WEIGHTED HYBRID PCA: LEARNING SHARED STRUCTURE UNDER BATCH EFFECTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Principal component analysis (PCA) is widely used to extract low-dimensional structure from high-dimensional data. However, batch effects, such as those due to experimental timing, equipment, or data sources, may dominate the leading principal components and obscure the shared variation of interest. When the corresponding labels are only partially observed, removing these effects without discarding useful observations becomes difficult. We model this setting using a factor model with a shared factor loading space and different intercepts across batches. We then propose weighted hybrid PCA, which combines covariance estimates from labeled and unlabeled observations. Using the observed labels, we select a weight from a safe set that excludes additional components caused by between-population variation. Under suitable conditions, the selected weight asymptotically minimizes both the factor detectability threshold and the factor loading space estimation error over the safe set. The lower threshold allows weaker factors to be detected. Simulations show substantially lower factor loading space estimation error for weights inside the safe set than for those outside it. The selected weight achieves the lowest estimation error, close to the oracle benchmark. On two A549 single-cell RNA perturbation datasets with p53-related expression responses, weighted hybrid PCA more efficiently captures the common within-population variation than the other methods and gives the highest average number and proportion of p53-associated principal components.

1 INTRODUCTION

Principal component analysis (PCA) is a standard tool for estimating dominant low-dimensional structure in high-dimensional data (Jolliffe & Cadima, 2016). Under a factor model, PCA is commonly used to estimate the factor loading space (Bai, 2003; Fan et al., 2013). We study this problem in a partially labeled multi-population setting, where the populations share the same factor loading space but may have different intercept vectors, and population memberships are observed for only a subset of the observations. A motivating example is gene expression data collected across multiple batches. Batch effects may shift expression levels and obscure the biological variation shared across batches. When population memberships are observed, the population-specific intercepts can be removed by centering observations within each population. For unlabeled observations, however, the resulting between-population variation remains and may introduce dominant principal directions outside the factor loading space. Discarding the unlabeled observations avoids this additional variation but also discards observations that contain information about the factor loading space. The central problem is therefore how to exploit unlabeled observations while controlling the effect of between-population variation on factor loading space estimation.

Consider N observations from a mixture of K populations:

$$\boldsymbol{x}_i = \boldsymbol{\mu}_{g_i} + \boldsymbol{U}\boldsymbol{f}_i + \sigma \boldsymbol{z}_i, \qquad i = 1, \ldots, N, \tag{1}$$

where $\mathbb{P}\{g_i = k\} = \pi_k$, $\pi_k \in (0, 1)$, and $\sum_{k=1}^{K} \pi_k = 1$. Here $\boldsymbol{\mu}_k \in \mathbb{R}^p$ is the intercept vector for population k, $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r) \in \mathbb{R}^{p \times r}$ has orthonormal columns spanning the factor loading space, $\boldsymbol{f}_i \in \mathbb{R}^r$ is the factor score, and $\boldsymbol{z}_i \in \mathbb{R}^p$ is the idiosyncratic noise. Here $\sigma > 0$ is the common noise scale. Dividing all observations by σ leaves the factor loading space unchanged. For the theoretical analysis, we set $\sigma = 1$ without loss of generality and retain the same notation for

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the rescaled quantities. We select m observations uniformly at random and reveal their population memberships, leaving $n = N - m$ observations unlabeled. Denote the labeled observations by $\{(\boldsymbol{x}_i^a, g_i^a)\}_{i=1}^m$ and the unlabeled observations by $\{\boldsymbol{x}_i^u\}_{i=1}^n$.

The factor scores are centered with covariance $\boldsymbol{\Theta} = \mathrm{diag}(\theta_1, \ldots, \theta_r)$, where $\theta_1 > \cdots > \theta_r > 0$ are the factor strengths. The idiosyncratic noise is centered with identity covariance and is uncorrelated with the factor scores. It follows that the common within-population covariance is

$$\boldsymbol{\Sigma} = \boldsymbol{I}_p + \boldsymbol{U}\boldsymbol{\Theta}\boldsymbol{U}^\top. \tag{2}$$

Thus, the factor loading space is the leading r-dimensional principal subspace of $\boldsymbol{\Sigma}$, with $\boldsymbol{u}_a$ associated with the eigenvalue $1 + \theta_a$, for $a = 1, \ldots, r$.

Let $\bar{\boldsymbol{\mu}} = \sum_{k=1}^K \pi_k \boldsymbol{\mu}_k$ denote the mixture mean. The population-specific intercepts induce the between-population covariance

$$\boldsymbol{\Sigma}_\mu = \sum_{k=1}^K \pi_k (\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}})^\top. \tag{3}$$

For the labeled observations, we center within each population and form

$$\boldsymbol{S}_{\mathrm{lab}} = \frac{1}{m} \sum_{k=1}^K \sum_{i:g_i^a = k} (\boldsymbol{x}_i^a - \bar{\boldsymbol{x}}_k^a)(\boldsymbol{x}_i^a - \bar{\boldsymbol{x}}_k^a)^\top, \qquad \bar{\boldsymbol{x}}_k^a = \frac{1}{m_k} \sum_{i:g_i^a = k} \boldsymbol{x}_i^a,$$

where $m_k = \sum_{i=1}^m \mathbf{1}_{\{g_i^a = k\}}$. Thus, $\boldsymbol{S}_{\mathrm{lab}}$ estimates the common within-population covariance $\boldsymbol{\Sigma}$.

For the unlabeled observations, we use the globally centered covariance

$$\boldsymbol{S}_{\mathrm{mix}} = \frac{1}{n} \sum_{i=1}^n (\boldsymbol{x}_i^u - \bar{\boldsymbol{x}}^{\,u})(\boldsymbol{x}_i^u - \bar{\boldsymbol{x}}^{\,u})^\top, \qquad \bar{\boldsymbol{x}}^{\,u} = \frac{1}{n} \sum_{i=1}^n \boldsymbol{x}_i^u.$$

The corresponding population covariance is $\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_\mu$.

If all population memberships were observed, all N observations could be centered within each population, yielding the oracle covariance matrix

$$\boldsymbol{S}_{\mathrm{or}} = \frac{1}{N} \sum_{k=1}^K \sum_{i:g_i = k} (\boldsymbol{x}_i - \bar{\boldsymbol{x}}_k)(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_k)^\top, \quad \bar{\boldsymbol{x}}_k = \frac{1}{N_k} \sum_{i:g_i = k} \boldsymbol{x}_i,$$

where $N_k = \sum_{i=1}^N \mathbf{1}_{\{g_i = k\}}$. We use $\boldsymbol{S}_{\mathrm{or}}$ as an oracle benchmark.

We combine $\boldsymbol{S}_{\mathrm{lab}}$ and $\boldsymbol{S}_{\mathrm{mix}}$ through

$$\boldsymbol{S}_\alpha = \alpha \boldsymbol{S}_{\mathrm{lab}} + (1 - \alpha)\boldsymbol{S}_{\mathrm{mix}}, \qquad \alpha \in [0, 1]. \tag{4}$$

We refer to PCA based on $\boldsymbol{S}_\alpha$ as weighted hybrid PCA. Decreasing α makes greater use of the unlabeled observations, but also increases the contribution of between-population variation.

In high dimensions, additional observations can lower the factor detectability threshold. A direction is detectable when its strength exceeds this threshold, so that the corresponding sample eigenvalue separates from the spectral bulk and the sample eigenvector has nonvanishing asymptotic alignment with the population direction (Baik et al., 2005; Baik & Silverstein, 2006; Paul, 2007). For $\boldsymbol{S}_\alpha$, the threshold depends on α and, in the absence of between-population variation, is asymptotically minimized at the natural weight $\alpha_0 = m/N$. Giving the unlabeled observations too much weight, however, may allow between-population variation to generate additional separated eigenvalues.

In $\boldsymbol{S}_\alpha$, the between-population covariance enters with weight $1 - \alpha$. We call α safe if this weighted contribution does not generate additional separated eigenvalues. Weighted hybrid PCA selects the weight closest to the natural weight α_0 from the estimated safe set.

Related work has studied spectral effects of population mean differences in high-dimensional multi-population models. Li & Zhu (2023) characterized the fluctuations of sample eigenvalues arising from mean differences in Gaussian mixtures, while Li et al. (2024) studied such outlier eigenvalues for a renormalized sample covariance matrix and applied them to estimating the number of

populations and assessing clustering results. Our goal is different: we aim to estimate the factor loading space while preventing between-population variation from introducing additional principal components. Robust PCA methods address sparse or other forms of contamination (Candès et al., 2011; Xu et al., 2012; Cai et al., 2019; Jambulapati et al., 2020; Paul et al., 2024), rather than the between-population variation considered here. More closely related is the mean-shift PCA method of Li et al. (2026), which adds a knockoff mean and uses the stability of sample eigenvalues to distinguish covariance components from mean-shift-induced components without exploiting available population memberships. We instead use the available memberships to estimate the strength of the between-population covariance and guide the weighting of the labeled and unlabeled observations.

Under an asymptotic orthogonality condition and suitable signal strengths, safe weights consistently select the number of factors and exclude additional components caused by between-population variation. We also characterize the factor loading space estimation error. The selected weight asymptotically minimizes both the factor detectability threshold and the factor loading space estimation error over the estimated safe set. The lower threshold allows weaker factors to be detected. In this case, weighted hybrid PCA attains the same asymptotic estimation error as oracle PCA when the natural weight α_0 is safe. Without the orthogonality condition, safe weights still consistently select the number of factors, and we characterize the factor loading space estimation error.

Simulations show substantially lower factor loading space estimation error for weights inside the estimated safe set than for those outside it. The selected weight achieves the lowest observed error, close to the oracle benchmark. We also evaluate the methods using two A549 single-cell RNA perturbation datasets with p53-related expression responses. In both datasets, weighted hybrid PCA more efficiently captures the common within-population variation than the other methods and gives the highest average number and proportion of p53-associated principal components.

Notation. For a real symmetric matrix $\boldsymbol{A}$, its eigenvalues are ordered as $\lambda_1(\boldsymbol{A}) \geq \lambda_2(\boldsymbol{A}) \geq \cdots$. We use $\|\cdot\|$ for the Euclidean norm of vectors and the spectral norm of matrices, and $\|\cdot\|_F$ for the Frobenius norm. Let $\mathbb{C}^+ = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$. Unless otherwise stated, scalar model parameters not indexed by p are held fixed as $p \to \infty$. For convergent deterministic population scalar parameters, we use the same notation for the sequence and its limit whenever no ambiguity arises.

2 WEIGHTED HYBRID PCA

We first introduce the quantities used in the procedure; their formal definitions are given in Section 3. Let γ_μ denote the strength of the between-population covariance. For a weight $\alpha \in [0,1]$, let t_α denote the between-population separation threshold. A weight is safe when $(1-\alpha)\gamma_\mu \leq t_\alpha$, so that the weighted between-population covariance does not generate additional separated eigenvalues. The factor loading space is detectable when $\theta_r > \vartheta_\alpha$, where ϑ_α is the factor detectability threshold. Algorithm 1 summarizes the procedure, with the required empirical quantities defined below.

Algorithm 1 Weighted hybrid PCA

1: Compute $\widehat{\gamma}_\mu$ from (5).
2: Compute $\widehat{s}_\alpha$ from (6), obtain $\widehat{t}_\alpha$, and construct $\widehat{\mathcal{A}}_N$ from (7).
3: Select α_* by (8).
4: Form $\boldsymbol{S}_{\alpha_*}$ and compute its ordered eigenvalues and eigenvectors.
5: Compute $\widehat{\lambda}_{+,\alpha_*}$ from (9) and determine $\widehat{r}_{\alpha_*}$ from (10).
6: Estimate the factor loading space by $\operatorname{span}\{\widehat{\boldsymbol{q}}_1^{\alpha_*}, \ldots, \widehat{\boldsymbol{q}}_{\widehat{r}_{\alpha_*}}^{\alpha_*}\}$.

To construct the empirical safe set, we first estimate γ_μ from the labeled observations. Define

$$\widehat{\boldsymbol{C}} = \operatorname{diag}(\widehat{\boldsymbol{\pi}}) - \widehat{\boldsymbol{\pi}}\widehat{\boldsymbol{\pi}}^\top, \qquad \widehat{\boldsymbol{G}} = (\bar{\boldsymbol{x}}_1^a, \ldots, \bar{\boldsymbol{x}}_K^a)^\top (\bar{\boldsymbol{x}}_1^a, \ldots, \bar{\boldsymbol{x}}_K^a) - \operatorname{diag}\left(\frac{\widehat{\tau}_1}{m_1}, \ldots, \frac{\widehat{\tau}_K}{m_K}\right),$$

where $\widehat{\boldsymbol{\pi}} = (m_1/m, \ldots, m_K/m)^\top$ and $\widehat{\tau}_k = \sum_{i:g_i^a=k} \|\boldsymbol{x}_i^a - \bar{\boldsymbol{x}}_k^a\|^2/m_k$. With $\widehat{\sigma}^2 = \operatorname{tr}(\boldsymbol{S}_{\mathrm{lab}})/p$, the estimator is

$$\widehat{\gamma}_\mu = \frac{1}{\widehat{\sigma}^2} \lambda_{\max}\left(\widehat{\boldsymbol{C}}^{1/2}\widehat{\boldsymbol{G}}\widehat{\boldsymbol{C}}^{1/2}\right). \tag{5}$$

For each $\alpha \in [0,1]$, we estimate t_α as follows. Set $y_N = p/N$, $\widehat{d}_{0,\alpha} = \alpha/\alpha_0$, and $\widehat{d}_{1,\alpha} = (1-\alpha)/(1-\alpha_0)$. Let $\widehat{s}_\alpha$ be the unique solution in $\big(0, 1/\{y_N \max(\widehat{d}_{0,\alpha}, \widehat{d}_{1,\alpha})\}\big)$ to

$$\frac{1}{\widehat{s}_\alpha^2} = y_N \left\{ \frac{\alpha_0 \widehat{d}_{0,\alpha}^2}{(1 - y_N \widehat{d}_{0,\alpha}\widehat{s}_\alpha)^2} + \frac{(1-\alpha_0)\widehat{d}_{1,\alpha}^2}{(1 - y_N \widehat{d}_{1,\alpha}\widehat{s}_\alpha)^2} \right\}. \tag{6}$$

Define $\widehat{t}_\alpha = 1/\widehat{s}_\alpha - y_N \widehat{d}_{1,\alpha}$. The empirical safe set is

$$\widehat{\mathcal{A}}_N = \left\{ \alpha \in [0,1] : (1-\alpha)\widehat{\gamma}_\mu \le \widehat{t}_\alpha \right\}. \tag{7}$$

Among the weights in the empirical safe set, we choose the one closest to the natural weight α_0:

$$\alpha_* = \arg\min_{\alpha \in \widehat{\mathcal{A}}_N} |\alpha - \alpha_0|. \tag{8}$$

As shown in Theorem 2, this choice asymptotically minimizes the factor detectability threshold over $\widehat{\mathcal{A}}_N$. When the factor loading space is detectable, it also minimizes the asymptotic factor loading space estimation error over this set.

To determine the number of directions retained at a given weight α, we estimate the right edge of the spectral bulk of $\boldsymbol{S}_\alpha/\sigma^2$ by

$$\widehat{\lambda}_{+,\alpha} = \frac{1}{\widehat{s}_\alpha} + \frac{\alpha}{1 - y_N\, \widehat{d}_{0,\alpha}\widehat{s}_\alpha} + \frac{1-\alpha}{1 - y_N\, \widehat{d}_{1,\alpha}\widehat{s}_\alpha}. \tag{9}$$

We then set

$$\widehat{r}_\alpha = \#\left\{ j : \frac{1}{\widehat{\sigma}^2}\lambda_j(\boldsymbol{S}_\alpha) > \widehat{\lambda}_{+,\alpha} + \frac{1}{\log p} \right\}, \tag{10}$$

and define the orthogonal projector onto the estimated factor loading space by $\widehat{\boldsymbol{P}}_\alpha = \sum_{a=1}^{\widehat{r}_\alpha} \widehat{\boldsymbol{q}}_a^\alpha (\widehat{\boldsymbol{q}}_a^\alpha)^\top$, where the $\widehat{\boldsymbol{q}}_a^\alpha$ are orthonormal eigenvectors associated with $\lambda_a(\boldsymbol{S}_\alpha)$.

3 Theoretical guarantees

Assumption 1. *$p, m, n \to \infty$, $y_n = p/n \to y_1 \in (0,\infty)$ and $y_m = p/m \to y_2 \in (0,\infty)$.*

Under Assumption 1, $y_N = p/N \to y = y_1 y_2/(y_1+y_2)$ and $\alpha_0 = m/N \to \beta = y_1/(y_1+y_2)$.

Assumption 2. *The observations $\{\boldsymbol{x}_i\}_{i=1}^N$ are i.i.d. and follow* (1)*. The memberships, factor scores, idiosyncratic noise vectors, and the set of labeled indices are mutually independent. For some $\varepsilon > 0$, the components of $\boldsymbol{f}_i$ are independent, with $\mathbb{E}f_{ij} = 0$, $\mathbb{E}f_{ij}^2 = \theta_j$, and $\mathbb{E}|f_{ij}|^{4+\varepsilon} < \infty$. The entries of $\boldsymbol{z}_i$ are i.i.d., with $\mathbb{E}z_{ij} = 0$, $\mathbb{E}z_{ij}^2 = 1$, and $\mathbb{E}|z_{ij}|^{4+\varepsilon} < \infty$.*

Under the moment condition in Assumption 2, the results of Ding & Yang (2021) give a polynomial convergence rate $N^{-\varepsilon}$ near the spectral edge. Accordingly, the $1/\log p$ buffer applies for any $\varepsilon > 0$ allowed by the assumption.

Assumption 3. *The between-population covariance defined in* (3) *satisfies*

$$\|\boldsymbol{\Sigma}_\mu\| = O(1), \qquad \|\boldsymbol{P}_s \boldsymbol{\Sigma}_\mu\| = o\!\left(\frac{1}{\log p}\right),$$

where $\boldsymbol{P}_s = \boldsymbol{U}\boldsymbol{U}^\top$ is the orthogonal projector onto the factor loading space.

Remark 1. *The asymptotic orthogonality condition in Assumption 3 reflects a setting where between-population variation does not carry information about the shared factor loading space. This is consistent with Li et al. (2026), where the mean-shift directions are randomized independently of the underlying covariance structure. Under their random-direction assumption, this condition holds asymptotically for any fixed finite-dimensional factor loading space.*

Let $\gamma_\mu = \lambda_{\max}(\boldsymbol{\Sigma}_\mu)$ denote the strength of the between-population covariance. For $\alpha \in [0,1]$, define $d_{0,\alpha} = \alpha/\beta$ and $d_{1,\alpha} = (1-\alpha)/(1-\beta)$. The corresponding limiting distribution of the observation weights is

$$H_\alpha = \beta \delta_{d_{0,\alpha}} + (1-\beta)\delta_{d_{1,\alpha}},$$

where δ_x denotes the Dirac measure at x.

For each $\alpha \in [0,1]$, let F_α denote the limiting empirical spectral distribution of $\boldsymbol{S}_\alpha$. Its Stieltjes transform m_α is characterized by

$$z = -\frac{1}{m_\alpha(z)} + \int \frac{t}{1 + ytm_\alpha(z)}\, dH_\alpha(t), \qquad z, m_\alpha(z) \in \mathbb{C}^+. \tag{11}$$

Denote the right edge of $\mathrm{supp}(F_\alpha)$ by $\lambda_{+,\alpha}$ and set $s_\alpha = -m_\alpha(\lambda_{+,\alpha})$. The between-population separation threshold and factor detectability threshold are

$$t_\alpha = \frac{1}{s_\alpha} - yd_{1,\alpha}, \qquad \vartheta_\alpha = \frac{1}{\lambda_{+,\alpha} s_\alpha - 1}.$$

Lemma 1 in the Appendix establishes the consistency of $\widehat{\gamma}_\mu$, $\widehat{s}_\alpha$, $\widehat{t}_\alpha$, and $\widehat{\lambda}_{+,\alpha}$.

For $\theta > \vartheta_\alpha$, let $\rho_\alpha(\theta)$ denote the unique solution in $(\lambda_{+,\alpha}, \infty)$ to

$$1 + \rho_\alpha(\theta) m_\alpha(\rho_\alpha(\theta))\, \frac{\theta}{1+\theta} = 0. \tag{12}$$

Define

$$\kappa_\alpha(\theta) = \frac{m_\alpha^2(\rho_\alpha(\theta))}{(1+\theta)m_\alpha'(\rho_\alpha(\theta)) - \theta m_\alpha^2(\rho_\alpha(\theta))}.$$

Here, $\rho_\alpha(\theta)$ is the limiting location of the corresponding separated sample eigenvalue. The quantity $\kappa_\alpha(\theta)$ is the limiting squared alignment between the sample eigenvector and the population factor loading direction.

Theorem 1. *Under Assumptions 1, 2, and 3, let $\alpha \in \widehat{\mathcal{A}}_N$ satisfy $\theta_r > \vartheta_\alpha$. Then $\widehat{r}_\alpha = r$ with probability tending to one, and*

$$\left\| \widehat{\boldsymbol{P}}_\alpha - \boldsymbol{P}_s \right\|_F^2 - 2\sum_{a=1}^{r} (1 - \kappa_\alpha(\theta_a)) \xrightarrow{P} 0.$$

Theorem 2. *Under Assumptions 1, 2, and 3, let α_* be selected by* (8)*. Then*

$$\vartheta_{\alpha_*} - \min_{\alpha \in \widehat{\mathcal{A}}_N} \vartheta_\alpha \xrightarrow{P} 0.$$

For any $\alpha \in \widehat{\mathcal{A}}_N$ satisfying $\theta_r > \vartheta_\alpha$,

$$\left\| \widehat{\boldsymbol{P}}_{\alpha_*} - \boldsymbol{P}_s \right\|_F^2 \le \left\| \widehat{\boldsymbol{P}}_\alpha - \boldsymbol{P}_s \right\|_F^2 + o_p(1).$$

Moreover, if $\theta_r > \vartheta_{\alpha_0}$, then, on the event $\{\alpha_0 \in \widehat{\mathcal{A}}_N\}$,

$$\left\| \widehat{\boldsymbol{P}}_{\alpha_*} - \boldsymbol{P}_s \right\|_F^2 - \left\| \widehat{\boldsymbol{P}}_{\mathrm{or}} - \boldsymbol{P}_s \right\|_F^2 \xrightarrow{P} 0,$$

where $\widehat{\boldsymbol{P}}_{\mathrm{or}}$ is the orthogonal projector onto the span of the leading r eigenvectors of $\boldsymbol{S}_{\mathrm{or}}$.

Remark 2. *Theorem 1 shows that safe weighting selects the correct number of factors and excludes additional components caused by between-population variation. Theorem 2 shows that the selected weight asymptotically minimizes the factor detectability threshold over the estimated safe set, allowing weaker factors to be detected. When the factor loading space is detectable, it also minimizes the asymptotic factor loading space estimation error over this set. This error matches that of oracle PCA when the natural weight α_0 is safe.*

We next characterize the factor loading space estimation error without the asymptotic orthogonality condition in Assumption 3. For a given weight α, define

$$a(z) = -zm_\alpha(z) - 1, \qquad b(z) = \frac{-(1-\alpha)m_\alpha(z)}{1 + yd_{1,\alpha}m_\alpha(z)},$$

and

$$\boldsymbol{K}(z) = \boldsymbol{I}_p - a(z)\boldsymbol{U}\boldsymbol{\Theta}\boldsymbol{U}^\top - b(z)\boldsymbol{\Sigma}_\mu.$$

Let $\mathcal{Z}$ be the set of distinct values among the r largest real roots of $\det \boldsymbol{K}(z) = 0$ above $\lambda_{+,\alpha}$, where roots are counted with multiplicity. For each $\rho \in \mathcal{Z}$, let $\boldsymbol{V}_\rho$ have orthonormal columns spanning $\ker \boldsymbol{K}(\rho)$, and define

$$c_\rho = -m_\alpha(\rho)\, \mathrm{tr}\left[\left\{ \boldsymbol{V}_\rho^\top \boldsymbol{K}'(\rho) \boldsymbol{V}_\rho \right\}^{-1} \boldsymbol{V}_\rho^\top \boldsymbol{P}_s \boldsymbol{V}_\rho \right].$$

Theorem 3. *Under Assumptions 1 and 2, suppose* $\|\boldsymbol{\Sigma}_\mu\| = O(1)$. *Let* α_* *be selected by* (8). *Then*

$$\vartheta_{\alpha_*} - \min_{\alpha \in \widehat{\mathcal{A}}_N} \vartheta_\alpha \xrightarrow{P} 0.$$

For any $\alpha \in \widehat{\mathcal{A}}_N$ *satisfying* $\theta_r > \vartheta_\alpha$, $\widehat{r}_\alpha = r$ *with probability tending to one, and*

$$\left\| \widehat{\boldsymbol{P}}_\alpha - \boldsymbol{P}_s \right\|_F^2 - 2\left(r - \sum_{\rho \in \mathcal{Z}} c_\rho \right) \xrightarrow{P} 0.$$

Remark 3. *Theorem 3 removes the asymptotic orthogonality requirement from the factor number selection result. The selected weight still asymptotically minimizes the factor detectability threshold over the estimated safe set. Without the orthogonality condition, the factor loading space estimation error is characterized through the roots of* $\det \boldsymbol{K}(z) = 0$*, which account for the interaction between the factor loading space and between-population variation.*

4 Numerical studies

4.1 Simulation

We evaluate the proposed weight selection procedure through simulations. Set $p = 400$, $n = 200$, $m \in \{400, 600, 800\}$, and consider $K = 4$ populations with equal mixing proportions $\pi_k = 1/4$. We take $\boldsymbol{U} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ and $\boldsymbol{\Theta} = \mathrm{diag}(5, 4, 3)$. The factor scores and idiosyncratic noise vectors are generated independently from $N(\boldsymbol{0}, \boldsymbol{\Theta})$ and $N(\boldsymbol{0}, \boldsymbol{I}_p)$, respectively. For each replication, we independently draw an orthonormal triple $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ uniformly at random. For $\delta \in \{1.8, 2.2, 3.5, 5\}$, we set

$$\boldsymbol{\mu}_1 = \sqrt{\delta}\,(\boldsymbol{w}_1 + \boldsymbol{w}_2 + \boldsymbol{w}_3), \qquad \boldsymbol{\mu}_2 = \sqrt{\delta}\,(\boldsymbol{w}_1 - \boldsymbol{w}_2 - \boldsymbol{w}_3),$$
$$\boldsymbol{\mu}_3 = \sqrt{\delta}\,(-\boldsymbol{w}_1 + \boldsymbol{w}_2 - \boldsymbol{w}_3), \quad \boldsymbol{\mu}_4 = \sqrt{\delta}\,(-\boldsymbol{w}_1 - \boldsymbol{w}_2 + \boldsymbol{w}_3).$$

Thus, $\boldsymbol{\Sigma}_\mu = \delta \sum_{j=1}^3 \boldsymbol{w}_j \boldsymbol{w}_j^\top$, so the between-population covariance has strength $\gamma_\mu = \delta$. We vary α from 0 to 1 in increments of 0.02. For each α, we measure the factor loading space estimation error by $\|\widehat{\boldsymbol{P}}_\alpha - \boldsymbol{P}_s\|_F^2$.

The solid curves in Figure 1 show the average error over 100 independent replications. The background shading indicates the proportion of replications in which each weight belongs to the empirical safe set, with darker shades corresponding to larger proportions. The stars mark the median selected weight α_*. As a benchmark, the horizontal dashed lines show the average factor loading space estimation error of oracle PCA, which uses the true population memberships for within-population centering of all $N = m + n$ observations and the true factor dimension $r = 3$.

Figure 1 shows a consistent pattern across the three labeled observation sizes. As δ increases, the empirical safe region shifts toward larger values of α. For $\delta = 1.8$ and 2.2, the natural weight α_0 is safe with high probability, and the selected weight remains near α_0. For $\delta = 3.5$ and 5, α_0 is typically unsafe, and the selected weight lies near the lower boundary of the empirical safe region. The sharp drop in factor loading space estimation error near this boundary is consistent with the exclusion of additional separated eigenvalues caused by between-population variation. Beyond the selected weight, the error increases as the unlabeled observations receive less weight. Across all settings, the selected weight achieves the lowest observed error, close to the oracle benchmark. Larger labeled sample sizes reduce both the oracle benchmark and the minimum estimation error.

4.2 Single-cell RNA perturbation data

We evaluate weighted hybrid PCA using two A549 drug perturbation subsets from the sci-Plex single-cell RNA sequencing experiments of Srivatsan et al. (2020): Nutlin-3a and Pirarubicin. We use 2,000 Nutlin-3a cells from six batches and 615 Pirarubicin cells from four batches. Both subsets contain cells exposed to multiple drug doses and controls. We treat each batch as a population and aim to estimate the factor loading space shared across populations. Nutlin-3a activates p53 (Vassilev et al., 2004), and the original study identified p53 signaling as a major response to Nutlin-3a. Pirarubicin has also been linked to p53 regulation (Tang et al., 2021). We therefore evaluate

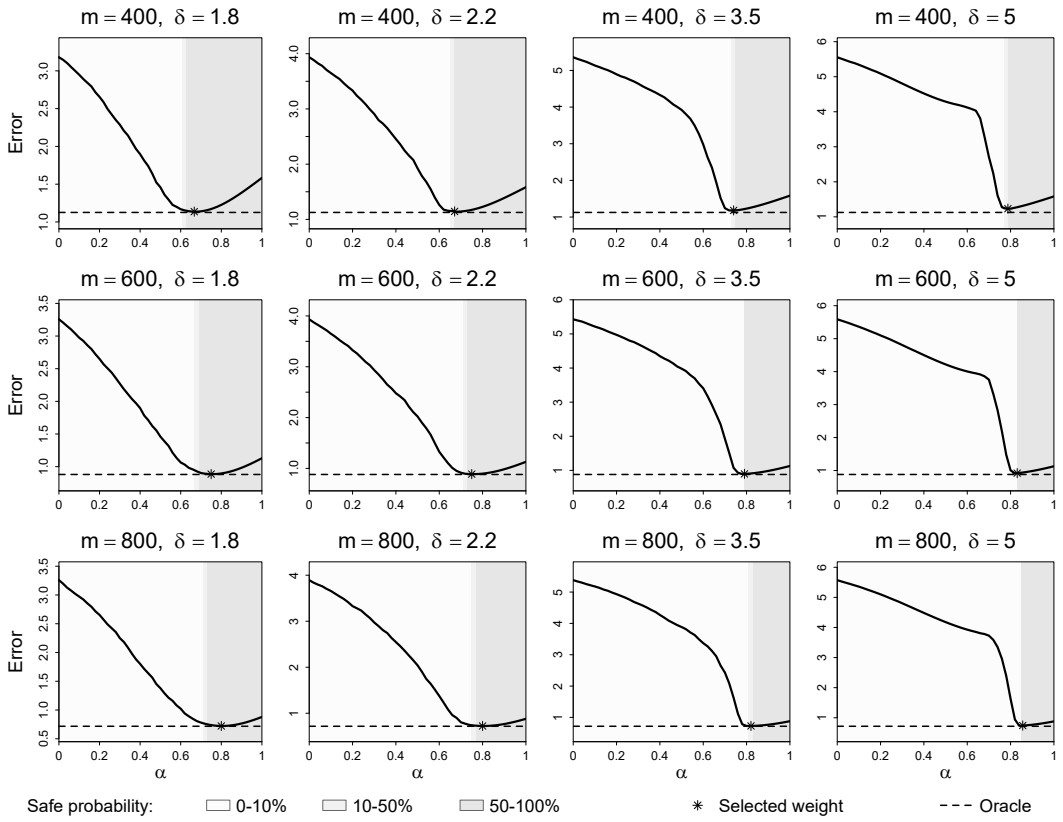

Figure 1: Factor loading space estimation error of weighted hybrid PCA for different numbers of labeled observations and between-population covariance strengths. The background shading indicates the proportion of replications in which each weight belongs to the empirical safe set, the stars mark the median selected weights, and the horizontal dashed lines show the oracle benchmark.

the methods in both datasets according to the common within-population variation captured by the estimated factor loading space and the association of the selected principal components with the p53 response.

For each dataset, we reserve about 20% of the cells as a fixed evaluation set. From the remaining cells, we sample 200 cells for training and reveal population memberships for 10% of the training cells within each batch. This sampling and labeling process is repeated 100 times. Counts are normalized to counts per million (CPM) and transformed as $\log_2(\mathrm{CPM}+1)$. In each repetition, we filter genes and select the top 40% most variable genes using only the training cells. All methods use the same selected genes.

For both datasets, we compare labeled PCA at $\alpha = 1$, natural hybrid PCA at $\alpha = \alpha_0$, weighted hybrid PCA at $\alpha = \alpha_*$, and mean-shift PCA of Li et al. (2026) with its default settings. Labeled PCA and the two hybrid methods select the number of principal components using (10). We evaluate the selected principal components on the evaluation cells. We first center the evaluation cells within each batch and let $\boldsymbol{S}_{\mathrm{eval}}$ denote the resulting covariance matrix. For the estimated factor loading space with orthonormal basis $\widehat{\boldsymbol{U}}$, we define the held-out variance efficiency as

$$\mathrm{HVE} = \frac{\mathrm{tr}\left(\widehat{\boldsymbol{U}}^\top \boldsymbol{S}_{\mathrm{eval}} \widehat{\boldsymbol{U}}\right)}{\sum_{j=1}^{\mathrm{rank}(\widehat{\boldsymbol{U}})} \lambda_j(\boldsymbol{S}_{\mathrm{eval}})}.$$

This quantity measures the fraction of the maximum variance in $\boldsymbol{S}_{\mathrm{eval}}$ captured by a subspace of the same dimension. We also use PCGSE (Frost et al., 2015) with the Hallmark p53 gene set (Liberzon

Table 1: Results for the Nutlin-3a and Pirarubicin data. Values are reported as mean ± standard deviation over 100 repetitions.

Metric	Labeled PCA	Natural hybrid	Weighted hybrid	Mean-shift PCA
Nutlin-3a				
Weight	1.000	0.100	0.503 ± 0.059	–
Selected PCs	9.98 ± 1.00	37.43 ± 5.27	10.19 ± 0.99	3.09 ± 1.33
HVE	0.339 ± 0.055	0.431 ± 0.023	$\mathbf{0.505 \pm 0.032}$	0.117 ± 0.055
PCs with p53	4.86 ± 1.56	4.31 ± 1.69	$\mathbf{5.46 \pm 1.64}$	1.00 ± 0.75
Proportion with p53 (%)	48.63 ± 14.69	11.65 ± 4.85	$\mathbf{53.38 \pm 14.53}$	33.83 ± 27.06
Pirarubicin				
Weight	1.000	0.100	0.327 ± 0.050	–
Selected PCs	4.52 ± 0.54	11.23 ± 1.59	5.01 ± 0.58	4.39 ± 1.45
HVE	0.277 ± 0.044	0.372 ± 0.024	$\mathbf{0.450 \pm 0.027}$	0.102 ± 0.032
PCs with p53	0.84 ± 0.84	1.45 ± 0.64	$\mathbf{1.49 \pm 0.83}$	0.68 ± 0.76
Proportion with p53 (%)	18.68 ± 18.46	13.17 ± 6.18	$\mathbf{29.77 \pm 16.38}$	15.57 ± 18.18

et al., 2015) to measure the association of the selected principal components with the p53 response. A principal component is counted as p53-associated if its nominal p value is below 0.05.

Table 1 summarizes the results for both datasets. Weighted hybrid PCA achieves the highest HVE and gives the highest average number and proportion of p53-associated principal components in both datasets. For Pirarubicin, natural and weighted hybrid PCA identify nearly the same number of p53-associated components, but natural hybrid PCA selects many more components overall. This is consistent with our theory that the natural weight may retain additional directions caused by between-population variation. Mean-shift PCA, which relies on a random-direction assumption for the mean-shift directions (Li et al., 2026), shows comparatively weaker empirical performance in these experiments.

5 Conclusion

Weighted hybrid PCA uses observed population memberships to determine how much to rely on unlabeled observations. Under the model assumptions, the selected weight asymptotically minimizes the factor detectability threshold over the estimated safe set, allowing weaker factors to be detected. When the factor loading space is detectable, it also minimizes the asymptotic factor loading space estimation error over this set. Simulations support the gains in factor loading space estimation. On two A549 single-cell RNA perturbation datasets with p53-related expression responses, the method more efficiently captures the common within-population variation than the other methods and gives the highest average number and proportion of p53-associated principal components.

Reproducibility statement

Algorithm 1 and Section 2 specify the estimation and component selection procedures. Section 3 states the assumptions, and Appendix A provides the proofs. Section 4.1 describes the simulation settings. Section 4.2 identifies the data sources and describes the preprocessing, sampling scheme, and evaluation criteria for the empirical studies. The supplementary material provides the empirical data and code used to reproduce the reported experimental results.

AI use statement

ChatGPT was used for language polishing, assistance with experimental code development, and checking proofs for correctness and consistency. The authors take responsibility for all theoretical claims, experimental results, and conclusions.

REFERENCES

Jushan Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003.

Jinho Baik and Jack W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.

Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

Hanqin Cai, Jianfeng Cai, and Ke Wei. Accelerated alternating projections for robust principal component analysis. *Journal of Machine Learning Research*, 20(20):1–33, 2019.

Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):1–37, 2011.

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

Jianqing Fan, Yuan Liao, and Martina Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(4):603–680, 2013.

H. Robert Frost, Zhigang Li, and Jason H. Moore. Principal component gene set enrichment (PCGSE). *BioData Mining*, 8(1):25, 2015.

Arun Jambulapati, Jerry Li, and Kevin Tian. Robust sub-gaussian principal component analysis and width-independent schatten packing. *Advances in Neural Information Processing Systems*, 33: 15689–15701, 2020.

Ian T. Jolliffe and Jorge Cadima. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

Mengda Li, Zeng Li, and Jianfeng Yao. Mean-shift PCA by knockoff mean. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026.

Weiming Li and Junpeng Zhu. CLT for spiked eigenvalues of a sample covariance matrix from high-dimensional Gaussian mean mixtures. *Journal of Multivariate Analysis*, 193:105127, 2023.

Weiming Li, Zeng Li, and Junpeng Zhu. On spiked eigenvalues of a renormalized sample covariance matrix from multi-population. *arXiv preprint arXiv:2409.08715*, 2024.

Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. The molecular signatures database hallmark gene set collection. *Cell Systems*, 1(6): 417–425, 2015.

Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, 2007.

Debolina Paul, Saptarshi Chakraborty, and Swagatam Das. Robust principal component analysis: A median of means approach. *IEEE Transactions on Neural Networks and Learning Systems*, 35 (11):16788–16800, 2024.

Sanjay R. Srivatsan, José L. McFaline-Figueroa, Vijay Ramani, Lauren Saunders, Junyue Cao, Jonathan Packer, Hannah A. Pliner, Dana L. Jackson, Riza M. Daza, Lena Christiansen, Fan Zhang, Frank Steemers, Jay Shendure, and Cole Trapnell. Massively multiplex chemical transcriptomics at single-cell resolution. *Science*, 367(6473):45–51, 2020.

Jun Tang, Weiyan Peng, Yixiao Feng, Xin Le, Kang Wang, Qin Xiang, Lili Li, Yan Wang, Can Xu, Junhao Mu, Ke Xu, Ping Ji, Qian Tao, Ailong Huang, Chu-Xia Deng, Yong Lin, and Tingxiu Xiang. Cancer cells escape p53's tumor suppression through ablation of ZDHHC1-mediated p53 palmitoylation. *Oncogene*, 40(35):5416–5426, 2021.

Lyubomir T. Vassilev, Binh T. Vu, Bradford Graves, Daisy Carvajal, Frank Podlaski, Zoran Filipovic, Norman Kong, Ursula Kammlott, Christine Lukacs, Christian Klein, Nader Fotouhi, and Emily A. Liu. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*, 303(5659):844–848, 2004.

Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Robust PCA via outlier pursuit. *IEEE Transactions on Information Theory*, 58(5):3047–3064, 2012.

A PROOFS FOR WEIGHTED HYBRID PCA

Lemma 1. *Under Assumptions 1 and 2, suppose* $\|\mathbf{\Sigma}_\mu\| = O(1)$. *Then*

$$\widehat{\sigma}^2 - 1 = O_p(p^{-1}) \qquad \widehat{\gamma}_\mu - \gamma_\mu = o_p\left(\frac{1}{\log p}\right).$$

Moreover, uniformly over $\alpha \in [0,1]$,

$$\sup_{\alpha\in[0,1]} |\widehat{s}_\alpha - s_\alpha| \longrightarrow 0, \qquad \sup_{\alpha\in[0,1]} \left|\widehat{t}_\alpha - t_\alpha\right| \longrightarrow 0, \qquad \sup_{\alpha\in[0,1]} \left|\widehat{\lambda}_{+,\alpha} - \lambda_{+,\alpha}\right| \longrightarrow 0.$$

Finally, $\widehat{\mathcal{A}}_N$ *is a closed interval with right endpoint* 1.

Proof of Lemma 1. For $k = 1, \ldots, K$, set $\mathcal{I}_k = \{i : g_i^a = k\}$, and write

$$\boldsymbol{M} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K), \qquad \boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)^\top, \qquad \boldsymbol{C} = \mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^\top.$$

Define

$$\boldsymbol{B}_\mu = \boldsymbol{C}^{1/2}\boldsymbol{M}^\top\boldsymbol{M}\boldsymbol{C}^{1/2}, \qquad \widehat{\boldsymbol{B}}_\mu = \frac{1}{\widehat{\sigma}^2}\widehat{\boldsymbol{C}}^{1/2}\widehat{\boldsymbol{G}}\widehat{\boldsymbol{C}}^{1/2}.$$

Since

$$\mathbf{\Sigma}_\mu = \boldsymbol{M}\boldsymbol{C}\boldsymbol{M}^\top,$$

the nonzero eigenvalues of $\boldsymbol{B}_\mu$ and $\mathbf{\Sigma}_\mu$ coincide. It therefore suffices to bound $\|\widehat{\boldsymbol{B}}_\mu - \boldsymbol{B}_\mu\|$.

Both matrices are invariant under a common translation of the population means. We may therefore assume $\bar{\boldsymbol{\mu}} = \mathbf{0}$, under which

$$\mathbf{\Sigma}_\mu = \sum_{k=1}^K \pi_k \boldsymbol{\mu}_k \boldsymbol{\mu}_k^\top, \qquad \|\boldsymbol{\mu}_k\|^2 \le \frac{\|\mathbf{\Sigma}_\mu\|}{\pi_k} = O(1).$$

By the law of large numbers,

$$\frac{m_k}{m} \xrightarrow{\text{a.s.}} \pi_k, \qquad k = 1, \ldots, K,$$

so $m_k \asymp m$ with probability tending to one. We condition on the labeled population memberships throughout the remaining argument; all expectations below are conditional expectations.

For $k = 1, \ldots, K$, write

$$\boldsymbol{y}_i = \boldsymbol{U}\boldsymbol{f}_i + \boldsymbol{z}_i, \qquad \bar{\boldsymbol{y}}_k = \frac{1}{m_k}\sum_{i\in\mathcal{I}_k} \boldsymbol{y}_i.$$

Then

$$\mathbb{E}\boldsymbol{y}_i = \mathbf{0}, \qquad \mathbb{E}\boldsymbol{y}_i\boldsymbol{y}_i^\top = \mathbf{\Sigma}.$$

Since r is fixed and $\|\mathbf{\Theta}\| = O(1)$, Assumption 2 also gives

$$\mathrm{tr}(\mathbf{\Sigma}^2) = O(p), \qquad \mathbb{E}\|\boldsymbol{y}_i\|^4 = O(p^2), \qquad \mathrm{Var}\big(\|\boldsymbol{y}_i\|^2\big) = O(p).$$

For $k \neq \ell$,

$$\bar{\boldsymbol{x}}_k^a = \boldsymbol{\mu}_k + \bar{\boldsymbol{y}}_k,$$

and hence

$$\widehat{G}_{k\ell} - \boldsymbol{\mu}_k^\top\boldsymbol{\mu}_\ell = \boldsymbol{\mu}_k^\top\bar{\boldsymbol{y}}_\ell + \bar{\boldsymbol{y}}_k^\top\boldsymbol{\mu}_\ell + \bar{\boldsymbol{y}}_k^\top\bar{\boldsymbol{y}}_\ell.$$

The first two terms are $O_p(m^{-1/2})$, while independence gives

$$\mathbb{E}\left|\bar{\boldsymbol{y}}_k^\top\bar{\boldsymbol{y}}_\ell\right|^2 = \frac{1}{m_k m_\ell}\mathrm{tr}(\mathbf{\Sigma}^2) = O\left(\frac{p}{m^2}\right).$$

Hence

$$\widehat{G}_{k\ell} - \boldsymbol{\mu}_k^\top\boldsymbol{\mu}_\ell = O_p\left(m^{-1/2} + \frac{\sqrt{p}}{m}\right).$$

For the diagonal entries,

$$\widehat{G}_{kk} - \boldsymbol{\mu}_k^\top \boldsymbol{\mu}_k = 2\boldsymbol{\mu}_k^\top \bar{\boldsymbol{y}}_k + \left(1 + \frac{1}{m_k}\right) \bar{\boldsymbol{y}}_k^\top \bar{\boldsymbol{y}}_k - \frac{1}{m_k^2} \sum_{i \in \mathcal{I}_k} \|\boldsymbol{y}_i\|^2.$$

The first term is $O_p(m^{-1/2})$. The sum of the remaining two terms has expectation

$$\frac{\operatorname{tr}(\boldsymbol{\Sigma})}{m_k^2} = O\left(\frac{p}{m^2}\right).$$

Moreover, independence and the preceding moment bounds give

$$\operatorname{Var}\left(\bar{\boldsymbol{y}}_k^\top \bar{\boldsymbol{y}}_k\right) = O\left(\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{m_k^2} + \frac{\mathbb{E}\|\boldsymbol{y}_i\|^4}{m_k^3}\right) = O\left(\frac{p}{m^2}\right),$$

and

$$\operatorname{Var}\left(\frac{1}{m_k^2} \sum_{i \in \mathcal{I}_k} \|\boldsymbol{y}_i\|^2\right) = O\left(\frac{p}{m^3}\right).$$

Hence the variance of the sum of the remaining two terms is $O(p/m^2)$. Hence

$$\widehat{G}_{kk} - \boldsymbol{\mu}_k^\top \boldsymbol{\mu}_k = O_p\left(m^{-1/2} + \frac{\sqrt{p}}{m}\right).$$

Therefore, since K is fixed,

$$\|\widehat{\boldsymbol{G}} - \boldsymbol{M}^\top \boldsymbol{M}\| = O_p\left(m^{-1/2} + \frac{\sqrt{p}}{m}\right).$$

Moreover, since K is fixed,

$$\|\widehat{\boldsymbol{C}} - \boldsymbol{C}\| = O_p(m^{-1/2}).$$

Let $\mathbf{1}_d$ denote the all-one vector in $\mathbb{R}^d$. The matrices $\widehat{\boldsymbol{C}}$ and $\boldsymbol{C}$ have the common null space $\operatorname{span}\{\mathbf{1}_K\}$, and the nonzero eigenvalues of $\boldsymbol{C}$ are bounded away from zero. Hence

$$\|\widehat{\boldsymbol{C}}^{1/2} - \boldsymbol{C}^{1/2}\| = O_p(m^{-1/2}).$$

Also,

$$\widehat{\sigma}^2 = \frac{1}{mp} \sum_{i=1}^{m} \|\boldsymbol{y}_i\|^2 - \frac{1}{p} \sum_{k=1}^{K} \frac{m_k}{m} \|\bar{\boldsymbol{y}}_k\|^2.$$

Since $\operatorname{tr}(\boldsymbol{\Sigma}) = p + O(1)$ and $m \asymp p$, the first term is $1 + O_p(p^{-1})$ by the variance bound above. The second term is nonnegative and has conditional expectation $K \operatorname{tr}(\boldsymbol{\Sigma})/(mp) = O(p^{-1})$. Hence $\widehat{\sigma}^2 - 1 = O_p(p^{-1})$.

Consequently, since $\|\boldsymbol{M}^\top \boldsymbol{M}\| = O(1)$ and $\widehat{\sigma}^2 - 1 = O_p(p^{-1})$,

$$\|\widehat{\boldsymbol{B}}_\mu - \boldsymbol{B}_\mu\| = O_p\left(m^{-1/2} + \frac{\sqrt{p}}{m}\right).$$

Under Assumption 1, $m \asymp p$. Hence Weyl's inequality gives

$$\widehat{\gamma}_\mu - \gamma_\mu = O_p(p^{-1/2}) = o_p\left(\frac{1}{\log p}\right),$$

which proves the stated rates.

We next establish the uniform consistency of the edge quantities. Define the edge maps

$$x_\alpha(u) = -\frac{1}{u} + \frac{\alpha}{1 + y d_{0,\alpha} u} + \frac{1 - \alpha}{1 + y d_{1,\alpha} u}$$

and

$$\widehat{x}_\alpha(u) = -\frac{1}{u} + \frac{\alpha}{1 + y_N \widehat{d}_{0,\alpha} u} + \frac{1 - \alpha}{1 + y_N \widehat{d}_{1,\alpha} u}. \tag{13}$$

Since $\alpha_0 \to \beta \in (0,1)$, α_0 and $1-\alpha_0$ are bounded away from zero, $\widehat{d}_{0,\alpha}$ and $\widehat{d}_{1,\alpha}$ are uniformly bounded over $\alpha \in [0,1]$, and

$$\sup_{\alpha\in[0,1]} \left|\widehat{d}_{j,\alpha} - d_{j,\alpha}\right| \longrightarrow 0, \qquad j = 0, 1.$$

On the admissible interval

$$\left(-\frac{1}{y_N \max\{\widehat{d}_{0,\alpha}, \widehat{d}_{1,\alpha}\}}, 0\right),$$

direct differentiation gives $\widehat{x}''_\alpha(u) > 0$, while $\widehat{x}'_\alpha(u)$ tends to $-\infty$ and $+\infty$ at the left and right endpoints, respectively. Hence $\widehat{x}'_\alpha$ has a unique zero $-\widehat{s}_\alpha$, which is equivalent to (6).

The same argument applies to x_α. To obtain uniform regularity, set

$$M_\alpha = \max\{d_{0,\alpha}, d_{1,\alpha}\}.$$

The edge equation is

$$1 = y s_\alpha^2 \int \frac{t^2}{(1 - y t s_\alpha)^2}\, dH_\alpha(t).$$

Since y, $d_{0,\alpha}$, and $d_{1,\alpha}$ are uniformly bounded and H_α has a positive mass on a positive atom uniformly in α, this identity implies that s_α is bounded above and away from zero uniformly in α. Indeed, the upper bound follows from

$$1 \geq c s_\alpha^2,$$

whereas $s_\alpha \to 0$ along any subsequence would make the right-hand side of the edge equation tend to zero, a contradiction.

The mass assigned by H_α to the atom M_α is bounded away from zero uniformly in α. Hence

$$1 \geq c y M_\alpha^2 s_\alpha^2 \left(1 - y M_\alpha s_\alpha\right)^{-2},$$

and the preceding lower bound on s_α, together with $M_\alpha \geq 1$, yields

$$1 - y M_\alpha s_\alpha \geq c$$

uniformly over $\alpha \in [0,1]$. The same argument applies to $\widehat{s}_\alpha$ for all sufficiently large N.

Since $y_N \to y$, the preceding bounds and the uniform convergence of $\widehat{d}_{j,\alpha}$ imply

$$\sup_{\alpha\in[0,1]} |\widehat{x}'_\alpha(-s_\alpha) - x'_\alpha(-s_\alpha)| \longrightarrow 0.$$

Furthermore, $\widehat{x}''_\alpha$ is uniformly bounded away from zero between $-s_\alpha$ and $-\widehat{s}_\alpha$. Since $x'_\alpha(-s_\alpha) = \widehat{x}'_\alpha(-\widehat{s}_\alpha) = 0$, the mean value theorem yields

$$\sup_{\alpha\in[0,1]} |\widehat{s}_\alpha - s_\alpha| \longrightarrow 0.$$

It follows directly from the definitions that

$$\sup_{\alpha\in[0,1]} \left|\widehat{\lambda}_{+,\alpha} - \lambda_{+,\alpha}\right| \longrightarrow 0, \qquad \sup_{\alpha\in[0,1]} \left|\widehat{t}_\alpha - t_\alpha\right| \longrightarrow 0.$$

It remains to characterize the safe-set boundary. Set $\widehat{a}_\alpha = \alpha\widehat{s}_\alpha/(1 - \alpha y_m \widehat{s}_\alpha)$ and $\widehat{b}_\alpha = (1-\alpha)\widehat{s}_\alpha/\{1 - (1-\alpha) y_n \widehat{s}_\alpha\}$. Since $y_N \widehat{d}_{0,\alpha} = \alpha y_m$ and $y_N \widehat{d}_{1,\alpha} = (1-\alpha) y_n$, (6) gives

$$y_m \widehat{a}_\alpha^2 + y_n \widehat{b}_\alpha^2 = 1.$$

For $\alpha \in (0,1)$, the definitions also give

$$\frac{\alpha}{1-\alpha} = \frac{\widehat{a}_\alpha(1 + y_n \widehat{b}_\alpha)}{\widehat{b}_\alpha(1 + y_m \widehat{a}_\alpha)}.$$

Along the preceding ellipse, $\widehat{a}_\alpha = \{(1 - y_n \widehat{b}_\alpha^2)/y_m\}^{1/2}$, and the right-hand side is strictly decreasing in $\widehat{b}_\alpha$. Hence $\widehat{b}_\alpha$ is continuous and strictly decreasing in α, with $\widehat{b}_0 = 1/\sqrt{y_n}$ and $\widehat{b}_1 = 0$.

For $\alpha < 1$, $\widehat{t}_\alpha = (1-\alpha)/\widehat{b}_\alpha$, while the safe condition holds automatically at $\alpha = 1$. Therefore

$$\widehat{\mathcal{A}}_N = \left\{\alpha \in [0,1] : \widehat{\gamma}_\mu \widehat{b}_\alpha \leq 1\right\}.$$

Since $\widehat{b}_\alpha$ is continuous and strictly decreasing, $\widehat{\mathcal{A}}_N$ is a closed interval with right endpoint 1, which completes the proof. □

Proof of Theorem 1. Set $b_p = 1/\log p$. By the definition of $\widehat{\mathcal{A}}_N$ and Lemma 1,

$$\sup_{\alpha\in\widehat{\mathcal{A}}_N} \left[(1-\alpha)\gamma_\mu - \widehat{t}_\alpha\right]_+ \le |\widehat{\gamma}_\mu - \gamma_\mu| = o_p(b_p).$$

With probability tending to one, every population is represented among the labeled observations. On this event, after reordering the labeled observations by population, let

$$\boldsymbol{\Phi}_{\mathrm{lab}} = \mathrm{diag}\left(\boldsymbol{I}_{m_1} - \frac{1}{m_1}\mathbf{1}_{m_1}\mathbf{1}_{m_1}^\top, \ldots, \boldsymbol{I}_{m_K} - \frac{1}{m_K}\mathbf{1}_{m_K}\mathbf{1}_{m_K}^\top\right), \qquad \boldsymbol{\Phi}_{\mathrm{mix}} = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top.$$

Define

$$\boldsymbol{B}_{\alpha,N} = \mathrm{diag}\left(\widehat{d}_{0,\alpha}\boldsymbol{\Phi}_{\mathrm{lab}}, \widehat{d}_{1,\alpha}\boldsymbol{\Phi}_{\mathrm{mix}}\right).$$

Then

$$F^{\boldsymbol{B}_{\alpha,N}} = \frac{m-K}{N}\delta_{\widehat{d}_{0,\alpha}} + \frac{n-1}{N}\delta_{\widehat{d}_{1,\alpha}} + \frac{K+1}{N}\delta_0 \Rightarrow H_\alpha.$$

Write

$$\boldsymbol{\mathcal{F}}_N = (\boldsymbol{f}_1^a, \ldots, \boldsymbol{f}_m^a, \boldsymbol{f}_1^u, \ldots, \boldsymbol{f}_n^u), \qquad \boldsymbol{Z}_N = (\boldsymbol{z}_1^a, \ldots, \boldsymbol{z}_m^a, \boldsymbol{z}_1^u, \ldots, \boldsymbol{z}_n^u),$$

and

$$\boldsymbol{\mathcal{M}}_N = (\boldsymbol{\mu}_{g_1^a}, \ldots, \boldsymbol{\mu}_{g_m^a}, \boldsymbol{\mu}_{g_1^u}, \ldots, \boldsymbol{\mu}_{g_n^u}), \qquad \boldsymbol{X}_N = (\boldsymbol{x}_1^a, \ldots, \boldsymbol{x}_m^a, \boldsymbol{x}_1^u, \ldots, \boldsymbol{x}_n^u).$$

By (1),

$$\frac{1}{\sqrt{N}}\boldsymbol{X}_N\boldsymbol{B}_{\alpha,N}^{1/2} = \boldsymbol{Y}_{\alpha,N} + \boldsymbol{U}\boldsymbol{W}_{\alpha,N}^\top + \frac{1}{\sqrt{N}}\boldsymbol{\mathcal{M}}_N\boldsymbol{B}_{\alpha,N}^{1/2},$$

where

$$\boldsymbol{Y}_{\alpha,N} = \frac{1}{\sqrt{N}}\boldsymbol{Z}_N\boldsymbol{B}_{\alpha,N}^{1/2}, \qquad \boldsymbol{W}_{\alpha,N} = \frac{1}{\sqrt{N}}\boldsymbol{B}_{\alpha,N}^{1/2}\boldsymbol{\mathcal{F}}_N^\top.$$

Consequently,

$$\boldsymbol{S}_\alpha = \left(\boldsymbol{Y}_{\alpha,N} + \boldsymbol{U}\boldsymbol{W}_{\alpha,N}^\top + \frac{1}{\sqrt{N}}\boldsymbol{\mathcal{M}}_N\boldsymbol{B}_{\alpha,N}^{1/2}\right)\left(\boldsymbol{Y}_{\alpha,N} + \boldsymbol{U}\boldsymbol{W}_{\alpha,N}^\top + \frac{1}{\sqrt{N}}\boldsymbol{\mathcal{M}}_N\boldsymbol{B}_{\alpha,N}^{1/2}\right)^\top.$$

Since both centering projections annihilate a common translation, we may assume $\bar{\boldsymbol{\mu}} = \mathbf{0}$. Write

$$\boldsymbol{\mathcal{M}}_u = (\boldsymbol{\mu}_{g_1^u}, \ldots, \boldsymbol{\mu}_{g_n^u}).$$

The labeled mean block is annihilated by $\boldsymbol{\Phi}_{\mathrm{lab}}$, and hence

$$\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{M}}_N\boldsymbol{B}_{\alpha,N}^{1/2} = \sqrt{\widehat{d}_{1,\alpha}}\,\frac{1}{\sqrt{N}}\left(\mathbf{0}, \boldsymbol{\mathcal{M}}_u\boldsymbol{\Phi}_{\mathrm{mix}}\right).$$

Choose a singular-value decomposition

$$\frac{1}{\sqrt{N}}\left(\mathbf{0}, \boldsymbol{\mathcal{M}}_u\boldsymbol{\Phi}_{\mathrm{mix}}\right) = \boldsymbol{F}_{\mu,N}^{(0)}\boldsymbol{R}_{\mu,N}^\top$$

and set

$$\boldsymbol{F}_{\mu,N} = \sqrt{\widehat{d}_{1,\alpha}}\,\boldsymbol{F}_{\mu,N}^{(0)}.$$

Then

$$\frac{1}{\sqrt{N}}\boldsymbol{\mathcal{M}}_N\boldsymbol{B}_{\alpha,N}^{1/2} = \boldsymbol{F}_{\mu,N}\boldsymbol{R}_{\mu,N}^\top,$$

where $\boldsymbol{R}_{\mu,N}^\top\boldsymbol{R}_{\mu,N} = \boldsymbol{I}$, $\boldsymbol{F}_{\mu,N}^\top\boldsymbol{F}_{\mu,N}$ is diagonal, and

$$\boldsymbol{B}_{\alpha,N}\boldsymbol{R}_{\mu,N} = \widehat{d}_{1,\alpha}\boldsymbol{R}_{\mu,N}.$$

Moreover,

$$\boldsymbol{F}_{\mu,N}\boldsymbol{F}_{\mu,N}^\top = (1-\alpha)\frac{1}{n}\boldsymbol{\mathcal{M}}_u\boldsymbol{\Phi}_{\mathrm{mix}}\boldsymbol{\mathcal{M}}_u^\top,$$

and therefore

$$\sup_{\alpha\in[0,1]} \left\|\boldsymbol{F}_{\mu,N}\boldsymbol{F}_{\mu,N}^\top - (1-\alpha)\boldsymbol{\Sigma}_\mu\right\| = O_p(N^{-1/2}) = o_p(b_p).$$

Combining this bound with the preceding safe-set estimate gives

$$\sup_{\alpha\in\widehat{\mathcal{A}}_N}\left[\|\boldsymbol{F}_{\mu,N}\|^2-\widehat{t}_\alpha\right]_+=o_p(b_p).$$

By Assumption 3,

$$\sup_{\alpha\in[0,1]}\|\boldsymbol{U}^\top\boldsymbol{F}_{\mu,N}\|^2=o_p(b_p).$$

Furthermore, conditional on the population memberships,

$$\boldsymbol{W}_{\alpha,N}^\top\boldsymbol{W}_{\alpha,N}=\frac{1}{N}\boldsymbol{\mathcal{F}}_N\boldsymbol{B}_{\alpha,N}\boldsymbol{\mathcal{F}}_N^\top.$$

Since

$$\frac{1}{N}\operatorname{tr}(\boldsymbol{B}_{\alpha,N})=\frac{m-K}{N}\widehat{d}_{0,\alpha}+\frac{n-1}{N}\widehat{d}_{1,\alpha}=1+O(N^{-1})$$

uniformly over $\alpha\in[0,1]$, the independence of the entries of $\boldsymbol{\mathcal{F}}_N$, the boundedness of $\|\boldsymbol{B}_{\alpha,N}\|$, and the fixed dimension r give

$$\sup_{\alpha\in[0,1]}\left\|\boldsymbol{W}_{\alpha,N}^\top\boldsymbol{W}_{\alpha,N}-\boldsymbol{\Theta}\right\|=O_p(N^{-1/2}).$$

Here the uniformity directly because $\boldsymbol{B}_{\alpha,N}$ is affine in α.

Moreover,

$$\boldsymbol{B}_{\alpha,N}^{1/2}\boldsymbol{R}_{\mu,N}=\sqrt{\widehat{d}_{1,\alpha}}\,\boldsymbol{R}_{\mu,N},$$

and hence

$$\boldsymbol{W}_{\alpha,N}^\top\boldsymbol{R}_{\mu,N}=\frac{\sqrt{\widehat{d}_{1,\alpha}}}{\sqrt{N}}\boldsymbol{\mathcal{F}}_N\boldsymbol{R}_{\mu,N}.$$

Conditional on the memberships, $\boldsymbol{R}_{\mu,N}$ is independent of the factor scores and has orthonormal columns. Since r and K are fixed,

$$\sup_{\alpha\in[0,1]}\left\|\boldsymbol{W}_{\alpha,N}^\top\boldsymbol{R}_{\mu,N}\right\|=O_p(N^{-1/2}).$$

Finally,

$$\sup_{\alpha\neq\alpha'}\frac{\|\boldsymbol{B}_{\alpha,N}^{1/2}-\boldsymbol{B}_{\alpha',N}^{1/2}\|}{|\alpha-\alpha'|^{1/2}}=O(1).$$

Since

$$\frac{1}{\sqrt{N}}\|\boldsymbol{Z}_N\|=O_p(1),\qquad\frac{1}{\sqrt{N}}\|\boldsymbol{\mathcal{F}}_N\|=O_p(1),$$

we also have

$$\sup_{\alpha\neq\alpha'}\frac{\|\boldsymbol{Y}_{\alpha,N}-\boldsymbol{Y}_{\alpha',N}\|}{|\alpha-\alpha'|^{1/2}}=O_p(1),\qquad\sup_{\alpha\neq\alpha'}\frac{\|\boldsymbol{W}_{\alpha,N}-\boldsymbol{W}_{\alpha',N}\|}{|\alpha-\alpha'|^{1/2}}=O_p(1).$$

With the common choice of $\boldsymbol{R}_{\mu,N}$,

$$\sup_{\alpha\neq\alpha'}\frac{\|\boldsymbol{F}_{\mu,N}(\alpha)-\boldsymbol{F}_{\mu,N}(\alpha')\|}{|\alpha-\alpha'|^{1/2}}=O_p(1).$$

Set

$$\boldsymbol{T}_{\alpha,N}=\boldsymbol{Y}_{\alpha,N}\boldsymbol{Y}_{\alpha,N}^\top,\qquad\boldsymbol{G}_{\alpha,N}(z)=(\boldsymbol{T}_{\alpha,N}-z\boldsymbol{I}_p)^{-1},$$

and

$$\boldsymbol{Q}_{\alpha,N}(z)=-\boldsymbol{I}_N+\boldsymbol{Y}_{\alpha,N}^\top\boldsymbol{G}_{\alpha,N}(z)\boldsymbol{Y}_{\alpha,N}.$$

Let

$$\boldsymbol{C}_N=(\boldsymbol{U},\boldsymbol{F}_{\mu,N}),\qquad\boldsymbol{V}_N=(\boldsymbol{W}_{\alpha,N},\boldsymbol{R}_{\mu,N}).$$

Then

$$\frac{1}{\sqrt{N}}\boldsymbol{X}_N\boldsymbol{B}_{\alpha,N}^{1/2} = \boldsymbol{Y}_{\alpha,N} + \boldsymbol{C}_N\boldsymbol{V}_N^\top.$$

With

$$\boldsymbol{H}_N = (\boldsymbol{Y}_{\alpha,N}\boldsymbol{V}_N, \boldsymbol{C}_N), \qquad \boldsymbol{K}_N = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{V}_N^\top\boldsymbol{V}_N \end{pmatrix},$$

a direct expansion gives

$$\boldsymbol{S}_\alpha = \boldsymbol{T}_{\alpha,N} + \boldsymbol{H}_N\boldsymbol{K}_N\boldsymbol{H}_N^\top.$$

The matrix $\boldsymbol{K}_N$ is invertible, with

$$\boldsymbol{K}_N^{-1} = \begin{pmatrix} -\boldsymbol{V}_N^\top\boldsymbol{V}_N & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{0} \end{pmatrix}.$$

For $z \notin \operatorname{spec}(\boldsymbol{T}_{\alpha,N})$, define

$$\boldsymbol{\mathcal{L}}_{\alpha,N}(z) = \boldsymbol{K}_N^{-1} + \boldsymbol{H}_N^\top\boldsymbol{G}_{\alpha,N}(z)\boldsymbol{H}_N.$$

The matrix determinant lemma gives

$$\det(\boldsymbol{S}_\alpha - z\boldsymbol{I}_p) = \det(\boldsymbol{T}_{\alpha,N} - z\boldsymbol{I}_p)\det(\boldsymbol{K}_N)\det\boldsymbol{\mathcal{L}}_{\alpha,N}(z).$$

Using

$$\boldsymbol{T}_{\alpha,N}\boldsymbol{G}_{\alpha,N}(z) = \boldsymbol{I}_p + z\boldsymbol{G}_{\alpha,N}(z),$$

the kernel can equivalently be written as

$$\boldsymbol{\mathcal{L}}_{\alpha,N}(z) = \begin{pmatrix} \boldsymbol{V}_N^\top\boldsymbol{Q}_{\alpha,N}(z)\boldsymbol{V}_N & \boldsymbol{I} + \boldsymbol{V}_N^\top\boldsymbol{Y}_{\alpha,N}^\top\boldsymbol{G}_{\alpha,N}(z)\boldsymbol{C}_N \\ \boldsymbol{I} + \boldsymbol{C}_N^\top\boldsymbol{G}_{\alpha,N}(z)\boldsymbol{Y}_{\alpha,N}\boldsymbol{V}_N & \boldsymbol{C}_N^\top\boldsymbol{G}_{\alpha,N}(z)\boldsymbol{C}_N \end{pmatrix}.$$

For each deterministic $\alpha \in [0,1]$, let $m_{\alpha,N}^\circ(z)$ denote the deterministic Stieltjes transform associated with $F^{\boldsymbol{B}_{\alpha,N}}$, characterized by

$$z = -\frac{1}{m_{\alpha,N}^\circ(z)} + \frac{1}{N}\operatorname{tr}\left[\boldsymbol{B}_{\alpha,N}\left(\boldsymbol{I}_N + y_N m_{\alpha,N}^\circ(z)\boldsymbol{B}_{\alpha,N}\right)^{-1}\right].$$

Set

$$q_{\alpha,N}^\circ(z) = -\frac{1}{N}\operatorname{tr}\left[\boldsymbol{B}_{\alpha,N}\left(\boldsymbol{I}_N + y_N m_{\alpha,N}^\circ(z)\boldsymbol{B}_{\alpha,N}\right)^{-1}\right].$$

Then

$$q_{\alpha,N}^\circ(z) = -z - \frac{1}{m_{\alpha,N}^\circ(z)}.$$

Conditional on the population memberships and factor scores, $\boldsymbol{B}_{\alpha,N}$, $\boldsymbol{W}_{\alpha,N}$, and $\boldsymbol{R}_{\mu,N}$ are deterministic with respect to the noise matrix. The anisotropic local laws in Theorems S.3.9 and S.3.12 of Ding & Yang (2021), applied with $A = \boldsymbol{I}_p$, $B = \boldsymbol{B}_{\alpha,N}$, and $X = N^{-1/2}\boldsymbol{Z}_N$, together with the truncation argument in Section S.2 of its supplement, give

$$\boldsymbol{v}^\top\boldsymbol{G}_{\alpha,N}(z)\boldsymbol{w} = m_{\alpha,N}^\circ(z)\boldsymbol{v}^\top\boldsymbol{w} + o_p(1),$$

$$\boldsymbol{v}^\top\boldsymbol{G}_{\alpha,N}(z)\boldsymbol{Y}_{\alpha,N}\boldsymbol{r} = o_p(1),$$

and

$$\boldsymbol{r}^\top\boldsymbol{Q}_{\alpha,N}(z)\boldsymbol{s} = -\boldsymbol{r}^\top\left(\boldsymbol{I}_N + y_N m_{\alpha,N}^\circ(z)\boldsymbol{B}_{\alpha,N}\right)^{-1}\boldsymbol{s} + o_p(1)$$

uniformly on fixed compact spectral domains separated from the deterministic bulk.

Use a common entrywise truncation of $\boldsymbol{Z}_N$ for all α. The local-law bounds for the truncated matrix hold with arbitrarily high polynomial probability. A polynomial mesh in α, together with the resolvent identity and the preceding Hölder estimates, extends these estimates uniformly to $\alpha \in [0,1]$. Theorem S.3.12 gives errors $o_p(b_p^{1/2})$ on the bounded spectral domains used below, with real part at least b_p above the finite-N right edge.

We next evaluate the factor-score blocks. Conditional on the population memberships,

$$\boldsymbol{W}_{\alpha,N}^\top\left(\boldsymbol{I}_N + y_N m_{\alpha,N}^\circ(z)\boldsymbol{B}_{\alpha,N}\right)^{-1}\boldsymbol{W}_{\alpha,N}$$

is a fixed-dimensional quadratic form in the independent factor scores. The finite $4+\varepsilon$ moments and the boundedness of the deterministic matrix in this quadratic form give

$$\boldsymbol{W}_{\alpha,N}^{\top}\boldsymbol{Q}_{\alpha,N}(z)\boldsymbol{W}_{\alpha,N} = q_{\alpha,N}^{\circ}(z)\boldsymbol{\Theta} + o_p(1),$$

with error $o_p(b_p^{1/2})$ on the edge-scale domains.

Since

$$\boldsymbol{B}_{\alpha,N}\boldsymbol{R}_{\mu,N} = \widehat{d}_{1,\alpha}\boldsymbol{R}_{\mu,N},$$

we likewise have

$$\boldsymbol{R}_{\mu,N}^{\top}\boldsymbol{Q}_{\alpha,N}(z)\boldsymbol{R}_{\mu,N} = -\frac{1}{1+y_N\widehat{d}_{1,\alpha}m_{\alpha,N}^{\circ}(z)}\boldsymbol{I} + o_p(1).$$

Moreover,

$$\boldsymbol{W}_{\alpha,N}^{\top}\boldsymbol{Q}_{\alpha,N}(z)\boldsymbol{R}_{\mu,N} = o_p(b_p^{1/2}),$$

because the corresponding deterministic term is proportional to

$$\boldsymbol{W}_{\alpha,N}^{\top}\boldsymbol{R}_{\mu,N} = O_p(N^{-1/2}).$$

On the left side, Assumption 3 and the mean-matrix approximation give

$$\|\boldsymbol{U}^{\top}\boldsymbol{F}_{\mu,N}\| = o_p(b_p^{1/2}).$$

Together with the anisotropic local law and the mixed resolvent estimates, this gives

$$\boldsymbol{U}^{\top}\boldsymbol{G}_{\alpha,N}(z)\boldsymbol{F}_{\mu,N} = o_p(b_p^{1/2}),$$

$$\boldsymbol{W}_{\alpha,N}^{\top}\boldsymbol{Y}_{\alpha,N}^{\top}\boldsymbol{G}_{\alpha,N}(z)\boldsymbol{F}_{\mu,N} = o_p(b_p^{1/2}),$$

and

$$\boldsymbol{R}_{\mu,N}^{\top}\boldsymbol{Y}_{\alpha,N}^{\top}\boldsymbol{G}_{\alpha,N}(z)\boldsymbol{U} = o_p(b_p^{1/2})$$

uniformly on the edge-scale domains. Thus all factor–mean off-diagonal blocks of $\boldsymbol{\mathcal{L}}_{\alpha,N}(z)$ are $o_p(b_p^{1/2})$.

On each fixed compact domain separated from the limiting bulk,

$$m_{\alpha,N}^{\circ}(z) \longrightarrow m_{\alpha}(z), \qquad q_{\alpha,N}^{\circ}(z) \longrightarrow q_{\alpha}(z) = -z - \frac{1}{m_{\alpha}(z)}$$

along every convergent subsequence of α.

From any subsequence, first extract a further subsequence on which the preceding uniform $o_p(1)$ terms, including the normalized $o_p(b_p)$ and $o_p(b_p^{1/2})$ terms, converge to zero almost surely. On this event, compactness allows us, for each realization, to pass to a further subsequence along which α and the eigenvalues of $\boldsymbol{F}_{\mu,N}^{\top}\boldsymbol{F}_{\mu,N}$ converge, with the matrix dimension fixed. Denote the resulting nonnegative limits, counted with multiplicity, by $\xi_1,\dots,\xi_L$, where $L \le K-1$. The safe-set bound above implies

$$\xi_{\ell} \le t_{\alpha}, \qquad \ell = 1,\dots,L.$$

After a fixed permutation of coordinates, the limit of $\boldsymbol{\mathcal{L}}_{\alpha,N}(z)$ is block diagonal. The block corresponding to factor a is

$$\boldsymbol{L}_{a,\alpha}^{\mathrm{fac}}(z) = \begin{pmatrix} \theta_a q_{\alpha}(z) & 1 \\ 1 & m_{\alpha}(z) \end{pmatrix},$$

with determinant

$$\det \boldsymbol{L}_{a,\alpha}^{\mathrm{fac}}(z) = -(1+\theta_a)\left(1 + z m_{\alpha}(z)\frac{\theta_a}{1+\theta_a}\right).$$

The block corresponding to the limiting mean eigenvalue ξ_{ℓ} is

$$\boldsymbol{L}_{\ell,\alpha}^{\mathrm{mean}}(z) = \begin{pmatrix} -\dfrac{1}{1+y d_{1,\alpha} m_{\alpha}(z)} & 1 \\ 1 & \xi_{\ell} m_{\alpha}(z) \end{pmatrix},$$

with determinant

$$\det \boldsymbol{L}_{\ell,\alpha}^{\mathrm{mean}}(z) = -\left(1 + \frac{\xi_\ell m_\alpha(z)}{1 + y d_{1,\alpha} m_\alpha(z)}\right).$$

For $z > \lambda_{+,\alpha}$,

$$1 + \frac{\xi_\ell m_\alpha(z)}{1 + y d_{1,\alpha} m_\alpha(z)} = 0 \quad \Longleftrightarrow \quad \xi_\ell = -\frac{1}{m_\alpha(z)} - y d_{1,\alpha} > \frac{1}{s_\alpha} - y d_{1,\alpha} = t_\alpha.$$

Hence the mean blocks have no zeros in $(\lambda_{+,\alpha}, \infty)$.

For the factor blocks,

$$\frac{d}{dz}\left\{-\frac{1}{z m_\alpha(z)}\right\} = \frac{\displaystyle\int \frac{x}{(x-z)^2}\, dF_\alpha(x)}{z^2 m_\alpha^2(z)} > 0.$$

Since

$$-\frac{1}{\lambda_{+,\alpha} m_\alpha(\lambda_{+,\alpha})} = \frac{1}{\lambda_{+,\alpha} s_\alpha},$$

the equation

$$1 + z m_\alpha(z) \frac{\theta_a}{1 + \theta_a} = 0$$

has a solution in $(\lambda_{+,\alpha}, \infty)$ if and only if

$$\theta_a > \frac{1}{\lambda_{+,\alpha} s_\alpha - 1} = \vartheta_\alpha.$$

Since $\theta_r > \vartheta_\alpha$, this equation has a unique simple root $\rho_\alpha(\theta_a)$ for each $a = 1, \ldots, r$. Moreover,

$$\rho_\alpha(\theta_1) > \cdots > \rho_\alpha(\theta_r) > \lambda_{+,\alpha}.$$

Fix a sufficiently small $\eta > 0$ and set

$$J_a = \left(\rho_\alpha(\theta_a) - \eta, \rho_\alpha(\theta_a) + \eta\right), \qquad a = 1, \ldots, r,$$

so that the intervals are pairwise disjoint and contained in $(\lambda_{+,\alpha}, \infty)$. The limiting kernel has exactly one simple zero in each corresponding complex disk. The edge-rigidity result in Theorem S.3.11 of Ding & Yang (2021), together with the preceding uniform kernel approximation and Rouché's theorem, therefore gives exactly one eigenvalue of $\boldsymbol{S}_\alpha$ in each J_a with probability tending to one.

We next exclude additional eigenvalues above the selection threshold. The self-consistent map associated with $F^{\boldsymbol{B}_{\alpha,N}}$ is

$$x_{\alpha,N}^{\circ}(u) = -\frac{1}{u} + \frac{(m-K)\widehat{d}_{0,\alpha}/N}{1 + y_N \widehat{d}_{0,\alpha} u} + \frac{(n-1)\widehat{d}_{1,\alpha}/N}{1 + y_N \widehat{d}_{1,\alpha} u}.$$

Since

$$\frac{m-K}{N}\widehat{d}_{0,\alpha} = \alpha - \frac{K}{N}\widehat{d}_{0,\alpha}, \qquad \frac{n-1}{N}\widehat{d}_{1,\alpha} = 1 - \alpha - \frac{1}{N}\widehat{d}_{1,\alpha},$$

we have, uniformly on a fixed neighborhood of $-\widehat{s}_\alpha$ separated from the poles,

$$x_{\alpha,N}^{\circ}(u) - \widehat{x}_\alpha(u) = O(N^{-1}), \qquad {x_{\alpha,N}^{\circ}}'(u) - \widehat{x}_\alpha'(u) = O(N^{-1}).$$

Since $\widehat{x}_\alpha''(-\widehat{s}_\alpha)$ is uniformly bounded away from zero, the corresponding critical point and right edge differ from $-\widehat{s}_\alpha$ and $\widehat{\lambda}_{+,\alpha}$, respectively, by $O(N^{-1})$.

On the right-edge branch at distance at least b_p from the edge, the quadratic edge expansion gives

$$|{x_{\alpha,N}^{\circ}}'(u)| \geq c b_p^{1/2}.$$

Hence the inverse branches of $x_{\alpha,N}^{\circ}$ and $\widehat{x}_\alpha$ differ by

$$O(N^{-1} b_p^{-1/2}) = o(b_p^{1/2}).$$

Thus the finite-N corrections do not affect any of the $b_p^{1/2}$-scale kernel bounds below. The edge-rigidity result in Theorem S.3.11 of Ding & Yang (2021), together with the same truncation reduction, gives

$$\lambda_1(\boldsymbol{T}_{\alpha,N}) \le \widehat{\lambda}_{+,\alpha} + o_p(b_p)$$

uniformly over $\alpha \in [0,1]$. Furthermore,

$$\|\boldsymbol{S}_\alpha\|^{1/2} \le \|\boldsymbol{Y}_{\alpha,N}\| + \|\boldsymbol{W}_{\alpha,N}\| + \|\boldsymbol{F}_{\mu,N}\|.$$

The uniform edge bound, the uniform convergence of $\boldsymbol{W}_{\alpha,N}^\top \boldsymbol{W}_{\alpha,N}$ to $\boldsymbol{\Theta}$, and the uniform mean-matrix bound imply that, for some sufficiently large fixed M,

$$\mathbb{P}\left\{\sup_{\alpha\in[0,1]} \|\boldsymbol{S}_\alpha\| < M\right\} \longrightarrow 1.$$

Set

$$I_N = [\widehat{\lambda}_{+,\alpha} + b_p, M] \setminus \bigcup_{a=1}^{r} J_a.$$

For $z \in I_N$, let $u(z) \in (-\widehat{s}_\alpha, 0)$ be the right-edge solution of

$$z = \widehat{x}_\alpha(u(z)).$$

Let

$$u_N^\circ(z) = m_{\alpha,N}^\circ(z)$$

denote the corresponding right-edge solution of

$$z = x_{\alpha,N}^\circ(u_N^\circ(z)).$$

The preceding inverse-branch comparison gives

$$\sup_{z\in I_N} |u_N^\circ(z) - u(z)| = o(b_p^{1/2}).$$

The quadratic edge expansion gives

$$u(z) + \widehat{s}_\alpha \ge c b_p^{1/2}.$$

Moreover,

$$0 < c \le 1 + y_N \widehat{d}_{1,\alpha} u(z) \le 1.$$

For every positive eigenvalue $\xi_{\ell,N}$ of $\boldsymbol{F}_{\mu,N}\boldsymbol{F}_{\mu,N}^\top$,

$$1 + \left(\xi_{\ell,N} + y_N \widehat{d}_{1,\alpha}\right) u(z) = \frac{u(z) + \widehat{s}_\alpha}{\widehat{s}_\alpha} + \left(\xi_{\ell,N} - \widehat{t}_\alpha\right) u(z).$$

Since

$$\sup_{\alpha\in\widehat{\mathcal{A}}_N} \left[\|\boldsymbol{F}_{\mu,N}\|^2 - \widehat{t}_\alpha\right]_+ = o_p(b_p),$$

it follows that

$$1 + \frac{\xi_{\ell,N} u(z)}{1 + y_N \widehat{d}_{1,\alpha} u(z)} \ge c b_p^{1/2}$$

uniformly over $\alpha \in \widehat{\mathcal{A}}_N$, $z \in I_N$, and all ℓ, with probability tending to one.

For $a = 1, \ldots, r$, exclusion of J_a gives

$$\left|1 + z u(z) \frac{\theta_a}{1 + \theta_a}\right| \ge c$$

uniformly on I_N. Since

$$\sup_{z\in I_N} |u_N^\circ(z) - u(z)| = o(b_p^{1/2}),$$

the corresponding finite-N factor blocks satisfy

$$\left|1 + z u_N^{\circ}(z) \frac{\theta_a}{1+\theta_a}\right| \geq c b_p^{1/2}, \qquad a = 1, \dots, r,$$

uniformly on I_N, with probability tending to one. Likewise, for every positive eigenvalue $\xi_{\ell,N}$ of $\boldsymbol{F}_{\mu,N}\boldsymbol{F}_{\mu,N}^{\top}$,

$$1 + \frac{\xi_{\ell,N} u_N^{\circ}(z)}{1 + y_N \widehat{d}_{1,\alpha} u_N^{\circ}(z)} \geq c b_p^{1/2}.$$

Therefore the absolute values of the determinants of the finite-N diagonal factor and mean blocks are bounded below by $c b_p^{1/2}$. Since these blocks have uniformly bounded operator norm, their smallest singular values are also bounded below by $c b_p^{1/2}$. Since all off-diagonal blocks and local-law errors are $o_p(b_p^{1/2})$, the full kernel satisfies

$$\inf_{z \in I_N} \sigma_{\min}\left(\boldsymbol{\mathcal{L}}_{\alpha,N}(z)\right) \geq c b_p^{1/2}$$

with probability tending to one. The determinant identity therefore shows that

$$\mathbb{P}\left\{\operatorname{spec}(\boldsymbol{S}_\alpha) \cap I_N \neq \varnothing\right\} \to 0.$$

The preceding exclusion argument remains valid with b_p replaced by any fixed positive multiple of b_p. Hence

$$\left[\lambda_{r+1}(\boldsymbol{S}_\alpha) - \widehat{\lambda}_{+,\alpha}\right]_+ = o_p(b_p).$$

By Lemma 1, uniformly over $\alpha \in [0,1]$,

$$\widehat{\sigma}^2\left(\widehat{\lambda}_{+,\alpha} + b_p\right) = \widehat{\lambda}_{+,\alpha} + b_p + o_p(b_p).$$

The cutoff in (10) therefore exceeds $\lambda_{r+1}(\boldsymbol{S}_\alpha)$ with probability tending to one and converges to $\lambda_{+,\alpha}$. Together with the localization in the intervals J_a, this gives

$$\mathbb{P}\{\widehat{r}_\alpha \neq r\} \to 0.$$

It remains to establish the subspace limit. Fix $a \in \{1, \dots, r\}$ and a positively oriented contour $\mathcal{C}_a$ enclosing $\rho_\alpha(\theta_a)$ and no other limiting outlier. The Woodbury identity and the preceding uniform kernel approximation give, for $b = 1, \dots, r$,

$$\boldsymbol{u}_b^{\top}(\boldsymbol{S}_\alpha - z\boldsymbol{I}_p)^{-1}\boldsymbol{u}_b = m_\alpha(z) - \frac{\theta_b m_\alpha^2(z) q_\alpha(z)}{\theta_b m_\alpha(z) q_\alpha(z) - 1} + o_p(1)$$

uniformly on $\mathcal{C}_a$. Since $\mathcal{C}_a$ contains exactly $\lambda_a(\boldsymbol{S}_\alpha)$ with probability tending to one,

$$\left|\boldsymbol{u}_b^{\top}\widehat{\boldsymbol{q}}_a^{\alpha}\right|^2 = -\frac{1}{2\pi \mathrm{i}} \oint_{\mathcal{C}_a} \boldsymbol{u}_b^{\top}(\boldsymbol{S}_\alpha - z\boldsymbol{I}_p)^{-1}\boldsymbol{u}_b \, dz + o_p(1).$$

Using $q_\alpha(z) = -z - 1/m_\alpha(z)$, we obtain

$$\begin{aligned} &- \operatorname*{Res}_{z=\rho_\alpha(\theta_a)} \left[m_\alpha(z) - \frac{\theta_b m_\alpha^2(z) q_\alpha(z)}{\theta_b m_\alpha(z) q_\alpha(z) - 1}\right] \\ &\quad = \mathbf{1}_{\{a=b\}} \frac{m_\alpha^2(\rho_\alpha(\theta_a))}{(1+\theta_a) m_\alpha'(\rho_\alpha(\theta_a)) - \theta_a m_\alpha^2(\rho_\alpha(\theta_a))} = \mathbf{1}_{\{a=b\}} \kappa_\alpha(\theta_a). \end{aligned}$$

Consequently,

$$\left|\boldsymbol{u}_b^{\top}\widehat{\boldsymbol{q}}_a^{\alpha}\right|^2 - \mathbf{1}_{\{a=b\}}\kappa_\alpha(\theta_a) \xrightarrow{P} 0, \qquad a, b = 1, \dots, r.$$

On the event $\widehat{r}_\alpha = r$,

$$\begin{aligned} \left\|\widehat{\boldsymbol{P}}_\alpha - \boldsymbol{P}_s\right\|_F^2 &= 2r - 2\operatorname{tr}\left(\boldsymbol{P}_s \widehat{\boldsymbol{P}}_\alpha\right) \\ &= 2r - 2\sum_{a=1}^{r}\sum_{b=1}^{r} \left|\boldsymbol{u}_b^{\top}\widehat{\boldsymbol{q}}_a^{\alpha}\right|^2. \end{aligned}$$

Since $\mathbb{P}\{\widehat{r}_\alpha = r\} \to 1$,

$$\left\|\widehat{\boldsymbol{P}}_\alpha - \boldsymbol{P}_s\right\|_F^2 - 2\left\{r - \sum_{a=1}^{r} \kappa_\alpha(\theta_a)\right\} \xrightarrow{P} 0.$$

The preceding bounds are uniform over any compact subset $\mathcal{K}$ of

$$\{\alpha \in [0,1] : \vartheta_\alpha < \theta_r\}.$$

Hence

$$\sup_{\alpha \in \widehat{\mathcal{A}}_N \cap \mathcal{K}} \left[|\widehat{r}_\alpha - r| + \left|\left\|\widehat{\boldsymbol{P}}_\alpha - \boldsymbol{P}_s\right\|_F^2 - 2\left\{r - \sum_{a=1}^{r} \kappa_\alpha(\theta_a)\right\}\right|\right] \xrightarrow{P} 0.$$

□

Proof of Theorem 2. Set

$$a_\alpha = \frac{\alpha s_\alpha}{1 - \alpha y_2 s_\alpha}, \qquad b_\alpha = \frac{(1-\alpha) s_\alpha}{1 - (1-\alpha) y_1 s_\alpha}.$$

The right-edge equation gives

$$y_2 a_\alpha^2 + y_1 b_\alpha^2 = 1, \qquad \lambda_{+,\alpha} s_\alpha - 1 = a_\alpha + b_\alpha.$$

Hence

$$\vartheta_\alpha = \frac{1}{a_\alpha + b_\alpha}.$$

By the Cauchy–Schwarz inequality,

$$(a_\alpha + b_\alpha)^2 \le \left(\frac{1}{y_2} + \frac{1}{y_1}\right)\left(y_2 a_\alpha^2 + y_1 b_\alpha^2\right) = \frac{1}{y},$$

and therefore $\vartheta_\alpha \ge \sqrt{y}$. Equality holds if and only if $y_2 a_\alpha = y_1 b_\alpha$. For $\alpha \in (0,1)$,

$$\frac{\alpha}{1-\alpha} = \frac{a_\alpha(1 + y_1 b_\alpha)}{b_\alpha(1 + y_2 a_\alpha)},$$

so equality is attained if and only if $\alpha = \beta$.

As in the proof of Lemma 1, b_α is strictly decreasing in α. Along the ellipse $y_2 a_\alpha^2 + y_1 b_\alpha^2 = 1$,

$$\frac{d}{db_\alpha}(a_\alpha + b_\alpha) = 1 - \frac{y_1 b_\alpha}{y_2 a_\alpha}.$$

It follows that ϑ_α is strictly decreasing on $[0, \beta]$ and strictly increasing on $[\beta, 1]$.

For $\theta > \vartheta_\alpha$, put

$$s = -m_\alpha(\rho_\alpha(\theta)), \qquad u = \frac{\alpha s}{1 - \alpha y_2 s}, \qquad v = \frac{(1-\alpha)s}{1 - (1-\alpha) y_1 s}, \qquad Q = y_2 u^2 + y_1 v^2.$$

The outlier equation and (11) give

$$u + v = \theta^{-1}, \qquad Q < 1, \qquad m'_\alpha(\rho_\alpha(\theta)) = \frac{s^2}{1-Q}, \qquad \kappa_\alpha(\theta) = \frac{1-Q}{1+\theta Q}.$$

For $\alpha \in (0,1)$,

$$\frac{\alpha}{1-\alpha} = \frac{u(1 + y_1 v)}{v(1 + y_2 u)}.$$

Along $v = \theta^{-1} - u$,

$$\frac{d}{du} \log \frac{u(1 + y_1 v)}{v(1 + y_2 u)} = \frac{1}{u(1 + y_2 u)} + \frac{1}{v(1 + y_1 v)} > 0.$$

Thus u is increasing in α, with $u = \beta/\theta$ at $\alpha = \beta$. Since

$$Q = \frac{y}{\theta^2} + (y_1 + y_2)\left(u - \frac{\beta}{\theta}\right)^2, \qquad \frac{d}{dQ}\frac{1-Q}{1+\theta Q} = -\frac{1+\theta}{(1+\theta Q)^2},$$

$\kappa_\alpha(\theta)$ is increasing on the detectable part of $[0, \beta]$ and decreasing on the detectable part of $[\beta, 1]$. The endpoint values follow by continuity.

By Lemma 1, $\widehat{\mathcal{A}}_N$ is a closed interval with right endpoint 1. Define

$$\widetilde{\alpha} = \max\{\min \widehat{\mathcal{A}}_N, \beta\}.$$

Then

$$\alpha_* = \max\{\min \widehat{\mathcal{A}}_N, \alpha_0\}, \qquad |\alpha_* - \widetilde{\alpha}| \le |\alpha_0 - \beta| \longrightarrow 0.$$

The threshold monotonicity and continuity imply

$$\vartheta_{\widetilde{\alpha}} = \min_{\alpha \in \widehat{\mathcal{A}}_N} \vartheta_\alpha, \qquad \vartheta_{\alpha_*} - \vartheta_{\widetilde{\alpha}} = o_p(1).$$

Let $\alpha \in \widehat{\mathcal{A}}_N$ satisfy $\vartheta_\alpha < \theta_r$. Since

$$\vartheta_{\widetilde{\alpha}} \le \vartheta_\alpha < \theta_r, \qquad \vartheta_{\alpha_*} - \vartheta_{\widetilde{\alpha}} = o_p(1),$$

the weights α, $\widetilde{\alpha}$, and α_* lie with probability tending to one in a compact subset of

$$\{\gamma \in [0,1] : \vartheta_\gamma < \theta_r\}.$$

The preceding monotonicity and continuity give

$$\kappa_{\widetilde{\alpha}}(\theta_a) \ge \kappa_\alpha(\theta_a), \qquad \kappa_{\alpha_*}(\theta_a) - \kappa_{\widetilde{\alpha}}(\theta_a) = o_p(1), \qquad a = 1, \dots, r.$$

The uniform conclusion in the proof of Theorem 1 therefore yields $\widehat{r}_{\alpha_*} = \widehat{r}_\alpha = r$ with probability tending to one and

$$\begin{aligned}
&\left\|\widehat{\boldsymbol{P}}_{\alpha_*} - \boldsymbol{P}_s\right\|_F^2 - \left\|\widehat{\boldsymbol{P}}_{\alpha} - \boldsymbol{P}_s\right\|_F^2 \\
&\qquad = 2\sum_{a=1}^{r}\{\kappa_\alpha(\theta_a) - \kappa_{\alpha_*}(\theta_a)\} + o_p(1) \le o_p(1).
\end{aligned}$$

Suppose now that $\theta_r > \sqrt{y}$. On the event $\{\alpha_0 \in \widehat{\mathcal{A}}_N\}$, $\alpha_* = \alpha_0 \to \beta$. At $\alpha = \beta$,

$$\vartheta_\beta = \sqrt{y}, \qquad \kappa_\beta(\theta_a) = \frac{\theta_a^2 - y}{\theta_a(\theta_a + y)}.$$

The same uniform conclusion gives

$$\mathbf{1}_{\{\alpha_0 \in \widehat{\mathcal{A}}_N\}}\left[\left\|\widehat{\boldsymbol{P}}_{\alpha_*} - \boldsymbol{P}_s\right\|_F^2 - 2\sum_{a=1}^{r}\left(1 - \frac{\theta_a^2 - y}{\theta_a(\theta_a + y)}\right)\right] \xrightarrow{P} 0.$$

For the oracle covariance, let N_k denote the number of observations from population k. With probability tending to one, all N_k are positive. After reordering the columns of $\boldsymbol{Z}_N$ and $\boldsymbol{\mathcal{F}}_N$ by population, set

$$\boldsymbol{P}_{\mathrm{or},N} = \mathrm{diag}\left(\boldsymbol{\Phi}_{N_1}, \dots, \boldsymbol{\Phi}_{N_K}\right), \qquad \boldsymbol{\Phi}_q = \boldsymbol{I}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top,$$

and

$$\boldsymbol{Y}_{\mathrm{or},N} = N^{-1/2}\boldsymbol{Z}_N\boldsymbol{P}_{\mathrm{or},N}, \qquad \boldsymbol{W}_{\mathrm{or},N} = N^{-1/2}\boldsymbol{P}_{\mathrm{or},N}\boldsymbol{\mathcal{F}}_N^\top.$$

Then

$$\boldsymbol{S}_{\mathrm{or}} = \left(\boldsymbol{Y}_{\mathrm{or},N} + \boldsymbol{U}\boldsymbol{W}_{\mathrm{or},N}^\top\right)\left(\boldsymbol{Y}_{\mathrm{or},N} + \boldsymbol{U}\boldsymbol{W}_{\mathrm{or},N}^\top\right)^\top.$$

Conditional on the population memberships, the projection is independent of the factor scores and noise vectors, and

$$\|\boldsymbol{P}_{\mathrm{or},N}\| = 1, \qquad F^{\boldsymbol{P}_{\mathrm{or},N}} = \left(1 - \frac{K}{N}\right)\delta_1 + \frac{K}{N}\delta_0 \Rightarrow \delta_1.$$

Moreover,

$$\boldsymbol{W}_{\mathrm{or},N}^{\top}\boldsymbol{W}_{\mathrm{or},N} = \boldsymbol{\Theta} + O_p(N^{-1/2}).$$

The projected resolvent and residue calculation in the proof of Theorem 1, with $H = \delta_1$ and no mean block, gives

$$\operatorname{tr}\left(\boldsymbol{P}_s\widehat{\boldsymbol{P}}_{\mathrm{or}}\right) - \sum_{a=1}^{r}\frac{\theta_a^2 - y}{\theta_a(\theta_a + y)} \xrightarrow{P} 0.$$

Since both projectors have rank r,

$$\left\|\widehat{\boldsymbol{P}}_{\mathrm{or}} - \boldsymbol{P}_s\right\|_F^2 - 2\sum_{a=1}^{r}\left(1 - \frac{\theta_a^2 - y}{\theta_a(\theta_a + y)}\right) \xrightarrow{P} 0.$$

Combining the preceding two error limits proves the last assertion. □

Proof of Theorem 3. The monotonicity calculation in the proof of Theorem 2 remains valid under the present assumptions: ϑ_α is strictly decreasing on $[0, \beta]$ and strictly increasing on $[\beta, 1]$. By Lemma 1, $\widehat{\mathcal{A}}_N$ is a closed interval with right endpoint 1. Let

$$\widetilde{\alpha} = \max\{\min\widehat{\mathcal{A}}_N, \beta\}.$$

Then

$$\vartheta_{\widetilde{\alpha}} = \min_{\alpha\in\widehat{\mathcal{A}}_N}\vartheta_\alpha.$$

Moreover,

$$\alpha_* = \max\{\min\widehat{\mathcal{A}}_N, \alpha_0\}, \qquad |\alpha_* - \widetilde{\alpha}| \le |\alpha_0 - \beta| \longrightarrow 0.$$

Hence, by continuity of ϑ_α,

$$\vartheta_{\alpha_*} - \min_{\alpha\in\widehat{\mathcal{A}}_N}\vartheta_\alpha \xrightarrow{P} 0.$$

Fix $\alpha \in \widehat{\mathcal{A}}_N$ satisfying $\theta_r > \vartheta_\alpha$. For the remaining assertions, we retain the notation of the proof of Theorem 1 and set

$$\boldsymbol{A}_{\alpha,N} = \boldsymbol{Y}_{\alpha,N} + \boldsymbol{F}_{\mu,N}\boldsymbol{R}_{\mu,N}^{\top}.$$

The mean perturbation bound established in that proof does not require asymptotic orthogonality. Applied with the factor term removed, it gives

$$\sup_{\alpha\in\widehat{\mathcal{A}}_N}\left[\lambda_1(\boldsymbol{A}_{\alpha,N}\boldsymbol{A}_{\alpha,N}^{\top}) - \widehat{\lambda}_{+,\alpha}\right]_+ = o_p(b_p).$$

Since

$$\boldsymbol{S}_\alpha = (\boldsymbol{A}_{\alpha,N} + \boldsymbol{U}\boldsymbol{W}_{\alpha,N}^{\top})(\boldsymbol{A}_{\alpha,N} + \boldsymbol{U}\boldsymbol{W}_{\alpha,N}^{\top})^{\top}$$

and $\operatorname{rank}(\boldsymbol{U}\boldsymbol{W}_{\alpha,N}^{\top}) \le r$, the singular value rank inequality yields

$$\lambda_{r+1}(\boldsymbol{S}_\alpha) \le \lambda_1(\boldsymbol{A}_{\alpha,N}\boldsymbol{A}_{\alpha,N}^{\top}) \le \widehat{\lambda}_{+,\alpha} + o_p(b_p).$$

The cutoff comparison in the proof of Theorem 1 therefore gives $\widehat{r}_\alpha \le r$ simultaneously over $\alpha \in \widehat{\mathcal{A}}_N$, with probability tending to one.

Define

$$\boldsymbol{J}(z) = \begin{pmatrix}\left(-z - m_\alpha(z)^{-1}\right)\boldsymbol{\Theta} & \boldsymbol{0}\\ \boldsymbol{0} & -\left(1 + yd_{1,\alpha}m_\alpha(z)\right)^{-1}\boldsymbol{I}\end{pmatrix}.$$

The uniform resolvent estimates in the proof of Theorem 1, with the block $\boldsymbol{U}^{\top}\boldsymbol{F}_{\mu,N}$ retained, give

$$\boldsymbol{\mathcal{L}}_{\alpha,N}(z) = \begin{pmatrix}\boldsymbol{J}(z) & \boldsymbol{I}\\ \boldsymbol{I} & m_\alpha(z)\boldsymbol{C}_N^{\top}\boldsymbol{C}_N\end{pmatrix} + o_p(1).$$

Here and below, these approximations are uniform in α on compact spectral domains at a fixed positive distance from the limiting bulk. Using the mean matrix approximation from the same proof,

$$\boldsymbol{F}_{\mu,N}\boldsymbol{F}_{\mu,N}^{\top} = (1 - \alpha)\boldsymbol{\Sigma}_\mu + o_p(1),$$

we obtain

$$m_\alpha(z)\boldsymbol{C}_N\boldsymbol{J}(z)\boldsymbol{C}_N^\top = a(z)\boldsymbol{U}\boldsymbol{\Theta}\boldsymbol{U}^\top + b(z)\boldsymbol{\Sigma}_\mu + o_p(1) = \boldsymbol{I}_p - \boldsymbol{K}(z) + o_p(1).$$

The Schur complement and Sylvester's determinant identity therefore imply

$$\det(\boldsymbol{K}_N)\det\boldsymbol{\mathcal{L}}_{\alpha,N}(z) = \det\boldsymbol{K}(z) + o_p(1).$$

By Lemma 1,

$$(1-\alpha)\gamma_\mu \le t_\alpha + o_p(1).$$

Moreover,

$$a(\lambda_{+,\alpha}) = \vartheta_\alpha^{-1}, \qquad b(\lambda_{+,\alpha}) = \frac{1-\alpha}{t_\alpha}.$$

For real $x > \lambda_{+,\alpha}$,

$$a'(x) = -\int \frac{t}{(t-x)^2}\,dF_\alpha(t) < 0, \qquad b'(x) = -\frac{(1-\alpha)m'_\alpha(x)}{\{1+yd_{1,\alpha}m_\alpha(x)\}^2} \le 0.$$

Since $\theta_r > \vartheta_\alpha$, there exist fixed $\varepsilon, c > 0$ such that

$$\lambda_r\{\boldsymbol{I}_p - \boldsymbol{K}(\lambda_{+,\alpha}+2\varepsilon)\} \ge a(\lambda_{+,\alpha}+2\varepsilon)\theta_r \ge 1+c,$$
$$\lambda_{r+1}\{\boldsymbol{I}_p - \boldsymbol{K}(\lambda_{+,\alpha}+\varepsilon)\} \le b(\lambda_{+,\alpha}+\varepsilon)\gamma_\mu \le 1-c+o_p(1).$$

Thus $\boldsymbol{I}_p - \boldsymbol{K}(x)$ decreases to zero as $x \to \infty$, and its eigenvalues are nonincreasing in x. At any root $\rho > \lambda_{+,\alpha}$, every nonzero $\boldsymbol{v} \in \ker\boldsymbol{K}(\rho)$ satisfies

$$\|\boldsymbol{v}\|^2 = a(\rho)\boldsymbol{v}^\top\boldsymbol{U}\boldsymbol{\Theta}\boldsymbol{U}^\top\boldsymbol{v} + b(\rho)\boldsymbol{v}^\top\boldsymbol{\Sigma}_\mu\boldsymbol{v},$$

and hence

$$\boldsymbol{v}^\top\boldsymbol{K}'(\rho)\boldsymbol{v} > 0.$$

Therefore exactly r roots, counted with multiplicity, lie above $\lambda_{+,\alpha}+\varepsilon$, and all lie above $\lambda_{+,\alpha}+2\varepsilon$. These are precisely the roots represented by $\mathcal{Z}$.

For nonreal z,

$$\operatorname{Im} a(z) = -\operatorname{Im} z\int \frac{t}{|t-z|^2}\,dF_\alpha(t), \quad \operatorname{Im} b(z) = -\frac{(1-\alpha)\operatorname{Im} m_\alpha(z)}{|1+yd_{1,\alpha}m_\alpha(z)|^2}.$$

These identities imply that $\boldsymbol{K}(z)$ is nonsingular when $\operatorname{Im} z \neq 0$.

Choose a positively oriented rectangular contour Γ_α with left boundary $\operatorname{Re} z = \lambda_{+,\alpha}+\varepsilon$, right boundary at a sufficiently large fixed value, and horizontal boundaries at fixed nonzero imaginary parts. It encloses exactly the r roots above. The preceding gaps, nonreal invertibility, and compactness of the bounded Gram matrices of fixed dimension give

$$\sup_{z\in\Gamma_\alpha}\|\boldsymbol{K}(z)^{-1}\| = O_p(1).$$

The determinant identity and the Rouché argument used in the proof of Theorem 1 then show that Γ_α encloses exactly r eigenvalues of $\boldsymbol{S}_\alpha$. Since $\widehat{\sigma}^2 \to 1$, $\widehat{\lambda}_{+,\alpha} \to \lambda_{+,\alpha}$, and $b_p \to 0$, these eigenvalues exceed $\widehat{\sigma}^2(\widehat{\lambda}_{+,\alpha}+2b_p)$ with probability tending to one. Together with the upper bound already proved, this yields

$$\mathbb{P}\{\widehat{r}_\alpha = r\} \longrightarrow 1.$$

On Γ_α, the Woodbury identity and the same resolvent estimates give

$$\begin{aligned}\boldsymbol{U}^\top(\boldsymbol{S}_\alpha - z\boldsymbol{I}_p)^{-1}\boldsymbol{U} &= m_\alpha(z)\boldsymbol{I}_r - m_\alpha(z)^2\boldsymbol{U}^\top\boldsymbol{C}_N\left\{m_\alpha(z)\boldsymbol{C}_N^\top\boldsymbol{C}_N - \boldsymbol{J}(z)^{-1}\right\}^{-1}\boldsymbol{C}_N^\top\boldsymbol{U} + o_p(1)\\ &= m_\alpha(z)\boldsymbol{U}^\top\boldsymbol{K}(z)^{-1}\boldsymbol{U} + o_p(1).\end{aligned}$$

For each $\rho \in \mathcal{Z}$, the positivity proved above gives

$$\boldsymbol{V}_\rho^\top\boldsymbol{K}'(\rho)\boldsymbol{V}_\rho \succ 0.$$

Taking the Schur complement with respect to $\ker \boldsymbol{K}(\rho)$ and its orthogonal complement yields

$$\operatorname*{Res}_{z=\rho} \boldsymbol{K}(z)^{-1} = \boldsymbol{V}_\rho \left\{ \boldsymbol{V}_\rho^\top \boldsymbol{K}'(\rho) \boldsymbol{V}_\rho \right\}^{-1} \boldsymbol{V}_\rho^\top.$$

This formula also holds when the root has multiplicity greater than one. Therefore,

$$\operatorname{tr}(\boldsymbol{P}_s \widehat{\boldsymbol{P}}_\alpha) = -\frac{1}{2\pi i} \oint_{\Gamma_\alpha} m_\alpha(z) \operatorname{tr}\{\boldsymbol{P}_s \boldsymbol{K}(z)^{-1}\} \, dz + o_p(1) = \sum_{\rho \in \mathcal{Z}} c_\rho + o_p(1).$$

Finally, on the event $\widehat{r}_\alpha = r$,

$$\|\widehat{\boldsymbol{P}}_\alpha - \boldsymbol{P}_s\|_F^2 = 2r - 2 \operatorname{tr}(\boldsymbol{P}_s \widehat{\boldsymbol{P}}_\alpha) = 2 \left(r - \sum_{\rho \in \mathcal{Z}} c_\rho \right) + o_p(1).$$

□