# PRINCIPAL COMPONENTS IN LINEAR MIXED MODELS WITH GENERAL BULK

BY ZHOU FAN[1], YI SUN[2] AND ZHICHAO WANG[3]

[1]*Department of Statistics and Data Science, Yale University, zhou.fan@yale.edu*
[2]*Department of Statistics, University of Chicago, yisun@statistics.uchicago.edu*
[3]*Department of Mathematics, University of California, San Diego, zhw036@ucsd.edu*

We study the principal components of covariance estimators in multivariate mixed-effects linear models. We show that, in high dimensions, the principal eigenvalues and eigenvectors may exhibit bias and aliasing effects that are not present in low-dimensional settings. We derive the first-order limits of the principal eigenvalue locations and eigenvector projections in a high-dimensional asymptotic framework, allowing for general population spectral distributions for the random effects and extending previous results from a more restrictive spiked model. Our analysis uses free probability techniques, and we develop two general tools of independent interest—strong asymptotic freeness of GOE and deterministic matrices and a free deterministic equivalent approximation for bilinear forms of resolvents.

**1. Introduction.** Principal components analysis (PCA) is a commonly used technique for identifying linear low-rank structure in high-dimensional data [35]. For $n$ independent samples in a comparably large dimension $p$, it is now well established that the principal components of the sample covariance matrix may be inaccurate for their population counterparts [33]. A body of work has quantified the behavior of PCA in this setting [1–3, 9, 32, 47], connecting to the Marcenko–Pastur and Tracy–Widom laws of asymptotic random matrix theory [41, 54]. We refer readers to the review articles [34, 48] for more discussion and references to this and related lines of work.

Similar phenomena occur in statistical models where samples are not independent, but instead exhibit complex dependence structure [16, 38, 57, 60]. However, the behavior of PCA in many such models is less well understood. In this work, we consider the setting of mixed effects linear models [51], where dependence across observed samples arises via linear combinations of independent latent variables. These models are commonly used in statistical genetics to model quantitative phenotypes in related individuals [39]. We study the behavior of principal eigenvalues and eigenvectors of MANOVA covariance estimates for the random effects.

Our main results quantify several spectral bias and aliasing phenomena that may occur in high-dimensional applications. In particular, we show that large principal eigenvalues in the covariance of one random effect may bias the principal eigenvectors and also yield spurious eigenvalues in the estimated covariances of the other effects. These phenomena are unique to mixed-effects models, and they do not arise in similar spiked models of sample covariance matrices for independent samples [2, 3, 47]. In [25], such phenomena for mixed models were first described under an "isotropic noise" assumption, where the population covariance of each random effect is a low-rank perturbation of the identity. Our work extends these results to the setting of general population spectral distributions for the random effects. We derive

generalizations of the first-order limits for eigenvalues and eigenvector projections in [25] involving quantities appearing in the fixed-point equations for the empirical spectral law in [24]. We describe these results in Section 2.

Our proofs are very different from the analytic approach of [25]. Instead, they are based in free probability theory and its connection to random matrices [42, 55]. Our work also establishes two general results in this area—strong asymptotic freeness of independent GOE and deterministic matrices and a method of deriving anisotropic resolvent approximations using free deterministic equivalents [52]. We describe these in Section 4.

The connection between free probability and random matrices was introduced in [55] for deterministic and GUE matrices and has been extended to many other matrix models [7, 17, 20, 21, 23, 30, 52, 56]. Strong asymptotic freeness extending the approximation from the trace to the operator norm was first proven in [28] for GUE matrices and extended to other models in [6, 18, 40, 50]. Free probability techniques have recently been applied to study outlier eigenvalues in other matrix models [4, 5] and spectral behavior in other statistical applications, including autocovariance estimates for high-dimensional time series [10–13, 15] and sketching methods for linear regression [22]. The tools we develop may be of broader interest to the analysis of structured random matrices arising in other applications.

## 2. Probabilistic results in the linear mixed model.
Extending the representation of [49] to a multivariate setting, we consider the mixed-effects linear model

$$(1) \qquad Y = X\beta + U_1\alpha_1 + \cdots + U_k\alpha_k \in \mathbb{R}^{n \times p},$$

where $Y$ contains $n$ dependent observations in dimension $p$, each a combination of fixed effects $X\beta$ and random effects constituting the rows of $\alpha_1, \ldots, \alpha_k$. Here,

- $X \in \mathbb{R}^{n \times m}$ is an observed design matrix of a small number $m$ of fixed effects, with unknown regression coefficients $\beta \in \mathbb{R}^{m \times p}$.
- For each $r = 1, \ldots, k$, the matrix $\alpha_r \in \mathbb{R}^{n_r \times p}$ is unobserved, its rows constituting $n_r$ i.i.d. realizations of a $p$-dimensional random effect.
- Each $U_r \in \mathbb{R}^{n \times n_r}$ is a known, deterministic incidence matrix specified by the model design.

We study the behavior of PCA for estimates of the *variance components*, which are the covariance matrices $\Sigma_1, \ldots, \Sigma_k$ for the random effects in $\alpha_1, \ldots, \alpha_k$.

In quantitative genetics, $U_1, \ldots, U_k$ may encode a classification design, as commonly used in twin/sibling studies and breeding experiments. Examples are discussed in [24, 25]. In genomewide association study designs, $U_1, \ldots, U_k$ may contain genotype measurements at a set of single-nucleotide polymorphisms (SNPs) [37, 59]. It has been recognized since [27, 58] that variance components in these models can provide a decomposition of the total population variance of quantitative phenotypes into constituent genetic and nongenetic effects, yielding estimates of heritability. In high-dimensional applications, including the analysis of gene expression traits and other molecular phenotypes, the principal eigenvectors of the genetic components may indicate phenotypic subspaces near which responses to selection or random mutational drift are likely to be constrained [14, 19, 31]. Principal eigenvectors of the nongenetic components may correspond to hidden experimental confounders, to be removed before performing downstream analyses [36, 53].

As $\alpha_1, \ldots, \alpha_k$ are not individually observed, one cannot construct the usual sample covariance estimator for $\Sigma_1, \ldots, \Sigma_k$. Instead, each $\Sigma_r$ may be classically estimated by a MANOVA estimator of the form

$$\widehat{\Sigma}_r = Y^\mathsf{T} B_r Y,$$

where the symmetric matrix $B_r$ is chosen to satisfy the properties

$$B_r X = 0, \qquad \mathbb{E}[Y^\mathsf{T} B_r Y] = \Sigma_r.$$

Such an estimator $\widehat{\Sigma}_r$ is unbiased and equivariant to rotations of coordinates in $\mathbb{R}^p$—these properties are analogous to those holding for a sample covariance matrix for independent samples. For example, in a balanced one-way classification design, the within-group covariance matrix $\Sigma_1$ is estimated by the MANOVA estimator $\widehat{\Sigma}_1 = Y^\top B_1 Y$ where $B_1$ is a scaled difference of two orthogonal projections, the first onto a subspace of group means and the second onto its orthogonal complement. See Appendix A.2 in the Supplementary Material [26] for further details of this example.

Our main results, Theorems 2.5 and 2.6 below, characterize the first-order limiting behavior of the principal eigenvalues and eigenvectors of any such matrix $\widehat{\Sigma} = Y^\top B Y$ in a high-dimensional asymptotic framework.

2.1. *Model assumptions.* We assume that the random effects arise in the following way.

ASSUMPTION 2.1. The matrices $\alpha_1, \ldots, \alpha_k$ are independent. The rows of each $\alpha_r$ are independent, with the $i$th row given by

$$\sum_{j=1}^{\ell_r} \gamma_j^{(r)} \xi_{ij}^{(r)} + \varepsilon_i^{(r)}.$$

Here, $\gamma_1^{(r)}, \ldots, \gamma_{\ell_r}^{(r)} \in \mathbb{R}^p$ are $\ell_r$ deterministic vectors, and $\xi_{ij}^{(r)} \in \mathbb{R}$ are independent random variables satisfying

$$\mathbb{E}[\xi_{ij}^{(r)}] = 0, \qquad \mathbb{E}[(\xi_{ij}^{(r)})^2] = 1, \qquad \mathbb{E}[|\xi_{ij}^{(r)}|^k] \le C_k$$

for all $k \ge 1$ and some constants $C_k > 0$. For a covariance $\mathring{\Sigma}_r \in \mathbb{R}^{p \times p}$, the noise $\varepsilon_i^{(r)} \in \mathbb{R}^p$ is Gaussian with distribution $\varepsilon_i^{(r)} \sim \mathcal{N}(0, \mathring{\Sigma}_r)$.

Stacking $\gamma_1^{(r)}, \ldots, \gamma_{\ell_r}^{(r)}$ as the rows of

$$(2) \qquad \Gamma_r = \begin{pmatrix} - & \gamma_1^{(r)} & - \\ & \vdots & \\ - & \gamma_{\ell_r}^{(r)} & - \end{pmatrix} \in \mathbb{R}^{\ell_r \times p},$$

each $\alpha_r$ has independent rows with mean 0 and covariance of the spiked form

$$(3) \qquad \Sigma_r = \Gamma_r^\mathsf{T} \Gamma_r + \mathring{\Sigma}_r.$$

The leading term $\Gamma_r^\mathsf{T} \Gamma_r$ induces up to $\ell_r$ "signal" eigenvalues that separate from the "noise" eigenvalues of $\mathring{\Sigma}_r$. Our results should be interpreted in the setting where the noise covariance $\mathring{\Sigma}_r$ does not itself have isolated eigenvalues that separate from the bulk of its eigenvalue distribution.

As a compromise between generality of the model and simplicity of the analysis, Assumption 2.1 follows the approach in [43] and imposes a Gaussian assumption on $\varepsilon_i^{(r)}$ but not on $\xi_{ij}^{(r)}$. The signal directions $\gamma_1^{(r)}, \ldots, \gamma_{\ell_r}^{(r)}$ are not required to be orthogonal for each $r$. It is likely that our theoretical results in Theorems 2.4, 2.5 and 2.6 all remain correct under a milder moment assumption for this noise $\varepsilon_i^{(r)}$, and it may be possible to prove such an extension using cumulant expansions of the remainder terms in the Gaussian integration-by-parts formula, as done in [6]. However, we will not pursue this direction in the current work.

For the linear mixed model (1), we study an asymptotic framework summarized as follows.

ASSUMPTION 2.2. The dimensions $n, p, n_1, \ldots, n_k \to \infty$, where $k$ is a fixed constant. There are universal constants $C, c > 0$ such that for each $r = 1, \ldots, k$:

- $c < p/n < C$ and $c < n_r/n < C$,
- $\|U_r\| < C$ and $\|B\| < C/n$,
- $\|\Gamma_r\| < C$, $\|\mathring{\Sigma}_r\| < C$ and $\ell_r < C$.

Thus, the number of samples is proportional to the number of realizations of each random effect (and also to the dimension $p$). This and the assumption $\|U_r\| < C$ are discussed in greater detail in [24, 25], and hold for many classification and experimental designs. The scaling $\|B\| < C/n$ is usual for MANOVA estimators, to yield $\widehat{\Sigma}_r$ on the same scale as its estimand $\Sigma_r$.

The last statement implies a bounded number of signal eigenvalues in each variance component, where each eigenvalue remains bounded in size. It is an important open problem to extend our results beyond this setting.

2.2. *Bulk eigenvalue distribution.* Under the above assumptions, a characterization of a deterministic approximation for the empirical eigenvalue distribution of

$$\widehat{\Sigma} = Y^\top BY \tag{4}$$

was derived in [24]. We review this result here.

Consider the setting of no signal, meaning $\ell_r = 0$ and $\Sigma_r = \mathring{\Sigma}_r$ for each $r = 1, \ldots, k$. We introduce the notations $n_+ = n_1 + \cdots + n_k$ and

$$F_{rs} = \sqrt{n_r n_s} U_r^\top B U_s \in \mathbb{R}^{n_r \times n_s}, \qquad F = (F_{rs})_{r,s=1}^k \in \mathbb{R}^{n_+ \times n_+},$$

$$\operatorname{diag}_n(\mathbf{a}) = \operatorname{diag}(a_1 \operatorname{Id}_{n_1}, \ldots, a_k \operatorname{Id}_{n_k}) \in \mathbb{R}^{n_+ \times n_+}, \tag{5}$$

$$\mathbf{b} \cdot \mathring{\Sigma} = b_1 \mathring{\Sigma}_1 + \cdots + b_k \mathring{\Sigma}_k \in \mathbb{R}^{p \times p}.$$

Let $\operatorname{Tr}_r$ be the trace of the $(r, r)$ block (of size $n_r \times n_r$) in the $k \times k$ matrix block decomposition corresponding to $\mathbb{C}^{n_+} = \mathbb{C}^{n_1} \oplus \cdots \oplus \mathbb{C}^{n_k}$. The Stieltjes transform of a measure $\mu$ is $m(z) = \int (x - z)^{-1} d\mu(x)$.

THEOREM 2.3 ([24]). *Suppose Assumptions* 2.1 *and* 2.2 *hold, and* $\ell_r = 0$ *for each* $r = 1, \ldots, k$. *Let* $\widehat{\Sigma}$ *be as in* (4), *and let* $\widehat{\mu} = p^{-1} \sum_{i=1}^p \delta_{\lambda_i(\widehat{\Sigma})}$ *be the empirical distribution of its eigenvalues.*

*For each* $z \in \mathbb{C}^+$, *there exist unique $z$-dependent values* $a_1, \ldots, a_k \in \mathbb{C}^+ \cup \{0\}$ *and* $b_1, \ldots, b_k \in \overline{\mathbb{C}^+}$ *that satisfy the equations*

$$a_r = -n_r^{-1} \operatorname{Tr}((z \operatorname{Id} + \mathbf{b} \cdot \mathring{\Sigma})^{-1} \mathring{\Sigma}_r), \tag{6}$$

$$b_r = -n_r^{-1} \operatorname{Tr}_r((\operatorname{Id} + F \operatorname{diag}_n(\mathbf{a}))^{-1} F). \tag{7}$$

*The function* $m_0 : \mathbb{C}^+ \to \mathbb{C}^+$ *defined by*

$$m_0(z) = -p^{-1} \operatorname{Tr}((z \operatorname{Id} + \mathbf{b} \cdot \mathring{\Sigma})^{-1}) \tag{8}$$

*is the Stieltjes transform of a deterministic probability measure $\mu_0$ on $\mathbb{R}$, for which* $\widehat{\mu} - \mu_0 \to 0$ *weakly almost surely.*

The distribution $\mu_0$ is an $n$-dependent deterministic equivalent measure [29] for the empirical eigenvalue distribution of $\widehat{\Sigma}$. An example is depicted in Figure 1. It is defined by the noise covariances $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$ and the structure of the linear model (1), via the fixed-point equations (6)–(8).

2.3. *Noise eigenvalues stick to the support.* For any $\delta > 0$, denote the $\delta$-neighborhood of the support of the above law $\mu_0$ as

$$\text{supp}(\mu_0)_\delta = \{x \in \mathbb{R} : \text{dist}(x, \text{supp}(\mu_0)) < \delta\}.$$

We first strengthen the weak convergence statement of Theorem 2.3 to show that in the same setting of no signal, all eigenvalues of $\widehat{\Sigma}$ belong to $\text{supp}(\mu_0)_\delta$ for any fixed $\delta > 0$ and large $n$.

THEOREM 2.4. *Suppose Assumptions 2.1 and 2.2 hold, and $\ell_r = 0$ for each $r = 1, \ldots, k$. Let $\widehat{\Sigma}$ be as in (4). Then for any constant $\delta > 0$, almost surely for all large $n$,*

$$\text{spec}(\widehat{\Sigma}) \subset \text{supp}(\mu_0)_\delta.$$

We defer the proof to Appendix E.1. The proof is an application of a strong asymptotic freeness result for GOE and deterministic matrices, which we describe in Section 4.

2.4. *Limits of signal eigenvalues and eigenvectors.* We now consider the setting where $\ell_s \neq 0$ for at least one component $s \in \{1, \ldots, k\}$. This may induce "outlier" eigenvalues of $\widehat{\Sigma}$ that separate from the support of $\mu_0$—these and their eigenvectors are typically the focus of analysis in PCA. (The component where $\ell_s \neq 0$ may or may not be the component estimated by $\widehat{\Sigma} \equiv \widehat{\Sigma}_r$.)

Our main results describe the first-order limits of these eigenvalues and eigenvectors. This description involves the $z$-dependent quantities $\{b_r\}_{r=1}^k$ from Theorem 2.3. We check in Proposition E.1 that each $b_r$ extends as an analytic function in $z$ to all of $\mathbb{C} \setminus \text{supp}(\mu_0)$, and we denote this extension by $b_r(z)$. Let us write as shorthand

$$\Gamma = \begin{pmatrix} \Gamma_1 \\ \vdots \\ \Gamma_k \end{pmatrix} \in \mathbb{R}^{\ell_+ \times p}, \quad \ell_+ = \ell_1 + \cdots + \ell_k,$$

where $\Gamma_r \in \mathbb{R}^{\ell_r \times p}$ are as defined in (2). For $\lambda \in \mathbb{R} \setminus \text{supp}(\mu_0)$, let us denote

$$(9) \qquad \mathbf{b} \cdot \overset{\circ}{\Sigma} = \sum_{r=1}^k b_r(\lambda) \overset{\circ}{\Sigma}_r, \qquad \text{diag}_\ell(\mathbf{b}) = \text{diag}(b_1(\lambda) \, \text{Id}_{\ell_1}, \ldots, b_k(\lambda) \, \text{Id}_{\ell_k}).$$

Then, in the asymptotic limit, the outlier eigenvalue locations are approximated by the deterministic multiset

$$(10) \qquad \Lambda_0 = [\lambda \in \mathbb{R} \setminus \text{supp}(\mu_0) : 0 = \det T(\lambda)]$$

where, for $\lambda \in \mathbb{R} \setminus \text{supp}(\mu_0)$, we define

$$(11) \qquad T(\lambda) = \text{Id} + \Gamma(\lambda \, \text{Id} + \mathbf{b} \cdot \overset{\circ}{\Sigma})^{-1} \Gamma^{\mathsf{T}} \, \text{diag}_\ell(\mathbf{b}) \in \mathbb{R}^{\ell_+ \times \ell_+}.$$

The roots of the equation $0 = \det T(\lambda)$ are counted with their analytic multiplicities in this multiset.

THEOREM 2.5. *Suppose Assumptions 2.1 and 2.2 hold, let $\widehat{\Sigma}$ be as in (4), and let $\Lambda_0$ be defined by (10). Fix any constant $\delta > 0$. Almost surely as $n \to \infty$, there exist $\Lambda_\delta \subseteq \Lambda_0$ and $\widehat{\Lambda}_\delta \subseteq \text{spec}(\widehat{\Sigma})$, where $\Lambda_\delta$ and $\widehat{\Lambda}_\delta$, respectively, contain all elements of $\Lambda_0$ and $\text{spec}(\widehat{\Sigma})$ outside $\text{supp}(\mu_0)_\delta$, such that*

$$\text{ordered-dist}(\Lambda_\delta, \widehat{\Lambda}_\delta) \to 0.$$

Here, for two finite multisets $A, B \subset \mathbb{R}$, we denote

$$\text{ordered-dist}(A, B) = \begin{cases} \infty & \text{if } |A| \neq |B|, \\ \max_i \{|a_{(i)} - b_{(i)}|\} & \text{if } |A| = |B|, \end{cases}$$

where $a_{(i)}$ and $b_{(i)}$ are the ordered values of $A$ and $B$ counting multiplicity. We state the result as a matching of $\text{spec}(\widehat{\Sigma})$ and $\Lambda_0$, rather than convergence of $\text{spec}(\widehat{\Sigma})$ to $\Lambda_0$, as $\Lambda_0$ is also $n$-dependent. A phase-transition phenomenon analogous to that of [2] is implicit in this result, in that the cardinality of the multiset $\Lambda_0$ may transition from 0 to a positive value with the increase of signal strength in $\Gamma$.

For the corresponding outlier eigenvectors of $\widehat{\Sigma}$, the following characterizes their inner products with the signal vectors $\gamma_i^{(r)}$ that constitute the rows of $\Gamma_1, \ldots, \Gamma_k$. We denote, in addition to (9), $\partial_\lambda$ as the derivative in $\lambda$, and

$$\text{diag}_\ell(\mathbf{b}') = \partial_\lambda \text{diag}_\ell(\mathbf{b}) = \text{diag}(b_1'(\lambda) \, \text{Id}_{\ell_1}, \ldots, b_k'(\lambda) \, \text{Id}_{\ell_k}).$$

THEOREM 2.6.   *In the setting of Theorem 2.5, let $\lambda \in \Lambda_0$ be any element of multiplicity 1 such that $|\lambda - \lambda'| > \delta$ for all other $\lambda' \in \Lambda_0$, and $\text{dist}(\lambda, \text{supp}(\mu_0)) > \delta$. Let $u \in \ker T(\lambda) \subset \mathbb{R}^{\ell_+}$ be a unit vector, and let $\widehat{v}$ be the unit eigenvector for the eigenvalue $\widehat{\lambda}$ of $\widehat{\Sigma}$ closest to $\lambda$. Almost surely as $n \to \infty$, for some choice of sign of $\widehat{v}$,*

$$\Gamma \widehat{v} - \alpha^{-1/2} u \to 0,$$

*where $\alpha > 0$ is the scalar quantity defined by*

$$(12) \qquad \alpha = u^\mathsf{T} \big( -\text{diag}_\ell(\mathbf{b}) \Gamma \cdot \partial_\lambda [(\lambda \, \text{Id} + \mathbf{b} \cdot \mathring{\Sigma})^{-1}] \cdot \Gamma^\mathsf{T} \, \text{diag}_\ell(\mathbf{b}) + \text{diag}_\ell(\mathbf{b}') \big) u.$$

We show in the proof that $\ker T(\lambda)$ has dimension 1, so $u \in \ker T(\lambda)$ is unique up to sign. The above states that the inner products of the sample eigenvector $\widehat{v}$ with the true signal vectors $\gamma_i^{(r)}$ are approximately a scalar multiple of the entries of this vector $u$.

We verify in Appendix A.3 that if $\mathring{\Sigma}_r = \sigma_r^2 \, \text{Id}$ for each $r = 1, \ldots, k$, then Theorems 2.5 and 2.6 coincide with the first-order results in [25].

## 3. Implications for principal components analysis.
Theorems 2.5 and 2.6 imply several qualitative phenomena for the behavior of PCA for classical MANOVA covariance estimators in high-dimensional linear mixed models. They imply that in the asymptotic regime of Assumption 2.2, the naive sample eigenvalues and eigenvectors are inconsistent for their population counterparts, and they lead to open questions about how to obtain improved estimates in these settings. A full exploration of these questions is outside the scope of this work, but we provide some discussion of these phenomena and inferential challenges in this section.

3.1. *Qualitative phenomena.*   To illustrate the phenomena that are implied by Theorems 2.5 and 2.6, we will focus our discussion on a simple example.

Consider any mixed model (1) with $k \geq 2$ components. Suppose that $\Sigma_1$ and $\Sigma_2$ each have a rank-one signal, and $\Sigma_r$ has no signals for $r \geq 3$; that is, $\ell_1 = \ell_2 = 1$, and $\ell_r = 0$ for each $r \geq 3$. Suppose further that $\widehat{\Sigma} \equiv \widehat{\Sigma}_1$ is an unbiased MANOVA estimate of $\Sigma_1$. Denote

$$\gamma_1^{(1)} \equiv \sqrt{\mu_1} v_1, \qquad \gamma_1^{(2)} \equiv \sqrt{\mu_2} v_2$$

as the rows of $\Gamma_1$ and $\Gamma_2$, where $v_1, v_2 \in \mathbb{R}^p$ are unit vectors. For simplicity of interpretation, let us assume that $\mathring{\Sigma}_r v_1 = \mathring{\Sigma}_r v_2 = 0$ for every $r$. This implies by (3) that $\mu_1, \mu_2$ are the signal

eigenvalues in $\Sigma_1$, $\Sigma_2$, with eigenvectors $v_1$, $v_2$. (Note that our results do not require the matrices $\mathring{\Sigma}_r$ to be of full rank.) We set

$$\rho = \langle v_1, v_2 \rangle, \qquad \mu = \max(\mu_1, \mu_2).$$

We also define two $O(1)$ quantities $c_1$, $c_2$ by

$$(13) \qquad c_r = \sum_{t=1}^{k} \mathrm{Tr}(U_r^\top B U_t)(U_t B U_r^\top) \cdot \mathrm{Tr}\, \mathring{\Sigma}_t.$$

*Eigenvalue bias.* Theorem 2.5 reveals that principal eigenvalues of $\widehat{\Sigma}_1$ are biased upwards for the true eigenvalues of $\Sigma_1$. Assuming that $\mu_1$ is large and $\mu_2 \lesssim \mu_1$, we show in Appendix A.1 that the largest root of $0 = \det T(\lambda)$ has the large-$\mu$ expansion

$$(14) \qquad \lambda = \mu_1 + \mathrm{bias}, \quad \mathrm{bias} = c_1 + c_2 \frac{\mu_2 \rho^2}{\mu_1} + o_\mu(1),$$

where $o_\mu(1) \to 0$ as $\mu_1 \to \infty$. Thus, for large but fixed $\mu_1$ and $\mu_2 \lesssim \mu_1$, as $n, p \to \infty$, the sample eigenvalue is upward biased by approximately $c_1 + c_2 \mu_2 \rho^2 / \mu_1$. Here, the first term is a constant depending on the model design and level of noise, and the second term arises as an extra bias if $\mu_2$ is also large and the signal eigenvector of $\Sigma_2$ is aligned with that of $\Sigma_1$.

*Eigenvalue aliasing.* Theorem 2.5 also reveals that $\widehat{\Sigma}_1$ can have spurious "aliased" outlier eigenvalues that are not caused by signal in $\Sigma_1$, but rather by signal in $\Sigma_2$. Suppose $\mu_1 = 0$, but $\mu_2$ is large. We show in Appendix A.1 that $0 = \det T(\lambda)$ has two roots given by

$$(15) \qquad \lambda = \pm \sqrt{c_2 \mu_2} + o_\mu(1),$$

where $o_\mu(1) \to 0$ as $\mu_2 \to \infty$. Thus, $\widehat{\Sigma}_1$ has two aliased outlier eigenvalues of opposite signs. For large but fixed $\mu_2$, as $n, p \to \infty$, these aliased eigenvalues are of size proportional to $\sqrt{\mu_2}$.

*Eigenvector bias.* Theorem 2.6 implies that the sample eigenvectors of $\widehat{\Sigma}_1$ may be biased in the signal direction of $\Sigma_2$. Suppose $\mu_1 \asymp \mu_2$ are both large, and $\rho$ is bounded away from $\pm 1$. For the sample eigenvector $\widehat{v}$ corresponding to the eigenvalue described by (14), we show in Appendix A.1 that the deterministic approximation for $\Gamma \widehat{v}$ in Theorem 2.6 is

$$(16) \qquad \alpha^{-1/2} u = \Gamma v_1 + O_\mu(1/\sqrt{\mu}).$$

Here, the $O_\mu(1/\sqrt{\mu})$ term captures the error between $\widehat{v}$ and the true eigenvector $v_1$. To better understand this error, let us define a vector $w \in \mathbb{R}^2$ so that $\Gamma^\top w$ is the unit vector parallel to the component of $v_2$ orthogonal to $v_1$. We show in Appendix A.1 that the approximation for $\langle \Gamma^\top w, \widehat{v} \rangle = w^\top (\Gamma \widehat{v})$ has the large-$\mu$ expansion

$$(17) \qquad w^\top (\alpha^{-1/2} u) = \frac{c_2 \mu_2}{\mu_1^2} \rho \sqrt{1 - \rho^2} + o_\mu(1/\mu).$$

Thus, for large but fixed $\mu_1 \asymp \mu_2$, as $n, p \to \infty$, $\widehat{v}$ is biased in the direction $\Gamma^\top w$ which is orthogonal to $v_1$, of size approximately $(c_2 \mu_2 / \mu_1^2) \rho \sqrt{1 - \rho^2}$.

Figure 1 and Table 1 illustrate these phenomena in a more complex setting for a balanced one-way layout design, corresponding for example to a twin study with $n = 1600$ individuals in $n_1 = 800$ twin pairs, and $p = 3200$ traits. We simulate a rank-3 signal component $32 e_1 e_1^\top + 16 e_2 e_2^\top + 8 e_3 e_3^\top$ in $\Sigma_1$ and a rank-2 signal component $32 w w^\top + 64 e_4 e_4^\top$ in $\Sigma_2$, where $w = (e_1 + e_2 + e_3)/\sqrt{3}$, and we sample all remaining eigenvalues of $\Sigma_1$, $\Sigma_2$ from Exponential(1). Additional details are provided in Appendix A.2.

Figure 1 displays sample eigenvalues of the MANOVA estimate $\widehat{\Sigma}_1$, with numerically computed roots of $0 = \det T(\lambda)$. There are 4 positive and 2 negative roots. Of these, the 3rd
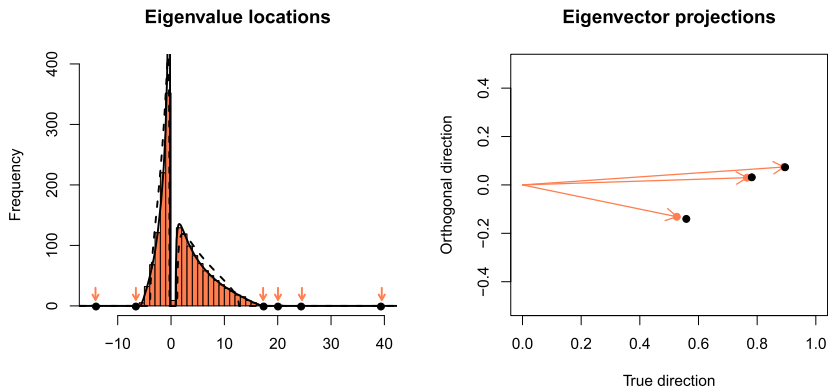
FIG. 1. *Left*: *Histogram of sample eigenvalues of the MANOVA estimate for* $\Sigma_1$ *in a one-way layout design, averaged across* 1000 *simulations, with the four largest and two smallest eigenvalues indicated by red arrows. Six black dots indicate roots of* $0 = \det T(\lambda)$, *predicting the asymptotic locations of these eigenvalues. Overlaid are the densities of the bulk law* $\mu_0$ (*solid black*) *and of* $\mu_0$ *computed from an isotropic-noise approximation* (*dashed black*). *Right*: *Inner-product of each of three principal sample eigenvectors* ($\hat{v}_j : j = 1, 2, 3$) *with the true population eigenvector* $e_j$ (*horizontal axis*) *and an orthogonal direction* $w_j$ *partially aligned with* $\Sigma_2$ (*vertical axis*). *Empirical averages across* 1000 *simulations* (*red dots/arrows*) *are overlaid with the predictions of Theorem* 2.6 (*black dots*).

largest and the smallest (negative) root are attributed to aliasing from $e_4$ in $\Sigma_2$—their sample eigenvectors are predicted by Theorem 2.6 to be orthogonal to $\{e_1, e_2, e_3\}$. The 1st, 2nd and 4th largest roots correspond to the true eigenvalues 32, 16 and 8, each observed with upward bias. For each of the three corresponding sample eigenvectors $\hat{v}_j$, Figure 1 displays its predicted and simulated alignment with the true direction $e_j$ and with the orthogonal direction $w_j$ obtained by residualizing $e_j$ out of $w$. The values of these predicted and simulated eigenvalues and eigenvector alignments are also summarized in Table 1.

3.2. *Improved estimation of principal components.* The preceding phenomena indicate that the sample eigenvalues and eigenvectors of classical MANOVA estimates for $\Sigma_1, \ldots, \Sigma_k$ are inconsistent in the regime of Assumption 2.2. The estimated eigenvalues in $\hat{\Sigma}_r$ may have upward bias, the estimated eigenvectors may be biased towards eigenvectors of other components $\Sigma_s$ for $s \neq r$, and the number of apparent signal principal eigenvectors in $\hat{\Sigma}_r$ may even be incorrect due to aliasing effects from these other components.

The probabilistic results of Theorems 2.5 and 2.6 also suggest a possible route for improved estimation of the principal eigenvalues and eigenvectors: The observed signal eigenvalues of a matrix $\hat{\Sigma} = Y^\top BY$, while inconsistent for the true signal eigenvalues of $\Sigma_1, \ldots, \Sigma_k$, do nonetheless provide some information about these matrices. As indicated by

TABLE 1
*Three principal eigenvalues* $\hat{\lambda}$ *and eigenvector inner-products* $\hat{v}_j^\top e_j$ *and* $\hat{v}_j^\top w_j$ *displayed in Figure* 1. *Theoretical predictions are computed from Theorems* 2.5 *and* 2.6, *using the true noise covariances* (*left*) *and an isotropic-noise approximation* (*middle*). *Observed values* (*right*) *are averaged over* 1000 *simulations*

| | Predicted (true) | | | Predicted (isotropic) | | | Observed | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\lambda}$ | $\hat{v}^\top e_j$ | $\hat{v}^\top w_j$ | $\hat{\lambda}$ | $\hat{v}^\top e_j$ | $\hat{v}^\top w_j$ | $\hat{\lambda}$ | $\hat{v}^\top e_j$ | $\hat{v}^\top w_j$ |
| $\mu_1 = 32$ | 39.35 | 0.90 | 0.07 | 40.22 | 0.90 | 0.07 | 39.51 | 0.89 | 0.07 |
| $\mu_2 = 16$ | 24.39 | 0.78 | 0.03 | 25.05 | 0.81 | 0.03 | 24.54 | 0.77 | 0.03 |
| $\mu_3 = 8$ | 17.36 | 0.56 | $-0.14$ | 17.37 | 0.71 | $-0.16$ | 17.26 | 0.53 | $-0.13$ |

Theorem 2.5, they correspond approximately to roots of the equation $0 = \det T(\lambda)$. This matrix $T(\lambda)$ in (11) depends on:

1. The spectra of the $k$ noise covariances $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$, and the alignments of their eigenvectors across these $k$ different components.
2. The alignments of the rows of $\Gamma$ (the true signal vectors) with these noise covariances $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$.
3. The sizes of the true signal eigenvalues and the alignments of the signal vectors across these $k$ components, which are related to the magnitudes and inner-products of the rows of $\Gamma$.

Under parametric modeling assumptions for the noise covariance matrices $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$, the observed outlier eigenvalues $\widehat{\lambda}$ for matrices of the form $\widehat{\Sigma} = Y^\top B Y$ can yield estimating equations $0 = \det T(\widehat{\lambda})$ for the true signal eigenvalues in each component $\Sigma_1, \ldots, \Sigma_k$ (as well as for the cross-component alignments of their corresponding signal eigenvectors). Furthermore, if an estimation matrix $B$ for $\widehat{\Sigma} = Y^\top B Y$ can be chosen such that the vector $u \in \ker T(\widehat{\lambda})$ is proportional to $\Gamma v$, where $\widehat{\lambda}$ is the observed eigenvalue of $\widehat{\Sigma}$ and $v$ is a true signal eigenvector of $\Sigma_r$. Then Theorem 2.6 indicates that the corresponding sample eigenvector $\widehat{v}$ of $\widehat{\Sigma}$ approximately satisfies $\Gamma \widehat{v} \propto \Gamma v$, so that $\widehat{v}$ is not asymptotically biased toward the signal direction of a different variance component $\Sigma_s$. This debiasing can, for example, lead to asymptotically consistent estimates of linear functionals of this true eigenvector $v$.

These ideas were implemented and analyzed in [25] in the simplest parametric setting where $\mathring{\Sigma}_r = \sigma_r^2 \, \mathrm{Id}$, for each $r = 1, \ldots, k$ and some scalar variance parameters $\sigma_1^2, \ldots, \sigma_k^2$. In this setting, [25] proposed a specific algorithm to solve the estimating equations $0 = \det T(\widehat{\lambda})$ arising from a parametric family of matrices $\widehat{\Sigma} = Y^\top B(\theta_1, \ldots, \theta_k) Y$ to yield estimates of all sufficiently large signal eigenvalues of $\Sigma_1, \ldots, \Sigma_k$. These estimated eigenvalues were shown to be asymptotically consistent in the high-dimensional regime of Assumption 2.2, at a parametric $1/\sqrt{n}$ rate. Furthermore, for each corresponding signal eigenvector $v$, [25] demonstrated how to obtain a specific estimation matrix $B(\theta_1, \ldots, \theta_k)$ for which the vector $u \in \ker T(\widehat{\lambda})$ indeed satisfies $u \propto \Gamma v$, and thus the algorithm returns a debiased estimate of this true eigenvector $v$. We refer readers to [25] for further details.

When $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$ are not isotropic, we believe that nonparametric estimation of their spectra and eigenvector alignments may be challenging. However, in certain more parametric contexts—for example, when $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$ capture known autocovariance structure across temporal variables or known genetic correlation structure across quantitative traits, up to a small number of unknown parameters—it may be possible to develop an estimation procedure similar to that of [25], which first estimates these parameters that describe the noise structure in $\mathring{\Sigma}_1, \ldots, \mathring{\Sigma}_k$, and then estimates the principal eigenvalues and eigenvectors of interest, using more general estimating equations that are derived from our results in Theorems 2.5 and 2.6. We leave a more detailed exploration of this possibility to future work.

Isotropic noise is often assumed in practice [46], and our results also provide an understanding of the error that may arise in the original method of [25] due to model misspecification. Figure 1 displays simulated eigenvalue densities $\mu_0$ computed using the true matrices $\mathring{\Sigma}_r$, which have exponentially decaying spectra, versus using their isotropic noise approximations with $\sigma_r^2 = p^{-1} \operatorname{Tr} \mathring{\Sigma}_r$. Table 1 compares the corresponding eigenvalue and eigenvector alignment predictions. We observe that the predictions of Theorems 2.5 and 2.6 for large outliers are very close to those under the isotropic noise approximation. This may also be understood from the calculations in the preceding section for large $\mu_1, \mu_2$, as the dependence of $c_1$, $c_2$ in (13) on $\mathring{\Sigma}_1, \mathring{\Sigma}_2$ is only through their trace. This suggests that the estimation procedure in [25] may be reasonably accurate for the larger principal eigenvalues and their associated eigenvectors. For eigenvalues closer to the support of the noise spectrum, the predictions of Theorems 2.5 and 2.6 using the true noise covariances $\mathring{\Sigma}_r$ are more accurate than those assuming isotropic noise, suggesting that inference for these principal components may be improved by better parametric modeling of the noise structure.

**4. Free probability results.** Our proofs use the connection between free probability and random matrices. Introducing representations of $U_r$, $\alpha_r$, and $B$ detailed in Section 5.1, our matrix model $\widehat{\Sigma}$ may be written as

$$(18) \qquad \widehat{\Sigma} = W + P, \quad W = \sum_{r=1}^{k} \sum_{s=1}^{k} H_r^\mathsf{T} G_r^\mathsf{T} F_{rs} G_s H_s,$$

for deterministic matrices $\{H_1, \dots, H_k\}$ and $\{F_{11}, F_{12}, \dots, F_{kk}\}$, independent matrices $\{G_1, \dots, G_k\}$ with i.i.d. Gaussian entries, and a fixed-rank perturbation $P$ (depending on $G_1, \dots, G_k$). We study the spectrum of $W$ by introducing an asymptotic approximation

$$w = \sum_{r=1}^{k} \sum_{s=1}^{k} h_r^* g_r^* f_{rs} g_s h_s,$$

where $h_r$, $g_r$, $f_{rs}$ belong to a von Neumann algebra and are conditionally free (i.e., free with amalgamation) over a diagonal subalgebra [7]. This method was also used in [24] to derive the fixed-point equations (6)–(8) in Theorem 2.3.

Our analysis develops several new tools and results in free probability theory. In this section, we state these results independent of the specific model (1), as they are of general interest for analyzing structured random matrices in other applications. We defer proofs to Appendices B, C and D.

4.1. *Augmented Cauchy and R-transforms.* We call $(\mathcal{A}, \tau)$ a *von Neumann probability space* ($W^*$-probability space) if $\mathcal{A}$ is a von Neumann algebra and $\tau : \mathcal{A} \to \mathbb{C}$ a positive, faithful, normal trace. For a von Neumann subalgebra $\mathcal{B} \subset \mathcal{A}$, we denote by

$$\tau^{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$$

the (unique) conditional expectation satisfying $\tau(\tau^{\mathcal{B}}(a)) = \tau(a)$.

We review the following definitions of $\mathcal{B}$-valued Cauchy- and $\mathcal{R}$-transforms: For each $l \geq 1$, let NC($l$) be the space of noncrossing partitions of $1, \dots, l$. For $\pi \in$ NC($l$), denote by $\kappa_\pi^{\mathcal{B}}(a_1, \dots, a_l)$ the noncrossing cumulant corresponding to $\pi$. These satisfy the moment-cumulant relations

$$(19) \qquad \tau^{\mathcal{B}}(a_1 a_2 \dots a_l) = \sum_{\pi \in \mathrm{NC}(l)} \kappa_\pi^{\mathcal{B}}(a_1, a_2, \dots, a_l).$$

Define the $\mathcal{B}$-valued Cauchy- and $\mathcal{R}$-transform of $a \in \mathcal{A}$ by

$$G_a^{\mathcal{B}}(b) = \tau^{\mathcal{B}}\big((b - a)^{-1}\big) = \sum_{l \geq 0} b^{-1}\big(ab^{-1}\big)^l, \qquad \mathcal{R}_a^{\mathcal{B}}(b) = \sum_{l \geq 1} \kappa_l^{\mathcal{B}}(a, ba, \dots, ba),$$

the former for all invertible $b \in \mathcal{B}$ with $\|b^{-1}\|$ sufficiently small and the latter for all $b \in \mathcal{B}$ with $\|b\|$ sufficiently small.[1] The moment-cumulant relations (19) yield the identity

$$(20) \qquad G_a^{\mathcal{B}}(b) = \big(b - \mathcal{R}_a^{\mathcal{B}}(G_a^{\mathcal{B}}(b))\big)^{-1}$$

for invertible $b \in \mathcal{B}$ with $\|b^{-1}\|$ sufficiently small. We refer the reader to [42], Chapter 9, for additional background and details.

---

[1]Note that, following conventions in free probability, we take the opposite sign for $G_a^{\mathcal{B}}(b)$ here as for the Stieltjes transform used in Section 2.2.

For our computations in Section 5, we will make use of the following "left-augmented" Cauchy- and $R$-transforms, defined for $a_1, a \in \mathcal{A}$ and $b \in \mathcal{B}$ by the mixed moments and mixed cumulants

$$(21) \qquad G_{a_1,a}^{\mathcal{B}}(b) = \tau^{\mathcal{B}}\big(a_1(b-a)^{-1}\big) = \sum_{l \geq 0} \tau^{\mathcal{B}}\big(a_1 b^{-1}(ab^{-1})^l\big),$$

$$(22) \qquad R_{a_1,a}^{\mathcal{B}}(b) = \sum_{l \geq 1} \kappa_l^{\mathcal{B}}(a_1, ba, \ldots, ba).$$

The following identity is then also a consequence of (19), and we provide a short proof in Appendix B.

LEMMA 4.1. *For $a_1, a \in \mathcal{A}$ and all invertible $b \in \mathcal{B}$ with $\|b^{-1}\|$ sufficiently small,*

$$(23) \qquad G_{a_1,a}^{\mathcal{B}}(b) = R_{a_1,a}^{\mathcal{B}}(G_a^{\mathcal{B}}(b))G_a^{\mathcal{B}}(b).$$

4.2. *Strong asymptotic freeness of GOE and deterministic matrices.* We establish a strong asymptotic freeness result for GOE and deterministic matrices, which is the real analogue of the GUE result in [40]. The proof is provided in Appendix C.

Fix integers $p, q \geq 0$. Let $X_1, \ldots, X_p \in \mathbb{R}^{N \times N}$ be independent GOE matrices, with diagonal entries distributed as $\mathcal{N}(0, 2/N)$ and off-diagonal entries as $\mathcal{N}(0, 1/N)$. Let $Y_1, \ldots, Y_q \in \mathbb{C}^{N \times N}$ be deterministic matrices. Denote $\mathbf{X}_N = (X_1, \ldots, X_p)$ and $\mathbf{Y}_N = (Y_1, \ldots, Y_q)$. Let $\mathrm{tr}_N = N^{-1} \mathrm{Tr}$ be the normalized matrix trace on $\mathbb{C}^{N \times N}$.

Consider an $N$-dependent von Neumann probability space $(\mathcal{A}_N, \tau_N)$. Suppose $\mathcal{A}_N$ contains $x_1, \ldots, x_p$ and $Y_1, \ldots, Y_q$, where $x_1, \ldots, x_p$ are free semicircular elements also free of $Y_1, \ldots, Y_q$, and $\tau_N \equiv \mathrm{tr}_N$ restricted to the von Neumann subalgebra $\langle Y_1, \ldots, Y_q \rangle$. Denote $\mathbf{x} = (x_1, \ldots, x_p)$.

THEOREM 4.2. *Suppose $\|Y_j\| \leq C$ for all $j = 1, \ldots, q$ and a constant $C > 0$. Then for any fixed noncommutative self-adjoint $*$-polynomial $Q$ in $p + q$ variables, and any constant $\delta > 0$, almost surely for all large $N$,*

$$(24) \qquad \mathrm{spec}\big(Q(\mathbf{X}_N, \mathbf{Y}_N)\big) \subset \mathrm{spec}\big(Q(\mathbf{x}, \mathbf{Y}_N)\big)_\delta.$$

Here, $\mathrm{spec}(Q(\mathbf{X}_N, \mathbf{Y}_N))$ are the eigenvalues of the self-adjoint random matrix $Q(\mathbf{X}_N, \mathbf{Y}_N) \in \mathbb{C}^{N \times N}$, and $\mathrm{spec}(Q(\mathbf{x}, \mathbf{Y}_N))_\delta$ is the $\delta$-neighborhood of the spectrum of the operator $Q(\mathbf{x}, \mathbf{Y}_N) \in \mathcal{A}_N$.

For our application, we will apply strong asymptotic freeness directly in the above form. However, we may also obtain as a corollary the following more usual statement, by the arguments of [40], Section 7.

THEOREM 4.3. *Let $\mathbf{x} = (x_1, \ldots, x_p)$ and $\mathbf{y} = (y_1, \ldots, y_q)$ be elements of a fixed von Neumann probability space $(\mathcal{A}, \tau)$, such that $x_1, \ldots, x_p$ are free semicircular elements also free from $\mathbf{y}$. Assume that almost surely as $N \to \infty$, for any fixed self-adjoint $*$-polynomial $P$ in $q$ variables,*

$$\mathrm{tr}_N\big[P(\mathbf{Y}_N)\big] \to \tau\big(P(\mathbf{y})\big) \quad and \quad \|P(\mathbf{Y}_N)\| \to \|P(\mathbf{y})\|.$$

*Then, almost surely for any self-adjoint $*$-polynomial $Q$ in $p + q$ variables,*

$$(25) \qquad \mathrm{tr}_N\big[Q(\mathbf{X}_N, \mathbf{Y}_N)\big] \to \tau\big(Q(\mathbf{x}, \mathbf{y})\big) \quad and \quad \|Q(\mathbf{X}_N, \mathbf{Y}_N)\| \to \|Q(\mathbf{x}, \mathbf{y})\|.$$

4.3. *Resolvent approximation using free deterministic equivalents.* We also establish a method of approximating bilinear forms in resolvents using the free deterministic equivalent framework of [52].

Fix integers $p, q \geq 0$. We study the resolvent $R(z)$ of a random matrix

$$(26) \qquad W = Q(H_1, \ldots, H_p, B_1, \ldots, B_q) \in \mathbb{C}^{N \times N},$$

where $Q$ is any self-adjoint $*$-polynomial, $H_1, \ldots, H_p$ are deterministic, and $B_1, \ldots, B_q$ are random matrices orthogonally invariant in law. For spectral arguments $z$ with constant separation from $\mathrm{spec}(W)$, and any deterministic unit vectors $u, v \in \mathbb{C}^N$, we will show an approximation

$$u^* R(z) v \approx u^* R_0(z) v,$$

where $R_0(z)$ is a deterministic matrix defined by a free deterministic equivalent model.

We consider a setup that will allow us to study rectangular matrices, following [7]: Let $\mathcal{A}_1 = \mathbb{C}^{N \times N}$ and $\tau_1 = N^{-1} \mathrm{Tr}$. Fix $k \geq 1$, let $N = N_1 + \cdots + N_k$, and consider the associated $k \times k$ block decomposition of $\mathcal{A}_1$. Define mutually orthogonal projections $P_1, \ldots, P_k \in \mathcal{A}_1$ by

$$P_r = \mathrm{diag}(0, \ldots, 0, \mathrm{Id}_{N_r}, 0, \ldots, 0)$$

with $\mathrm{Id}_{N_r}$ in the $r$th diagonal block. Then $(\mathcal{A}_1, \tau_1, P_1, \ldots, P_k)$ is a rectangular probability space in the sense of [7]. Define the subalgebra $\mathcal{D} \subset \mathcal{A}_1$ generated by $P_1, \ldots, P_k$, given explicitly by

$$\mathcal{D} = \{z_1 P_1 + \cdots + z_k P_k : z_1, \ldots, z_k \in \mathbb{C}\}.$$

Define also the space of block-diagonal orthogonal matrices

$$\mathcal{O} = \{\mathrm{diag}(O_1, \ldots, O_k) : O_r \in \mathbb{R}^{N_r \times N_r}, O_r^\mathsf{T} O_r = \mathrm{Id} \text{ for each } r\}.$$

Consider $H_1, \ldots, H_p, B_1, \ldots, B_q \in \mathcal{A}_1$, where $H_1, \ldots, H_p$ are deterministic, and $(B_1, \ldots, B_q)$ is random and equal in joint law to $(O B_1 O^\mathsf{T}, \ldots, O B_q O^\mathsf{T})$ for all $O \in \mathcal{O}$. For a self-adjoint $*$-polynomial $Q$ in $p + q$ arguments with coefficients in $\mathcal{D}$, define $W$ by (26), and define its resolvent

$$R(z) = (W - z\,\mathrm{Id})^{-1}.$$

To define the approximation $R_0(z)$, we construct a free deterministic equivalent model: Let $(\mathcal{A}_2, \tau_2)$ be a second von Neumann probability space, where $\mathcal{D} \subset \mathcal{A}_2$ and $\tau_1 \equiv \tau_2$ restricted to $\mathcal{D}$. Let $\mathcal{A}_2$ have elements $b_1, \ldots, b_q$ satisfying

$$(27) \qquad N^{-1} \mathrm{Tr}\big(P(B_1, \ldots, B_q)\big) - \tau_2\big(P(b_1, \ldots, b_q)\big) \to 0$$

almost surely as $N \to \infty$, for any fixed $*$-polynomial $P$ with coefficients in $\mathcal{D}$. Define the von Neumann amalgamated free product over $\mathcal{D}$,

$$(\mathcal{A}, \tau) = (\mathcal{A}_1, \tau_1) *_{\mathcal{D}} (\mathcal{A}_2, \tau_2),$$

so that $(H_1, \ldots, H_p)$ is free of $(b_1, \ldots, b_q)$ with amalgamation over $\mathcal{D}$. Define the free deterministic equivalent approximation to $W$ by

$$w = Q(H_1, \ldots, H_p, b_1, \ldots, b_q) \in \mathcal{A}.$$

Finally, let $\mathcal{H} = \langle H_1, \ldots, H_p, \mathcal{D} \rangle$ be the generated von Neumann subalgebra of $\mathcal{A}$, and let $\tau^{\mathcal{H}} : \mathcal{A} \to \mathcal{H}$ be the conditional expectation onto $\mathcal{H}$ that satisfies $\tau(\tau^{\mathcal{H}}(a)) = \tau(a)$. Importantly, note that for any $a \in \mathcal{A}$,

$$\tau^{\mathcal{H}}(a) \in \mathcal{H} \subset \mathcal{A}_1 \equiv \mathbb{C}^{N \times N},$$

so that $\tau^{\mathcal{H}}(a)$ is an $N \times N$ matrix. We define the free deterministic approximation $R_0(z)$ of $R(z)$ by

$$(28) \qquad R_0(z) = \tau^{\mathcal{H}}((w - z)^{-1}).$$

We now state our approximation result, whose proof is in Appendix D.

THEOREM 4.4 (Resolvent approximation). *For some constants $C, c > 0$, suppose $c < N_r/N < C$, $\|H_i\| < C$ and $\|B_j\| < C$ for all $r, i, j$, almost surely for all large $N$. Fix any constant $\delta > 0$ and set*

$$\mathbb{D} = \{z \in \mathbb{C} : \text{dist}(z, \text{spec}(w)) \geq \delta \text{ and } \text{dist}(z, \text{spec}(W)) \geq \delta\}.$$

*Then for any (sequence of) deterministic unit vectors $u, v \in \mathbb{C}^N$, almost surely as $N \to \infty$,*

$$(29) \qquad \sup_{z \in \mathbb{D}} |u^* R(z) v - u^* R_0(z) v| \to 0.$$

Taking $k = 1$ yields a result for square orthogonally invariant matrices, where $(\mathcal{A}, \tau)$ is the von Neumann free product over $\mathcal{D} \equiv \mathbb{C}$. We consider $k \geq 2$ to encompass applications with rectangular matrices, where each $H_i$, $B_j$ typically has a single off-diagonal block which is nonzero. We are then interested in $*$-polynomials $Q$ that are $(1, 1)$-simple, that is, $W$ and $w$ satisfy

$$W = P_1 W P_1, \qquad w = P_1 w P_1.$$

Denote by $W_{11} \in \mathbb{C}^{N_1 \times N_1}$ the $(1, 1)$-block of $W$. Corresponding to $\mathbb{C}^{N_1 \times N_1}$ is a "compressed algebra" $\mathcal{A}^c = \{P_1 a P_1 : a \in \mathcal{A}\}$ with unit $P_1$ [52]. Denote by $w_{11} \in \mathcal{A}^c$ and $\text{spec}(w_{11})$ the element $w$ and its spectrum, viewed as a self-adjoint operator in $\mathcal{A}^c$. We then have the following corollary.

COROLLARY 4.5. *In the setting of Theorem 4.4, suppose in addition that $W = P_1 W P_1$ and $w = P_1 w P_1$, and let $W_{11}$ and $w_{11}$ be as above. Let $(R_0(z))_{11} \in \mathbb{C}^{N_1 \times N_1}$ be the $(1, 1)$-block of $R_0(z) = \tau^{\mathcal{H}}((w - z)^{-1})$, and set*

$$\mathbb{D}_1 = \{z \in \mathbb{C} : \text{dist}(z, \text{spec}(w_{11})) \geq \delta \text{ and } \text{dist}(z, \text{spec}(W_{11})) \geq \delta\}.$$

*Then for any (sequence of) deterministic unit vectors $u_1, v_1 \in \mathbb{C}^{N_1}$, almost surely as $N \to \infty$,*

$$(30) \qquad \sup_{z \in \mathbb{D}_1} |u_1^* (W_{11} - z \, \text{Id})^{-1} v_1 - u_1^* (R_0(z))_{11} v_1| \to 0.$$

**5. Analysis of the linear mixed model.** In this section, we give a high-level outline of the proofs of Theorems 2.5 and 2.6, which follow the perturbative approach of [8]. We present the main steps of the computations, deferring technical details to Appendix E.

We assume implicitly throughout that Assumptions 2.1 and 2.2 hold. We denote by $C, c > 0$ constants which may change from instance to instance. We fix a constant $\delta > 0$, and define

$$U_\delta = \{z \in \mathbb{C} : \text{dist}(z, \text{supp}(\mu_0)) > \delta\}.$$

We denote $\|X\|_\infty = \max_{i, j} |X_{i, j}|$. For $n$-dependent matrices $X_1(z), X_2(z)$ of the same (bounded) dimension, we write

$$X_1(z) \sim X_2(z)$$

if almost surely as $n \to \infty$, we have

$$\sup_{z \in U_\delta} \|X_1(z) - X_2(z)\|_\infty \to 0.$$

5.1. *Model and deterministic equivalent measure.* We first clarify the form of $\widehat{\Sigma}$ and the free probability interpretation of the measure $\mu_0$. Introducing $\Gamma_r \in \mathbb{R}^{\ell_r \times p}$ as in (2), and defining

$$\Xi_r = \frac{1}{\sqrt{n_r}}(\xi_{ij}^{(r)})_{i,j} \in \mathbb{R}^{n_r \times \ell_r}, \qquad E_r = \begin{pmatrix} - & \varepsilon_1^{(r)} & - \\ & \vdots & \\ - & \varepsilon_n^{(r)} & - \end{pmatrix} \in \mathbb{R}^{n_r \times p},$$

the random effect matrix $\alpha_r$ is written concisely as

(31)                                    $$\alpha_r = \sqrt{n_r}\,\Xi_r \Gamma_r + E_r.$$

Write further

(32)                                    $$E_r = \sqrt{n_r}\,G_r H_r,$$

where $G_r \in \mathbb{R}^{n_r \times p}$ has i.i.d. $\mathcal{N}(0, 1/n_r)$ entries and $H_r = \overset{\circ}{\Sigma}{}^{1/2} \in \mathbb{R}^{p \times p}$. Then, when $\ell_r = 0$ and $\alpha_r = E_r$ for all $r$, we obtain

$$\widehat{\Sigma} = W \equiv \sum_{r,s=1}^{k} H_r^{\mathsf{T}} G_r^{\mathsf{T}} F_{rs} G_s H_s,$$

where $\{F_{rs} : r, s = 1, \ldots, k\}$ are defined in (5). More generally, we have

(33)                $$\widehat{\Sigma} = \sum_{r,s=1}^{k} (\Xi_r \Gamma_r + G_r H_r)^{\mathsf{T}} F_{rs} (\Xi_s \Gamma_s + G_s H_s) = W + P$$

for $W$ as above, and for the low-rank perturbation

(34)        $$P = \sum_{r,s=1}^{k} (\Gamma_r^{\mathsf{T}} \Xi_r^{\mathsf{T}} F_{rs} G_s H_s + H_r^{\mathsf{T}} G_r^{\mathsf{T}} F_{rs} \Xi_s \Gamma_s + \Gamma_r^{\mathsf{T}} \Xi_r^{\mathsf{T}} F_{rs} \Xi_s \Gamma_s).$$

The proof of Theorem 2.3 in [24] used a free probability approach. As the matrices $G_r$ and $F_{rs}$ in this model are rectangular, asymptotic freeness was formally expressed using the ideas of [7], by embedding these matrices in a larger square matrix, and establishing asymptotic freeness with amalgamation over a subalgebra generated by block-identity matrices along the diagonal.

More specifically, the proof in [24], Section 4, illustrates that $\mu_0$ is a spectral measure in the following model: Set $N = (k + 1)p + n_1 + \cdots + n_k$. Embed $\{F_{rs}, G_r, H_r : r, s = 1, \ldots, k\}$ into $\mathbb{C}^{N \times N}$ by zero-padding, in the following blocks of the $(2k + 1) \times (2k + 1)$ block decomposition for $\mathbb{C}^N = \mathbb{C}^p \oplus \cdots \oplus \mathbb{C}^p \oplus \mathbb{C}^{n_1} \cdots \oplus \mathbb{C}^{n_k}$:

(35)    $$\begin{pmatrix} & H_1^* & \cdots & H_k^* & & & \\ H_1 & & & & & G_1^* & \\ \vdots & & & & & & \ddots \\ H_k & & & & & & & G_k^* \\ & G_1 & & & F_{11} & \cdots & F_{1k} \\ & & \ddots & & \vdots & \ddots & \vdots \\ & & & G_k & F_{k1} & \cdots & F_{kk} \end{pmatrix}$$

Denote by $\tilde{F}_{rs}$, $\tilde{G}_r$ and $\tilde{H}_r \in \mathbb{C}^{N \times N}$ these embedded matrices. Consider the mutually orthogonal projections

$$P_0 = \mathrm{diag}(\mathrm{Id}_p, 0, \ldots, 0), \quad \ldots, \quad P_{2k} = \mathrm{diag}(0, \ldots, 0, \mathrm{Id}_{n_k})$$

corresponding to the $2k + 1$ diagonal blocks of $\mathbb{C}^{N \times N}$. Then the block structure of this embedding induces an asymptotic freeness of the families $\{F_{rs}\}$, $\{H_r\}$, and individual matrices $G_1, \ldots, G_k$ with amalgamation over the subalgebra generated by these projections $\{P_0, \ldots, P_{2k}\}$.

Let $(\mathcal{A}, \tau)$ be a von Neumann probability space containing mutually orthogonal projections $p_0, p_1, \ldots, p_{2k}$, which analogously satisfy $\tau(p_0) = \cdots = \tau(p_k) = p/N$ and $\tau(p_{k+r}) = n_r/N$ for each $r = 1, \ldots, k$. Let $(\mathcal{A}, \tau)$ also contain $\{f_{rs}, g_r, h_r : r, s = 1, \ldots, k\}$ such that:

1. $p_{k+r} f_{rs} p_{k+s}$, $p_{k+r} g_r p_r = g_r$, and $p_r h_r p_0 = h_r$.
2. For any noncommutative $*$-polynomial $Q$ of $k$ variables,

$$\tau\big(Q(h_1, \ldots, h_k)\big) = N^{-1} \operatorname{Tr} Q(\tilde{H}_1, \ldots, \tilde{H}_k).$$

Similarly, for any noncommutative $*$-polynomial $Q$ of $k^2$ variables,

$$\tau\big(Q(f_{11}, f_{12}, \ldots, f_{kk})\big) = N^{-1} \operatorname{Tr} Q(\tilde{F}_{11}, \tilde{F}_{12}, \ldots, \tilde{F}_{kk}).$$

3. For each $r \in \{1, \ldots, k\}$ and $l \geq 0$,

$$\frac{N}{p} \tau\big((g_r^* g_r)^l\big) = \int x^l \nu_{p/n_r}(dx)$$

where $\nu$ is the Marcenko–Pastur law with parameter $p/n_r$.

4. The families $\{f_{rs} : r, s = 1, \ldots, k\}$, $\{h_r : r = 1, \ldots, k\}$, and individual elements $g_1, \ldots, g_k$ are free with amalgamation over the von Neumann subalgebra $\mathcal{D} = \langle p_0, \ldots, p_{2k} \rangle$.

Define a free deterministic equivalent for $W$ by

$$(36) \qquad w = \sum_{r,s=1}^{k} h_r^* g_r^* f_{rs} g_s h_s.$$

Only the $(0, 0)$-block of $w$ is nonzero—this corresponds to $W$ belonging to the $(0, 0)$ block in the embedded space (35). Thus $w$ is an element of the compressed algebra $\mathcal{A}^c = \{a \in \mathcal{A} : a = p_0 a p_0\}$, which has unit $p_0$ and trace $\tau^c(a) = (N/p)\tau(p_0 a p_0)$. The analysis of [24], Section 4, shows that the law $\mu_0$ in Theorem 2.3 is the $\tau^c$-distribution of $w$. This means that for any continuous function $f : \mathbb{R} \to \mathbb{C}$, we have

$$\int f(x) \, d\mu_0(x) = \tau^c(f(w))$$

where $f(w)$ is defined by the functional calculus on $\mathcal{A}^c$. Since $\tau$ is a faithful trace, so is $\tau^c$ as a trace on $\mathcal{A}^c$, and thus (cf. [45], Propositions 3.13 and 3.15)

$$(37) \qquad \operatorname{supp}(\mu_0) = \operatorname{spec}(w)$$

where $\operatorname{spec}(w)$ is the spectrum of $w$ as an element of $\mathcal{A}^c$.

5.2. *Master equation.* Following [8], we first establish a "master equation" characterizing outlier eigenvalues of $\widehat{\Sigma}$.

Recall the form (33) for $\widehat{\Sigma}$. Letting $\ell$ be the rank of $\Gamma$ (so $\ell \leq \ell_+$), write

$$\Gamma = \tilde{\Gamma} Q^\mathsf{T},$$

where $Q \in \mathbb{R}^{p \times \ell}$ contains the right singular vectors of $\Gamma$. We have $Q^\mathsf{T} Q = \operatorname{Id}_\ell$ and $\|\tilde{\Gamma}\| \leq C$. Denote the resolvent of $W$ by

$$R(z) = (W - z \operatorname{Id})^{-1}.$$

Define the block-diagonal matrices

$$
\Xi = \begin{bmatrix} \Xi_1 & & & \\ & \Xi_2 & & \\ & & \ddots & \\ & & & \Xi_k \end{bmatrix} \in \mathbb{R}^{n_+ \times \ell_+}, \qquad G = \begin{bmatrix} G_1 & & & \\ & G_2 & & \\ & & \ddots & \\ & & & G_k \end{bmatrix} \in \mathbb{R}^{n_+ \times kp}.
$$

Finally, define $H \in \mathbb{R}^{kp \times p}$ as the vertical stacking of $\{H_r\}_{r=1}^k$, and set

$$
S(z) = \Xi^\mathsf{T} F G H R(z) Q.
$$

In Appendix E.2, we write the low-rank perturbation matrix $P$ in (34) as $P = P_1 P_2$ for two rectangular matrices $P_1$ and $P_2$. We then apply the identity $\det(\mathrm{Id} + R(z) P_1 P_2) = \det(\mathrm{Id} + P_2 R(z) P_1)$ to obtain the following result.

LEMMA 5.1. *The eigenvalues of $\widehat{\Sigma}$ which are not eigenvalues of $W$ are the roots of* $\det \widehat{K}(z) = 0$, *where*

$$
(38) \quad \widehat{K}(z) = \mathrm{Id} + \begin{bmatrix} S(z) \cdot \widetilde{\Gamma}^\mathsf{T} & \Xi^\mathsf{T} F G H R(z) H^\mathsf{T} G^\mathsf{T} F \Xi \cdot \widetilde{\Gamma} + S(z) \cdot \widetilde{\Gamma}^\mathsf{T} \Xi^\mathsf{T} F \Xi \widetilde{\Gamma} \\ Q^\mathsf{T} R(z) Q \cdot \widetilde{\Gamma}^\mathsf{T} & S(z)^\mathsf{T} \cdot \widetilde{\Gamma} + Q^\mathsf{T} R(z) Q \cdot \widetilde{\Gamma}^\mathsf{T} \Xi^\mathsf{T} F \Xi \widetilde{\Gamma} \end{bmatrix}.
$$

Denote the four blocks of this matrix as $\widehat{K}_{11}$, $\widehat{K}_{12}$, $\widehat{K}_{21}$, $\widehat{K}_{22}$. When $\widehat{K}_{11}$ is invertible, the condition $\det \widehat{K}(z) = 0$ is equivalent to $\det \widehat{T}(z) = 0$ for the Schur complement

$$
(39) \qquad\qquad \widehat{T}(z) = \widehat{K}_{22}(z) - \widehat{K}_{21}(z) \widehat{K}_{11}(z)^{-1} \widehat{K}_{12}(z).
$$

Observe that each matrix $\Xi^\top (\dots) \Xi$ in the definition of $\widehat{K}$ has bounded dimension $\ell_+ \times \ell_+$, each matrix $\Xi^\top (\dots) Q$ has bounded dimension $\ell_+ \times \ell$, and $\Xi$ is independent of $G$ and $R(z)$. Then, conditioning on $G$ and $R(z)$ and applying concentration inequalities for linear and bilinear forms in $\Xi$, we obtain that $\widehat{T}(z)$ is approximated by a matrix

$$
(40) \qquad \check{T}(z) = \mathrm{Id} + Q^\mathsf{T} R(z) Q \cdot \widetilde{\Gamma}^\mathsf{T} \left( \sum_{r=1}^k n_r^{-1} \mathrm{Tr}_r \big[ F - FGHR(z)H^\mathsf{T}G^\mathsf{T}F \big] \mathrm{Id}_{\ell_r} \right) \widetilde{\Gamma}.
$$

This is formalized in the following result, proven in Appendix E.2.

LEMMA 5.2. *We have that $S(z) \sim 0$, $\widehat{K}_{11}(z) \sim \mathrm{Id}_{\ell_+}$, and $\widehat{T}(z) \sim \check{T}(z)$.*

The outlier eigenvalues of $\widehat{\Sigma}$ will be approximate roots of $0 = \det \check{T}(z)$, where this matrix $\check{T}(z)$ no longer depends on the randomness in $\Xi$.

5.3. *Approximation by deterministic equivalents.* The main step of the proof is to approximate the $G$- and $R(z)$-dependent terms appearing in (40) by deterministic quantities. We do this using a free deterministic equivalent approach. Define

$$
\widetilde{T}(z) = \mathrm{Id} + Q^\mathsf{T} (z\,\mathrm{Id} + \mathbf{b} \cdot \mathring{\Sigma})^{-1} Q \cdot \widetilde{\Gamma}^\mathsf{T} \mathrm{diag}_\ell(\mathbf{b}) \widetilde{\Gamma},
$$

with notation as in Theorem 2.5. Our goal is to show the following lemma.

LEMMA 5.3. *We have $\check{T}(z) \sim \widetilde{T}(z)$.*

This requires approximating the two terms in $\check{T}$ by those in $\widetilde{T}$. For approximating the first term, as perhaps can be guessed from the form of the Stieltjes transform (8), the matrix $-(z\,\mathrm{Id} + \mathbf{b} \cdot \mathring{\Sigma})^{-1}$ is a deterministic equivalent for the resolvent $R(z)$. We verify this in the following result, using the resolvent approximation techniques in Section 4.3 and Theorem 4.4.

PROPOSITION 5.4.    *We have* $Q^{\mathsf{T}} R(z) Q \sim -Q^{\mathsf{T}} (z \operatorname{Id} + \mathbf{b} \cdot \mathring{\Sigma})^{-1} Q$.

PROOF.    The von Neumann probability space $(\mathcal{A}, \tau)$ in Section 5.1 may be constructed as follows: Let $(\mathcal{A}_1, \tau_1) = (\mathbb{C}^{N \times N}, N^{-1} \operatorname{Tr})$, containing the embeddings of the matrices $H_1, \ldots, H_k$ and $P_0, \ldots, P_{2k}$. Denote these elements of $\mathcal{A}_1$ also by $h_r$ and $p_r$. Construct a von Neumann probability space $(\mathcal{A}_2, \tau_2)$ also containing $p_0, \ldots, p_{2k}$ and elements $\{f_{rs}, g_r : r, s = 1, \ldots, k\}$ satisfying all required conditions on their joint law under $\tau_2$. Let $(\mathcal{A}, \tau)$ be the von Neumann amalgamated free product over $\langle p_0, \ldots, p_{2k} \rangle$.

Let $w = \sum_{r,s} h_r^* g_r^* f_{rs} g_s h_s \in \mathcal{A}$. By Corollary 4.5 applied to each pair of columns of $Q$, we find that

$$Q^{\mathsf{T}} R(z) Q \sim Q^{\mathsf{T}} P_0 \tau^{\mathcal{H}} ((w - z)^{-1}) P_0 Q,$$

where $P_0 \tau^{\mathcal{H}} ((w - z)^{-1}) P_0$ is identified with its $(0,0)$-block as an element of $\mathbb{C}^{p \times p}$. This $\tau^{\mathcal{H}}$ trace was computed in [24], equation (4.12), to be

$$\tau^{\mathcal{H}} ((w - z)^{-1}) = -\left( z + \sum_{r=1}^{k} h_r^* h_r b_r(z) \right)^{-1},$$

using the identification $\beta_r(z) = -b_r(z)$ at the conclusion of the proof of [24], Lemma 4.4. The $(0,0)$-block of this matrix is exactly

$$-\left( z + \sum_{r=1}^{k} H_r^{\mathsf{T}} H_r b_r(z) \right)^{-1} = -(z + \mathbf{b} \cdot \mathring{\Sigma})^{-1}. \qquad \square$$

Lemma 5.3 now follows by applying Proposition 5.4 and the following approximation for the second term of $\check{T}(z)$.

PROPOSITION 5.5.    *For each* $t \in \{1, \ldots, k\}$, *we have*
$$n_t^{-1} \operatorname{Tr}_t [F - FGHR(z)H^{\mathsf{T}}G^{\mathsf{T}}F] \sim -b_t(z).$$

In the remainder of this section, we prove Proposition 5.5. We apply a computation using the augmented Cauchy- and $\mathcal{R}$-transforms of Section 4.1. In the von Neumann probability space $(\mathcal{A}, \tau)$ of Section 5.1, let $\mathcal{H} = \langle h_1, \ldots, h_k \rangle$, $\mathcal{G} = \langle g_1, \ldots, g_k \rangle$, $\mathcal{F} = \langle f_{11}, f_{12}, \ldots, f_{kk} \rangle$ and $\mathcal{D} = \langle p_0, \ldots, p_{2k} \rangle$ be the generated von Neumann subalgebras of $\mathcal{A}$. Define the elements

$$(41) \qquad w = \sum_{r,s=1}^{k} h_r^* g_r^* f_{rs} g_s h_s, \qquad v = \sum_{r,s=1}^{k} g_r^* f_{rs} g_s, \qquad u = \sum_{r,s=1}^{k} f_{rs}.$$

For any $r, s, t \in \{1, \ldots, k\}$ define

$$a_{rts} = h_r^* g_r^* f_{rt} f_{ts} g_s h_s, \qquad b_{rts} = g_r^* f_{rt} f_{ts} g_s, \qquad c_{rts} = f_{rt} f_{ts}.$$

Our goal is to compute

$$\sum_{r,s=1}^{k} \tau (f_{ts} g_s h_s (w - z)^{-1} h_r^* g_r^* f_{rt}) = \sum_{s,t=1}^{k} \tau (a_{rts} (w - z)^{-1}),$$

which is the free approximation for $\operatorname{Tr}_t FGHR(z)H^{\mathsf{T}}G^{\mathsf{T}}F$.

For $a \in \mathcal{A}$ and $h \in \mathcal{H}$, define the $\mathcal{H}$-valued conditional expectation $\tau^{\mathcal{H}}(a)$, Cauchy-transform $G_a^{\mathcal{H}}(h)$, and $\mathcal{R}$-transform $\mathcal{R}_a^{\mathcal{H}}(h)$, and similarly for $\mathcal{G}$ and $\mathcal{D}$. For each $i \in \{0, \ldots, 2k\}$, denote

$$\tau_i(a) = \tau(p_i)^{-1} \tau(p_i a p_i)$$

and note that $\tau^{\mathcal{D}}(a) = \sum_i \tau_i(a) p_i$. For a sufficiently large constant $C > 0$, define

$$\mathbb{D} = \{z \in \mathbb{C} : |z| > C\}.$$

We define the following analytic functions $\{\alpha_i\}_{i=0}^{2k}$, $\{\beta_i\}_{i=0}^{2k}$, $\{d_i\}_{i=0}^{2k}$, $\{\gamma_j\}_{j=0}^{2k}$, $\{\delta_j\}_{j=0}^{2k}$, and $\{e_j\}_{j=0}^{2k}$ on $\mathbb{D}$, also used in [24]: For $i = 1, \ldots, k$, define

$$(42) \qquad \alpha_i = \tau_i(h_i G_w^{\mathcal{H}}(z) h_i^*), \qquad \beta_i = \tau_i\left(R_v^{\mathcal{D}}\left(\sum_{i=1}^{k} \alpha_i p_i\right)\right).$$

Set $\alpha_0 = \alpha_{k+1} = \cdots = \alpha_{2k} = |z|^{-1}$ and $\beta_0 = \beta_{k+1} = \cdots = \beta_{2k} = 0$, and

$$d_i = \alpha_i^{-1} + \beta_i, \qquad d = \sum_{i=0}^{2k} d_i p_i.$$

Now, for $j = 1, \ldots, k$, define

$$(43) \qquad \gamma_{j+k} = \tau_{j+k}(g_j G_v^{\mathcal{G}}(d) g_j^*), \qquad \delta_{j+k} = \tau_{j+k}\left(R_u^{\mathcal{D}}\left(\sum_{j=k+1}^{2k} \gamma_j p_j\right)\right).$$

Set $\gamma_0 = \gamma_1 = \cdots = \gamma_k = |z|^{-1}$ and $\delta_0 = \delta_1 = \cdots = \delta_k = 0$, and

$$e_j = \gamma_j^{-1} + \delta_j, \qquad e = \sum_{j=0}^{2k} e_j p_j.$$

The following identities are shown in [24].

PROPOSITION 5.6. *For all $z \in \mathbb{D}$,*

(a) $\sum_{i=0}^{2k} \alpha_i p_i = G_v^{\mathcal{D}}(d)$.
(b) $\sum_{j=0}^{2k} \gamma_j p_j = G_u^{\mathcal{D}}(e)$.
(c) *The quantities $a_r = -\frac{p\alpha_r}{n_r}$ and $b_r = -\beta_r$ satisfy the relations* (6)–(7).
(d) *For $r = 1, \ldots, k$, we have $e_{r+k} = -a_r^{-1}$.*

PROOF. (a) follows from [24], equation (4.15), (b) follows from [24], equation (4.21), (c) is shown at the end of the proof of [24], Lemma 4.4, and (d) follows from [24], equation (4.28). □

The following identities are similar to [24], Lemma 4.3.

PROPOSITION 5.7. *We have*

$$R_{a_{rts},w}^{\mathcal{H}}(G_w^{\mathcal{H}}(z)) = h_r^* h_r \tau_r[R_{b_{rts},v}^{\mathcal{D}}(G_v^{\mathcal{D}}(d))],$$
$$R_{b_{rts},v}^{\mathcal{G}}(G_v^{\mathcal{G}}(d)) = g_r^* g_r \tau_{r+k}[R_{c_{rts},u}^{\mathcal{D}}(G_u^{\mathcal{D}}(e))].$$

PROOF. For the first equality, notice that for $c = G_w^{\mathcal{H}}(z)$, we have

$$\kappa_l^{\mathcal{H}}(a_{rts}, cw, \ldots, cw)$$

$$= \sum_{\substack{r_2, \ldots, r_l = 1 \\ s_2, \ldots, s_l = 1}}^{k} \kappa_l^{\mathcal{H}}(h_r^* g_r^* f_{rt} f_{ts} g_s h_s, ch_{r_2}^* g_{r_2}^* f_{r_2 s_2} g_{s_2} h_{s_2}, \ldots, ch_{r_l}^* g_{r_l}^* f_{r_l s_l} g_{s_l} h_{s_l})$$

$$= \sum_{\substack{r_2,\ldots,r_l=1 \\ s_2,\ldots,s_l=1}}^{k} h_r^* \kappa_l^{\mathcal{H}} (g_r^* f_{rt} f_{ts} g_s, h_s c h_{r_2}^* g_{r_2}^* f_{r_2 s_2} g_{s_2}, \ldots, h_{s_{l-1}} c h_{r_l}^* g_{r_l}^* f_{r_l s_l} g_{s_l}) h_{s_l}$$

$$= \sum_{\substack{r_2,\ldots,r_l=1 \\ s_2,\ldots,s_l=1}}^{k} h_r^* \kappa_l^{\mathcal{D}} (g_r^* f_{rt} f_{ts} g_s, \tau^{\mathcal{D}}(h_s c h_{r_2}^*) g_{r_2}^* f_{r_2 s_2} g_{s_2}, \ldots, \tau^{\mathcal{D}}(h_{s_{l-1}} c h_{r_l}^*) g_{r_l}^* f_{r_l s_l} g_{s_l}) h_{s_l},$$

where we apply [44], Theorem 3.6, and $\mathcal{D}$-freeness of $\{\mathcal{F}, \mathcal{G}\}$ and $\mathcal{H}$ in the last step. Notice now that $\tau^{\mathcal{D}}(h_s c h_r^*) = 0$ unless $s = r$, that for any $d' \in \mathcal{D}$ we have $h_r^* d' h_r = h_r^* h_r \tau_r(d')$, and that

$$\tau^{\mathcal{D}}(h_r c h_r^*) g_r^* = \tau_r(h_r c h_r^*) p_r g_r^* = \left(\sum_{i=0}^{2k} \alpha_i p_i\right) g_r^*.$$

Therefore, applying Proposition 5.6(a) and defining $c' = G_v^{\mathcal{D}}(d)$, the above is equal to

$$h_r^* h_r \sum_{r_3,\ldots,r_l=1}^{k} \tau_r (\kappa_l^{\mathcal{D}} (g_r^* f_{rt} f_{ts} g_s, c' g_s^* f_{sr_3} g_{r_3}, c' g_{r_3}^* f_{r_3 r_4} g_{r_4}, \ldots, c' g_{r_l}^* f_{r_l r} g_r)).$$

On the other hand, using $g_s = g_s p_s$ and $p_s c' p_r = 0$ unless $s = r$, we have

$$\kappa_l^{\mathcal{D}} (b_{rts}, c'v, \ldots, c'v)$$

$$= \sum_{\substack{r_2,\ldots,r_l=1 \\ s_2,\ldots,s_l=1}}^{k} \kappa_l^{\mathcal{D}} (g_r^* f_{rt} f_{ts} g_s, c' g_{r_2}^* f_{r_2 s_2} g_{s_2}, \ldots, c' g_{r_l}^* f_{r_l s_l} g_{s_l})$$

$$= \sum_{\substack{r_2,\ldots,r_l=1 \\ s_2,\ldots,s_l=1}}^{k} \kappa_l^{\mathcal{D}} (g_r^* f_{rt} f_{ts} g_s, p_s c' p_{r_2} g_{r_2}^* f_{r_2 s_2} g_{s_2}, \ldots, p_{s_{l-1}} c' p_{r_l} g_{r_l}^* f_{r_l s_l} g_{s_l})$$

$$= \sum_{r_3,\ldots,r_l=1}^{k} \kappa_l^{\mathcal{D}} (g_r^* f_{rt} f_{ts} g_s, c' g_s^* f_{sr_3} g_{r_3}, c' g_{r_3}^* f_{r_3 r_4} g_{r_4}, \ldots, c' g_{r_l}^* f_{r_l r} g_r).$$

Comparing with the above,

$$\kappa_l^{\mathcal{H}} (a_{rts}, cw, \ldots, cw) = h_r^* h_r \tau_r (\kappa_l^{\mathcal{D}} (b_{rts}, c'v, \ldots, c'v)).$$

Summing over $l \geq 1$ yields the first identity. The proof of the second identity is exactly parallel, using Proposition 5.6(b) in place of Proposition 5.6(a). $\square$

PROPOSITION 5.8. *We have*

$$\tau(a_{rts}(z - w)^{-1}) = \tau(c_{rts}(e - u)^{-1}).$$

PROOF. Note first that

$$\tau(a_{rts}(z - w)^{-1}) = \tau(G_{a_{rts},w}^{\mathcal{H}}(z)) = \tau(p_0)\tau_0(G_{a_{rts},w}^{\mathcal{H}}(z)).$$

Substituting the expression of Proposition 5.7 into the identity

$$G_{a_{rts},w}^{\mathcal{H}}(z) = R_{a_{rts},w}^{\mathcal{H}}(G_w^{\mathcal{H}}(z)) G_w^{\mathcal{H}}(z)$$

of Lemma 4.1, we find that

$$G_{a_{rts},w}^{\mathcal{H}}(z) = h_r^* h_r \cdot G_w^{\mathcal{H}}(z) \tau_r [R_{b_{rts},v}^{\mathcal{D}}(G_v^{\mathcal{D}}(d))],$$

from which we obtain

$$\tau_0[G^{\mathcal{H}}_{a_{rts},w}(z)] = \tau_0[h_r^* h_r G^{\mathcal{H}}_w(z)]\tau_r[R^{\mathcal{D}}_{b_{rts},v}(G^{\mathcal{D}}_v(d))].$$

Noting that $\tau_0[h_r^* h_r G^{\mathcal{H}}_w(z)] = \frac{\tau(p_r)}{\tau(p_0)}\alpha_r$, we obtain

$$\tau_0[G^{\mathcal{H}}_{a_{rts},w}(z)] = \frac{\tau(p_r)}{\tau(p_0)}\tau_r[R^{\mathcal{D}}_{b_{rts},v}(G^{\mathcal{D}}_v(d))\alpha_r]$$

$$= \frac{\tau(p_r)}{\tau(p_0)}\tau_r[R^{\mathcal{D}}_{b_{rts},v}(G^{\mathcal{D}}_v(d))G^{\mathcal{D}}_v(d)]$$

$$= \frac{\tau(p_r)}{\tau(p_0)}\tau_r[G^{\mathcal{D}}_{b_{rts},v}(d)] = \frac{\tau(p_r)}{\tau(p_0)}\tau_r[G^{\mathcal{G}}_{b_{rts},v}(d)],$$

where in the second equality we replace $\alpha_r$ by $G^{\mathcal{D}}_v(d) = \sum_{i=0}^{2k}\alpha_i p_i$. Substituting Proposition 5.7 into the identity

$$G^{\mathcal{G}}_{b_{rts},v}(d) = R^{\mathcal{G}}_{b_{rts},v}(G^{\mathcal{G}}_v(d))G^{\mathcal{G}}_v(d),$$

we find that

$$G^{\mathcal{G}}_{b_{rts},v}(d) = g_r^* g_r G^{\mathcal{G}}_v(d)\tau_{r+k}[R^{\mathcal{D}}_{c_{rts},u}(G^{\mathcal{D}}_u(e))].$$

Noting that $\tau_r(g_r^* g_r G^{\mathcal{G}}_v(d)) = \frac{\tau(p_{r+k})}{\tau(p_r)}\gamma_{r+k}$, we find similarly that

$$\tau_r[G^{\mathcal{G}}_{b_{rts},v}(d)] = \frac{\tau(p_{r+k})}{\tau(p_r)}\tau_{r+k}[R^{\mathcal{D}}_{c_{rts},u}(G^{\mathcal{D}}_u(e))\gamma_{r+k}]$$

$$= \frac{\tau(p_{r+k})}{\tau(p_r)}\tau_{r+k}[R^{\mathcal{D}}_{c_{rts},u}(G^{\mathcal{D}}_u(e))G^{\mathcal{D}}_u(e)] = \frac{\tau(p_{r+k})}{\tau(p_r)}\tau_{r+k}[G^{\mathcal{D}}_{c_{rts},u}(e)].$$

Putting everything together, we conclude that

$$\tau(a_{rts}(z-w)^{-1}) = \tau(p_{r+k})\tau_{r+k}[G^{\mathcal{D}}_{c_{rts},u}(e)] = \tau(c_{rts}(e-u)^{-1}). \qquad \square$$

Applying the definitions of $a_{rts}$ and $c_{rts}$, the asymptotic freeness result in [25], Theorem 3.9, and Proposition 5.6(d), the above implies

$$\frac{1}{n_t}\mathrm{Tr}_t[FGHR(z)H^\top G^\top F] \sim \frac{1}{n_t}\mathrm{Tr}_t(F(\mathrm{diag}_n(\mathbf{a}^{-1}) + F)^{-1}F),$$

and Proposition 5.5 now follows from the Woodbury matrix identity. We defer these details to Appendix E.2.

5.4. *Outlier eigenvectors and eigenvectors.* Combining Lemmas 5.2 and 5.3, we have shown that $\widehat{T} \sim \widetilde{T}$. Recalling $\Gamma = \widetilde{\Gamma}Q^\top$ and using $\det(\mathrm{Id} + AB) = \det(\mathrm{Id} + BA)$, we see that the roots of $0 = \det\widetilde{T}(z)$ are the same as those of $0 = \det T(z)$. Then Theorem 2.5 follows from an application of Hurwitz's theorem. We defer the technical details of this argument to Appendix E.3.

The proof of Theorem 2.6 uses the following two results, whose proofs are deferred to Appendix E.4.

PROPOSITION 5.9. *In the setting of Theorem 2.6, $\ker\widetilde{T}(\lambda)$ has dimension exactly 1, and each other singular value of $\widetilde{T}(\lambda)$ is at least a constant $c \equiv c(\delta) > 0$.*

PROPOSITION 5.10. *Denote by $S'(z)$ and $R'(z)$ the derivatives of $S(z)$ and $R(z)$ with respect to $z$. Then*

$$S'(z) \sim 0, \qquad Q^{\mathsf{T}} R'(z) Q \sim -Q^{\mathsf{T}} \partial_z \big[ (z \operatorname{Id} + \mathbf{b} \cdot \mathring{\Sigma})^{-1} \big] Q,$$

$$n_t^{-1} \operatorname{Tr}_t \big[ F G H R'(z) H^{\mathsf{T}} G^{\mathsf{T}} F \big] \sim b_t'(z).$$

PROOF OF THEOREM 2.6. Since $(\widehat{\lambda}, \widehat{v})$ is an eigenvalue-eigenvector pair, we have that $\widehat{\lambda}\widehat{v} = \widehat{\Sigma}\widehat{v} = W\widehat{v} + P\widehat{v}$, which implies that

(44) $$0 = (\operatorname{Id} + R(\widehat{\lambda})P)\widehat{v}.$$

Define

$$\widehat{v}_1 = \Xi^{\mathsf{T}} F G H \widehat{v} \quad \text{and} \quad \widehat{v}_2 = Q^{\mathsf{T}} \widehat{v}.$$

Multiplying (44) on the left by $\big[ \begin{smallmatrix} \Xi^{\mathsf{T}} FGH \\ Q^{\mathsf{T}} \end{smallmatrix} \big]$ and recalling (38), we obtain

(45) $$0 = \widehat{K}(\widehat{\lambda}) \begin{pmatrix} \widehat{v}_1 \\ \widehat{v}_2 \end{pmatrix}.$$

Eliminating $\widehat{v}_1$ in this system of equations, we get $0 = \widehat{T}(\widehat{\lambda})\widehat{v}_2$ for the Schur complement $\widehat{T}$ from (39). We show in Proposition E.6 that $\widetilde{T}(z)$ is bounded over $U_\delta$. Then so is $\widetilde{T}'(z)$, by the Cauchy integral formula. Applying a Taylor expansion and the results $\widehat{\lambda} - \lambda \to 0$ and $\widehat{T} \sim \widetilde{T}$ from Theorem 2.5 and Lemmas 5.2 and 5.3, almost surely $\|\widehat{T}(\widehat{\lambda}) - \widetilde{T}(\lambda)\| \to 0$. So also

$$\|\widehat{T}(\widehat{\lambda})^{\mathsf{T}}\widehat{T}(\widehat{\lambda}) - \widetilde{T}(\lambda)^{\mathsf{T}}\widetilde{T}(\lambda)\| \to 0.$$

Applying this to $\widehat{v}_2$, we find that $\|\widetilde{T}(\lambda)^{\mathsf{T}}\widetilde{T}(\lambda)\widehat{v}_2\| \to 0$, which implies by Proposition 5.9 and the Davis–Kahan theorem that

(46) $$\widehat{v}_2 - \|\widehat{v}_2\| v_2 \to 0,$$

where $v_2$ is a unit vector in $\ker \widetilde{T}(\lambda)$ with an appropriate choice of sign.

We now compute the limit of $\|\widehat{v}_2\|$. By (44) and the definition of $P$, we see that

(47) $$-\widehat{v} = R(\widehat{\lambda})\big(Q\widetilde{\Gamma}^{\mathsf{T}}\widehat{v}_1 + (H^{\mathsf{T}}G^{\mathsf{T}}F\Xi\widetilde{\Gamma} + Q\widetilde{\Gamma}^{\mathsf{T}}\Xi^{\mathsf{T}}F\Xi\widetilde{\Gamma})\widehat{v}_2\big).$$

On the other hand, in equation (45), we may solve for $\widehat{v}_1$ to obtain $\widehat{v}_1 = -\widehat{K}_{11}(\widehat{\lambda})^{-1}\widehat{K}_{12}(\widehat{\lambda})\widehat{v}_2$ when $\widehat{K}_{11}(\widehat{\lambda})$ is invertible. Substituting into (47),

(48) $$\widehat{v} = R(\widehat{\lambda})\big(M_1(\widehat{\lambda}) + M_2(\widehat{\lambda})\big)\widehat{v}_2$$

for the matrices

$$M_1(\widehat{\lambda}) = Q\widetilde{\Gamma}^{\mathsf{T}}\widehat{K}_{11}(\widehat{\lambda})^{-1}\widehat{K}_{12}(\widehat{\lambda}) - Q\widetilde{\Gamma}^{\mathsf{T}}\Xi^{\mathsf{T}}F\Xi\widetilde{\Gamma}, \qquad M_2(\widehat{\lambda}) = -H^{\mathsf{T}}G^{\mathsf{T}}F\Xi\widetilde{\Gamma}.$$

Note that $M_1'(z)$, $M_2'(z)$, $R'(z)$ are also bounded over $U_\delta$, on a high-probability event when $\operatorname{spec}(W) \subset \operatorname{supp}(\mu_0)_{\delta/2}$ and $\|\Xi\|, \|G\| < C$. Taking the squared norm of (48) on both sides and applying $\widehat{\lambda} - \lambda \to 0$ and a Taylor expansion,

(49) $$1 = \sum_{i,j=1}^{2} \widehat{v}_2^{\mathsf{T}} M_i(\widehat{\lambda})^{\mathsf{T}} R(\widehat{\lambda})^2 M_j(\widehat{\lambda})\widehat{v}_2 = \sum_{i,j=1}^{2} \widehat{v}_2^{\mathsf{T}} M_i(\lambda)^{\mathsf{T}} R(\lambda)^2 M_j(\lambda)\widehat{v}_2 + o(1).$$

Applying Lemma 5.2 and Propositions E.3 and 5.5, we find that

$$Q^{\mathsf{T}} M_1(z) \sim \widetilde{\Gamma}^{\mathsf{T}}\Xi^{\mathsf{T}} FGH R(z) H^{\mathsf{T}}G^{\mathsf{T}}F\Xi\widetilde{\Gamma} - \widetilde{\Gamma}^{\mathsf{T}}\Xi^{\mathsf{T}}F\Xi\widetilde{\Gamma} \sim \widetilde{\Gamma}^{\mathsf{T}} \operatorname{diag}_\ell(\mathbf{b}(z))\widetilde{\Gamma}.$$

Also, noting that $R(z)^2 = R'(z)$ and applying Proposition 5.10,

$$Q^{\mathsf{T}} R(z)^2 Q \sim Q^{\mathsf{T}} R'(z) Q \sim -Q^{\mathsf{T}} \partial_z \big[ (z \operatorname{Id} + \mathbf{b}(z) \cdot \mathring{\Sigma})^{-1} \big] Q.$$

Combining these, applying $\Gamma = \widetilde{\Gamma} Q^{\mathsf{T}}$, and setting $\widehat{u} = \widetilde{\Gamma} \widehat{v}_2 = \Gamma \widehat{v}$, we get

$$
\begin{aligned}
(50) \qquad & \widehat{v}_2^{\mathsf{T}} M_1(\lambda)^{\mathsf{T}} R(\lambda)^2 M_1(\lambda) \widehat{v}_2 \\
&= -\widehat{u}^{\mathsf{T}} \operatorname{diag}_\ell(\mathbf{b}) \Gamma \cdot \partial_\lambda \big[ (\lambda \operatorname{Id} + \mathbf{b} \cdot \mathring{\Sigma})^{-1} \big] \cdot \Gamma^{\mathsf{T}} \operatorname{diag}_\ell(\mathbf{b}) \widehat{u} + o(1),
\end{aligned}
$$

where we write as shorthand $\mathbf{b} \equiv \mathbf{b}(\lambda)$. Applying $R(z)^2 = R'(z)$ and Propositions E.3 and 5.10, we also get $\Xi^{\mathsf{T}} F G H R(z)^2 H^{\mathsf{T}} G^{\mathsf{T}} F \Xi \sim \operatorname{diag}_\ell(\mathbf{b}'(z))$, and hence

$$
(51) \qquad \widehat{v}_2^{\mathsf{T}} M_2(\lambda)^{\mathsf{T}} R(\lambda)^2 M_2(\lambda) \widehat{v}_2 = \widehat{u}^{\mathsf{T}} \operatorname{diag}_\ell(\mathbf{b}') \widehat{u} + o(1).
$$

Finally, applying $S'(z) \sim 0$ from Proposition 5.10, we get $Q^{\mathsf{T}} R(z)^2 H^{\mathsf{T}} G^{\mathsf{T}} F \Xi \sim 0$, and hence

$$
(52) \qquad \widehat{v}_2^{\mathsf{T}} M_1(\lambda)^{\mathsf{T}} R(\lambda)^2 M_2(\lambda) \widehat{v}_2 \to 0.
$$

Then substituting (50), (51) and (52) into (49),

$$
(53) \qquad 1 = \widehat{u}^{\mathsf{T}} \big( -\operatorname{diag}_\ell(\mathbf{b}) \Gamma \cdot \partial_\lambda [ (\lambda \operatorname{Id} + \mathbf{b} \cdot \mathring{\Sigma})^{-1} ] \cdot \Gamma^{\mathsf{T}} \operatorname{diag}_\ell(\mathbf{b}) + \operatorname{diag}_\ell(\mathbf{b}') \big) \widehat{u} + o(1).
$$

Multiplying (46) on the left by $\widetilde{\Gamma}$, we find that

$$
(54) \qquad \widehat{u} - \|\widehat{v}_2\| \widetilde{\Gamma} v_2 \to 0.
$$

Define $\tilde{u} = \widetilde{\Gamma} v_2$, and note that $\tilde{u}$ is a nonzero vector in $\ker T(\lambda)$ because $v_2$ is a unit vector in $\ker \widetilde{T}(\lambda)$. Then $u = \tilde{u}/\|\tilde{u}\|$ is a unit vector in $\ker T(\lambda)$, which is unique up to sign by Proposition 5.9. Substituting (54) into (53) and recalling the definition of $\alpha$ in Theorem 2.6, we find that

$$
1 = \|\widehat{v}_2\|^2 \|\tilde{u}\|^2 \cdot \alpha + o(1).
$$

Writing (54) as $\widehat{u} - \|\widehat{v}_2\| \|\tilde{u}\| u \to 0$ and substituting $\alpha^{-1/2}$ for $\|\widehat{v}_2\| \|\tilde{u}\|$ concludes the proof. □

## SUPPLEMENTARY MATERIAL

**Supplementary Appendices** (DOI: 10.1214/20-AOS2010SUPP; .pdf). The Appendices contain additional details for Sections 3.1 and 3.2, proofs of the free probability results in Section 4, a proof of Theorem 2.4 and details of the proofs of Theorems 2.5 and 2.6 omitted from Section 5.

## REFERENCES

[1] BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686 https://doi.org/10.1016/j.jmva.2011.10.009

[2] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

[3] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003

[4] BELINSCHI, S., BERCOVICI, H. and CAPITAINE, M. (2017). On the outlying eigenvalues of a polynomial in large independent random matrices. Preprint. Available at arXiv:1703.08102.

[5] BELINSCHI, S. T., BERCOVICI, H., CAPITAINE, M. and FÉVRIER, M. (2017). Outliers in the spectrum of large deformed unitarily invariant models. *Ann. Probab.* **45** 3571–3625. MR3729610 https://doi.org/10.1214/16-AOP1144

[6] BELINSCHI, S. T. and CAPITAINE, M. (2017). Spectral properties of polynomials in independent Wigner and deterministic matrices. *J. Funct. Anal.* **273** 3901–3963. MR3711884 https://doi.org/10.1016/j.jfa.2017.07.010

[7] BENAYCH-GEORGES, F. (2009). Rectangular random matrices, related convolution. *Probab. Theory Related Fields* **144** 471–515. MR2496440 https://doi.org/10.1007/s00440-008-0152-z

[8] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007

[9] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.* **111** 120–135. MR2944410 https://doi.org/10.1016/j.jmva.2012.04.019

[10] BHATTACHARJEE, M. and BOSE, A. (2016). Large sample behaviour of high dimensional autocovariance matrices. *Ann. Statist.* **44** 598–628. MR3476611 https://doi.org/10.1214/15-AOS1378

[11] BHATTACHARJEE, M. and BOSE, A. (2016). Polynomial generalizations of the sample variance-covariance matrix when $pn^{-1} \to 0$. *Random Matrices Theory Appl.* **5** 1650014, 41. MR3569842 https://doi.org/10.1142/S2010326316500143

[12] BHATTACHARJEE, M. and BOSE, A. (2017). Matrix polynomial generalizations of the sample variance-covariance matrix when $pn^{-1} \to y \in (0, \infty)$. *Indian J. Pure Appl. Math.* **48** 575–607. MR3741695 https://doi.org/10.1007/s13226-017-0247-2

[13] BHATTACHARJEE, M. and BOSE, A. (2019). Joint convergence of sample autocovariance matrices when $p/n \to 0$ with application. *Ann. Statist.* **47** 3470–3503. MR4025749 https://doi.org/10.1214/18-AOS1785

[14] BLOWS, M. W. and MCGUIGAN, K. (2015). The distribution of genetic variance across phenotypic space and the response to selection. *Mol. Ecol.* **24** 2056–2072.

[15] BOSE, A. and BHATTACHARJEE, M. (2019). *Large Covariance and Autocovariance Matrices. Monographs on Statistics and Applied Probability* **162**. CRC Press, Boca Raton, FL. MR3821539

[16] BURDA, Z., JURKIEWICZ, J. and WACŁAW, B. (2005). Spectral moments of correlated Wishart matrices. *Phys. Rev. E (3)* **71** 026111, 11. MR2139960 https://doi.org/10.1103/PhysRevE.71.026111

[17] CAPITAINE, M. and CASALIS, M. (2004). Asymptotic freeness by generalized moments for Gaussian and Wishart matrices. Application to beta random matrices. *Indiana Univ. Math. J.* **53** 397–431. MR2060040 https://doi.org/10.1512/iumj.2004.53.2325

[18] CAPITAINE, M. and DONATI-MARTIN, C. (2007). Strong asymptotic freeness for Wigner and Wishart matrices. *Indiana Univ. Math. J.* **56** 767–803. MR2317545 https://doi.org/10.1512/iumj.2007.56.2886

[19] COLLET, J. M., MCGUIGAN, K., ALLEN, S. L., CHENOWETH, S. F. and BLOWS, M. W. (2018). Mutational pleiotropy and the strength of stabilizing selection within and between functional modules of gene expression. *Genetics* **208** 1601–1616. https://doi.org/10.1534/genetics.118.300776

[20] COLLINS, B. (2003). Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability. *Int. Math. Res. Not.* **17** 953–982. MR1959915 https://doi.org/10.1155/S107379280320917X

[21] COLLINS, B. and ŚNIADY, P. (2006). Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Comm. Math. Phys.* **264** 773–795. MR2217291 https://doi.org/10.1007/s00220-006-1554-3

[22] DOBRIBAN, E. and LIU, S. (2018). A new theory for sketching in linear regression. Preprint. Available at arXiv:1810.06089.

[23] DYKEMA, K. (1993). On certain free product factors via an extended matrix model. *J. Funct. Anal.* **112** 31–60. MR1207936 https://doi.org/10.1006/jfan.1993.1025

[24] FAN, Z. and JOHNSTONE, I. M. (2019). Eigenvalue distributions of variance components estimators in high-dimensional random effects models. *Ann. Statist.* **47** 2855–2886. MR3988775 https://doi.org/10.1214/18-AOS1767

[25] FAN, Z., JOHNSTONE, I. M. and SUN, Y. (2018). Spiked covariances and principal components analysis in high-dimensional random effects models. Preprint. Available at arXiv:1806.09529.

[26] FAN, Z., SUN, Y. and WANG, Z. (2021). Supplement to "Principal components in linear mixed models with general bulk." https://doi.org/10.1214/20-AOS2010SUPP

[27] FISHER, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52** 399–433.

[28] HAAGERUP, U. and THORBJØRNSEN, S. (2005). A new application of random matrices: Ext($C^*_{\text{red}}(F_2)$) is not a group. *Ann. of Math. (2)* **162** 711–775. MR2183281 https://doi.org/10.4007/annals.2005.162.711

[29] HACHEM, W., LOUBATON, P. and NAJIM, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. MR2326235 https://doi.org/10.1214/105051606000000925

[30] HIAI, F. and PETZ, D. (2000). Asymptotic freeness almost everywhere for random matrices. *Acta Sci. Math.* (*Szeged*) **66** 809–834. MR1804226

[31] HINE, E. and BLOWS, M. W. (2006). Determining the effective dimensionality of the genetic variance–covariance matrix. *Genetics* **173** 1135–1144.

[32] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

[33] JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448 https://doi.org/10.1198/jasa.2009.0121

[34] JOHNSTONE, I. M. and PAUL, D. (2018). PCA in high dimensions: An orientation. *Proc. IEEE* **106** 1277–1292. https://doi.org/10.1109/JPROC.2018.2846730

[35] JOLLIFFE, I. T. (1986). *Principal Component Analysis*. *Springer Series in Statistics*. Springer, New York. MR0841268 https://doi.org/10.1007/978-1-4757-1904-8

[36] LEEK, J. T. and STOREY, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3** e161.

[37] LI, G. and ZHU, H. (2013). Genetic studies: The linear mixed models in genome-wide association studies. *Open Bioinform. J.* **7** 27–33. MR3172942 https://doi.org/10.2174/1875036201307010027

[38] LIU, H., AUE, A. and PAUL, D. (2015). On the Marčenko–Pastur law for linear time series. *Ann. Statist.* **43** 675–712. MR3319140 https://doi.org/10.1214/14-AOS1294

[39] LYNCH, M. and WALSH, B. (1998). *Genetics and Analysis of Quantitative Traits* **1**. Sinauer, Sunderland, MA.

[40] MALE, C. (2012). The norm of polynomials in large random and deterministic matrices. *Probab. Theory Related Fields* **154** 477–532. MR3000553 https://doi.org/10.1007/s00440-011-0375-2

[41] MARCENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sb. Math.* **1** 457–483.

[42] MINGO, J. A. and SPEICHER, R. (2017). *Free Probability and Random Matrices. Fields Institute Monographs* **35**. Springer, New York. MR3585560 https://doi.org/10.1007/978-1-4939-6942-5

[43] NADLER, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann. Statist.* **36** 2791–2817. MR2485013 https://doi.org/10.1214/08-AOS618

[44] NICA, A., SHLYAKHTENKO, D. and SPEICHER, R. (2002). Operator-valued distributions. I. Characterizations of freeness. *Int. Math. Res. Not.* **29** 1509–1538. MR1907203 https://doi.org/10.1155/S1073792802201038

[45] NICA, A. and SPEICHER, R. (2006). *Lectures on the Combinatorics of Free Probability. London Mathematical Society Lecture Note Series* **335**. Cambridge Univ. Press, Cambridge. MR2266879 https://doi.org/10.1017/CBO9780511735127

[46] PATTERSON, N., PRICE, A. L. and REICH, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* **2** e190. https://doi.org/10.1371/journal.pgen.0020190

[47] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

[48] PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718 https://doi.org/10.1016/j.jspi.2013.09.005

[49] RAO, C. R. (1972). Estimation of variance and covariance components in linear models. *J. Amer. Statist. Assoc.* **67** 112–115. MR0314185

[50] SCHULTZ, H. (2005). Non-commutative polynomials of independent Gaussian random matrices. The real and symplectic cases. *Probab. Theory Related Fields* **131** 261–309. MR2117954 https://doi.org/10.1007/s00440-004-0366-7

[51] SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (2006). *Variance Components. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2298115

[52] SPEICHER, R. and VARGAS, C. (2012). Free deterministic equivalents, rectangular random matrix models, and operator-valued free probability theory. *Random Matrices Theory Appl.* **1** 1150008, 26. MR2934714 https://doi.org/10.1142/S2010326311500080

[53] STEGLE, O., PARTS, L., PIIPARI, M., WINN, J. and DURBIN, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7** 500–507. https://doi.org/10.1038/nprot.2011.457

[54] TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083

[55] VOICULESCU, D. (1991). Limit laws for random matrices and free products. *Invent. Math.* **104** 201–220. MR1094052 https://doi.org/10.1007/BF01245072

[56] VOICULESCU, D. (1998). A strengthened asymptotic freeness result for random matrices with applications to free entropy. *Int. Math. Res. Not.* **1** 41–63. MR1601878 https://doi.org/10.1155/S107379289800004X

[57] WANG, L., AUE, A. and PAUL, D. (2017). Spectral analysis of sample autocovariance matrices of a class of linear time series in moderately high dimensions. *Bernoulli* **23** 2181–2209. MR3648029 https://doi.org/10.3150/16-BEJ807

[58] WRIGHT, S. (1935). The analysis of variance and the correlations between relatives with respect to deviations from an optimum. *J. Genet.* **30** 243–256.

[59] YANG, J., LEE, S. H., GODDARD, M. E. and VISSCHER, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88** 76–82.

[60] ZHANG, L. (2006). Spectral analysis of large dimensional random matrices Ph.D. thesis, National Univ. Singapore.