

# SPECTRAL ANALYSIS OF GRAM MATRICES WITH MISSING AT RANDOM OBSERVATIONS: CONVERGENCE, CENTRAL LIMIT THEOREMS, AND APPLICATIONS IN STATISTICAL INFERENCE

BY HUIQIN LI<sup>1,a</sup>, GUANGMING PAN<sup>2,c</sup>, YANQING YIN<sup>1,b</sup> AND WANG ZHOU<sup>3,d</sup>

<sup>1</sup>*School of Mathematics and Statistics, Chongqing University, <sup>a</sup>lihq118@nenu.edu.cn, <sup>b</sup>yinyq@cqu.edu.cn*

<sup>2</sup>*Division of Mathematical Sciences, Nanyang Technological University, <sup>c</sup>gmpan@ntu.edu.sg*

<sup>3</sup>*Department of Statistics and Data Science, National University of Singapore, <sup>d</sup>wangzhou@nus.edu.sg*

Motivated by the statistical inference using the Gram matrix in the context of missing at random observations, this paper investigates the spectral properties of the random matrices  $\mathbf{S}_n = \frac{1}{n} \mathbf{Z} \mathbf{Z}^*$ , where  $\mathbf{Z} = \mathbf{D} \circ (\Sigma^{1/2} \mathbf{X})$  represents a Hadamard random matrix with entries determined by independent Bernoulli variables  $\mathbf{D}$ . Operating within the high-dimensional framework, we establish the convergence of the empirical spectral distribution of  $\mathbf{S}_n$  to a well-defined limiting distribution. In addition, we explore the impact of the missing mechanism on the second-order properties of the spectral distribution of the Gram matrix  $\mathbf{S}_n$ . We establish the central limit theorem for the linear spectral statistics of  $\mathbf{S}_n$ , shedding light on their fluctuations. Surprisingly, our analysis reveals that even in the ideal Gaussian distribution scenario, the fluctuations of statistics generated by eigenvalues are influenced by the eigenvectors of the population covariance matrix in the missing-at-random case. This discovery uncovers a remarkable phenomenon that starkly contrasts with the classical case. Subsequently, we demonstrate the practical application of our central limit theorem in hypothesis testing for the population covariance matrix.

**1. Introduction and motivation.** The analysis of high-dimensional data has gained significant traction across various fields, including genomics, finance, image processing, and social networks. See, for instance, Chen and Qin [6], Cai, Liu and Xia [4], Wang, Peng and Li [19], Fan, Guo and Zheng [9]. However, real-world data collection processes are often imperfect, resulting in missing observations that can profoundly affect the reliability and validity of statistical analysis.

Missing data poses a pervasive problem in numerous domains. For example, in healthcare, patient data frequently contains missing values due to incomplete medical records or patients dropping out of studies. Financial datasets may exhibit missing data due to irregular trading patterns or incomplete financial records. Similarly, in the social sciences, nonresponse in surveys or incomplete data collection processes can lead to missing data. These instances highlight the practical relevance of addressing the challenges posed by missing data and emphasize the necessity for robust methodologies to handle such scenarios. A pragmatic approach to addressing missing data is the exclusion of variables with incomplete observations from the analysis, thereby restricting the analysis to a subset of fully observed variables. However, in the context of gene expression data, where a substantial proportion, approximately 90%, of genes exhibit missing values, this strategy would yield an insufficient number of variables to conduct a valid statistical analysis. Disregarding variables with only a few missing observations would also lead to a squandering of valuable information that could otherwise

---

Received September 2023; revised April 2024.

*MSC2020 subject classifications.* Primary 62H15, 62B20; secondary 62D10.

*Key words and phrases.* Sample covariance matrix, missing observations, high-dimensionality, limiting spectral distribution, central limit theorem, random matrix theory.

be utilized effectively. Considering these challenges, alternative methodologies need to be explored to handle missing data in the fields such as gene expression analysis effectively, ensuring the utilization of available information while maintaining statistical integrity. Significant progress has been made in this field, with numerous efforts dedicated to advancing the research in this direction. Noteworthy contributions can be found in the works of Lounici [13], Chen et al. [7], Cai and Zhang [5] and references therein.

In the realm of high-dimensional data analysis, Random matrix theory (RMT) assumes a pivotal role by offering indispensable tools and theoretical frameworks to comprehend intricate matrices encountered across diverse applications. RMT has demonstrated its applicability in a broad spectrum of fields, including wireless communications, signal processing, finance, genomics, and machine learning. By exploring the spectral properties of random matrices that arise in various high-dimensional statistical scenarios, RMT equips researchers with powerful statistical tools to make inference about the underlying data structure and extract meaningful information. This becomes particularly valuable when dealing with large dimensions and when the matrix entries exhibit specific randomness or correlation structures. For further exploration of this research direction, we direct readers to the works of Bai and Silverstein [3], Hu et al. [10], and other relevant sources.

In this study, our primary objective is to establish a profound connection between random matrix theory (RMT) and the analysis of incomplete data. We achieve this by investigating the spectral properties of high-dimensional Gram matrices in the presence of missing observations. Our goal is to derive theoretical results that thoroughly characterize the behavior of the spectrum of these matrices under missing-at-random mechanisms. By harnessing the tools and insights offered by RMT, we can gain a deeper understanding of how missing probabilities impact both the first and second-order limiting properties of the Gram matrix.

Our work represents a significant expansion upon the existing literature, effectively addressing a crucial research gap in this field. While previous studies, exemplified by the work of Jurczak and Rohde [11], have made valuable contributions by establishing first-order results for the spectrum of a covariance-type matrix under missing observations, our research takes a notable stride forward. We delve into the application of random matrix theory in the context of high-dimensional statistics, seeking to unlock its full potential and provide a comprehensive understanding of the effects of missing data. This advancement is pivotal in enhancing the accuracy and reliability of statistical inference, particularly in complex and high-dimensional settings. The contributions of this paper can be summarized as follows:

1. We offer a comprehensive examination of the spectral properties of high-dimensional Gram matrices when confronted with missing observations, thereby extending the current knowledge derived from random matrix theory. Due to the absence of a prior bound on the spectral norm, we employ a combination of  $\epsilon$ -net arguments and the analytic method of Stieltjes transform to achieve the separation of the spectrum.
2. We derive explicit expressions for the eigenvalue distribution and the central limit theorem (CLT) for the limiting spectral statistics (LSS) under missing at random mechanisms. These expressions provide a deeper understanding of how missingness influences the spectral behavior of the Gram matrices.
3. We discuss the test procedure designed specifically for testing the equality of the population covariance matrix with a given matrix. This procedure offers an approach to assessing covariance structures in the presence of missing data.

In this paper, we adopt the following notation conventions:

- The spectral norm of a matrix  $\mathbf{A}$  is denoted as  $\|\mathbf{A}\|_S$  and the determinant of a matrix  $\mathbf{A}$  is denoted as  $\det \mathbf{A}$ .
- The Euclidean norm of a vector  $\boldsymbol{\alpha}$  is represented as  $\|\boldsymbol{\alpha}\|$ .

- The symbols  $c$  and  $C$  denote absolute constants, which may vary in different instances.
- The notation  $\mathbf{e}_k$  refers to the  $k$ th column of an identity matrix.
- The transpose of a vector  $\boldsymbol{\alpha}$  is denoted as  $\boldsymbol{\alpha}^T$ .
- If for any  $\ell > 0$ ,  $1 - \mathbb{P}(A_n) = o(n^{-\ell})$  as  $n \rightarrow \infty$ , then we say  $A_n$  occurs with high probability (w.h.p).

Please note that these notation conventions have been employed consistently throughout the paper.

The paper is structured as follows. Section 2 presents theoretical results on the spectral properties of the Gram matrix under missing at random. In Section 3, we examine the test for covariance structure in the presence of missing data. Technical proofs are deferred to Section 4, with supplementary details provided in the supplementary material [12]. Section 5 contains a list of necessary lemmas to be utilized.

**2. Main theoretical results.** This section presents our main theorem. First, we introduce the Gram random matrix model that we investigate for handling missing at random data. Additionally, we provide background information on random matrix theory. Subsequently, we present definitions, model assumptions, and lemmas that play critical roles in establishing our theorem. Finally, we present our main theorems individually.

**2.1. Model description.** We begin by introducing some primary knowledge in random matrix theory. Suppose  $\mathbf{A}_n$  is an  $n \times n$  Hermitian matrix. Then the *empirical spectral distribution* (ESD) of  $\mathbf{A}_n$  is defined by

$$F^{\mathbf{A}_n}(x) = \frac{1}{n} \sum_{j=1}^n I(\lambda_j \leq x).$$

In the asymptotic scenario where  $n$  tends to infinity, if the limit of  $F^{\mathbf{A}_n}(x)$  exists, it is denoted as the *Limit Spectral Distribution* (LSD). This represents the distribution that emerges as a result of this convergence in the context of spectral analysis. The *Stieltjes transform* of  $F^{\mathbf{A}}(x)$  is given by

$$m_{F^{\mathbf{A}}}(z) = \int \frac{1}{x - z} dF^{\mathbf{A}}(x) = \frac{1}{n} \text{tr}(\mathbf{A} - z\mathbf{I}_n)^{-1}, \quad z \in \mathbb{C}^+.$$

Consider the random vector  $\mathbf{y} = \boldsymbol{\Sigma}_n^{1/2} \mathbf{x}$ , where  $\mathbf{x}$  represents a  $p$ -dimensional random vector whose entries are independent and identically distributed (i.i.d.) random variables with zero mean and unit variance. In this context,  $\mathbf{y}$  typically serves as a model for a  $p$ -dimensional population with covariance matrix  $\boldsymbol{\Sigma}_n$ . The well-known M-P law, originally presented by Marčenko and Pastur [14] and extensively studied in subsequent works such as Wachter [18] and Silverstein [16], establishes that when the ratio  $p/n$  tends to  $\rho \in (0, \infty)$ , the distribution  $F^{\boldsymbol{\Sigma}_n}$  converges to  $\mathfrak{H}$ , and the population covariance matrix is bounded in the spectral norm, the ESD of the sample covariance matrix converges weakly to a nonrandom probability density function.

To incorporate missing at random mechanisms, we introduce the random vector  $\mathbf{d} = (d_1, \dots, d_p)^T$ , where each  $d_j$  is an independent Bernoulli random variable  $B(1, p_j)$ , for  $j = 1, \dots, p$ . Consequently, we consider the Hadamard product of the random vector  $\mathbf{z} = \mathbf{d} \circ \mathbf{y}$  as the population with missing probability. This population is referred to as the *incomplete population*, distinguishing it from the *complete population*  $\mathbf{y}$ . To conduct statistical inference, we draw a sample of  $n$  observations from the incomplete population, denoted as  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$ .

Accordingly, letting  $\mathbf{d}_j = (d_{1j}, \dots, d_{pj})^T$  and  $\mathbb{D}_j = \text{diag}(d_{1j}, \dots, d_{pj})$ , we compute the corresponding Gram matrix as

$$\mathbf{S}_n = \frac{1}{n} \sum_{j=1}^n \mathbf{z}_j \mathbf{z}_j^T = \frac{1}{n} \mathbf{Z} \mathbf{Z}^T = \frac{1}{n} \sum_{j=1}^n \mathbb{D}_j \mathbf{y}_j \mathbf{y}_j^T \mathbb{D}_j = \frac{1}{n} (\mathbf{D}_n \circ \boldsymbol{\Sigma}_n^{1/2} \mathbf{X}_n) (\mathbf{D}_n \circ \boldsymbol{\Sigma}_n^{1/2} \mathbf{X}_n)^T,$$

where  $\mathbf{X}_n = (\mathbf{x}_1, \dots, \mathbf{x}_n) = (x_{jk})$  and  $\mathbf{D}_n = (d_{jk})$ . In the context of statistical applications, we introduce an auxiliary matrix  $\Theta_n$ . This auxiliary matrix is arbitrary, as long as it satisfies the specific assumptions outlined in our main results. The form of this matrix may vary case by case in statistical applications, depending on the particular problem under consideration. An example illustrating this variability is provided in Corollary 2.1, where the corresponding result is applied to covariance matrix testing, offering insight into this aspect. Our main theorem addresses the LSD and CLT for LSS, with a particular emphasis on the product matrix  $\mathbf{S}_n \Theta_n$ . Note that one shall obtain the corresponding results of  $\mathbf{S}_n$  by setting  $\Theta_n = \mathbf{I}$ .

**2.2. Definitions and assumptions.** Prior to presenting our main theorems, we will provide definitions and assumptions. Define  $\mathbb{P} = \mathbb{E} \mathbb{D}_1 = \text{diag}(\mathfrak{p}_1, \dots, \mathfrak{p}_p)$  and

$$\begin{aligned} \mathbb{P}^{(2)} &= \mathbb{E}(\mathbb{D}_j - \mathbb{P})^2 = \mathbb{P} - \mathbb{P}^2, & \mathbb{P}^{(3)} &= \mathbb{E}(\mathbb{D}_j - \mathbb{P})^3 = \mathbb{P} - 3\mathbb{P}^2 + 2\mathbb{P}^3, \\ \mathbb{P}^{(4)} &= \mathbb{E}(\mathbb{D}_j - \mathbb{P})^4 = \mathbb{P} - 4\mathbb{P}^2 + 6\mathbb{P}^3 - 3\mathbb{P}^4, & \mathbb{P}^{(5)} &= (\mathbb{P}^{(4)} - 3(\mathbb{P}^{(2)})^2). \end{aligned}$$

Recall that a random variable  $X$  is considered sub-Gaussian if there exists a constant  $C$  such that:

$$\mathbb{E} \exp(X^2/C^2) \leq 2.$$

We define its sub-Gaussian norm as:

$$\|X\|_g = \inf\{c > 0 : \mathbb{E} \exp(X^2/c^2) \leq 2\}.$$

We now introduce the following assumptions.

**Assumption A:** The random variables  $x_{jk}$ ,  $j, k = 1, 2, \dots$  are independent and identically sub-Gaussian with zero mean, unit variance and sub-Gaussian norm  $\|x\|_g$ . Let  $\mathbb{E} |x_{11}|^4 = \nu_4$ .

**Assumption B:** The matrices  $\mathbf{X}_n$  and  $\mathbf{D}_n$  are independent, and  $\Theta_n$  is a nonrandom matrix.

**Assumption C:** The convergence regime is  $y_n = p/n \rightarrow y \in (0, +\infty)$ .

**Assumption D:** The spectral distribution  $H_n$  of the matrix  $\Psi_n = \Theta_n^{1/2} \mathbf{T}_n \Theta_n^{1/2}$  where  $\mathbf{T}_n = \mathbb{P} \boldsymbol{\Sigma}_n \mathbb{P} + \mathbb{P}^{(2)} \circ \boldsymbol{\Sigma}_n$  weakly converges to a probability distribution  $H$ , as  $p \rightarrow \infty$ , referred as the incomplete population spectral distribution (IPSD). Let the spectral norm of the sequence  $(\boldsymbol{\Sigma}_n \Theta_n)$  be uniformly bounded.

**REMARK 2.1.** We would like to discuss two points in order. First, the assumption of sub-Gaussianity is connected to Lemma 2.3, which plays a crucial role in our analysis. In the context of missing at random scenarios, establishing a high-probability bound for the spectral norm of the Gram matrix  $\mathbf{S}_n$  proves to be challenging under general moment conditions. This difficulty arises due to the limitations of applying the standard moment method. The sub-Gaussianity assumption is instrumental in overcoming this challenge by ensuring that the entries possess controlled tails. This controlled tail behavior is crucial for the application of concentration inequalities. With the sub-Gaussianity assumption, we can establish Lemma 2.3 using an  $\varepsilon$ -net argument. It's worth noting that the use of sub-Gaussianity is a common practice in high-dimensional statistics, serving to control tail probabilities and ensure the concentration of random matrices. Furthermore, we would like to emphasize that while the existence of a high-probability bound for  $\mathbf{S}_n$  under general distributions is not guaranteed,

our CLT results would seamlessly align with any advancements in this direction, provided such bounds can be established.

Second, the specific matrix  $\mathbf{T}_n$  introduced in Assumption D takes on the role of the population covariance matrix for incomplete data. This assignment is elucidated through the following calculations:

$$\begin{aligned} & \mathbb{E} \mathbb{D}_1 \mathbf{y}_1 \mathbf{y}_1^T \mathbb{D}_1 \\ &= \mathbb{E}((\mathbb{D}_1 - \mathbb{P}) \mathbf{y}_1 \mathbf{y}_1^T (\mathbb{D}_1 - \mathbb{P}) + (\mathbb{D}_1 - \mathbb{P}) \mathbf{y}_1 \mathbf{y}_1^T \mathbb{P} + \mathbb{P} \mathbf{y}_1 \mathbf{y}_1^T (\mathbb{D}_1 - \mathbb{P}) + \mathbb{P} \mathbf{y}_1 \mathbf{y}_1^T \mathbb{P}) \\ &= \mathbb{P}^{(2)} \circ \boldsymbol{\Sigma}_n + \mathbb{P} \boldsymbol{\Sigma}_n \mathbb{P}. \end{aligned}$$

Under the scenario of complete data where  $\mathbf{p}_j = 1$ ,  $1 \leq j \leq p$ , we have  $\mathbb{P} = \mathbf{I}_p$  and  $\mathbb{P}^{(2)} = \mathbf{0}$ , resulting in the degeneration of  $\mathbf{T}_n$  to  $\boldsymbol{\Sigma}_n$ .

**2.3. Lemmas on quadratic forms.** This subsection is dedicated to presenting lemmas concerning the quadratic forms of the incomplete population  $\mathbf{z}$ . These lemmas provide valuable insights into the effects of missing data thus play a central role in analyzing the spectral properties of the missing data model. For the  $p$ -dimensional random vector  $\mathbf{z} \in \mathbb{R}^p$ , let us recall the regular complex matrix field  $\mathbb{C}^{p \times p}$ , which consists of all  $p$ -dimensional nonrandom matrices. We define

$$\mathcal{Q}(\mathbf{z}, \mathcal{M}) = \mathbf{z}^T \mathcal{M} \mathbf{z}, \quad \mathcal{M} \in \mathbb{C}^{p \times p}.$$

Within this context, it is our primary objective to present the mean and variance-covariance profile of  $\mathcal{Q}(\mathbf{z}, \cdot)$ . These profiles play a fundamental role in various aspects of our analysis. First, they are instrumental in establishing the convergence of the ESD. Second, by leveraging the variance-covariance profiles, we can ascertain the limiting parameters that govern the second-order fluctuations of the Gram matrix.

**LEMMA 2.1** (First two moments of  $\mathcal{Q}(\mathbf{z}, \cdot)$ ). *Assuming  $\mathbf{z}$  follows our model, we consider any  $p \times p$  nonrandom matrix  $\mathbf{A}$  and a nonrandom symmetric matrix  $\mathbf{B}$ . In this context, we establish the following result:  $\mathbb{E} \mathcal{Q}(\mathbf{z}, \mathbf{A}) = \text{tr} \mathbf{A} \mathbf{T}_n$  and*

$$\mathbb{E}[(\mathcal{Q}(\mathbf{z}, \mathbf{A}) - \text{tr} \mathbf{A} \mathbf{T}_n)(\mathcal{Q}(\mathbf{z}, \mathbf{B}) - \text{tr} \mathbf{B} \mathbf{T}_n)] = \mathcal{I}(\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_n, \mathbb{P}) + (v_4 - 3) \mathcal{I} \mathcal{I}(\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_n, \mathbb{P}),$$

where the terms that are independent of the kurtosis of the underlying distribution are given by  $\mathcal{I}(\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_n, \mathbb{P}) = \mathcal{I}_1(\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_n, \mathbb{P}) + \mathcal{I}_2(\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_n, \mathbb{P})$  with  $\mathcal{I}_1(\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_n, \mathbb{P}) = 2 \text{tr}(\mathbf{T}_n \mathbf{A} \mathbf{T}_n \mathbf{B})$  and

$$\begin{aligned} & \mathcal{I}_2(\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_n, \mathbb{P}) \\ &= 4 \text{tr}[\mathbb{P}^{(2)}(\mathbf{A} \circ \mathbf{B}) \mathbb{P}^{(2)}(\boldsymbol{\Sigma}_n \circ \boldsymbol{\Sigma}_n)] + 2 \text{tr}[(\mathbf{A} \circ \mathbb{P}^{(2)}) \boldsymbol{\Sigma}_n (\mathbf{B} \circ \mathbb{P}^{(2)}) \boldsymbol{\Sigma}_n] \\ &+ 6 \text{tr}(\mathbb{P}^{(3)} \circ \mathbf{A} \circ \boldsymbol{\Sigma}_n \circ (\mathbf{B} \mathbb{P} \boldsymbol{\Sigma}_n)) + 2 \text{tr}((\mathbb{P}^{(2)} \circ \mathbf{A}) \boldsymbol{\Sigma}_n \mathbb{P} \mathbf{B} \mathbb{P} \boldsymbol{\Sigma}_n) \\ &+ 2 \text{tr}((\mathbb{P}^{(2)} \circ \mathbf{B}) \boldsymbol{\Sigma}_n \mathbb{P} \mathbf{A} \mathbb{P} \boldsymbol{\Sigma}_n) + 3 \text{tr}[\mathbb{P}^{(3)} \circ \mathbf{B} \circ \boldsymbol{\Sigma}_n \circ (\mathbf{A} \mathbb{P} \boldsymbol{\Sigma}_n + \mathbf{A}^T \mathbb{P} \boldsymbol{\Sigma}_n)] \\ &+ 3 \text{tr}[\mathbf{A} \circ \boldsymbol{\Sigma}_n \circ \mathbf{B} \circ \boldsymbol{\Sigma}_n \circ \mathbb{P}^{(5)}] + 4 \text{tr}[\mathbb{P}^{(2)} \circ (\mathbf{A} \mathbb{P} \boldsymbol{\Sigma}_n + \mathbf{A}^T \mathbb{P} \boldsymbol{\Sigma}_n) \circ (\mathbf{B} \mathbb{P} \boldsymbol{\Sigma}_n)], \end{aligned}$$

while the terms that are dependent on the kurtosis of the underlying distribution are expressed as

$$\begin{aligned} & \mathcal{I} \mathcal{I}(\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_n, \mathbb{P}) \\ &= \text{tr}[(\boldsymbol{\Sigma}_n^{1/2} (\mathbb{P} \mathbf{A} \mathbb{P} + \mathbb{P}^{(2)} \circ \mathbf{A}) \boldsymbol{\Sigma}_n^{1/2}) \circ (\boldsymbol{\Sigma}_n^{1/2} (\mathbb{P} \mathbf{B} \mathbb{P} + \mathbb{P}^{(2)} \circ \mathbf{B}) \boldsymbol{\Sigma}_n^{1/2})] \end{aligned}$$

$$\begin{aligned}
& + 2\text{tr}[\mathbb{P}^{(2)}(\mathbf{A} \circ \mathbf{B})\mathbb{P}^{(2)}(\boldsymbol{\Sigma}_n^{1/2} \circ \boldsymbol{\Sigma}_n^{1/2})^2] + \text{tr}[\mathbb{P}^{(5)} \circ \mathbf{A} \circ \mathbf{B} \circ (\boldsymbol{\Sigma}_n^{1/2} \circ \boldsymbol{\Sigma}_n^{1/2})^2] \\
& + 2\text{tr}[(\boldsymbol{\Sigma}_n^{1/2} \circ \boldsymbol{\Sigma}_n^{1/2} \circ \boldsymbol{\Sigma}_n^{1/2})(\mathbf{A} \circ \mathbb{P}^{(3)})\mathbf{B}\mathbb{P}\boldsymbol{\Sigma}_n^{1/2}] \\
& + \text{tr}[(\boldsymbol{\Sigma}_n^{1/2} \circ \boldsymbol{\Sigma}_n^{1/2} \circ \boldsymbol{\Sigma}_n^{1/2})(\mathbf{B} \circ \mathbb{P}^{(3)})(\mathbf{A} + \mathbf{A}^T)\mathbb{P}\boldsymbol{\Sigma}_n^{1/2}] \\
& + 2\text{tr}[(\boldsymbol{\Sigma}_n^{1/2} \circ \boldsymbol{\Sigma}_n^{1/2})\mathbb{P}^{(2)}((\mathbf{A}\mathbb{P}\boldsymbol{\Sigma}_n^{1/2} + \mathbf{A}^T\mathbb{P}\boldsymbol{\Sigma}_n^{1/2}) \circ (\mathbf{B}\mathbb{P}\boldsymbol{\Sigma}_n^{1/2}))].
\end{aligned}$$

Furthermore, we proceed to present a lemma regarding the higher-order moments of  $\mathcal{Q}(\mathbf{z}, \cdot)$ . This lemma also plays crucial role in our analysis, serving multiple purposes. First, it enables us to derive some necessary high probability results. Second, it facilitates the control of negligible terms in our analysis.

**LEMMA 2.2** (High-order moments of  $\mathcal{Q}(\mathbf{z}, \cdot)$ ). *Assume  $\mathbf{z}$  follows our model. Let  $\mathbf{A} = (a_{jk})$  be a  $p \times p$  nonrandom symmetric matrix and  $\mathbb{E}|x_j|^\ell \leq v_\ell$ . Then for  $\ell > 2$ ,*

$$\mathbb{E}|\mathcal{Q}(\mathbf{z}, \mathbf{A}) - \text{tr}(\mathbf{A}\mathbf{T}_n)|^\ell \leq C_\ell[(n\|\mathbf{A}\|^2)^{\ell/2} + v_{2\ell}n\|\mathbf{A}\|^\ell],$$

where  $C_\ell$  is a constant depending on  $\ell$  only.

The proofs of the aforementioned lemmas are deferred to the supplement material.

**2.4. Limiting spectral distribution.** Note that for a random matrix, both its ESD and the corresponding probability measure are inherently stochastic. To ensure clarity and comprehensiveness in our statement, we introduce the following definitions.

**DEFINITION 2.1** (Convergence with high probability). We say that a sequence of random variables  $x_n$  converges with high probability to  $x$ , denoted as  $x_n \xrightarrow{\text{w.h.p.}} x$ , if for any  $\epsilon > 0$  and  $\ell > 0$ ,

$$\mathbb{P}(|x_n - x| > \epsilon) = o(n^{-\ell}), \quad n \rightarrow \infty.$$

**DEFINITION 2.2** (Convergence with high probability of ESD). Consider a sequence of random matrices  $\mathbf{A}_n$  with empirical spectral distributions denoted by  $F^{\mathbf{A}_n}$ . Adopting the usual definition of Kolmogorov distance  $\|F - G\|_\infty = \sup_x |F(x) - G(x)|$  for two distribution functions  $F$  and  $G$ , we say that the sequence  $F^{\mathbf{A}_n}$  converges with high probability to  $F$ , denoted as  $F^{\mathbf{A}_n} \xrightarrow{\text{w.h.p.}} F$ , if

$$\|F^{\mathbf{A}_n} - F\|_\infty \xrightarrow{\text{w.h.p.}} 0.$$

The forthcoming theorem establishes the asymptotic distribution of  $F^{\mathbf{S}_n \Theta_n}(x)$ , providing insights into the behavior of the Gram matrix under missing at random mechanisms.

**THEOREM 2.1** (LSD). *Under Assumptions A-D, the ESD  $F^{\mathbf{S}_n \Theta_n}(x)$  converges to the limiting spectral distribution  $F$  with high probability. The Stieltjes transform  $m(z)$  associated with  $F$  is analytic in  $\mathbb{C}^+$ . For each  $z \in \mathbb{C}^+$ ,  $m = m(z)$  represents a unique solution to the equation:*

$$(2.1) \quad m = \int \frac{1}{t(1 - y - yzm) - z} dH(t),$$

where uniqueness is guaranteed within the set  $\{m \in \mathbb{C}^+ : -\frac{1-y}{z} + ym \in \mathbb{C}^+\}$ .

As observed, the limiting spectral distribution of the Gram matrix, under the presence of missing at random, still exhibits characteristics akin to the M-P law. However, a notable distinction arises in the distribution parameter  $H$ , which transitions from the complete population to the incomplete population due to the influence of missing at random. This phenomenon bears resemblance to the behavior observed in sample correlation matrices, see Karoui [8].

**REMARK 2.2.** We would like to discuss some points. First, we would like to emphasize that our Gram matrix model differs from the one in Jurczak and Rohde [11] in terms of normalization parameters, using  $n$  instead of  $N_{jk} = 1 \vee \#\{t \in \{1, \dots, n\} : d_{jt}d_{kt} = 1\}$  for the entries of the Gram matrix  $(\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n)(\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n)^T$ . This difference in normalization brings about technical challenges in studying the spectral properties of the Gram matrix, especially under nondiagonal covariance matrix scenarios, as shown in Jurczak and Rohde [11]. The dependence introduced by  $N_{jk}$  breaks the independent structure of the columns in  $(\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n)$ , creating challenges in applying standard techniques for deriving the second-order spectral properties of the Gram matrix. In contrast, our studied Gram model maintains both the independent structure between columns of the data matrix and the conventional covariance matrix form. This characteristic not only facilitates a more straightforward investigation of the second-order limit but also enhances the overall clarity of the theoretical procedure. Second, it is evident that setting the auxiliary matrix  $\Theta_n$  to the identity yields a distinct LSD from the one in Jurczak and Rohde [11]. More specifically, under the null case where  $\Sigma_n^{1/2}$  is the identity and assuming equal missing probability, our Gram matrix model remains nonnegative definite, whereas the matrix model in Jurczak and Rohde [11] could result in negative eigenvalues for a small enough missing probability. Importantly, our model ensures the applicability of theoretical results to statistical applications, as shown in Section 3. Finally, the establishment of convergence with high probability beyond almost surely meets the requirement in the proof of Theorem 2.2, which, in turn, fulfills the requirement in the proof of Theorem 2.3.

Let  $F^{y,H}$  represent the distribution  $F$ , and let  $F^{y_n, H_n}$  be obtained by replacing  $y$  and  $H$  with  $y_n$  and  $H_n$ , respectively. We denote  $m_n^0(z)$  as  $m_{F^{y_n, H_n}}^0(z)$ , which satisfies the equation (2.1). In other words, we have

$$(2.2) \quad \underline{m}_n^0(z) = \frac{1}{y_n \int \frac{t}{t \underline{m}_n^0(z) + 1} dH_n(t) - z}$$

and

$$(2.3) \quad z = -\frac{1}{\underline{m}_n^0} + y \int \frac{t}{t \underline{m}_n^0 + 1} dH_n(t).$$

Here,  $-z \underline{m}_n^0(z) = 1 - y_n - y_n z m_n^0(z)$ .

**2.5. The separation of spectrum.** The subsequent theorem provides further understanding regarding the convergence of sample eigenvalues. We present a comprehensive theorem that demonstrates the absence of eigenvalues with high probability outside the limiting support. This result is also fundamental in establishing the central limit theorem for the LSS as it provides a high probability bound on the spectral norm of  $\mathbf{S}_n$ .

**THEOREM 2.2 (No outside eigenvalues).** *Consider the gram matrix  $\mathbf{S}_n \Theta_n$ . Assume that the interval  $[a, b]$  with  $a > 0$  lies outside the support of  $F^{c,H}$  and  $F^{c_n, H_n}$  for all large  $n$ . Then with high probability, no eigenvalue of  $\mathbf{S}_n \Theta_n$  appears in  $[a, b]$  for all large  $n$ .*



To prove Theorem 2.2, we adopt the strategy employed by Bai and Silverstein [2]. Specifically, we begin by establishing a primary bound on the spectral norm of  $\mathbf{S}_n$ . Subsequently, we complete the proof of this theorem. The proof of Theorem 2.2 is postponed to Section 4. We are in position to introduce a lemma that bounds the spectral norm of  $\mathbf{S}_n$  by a constant-order quantity.

LEMMA 2.3. *Under the assumptions of Theorem 2.2, with high probability, we have*

$$\|\mathbf{S}_n\|_s \leq C(1 + \sqrt{y})^2.$$

To prove the above lemma, we employ the  $\varepsilon$ -net argument. This approach draws inspiration from the proof of Theorem 4.4.5 in Vershynin [17], which considers a random vector with i.i.d. sub-Gaussian entries. However, our model deals with more general variables. To begin, we introduce some necessary definitions and lemmas, the proofs of which can be found in Vershynin [17].

DEFINITION 2.3 ( $\varepsilon$ -net). Let  $(T, d)$  be a metric space. Consider a subset  $K \subset T$  and let  $\varepsilon > 0$ . A subset  $\mathcal{N} \subseteq K$  is called an  $\varepsilon$ -net of  $K$  if every point in  $K$  is within distance  $\varepsilon$  of some point of  $\mathcal{N}$ , that is,

$$\forall x \in K, \quad \exists x_0 \in \mathcal{N} : d(x, x_0) \leq \varepsilon.$$

DEFINITION 2.4 (Covering numbers). The smallest possible cardinality of an  $\varepsilon$ -net of  $K$  is called the *covering number* of  $K$  and is denoted  $\mathcal{N}(K, d, \varepsilon)$ . Equivalently,  $\mathcal{N}(K, d, \varepsilon)$  is the smallest number of closed balls with centers in  $K$  and radii  $\varepsilon$  whose union covers  $K$ .

LEMMA 2.4 (Covering numbers of the Euclidean sphere). *The covering numbers of the unit Euclidean sphere  $\mathbb{S}^{p-1}$  with Euclidean distance satisfy that for any  $\varepsilon > 0$ :*

$$\left(\frac{1}{\varepsilon}\right)^p \leq \mathcal{N}(\mathbb{S}^{p-1}, \varepsilon) \leq \left(\frac{2}{\varepsilon} + 1\right)^p.$$

LEMMA 2.5 (Quadratic form on a net). *Let  $\mathbf{A}$  be a  $p \times n$  matrix and  $\varepsilon \in [0, 1/2)$ . For any  $\varepsilon$ -net  $\mathcal{P}$  of the sphere  $\mathbb{S}^{p-1}$  and any  $\varepsilon$ -net  $\mathcal{N}$  of the sphere  $\mathbb{S}^{n-1}$ , we have*

$$\sup_{\alpha \in \mathcal{P}, \beta \in \mathcal{N}} \alpha^T \mathbf{A} \beta \leq \|\mathbf{A}\| \leq \frac{1}{1 - 2\varepsilon} \sup_{\alpha \in \mathcal{P}, \beta \in \mathcal{N}} \alpha^T \mathbf{A} \beta.$$

PROOF OF LEMMA 2.3. Our objective is to ensure control over  $\alpha^T (\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n) \beta$  for all vectors  $\alpha$  and  $\beta$  belonging to the unit sphere, as defined by the operator norm. To accomplish this, we adopt a two-step approach. First, we establish a rigorous control over  $\alpha^T (\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n) \beta$  for fixed vectors  $\alpha$  and  $\beta$ . This step allows us to obtain precise bounds on the expression. Second, we discretize the unit sphere by utilizing a carefully chosen net, which enables us to approximate the sphere with a finite set of representative points. By leveraging this discretization, we can then employ a union bound over all pairs of  $\alpha$  and  $\beta$  within the net. By taking into account the tight control established in the first step and the discretization scheme in the second step, we can confidently conclude the proof.



2.5.1. *Bound for fixed vectors.* Let  $r_{ij} = \mathbf{e}_i^T \Sigma_n^{1/2} \mathbf{e}_j$ . For any fixed  $\alpha \in \mathbb{S}^{p-1}$  and  $\beta \in \mathbb{S}^{n-1}$ , we may write the expression  $\alpha^T (\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n) \beta$  as follows:

$$\alpha^T (\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n) \beta = \sum_{j=1}^n \sum_{k=1}^p \left( \sum_{i=1}^p d_{ij} \alpha_i r_{ik} \right) \beta_j x_{kj} \triangleq \sum_{j=1}^n \sum_{k=1}^p \varrho_{kj} \beta_j x_{kj},$$

where  $\varrho_{kj} = \sum_{i=1}^p d_{ij} \alpha_i r_{ik}$  is a weighted sum of independent Bernoulli variables  $d_{ij}$ . To bound  $\sum_{j=1}^n \sum_{k=1}^p \varrho_{kj} \beta_j x_{kj}$ , we observe that the variables  $x_{kj}$  are independent sub-Gaussian variables with sub-Gaussian norm  $\|x\|_g$ . By utilizing the concentration properties of sub-Gaussian random variables (Lemma 5.2) and the properties of sub-Gaussian norms (Lemma 5.1), for any  $\Upsilon \geq 0$  and for all choices of  $d_{ij}^\odot \in \{0, 1\}^{pn}$ , we have

$$\mathbb{P}\{\alpha^T (\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n) \beta \geq \Upsilon | d_{ij} = d_{ij}^\odot, i = 1, \dots, p, j = 1, \dots, n\} \leq 2 \exp(-c\Upsilon^2/C_g^2),$$

where

$$\begin{aligned} C_g^2 &\leq C \sum_{j=1}^n \sum_{k=1}^p \varrho_{kj}^2 \beta_j^2 = C \sum_{j=1}^n \sum_{k=1}^p \sum_{i_1, i_2=1}^p d_{i_1 j}^\odot d_{i_2 j}^\odot \alpha_{i_1} \alpha_{i_2} r_{i_1 k} r_{i_2 k} \beta_j^2 \\ &\leq C \alpha^T \Sigma_n \alpha \leq C \|\Sigma_n\|_s. \end{aligned}$$

Thus, we arrive at

$$\begin{aligned} &\mathbb{P}\{\alpha^T (\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n) \beta \geq \Upsilon\} \\ &= \sum_{d_{ij}^\odot \in \{0, 1\}^{pn}} \mathbb{P}\{\alpha^T (\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n) \beta \geq \Upsilon | d_{ij} = d_{ij}^\odot\} \mathbb{P}\{d_{ij} = d_{ij}^\odot\} \leq \exp(-C\Upsilon^2), \end{aligned}$$

for some constant  $C$ .

2.5.2. *Union bound:.* Let  $\varepsilon = 1/4$ . Utilizing Lemma 2.4, we can construct an  $\varepsilon$ -net denoted as  $\mathcal{P}$  for the sphere  $\mathbb{S}^{p-1}$  and an  $\varepsilon$ -net denoted as  $\mathcal{N}$  for the sphere  $\mathbb{S}^{n-1}$ , where the cardinalities are bounded as follows:

$$|\mathcal{P}| \leq 9^p \quad \text{and} \quad |\mathcal{N}| \leq 9^n.$$

By further employing Lemma 2.5, we can establish an upper bound on the spectral norm of  $\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n$  using these nets as follows:

$$\|\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n\|_s \leq 2 \max_{\alpha \in \mathcal{P}, \beta \in \mathcal{N}} \alpha^T (\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n) \beta.$$

By utilizing the bound for fixed vectors, we obtain:

$$\begin{aligned} &\mathbb{P}\{\|\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n\|_s \geq 2C_1(\sqrt{p} + \sqrt{n})\} \\ &\leq \sum_{\alpha \in \mathcal{P}, \beta \in \mathcal{N}} \mathbb{P}\{\alpha^T (\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n) \beta \geq C_1(\sqrt{p} + \sqrt{n})\} \leq 9^{p+n} \exp(-CC_1^2(\sqrt{p} + \sqrt{n})^2). \end{aligned}$$

If we choose the constant  $C_1$  to be sufficiently large, we can conclude that:

$$\mathbb{P}\{\|\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n\|_s \geq 2C_1(\sqrt{p} + \sqrt{n})\} \leq \exp(-n).$$

Finally, we can conclude that with high probability, the spectral norm of  $\mathbf{S}_n$  is bounded as follows:

$$\|\mathbf{S}_n\|_s \leq C(1 + \sqrt{y})^2.$$

Hence, the proof of this lemma is complete.  $\square$

2.6. *CLT for the LSS.* Let us consider the LSS of  $\mathbf{S}_n \Theta_n$  associated with a test function  $f$ . This statistic is given by the integral:

$$L_n(f, \mathbf{S}_n \Theta_n) = \int f(x) dF^{\mathbf{S}_n \Theta_n}(x).$$

This kind of statistics arise in many statistical problems, especially in hypotheses tests considering the covariance structure of the population. In order to establish the second-order convergence of the LSS, a centralization step is performed to eliminate the leading order term. This is achieved by subtracting the integral of  $f(x)$  with respect to  $dF^{c_n, H_n}(x)$  from the integral of  $f(x)$  with respect to  $dF^{\mathbf{S}_n \Theta_n}(x)$ . Consequently, the centralized LSS can be expressed as follows:

$$L^c(f, \mathbf{S}_n \Theta_n) = \int f(x) d[F^{\mathbf{S}_n \Theta_n}(x) - F^{c_n, H_n}(x)].$$

Here,  $F^{\mathbf{S}_n \Theta_n}(x) - F^{c_n, H_n}(x)$  provides a measure of the deviation from the finite-dimensional approximation of the LSD. By examining the convergence properties and establishing the second-order convergence of this quantity, we can gain insights into the behavior of the LSS with respect to the underlying parameters and better understand its statistical properties.

We now present the formal results that establish the limiting fluctuation of a collection of LSSs in detail. This convergence theorem allows us to make reliable inference about the underlying population parameters based on the observed LSSs.

**THEOREM 2.3 (CLT for LSS).** *Let  $f_1, \dots, f_\kappa$  be functions on  $\mathbb{R}$  analytic on an open interval containing  $[\liminf_n \lambda_{\min}^{\Sigma_n \Theta_n} I_{(0,1)}(y)(1 - \sqrt{y})^2, \limsup_n \lambda_{\max}^{\Sigma_n \Theta_n} (1 + \sqrt{y})^2]$ . Under Assumptions A-D, we have that the  $\kappa$  dimensional random vector  $(L^c(f_j, \mathbf{S}_n \Theta_n))_{j=1}^\kappa$  is tight and converges weakly to a Gaussian vector  $(X_{f_j})_{j=1}^\kappa$ . The respective expectation is*

$$(2.4) \quad \mathbb{E}X_f = -\frac{1}{2\pi i} \oint f(z) \left\{ \frac{y \int \frac{m^3(z)t^2}{(\underline{m}(z)t+1)^3} dH(t)}{(1 - y \int \frac{m^2(z)t^2}{(\underline{m}(z)t+1)^2} dH(t))^2} + \frac{\underline{m}^3(z)a(z)}{1 - y \int \frac{m^2(z)t^2}{(\underline{m}(z)t+1)^2} dH(t)} \right\} dz,$$

while the covariance is given by

$$(2.5) \quad \begin{aligned} \text{Cov}(X_f, X_g) &= -\frac{1}{2\pi^2} \oint \oint f(z_1)g(z_2) \left( \frac{\frac{dm(z_1)}{dz_1} \frac{dm(z_2)}{dz_2}}{(\underline{m}(z_2) - \underline{m}(z_1))^2} - \frac{1}{(z_1 - z_2)^2} \right) dz_1 dz_2 \\ &\quad - \frac{1}{4\pi^2} \oint \oint f(z_1)g(z_2) \frac{\partial^2 \underline{m}(z_1) \underline{m}(z_2) d_2(z_1, z_2)}{\partial z_2 \partial z_1} dz_1 dz_2, \end{aligned}$$

where  $f, g \in \{f_1, \dots, f_\kappa\}$ . The limiting parameters can be expressed as follows by denoting  $\mathbb{T}_n^{(1)}(\Theta_n, \Sigma_n, \mathbb{P}_n, z) = \Theta_n^{1/2}(\underline{m}(z)\Psi_n + \mathbf{I}_p)^{-1}\Theta_n^{1/2}$  and  $\mathbb{T}_n^{(2)}(\Theta_n, \Sigma_n, \mathbb{P}_n, z) = \Theta_n^{1/2}(\underline{m}(z)\Psi_n + \mathbf{I}_p)^{-2}\Theta_n^{1/2}$ :

$$\begin{aligned} a(z) &= \lim_{n \rightarrow \infty} n^{-1} \mathcal{I}_2(\mathbb{T}_n^{(1)}(\Theta_n, \Sigma_n, \mathbb{P}_n, z), \mathbb{T}_n^{(2)}(\Theta_n, \Sigma_n, \mathbb{P}_n, z), \Sigma_n, \mathbb{P}_n) \\ &\quad + (v_4 - 3) \lim_{n \rightarrow \infty} n^{-1} \mathcal{II}(\mathbb{T}_n^{(1)}(\Theta_n, \Sigma_n, \mathbb{P}_n, z), \mathbb{T}_n^{(2)}(\Theta_n, \Sigma_n, \mathbb{P}_n, z), \Sigma_n, \mathbb{P}_n). \end{aligned}$$

$$\begin{aligned} d_2(z_1, z_2) &= \lim_{n \rightarrow \infty} n^{-1} \mathcal{I}_2(\mathbb{T}_n^{(1)}(\Theta_n, \Sigma_n, \mathbb{P}_n, z_1), \mathbb{T}_n^{(1)}(\Theta_n, \Sigma_n, \mathbb{P}_n, z_2), \Sigma_n, \mathbb{P}_n) \\ &\quad + (v_4 - 3) \lim_{n \rightarrow \infty} n^{-1} \mathcal{II}(\mathbb{T}_n^{(1)}(\Theta_n, \Sigma_n, \mathbb{P}_n, z_1), \mathbb{T}_n^{(1)}(\Theta_n, \Sigma_n, \mathbb{P}_n, z_2), \Sigma_n, \mathbb{P}_n). \end{aligned}$$

Moreover, both in equation (2.4) and (2.5), the contours are closed and taken in the positive direction in the complex plane. Each contour encloses the support of  $F^{y,H}$ , with the two contours in equation (2.5) assumed to be nonoverlapping.

REMARK 2.3. A thorough comparison between our theorem and Theorem 1.1 in Bai and Silverstein [3] reveals the noteworthy impact of missing at random on the CLT for LSSs. It becomes apparent that the inclusion of additional terms in both the means and variance-covariance functions can introduce significant bias if one simply treats  $\mathbf{T}_n$  as the population covariance matrix for incomplete data and applies the ordinary CLT without considering the influence of missing at random on the second-order properties under the nonnull case. In contrast to the case of complete observations, where the CLT for LSS is primarily influenced by the eigenvalues of the population covariance matrix  $\Sigma$  under the assumption of a Gaussian distribution, the situation changes when dealing with missing observations. Even under the ideal assumption of Gaussianity, the CLT for LSS under the nonnull case is strongly affected by the eigenvectors of  $\Sigma_n$  and  $\Theta_n$ , not just their eigenvalues.

To provide a more lucid explanation and isolate the impact of the auxiliary matrix  $\Theta_n$ , we set  $\Theta_n = \mathbf{I}_p$  in the aforementioned CLT. In this case, we have

$$a(z) = \lim_{n \rightarrow \infty} n^{-1} \mathcal{I}_2((\underline{m}(z)\mathbf{T}_n + \mathbf{I}_p)^{-1}, (\underline{m}(z)\mathbf{T}_n + \mathbf{I}_p)^{-2}, \Sigma_n, \mathbb{P}_n) \\ + (\nu_4 - 3) \lim_{n \rightarrow \infty} n^{-1} \mathcal{II}((\underline{m}(z)\mathbf{T}_n + \mathbf{I}_p)^{-1}, (\underline{m}(z)\mathbf{T}_n + \mathbf{I}_p)^{-2}, \Sigma_n, \mathbb{P}_n).$$

Taking into account the definitions of  $\mathcal{I}_2$  and  $\mathcal{II}$  in Lemma 2.1, we observe that  $a(z)$  is determined by numerous terms and affected not only by the eigenvalues of  $\Sigma_n$  but also by its eigenvectors. An illustrative example of such a term is

$$\text{tr}[\mathbb{P}^{(2)}((\underline{m}(z)\mathbf{T}_n + \mathbf{I}_p)^{-1} \circ (\underline{m}(z)\mathbf{T}_n + \mathbf{I}_p)^{-2})\mathbb{P}^{(2)}(\Sigma_n \circ \Sigma_n)].$$

The discussion on  $d_2(z_1, z_2)$  is similar. For a more profound understanding, it is crucial to note that in the case of complete observations under Gaussian distribution, the influence on the LSSs from the left eigenmatrix of  $\Sigma_n^{1/2}$  is nullified by the matching eigenvalues of  $\mathbf{A}\mathbf{B}$  and  $\mathbf{B}\mathbf{A}$  for any square matrices  $\mathbf{A}$  and  $\mathbf{B}$ . Similarly, the effect on the LSSs from the right eigenmatrix of  $\Sigma_n^{1/2}$  is nullified by the orthogonal invariance of the Gaussian distribution. However, the missing at random scenario significantly alters these dynamics by preventing the cancellation effect from the left eigenmatrix. This uniqueness characterizes the missing at random case as distinct and notable in its impact on the LSSs.

This observation underscores the importance of accounting for missingness when characterizing the asymptotic behavior of LSSs and highlights the need for specialized methodologies in high dimensional statistics that appropriately address the effects of missing data.

REMARK 2.4. To establish the consistency of our CLT with those derived for complete data in previous works, consider setting  $\Theta_n = \mathbf{I}_p$ . In the case of complete data, where  $p_j = 1$  for  $1 \leq j \leq p$ , we have  $\mathbb{P} = \mathbf{I}_p$  and  $\mathbb{P}^{(2)} = \mathbb{P}^{(3)} = \mathbb{P}^{(4)} = \mathbb{P}^{(5)} = \mathbf{0}$ . Consequently, this leads to the disappearance of the term  $\mathcal{I}_2$  and the reduction of the term  $\mathcal{II}$  in  $a(z)$  to:

$$\text{tr}[(\Sigma_n^{1/2}(\underline{m}(z)\Sigma_n + \mathbf{I}_p)^{-1}\Sigma_n^{1/2}) \circ (\Sigma_n^{1/2}(\underline{m}(z)\Sigma_n + \mathbf{I}_p)^{-2}\Sigma_n^{1/2})],$$

as well as the simplification of the term  $\mathcal{II}$  in  $d_2(z_1, z_2)$  to:

$$\text{tr}[(\Sigma_n^{1/2}(\underline{m}(z)\Sigma_n + \mathbf{I}_p)^{-1}\Sigma_n^{1/2}) \circ (\Sigma_n^{1/2}(\underline{m}(z)\Sigma_n + \mathbf{I}_p)^{-1}\Sigma_n^{1/2})].$$

This results in the degeneration of our CLT to the classical one under complete data. Refer to Theorem 1.1 in Bai and Silverstein [3] for the case  $\nu_4 = 3$  and Theorem 1.4 in Pan and Zhou [15] for a more general scenario.

We present a corollary for the renormalized case where  $\Theta_n = \mathbf{T}_n^{-1}$ , leading to  $\Psi_n = \mathbf{I}_p$ . It is demonstrated that in this scenario, the limiting parameters in the general CLT manifest a more accessible form and offer computational simplicity. Furthermore, we apply this corollary to a test problem in the subsequent section.

**COROLLARY 2.1 (Renormalized case).** *Under the Assumptions of 2.3, assume additionally that  $\Theta_n = \mathbf{T}_n^{-1}$ , then we have the mean and (co)variance functions as*

$$\begin{aligned} EX_{f_\ell} &= \lim_{r \rightarrow 1^+} \frac{1}{2\pi i} \oint_{|\xi|=1} f_\ell(|1 + \sqrt{y}\xi|^2) \left( \frac{\xi}{\xi^2 - r^{-2}} - \frac{1}{\xi} \right) d\xi \\ &\quad + \frac{a_{\mathbb{P}, \Sigma}}{2\pi i} \oint_{|\xi|=1} \frac{f_\ell(|1 + \sqrt{y}\xi|^2)}{\xi^3} d\xi \end{aligned}$$

and

$$\begin{aligned} \text{Cov}(X_{f_{\ell_1}}, X_{f_{\ell_2}}) &= \lim_{r \rightarrow 1^+} \frac{-1}{2\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \frac{f_{\ell_1}(|1 + \sqrt{y}\xi_1|^2) f_{\ell_2}(|1 + \sqrt{y}\xi_2|^2)}{(\xi_1 - r\xi_2)^2} d\xi_2 d\xi_1 \\ &\quad - \frac{a_{\mathbb{P}, \Sigma}}{4\pi^2} \oint_{|\xi_1|=1} \frac{f_{\ell_1}(|1 + \sqrt{y}\xi_1|^2)}{\xi_1^2} d\xi_1 \oint_{|\xi_2|=1} \frac{f_{\ell_2}(|1 + \sqrt{y}\xi_2|^2)}{\xi_2^2} d\xi_2, \end{aligned}$$

where  $\ell, \ell_1, \ell_2 \in \{1, \dots, \kappa\}$ , the contour  $\oint$  is anticlockwise and

$$a_{\mathbb{P}, \Sigma} = \lim_{n \rightarrow \infty} p^{-1} \mathcal{I}_2(\mathbf{T}_n^{-1}, \mathbf{T}_n^{-1}, \Sigma_n, \mathbb{P}_n) + (v_4 - 3) \lim_{n \rightarrow \infty} p^{-1} \mathcal{II}(\mathbf{T}_n^{-1}, \mathbf{T}_n^{-1}, \Sigma_n, \mathbb{P}_n).$$

**3. Testing for covariance structure under missing observations.** The problem of testing for covariance matrix equality plays a crucial role in various statistical applications, including multivariate analysis, portfolio optimization, and pattern recognition. In high-dimensional settings, testing for covariance structure becomes particularly challenging due to the curse of dimensionality. Furthermore, when missing observations are present, the problem becomes even more complex. In this section, we discuss the application of our theoretical results for testing the equality of high-dimensional covariance matrices to a given matrix in the presence of missing at random.

**3.1. Description of the test problem.** We adopt the data model discussed in the previous section. Specifically, we draw an incomplete sample of  $n$  observations from a centered  $p$ -dimensional population  $\mathbf{y} = (y_1, \dots, y_p)^T$ . The missingness is governed by a missing probability matrix  $\mathbb{P}$ , which determines the probability of each component being observed. Similarly to Lounici [13], where the estimation of the covariance matrix is considered, we assume that the missing probabilities are known. For  $1 \leq i \leq p$  and  $1 \leq j \leq n$ , we define  $d_{ij} = 1$  if the  $i$ th variate of the  $j$ th sample is observed, and  $d_{ij} = 0$  otherwise. Let  $\mathbb{D}_j = \text{diag}(d_{1j}, \dots, d_{pj})$ . We compute the corresponding Gram matrix as follows:

$$\hat{\mathbf{S}}_n = \frac{1}{n} \sum_{j=1}^n \mathbb{D}_j \mathbf{y}_j \mathbf{y}_j^T \mathbb{D}_j.$$

We consider the null and alternative hypotheses for testing the equality of high-dimensional covariance matrices, denoted by  $\Sigma = \mathbf{E} \mathbf{y} \mathbf{y}^T$ :

$$H_0 : \Sigma = \Sigma_0 \quad \text{vs.} \quad H_1 : \Sigma \neq \Sigma_0.$$

Here,  $\Sigma_0$  represents the given covariance matrix against which we are testing the equality.

**3.2. Test procedure under missing observations.** In this subsection, we describe the test procedure for handling missing observations. Specifically, we define the matrix  $\mathbf{T}_0 = \mathbb{P}\boldsymbol{\Sigma}_0\mathbb{P} + \mathbb{P}^{(2)} \circ \boldsymbol{\Sigma}_0$ , where  $\mathbb{P}$  represents the missing probability matrix. We consider two test statistics,

$$\mathcal{T}_F = \text{tr}(\mathbf{T}_0^{-1}\hat{\mathbf{S}}_n - \mathbf{I}_p)^2, \quad \mathcal{T}_L = \log \det \mathbf{T}_0^{-1}\hat{\mathbf{S}}_n.$$

We now present the following theorem regarding the null distribution of these two statistics.

**THEOREM 3.1 (Test statistics).** *Under  $H_0$ , we have*

$$\begin{aligned} \sigma_L^{-1}(\mathcal{T}_L - \mu_L) &\rightarrow N(0, 1), \quad p/n \in (0, 1), \\ \sigma_F^{-1}(\mathcal{T}_F - \mu_F) &\rightarrow N(0, 1), \end{aligned}$$

where

$$\begin{aligned} \mu_L &= n(y_n - 1) \log(1 - y_n) - p + \frac{\log(1 - y_n)}{2} - \frac{y_n a_{\mathbb{P}, \boldsymbol{\Sigma}}}{2}, \\ \mu_F &= py_n + y_n(a_{\mathbb{P}, \boldsymbol{\Sigma}} + 1), \\ \sigma_L^2 &= -2 \log(1 - y_n) + y_n a_{\mathbb{P}, \boldsymbol{\Sigma}}, \quad \sigma_F^2 = 4y_n^2 + 4y_n^3(a_{\mathbb{P}, \boldsymbol{\Sigma}} + 2). \end{aligned}$$

According to the above theorem, given test level  $\alpha$ , we reject  $H_0$  based on the statistics  $\mathcal{T}_L, \mathcal{T}_F$  on the following region:

$$(3.1) \quad \{\mathbb{D}_1, \dots, \mathbb{D}_n, \mathbf{y}_1, \dots, \mathbf{y}_n : \sigma_L^{-1}|\mathcal{T}_L - \mu_L| > z_{1-\alpha/2}\},$$

$$(3.2) \quad \{\mathbb{D}_1, \dots, \mathbb{D}_n, \mathbf{y}_1, \dots, \mathbf{y}_n : \sigma_F^{-1}|\mathcal{T}_F - \mu_F| > z_{1-\alpha/2}\},$$

where  $z_{1-\alpha/2}$  is the  $1 - \alpha/2$  quantile of the standard normal distribution.

**3.3. Numerical study.** In this section, we present a simulation study to assess the performance of the proposed statistics for testing covariance structures under missing observations.

**3.3.1. Experimental setup.** We consider populations with dimensions  $p \in \{64, 128, 256, 512\}$  and dimension-to-sample size ratios of  $y_n \in \{0.25, 0.5, 0.8\}$ , respectively. For each combination of  $(p, y_n)$ , we generate  $n$  i.i.d. data from a  $p$ -dimensional distribution with zero mean and covariance matrix  $\boldsymbol{\Sigma}_\theta = (s_{i,j,\theta})_{p \times p}$ .

The covariance matrix elements are defined as  $s_{i,j,\theta} = 0.65^{|i-j|+1} + \theta$ , where  $\theta$  differentiates the null hypothesis ( $\theta = 0$ ) from the alternative hypothesis ( $\theta \neq 0$ ).

To introduce missing observations, we generate a vector of missing probabilities  $\mathbf{b} = (\mathbf{p}_1, \dots, \mathbf{p}_p)^T$ , with  $\mathbf{p}_1, \dots, \mathbf{p}_p \sim \text{Uniform}(0.5, 1)$ . We consider two underlying distributions: Gaussian distribution, where  $x_{ij} \sim N(0, 1)$ , and Gamma distribution, where  $(x_{ij} + 2) \sim \Gamma(4, 0.5)$ .

The test statistics are computed under the null hypothesis ( $\theta = 0$ ). Each simulation experiment is replicated 10,000 times to ensure robustness of the results.

**3.3.2. Simulation results.** Table 1 and Table 2 present the simulation results for Gaussian distribution ( $\nu_4 = 3$ ) and Gamma distribution ( $\nu_4 = 4.5$ ), respectively. The tables display the empirical size and empirical power of the  $\mathcal{T}_L$  and  $\mathcal{T}_F$  statistics.

Overall, both  $\mathcal{T}_L$  and  $\mathcal{T}_F$  demonstrate favorable performance across all scenarios. They exhibit similar empirical sizes, indicating good control of the type I error rate. Regarding empirical power,  $\mathcal{T}_F$  consistently outperforms  $\mathcal{T}_L$  in all cases, providing higher power to detect departures from the null hypothesis.

These simulation results support the effectiveness of the proposed statistics for testing covariance structures under missing observations.

TABLE 1  
Percentages for empirical sizes and powers of the tests  $\mathcal{T}_L$  and  $\mathcal{T}_F$  under Gaussian distribution

$p$	$p/n$	Empirical sizes		Empirical powers			
		$\theta = 0$		$\theta = 0.01$		$\theta = 0.03$	
		$\mathcal{T}_L$	$\mathcal{T}_F$	$\mathcal{T}_L$	$\mathcal{T}_F$	$\mathcal{T}_L$	$\mathcal{T}_F$
64	0.8	5.08	6.20	6.09	8.46	11.41	23.15
	0.5	4.84	6.16	6.41	9.28	19.12	32.33
	0.25	5.24	5.85	8.88	11.03	36.08	49.74
128	0.8	5.05	5.83	8.59	12.17	28.96	62.13
	0.5	4.85	6.09	11.75	14.89	53.37	82.36
	0.25	5.32	4.77	20.05	22.11	83.96	98.72
256	0.8	5.03	5.43	16.75	30.07	72.72	99.75
	0.5	4.78	5.23	29.04	44.48	96.87	100
	0.25	5.25	5.26	54.23	73.05	99.98	100
512	0.8	4.99	5.23	46.39	87.31	100	100
	0.5	4.89	5.46	75.20	98.42	100	100
	0.25	5.26	4.95	98.02	100	100	100

**4. Proof of main theorems.** In this section, we provide a sketch of the proofs for the main theorems presented in this paper. To maintain the brevity of the main text, we defer the detailed and rigorous proofs to the supplement material. Here, we outline the key ideas and techniques employed in establishing these results.

**4.1. The proof of Theorem 2.1.** We proceed to present the proof of the convergence in high probability of the empirical spectral distribution (ESD). It is noteworthy that if our aim were to establish convergence in almost sure sense, the main theorem established in Bai and Zhou [1] would suffice considering the results of Lemma 2.1. However, we recognize the need for an additional intermediate result and the requirement of certain primary procedures that will be utilized in the proofs of subsequent theorems. Consequently, we choose to document the main proof procedures in this section.

TABLE 2  
Percentages for empirical sizes and powers of the tests  $\mathcal{T}_L$  and  $\mathcal{T}_F$  under Gamma distribution

$p$	$p/n$	Empirical sizes		Empirical powers			
		$\theta = 0$		$\theta = 0.01$		$\theta = 0.03$	
		$\mathcal{T}_L$	$\mathcal{T}_F$	$\mathcal{T}_L$	$\mathcal{T}_F$	$\mathcal{T}_L$	$\mathcal{T}_F$
64	0.8	4.57	8.71	5.18	10.49	9.64	21.42
	0.5	4.75	8.47	6.15	11.25	14.40	28.61
	0.25	4.93	8.64	7.33	13.01	27.29	45.35
128	0.8	4.71	6.75	7.29	12.16	24.10	52.20
	0.5	4.84	6.68	9.56	14.29	41.22	72.51
	0.25	5.16	7.04	15.35	21.80	71.54	96.80
256	0.8	4.94	5.67	13.78	24.76	62.51	98.53
	0.5	4.60	5.94	21.23	36.57	98.48	99.98
	0.25	5.15	5.77	40.71	65.39	99.65	100
512	0.8	5.13	5.58	36.26	76.20	100	100
	0.5	5.04	5.22	61.48	94.51	100	100
	0.25	5.36	4.95	90.74	100	100	100

Notice that

$$m_n(z) = m_{F\mathbf{S}_n}(z) = \frac{1}{p} \text{tr}(\mathbf{\Theta}_n^{1/2} \mathbf{S}_n \mathbf{\Theta}_n^{1/2} - z \mathbf{I}_p)^{-1},$$

where  $z = u + vi \in \mathbb{C}^+$ . Let  $\mathbf{G}_n = \mathbf{G}_n(z) = b_n(z) \Psi_n - z \mathbf{I}_p$ , where

$$b_n(z) = \frac{1}{1 + \frac{1}{n} \text{Etr}(\mathbf{M}^{-1} \Psi_n)}.$$

We proceed to prove the following steps:

S1 For any given  $z$ ,  $m_n(z) - \text{E}m_n(z) = o_{w.h.p.}(1)$ .

S2: For any given  $z$ ,  $\frac{1}{p} \text{Etr} \mathbf{G}_n^{-1} - \text{E}m_n(z) = o(1)$ .

S3: For any given  $z$ ,  $b_n(z) + z \hat{m}_n(z) = o_{w.h.p.}(1)$ . Then the result for LSD follows.

4.1.1. *Proof of S1:*. In this section, we show that

$$(4.1) \quad m_n(z) - \text{E}m_n(z) \xrightarrow{w.h.p.} 0.$$

Let  $\mathbb{D}_k$  represent a  $p \times p$  diagonal matrix that consists of the entries in the  $k$ th column of  $\mathbf{D}_n$ . Denote  $\mathbf{y}_k = \mathbf{\Sigma}_n^{1/2} \mathbf{x}_k$ ,  $\mathbf{S}_n^k = \mathbf{S}_n - \frac{1}{n} \mathbb{D}_k \mathbf{y}_k \mathbf{y}_k^T \mathbb{D}_k$ ,  $\beta_k(z) = \frac{1}{1 + n^{-1} \mathbf{y}_k^T \mathbb{D}_k \mathbf{\Theta}_n^{1/2} \mathbf{M}_k^{-1} \mathbf{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k}$ , with  $\mathbf{M} = \mathbf{M}(z) = \mathbf{\Theta}_n^{1/2} \mathbf{S}_n \mathbf{\Theta}_n^{1/2} - z \mathbf{I}_p$ ,  $\mathbf{M}_k = \mathbf{M}_k(z) = \mathbf{\Theta}_n^{1/2} \mathbf{S}_n^k \mathbf{\Theta}_n^{1/2} - z \mathbf{I}_p$ . Also let  $\text{E}_k(\cdot)$  denote the conditional expectation given  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \mathbb{D}_1, \dots, \mathbb{D}_k\}$ . Using the formula

$$(4.2) \quad (\mathbf{A} + \boldsymbol{\alpha} \boldsymbol{\beta}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1} \boldsymbol{\alpha} \boldsymbol{\beta}^T \mathbf{A}^{-1}}{1 + \boldsymbol{\beta}^T \mathbf{A}^{-1} \boldsymbol{\alpha}},$$

one finds

$$(4.3) \quad \begin{aligned} m_n(z) - \text{E}m_n(z) &= \frac{1}{p} \sum_{k=1}^n (\text{E}_k - \text{E}_{k-1}) [\text{tr}(\mathbf{M}^{-1}) - \text{tr}(\mathbf{M}_k^{-1})] \\ &= -\frac{1}{pn} \sum_{k=1}^n (\text{E}_k - \text{E}_{k-1}) \beta_k(z) \mathbf{y}_k^T \mathbb{D}_k \mathbf{\Theta}_n^{1/2} \mathbf{M}_k^{-2} \mathbf{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k. \end{aligned}$$

Note that for  $k = 1, \dots, n$ , the imaginary part of  $z(1 + n^{-1} \mathbf{y}_k^T \mathbb{D}_k \mathbf{\Theta}_n^{1/2} \mathbf{M}_k^{-1} \mathbf{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k)$  is not smaller than  $v$ , where  $v$  represents the imaginary part of  $z$ . Therefore, one can establish  $|\beta_k(z)| \leq \frac{|z|}{v}$ . Subsequently, by applying Lemma 5.3, Lemma 2.2 and the  $c_r$ -inequality, it follows that

$$(4.4) \quad \begin{aligned} \text{E}|m_n(z) - \text{E}m_n(z)|^\ell &\leq \frac{C_\ell n^{\ell/2} |z|^\ell}{p^\ell n^\ell v^\ell} \text{E}|\mathbf{y}_1^T \mathbb{D}_1 \mathbf{\Theta}_n^{1/2} \mathbf{M}_1^{-2} \mathbf{\Theta}_n^{1/2} \mathbb{D}_1 \mathbf{y}_1|^\ell \\ &\leq \frac{C_\ell n^{\ell/2}}{p^\ell} = O(n^{-\ell/2}). \end{aligned}$$

By the Chebyshev inequality, we then conclude  $m_n(z) - \text{E}m_n(z) \xrightarrow{w.h.p.} 0$ .

4.1.2. *Proof of S2:*. Write

$$(4.5) \quad \begin{aligned} \mathbf{G}_n^{-1} - \mathbf{M}^{-1} &= \frac{1}{n} \sum_{k=1}^n \mathbf{G}_n^{-1} \mathbf{\Theta}_n^{1/2} (\mathbb{D}_k \mathbf{y}_k \mathbf{y}_k^T \mathbb{D}_k - b_n(z) \mathbf{T}_n) \mathbf{\Theta}_n^{1/2} \mathbf{M}^{-1} \\ &= \frac{1}{n} \sum_{k=1}^n \beta_k(z) \mathbf{G}_n^{-1} \mathbf{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k \mathbf{y}_k^T \mathbb{D}_k \mathbf{\Theta}_n^{1/2} \mathbf{M}_k^{-1} - b_n(z) \mathbf{G}_n^{-1} \Psi_n \mathbf{M}^{-1}. \end{aligned}$$



Taking the trace and then the expectation, and dividing by  $p$ , it follows that

$$(4.6) \quad \begin{aligned} \frac{1}{p} \text{Etr} \mathbf{G}_n^{-1} - \text{E} m_n(z) &= \frac{1}{pn} \sum_{k=1}^n \text{E} [\beta_k(z) \mathbf{y}_k^T \mathbb{D}_k \boldsymbol{\Theta}_n^{1/2} \mathbf{M}_k^{-1} \mathbf{G}_n^{-1} \boldsymbol{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k] \\ &\quad - \frac{b_n(z)}{p} \text{tr} [\mathbf{G}_n^{-1} \Psi_n \mathbf{E} \mathbf{M}^{-1}]. \end{aligned}$$

It is easy to see by Lemma 2.2

$$\begin{aligned} &\left| \frac{1}{pn} \sum_{k=1}^n \text{E} [(\beta_k(z) - b_n(z)) \mathbf{y}_k^T \mathbb{D}_k \boldsymbol{\Theta}_n^{1/2} \mathbf{M}_k^{-1} \mathbf{G}_n^{-1} \boldsymbol{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k] \right| \\ &\leq \frac{1}{pn} \sum_{k=1}^n \text{E}^{1/2} |\mathbf{y}_k^T \mathbb{D}_k \boldsymbol{\Theta}_n^{1/2} \mathbf{M}_k^{-1} \mathbf{G}_n^{-1} \boldsymbol{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k|^2 \text{E}^{1/2} |\beta_k(z) - b_n(z)|^2 \leq \frac{C}{\sqrt{n}} \rightarrow 0, \end{aligned}$$

where we use the estimate

$$(4.7) \quad \begin{aligned} \text{E} |\beta_k(z) - b_n(z)|^2 &\leq \frac{C}{n^2} \text{E} |\mathbf{y}_k^T \mathbb{D}_k \boldsymbol{\Theta}_n^{1/2} \mathbf{M}_k^{-1} \boldsymbol{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k - \text{tr}(\mathbf{M}_k^{-1} \Psi_n)|^2 \\ &\quad + \frac{C}{n^2} \text{E} |\text{tr}(\mathbf{M}_k^{-1} \Psi_n) - \text{Etr}(\mathbf{M}^{-1} \Psi_n)|^2 \leq \frac{C}{n}. \end{aligned}$$

Therefore, combining the above inequality and (4.6), one finds

$$\begin{aligned} \frac{1}{p} \text{Etr} \mathbf{G}_n^{-1} - \text{E} m_n(z) &= \frac{b_n(z)}{p} \text{tr}[(\mathbf{E} \mathbf{M}_k^{-1} - \mathbf{E} \mathbf{M}^{-1}) \mathbf{G}_n^{-1} \Psi_n] + o(1) \\ &= \frac{b_n(z)}{pn} \sum_{k=1}^n \text{E} \left[ \frac{\frac{1}{n} \mathbf{y}_k^T \mathbb{D}_k \boldsymbol{\Theta}_n \mathbf{M}_k^{-1} \mathbf{G}_n^{-1} \boldsymbol{\Theta}_n^{1/2} \mathbf{T}_n \mathbf{M}_k^{-1} \boldsymbol{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k}{1 + \frac{1}{n} \mathbf{y}_k^T \mathbb{D}_k \boldsymbol{\Theta}_n^{1/2} \mathbf{M}_k^{-1} \boldsymbol{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k} \right] + o(1) \\ &= o(1). \end{aligned}$$

Consequently, we conclude that  $\frac{1}{p} \text{Etr} \mathbf{G}_n^{-1} - \text{E} m_n(z) = o(1)$ .

4.1.3. *Complete the proof.* Let  $\underline{\mathbf{S}}_n^\Theta = \frac{1}{n} (\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n)^T \Theta_n (\mathbf{D}_n \circ \Sigma_n^{1/2} \mathbf{X}_n)$ . Then we have  $F^{\underline{\mathbf{S}}_n^\Theta}(x) = 1 - y_n + y_n F^{\mathbf{S}_n^{\Theta_n}}(x)$  and

$$(4.8) \quad \underline{m}_n(z) = -\frac{1 - y_n}{z} + y_n m_n(z), \quad z \in \mathbb{C}^+,$$

where  $\underline{m}_n(z) = m_{F^{\underline{\mathbf{S}}_n^\Theta}}(z)$ . Let  $\widehat{m}_n(z) = \frac{1}{p} \text{tr} \mathbf{G}_n^{-1}(z)$  and  $\widehat{\underline{m}}_n(z) = -\frac{1 - y_n}{z} + y_n m_n^0(z)$ . By Sections 4.1.1 and 4.1.2, we conclude

$$(4.9) \quad \max\{m_n(z) - \widehat{m}_n(z), \underline{m}_n(z) - \widehat{\underline{m}}_n(z)\} \xrightarrow{w.h.p.} 0.$$

Recall  $\mathbf{M} + z \mathbf{I}_p = \frac{1}{n} \sum_{k=1}^n \boldsymbol{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k \mathbf{y}_k^T \mathbb{D}_k \boldsymbol{\Theta}_n^{1/2}$ . Multiplying the inverse of  $\mathbf{M}$  from the right on both sides yields  $\mathbf{I}_p + z \mathbf{M}^{-1} = \frac{1}{n} \sum_{k=1}^n \beta_k(z) \boldsymbol{\Theta}_n^{1/2} \mathbb{D}_k \mathbf{y}_k \mathbf{y}_k^T \mathbb{D}_k \boldsymbol{\Theta}_n^{1/2} \mathbf{M}_k^{-1}$ . Taking the trace on both sides and dividing by  $n$ , we find  $y_n + z y_n m_n(z) = 1 - \frac{1}{n} \sum_{k=1}^n \beta_k(z)$ . Together with (4.8), one gets

$$(4.10) \quad z \underline{m}_n(z) = -\frac{1}{n} \sum_{k=1}^n \beta_k(z).$$

Using (4.7) and (4.9), we deduce

$$b_n(z) = -z\mathbb{E}\underline{m}_n(z) + o\left(\frac{1}{\sqrt{n}}\right) = -z\widehat{m}_n(z) + o_{w.h.p.}(1).$$

Hence, we have

$$(4.11) \quad \widehat{m}_n(z) = \int \frac{1}{-z\widehat{m}_n(z)t - z} dH_n(t) + o_{w.h.p.}(1).$$

By  $|\widehat{m}_n(z)| < \frac{1}{v}$ , consider any convergent sequence  $(\widehat{m}_n(z))$  whose limit is denoted by  $m(z)$ . From (4.11), we get  $m(z) = \int \frac{1}{-zm(z)t - z} dH(t)$ , where  $\underline{m}(z) = -\frac{1-y}{z} + ym(z)$ . It can be shown that  $m(z)$  is the Stieltjes transform of a probability measure  $F$  on  $\mathbb{R}$ , analytic in  $\mathbb{C}^+$ . For each  $z \in \mathbb{C}^+$ ,  $m = m(z)$  is a solution to the equation

$$m = \int \frac{1}{t(1 - y - yzm) - z} dH(t),$$

which is unique in the set  $\{m \in \mathbb{C}^+ : -\frac{1-y}{z} + ym \in \mathbb{C}^+\}$ . The proof of this theorem is thus complete.

**4.2. Proof of Theorem 2.2.** The proof of the theorem at hand relies on analyzing the convergence of the resolvent. Our approach follows the method established by Bai and Silverstein [2], where the ordinary sample covariance matrix is considered. However, we aim to enhance their results from an almost sure convergence to a more precise notion, namely high probability under our model. Given the similarities between our proof process and theirs, we avoid duplicating their work and instead provide an overview, highlighting the nontrivial differences and our strategy for addressing these issues. It is important to note that readers can obtain a comprehensive understanding of the entire process by referring to the original paper.

The main objective is to achieve a convergence rate for  $m_n(z)$  when  $z$  is close, but not too close, to an interval outside the limiting support  $F^{y_n, H_n}$ . This can be accomplished by assigning an appropriate rate to the imaginary part, denoted by  $v_n = n^{-\delta}$ , where  $\delta \in (0, 1/68)$  is sufficiently small, such that  $q = (2\delta)^{-1}$  is a positive integer. Our goal is to prove the following expression:

$$(4.12) \quad \sup_{x \in [a, b]} nv_n |\underline{m}_n(x + iv_n) - \underline{m}_n^0(x + iv_n)| \xrightarrow{\text{w.h.p.}} 0.$$

Once this is established, we can multiply  $v_n$  by a constant  $k$  and obtain the following estimate:

$$\sup_{x \in [a, b]} |\underline{m}_n(x + i\sqrt{k}v_n) - \underline{m}_n^0(x + i\sqrt{k}v_n)| = o(1/(nv_n)) \quad \text{w.h.p.}$$

Next, by considering the imaginary part and taking differences, following the strategy in Section 6 of Bai and Silverstein [2] for better comprehension, we find, with high probability, the following expression holds:

$$\sup_{x \in [a, b]} \left| \int \frac{d(F^{\mathbf{S}_n}(t) - F^{y_n, H_n}(t))}{((x-t)^2 + v_n^2)((x-t)^2 + 2v_n^2) \cdots ((x-t)^2 + qv_n^2)} \right| = o(1).$$

Subsequently, we split up the integral and, with high probability, obtain the following expression:

$$\begin{aligned} & \sup_{x \in [a, b]} \left| \int \frac{I([a', b']^c) d(F^{\mathbf{S}_n}(t) - F^{y_n, H_n}(t))}{((x-t)^2 + v_n^2)((x-t)^2 + 2v_n^2) \cdots ((x-t)^2 + qv_n^2)} \right. \\ & \quad \left. + \sum_{t_j \in [a-\varepsilon, b+\varepsilon]} \frac{v_n^{2q}}{((x-t_j)^2 + v_n^2)((x-t_j)^2 + 2v_n^2) \cdots ((x-t_j)^2 + qv_n^2)} \right| = o(1). \end{aligned}$$

It follows that if each term in a subsequence satisfying the aforementioned equation contains at least one eigenvalue within the interval  $[a, b]$ , the sum in the equation will be uniformly bounded away from zero. As a result, the integral itself must also remain uniformly bounded away from zero. However, with high probability, the integral tends to zero as the integrand is bounded and, with high probability, both  $F^{\mathbf{B}_n}$  and  $F^{y_n, H_n}$  converge weakly to the same limit, which exhibits no mass on the interval  $[a - \varepsilon, b + \varepsilon]$ . Hence, for sufficiently large values of  $n$ , with high probability, no eigenvalues of  $\mathbf{B}_n$  will appear in the interval  $[a, b]$ . This conclusion establishes the proof of our theorem.

To achieve our goal, it is worth mentioning that all the mathematical tools presented in Section 2 of Bai and Silverstein [2] are applicable to our case. Since the underlying distribution is sub-Gaussian, there is no need to truncate the variables, and similar bounds for (3.1) and (3.2) can be achieved in our model by applying Lemma 2.2. Specifically, for any fixed real matrix  $\mathbf{A}$ , we have

$$\mathbb{E}|\mathbf{y}_1^T \mathbb{D}_1^T \mathbf{A} \mathbb{D}_1 \mathbf{y}_1 - \text{tr } \mathbf{A}|^q \leq C_q (\text{tr } \mathbf{A} \mathbf{A}^T)^{q/2},$$

where  $C_q$  depends on  $q$  and the moments of  $x_{11}$ . These high-order moment bounds for quadratic forms are useful for obtaining high probability results. To prove our main task (4.12), we follow similar steps as shown in Bai and Silverstein [2]. It is important to note that, to obtain high probability results, we only need to increase the order of moments whenever the moment bounds of quadratic forms are applied. Furthermore, by Lemma 2.3, we have obtained high probability bounds for the spectral norm. By carefully examining their proof of (4.1) and (5.1) in Bai and Silverstein [2], we find that all of their arguments are valid in our model. Moreover, we note that the almost sure result in (4.1) from their work can be improved to hold with high probability.

4.3. *The proof of Theorem 2.3.* Note that

$$(4.13) \quad \int f(x) dQ(x) = -\frac{1}{2\pi i} \oint f(z) m_Q(z) dz,$$

where  $Q$  is a cumulative distribution function (c.d.f.) and  $f$  is analytic on an open set containing the support of  $Q$ . The complex integral on the right-hand side is over any positively oriented contour enclosing the support of  $Q$  and on which  $f$  is analytic. Hence, the proof of Theorem 2.3 relies on establishing limiting results on

$$M_n(z) = p[m_n(z) - m_n^0(z)].$$

Let  $v_0$  be any positive number,  $x_r \in (\limsup_n \lambda_{\max}^{\Sigma_n \Theta_n} (1 + \sqrt{y})^2, \infty)$ . Let  $x_l$  be any negative number if  $\liminf_n \lambda_{\min}^{\Sigma_n \Theta_n} I_{(0,1)}(y)(1 - \sqrt{y})^2$  is zero. Otherwise choose  $x_l \in (0, \liminf_n \lambda_{\min}^{\Sigma_n \Theta_n} I_{(0,1)}(y)(1 - \sqrt{y})^2)$ . Let  $\mathcal{C}_u = \{x + iv_0 : x \in [x_l, x_r]\}$ . Define a contour  $\mathcal{C} = \{x_l + iv : v \in [0, v_0]\} \cup \mathcal{C}_u \cup \{x_r + iv : v \in [0, v_0]\}$ . To avoid dealing with small  $\Im z$ , we truncate  $M_n(z)$  on the contour  $\mathcal{C}$  of the complex plane. We choose a sequence  $\varepsilon_n$  decreasing to zero, satisfying, for some  $\alpha \in (0, 1)$ ,  $\varepsilon_n \geq n^{-\alpha}$ . Let  $\mathcal{C}_l = \{x_l + iv : v \in [n^{-1}\varepsilon_n, v_0]\}$  and  $\mathcal{C}_r = \{x_r + iv : v \in [n^{-1}\varepsilon_n, v_0]\}$ . Then, define  $\mathcal{C}_n = \mathcal{C}_l \cup \mathcal{C}_u \cup \mathcal{C}_r$ . For  $z = x + iv$ , we define the process  $\widehat{M}_n(\cdot)$  as

$$(4.14) \quad \widehat{M}_n(\cdot) = \begin{cases} M_n(z) & \text{for } z \in \mathcal{C}_n, \\ M_n(x_l + in^{-1}\varepsilon_n) & \text{for } x = x_l, v \in [0, n^{-1}\varepsilon_n], \\ M_n(x_r + in^{-1}\varepsilon_n) & \text{for } x = x_r, v \in [0, n^{-1}\varepsilon_n]. \end{cases}$$

It is observed that on the subset  $\mathcal{C}_n$  of  $\mathcal{C}$ ,  $M_n(\cdot)$  agrees with  $\widehat{M}_n(\cdot)$ . We are going to present the following CLT for the truncated process  $\widehat{M}_n(\cdot)$ . To be specific, we will establish the following lemma.

LEMMA 4.1. *Under the conditions of Theorem 2.3,  $\widehat{M}_n(z)$  forms a tight sequence and converges weakly to a two-dimensional Gaussian process  $M(\cdot)$  satisfying for  $z \in \mathcal{C} \cup \overline{\mathcal{C}}$  with  $\overline{\mathcal{C}} = \{\bar{z} : z \in \mathcal{C}\}$ ,*

$$(4.15) \quad EM(z) = \frac{y \int \frac{\underline{m}^3(z)t^2}{(\underline{m}(z)t+1)^3} dH(t)}{(1 - y \int \frac{\underline{m}^2(z)t^2}{(\underline{m}(z)t+1)^2} dH(t))^2} + \frac{\underline{m}^3(z)a(z)}{1 - y \int \frac{\underline{m}^2(z)t^2}{(\underline{m}(z)t+1)^2} dH(t)},$$

and for  $z_1, z_2 \in \mathcal{C} \cup \overline{\mathcal{C}}$ ,

$$(4.16) \quad \begin{aligned} \text{Cov}(M(z_1), M(z_2)) &= \frac{2\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_2) - \underline{m}(z_1))^2} - \frac{2}{(z_1 - z_2)^2} \\ &\quad + \frac{\partial^2}{\partial z_2 \partial z_1} \underline{m}(z_1)\underline{m}(z_2) d_2(z_1, z_2). \end{aligned}$$

Establishing the above lemma will straightforwardly lead to the CLT for LSS. Indeed, analogously to the arguments presented on pages 562–563 in Bai and Silverstein [3], one can establish that with probability 1, for  $f \in f_1, \dots, f_\kappa$ ,

$$\left| \oint f(z)(M_n(z) - \widehat{M}_n(z)) dz \right| \rightarrow 0.$$

Here, the complex integral is over  $z \in \mathcal{C} \cup \overline{\mathcal{C}}$ .

To accomplish the proof of the above lemma, we rewrite the expression for  $z \in \mathcal{C}_n$  as follows:

$$(4.17) \quad M_n(z) = p[m_n(z) - Em_n(z)] + p[Em_n(z) - m_n^0(z)] \triangleq M_{n1}(z) + M_{n2}(z).$$

Subsequently, we will focus on investigating the limiting properties of the resulting random processes  $M_{n1}(z)$  and  $M_{n2}(z)$ . We will establish the veracity of the following three facts.

Fact 1: The process  $M_{n1}(z)$  for  $z \in \mathcal{C}_n$  converges in distribution to a zero mean Gaussian process whose limiting covariance function agrees with (4.16) in Lemma 4.1. The main strategy for establishing this fact involves utilizing martingale difference decomposition and applying the CLT for martingales.

Fact 2: The deterministic process  $M_{n2}(z)$  for  $z \in \mathcal{C}_n$  converges uniformly to (4.15).

Fact 3: The process  $M_{n1}(z)$  for  $z \in \mathcal{C}_n$  is tight.

The detailed proof of above facts will be deferred to the supplementary material.

4.4. *Proof of Theorem 3.1.* By Corollary 2.1, we have that under  $H_0$ ,

$$\begin{aligned} \log \det \mathbf{T}_0^{-1} \hat{\mathbf{S}}_n - (n(y_n - 1) \log(1 - y_n) - p) &\xrightarrow{D} N(\varsigma_L, \vartheta_L), \\ \begin{pmatrix} \text{tr}(\mathbf{T}_0^{-1} \hat{\mathbf{S}}_n)^2 - (p + py_n) \\ \text{tr} \mathbf{T}_0^{-1} \hat{\mathbf{S}}_n - p \end{pmatrix} &\xrightarrow{D} N\left(\begin{pmatrix} \varsigma_2 \\ \varsigma_1 \end{pmatrix}, \begin{pmatrix} \vartheta_2 & \vartheta_{12} \\ \vartheta_{12} & \vartheta_2 \end{pmatrix}\right), \end{aligned}$$

where by the residual theorem

$$\begin{aligned} \varsigma_L &= \lim_{r \rightarrow 1^+} \frac{1}{2\pi i} \oint_{|\xi|=1} \log(|1 + \sqrt{y}\xi|^2) \left( \frac{\xi}{\xi^2 - r^{-2}} - \frac{1}{\xi} \right) d\xi \\ &\quad + \frac{a_{\mathbb{P}, \Sigma}}{2\pi i} \oint_{|\xi|=1} \frac{\log(|1 + \sqrt{y}\xi|^2)}{\xi^3} d\xi \\ &= \frac{\log(1 - y)}{2} - \frac{ya_{\mathbb{P}, \Sigma}}{2}, \end{aligned}$$

$$\varsigma_2 = \lim_{r \rightarrow 1^+} \frac{1}{2\pi i} \oint_{|\xi|=1} |1 + \sqrt{y}\xi|^4 \left( \frac{\xi}{\xi^2 - r^{-2}} - \frac{1}{\xi} \right) d\xi + \frac{a_{\mathbb{P}, \Sigma}}{2\pi i} \oint_{|\xi|=1} \frac{|1 + \sqrt{y}\xi|^4}{\xi^3} d\xi$$

$$= y + a_{\mathbb{P}, \Sigma} y,$$

$$\varsigma_1 = \lim_{r \rightarrow 1^+} \frac{1}{2\pi i} \oint_{|\xi|=1} |1 + \sqrt{y}\xi|^2 \left( \frac{\xi}{\xi^2 - r^{-2}} - \frac{1}{\xi} \right) d\xi + \frac{a_{\mathbb{P}, \Sigma}}{2\pi i} \oint_{|\xi|=1} \frac{|1 + \sqrt{y}\xi|^2}{\xi^3} d\xi = 0,$$

$$\vartheta_L = \lim_{r \rightarrow 1^+} \frac{-1}{2\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \frac{\log(|1 + \sqrt{y}\xi_1|^2) \log(|1 + \sqrt{y}\xi_2|^2)}{(\xi_1 - r\xi_2)^2} d\xi_2 d\xi_1$$

$$- \frac{a_{\mathbb{P}, \Sigma}}{4\pi^2} \oint_{|\xi_1|=1} \frac{\log(|1 + \sqrt{y}\xi_1|^2)}{\xi_1^2} d\xi_1 \oint_{|\xi_2|=1} \frac{\log(|1 + \sqrt{y}\xi_2|^2)}{\xi_2^2} d\xi_2$$

$$= -2\log(1 - y) + ya_{\mathbb{P}, \Sigma},$$

$$\vartheta_2 = \lim_{r \rightarrow 1^+} \frac{-1}{2\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \frac{|1 + \sqrt{y}\xi_1|^4 |1 + \sqrt{y}\xi_2|^4}{(\xi_1 - r\xi_2)^2} d\xi_2 d\xi_1$$

$$- \frac{a_{\mathbb{P}, \Sigma}}{4\pi^2} \oint_{|\xi_1|=1} \frac{|1 + \sqrt{y}\xi_1|^4}{\xi_1^2} d\xi_1 \oint_{|\xi_2|=1} \frac{|1 + \sqrt{y}\xi_2|^4}{\xi_2^2} d\xi_2$$

$$= 4y^2 + 4y(1 + y)^2(a_{\mathbb{P}, \Sigma} + 2),$$

$$\vartheta_1 = \lim_{r \rightarrow 1^+} \frac{-1}{2\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \frac{|1 + \sqrt{y}\xi_1|^2 |1 + \sqrt{y}\xi_2|^2}{(\xi_1 - r\xi_2)^2} d\xi_2 d\xi_1$$

$$- \frac{a_{\mathbb{P}, \Sigma}}{4\pi^2} \oint_{|\xi_1|=1} \frac{|1 + \sqrt{y}\xi_1|^2}{\xi_1^2} d\xi_1 \oint_{|\xi_2|=1} \frac{|1 + \sqrt{y}\xi_2|^2}{\xi_2^2} d\xi_2 = y(a_{\mathbb{P}, \Sigma} + 2),$$

$$\vartheta_{12} = \lim_{r \rightarrow 1^+} \frac{-1}{2\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \frac{|1 + \sqrt{y}\xi_1|^4 |1 + \sqrt{y}\xi_2|^2}{(\xi_1 - r\xi_2)^2} d\xi_2 d\xi_1$$

$$- \frac{a_{\mathbb{P}, \Sigma}}{4\pi^2} \oint_{|\xi_1|=1} \frac{|1 + \sqrt{y}\xi_1|^4}{\xi_1^2} d\xi_1 \oint_{|\xi_2|=1} \frac{|1 + \sqrt{y}\xi_2|^2}{\xi_2^2} d\xi_2 = 2y(1 + y)(a_{\mathbb{P}, \Sigma} + 2).$$

Then the results of this theorem can be obtained through elementary calculations.

## 5. Some auxiliary lemmas.

LEMMA 5.1 (Proposition 2.6.1 in Vershynin [17]). *Let  $x_1, \dots, x_n$  be independent, mean zero, sub-Gaussian random variables. Then  $\sum_{i=1}^n x_i$  is also a sub-Gaussian random variable, and*

$$\left\| \sum_{i=1}^n x_i \right\|_g \leq C \sum_{i=1}^n \|x_i\|_g,$$

where  $C$  is an absolute constant.

LEMMA 5.2 (Equation (2.14) in Vershynin [17]). *Let  $x$  be a mean zero sub-Gaussian random variable with sub-Gaussian norm  $\|x\|_g$ . Then*

$$\mathbb{P}\{|x| \geq t\} \leq 2 \exp(-ct^2/\|x\|_g^2) \quad \text{for all } t \geq 0.$$

LEMMA 5.3 (Burkholder’s inequality). *Let  $\{\mathbf{X}_k\}$  be a complex martingale difference sequence with respect to the increasing  $\sigma$ -field. Then, for  $p > 1$ ,*

$$\mathbb{E} \left| \sum_k \mathbf{X}_k \right|^p \leq K_p \mathbb{E} \left( \sum_k |\mathbf{X}_k|^2 \right)^{p/2}.$$

**Acknowledgments.** We extend our sincere appreciation to the Editor, Associate Editor, and the anonymous reviewers for their insightful feedback and valuable comments. Their constructive contributions have significantly enhanced the quality and clarity of this paper. We are truly grateful for their time and efforts in reviewing our work.

The authors are listed in alphabetical order.

Yanqing Yin serves as the corresponding author.

**Funding.** Yanqing Yin’s research was partially supported by grant NSFC 12271065.

Guangming Pan’s research was partially supported by a Tier 2 grant -MOE-T2EP20123-0007 and a grant RG76/21 at Nanyang Technological University.

Wang Zhou’s research was partially supported by a grant A-8001450-00-00 at the National University of Singapore.

## SUPPLEMENTARY MATERIAL

**Supplementary material for “Spectral analysis of gram matrices with missing at random observations: Convergence, central limit theorems, and applications in statistical inference”** (DOI: [10.1214/24-AOS2392SUPP](https://doi.org/10.1214/24-AOS2392SUPP); .pdf). This supplementary material contains the proof of Lemma 2.1 and Lemma 2.2 as well as detailed proof of Theorem 2.3.

## REFERENCES

- [1] BAI, Z. and ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. [MR2411613](#)
- [2] BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. [MR1617051](#) <https://doi.org/10.1214/aop/1022855421>
- [3] BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. [MR2040792](#) <https://doi.org/10.1214/aop/1078415845>
- [4] CAI, T. T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. [MR3164870](#) <https://doi.org/10.1111/rssb.12034>
- [5] CAI, T. T. and ZHANG, L. (2019). High dimensional linear discriminant analysis: Optimality, adaptive algorithm and missing data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 675–705. [MR3997097](#)
- [6] CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. [MR2604697](#) <https://doi.org/10.1214/09-AOS716>
- [7] CHEN, Y., FAN, J., MA, C. and YAN, Y. (2021). Bridging convex and nonconvex optimization in robust PCA: Noise, outliers and missing data. *Ann. Statist.* **49** 2948–2971. [MR4338899](#) <https://doi.org/10.1214/21-aos2066>
- [8] EL KAROUI, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19** 2362–2405. [MR2588248](#) <https://doi.org/10.1214/08-AAP548>
- [9] FAN, J., GUO, J. and ZHENG, S. (2022). Estimating number of factors by adjusted eigenvalues thresholding. *J. Amer. Statist. Assoc.* **117** 852–861. [MR4436317](#) <https://doi.org/10.1080/01621459.2020.1825448>
- [10] HU, J., LI, W., LIU, Z. and ZHOU, W. (2019). High-dimensional covariance matrices in elliptical distributions with application to spherical test. *Ann. Statist.* **47** 527–555. [MR3909941](#) <https://doi.org/10.1214/18-AOS1699>
- [11] JURČZAK, K. and ROHDE, A. (2017). Spectral analysis of high-dimensional sample covariance matrices with missing observations. *Bernoulli* **23** 2466–2532. [MR3648036](#) <https://doi.org/10.3150/16-BEJ815>
- [12] LI, H., PAN, G., YIN, Y. and ZHOU, W. (2024). Supplement to “Spectral analysis of gram matrices with missing at random observations: Convergence, central limit theorems, and applications in statistical inference.” <https://doi.org/10.1214/24-AOS2392SUPP>

- [13] LOUNICI, K. (2014). High-dimensional covariance matrix estimation with missing observations. *Bernoulli* **20** 1029–1058. [MR3217437](#) <https://doi.org/10.3150/12-BEJ487>
- [14] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR, Sb.* **1** 457–483.
- [15] PAN, G. M. and ZHOU, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270. [MR2418244](#) <https://doi.org/10.1214/07-AAP477>
- [16] SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. [MR1370408](#) <https://doi.org/10.1006/jmva.1995.1083>
- [17] VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. [MR3837109](#) <https://doi.org/10.1017/9781108231596>
- [18] WACHTER, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* **6** 1–18. [MR0467894](#) <https://doi.org/10.1214/aop/1176995607>
- [19] WANG, L., PENG, B. and LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist. Assoc.* **110** 1658–1669. [MR3449062](#) <https://doi.org/10.1080/01621459.2014.988215>