Statistica Sinica Preprint No: SS-2026-0076	
Title	Phase Transition of the Image Covariance Matrix and Inference for 2D-principal Components
Manuscript ID	SS-2026-0076
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202026.0076
Complete List of Authors	Shuhang Liu, Li Yang and Dandan Jiang
Corresponding Authors	Dandan Jiang
E-mails	jiangdd@xjtu.edu.cn
Notice: Accepted author version.	

PHASE TRANSITION OF THE IMAGE COVARIANCE MATRIX AND INFERENCE FOR *2D*-PRINCIPAL COMPONENTS

Shuhang Liu, Li Yang and Dandan Jiang*

School of Mathematics and Statistics

Xi'an Jiaotong University

Abstract: Image data exhibit local correlations and high-dimensional structure, with their covariance matrix capturing spatial dependencies and governing performance in feature extraction, dimensionality reduction, and reconstruction. Two-dimensional principal component analysis (*2D*-PCA) effectively preserves row/column-wise correlations, enabling accurate covariance estimation and identification of meaningful components. However, the limiting spectral behavior and phase transition phenomena of image covariance matrices in high-dimensional settings remain theoretically unexplored, compromising inference tasks such as principal component determination. To bridge this gap, we establish a unified theoretical framework for the asymptotic spectral behavior of the image covariance matrix under general non-i.i.d. assumptions, systematically characterizing its phase transition from non-spiked to spiked regimes and deriving corresponding central limit theorems for spectral statistics. Building on this theoretical

foundation, we propose a rigorous sequential hypothesis testing procedure to determine the number of principal components, offering provably consistent and prior-free detection thresholds that ensure statistical power in high dimensions. Our method is validated through extensive simulations and on real-world facial image datasets, demonstrating superior accuracy and robustness over existing approaches. The resulting framework provides a plug-and-play spectral module for diverse image processing and machine learning applications.

Key words and phrases: Random Matrix Theory, *2D*-PCA, Image Covariance Matrix.

1. Introduction

Image data possess rich spatial and textural information but involve high dimensionality, non-stationarity, and strong spatial correlations. Addressing these dependencies is crucial for tasks ranging from feature extraction (Zhao et al., 2021) and spatial modeling of textural patterns (Ding et al., 2022) to inverse problems like image denoising (Wang et al., 2021). The efficacy of these methods relies on capturing the spectral phase transition of the covariance structure, characterized by the emergence of informative "spiked" eigenvalues from the "bulk" noise spectrum (Xu et al., 2022). This transition defines the signal-to-noise boundary, intrinsic dimensionality, and statistical identifiability of structural patterns. Consequently, a rigorous

understanding of these limiting spectral behaviors is essential for developing robust statistical methodologies. Two-dimensional principal component analysis (*2D*-PCA), proposed by Yang et al. (2004), generalizes classical PCA to accommodate matrix-form data $\mathbf{W}_i \in \mathbb{R}^{h\times p}$ $(i = 1, \ldots, n)$. Unlike vectorization approaches, *2D*-PCA preserves the intrinsic matrix structure by constructing the image covariance matrix as

$$\mathbf{G}_t = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{W}_i - \overline{\mathbf{W}})^\top(\mathbf{W}_i - \overline{\mathbf{W}}), \tag{1.1}$$

where $\overline{\mathbf{W}} = 1/n\sum_{i=1}^{n}\mathbf{W}_i$ denotes the sample mean of the image matrices.

Since its inception, the *2D*-PCA framework has evolved to address diverse image analysis challenges. For instance, robust variants utilize joint-norm metrics to mitigate outlier sensitivity (Zhang et al., 2023), while methods like MURE-*2D*PCA employ data-driven low-rank approximations for inverse problems (Wang et al., 2021). Despite these practical advancements, the underlying spectral properties and phase transition phenomena remain largely underexplored. This limitation matters because the separation of eigenvalues governs the intrinsic robustness and identifiability of projection subspaces. Furthermore, determining the number of principal components is critical for downstream efficacy. In face recognition, this choice regulates the trade-off between within-class compactness and between-class separability (Zhao et al., 2021). In image denoising and background modeling, it

delineates the boundary between signal and noise or establishes the background rank against dynamic variations (Ades-Aron et al., 2021). Although recent studies such as sign-flip parallel analysis (SPA) address this issue (Li and Kuang, 2024), current methods remain largely heuristic. Consequently, a unified analytic spectral foundation is needed to quantify non-independent eigenstructures, advancing both inference accuracy and efficiency.

To address this fundamental need, we aim to establish a rigorous theoretical foundation for the spectral analysis of high-dimensional image data exhibiting complex dependencies. A direct application involves determining the number of principal components, a problem typically addressed via data-driven heuristics or Bayesian frameworks. Within data-driven methods, pioneering tools such as the Kaiser criterion and scree plots offer intuitive and computationally simple rules for rank selection. Subsequent refinements have included cross-validation techniques (Jin et al., 2021) and eigenvalue stabilization procedures (Ledoit and Wolf, 2020), which improve robustness in practice. More recent advances, such as the SPA method introduced by Li and Kuang (2024), provide a resampling-based calibration that adapts to empirical spectral variability. However, these methods often rely on subjective or heuristic rules, which can lead to limited robustness in high-dimensional settings and sensitivity to real-world variations in

imaging conditions, resulting in unstable estimations. Within the Bayesian framework, methods range from classical evidence maximization (Minka, 2000) to contemporary empirical Bayes models like SuSiE PCA (Yuan and Mancuso, 2023), which offer principled probabilistic criteria for rank determination by placing prior structures over the eigencomponents. While these approaches bring theoretical coherence, they often require vectorization of image data and assume isotropic noise, thereby overlooking the inter-matrix dependencies central to *2D* image analysis.

To overcome the limitations of the aforementioned works, we utilize random matrix theory (RMT) to characterize the underlying spectral phase transition of the image covariance matrix under complex dependencies. While RMT results are well-established for covariance matrices of random vectors (Bai and Yao, 2012; Bao et al., 2022; Jiang, 2023), the analogous theory for sequences of image matrices remains unexplored. The primary difficulty arises because the image covariance matrix inherently violates the independent and identically distributed (i.i.d.) data assumption, thereby invalidating classical limit spectral results such as the Marčenko-Pastur (M-P) law and existing central limit theorems (CLTs) (Bai and Silverstein, 2008) for spectral statistics. Furthermore, extending spectral theory from random vectors to image data introduces additional structural complexity.

At present, no existing framework fully characterizes the limiting spectral distribution (LSD) or the asymptotic behavior of spectral statistics for such image-based covariance matrices.

To render these inter-matrix dependencies mathematically tractable under an RMT framework, our analysis explicitly focuses on an adjusted formulation, denoted by $\mathbf{G} = n\{(n-1)h\}^{-1}\mathbf{G}_t$, where $\mathbf{G}_t$ is defined in (1.1). In response to this methodological imperative, this paper undertakes the following objectives: we first establish the asymptotic spectral behavior of the image covariance matrix $\mathbf{G}$ under two typical non-i.i.d. dependence structures arising in practice, as detailed in Section 2.1. Second, we characterize the phase transition from non-spiked to spiked regimes and derive the CLT for linear spectral statistics (LSS). Finally, we develop a sequential hypothesis testing procedure to determine the number of significant *2D*-principal components, ensuring statistical validity in high-dimensional settings. The main contributions of this work are summarized as follows:

1. We propose a typical but prevalent non-i.i.d. framework for high-dimensional image data that captures complex spatial dependencies, including non-stationarity, long-range dependence, and non-separable row–column interactions, thereby relaxing the i.i.d. assumptions.

2. We establish the CLT for the spectral statistics of the image covari-

ance matrix in the bulk phase under representative non-i.i.d. assumptions and characterize the phase transition from non-spiked to spiked regimes within a baseline model. This analysis fills a theoretical gap in the asymptotics of image sequences and captures the emergence of low-dimensional principal components, thereby revealing the dynamic spectral evolution of complex image transformations.

3. Building on our spectral theory, we develop a rigorously justified sequential testing strategy with theoretically derived prior-free and non-heuristic thresholds for determining the number of principal components. This method guarantees high-dimensional consistency and statistical detectability while enhancing robustness to non-separability, thereby reducing manual tuning.

This paper is organized as follows. Section 2 establishes the CLT for the LSS of $\mathbf{G}$ with non-i.i.d. structure in the bulk phase. Section 3 characterizes the phase transition for spikes of $\mathbf{G}$. Building on these theoretical foundations, Section 4 develops a sequential testing strategy for determining the number of *2D*-principal components. Sections 5 and 6 validate our findings through simulation studies and facial image applications. Finally, Section 7 concludes, with proofs in the Supplementary Materials.

2. Bulk Phase: CLT under non-i.i.d. structure

2.1 Model Formulation

For clarity of exposition, we express the image sample matrix $\mathbf{W}_i$ in terms of its row vectors $\boldsymbol{w}_{ji} \in \mathbb{R}^{1\times p}$, $j \in [1:h]$, as $\mathbf{W}_i = (\boldsymbol{w}_{1i}^\top, \boldsymbol{w}_{2i}^\top, \cdots, \boldsymbol{w}_{hi}^\top)_{h\times p}^\top$, $i \in [1:n]$. Here, the notation $[a:b]$ represents the set of all integers between a and b. Accordingly, the image covariance matrix $\mathbf{G}$, also referred to as the row-row covariance matrix, is formed of the row vector expression,

$$\mathbf{G} = \frac{1}{(n-1)h}\sum_{i=1}^{n}\sum_{j=1}^{h}(\boldsymbol{w}_{ji} - \bar{\boldsymbol{w}}_j)^\top(\boldsymbol{w}_{ji} - \bar{\boldsymbol{w}}_j), \tag{2.2}$$

where $\bar{\boldsymbol{w}}_j = 1/n\sum_{i=1}^{n}\boldsymbol{w}_{ji}$. Exploiting this representation, the jth row $\boldsymbol{w}_{ji}$ of each $\mathbf{W}_i$ is extracted to the sample matrix as $\check{\mathbf{W}}_j = (\boldsymbol{w}_{j1}^\top, \boldsymbol{w}_{j2}^\top, \cdots, \boldsymbol{w}_{jn}^\top)_{p\times n}$ for all $j \in [1:h]$, derived from the jth row-population.

Let the row-population sample covariance matrix $\mathbf{S}_j \in \mathbb{R}^{p\times p}$ of $\check{\mathbf{W}}_j$ be noted as $\mathbf{S}_j = 1/(n-1)\sum_{i=1}^{n}(\boldsymbol{w}_{ji} - \bar{\boldsymbol{w}}_j)^\top(\boldsymbol{w}_{ji} - \bar{\boldsymbol{w}}_j)$ for $j \in [1:h]$. In the high-dimensional asymptotic regime where $n, p \to \infty$ with $p/n \to c \in (0, +\infty)$, provided that for each j, the empirical spectral distribution (ESD) of $\mathbf{S}_j$ converges uniformly to its respective compactly supported LSD, the order of summation in (2.2) can be interchanged. Thus, $\mathbf{G}$ can be expressed as the average of the h row-population sample covariance matrices $\mathbf{S}_j$, i.e., $\mathbf{G} = 1/h\sum_{j=1}^{h}\mathbf{S}_j$.

To accommodate complex phenomena in real-world images, such as heteroscedastic image noise, spatial and channel variance, and correlation (Guo et al., 2019; Zhou et al., 2020), we allow the image covariance matrix to adopt a multi-population non-i.i.d. structure, thereby preserving intrinsic topological row correlations and lateral spatial information. To empirically corroborate this, additional facial image experiments in the Supplementary Materials illustrate the multi-population heterogeneity of $\mathbf{G}$.

Therefore, we consider $\mathbf{G}$ corresponding to τ row-population covariance matrices $\mathbf{\Sigma}^{(r)} = \mathbf{T}_r\mathbf{T}_r^\top$ for $r \in [1:\tau]$, where $\mathbf{T}_r$ is a $p \times p$ deterministic matrix. With the sample matrices $\check{\mathbf{W}}_j$ drawn from $\mathbf{\Sigma}^{(r)}$, the rth row-population sample matrix $\mathbf{W}_j^{(r)}$ is rewritten as:

$$\mathbf{W}_j^{(r)} \coloneqq \left(\boldsymbol{w}_{ji}^{(r)\top}\right)_{p\times n} = \mathbf{T}_r\mathbf{X}_j^{(r)} = \left(\mathbf{T}_r\boldsymbol{x}_{j1}^{(r)}, \ldots, \mathbf{T}_r\boldsymbol{x}_{jn}^{(r)}\right), \quad j \in [1:h_r], \tag{2.3}$$

where the entries $\{x_{jki}^{(r)}\}$ of $\boldsymbol{x}_{ji}^{(r)}$ are i.i.d. random variables with zero mean and unit variance for $k \in [1:p]$. The total number of samples h over all populations satisfies $\sum_{r=1}^{\tau} h_r = h$, where τ and h_r's are fixed integers independent of n and p. Natural images exhibit nonlocal self-similarity and redundancy, enabling patch group modeling via a unidirectional low-rank structure. The finite h assumption reflects the intrinsic spatial constraints on patch grouping, allowing the model to accurately capture the correlations inherent in the data (Chang et al., 2017). We utilize the sample covariance

$\mathbf{S}_j^{(r)} \coloneqq n^{-1}\mathbf{T}_r\mathbf{X}_j^{(r)}(\mathbf{X}_j^{(r)})^\top\mathbf{T}_r^\top$. For centered data, the factor n^{-1} is replaced by $(n-1)^{-1}$ to account for the loss of one degree of freedom, following the substitution principle (Zheng et al., 2015). Thus, $\mathbf{G}$ is represented as:

$$\mathbf{G} = \frac{1}{h}\sum_{r=1}^{\tau}\sum_{j=1}^{h_r}\mathbf{S}_j^{(r)}. \tag{2.4}$$

To characterize the non-i.i.d. structure of $\mathbf{G}$, we study the spectral properties of $\mathbf{S}_j^{(r)}$ and $\mathbf{S}_i^{(u)}$ through the following two pairwise cases:

Case 1: *Independent but not identically distributed.* For any $r \neq u$ within the range $[1:\tau]$, the populations are independent, implying sample independence. Consequently, $\mathbf{T}_r\mathbf{X}_j^{(r)}$ and $\mathbf{T}_u\mathbf{X}_i^{(u)}$ are independent for all $j \in [1:h_r]$ and $i \in [1:h_u]$;

Case 2: *Dependent but identically distributed.* For all $r \in [1:\tau]$, the samples within the same population share the same marginal distribution but are permitted to be dependent. Thus, $\mathbf{T}_r\mathbf{X}_j^{(r)}$ and $\mathbf{T}_r\mathbf{X}_i^{(r)}$ may be dependent for any $i \neq j \in [1:h_r]$. The dependence can be represented by the shared-component decomposition $\boldsymbol{x}_{jt}^{(r)} = \boldsymbol{x}_{0t}^{(r)} + \boldsymbol{\varepsilon}_{jt}^{(r)}$, and $\boldsymbol{x}_{it}^{(r)} = \boldsymbol{x}_{0t}^{(r)} + \boldsymbol{\varepsilon}_{it}^{(r)}$. Here $\boldsymbol{x}_{0t}^{(r)}$ is the latent component shared by samples from the same row-population, whereas $\boldsymbol{\varepsilon}_{jt}^{(r)}$ and $\boldsymbol{\varepsilon}_{it}^{(r)}$ are idiosyncratic components.

The dependence structure in Case 2 naturally accommodates the spatial characteristics of image data. By decomposing row features into a shared latent component and idiosyncratic residuals, it simultaneously captures

within-population cross-covariance and marginal variability. This aligns with established image models based on structural similarity, texture statistics, and nonlocal patch correlation (Ding et al., 2022; Zha et al., 2020). The formal assumptions for this structure are detailed below.

Assumption 1. For each $r \in [1:\tau]$ and $j \in [1:h_r]$, the entries $\{x_{jkt}^{(r)} : k \in [1:p], t \in [1:n]\}$ of $\mathbf{X}_j^{(r)}$ are i.i.d. with $\mathbb{E}(x_{j11}^{(r)}) = 0$, $\mathbb{E}|x_{j11}^{(r)}|^2 = 1$, and $\sup_n \mathbb{E}|x_{j11}^{(r)}|^4 < \infty$. For complex-valued variables, assume $\mathbb{E}\{(x_{j11}^{(r)})^2\} = 0$.

Assumption 2. For each $r \in [1:\tau]$, all $\ell, m \in [1:h_r]$, and all $t \in [1:n]$, assume $\mathbb{E}\boldsymbol{x}_{0t}^{(r)} = \mathbb{E}\boldsymbol{\varepsilon}_{\ell t}^{(r)} = \mathbf{0}$, $\mathbb{E}\boldsymbol{x}_{0t}^{(r)}(\boldsymbol{x}_{0t}^{(r)})^\top = \rho_r \mathbf{I}_p$, $\mathbb{E}\boldsymbol{\varepsilon}_{\ell t}^{(r)}(\boldsymbol{\varepsilon}_{mt}^{(r)})^\top = (1-\rho_r)\mathbf{I}_p \mathbf{1}_{\{\ell=m\}}$, $\mathbb{E}\boldsymbol{x}_{0t}^{(r)}(\boldsymbol{\varepsilon}_{\ell t}^{(r)})^\top = \mathbf{0}$, where $\rho_r \in [0,1]$, and $\{(x_{0kt}^{(r)}, \varepsilon_{1kt}^{(r)}, \ldots, \varepsilon_{h_r kt}^{(r)}) : k \in [1:p],\ t \in [1:n]\}$ are independent across the pairs (k,t).

2.2 CLT for Linear Spectral Statistic of G under non-i.i.d. cases

In statistical applications, the analytical objective of the *2D*-PCA framework in image processing is the trace statistic $\mathrm{tr}(\mathbf{G})$, which quantifies overall variability via the total scatter of the image covariance matrix $\mathbf{G}$ over the whole feature space (Yang et al., 2004). By letting $F_n^{\mathbf{G}}$ be the ESD of $\mathbf{G}$, this trace objective is mathematically formalized as the linear spectral statistic corresponding to the linear test function $f_0(x) = x$, given by $\int f_0(x)\mathrm{d}F_n^{\mathbf{G}} = 1/p \sum_{i=1}^p \lambda_i(\mathbf{G})$, where $\lambda_i(\mathbf{G})$ denotes the ith real eigenvalue

of the $p \times p$ matrix $\mathbf{G}$. By (2.4) and the linearity of the trace, the statistic admits the following decomposition:

$$\frac{1}{p}\sum_{i=1}^{p} h\lambda_i(\mathbf{G}) = \frac{1}{p}\operatorname{tr}\{h\mathbf{G}\} = \sum_{r=1}^{\tau}\sum_{j=1}^{h_r}\frac{1}{p}\operatorname{tr}\{\mathbf{S}_j^{(r)}\} = \sum_{r=1}^{\tau}\sum_{j=1}^{h_r}\frac{1}{p}\sum_{i=1}^{p}\lambda_i(\mathbf{S}_j^{(r)}). \quad (2.5)$$

To establish the CLT for this specific linear spectral statistic of $\mathbf{G}$ under non-i.i.d. cases, we analyze the empirical spectral processes of its constituent matrices $\mathbf{S}_j^{(r)}$. Specifically, under $f_0(x) = x$, this process is given by

$$\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f_0)\}_j = p\int x(F_n^{\mathbf{S}_j^{(r)}} - F^{c,H^{(r)}})(\mathrm{d}x), \text{ for each } j \in [1:h_r],\ r \in [1:\tau], \quad (2.6)$$

where $F_n^{\mathbf{S}_j^{(r)}}(x)$ is the ESD of the covariance matrix $\mathbf{S}_j^{(r)}$, and the LSD $F^{c,H^{(r)}}$ of sample covariance matrices from the same population is consistent. In accordance with (2.5) and (2.6), the empirical process $\boldsymbol{\mathcal{G}}_n(f_0)$ for $h\mathbf{G}$ can be exactly expressed as a linear combination of $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f_0)\}_j$. Thus,

$$\begin{aligned}\boldsymbol{\mathcal{G}}_n(f_0) &= \sum_{r=1}^{\tau}\sum_{j=1}^{h_r}\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f_0)\}_j = h\sum_{i=1}^{p}\lambda_i(\mathbf{G}) - p\sum_{r=1}^{\tau}h_r\int xF^{c,H^{(r)}}(\mathrm{d}x)\\ &= p\int x\big(hF_n^{\mathbf{G}} - \sum_{r=1}^{\tau}h_rF^{c,H^{(r)}}\big)(\mathrm{d}x). \quad (2.7)\end{aligned}$$

Recently, Jiang (2023) established the CLT for the LSS of a single general sample covariance matrix, showing that for any $f \in \mathcal{A}$, the class of analytic functions defined on an open subset of the complex plane, the corresponding empirical process $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j$ converges weakly to a Gaussian

limit with mean $\mu_f^{(r)}$ and variance $\nu_f^{(r)}$, given by

$$\mu_f^{(r)} = -\frac{q}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{c\int \underline{s}^r(z)^3x^2\{1+x\underline{s}^r(z)\}^{-3}\mathrm{d}H^{(r)}(x)}{\left[1-c\int \underline{s}^r(z)^2x^2\{1+x\underline{s}^r(z)\}^{-2}\mathrm{d}H^{(r)}(x)\right]^2}(\mathrm{d}z) \quad (2.8)$$
$$-\frac{\beta^{(r)}c}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{\underline{s}^r(z)^3\int x\{1+x\underline{s}^r(z)\}^{-1}\mathrm{d}H^{(r)}(x)\int\{1+x\underline{s}^r(z)\}^{-2}\mathrm{d}H^{(r)}(x)}{1-c\int \underline{s}^r(z)^2x^2\{1+x\underline{s}^r(z)\}^{-2}\mathrm{d}H^{(r)}(x)}(\mathrm{d}z),$$

$$\nu_f^{(r)} = -\frac{q+1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2}\frac{f(z_1)f(z_2)}{\{\underline{s}^r(z_1)-\underline{s}^r(z_2)\}^2}\mathrm{d}\underline{s}^r(z_1)\mathrm{d}\underline{s}^r(z_2) \quad (2.9)$$
$$-\frac{\beta^{(r)}c}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)f(z_2)\int\frac{x\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^r(z_1)\}^2}\int\frac{x\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^r(z_2)\}^2}\mathrm{d}\underline{s}^r(z_1)\mathrm{d}\underline{s}^r(z_2).$$

Here $\beta^{(r)} = \mathbb{E}|x_{jki}^{(r)}|^4 - q - 2$, where $q = 1$ for the real case and $q = 0$ for the complex case. Let $s^r(z)$ be the stieltjes transform of $F^{c,H^{(r)}}$, and $\underline{s}^r \equiv s_{\underline{F}^{c,H^{(r)}}}$ be the stieltjes transform of $\underline{F}^{c,H^{(r)}} \equiv (1-c)I_{[0,\infty)} + cF^{c,H^{(r)}}$.

Building upon this fundamental result, analyzing the linear combination in (2.7) reduces to establishing the joint asymptotic distribution of the constituent processes $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j$ under non-i.i.d. cases, as summarized in Lemma 1, with case-specific derivations deferred to the Supplementary Materials. We begin by stating Assumption 3 for regularity and Assumption 4, an analogue to condition (5) in Bai et al. (2007), to facilitate covariance calculations under within-population dependence.

Assumption 3. For each $r \in [1:\tau]$, $\mathbf{T}_r \coloneqq (\boldsymbol{\Sigma}^{(r)})^{1/2}$ is deterministic with $\sup_p \|\mathbf{T}_r\| < \infty$, and the ESD $H_n^{(r)}$ of $\boldsymbol{\Sigma}^{(r)}$ converges weakly to a probability distribution $H^{(r)}$.

Assumption 4. For each $r \in [1:\tau]$, assume that $\max_{1\le k\le p}\sup_z \left|\mathbf{e}_k^\top\{\mathbf{I}_p + \underline{s}^{(r)}(z)\mathbf{\Sigma}^{(r)}\}^{-1}\mathbf{e}_k - \int\{1+\underline{s}^{(r)}(z)x\}^{-1}\,\mathrm{d}H_n^{(r)}(x)\right| \to 0$, where $\mathbf{e}_k$ denotes the kth standard basis vector for $k \in [1:p]$.

Lemma 1. *Suppose Assumptions 1–4 hold, $p/n = c_n \to c \in (0,\infty)$ and let $f, g \in \mathcal{A}$ be analytic. For any indices (r,j) and (u,i), $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j + \{\boldsymbol{\mathcal{Q}}_n^{(u)}(g)\}_i \Rightarrow \mathcal{N}(\mu_{f,g}^{(r,u)}, \nu_{f,g}^{(r,u)})$, where the symbol $\Rightarrow$ denotes convergence in distribution, and the mean and variance terms $\left(\mu_{f,g}^{(r,u)}, \nu_{f,g}^{(r,u)}\right)$ are finite and depend on the underlying non-i.i.d. case; specifically,*

$$\left(\mu_{f,g}^{(r,u)}, \nu_{f,g}^{(r,u)}\right) := \begin{cases} \left(\mu_f^{(r)} + \mu_g^{(u)}, \nu_f^{(r)} + \nu_g^{(u)}\right), & \text{for Case 1:} \quad r \neq u, \\ \left(2\mu_f^{(r)}, 2\nu_f^{(r)} + 2\nu_{\mathrm{Cov}\, f_{ji}}^{(r)}\right), & \text{for Case 2:} \quad r = u,\ f = g. \end{cases}$$

Here $\mu_f^{(u)}, \nu_f^{(u)}$ are the analogues of $\mu_f^{(r)}, \nu_f^{(r)}$ defined in (2.8) and (2.9). Moreover, the calculations of the covariance $\nu_{\mathrm{Cov}\, f_{ji}}^{(r)}$ fluctuate based on the specific correlation structure considered, where

$$\begin{aligned} \nu_{\mathrm{Cov}\, f_{ji}}^{(r)} &= \mathrm{Cov}\left(\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j, \{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_i\right) \\ &= -\frac{1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)f(z_2)\partial_{z_1}\partial_{z_2}\,\mathrm{Cov}\left(\Xi_{n,j}^{(r)}(z_1), \Xi_{n,i}^{(r)}(z_2)\right)\mathrm{d}z_1\mathrm{d}z_2, \end{aligned} \tag{2.10}$$

where $\Xi_{n,j}^{(r)}(z_1) = \mathrm{tr}(\mathbf{S}_j^{(r)} - z_1\mathbf{I})^{-1} - ps^r(z_1)$, and $\Xi_{n,i}^{(r)}(z_2)$ is defined analogously. By leveraging the correlation exhibited between sample matrices $\mathbf{X}_i^{(r)}$ and $\mathbf{X}_j^{(r)}$ in Case 2, the covariance can be calculated by

$$\nu_{\mathrm{Cov}\, f_{ji}}^{(r)} = \nu_{f,\rho}^{(r)} + o(1), \tag{2.11}$$

where the cross-covariance is given by

$$\nu_{f,\rho}^{(r)} = -\frac{q+1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)f(z_2)\frac{\partial^2}{\partial z_1 \partial z_2}\Big[-\log\big\{1-\rho_r^2 c\,\underline{s}^{(r)}(z_1)\underline{s}^{(r)}(z_2) \times \int \frac{x^2\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_1)\}\{1+x\underline{s}^{(r)}(z_2)\}}\big\}\Big]\,\mathrm{d}z_1\mathrm{d}z_2 \qquad (2.12)$$
$$-\frac{\beta_\rho^{(r)} c}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)f(z_2)\int \frac{x\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_1)\}^2}\int \frac{x\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_2)\}^2}\,\mathrm{d}\underline{s}^{(r)}(z_1)\,\mathrm{d}\underline{s}^{(r)}(z_2).$$

Here, $\beta_\rho^{(r)} = \mathbb{E}\{(x_{jkt}^{(r)})^2(x_{ikt}^{(r)})^2\} - 1 - (q+1)\rho_r^2$ *is assumed to be common across all* $k \in [1:p]$ *and* $t \in [1:n]$ *within the* r*th row-population.*

Remark 1. Even without Assumption 4, explicit formulas can be derived for special cases. Under diagonal loading, the first terms in (2.9) and (2.12) remain unchanged, while for finite $\beta^{(r)}$ and $\beta_\rho^{(r)}$, the contour integrals in the second terms are replaced by

$$\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)f(z_2)\int \frac{x^2\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_1)\}^2\{1+x\underline{s}^{(r)}(z_2)\}^2}\,\mathrm{d}\underline{s}^{(r)}(z_1)\,\mathrm{d}\underline{s}^{(r)}(z_2).$$

In the particular case of $f_0(x) = x$ corresponding to the statistic $\boldsymbol{\mathcal{G}}_n(f_0)$ in (2.7), Lemma 1 explicitly characterizes the limiting covariance structure among its constituent processes $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f_0)\}_j$. Specifically, covariance contributions from pairs with $r \neq u$ vanish under Case 1 due to independence, whereas pairs with $r = u$ and $i \neq j$ yield the within-population cross-covariance under Case 2. Combining these results, we establish the limiting distribution of $\boldsymbol{\mathcal{G}}_n(f_0)$ in the following theorem.

Theorem 1. *Suppose Assumptions 1–4 hold and* $p/n = c_n \to c \in (0, \infty)$. *In the bulk regime,* $\boldsymbol{\mathcal{G}}_n(f_0)$ *converges weakly to the Gaussian distribution with mean* $\mu_x^{\mathbf{G}}$ *and variance* $\nu_x^{\mathbf{G}}$, *where*

$$\mu_x^{\mathbf{G}} = \sum_{r=1}^{\tau} h_r \mu_{f_0}^{(r)}, \quad \nu_x^{\mathbf{G}} = \sum_{r=1}^{\tau} h_r \nu_{f_0}^{(r)} + \sum_{r=1}^{\tau} h_r (h_r - 1) \nu_{f_0,\rho}^{(r)}. \tag{2.13}$$

To facilitate practical applications, we define the standardized statistic:

$$T_x^{\mathbf{G}} = (\nu_x^{\mathbf{G}})^{-\frac{1}{2}} \Big[h \sum_{i=1}^{p} \lambda_i(\mathbf{G}) - p \sum_{r=1}^{\tau} h_r \int x F^{c,H^{(r)}}(\mathrm{d}x) - \mu_x^{\mathbf{G}} \Big] \Rightarrow \mathcal{N}(0,1). \tag{2.14}$$

To illustrate, Example 1 derives explicit central term formulas for Theorem 1 assuming a simple $H^{(r)}$ across various structures of $\mathbf{G}$.

Example 1. Let $H^{(r)}$ be specified for $\tau = 3$, $r \in [1 : \tau]$ as $H^{(1)} = \delta_{\{1\}}$, $H^{(2)} = \delta_{\{2\}}/2 + \delta_{\{1\}}/2$, and $H^{(3)} = \delta_{\{3\}}/3 + \delta_{\{2\}}/3 + \delta_{\{1\}}/3$. Under several structures (h_1, h_2, h_3) considered in Scenarios 1–3, with $h = \sum h_r$ and sample eigenvalue $\lambda_i(\mathbf{G})$, the statistic in (2.14) simplifies to

$$T_x = \nu_x^{-\frac{1}{2}} \Big[h \sum_{i=1}^{p} \lambda_i(\mathbf{G}) - b_x - \mu_x \Big] \Rightarrow \mathcal{N}(0,1), \tag{2.15}$$

where, for Scenario 1 with $(h_1, h_2, h_3) = (2, 1, 1)$, the central term is $b_x = 11p/2$, $\mu_x = 0$. In the diagonal case, $\nu_x^{\mathrm{diag}} = c\{(q+1)(55/6 + 2\rho_1^2) + 2\beta^{(1)} + (5/2)\beta^{(2)} + (14/3)\beta^{(3)} + 2\beta_\rho^{(1)}\}$, whereas under Assumption 4, $\nu_x^{\mathrm{A4}} = c\{(q+1)(55/6 + 2\rho_1^2) + 2\beta^{(1)} + (9/4)\beta^{(2)} + 4\beta^{(3)} + 2\beta_\rho^{(1)}\}$. Its proof and analogous calculations for Scenarios 2–3 with $(h_1, h_2, h_3) = (1, 2, 1)$ and $(1, 1, 2)$ are deferred to the Supplementary Materials.

3. Spiked Phase: Phase Transition for the Spikes of G

This section investigates the extreme eigenvalues of $\mathbf{G}$ and their separation from the limiting spectral bulk under the non-i.i.d. cases. Following the construction in (2.4), the spectral behavior of $\mathbf{G}$ is inherently determined by the heterogeneous covariance matrices $\{\mathbf{\Sigma}^{(r)}\}_{r=1}^{\tau}$, which are characterized by generalized spiked models (Bai and Yao, 2012). Specifically, for each $r \in [1:\tau]$, these covariance matrices share a common orthogonal eigenbasis $\mathbf{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)$ and take the form $\mathbf{\Sigma}^{(r)} = \mathbf{U}(\mathbf{D}_0 + \mathbf{\Lambda}_r)\mathbf{U}^\top$ under the homogeneous non-spiked baseline. Here, $\mathbf{D}_0 = \operatorname{diag}(d_1, \ldots, d_p)$ is the non-spiked diagonal matrix with its diagonal entries sorted in descending order, where the ESD $H_n^{\mathbf{\Sigma}_0}$ of $\mathbf{\Sigma}_0 \coloneqq \mathbf{U}\mathbf{D}_0\mathbf{U}^\top$ converges to the LSD $H^{\mathbf{\Sigma}_0}$. Additionally, $\mathbf{\Lambda}_r = \operatorname{diag}(\lambda_1^{(r)}, \ldots, \lambda_{\kappa_r}^{(r)}, 0, \ldots, 0)$ is the spiked diagonal matrix with its non-zero entries sorted in descending order, where κ_r denotes the number of spiked eigenvalues for the rth population. We define the joint support of the spiked components across all populations as $\mathcal{M} = \{1, \ldots, M\}$, where $M \coloneqq \max_{1 \le r \le \tau} \kappa_r$. With this notation, for each $r \in [1:\tau]$,

$$\mathbf{\Sigma}^{(r)} = \mathbf{U} \operatorname{diag}(\sigma_1^{(r)}, \ldots, \sigma_p^{(r)})\mathbf{U}^\top, \quad \text{where } \sigma_m^{(r)} = \begin{cases} d_m + \lambda_m^{(r)}, & m \in [1:M], \\ d_m, & m \in [M+1:p]. \end{cases}$$

We concatenate the sample matrices from (2.3) along the sample dimension, and denote the resulting matrix by $\breve{\mathbf{W}} = \left(\mathbf{W}_1^{(1)}, \ldots, \mathbf{W}_{h_1}^{(1)}, \ldots, \mathbf{W}_1^{(\tau)}, \ldots, \mathbf{W}_{h_\tau}^{(\tau)}\right) \in$

$\mathbb{R}^{p\times N}$ with $N = nh$. Accordingly, (2.4) admits the equivalent form $\mathbf{G} = N^{-1}\check{\mathbf{W}}\check{\mathbf{W}}^\top$. Due to the spectral invariance of $\mathbf{G}$ under orthogonal transformations, its extreme eigenvalues are identical to those of $N^{-1}\check{\mathbf{W}}\check{\mathbf{W}}^\top$, where $\check{\mathbf{W}} \coloneqq \mathbf{U}^\top\check{\mathbf{W}}$, such that Assumption 2 is naturally imposed on the rotated observations $\mathbf{U}^\top\boldsymbol{x}_{jt}^{(r)}$. Consequently, the joint spiked components and the non-spiked components in $\check{\mathbf{W}}$ can be decoupled into separate submatrices, and we partition $\check{\mathbf{W}}$ as follows:

$$\check{\mathbf{W}} = \begin{pmatrix}\mathbf{Y}_1\\ \mathbf{Y}_0\end{pmatrix} \in \mathbb{R}^{p\times N}, \quad \mathbf{Y}_1 = \begin{pmatrix}\boldsymbol{y}_1\\ \vdots\\ \boldsymbol{y}_M\end{pmatrix} \in \mathbb{R}^{M\times N}, \quad \mathbf{Y}_0 = \begin{pmatrix}\boldsymbol{y}_{M+1}\\ \vdots\\ \boldsymbol{y}_p\end{pmatrix} \in \mathbb{R}^{p_0\times N}.$$

Here, the submatrix $\mathbf{Y}_1$ corresponds to the index set $\mathcal{M}$, and $\mathbf{Y}_0$ contains the remaining $p_0 = p - M$ noise components, whose corresponding sample covariance matrices will be examined separately in our subsequent analysis.

We first define the non-spiked sample covariance matrix $\mathbf{Q}_0 \coloneqq N^{-1}\mathbf{Y}_0\mathbf{Y}_0^\top$, whose rightmost spectral edge determines the upper boundary of the bulk support. We then introduce its dual matrix $\widetilde{\mathbf{Q}}_0 \coloneqq N^{-1}\mathbf{Y}_0^\top\mathbf{Y}_0$, which has the identical nonzero spectrum as $\mathbf{Q}_0$. It follows that $\mathbb{E}(\widetilde{\mathbf{Q}}_0) = N^{-1}\sum_{m=M+1}^{p}\mathbb{E}(\boldsymbol{y}_m^\top\boldsymbol{y}_m)$, where each summand satisfies, for $m \in [M+1:p]$,

$$\mathbb{E}(\boldsymbol{y}_m^\top\boldsymbol{y}_m) = d_m\mathbf{B}_0 \text{ with } \mathbf{B}_0 \coloneqq \operatorname{diag}\left(\mathbf{R}_{h_1}(\rho_1)\otimes\mathbf{I}_n, \ldots, \mathbf{R}_{h_\tau}(\rho_\tau)\otimes\mathbf{I}_n\right). \quad (3.16)$$

Here, $\otimes$ denotes the Kronecker product, $\mathbf{R}_{h_r}(\rho_r) = (1-\rho_r)\mathbf{I}_{h_r} + \rho_r\mathbf{1}_{h_r}\mathbf{1}_{h_r}^\top$, and the matrix $\mathbf{B}_0 \in \mathbb{R}^{N\times N}$ captures the within-population sample depen-

dence structure. Moreover, we define the diagonal matrix $\mathbf{A}_0 = \mathrm{diag}(d_{M+1}, \ldots, d_p)$. Based on these formulations, the bulk edge is characterized in the following lemma, leveraging the results from Ding and Yang (2021).

Lemma 2 (Bulk edge)**.** *Under Assumptions 1–4, let* $c_N = p_0/N$, $F_{p_0}^{\mathbf{A}_0}$ *and* $F_N^{\mathbf{B}_0}$ *denote the ESDs of* $\mathbf{A}_0$ *and* $\mathbf{B}_0$*, respectively. For any* $z \in \mathbb{C}^+$*, let* $s_{1,N}(z)$ *and* $s_{2,N}(z)$ *be the unique solutions in* $\mathbb{C}^+$ *to the system of self-consistent equations* $s_{1,N}(z) = c_N \int x[-z\{1 + xs_{2,N}(z)\}]^{-1} \mathrm{d}F_{p_0}^{\mathbf{A}_0}(x)$ *and* $s_{2,N}(z) = \int x[-z\{1 + xs_{1,N}(z)\}]^{-1} \mathrm{d}F_N^{\mathbf{B}_0}(x)$*. Let* F_{c_N} *be the deterministic probability measure associated with the ESD of* $\mathbf{Q}_0$*, whose stieltjes transform* $s_{c_N}(z) \coloneqq \int [-z\{1 + xs_{2,N}(z)\}]^{-1} \mathrm{d}F_{p_0}^{\mathbf{A}_0}(x)$*, and define its rightmost positive edge as* $\lambda_{+,N} = \sup\{\mathrm{supp}(F_{c_N}) \cap (0, \infty)\}$.

Next, we consider the corresponding summands associated with the spiked block $\mathbf{Y}_1$. In analogy with (3.16), for each $m \in [1 : M]$,

$$\mathbb{E}(\boldsymbol{y}_m^\top \boldsymbol{y}_m) = \Gamma_m \mathbf{B}_0 \text{ with } \boldsymbol{\Gamma}_m \coloneqq \mathrm{diag}\left(\sigma_m^{(1)} \mathbf{I}_{nh_1}, \ldots, \sigma_m^{(\tau)} \mathbf{I}_{nh_\tau}\right), \tag{3.17}$$

where $\boldsymbol{\Gamma}_m \in \mathbb{R}^{N \times N}$ incorporates the heterogeneous perturbation strengths across all row-populations. Thus, the multi-population heterogeneity affects the phase transition of $\mathbf{G}$ through the matrices $\boldsymbol{\Gamma}_m \mathbf{B}_0$ for $m \in [1 : M]$. Under the additional Assumptions 5, Theorem 2 gives the threshold criterion for separated sample eigenvalues above the bulk edge $\lambda_{+,N}$, together

with the corresponding deterministic outlier location equation.

Assumption 5. There exists a deterministic sequence $N^{-1/2} \le \phi_N \le N^{-c_\phi}$ for some $c_\phi > 0$, such that $\max_{1\le m\le M,\, 1\le r\le \tau,\, 1\le j\le h_r,\, 1\le t\le n} |x_{jmt}^{(r)}| \le N^{1/2}\phi_N$.

Theorem 2 (Threshold criterion)**.** *Under Assumptions 1–5, for each* $m \in [1:M]$*, define the function* $\psi_{m,N}(x) = -N^{-1}\operatorname{tr}\{\mathbf{\Gamma}_m \mathbf{B}_0 \mathcal{P}_0(x)\}$ *for* $x > \lambda_{+,N}$*, where the deterministic equivalent matrix* $\mathcal{P}_0(z) := \{-z(\mathbf{I}_N + s_{1,N}(z)\mathbf{B}_0)\}^{-1}$*. Then* $\psi_{m,N}$ *is positive, continuous, and strictly decreasing on* $(\lambda_{+,N}, \infty)$*, and* $\lim_{x\to\infty} \psi_{m,N}(x) = 0$*. The outlier location equation* $\psi_{m,N}(x) = 1$ *admits a unique solution* $\theta_{m,N} > \lambda_{+,N}$ *if and only if* $\lim_{x\downarrow\lambda_{+,N}} \psi_{m,N}(x) > 1$*. Accordingly, denote the set of supercritical indices as* $\mathcal{M}_+ = \{1 \le m \le M : \lim_{x\downarrow\lambda_{+,N}} \psi_{m,N}(x) > 1\}$*, and define the subcritical set as* $\mathcal{M}_0 = \mathcal{M} \setminus \mathcal{M}_+$*.*

Theorem 2 uses the boundary limit of $\psi_{m,N}$ at $\lambda_{+,N}$ to split $\mathcal{M}$ into the supercritical set $\mathcal{M}_+$, which generates separated sample outliers, and the subcritical set $\mathcal{M}_0$, whose eigenvalues stick to $\lambda_{+,N}$. To analyze the convergence of the outlier eigenvalues induced by the supercritical components in $\mathcal{M}_+$, let $\{\theta_{\ell,N}\}_{\ell=1}^{L}$ be the distinct deterministic roots, arranged in decreasing order. Partition $\mathcal{M}_+$ according to these roots by defining $\mathcal{M}_\ell = \{m \in \mathcal{M}_+ : \theta_{m,N} = \theta_{\ell,N}\}$, with $q_\ell = |\mathcal{M}_\ell|$ for $\ell \in [1:L]$, and set the total cardinality $K = \sum_{\ell=1}^{L} q_\ell$. The index sets of sample eigenvalue ranks corresponding to $\mathcal{M}_\ell$ are $\mathcal{I}_\ell = \{\omega_{\ell-1}+1, \ldots, \omega_\ell\}$, where $\omega_0 = 0$

and $\omega_\ell = \sum_{t=1}^{\ell} q_t$. Therefore, the phase transition of $\mathbf{G}$ is established as following theorem.

Theorem 3 (Phase transition)**.** *Suppose Assumptions 1–5 hold. Assume the separation condition that there exists a constant $c_0 > 0$ such that, for all sufficiently large N, $\min_{1\le\ell\le L}(\theta_{\ell,N} - \lambda_{+,N}) \ge c_0$ and $\min_{\ell\neq k} |\theta_{\ell,N} - \theta_{k,N}| \ge c_0$. As $n \to \infty$ with $N = nh$, the first-order limits of $\lambda_i(\mathbf{G})$ satisfy:*

(i) For each $\ell \in \{1, \ldots, L\}$, $\max_{i\in\mathcal{I}_\ell} |\lambda_i(\mathbf{G}) - \theta_{\ell,N}| \xrightarrow{a.s.} 0$;

(ii) For every fixed integer $\varpi > K$, $\max_{K+1\le j\le\varpi} \left|\lambda_j(\mathbf{G}) - \lambda_{+,N}\right| \xrightarrow{a.s.} 0$.

Driven by the phase transition in Theorem 3, exactly $K = |\mathcal{M}_+|$ sample outliers of $\mathbf{G}$ asymptotically separate from the bulk, providing the theoretical basis for determining K in the next section. Then, for heterogeneous non-spiked components $\mathbf{D}_r \neq \mathbf{D}_u$, the preprocessing alignment used in implementation is described in the Supplementary Materials, together with the proofs of Lemma 2 and Theorems 2–3.

4. Determining the number of *2D*-principal components in $\mathbf{G}$

In *2D*-PCA, strong underlying signals in $\mathbf{G}$ asymptotically detach from the bulk spectrum. Their total number, denoted by K in Theorem 3, provides a direct estimate of the effective number of *2D*-principal components. This quantity plays a key role in image processing: it distinguishes sig-

nal from noise in image denoising (Ades-Aron et al., 2021) and determines the dimensionality of feature maps in face recognition (Zhao et al., 2021). Accordingly, we focus on the following hypothesis testing problem:

$$\mathcal{H}_0 : K = K_0 \quad \text{versus} \quad \mathcal{H}_1 : K \neq K_0, \tag{4.18}$$

where K_0 represents the true number of supercritical spiked eigenvalues. The boundary case $K_0 = 0$ reduces to testing for the absence of separated spikes, implying that any true signals are subcritical and merge with the bulk. Guided by the statistical intuition in Jiang (2023), (4.18) can be tested by separating the detectable signals to evaluate the pure noise components. Thus, we propose the following test statistic:

$$\sum_{i=1}^{p} \lambda_i(\mathbf{G}) - \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}_\ell} \lambda_i(\mathbf{G}). \tag{4.19}$$

Here, the sample ranks $\mathcal{I}_\ell$ index the supercritical subset $\mathcal{M}_\ell$ characterizing the phase transition in Theorem 3. By subtracting these separated sample eigenvalues from the total linear spectral statistic, the evaluation under $\mathcal{H}_0$ is restricted exclusively to the residual bulk spectrum, capturing the pure noise behavior described by the CLT in Theorem 1. The ensuing asymptotic distribution is established in the following theorem.

Theorem 4. *Suppose Assumptions 1–5 hold and* $p/n = c_n \to c \in (0, \infty)$.

For the hypothesis (4.18)*, the test statistic* (4.19) *under* $\mathcal{H}_0$ *satisfies*

$$T_{x,H}(K_0) = \nu_{x,H}^{-\frac{1}{2}}\Big[\sum_{i=1}^{p}\lambda_i(\mathbf{G}) - \sum_{m=1}^{K_0}\lambda_m(\mathbf{G}) - b_{x,H}(K_0) - \mu_{x,H}\Big] \Rightarrow \mathcal{N}(0,1), \quad (4.20)$$

where $b_{x,H}(K_0) = p\int x\,\mathrm{d}F^{c,H^{\mathbf{\Sigma}_0}}(x) - \sum_{m=1}^{K_0}\theta_{m,N}$, $\mu_{x,H} = h^{-1}\mu_x^{\mathbf{G}}$ *and* $\nu_{x,H} = h^{-2}\nu_x^{\mathbf{G}}$*. Here* $\mu_x^{\mathbf{G}}$ *and* $\nu_x^{\mathbf{G}}$ *are defined in* (2.13)*.*

When $\mathbf{D}_0 = \mathbf{I}_p$, Theorem 4 admits deterministic roots $\theta_{m,N}$, where each $m \in [1:M]$ yields a unique root $\theta_{m,N} > \lambda_{+,N}$ of the simplified outlier location equation $\psi_{m,N}(x) = 1$ from Theorem 2, i.e.,

$$\frac{n}{xN}\sum_{r=1}^{\tau}\sigma_m^{(r)}\,\mathrm{tr}\left[\mathbf{R}_{h_r}(\rho_r)\left\{\mathbf{I}_{h_r} + s_{1,N}(x)\mathbf{R}_{h_r}(\rho_r)\right\}^{-1}\right] = 1, \quad (4.21)$$

where $s_{1,N}(z)$ and $s_{2,N}(z)$ satisfy the reduced self-consistent equations $s_{1,N}(z) = -c_N/[z\{1+s_{2,N}(z)\}]$ and $s_{2,N}(z) = -n/(zN)\sum_{r=1}^{\tau}\mathrm{tr}[\mathbf{R}_{h_r}(\rho_r)\{\mathbf{I}_{h_r}+s_{1,N}(z)\mathbf{R}_{h_r}(\rho_r)\}^{-1}]$. Building on (4.21), a sequential testing procedure is conducted on the hypothesis (4.18) to detect an inflection point in the empirical size curve, thereby identifying the true number of components K_0. The complete algorithm for the sequential tests is deferred to Algorithm 1. In practice, the population spiked eigenvalues $\sigma_m^{(r)}$ and the dependence parameter ρ_r are generally unknown and require estimation; see, for example, Ledoit and Wolf (2020); Jiang and Bai (2021), with details and the proof of Theorem 4 provided in the Supplementary Materials. The overall framework is visually summarized in Figure 1.

Algorithm 1 Sequential tests for determining the number of PCs

Require: Replications B; image covariance matrices $\{\mathbf{G}^{(b)}\}_{b=1}^{B}$; upper bound $K_{\max}$; significance level α.

Ensure: Empirical sizes $\{\widehat{\text{size}}(K_0)\}_{K_0=0}^{K_{\max}}$ and estimated number $\widehat{K}$.

Compute ordered eigenvalues $\lambda_1(\mathbf{G}^{(b)}) \geq \cdots \geq \lambda_p(\mathbf{G}^{(b)})$ for $b \in [1:B]$.
Solve the equations in (4.21) for $m \in [1:K_{\max}]$ to obtain the candidate deterministic outlier locations $\{\widehat{\theta}_{m,N}\}_{m=1}^{K_{\max}}$.
for $K_0 = 1$ to $K_{\max}$ **do**
    Compute $\widehat{b}_{x,1}(K_0) = p\int x\,\mathrm{d}F^{c,\widehat{H}}(x) - \sum_{m=1}^{K_0}\widehat{\theta}_{m,N}$.
    for $b = 1$ to B **do**
        Compute $T_{x,1}^{(b)}(K_0)$ via (4.20) using $\{\lambda_i(\mathbf{G}^{(b)})\}_{i=1}^{p}$, $\sum_{m=1}^{K_0}\lambda_m(\mathbf{G}^{(b)})$, $\widehat{b}_{x,1}(K_0)$.
        Set the rejection indicator $R_{K_0}^{(b)} = \mathbf{1}\{|T_{x,1}^{(b)}(K_0)| > z_{1-\alpha/2}\}$.
    end for
    Calculate empirical size $\widehat{\text{size}}(K_0) = B^{-1}\sum_{b=1}^{B} R_{K_0}^{(b)}$.
end for
Estimate $\widehat{K} = \arg\min_{0\leq K_0\leq K_{\max}} \widehat{\text{size}}(K_0)$.
return $\{\widehat{\text{size}}(K_0)\}_{K_0=0}^{K_{\max}}$ and $\widehat{K}$.

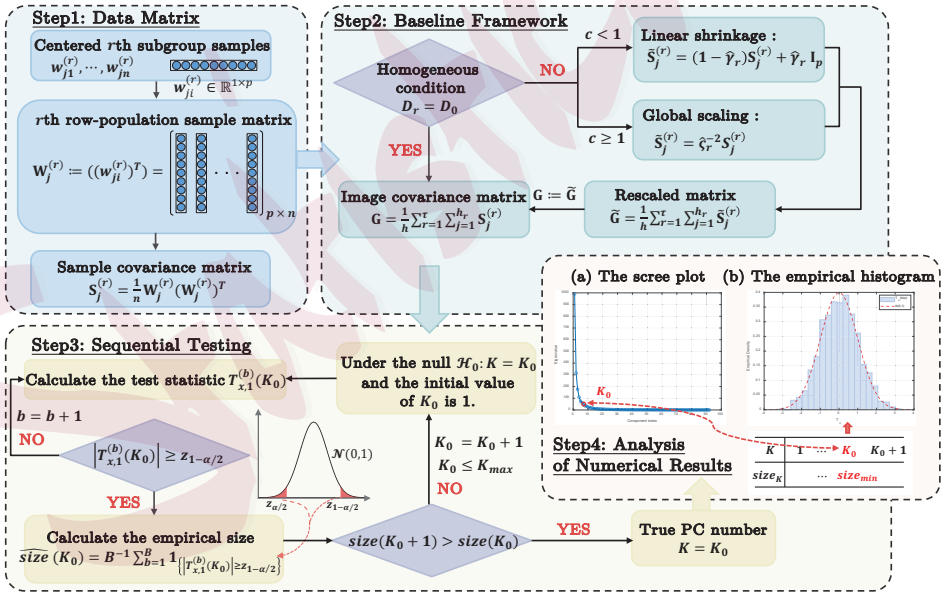

Figure 1: The framework for determining the PC number of $\mathbf{G}$.

5. Simulation Study

5.1 Numerical studies of Theorem 1

The following two models are used to evaluate the numerical performance of the CLT for the spectral statistics of $\mathbf{G}$ in the bulk phase under a non-i.i.d. structure. In both models, we set $\tau = 3$.

Model 1: Assume the population covariance matrix $\mathbf{\Sigma}^{(r)} = \mathbf{T}_r\mathbf{T}_r^\top$ and set $\mathbf{T}_r = \mathbf{L}_r^{1/2}$ for $r \in [1:\tau]$, where $\mathbf{L}_r$ is a diagonal matrix associated with the LSD $H^{(r)}$ in Example 1, which is specifically set as $\mathbf{L}_1 = \mathbf{I}_p$, $\mathbf{L}_2 = \mathrm{diag}(2\mathbf{I}_{p/2}, \mathbf{I}_{p/2})$, and $\mathbf{L}_3 = \mathrm{diag}(3\mathbf{I}_{p/3}, 2\mathbf{I}_{p/3}, \mathbf{I}_{p/3})$.

Model 2: Assume the non-diagonal structure $\mathbf{\Sigma}^{(r)} = \mathbf{T}_r\mathbf{T}_r^\top$ and set $\mathbf{T}_r = \mathbf{U}\mathbf{L}_r^{1/2}\mathbf{U}^\top$ for $r \in [1:\tau]$. Here, $\mathbf{L}_r$ is defined as in Model 1, and $\mathbf{U}$ is composed of eigenvectors of a $p \times p$ matrix $\mathbf{H}\mathbf{H}^\top$ with the entries of $\mathbf{H}$ being independently sampled from $\mathcal{N}(0,1)$.

Aligning with Cases 1–2, the sample matrices $\mathbf{W}_j^{(r)}$ generated via (2.3) maintain cross-population independence while permitting within-population dependence. To assess robustness, we examine two specific population settings. For illustration, Scenario 1 adopts $(h_1, h_2, h_3) = (2, 1, 1)$, where within-population dependence is introduced exclusively for $r = 1$.

Gaussian Assumption. (i) *Independent populations:* For $r = 2, 3$,

we generate $\mathbf{X}_1^{(r)}$ independently with i.i.d. $\mathcal{N}(0,1)$ entries and set $\mathbf{W}_1^{(r)} = \mathbf{T}_r\mathbf{X}_1^{(r)}$. (ii) *Dependent population:* For $r = 1$, the two dependent sample matrices are constructed as $\mathbf{W}_j^{(1)} = \mathbf{T}_1\mathbf{X}_j^{(1)}$ for $j = 1, 2$, where $\mathbf{X}_j^{(1)} = \sqrt{\rho_1}\,\mathbf{X}_0^{(1)} + \sqrt{1-\rho_1}\,\mathbf{Z}_j^{(1)}$. Here, the shared latent matrix $\mathbf{X}_0^{(1)}$ and idiosyncratic noises $\mathbf{Z}_1^{(1)}, \mathbf{Z}_2^{(1)}$ are mutually independent with i.i.d. $\mathcal{N}(0,1)$ entries, while $\rho_1 \in [0,1]$ quantifies the inter-sample dependence.

Gamma Assumption. (i) *Independent populations:* For $r = 2, 3$, we generate $\mathbf{X}_1^{(r)}$ with i.i.d. entries drawn from Gamma$(25, 0.2) - 5$ and set $\mathbf{W}_1^{(r)} = \mathbf{T}_r\mathbf{X}_1^{(r)}$. (ii) *Dependent population:* For $r = 1$, let $\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3$ be independent random matrices where entries $z_{ki}^{(1)} \sim$ Gamma$(25\rho_1, 0.2)$ and $z_{ki}^{(\ell)} \sim$ Gamma$(25(1-\rho_1), 0.2)$ for $\ell = 2, 3$. Defining $\mathbf{X}_1^{(1)} = \mathbf{Z}_1 + \mathbf{Z}_2 - 5\mathbf{1}_{p\times n}$ and $\mathbf{X}_2^{(1)} = \mathbf{Z}_1 + \mathbf{Z}_3 - 5\mathbf{1}_{p\times n}$, we obtain the observed samples $\mathbf{W}_j^{(1)} = \mathbf{T}_1\mathbf{X}_j^{(1)}$ for $j = 1, 2$, where $\rho_1 \in [0,1]$ determines the dependence strength.

To evaluate the tests in a diverse range of events, we examine four experimental circumstances in such settings. For each condition, the sample size is set to $n \in \{300, 600, 900\}$ with the dimensional ratio $c = p/n \in \{0.5, 1, 1.5\}$. All tests are assessed through 1000 replications at the significance level $\alpha = 0.05$. Table 1 summarizes the empirical sizes under Scenario 1 with $(h_1, h_2, h_3) = (2, 1, 1)$ for $\rho_1 \in \{0.5, 0.7, 0.9\}$. We make the following observations. First, the empirical sizes remain close to $\alpha = 0.05$

Table 1: Empirical size with Scenario 1 for varying ρ_1 in Example 1.

			Gaussian population						Gamma population					
			Model 1			Model 2			Model 1			Model 2		
n	c	p	$\rho_1=0.5$	$\rho_1=0.7$	$\rho_1=0.9$	$\rho_1=0.5$	$\rho_1=0.7$	$\rho_1=0.9$	$\rho_1=0.5$	$\rho_1=0.7$	$\rho_1=0.9$	$\rho_1=0.5$	$\rho_1=0.7$	$\rho_1=0.9$
300	0.5	150	0.055	0.051	0.056	0.056	0.060	0.046	0.052	0.052	0.058	0.048	0.045	0.049
	1.0	300	0.051	0.041	0.056	0.039	0.048	0.058	0.053	0.054	0.058	0.056	0.047	0.057
	1.5	450	0.052	0.038	0.054	0.040	0.059	0.045	0.047	0.051	0.050	0.052	0.050	0.048
600	0.5	300	0.052	0.035	0.052	0.051	0.065	0.043	0.064	0.054	0.053	0.061	0.055	0.041
	1.0	600	0.057	0.051	0.052	0.060	0.058	0.047	0.038	0.043	0.049	0.046	0.059	0.057
	1.5	900	0.041	0.062	0.051	0.050	0.046	0.049	0.052	0.053	0.048	0.047	0.050	0.049
900	0.5	450	0.053	0.049	0.047	0.054	0.045	0.044	0.046	0.047	0.052	0.057	0.052	0.051
	1.0	900	0.053	0.044	0.048	0.053	0.062	0.056	0.055	0.059	0.040	0.056	0.060	0.059
	1.5	1350	0.053	0.042	0.045	0.045	0.047	0.055	0.046	0.056	0.045	0.046	0.040	0.048

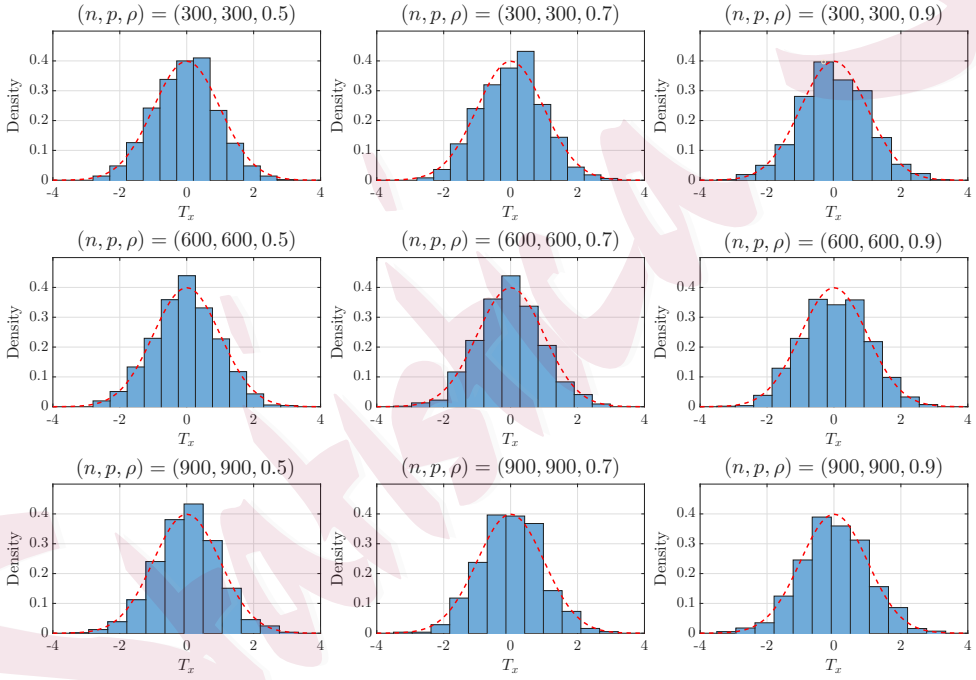

Figure 2: Empirical distributions of T_x in Model 1 under Gaussian.

across sample sizes n and all regimes of c. This stability in both low- and high-dimensional settings supports the proposed asymptotic approxi-

mation. Second, for all values of ρ_1 considered, the Type-I error is generally well controlled, with most entries lying within or near the 95% Monte Carlo error band. Moreover, when the within-population dependence becomes stronger, especially for $\rho_1 = 0.9$, the deviations from the Monte Carlo error band are mild, suggesting a tighter agreement with the asymptotic theory. These results indicate that the proposed asymptotic characterization remains applicable to model structures with complex dependence. Figure 2 shows that the empirical histograms of the test statistic for Model 1 ($c = 1$) align well with the standard normal limit, as indicated by the overlaid red dashed curve. Additional simulations for Models 1–2 across Scenarios 1–3 in Example 1 with $\rho_r = 0.9$ are reported in the Supplementary Materials.

5.2 Performance of PC number detection for G

In this section, we demonstrate the practical application and effectiveness of Theorem 4 for the hypothesis test (4.18) under two non-i.i.d. multi-population spiked models. In both models, we set $\tau = 4$.

Model 3: Under the homogeneous baseline, we consider the population spiked covariance matrix $\mathbf{\Sigma}^{(r)} = \mathbf{U}(\mathbf{I} + \mathbf{\Lambda}_r)\mathbf{U}^\top$ for $r \in [1 : \tau]$, where $\mathbf{U}$ is defined in Model 2. The spiked parts take the low-rank form $\mathbf{\Lambda}_1 = \mathrm{diag}(40, 35, 0, \ldots, 0)$, $\mathbf{\Lambda}_2 = \mathrm{diag}(40, 36, 28, 0, \ldots, 0)$, $\mathbf{\Lambda}_3 = \mathrm{diag}(40, 38, 29, 25, 20, 0, \ldots, 0)$,

and $\mathbf{\Lambda}_4 = \text{diag}(40, 36, 28, 25, 20, 18, 0, \ldots, 0)$ with $(K_1, K_2, K_3, K_4) = (2, 3, 5, 6)$. The random samples $\mathbf{X}_j^{(r)}$ are generated using the dependent Gaussian design from Section 5.1, with $h_r = (1, 2, 3, 4)$ and $\rho_r = 0.5$. We take $n \in \{600, 1200, 1800\}$ and $c \in \{0.5, 1, 1.5\}$.

Model 4: Assume the heterogeneous spiked covariance structure $\mathbf{\Sigma}^{(r)} = \mathbf{U}(\mathbf{D}_r + \mathbf{\Lambda}_r)\mathbf{U}^\top$ for $r \in [1 : \tau]$. The non-spiked parts are given by $\mathbf{D}_1 = \text{diag}(3\mathbf{I}_{p/3}, 2\mathbf{I}_{p/3}, \mathbf{I}_{p/3})$, $\mathbf{D}_2 = \text{diag}(2\mathbf{I}_{p/2}, \mathbf{I}_{p/2})$, $\mathbf{D}_3 = 3\mathbf{I}_p$, and $\mathbf{D}_4 = \mathbf{I}_p$. The spiked parts retain the low-rank structure specified in Model 3. The random samples $\mathbf{X}_j^{(r)}$ follow the dependent designs in Section 5.1, utilizing the Gaussian assumption for $r = 1, 2$ and the Gamma assumption for $r = 3, 4$. We take $n \in \{1800, 2400, 3000\}$ and $c \in \{0.5, 1, 1.5\}$.

For brevity, we focus on the results for Model 4, which features a more complex high-dimensional structure and multi-population heterogeneity. Results for Model 3 are provided in the Supplementary Materials. Table 2 reports the empirical sizes of the sequential test at level $\alpha = 0.05$ based on 1000 replications for Model 4. Across varying (n, p, c) configurations, we observe that the empirical size consistently reaches its minimum at the true $K_0 = 6$ (bold). Complementing this, the Kolmogorov-Smirnov (KS) distance serves as an auxiliary diagnostic, which similarly attains its minimum at this specific K_0, thereby corroborating the asymptotic validity of

Table 2: Empirical sizes for Model 4 under varying (n, c, p) and K_0.

n	c	p	$K_0 = 1$	$K_0 = 2$	$K_0 = 3$	$K_0 = 4$	$K_0 = 5$	$K_0 = 6$	$K_0 = 7$
1800	0.5	900	0.294 (0.167)	0.195 (0.113)	0.135 (0.101)	0.103 (0.062)	0.072 (0.041)	**0.050** (0.032)	0.809 (0.842)
	1.0	1800	0.164 (0.122)	0.121 (0.090)	0.107 (0.070)	0.086 (0.062)	0.068 (0.053)	**0.051** (0.033)	0.435 (0.621)
	1.5	2700	0.133 (0.114)	0.107 (0.078)	0.083 (0.070)	0.059 (0.053)	0.053 (0.046)	**0.047** (0.043)	0.250 (0.492)
2400	0.5	1200	0.231 (0.129)	0.159 (0.088)	0.110 (0.068)	0.075 (0.045)	0.055 (0.032)	**0.048** (0.036)	0.820 (0.848)
	1.0	2400	0.136 (0.078)	0.094 (0.055)	0.085 (0.034)	0.070 (0.039)	0.055 (0.028)	**0.047** (0.021)	0.441 (0.655)
	1.5	3600	0.137 (0.084)	0.103 (0.056)	0.094 (0.043)	0.077 (0.036)	0.067 (0.023)	**0.052** (0.019)	0.281 (0.509)
3000	0.5	1500	0.188 (0.129)	0.141 (0.098)	0.108 (0.085)	0.074 (0.073)	0.063 (0.049)	**0.049** (0.044)	0.783 (0.834)
	1.0	3000	0.112 (0.080)	0.082 (0.069)	0.061 (0.043)	0.055 (0.032)	0.052 (0.031)	**0.049** (0.027)	0.477 (0.653)
	1.5	4500	0.088 (0.076)	0.081 (0.060)	0.059 (0.050)	0.058 (0.045)	0.055 (0.044)	**0.050** (0.037)	0.266 (0.497)

Note: Empirical size (KS distance in parentheses).

the test statistic. These results indicate that the true K_0 is typically associated with an inflection of the empirical size curve and that the proposed method reliably identifies the number of principal components across non-i.i.d. multi-population settings in both low- and high-dimensional regimes.

Furthermore, Figure 3 illustrates that standard heuristics, such as the scree plot, Kaiser criterion, and 85% cumulative variance threshold, fail under strong dependence. As the dimension grows, the accumulation of bulk

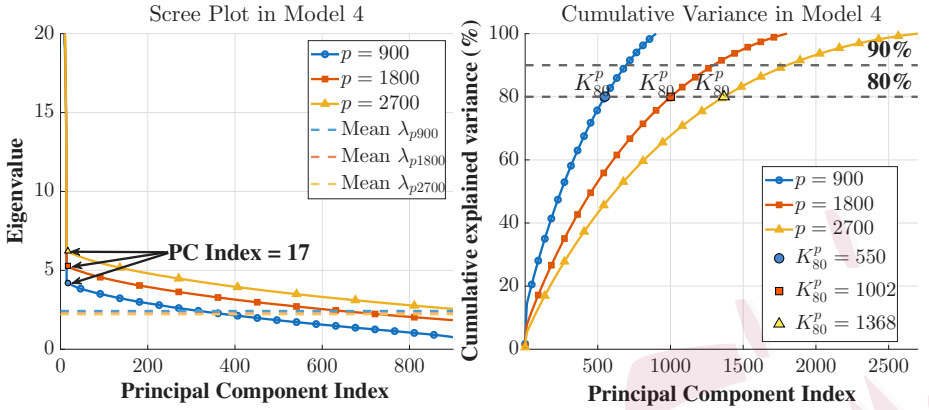

Figure 3: Scree plots and cumulative variance in Model 4.

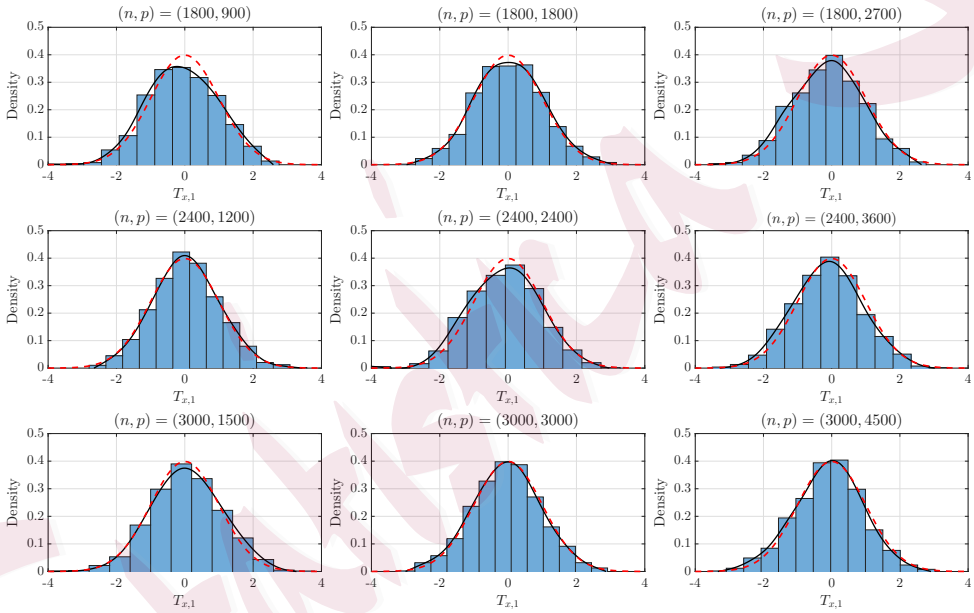

Figure 4: Empirical distributions of $T_{x,1}$ under the null in Model 4.

eigenvalues near the spectral edge blurs the separation between spikes and the bulk and shifts the apparent elbow to later components, causing visually based rules to absorb close spikes and overestimate K_0. In contrast, our

theoretically justified procedure effectively detects true outliers separating from the bulk support, accurately identifying $K_0 = 6$ regardless of the non-i.i.d. structure. Figure 4 corroborates this robustness, showing the empirical null distribution matches $\mathcal{N}(0, 1)$, validating the asymptotic approximation.

6. Real Data Analysis

In this section, we evaluate the proposed method using the ORL, FERET, and Extended Yale B facial image databases to cover varying complexities in dimensionality, sample size, and environmental factors. The ORL database contains 400 images sized 92×112 pixels and serves as a small-sample baseline characterized by pose variations (available at `http://cam-orl.co.uk/facedatabase.html`). The FERET database offers richer expression diversity and is analyzed in the Supplementary Materials. Finally, the Extended Yale B database (Georghiades et al., 2001) tests robustness against high dimensionality (480×640 pixels) and severe noise from illumination and uncropped backgrounds, utilizing a subset of 13 subjects with 50 images each. In the experimental setup, for the ORL database, $\mathbf{G}$ uses one random image from the first five samples per class (Figure 5), reserving the rest for testing. For the Extended Yale B database, training ratios of $\{0.1, 0.2, 0.3, 0.4\}$ are applied, leaving 390–585 testing images.

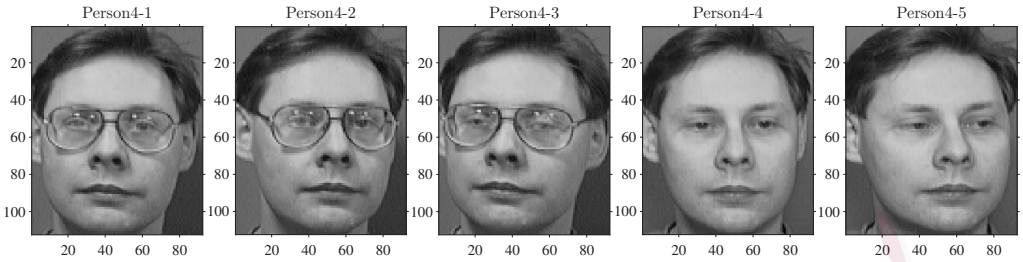

Figure 5: Five sample images from one subject in the ORL database.

Tables 3 and 4 report the empirical sizes and KS distances for the sequential test, based on 300 replications at $\alpha = 0.05$. Following the elbow pattern of the size curve and the KS distance, we identify $K_0 = 4$ for the ORL database and $K_0 = 10$ for Extended Yale B database. Table 5 summarizes these results alongside comparisons with the scree heuristic and the SPA methodology (Li and Kuang, 2024), specifically the SPA upper-edge (SPA-u) and SPA pairwise (SPA-p) thresholds.

On the ORL database, our sequential test selects $K_0 = 4$ (69.44%), matching SPA-p and surpassing SPA-u and the scree heuristic (68.06%). Unlike SPA-p, which averages thresholds across classes, our procedure employs a unified, theoretically derived cutoff. For the Extended Yale B

Table 3: Empirical size and KS distance for (4.18) in the ORL database.

	$K_0 = 1$	$K_0 = 2$	$K_0 = 3$	$K_0 = 4$	$K_0 = 5$	$K_0 = 6$
Empirical size	1.000	1.000	0.096	**0.056**	1.000	1.000
KS distance metric	(1.000)	(0.999)	(0.243)	(0.052)	(0.956)	(1.000)

Table 4: Empirical size and KS distance for (4.18) under varying ratios in the Extended Yale B database.

Ratio	$K_0=1$	$K_0=2$	$K_0=3$	$K_0=4$	$K_0=5$	$K_0=6$	$K_0=7$	$K_0=8$	$K_0=9$	$K_0=10$	$K_0=11$	$K_0=12$
0.1	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.836	0.083	**0.046**	1.000	1.000
	(1.000)	(1.000)	(1.000)	(0.999)	(0.999)	(0.989)	(0.987)	(0.855)	(0.195)	(0.045)	(0.867)	(0.995)
0.2	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.980	0.075	**0.053**	0.946	1.000
	(1.000)	(1.000)	(1.000)	(1.000)	(0.999)	(0.997)	(0.989)	(0.965)	(0.122)	(0.027)	(0.936)	(1.000)
0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.977	0.092	**0.046**	0.993	1.000
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(0.988)	(0.964)	(0.252)	(0.060)	(0.978)	(1.000)
0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.194	**0.043**	0.996	1.000
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(0.999)	(0.991)	(0.334)	(0.037)	(0.976)	(1.000)

Note: Empirical size (KS distance in parentheses).

Table 5: Selected K_0 and test-set classification accuracy for facial databases.

		SPA				**Scree**		**Ours**	
		SPA-u		**SPA-p**					
Database	Ratio	K_0	Acc.(%)	K_0	Acc.(%)	K_0	Acc.(%)	K_0	Acc.(%)
ORL	0.1	3	68.06	4	**69.44**	3	68.06	4	**69.44**
Extended Yale B	0.1	6	88.03	8	89.05	9	90.59	10	**91.28**
	0.2	6	92.50	8	93.26	8	93.26	10	**94.23**
	0.3	6	93.62	7	94.72	8	94.28	10	**95.82**
	0.4	6	96.41	8	96.15	9	96.41	10	**96.67**

database, our method consistently selects $K_0 = 10$ across training ratios 0.1–0.4, achieving the highest accuracies (91.28%–96.67%). In contrast, SPA-p is more sensitive to training ratios, with K_0 fluctuating between 7 and 8 and yielding lower accuracy, consistent with background heterogeneity and illumination variations that may affect the stability of resampling-based thresholding. Correspondingly, Figure 6 presents the comparative

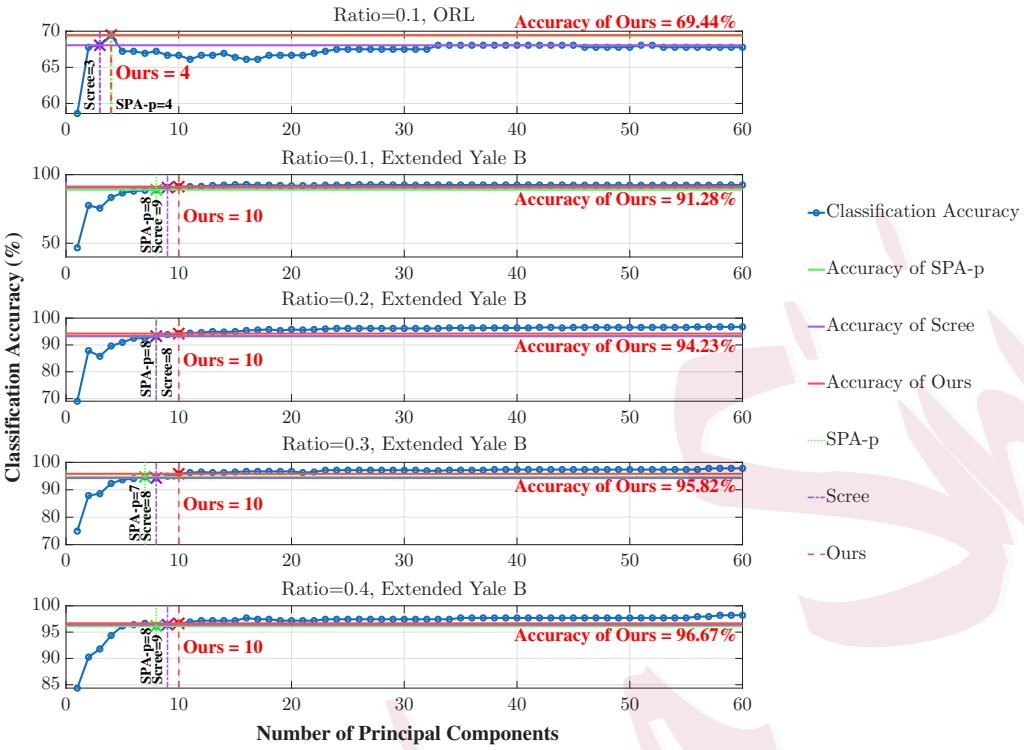

Figure 6: Test-set classification accuracy vs. number of PCs.

classification performance on both databases. As shown by the red solid curve, our sequential procedure provides stable selection and competitive accuracy under complex dependencies, achieving a reasonable trade-off between accuracy and the number of retained principal components K_0.

7. Conclusion

This paper develops an asymptotic framework for image covariance matrices under non-i.i.d. spatial dependence, proposing a tuning-free sequential test

to determine the number of *2D*-principal components. While current theory relies on the global scatter tr($\mathbf{G}$) to capture overall covariance variability, extending this to weighted statistics tr($\mathbf{CG}$) offers a natural progression. Employing specific projection matrices $\mathbf{C}$ isolates structural information within localized subspaces, which may be useful for weak signal detection and downstream tasks such as image compression. Exploring these anisotropic extensions (Yang, 2024) represents a valuable future direction. Practically, our procedure serves as a plug-and-play spectral module for *2D*-PCA variants and related matrix-structured learning tasks involving complex dependencies.

Supplementary Materials

The online Supplementary Materials include: technical proofs for Lemmas 1–2 and Theorems 2–4; calculations for Example 1; preprocessing details; additional simulations; and facial database experiments, specifically empirical support and analysis of the FERET dataset.

Acknowledgements

The authors are grateful to the Editor, the Associate Editors, and the referees for their review of the paper. Dandan Jiang was supported by

NSFC under Grant No. 12571311.

References

Ades-Aron, B., G. Lemberskiy, J. Veraart, and J. Golfinos (2021). Improved task-based functional MRI language mapping in patients with brain tumors through Marchenko–Pastur principal component analysis denoising. *Radiology 298*(2), 365–373.

Bai, Z. and J. W. Silverstein (2008). CLT for linear spectral statistics of large-dimensional sample covariance matrices. In *Advances in Statistics*, Singapore, pp. 281–333. World Scientific.

Bai, Z. and J. Yao (2012). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis 106*, 167–177.

Bai, Z. D., B. Miao, and G. Pan (2007). On asymptotics of eigenvectors of large sample covariance matrix. *The Annals of Probability 35*(4), 1532–1572.

Bao, Z., X. Ding, J. Wang, and K. Wang (2022). Statistical inference for principal components of spiked covariance matrices. *The Annals of Statistics 50*(2), 1144–1169.

Chang, Y., L. Yan, and S. Zhong (2017). Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4260–4268.

Ding, K., K. Ma, S. Wang, and E. P. Simoncelli (2022). Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine*

Intelligence 44(5), 2567–2581.

Ding, X. and F. Yang (2021). Spiked separable covariance matrices and principal components. *The Annals of Statistics 49*(2), 1113–1138.

Georghiades, A. S., P. N. Belhumeur, and D. J. Kriegman (2001). From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*(6), 643–660.

Guo, S., Z. Yan, K. Zhang, W. Zuo, and L. Zhang (2019). Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1712–1722.

Jiang, D. (2023). A universal test on spikes in a high-dimensional generalized spiked model and its applications. *Statistica Sinica 33*, 1749–1770.

Jiang, D. and Z. Bai (2021). Generalized four moment theorem and an application to CLT for spiked eigenvalues of high-dimensional covariance matrices. *Bernoulli 27*(1), 274–294.

Jin, S., K. Miao, and L. Su (2021). On factor models with random missing: EM estimation, inference, and cross validation. *Journal of Econometrics 222*(1), 745–777.

Ledoit, O. and M. Wolf (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics 48*(5), 3043–3065.

Li, Z. and Y. Kuang (2024). SPA: Signflip parallel analysis to optimize the number of principal components in two-dimensional PCA. *Journal of Data Science*.

Minka, T. P. (2000). Automatic choice of dimensionality for PCA. In *Advances in Neural Information Processing Systems*, Volume 13, Cambridge, MA, pp. 598–604. MIT Press.

Wang, M., X. Jiang, J. Gao, T. Wang, C. Hu, and F. Liu (2021). Minimum unbiased risk estimate based 2DPCA for color image denoising. *Neurocomputing 440*, 127–144.

Xu, S., J. Zhang, and C. Zhang (2022). Hyperspectral image denoising by low-rank models with hyper-laplacian total variation prior. *Signal Processing 201*, 108733.

Yang, F. (2024). Linear spectral statistics of eigenvectors of anisotropic sample covariance matrices. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 60*(4), 2767–2812.

Yang, J., D. Zhang, A. F. Frangi, and J.-y. Yang (2004). Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 26*(1), 131–137.

Yuan, D. and N. Mancuso (2023). SuSiE PCA: A scalable Bayesian variable selection technique for principal component analysis. *iScience 26*(11), 108181.

Zha, Z., X. Yuan, J. Zhou, C. Zhu, and B. Wen (2020). Image restoration via simultaneous nonlocal self-similarity priors. *IEEE Transactions on Image Processing 29*, 8561–8576.

Zhang, H., H. Bi, X. Wang, and P. Zhang (2023). A joint-norm distance metric 2DPCA for robust dimensionality reduction. *Information Sciences 640*, 119036.

Zhao, M., Z. Jia, Y. Cai, X. Chen, and D. Gong (2021). Advanced variations of two-dimensional

principal component analysis for face recognition. *Neurocomputing 452*, 653–664.

Zheng, S., Z. Bai, and J. Yao (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *The Annals of Statistics 43*(2), 546–591.

Zhou, Y., J. Jiao, H. Huang, Y. Wang, J. Wang, H. Shi, and T. Huang (2020). When AWGN-based denoiser meets real noises. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 34, pp. 13074–13081.

First author affiliation: School of Mathematics and Statistics, Xi'an Jiaotong University, No.28, Xianning West Road, Xi'an, Shaanxi, 710049, China.

Corresponding author. E-mail: suh_liu@stu.xjtu.edu.cn

Second author affiliation: School of Mathematics and Statistics, Xi'an Jiaotong University, No.28, Xianning West Road, Xi'an, Shaanxi, 710049, China.

Corresponding author. E-mail: yangl@xjtu.edu.cn

Third author affiliation: School of Mathematics and Statistics, Xi'an Jiaotong University, No.28, Xianning West Road, Xi'an, Shaanxi, 710049, China.

* Corresponding author. E-mail: jiangdd@xjtu.edu.cn