix $\mathbb{E}(\mathbf{B})$ whose entries are the expected values of the corresponding entries of $\mathbf{B}$, assuming that they all exist.

Let $\mathbf{B}_{m\times n}$ be a random matrix. Let $S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ denote the Euclidean sphere in $\mathbb{R}^n$. The largest singular value of $\mathbf{B}$ is by definition

$$s_{\max}(\mathbf{B}) = \|\mathbf{B}\| = \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|\mathbf{B}x\|_2}{\|x\|_2} = \sup_{x\in S^{n-1}} \|\mathbf{B}x\|_2. \tag{2.1}$$

Given an arbitrary matrix $\mathbf{B}$, the Hermitian dilation of $\mathbf{B}$ is defined by

$$\mathcal{H}(B) = \begin{bmatrix} 0 & \mathbf{B} \\ \mathbf{B}^* & 0 \end{bmatrix}. \tag{2.2}$$

It is true that $\lambda_{\max}(\mathcal{H}(\mathbf{B})) = \|\mathcal{H}(\mathbf{B})\| = \|\mathbf{B}\|$, where $\lambda_{\max}$ denotes the largest eigenvalue. Given a Hermitian matrix $\mathbf{H}_{d\times d}$ and a real-value function $f : \mathbb{R} \to \mathbb{R}$, then

$$f(\mathbf{H}) = \mathbf{U} \cdot f(\Lambda) \cdot \mathbf{U}^* = \mathbf{U} \begin{bmatrix} f(\lambda_1) & & & \\ & (\lambda_2) & & \\ & & \ddots & \\ & & & (\lambda_d) \end{bmatrix} \mathbf{U}^*,$$

where $\mathbf{H} = \mathbf{U}\Lambda\mathbf{U}^*$ is a diagonalization of $\mathbf{H}$, $\mathbf{U}$ is the unitary matrix and $\mathbf{U}^*$ stands for the Hermitian adjoint of $\mathbf{U}$. In particular, we define the map on diagonal matrix by applying the function to each diagonal entry. The relationship for real function f is the transfer rule. If $f(a) \le g(a)$ for $a \in I$, then $f(\mathbf{H}) \preceq g(\mathbf{H})$ for the eigenvalues of $\mathbf{H}$ lie in I, where the semi-definite partial order $\preceq$ is defined as follows:

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

Sub-Gaussian distributions are referring to a large class of probability distributions, e.g., normal random variables, Bernoulli and all bounded random variables.

Definition 2.1 *A real-valued random variable x is said to be sub-Gaussian if there exists $c > 0$ such that for every $t > 0$*

$$\mathbb{P}(|x| > t) \le 2\mathrm{e}^{-ct^2}. \tag{2.3}$$

Assuming the sub-Gaussian random variable's mean is zero, the following lemma presents equivalent conditions.

Lemma 2.2 *Let x be a mean zero (centered) random variable. The following statements are* *equivalent: (1) x is sub-Gaussian; and (2) $\exists b > 0$, $\forall \theta \in \mathbb{R}$, there holds that*

$$\mathbb{E}\mathrm{e}^{\theta x} \le \mathrm{e}^{b^2\theta^2/2}. \tag{2.4}$$

The Lemma and its proof can be referred to [16, Theorem 3.1].

There are more and more research interests lying in the sub-Gaussian distributions, including spectral properties of random matrices [17] and tail inequalities of sub-Gaussian random vectors [18].

3. Main results

In this section, we obtain a refined upper bound for the largest singular value (the norm) of sub-Gaussian matrix. We first give an upper bound for the norm of sub-Gaussian matrix by converting into a random sub-Gaussian variable.

Theorem 3.1 *Let $\mathbf{B}$ be an $m \times n$ random sub-Gaussian matrix. That is, its entries x_{ij} are i.i.d. centered random variables, and obey the sub-Gaussian distribution. Then there exists $c > 0$ such that for all $t \ge 0$,*

$$\mathbb{P}\{\|\mathbf{B}\| > t\} \le 2 \cdot 5^{(m+n)} \cdot \exp(-ct^2), \tag{3.1}$$

where c does not depend on m, n and t. The proof of Theorem 3.1 is similar to [9, Proposition 2.4], where $m = n$. Here we give the proof of the general case.

Proof The main idea of the proof of Theorem 3.1 is to convert the random matrix into a random variable, *i.e.*, $\langle \mathbf{B}x, y\rangle$ is a sub-Gaussian random variable. We then use the covering number to complete the proof.

$$\begin{aligned}\mathbb{P}(\|\mathbf{B}\| > t) \le& \mathbb{P}(\max_{\substack{x\in\mathcal{N}\\ y\in\mathcal{M}}}\langle \mathbf{B}x, y\rangle > \frac{t}{4}) \le \sum_{\substack{x\in\mathcal{N}\\ y\in\mathcal{M}}} \mathbb{P}(\langle \mathbf{B}x, y\rangle > \frac{t}{4})\\ \le& |\mathcal{N}||\mathcal{M}| \cdot \mathbb{P}(\langle \mathbf{B}x, y\rangle > \frac{t}{4}) \le 2 \cdot 5^{(m+n)} \cdot \exp(-ct^2),\end{aligned}$$

where $\mathcal{N}$, $\mathcal{M}$ are $\frac{1}{2}$-nets of S^{n-1}, S^{m-1} respectively, and the bounds on cardinality of the net are $|\mathcal{N}| \le (1 + 2/\epsilon)^n$ and $|\mathcal{M}| \le (1 + 2/\epsilon)^m$. That is to choose $\epsilon = \frac{1}{2}$, and the conclusion is proved. □

A minor shortcoming of above result is that when the matrix dimension increases, the result becomes very loose. Another method is to obtain the tail bound for matrix sub-Gaussian series. We first introduce the matrix sub-Gaussian moment generating function (mgf) bound.

Proposition 3.2 *Assume that $\mathbf{H}$ is a fixed Hermitian matrix and the random variable x obeys the centered sub-Gaussian distribution. Then there exists $b > 0$ such that*

$$\mathbb{E}\,\mathrm{e}^{x\theta\mathbf{H}} \preceq \mathrm{e}^{\theta^2 b^2 \mathbf{H}^2/2}, \tag{3.2}$$

where b does not depend on t or the dimension of $\mathbf{H}$[1]. According to the transfer rule, it is easy to get the proposition. Based on the mgf result (3.2), we develop a tail bound for the matrix

[1] The b appearing below is independent of the dimensions of the matrix and t.

sub-Gaussian series.

Theorem 3.3 *Consider a finite sequence $\{\mathbf{H}_k : k = 1, \dots, K\}$ of fixed Hermitian matrices*[2] *with dimension d, and $\{x_k : k = 1, \dots, K\}$ be a finite sequence of independent centered sub-Gaussian random variables. Compute the variance parameter*

$$\rho := \|\sum_k \mathbf{H}_k^2\|.$$

Then, for all $t \geq 0$,

$$\mathbb{P}\Big\{\lambda_{\max}\Big(\sum_k x_k \mathbf{H}_k\Big) \geq t\Big\} \leq d \cdot \exp(-\frac{t^2}{2b^2\rho}). \tag{3.3}$$

Proof It follows from Proposition 3.2 that, for any $\theta > 0$,

$$\begin{aligned}\mathbb{P}\Big\{\lambda_{\max}\Big(\sum_k x_k \mathbf{H}_k\Big) \geq t\Big\} &\leq \mathrm{e}^{-\theta t} \cdot \operatorname{tr}\exp\Big(\sum_k \log \mathrm{E}\mathrm{e}^{\theta x_k \mathbf{H}_k}\Big) \\ &\leq \mathrm{e}^{-\theta t} \cdot \operatorname{tr}\exp\Big(\frac{\theta^2 b^2}{2}\sum_k \mathbf{H}_k^2\Big) \leq \mathrm{e}^{-\theta t} \cdot d \cdot \lambda_{\max}\Big(\exp\Big(\frac{\theta^2 b^2}{2}\sum_k \mathbf{H}_k^2\Big)\Big) \\ &= d \cdot \exp\Big(-\theta t + \frac{\theta^2 b^2}{2}\lambda_{\max}\big(\sum_k \mathbf{H}_k^2\big)\Big) = d \cdot \exp(-\theta t + \frac{\theta^2 b^2}{2}\rho),\end{aligned}$$

where $\rho := \|\sum_k \mathbf{H}_k^2\|$, the first inequality follows from [10, Theorem 3.6]. This inequality holds for any positive θ, so we may take an infimum to complete the proof. The infimum is attained when $\theta = \frac{t}{b^2\rho}$. □

We apply above result to study the sum of rectangular matrix series by using matrices Hermitian dilation. The following is the general version of Theorem 3.3.

Corollary 3.4 *Consider a finite sequence $\{\mathbf{D}_k : k = 1, \dots, K\}$ of fixed matrices with dimension $m \times n$, and $\{x_k : k = 1, \dots, K\}$ be a finite sequence of independent centered sub-Gaussian random variables. Compute the variance parameter*

$$\rho := \max\Big\{\Big\|\sum_k \mathbf{D}_k\mathbf{D}_k^*\Big\| \Big\|\sum_k \mathbf{D}_k^*\mathbf{D}_k\Big\|\Big\}. \tag{3.4}$$

Then, for all $t \geq 0$,

$$\mathbb{P}\Big\{\Big\|\sum_k x_k \mathbf{D}_k\Big\| \geq t\Big\} \leq (m+n) \cdot \exp(-\frac{t^2}{2b^2\rho}). \tag{3.5}$$

Proof According to Hermitian dilation we know that

$$\Big\|\sum_k x_k \mathbf{D}_k\Big\| = \lambda_{\max}\Big(\mathcal{H}\Big(\sum_k x_k \mathbf{D}_k\Big)\Big) = \lambda_{\max}\Big(\sum_k x_k \mathcal{H}(\mathbf{D}_k)\Big).$$

We invoke Theorem 3.3 to obtain the tail bound for the sum of rectangular matrix series. The matrix variance parameter ρ satisfies the relation:

$$\rho = \Big\|\sum_k \mathcal{H}(\mathbf{D}_k)^2\Big\| = \begin{Vmatrix}\sum_k \mathbf{D}_k\mathbf{D}_k^* & 0 \\ 0 & \sum_k \mathbf{D}_k^*\mathbf{D}_k\end{Vmatrix} = \max\Big\{\Big\|\sum_k \mathbf{D}_k\mathbf{D}_k^*\Big\| \Big\|\sum_k \mathbf{D}_k^*\mathbf{D}_k\Big\|\Big\}$$

[2] The elements of fixed Hermitian matrices are not random variables, they are fixed. These matrices are also Hermitian.

This completes the proof. □

Based on the general version of tail bound for matrix sub-Gaussian series, we obtain another tail bound for the norm of the sub-Gaussian matrix.

Theorem 3.5 *Under the notations and conditions in Theorem 3.1. Then there holds that for all $t \geq 0$,*

$$\mathbb{P}\{\|\mathbf{B}\| > t\} \leq (m+n) \cdot \exp\Big(-\frac{t^2}{2b^2 m}\Big). \tag{3.6}$$

Proof In order to use Corollary 3.4, we decompose matrix as a matrix sub-Gaussian series:

$$\mathbf{B} = \sum_{ij} x_{ij} \mathbf{E}_{ij}, \quad i = 1, \ldots, m, \ j = 1, \ldots, n.$$

The matrix $\mathbf{E}_{ij}$ has an element one in the (i, j) position and zeros elsewhere. By calculating $\rho = m$, the conclusion is established by using Corollary 3.4. □

The combination of Theorems 3.1 and 3.5 leads to the following refined upper bound for the largest singular value (the soft edge) of sub-Gaussian matrix.

Theorem 3.6 *Follow the notations and conditions in Theorem 3.1. Then there holds that for all $t \geq 0$,*

$$\mathbb{P}\{\|\mathbf{B}\| > t\} \leq \begin{cases} (m+n) \cdot \exp(-\frac{t^2}{2b^2 m}), & 0 < t \leq \sqrt{\frac{2b^2 m}{1-2b^2 mc} \log \frac{m+n}{2 \cdot 5^{m+n}}}; \\ 2 \cdot 5^{(m+n)} \cdot \exp(-ct^2), & t > \sqrt{\frac{2b^2 m}{1-2b^2 mc} \log \frac{m+n}{2 \cdot 5^{m+n}}}. \end{cases} \tag{3.7}$$

Remark 3.7 It can be known from [19, Definition 1.2 and Lemma 1.3], $b^2 c = \frac{1}{2}$. Because of $m > 1$, it can be guaranteed to be meaningful in the root. In other words, there must be b and c which make the formula hold.

4. Application: Gaussian Toeplitz matrix

In this section, we use our theoretical findings to compute the tail bound of the Gaussian Toeplitz matrix. The Gaussian Toeplitz matrix is an example of Gaussian random matrix which has been widely used in various fields, e.g., differential equations, spline functions, and signal processing [20]. We consider an unsymmetric Gaussian Toeplitz matrix $\mathbf{T} \in \mathbb{C}^{d \times d}$ in the following form:

$$\mathbf{T} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{d-1} \\ \gamma_{-1} & \gamma_0 & \gamma_1 & \cdots & \gamma_{d-2} \\ \gamma_{-2} & \gamma_{-1} & \gamma_0 & \cdots & \gamma_{d-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{-(d-1)} & \gamma_{-(d-2)} & \gamma_{-(d-3)} & \cdots & \gamma_0 \end{bmatrix}, \tag{4.1}$$

where $\gamma_{-(d-1)}, \ldots, \gamma_{d-1}$ are independent standard normal variables. The Gaussian Toeplitz matrix $\mathbf{T}$ can be represented as a matrix Gaussian series:

$$\mathbf{T} = \gamma_0 \mathbf{I} + \sum_{j=1}^{d-1} \gamma_j \mathbf{C}^j + \sum_{j=1}^{d-1} \gamma_{-j} (\mathbf{C}^j)^T, \tag{4.2}$$

where $\mathbf{C}^j$ is the j-th power of $\mathbf{C}$ with

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix}.$$

Using the Theorem 3.6, we can compute the tail bound of the Gaussian Toeplitz matrix. First, we calculate

$$(\mathbf{C}^j)(\mathbf{C}^j)^T = \sum_{k=1}^{d-j} \mathbf{E}_{kk} \text{ and } (\mathbf{C}^j)^T(\mathbf{C}^j) = \sum_{k=j+1}^{d} \mathbf{E}_{kk}.$$

The matrix variance parameter $\rho = d$ can be calculated:

$$\mathbf{I}^2 + \sum_{j=1}^{d-1}(\mathbf{C}^j)(\mathbf{C}^j)^T + \sum_{j=1}^{d-1}(\mathbf{C}^j)^T(\mathbf{C}^j) = d\mathbf{I}_d.$$

For Gaussian matrix, $b = 1$, $c = \frac{1}{2}$. Through the application of Theorem 3.6, tail bound of the Gaussian Toeplitz matrix is presented, for all $t \geq 0$,

$$\mathbb{P}\{\|\mathbf{T}\| > t\} \leq \begin{cases} 2d \cdot \exp(-\frac{t^2}{2d}), & 0 < t \leq \sqrt{\frac{2d}{1-2d} \log \frac{2d}{2\cdot 5^d}}; \\ 2 \cdot 5^d \cdot \exp(-\frac{t^2}{2}), & t > \sqrt{\frac{2d}{1-2d} \log \frac{2d}{2\cdot 5^d}}. \end{cases} \tag{4.3}$$

5. Conclusion

In this paper, we first present the tail bounds for the largest singular value of sub-Gaussian matrix and matrix sub-Gaussian series. We then obtain a refined non-asymptotic tail bound for the largest singular value (the soft edge) of sub-Gaussian matrix. As an application, we finally compute the tail bound of Gaussian Toeplitz matrix.

References

[1] R. MUIRHEAD. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Inc., New York, 1982.

[2] P. BÜHLMANN, S. VAN DE GEER. *Statistics for High-dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg, 2011.

[3] N. HALKO, P. G. MARTINSSON, J. A. TROPP. *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*. SIAM Rev., 2011, **53**(2): 217–288.

[4] A. NAOR, O. REGEV, T. VIDICK. *Efficient rounding for the noncommutative grothendieck inequality*. Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing, ACM, 2013, 71–80.

[5] V. CHANDRASEKARAN, B. RECHT, P. A. PARRILO, et al. *The convex geometry of linear inverse problems*. Found. Comput. Math., 2012, **12**(6): 805–849.

[6] E. P. WIGNER. *On the distribution of the roots of certain symmetric matrices*. Ann. Math., 1958, 325–327.

[7] V. A. MARCHENKO, L. A. PASTUR. *Distribution of eigenvalues for some sets of random matrices*. Mat. Sb., 1967, **114**(4): 507–536.

[8] Zhidong BAI, Yongquan YIN. *Limit of the smallest eigenvalue of a large dimensional sample covariance matrix*. Ann. Probab., 1993, **21**(3): 1275–1294.

[9] R. VERSHYNIN. *Introduction to the non-asymptotic analysis of random matrices*. arXiv preprint arXiv:1011.3027, 2010.

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

[11] L. MACKEY, M. I. JORDAN, R. Y. CHEN, et al. *Matrix concentration inequalities via the method of exchangeable pairs*. Ann. Probab., 2014, **42**(3): 906–945.

[12] D. PAULIN, L. MACKEY, J. A. TROPP. *Deriving matrix concentration inequalities from kernel couplings*. arXiv preprint arXiv:1305.0612, 2013.

[13] D. HSU, S. M. KAKADE, Tong ZHANG. *Tail inequalities for sums of random matrices that depend on the intrinsic dimension*. Electron. Commun. Probab., 2012, **17**(14): 1–13.

[14] S. MINSKER. *On some extensions of Bernstein's inequality for self-adjoint operators*. Stat. Probab. Lett., 2017, **127**: 111–119.

[15] Chao ZHANG, Lei DU, Dacheng TAO. *Lsv-based tail inequalities for sums of random matrices*. Neural Comput., 2017, **29**(1): 247–262.

[16] O. RIVASPLATA. *Subgaussian random variables: An expository note*. Internet publication, PDF, 2012.

[17] A. E. LITVAK, A. PAJOR, M. RUDELSON, et al. *Smallest singular value of random matrices and geometry of random polytopes*. Adv. Math., 2005, **195**(2): 491–523.

[18] D. HSU, S. KAKADE, Tong ZHANG, et al. *A tail inequality for quadratic forms of subgaussian random vectors*. Electron. Commun. Probab., 2012, **17**(52): 1–6.

[19] P. RIGOLLET. *Sub-Gaussian Random Variables*. Available online: `https://ocw.mit.edu/courses/mathematics/18-s997-high-dimensional-statistics-spring-2015/lecture-notes/MIT18\_S997S15\_Chapter1.pdf`.

[20] R. M. GRAY. *Toeplitz and circulant matrices: A review*. Found. Trends Commun. Inf., 2006, **2**(3): 155–239.