The Annals of Statistics
2019, Vol. 47, No. 5, 2855–2886
https://doi.org/10.1214/18-AOS1767

EIGENVALUE DISTRIBUTIONS OF VARIANCE COMPONENTS ESTIMATORS IN HIGH-DIMENSIONAL RANDOM EFFECTS MODELS

BY ZHOU FAN[1] AND IAIN M. JOHNSTONE[2]

Yale University and Stanford University

We study the spectra of MANOVA estimators for variance component covariance matrices in multivariate random effects models. When the dimensionality of the observations is large and comparable to the number of realizations of each random effect, we show that the empirical spectra of such estimators are well approximated by deterministic laws. The Stieltjes transforms of these laws are characterized by systems of fixed-point equations, which are numerically solvable by a simple iterative procedure. Our proof uses operator-valued free probability theory, and we establish a general asymptotic freeness result for families of rectangular orthogonally invariant random matrices, which is of independent interest. Our work is motivated in part by the estimation of components of covariance between multiple phenotypic traits in quantitative genetics, and we specialize our results to common experimental designs that arise in this application.

1. Introduction. High-dimensional data exhibit phenomena unexpected from experience with a fixed number of variables. A well-studied example arises with n independent and identically distributed (i.i.d.) samples from a p-variate distribution with mean μ and covariance Σ. If p increases proportionately with n, then the eigenvalues of the sample covariance matrix are more dispersed than their population counterparts. Notably, this extra spreading, described by the celebrated Marcenko–Pastur equation (Marčenko and Pastur (1967), Silverstein (1995)), does not disappear in the limit of large p and n. For example, if $\Sigma = \mathrm{Id}$ and $p/n \to \gamma < 1$, then the limiting Marcenko–Pastur law is supported on $[(1-\sqrt{\gamma})^2, (1+\sqrt{\gamma})^2]$. This has many implications for statistical inference concerning Σ in high dimensions, which we discuss below.

The i.i.d. assumption, however, connotes a single level of variation in the data. In this paper, we begin study of high-dimensional data exhibiting several levels of variation, or *random effects*. In a simple example with two levels, the p-

Received November 2017; revised August 2018.

[1]Supported by a Hertz Foundation Fellowship and NDSEG Fellowship (DoD, AFOSR 32 CFR 168a).

[2]Supported in part by NIH Grant R01 EB001988 and NSF Grant DMS-1407813.

MSC2010 subject classifications. 62E20.

Key words and phrases. Random matrix theory, free probability, deterministic equivalents, covariance estimation.

dimensional observations may take the form

$$Y_{i,j} = \mu + \alpha_i + \varepsilon_{i,j}. \tag{1.1}$$

At the first level, there are $i = 1, \dots, I$ groups with i.i.d. random effects $\alpha_i \sim (0, \Sigma_1)$. The $j = 1, \dots, J_i$ observations within group i have independent second level effects $\varepsilon_{i,j} \sim (0, \Sigma_2)$, but as they share a common first level effect α_i, they are (perhaps strongly) correlated. For example, Yang et al. (2002) discusses multivariate examination response data for $n = \sum J_i \sim 50{,}000$ students in $I \sim 2500$ schools.

The goal of this paper is to describe analogs of the eigenvalue spreading phenomenon for the traditional (MANOVA) estimators of the covariance matrices Σ_1, Σ_2 and their multilevel extensions, Theorem 1.2. For $k = 2$ levels, the Marcenko–Pastur implicit equation is replaced by a system of $2k = 4$ equations. We show that this system can be solved numerically by a natural iterative scheme, Theorem 1.5. Our proof assumes that each random effect is Gaussian, although this assumption is likely inessential for the result, as discussed in Remark 1.6 below.

More generally, we study the multivariate mixed effects model

$$Y = X\beta + \sum_{r=1}^{k} U_r \alpha_r, \qquad \alpha_r \sim \mathcal{N}(0, \mathrm{Id}_{I_r} \otimes \Sigma_r), \tag{1.2}$$

the analogue of the univariate model studied in Rao (1971). Here, $Y \in \mathbb{R}^{n \times p}$ represents n observations of p traits, modeled as a sum of fixed effects $X\beta$ and k random effects $U_1\alpha_1, \dots, U_k\alpha_k$. (We may incorporate a residual error term ε by allowing $U_k = \mathrm{Id}$ and $\alpha_k = \varepsilon$.) The matrices $X \in \mathbb{R}^{n \times m}$ and $U_r \in \mathbb{R}^{n \times I_r}$ are known design and incidence matrices. Each $\alpha_r \in \mathbb{R}^{I_r \times p}$ is an unobserved random matrix with i.i.d. rows distributed as $\mathcal{N}(0, \Sigma_r)$, representing I_r independent realizations of the rth effect. The regression coefficients $\beta \in \mathbb{R}^{m \times p}$ and variance components $\Sigma_r \in \mathbb{R}^{p \times p}$ are unknown parameters.

We study estimators of Σ_r that are quadratic in Y and invariant to β, that is, estimators of the form

$$\hat{\Sigma}_r = Y^T B_r Y \qquad (B_r X = 0) \tag{1.3}$$

for symmetric matrices $B_r \in \mathbb{R}^{n \times n}$. In particular, model (1.2) encompasses nested and crossed classification designs, and (1.3) encompasses MANOVA estimators and MINQUEs. We discuss examples in Section 2 and Appendix A. Our main result shows that in a high-dimensional asymptotic regime, the spectra of these estimators are well approximated by deterministic laws, characterized by a certain generalization of the Marcenko–Pastur equation.

1.1. *Motivation from evolutionary genetics*. A primary motivation for our work comes from genetics, where it is common to decompose the population variance of phenotypic traits into its constituent components, for example, corresponding to additive effects of genetic alleles, residual nonadditive genetic effects and environmental effects (Lynch and Walsh (1998)). If natural or artificial selection acts on a trait, then genetics theory indicates that the response to selection is governed by this first additive genetic component of variance. More precisely, if an episode of selection changes the mean trait value by S, then the change in mean trait value $\Delta\mu$ inherited by the next generation is predicted by the "breeders' equation"

$$\Delta\mu = \sigma_A^2(\sigma^2)^{-1}S,$$

where σ_A^2 is the additive genetic component of the total variance σ^2 (Lush (1937)).

From a multivariate perspective, selection acting on one trait may induce an evolutionary response in genetically correlated traits (Blows (2007), Lande and Arnold (1983), Phillips and Arnold (1989)). Most of this correlation is likely due to pleiotropy, the influence of a single gene on multiple traits, and there is evidence that pleiotropic effects are widespread across the phenome (Barton (1990), McGuigan et al. (2014), Walsh and Blows (2009)). If selection changes the mean values of p traits by $S \in \mathbb{R}^p$, then the changes inherited by the next generation are predicted by

$$\Delta\mu = GP^{-1}S, \tag{1.4}$$

where $P \in \mathbb{R}^{p\times p}$ is the total phenotypic trait covariance and $G \in \mathbb{R}^{p\times p}$ is its additive genetic component (Lande (1979), Lande and Arnold (1983)).

Microarrays have enabled the measurements of thousands of quantitative phenotypes in a single study, providing an opportunity to better understand the extent of pleiotropy and the effective dimensionality of possible evolutionary response in the entire phenome of an organism (Blows et al. (2015), McGuigan et al. (2014)). In these high-dimensional settings, it becomes natural to interpret the breeders' equation (1.4) from a principal components perspective, where response to selection is understood via the principal eigenvectors of G and the alignment of the "selection gradient" $P^{-1}S$ with these eigenvectors (Blows and McGuigan (2015), Hine, McGuigan and Blows (2014), Kirkpatrick (2009), Walsh and Blows (2009)).

A central question is then how to perform inference on the spectral structure of G, or of more general components of covariance, in high dimensions from a limited sample of individuals. Linear mixed models (1.2) are commonly used to estimate G and other components of variance, ranging from classical studies where $U_1, \dots, U_k$ encode known kinship between samples (Fisher (1918), Wright (1935)) to modern genome-wide association studies where $U_1, \dots, U_k$ encode genotype information (Loh et al. (2015), Yang et al. (2011)). Recent work has explored in simulation the behavior of principal components analyses for such estimates (Blows and McGuigan (2015)). We initiate here a theoretical study of these questions, as a step toward developing new inferential procedures for this application.

1.2. *The Marcenko–Pastur equation and applications.* As an analogy, we review the Marcenko–Pastur equation describing sample eigenvalue dispersion in the setting of i.i.d. samples, along with a few of its implications for statistical inference in high dimensions. We refer the interested reader to Paul and Aue (2014) and the recent textbook (Yao, Zheng and Bai (2015)) for additional statistical applications.

Given $Y \in \mathbb{R}^{n \times p}$ consisting of n i.i.d. observations with distribution $\mathcal{N}(0, \Sigma)$, consider the sample covariance matrix $\hat{\Sigma} = n^{-1} Y^T Y$. Let $\mu_{\hat{\Sigma}} = p^{-1} \sum_{i=1}^{p} \delta_{\lambda_i(\hat{\Sigma})}$ denote the empirical spectral measure of $\hat{\Sigma}$.

THEOREM 1.1 (Marčenko and Pastur (1967), Silverstein (1995)). *Suppose $n, p \to \infty$ such that $c < p/n < C$ and $\|\Sigma\| < C$ for some constants $C, c > 0$. Then for each $z \in \mathbb{C}^+$, there exists a unique value $m_0(z) \in \{m \in \mathbb{C} : -(1 - p/n)z^{-1} + (p/n)m \in \mathbb{C}^+\}$ satisfying*

$$m_0(z) = \frac{1}{p} \operatorname{Tr}\left[\left(\left(1 - \frac{p}{n} - \frac{p}{n} z m_0(z)\right)\Sigma - z \operatorname{Id}_p\right)^{-1}\right], \tag{1.5}$$

and m_0 defines the Stieltjes transform of a (n, p, Σ-dependent) probability measure μ_0 on $\mathbb{R}$ such that $\mu_{\hat{\Sigma}} - \mu_0 \to 0$ weakly almost surely.

Theorem 1.1 is usually stated assuming convergence of p/n to $\gamma \in (0, \infty)$ and of the spectrum of Σ to a weak limit μ^*, in which case $\mu_{\hat{\Sigma}}$ converges to a limit μ_0 depending on γ and μ^*. The above statement is instead in a "deterministic equivalent" form Couillet, Debbah and Silverstein (2011), Hachem, Loubaton and Najim (2007), where μ_0 is defined by the finite-sample quantities p/n and Σ. We discuss this further in Remark 1.3.

The Marcenko–Pastur equation has many implications for statistical inference regarding Σ. One implication is in estimating the principal "signal" eigenvalues and eigenvectors of Σ. Sample eigenvalue dispersion leads to an upward bias in the sample locations of principal eigenvalues, and a quantitative description of this bias and of the error of the principal eigenvectors is closely connected to the Marcenko–Pastur equation (Bai and Yao (2012), Baik, Ben Arous and Péché (2005), Baik and Silverstein (2006), Benaych-Georges and Nadakuditi (2011), Paul (2007)). These results allow for consistent and debiased estimation of the principal eigenvalues and of low-dimensional projections of the eigenvectors, even as $n, p \to \infty$ proportionately.

A second application is in developing shrinkage estimates for the entire spectrum of Σ (Bai, Chen and Yao (2010), El Karoui (2008), Mestre (2008), Rao et al. (2008)) and for Σ itself under various matrix losses (Ledoit and Péché (2011), Ledoit and Wolf (2012)). Approaches for the former use various strategies to "invert" the mapping from Σ to μ_0 in the Marcenko–Pastur equation. For the latter, the Marcenko–Pastur equation plays a role in quantifying the risks of shrinkage estimates and in deriving the forms of optimal shrinkage procedures.

A third line of work pertains to testing sphericity or other spectral hypotheses regarding Σ (Dobriban (2017), Johnstone (2001), Onatski, Moreira and Hallin (2014)). Popular tests have been proposed based on the largest sample eigenvalue (Johnstone (2001), Soshnikov (2002)) or linear spectral statistics (Bai and Silverstein (2004)). The null distributions in such tests are related to the fluctuations of the empirical spectral measure around the Marcenko–Pastur law in local and global regimes.

Similar inferential questions are of interest pertaining to individual components of variance in genetics applications, but inferential procedures are less well developed in this setting. Developing such procedures is an interesting avenue for future work, and it will likely require an understanding of the bulk spectral law which is the focus of our current paper. Some results in this direction in the particular case of isotropic population variance component matrices are reported in Fan and Johnstone (2017), Fan, Johnstone and Sun (2018).

1.3. *Main result.* We consider asymptotics as $n, I_1, \ldots, I_k$ grow proportionately with p. For classification designs, this means that groups and subgroups of individuals remain bounded in size. This regime is relevant for experiments that estimate components of phenotypic covariance for reasons both of experimental practicality and of optimal design (Robertson (1959a, 1959b)).

Consider $\hat{\Sigma} = Y^T BY$ for symmetric $B \in \mathbb{R}^{n\times n}$ satisfying $BX = 0$. Define $I_+ = \sum_{r=1}^{k} I_r$,

$$U = \left(\sqrt{I_1}U_1 \mid \sqrt{I_2}U_2 \mid \cdots \mid \sqrt{I_k}U_k\right) \in \mathbb{R}^{n\times I_+}, \qquad F = U^T BU \in \mathbb{R}^{I_+\times I_+}.$$

For any $F \in \mathbb{C}^{I_+\times I_+}$, let $\operatorname{Tr}_r F$ denote the trace of its (r, r) block in the $k \times k$ block decomposition corresponding to $\mathbb{C}^{I_+} = \mathbb{C}^{I_1} \oplus \cdots \oplus \mathbb{C}^{I_k}$. For $a = (a_1, \ldots, a_k)$ and $b = (b_1, \ldots, b_k)$, define

$$D(a) = \operatorname{diag}(a_1 \operatorname{Id}_{I_1}, \ldots, a_k \operatorname{Id}_{I_k}) \in \mathbb{C}^{I_+\times I_+}, \qquad b \cdot \Sigma = b_1\Sigma_1 + \cdots + b_k\Sigma_k.$$

THEOREM 1.2. *Suppose $n, p, I_1, \ldots, I_k \to \infty$ such that $c < p/n < C$, $c < I_r/n < C$, $n\|B\| < C$, $\|\Sigma_r\| < C$, and $\|U_r\| < C$ for each $r = 1, \ldots, k$ and some constants $C, c > 0$. Then 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, for $r = 1, \ldots, k$, the equations*

$$a_r = -I_r^{-1} \operatorname{Tr}\left((z \operatorname{Id}_p + b \cdot \Sigma)^{-1}\Sigma_r\right), \tag{1.6}$$

$$b_r = -I_r^{-1} \operatorname{Tr}_r\left(\left[\operatorname{Id}_{I_+} + FD(a)\right]^{-1} F\right). \tag{1.7}$$

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

$$m_0(z) = -p^{-1} \operatorname{Tr}\big((z \operatorname{Id}_p + b \cdot \Sigma)^{-1}\big) \tag{1.8}$$

defines the Stieltjes transform of a probability measure μ_0 *on* $\mathbb{R}$ *such that* $\mu_{\hat{\Sigma}} - \mu_0 \to 0$ *weakly almost surely.*

REMARK 1.3. Here, μ_0 is a "deterministic equivalent" law defined directly by $\Sigma_1, \ldots, \Sigma_k$ and the model design for finite n and p. An asymptotic statement where μ_0 is a fixed limit would require not only that the spectral measures of $\Sigma_1, \ldots, \Sigma_k$ individually converge, but also that they convergence in a suitable joint sense, for example, convergence of $p^{-1} \operatorname{Tr} Q(\Sigma_1, \ldots, \Sigma_k)$ for each fixed polynomial Q. A similar requirement would be needed for convergence of polynomials in $(U_r^T B U_s : r, s = 1, \ldots, k)$, which depends on the sequence of model designs as $n, p \to \infty$. The deterministic equivalent form given above is simpler and arguably closer to applications in finite samples.

REMARK 1.4. When Y has n i.i.d. rows, the sample covariance $\hat{\Sigma} = n^{-1} Y^T Y$ corresponds to the special case of (1.2) with $k = 1$, $U_1 = \operatorname{Id}$, $\Sigma_1 = \Sigma$ and $B = n^{-1} \operatorname{Id}_n$. In this case, equations (1.6)–(1.8) reduce to

$$a_1 = -n^{-1} \operatorname{Tr}\big((z \operatorname{Id}_p + b_1 \Sigma)^{-1} \Sigma\big), \qquad b_1 = -(1 + a_1)^{-1}, \tag{1.9}$$

$$m_0(z) = -p^{-1} \operatorname{Tr}\big((z \operatorname{Id}_p + b_1 \Sigma)^{-1}\big), \tag{1.10}$$

which imply (by the identity $A^{-1} - (A + B)^{-1} = A^{-1} B (A + B)^{-1}$)

$$\begin{aligned} -1 - \frac{1}{b_1} = a_1 &= -\frac{z}{n b_1} \operatorname{Tr}\big((z \operatorname{Id}_p)^{-1} - (z \operatorname{Id}_p + b_1 \Sigma)^{-1}\big) \\ &= -\frac{p}{n b_1} + \frac{p z m_0(z)}{n b_1}. \end{aligned}$$

Hence $b_1 = -1 + (p/n) + (p/n) z m_0(z)$. Together with the above expression for $m_0(z)$, this recovers the Marcenko–Pastur equation (1.5).

In most cases, (1.6)–(1.8) do not admit a closed-form solution in $a_1, \ldots, a_k$, $b_1, \ldots, b_k$, and $m_0(z)$. However, these equations may be solved numerically.

THEOREM 1.5. *For each* $z \in \mathbb{C}^+$, *the values* a_r *and* b_r *in Theorem* 1.2 *are the limits, as* $t \to \infty$, *of the iterative procedure which arbitrarily initializes* $b_1^{(0)}, \ldots, b_k^{(0)} \in \overline{\mathbb{C}^+}$ *and iteratively computes (for* $t = 0, 1, 2, \ldots$) $a_r^{(t)}$ *from* $b_r^{(t)}$ *using* (1.6) *and* $b_r^{(t+1)}$ *from* $a_r^{(t)}$ *using* (1.7).

By the Stieltjes inversion formula, $\pi^{-1} \Im m_0(x + i\varepsilon)$ is the density of the convolution $\mu_0 \star \operatorname{Cauchy}(0, \varepsilon)$. This may be computed by the above procedure to numerically approximate μ_0; this is depicted in Figure 1, and a software implementation is available on the first author's website. We leave to future work the development of faster algorithms, such as in Dobriban (2015), for solving these fixed-point equations.

Theorems 1.2 and 1.5 are inspired by the study of similar models for wireless communication channels. In particular, Couillet, Debbah and Silverstein (2011) establishes analogous results for the matrix

$$S + \sum_{r=1}^{k} \Sigma_r^{1/2} G_r^* B_r G_r \Sigma_r^{1/2},$$

where $B_r \in \mathbb{C}^{n_r \times n_r}$ are positive semidefinite and diagonal. Earlier work of Zhang ((2006), Theorem 1.2.1) considers $k = 1$, $S = 0$, and arbitrary Hermitian B_1. For $S = 0$, this model is encompassed by our Theorem 4.1; however, we remark that these works do not require Gaussian G_r. In Dupuy and Loubaton (2011) and the earlier work of Moustakas and Simon (2007) using the replica method, the authors study the model

$$\sum_{r,s=1}^{k} \Sigma_r^{1/2} G_r^* T_r^{1/2} T_s^{1/2} G_s \Sigma_s^{1/2},$$

where Σ_r, T_r are positive semidefinite and G_r are complex Gaussian. This model is similar to ours, and we recover their result in Theorem 4.1 using a different proof. We note that Dupuy and Loubaton (2011) proves only mean convergence, whereas we also control the variance and prove convergence a.s. We use a free probability approach, which may be easier to generalize to other models.

1.4. *Overview of proof.* We use the tools of operator-valued free probability theory, in particular rectangular probability spaces and their connection to operator-valued freeness developed in Benaych-Georges (2009) and the free deterministic equivalents approach of Speicher and Vargas (2012).

Let us write α_r in (1.2) as $\alpha_r = G_r \Sigma_r^{1/2}$, where $G_r \in \mathbb{R}^{I_r \times p}$ has i.i.d. $\mathcal{N}(0, 1)$ entries. Then $\hat{\Sigma} = Y^T B Y$ takes the form

$$\hat{\Sigma} = \sum_{r,s=1}^{k} \Sigma_r^{1/2} G_r^T U_r^T B U_s G_s \Sigma_s^{1/2}.$$

We observe the following: If $O_0, O_1, \ldots, O_k \in \mathbb{R}^{p \times p}$ and $O_{k+r} \in \mathbb{R}^{I_r \times I_r}$ for each $r = 1, \ldots, k$ are real orthogonal matrices, then by rotational invariance of G_r, $\mu_{\hat{\Sigma}}$ remains invariant in law under the transformations

$$\Sigma_r^{1/2} \mapsto H_r := O_r^T \Sigma_r^{1/2} O_0, \qquad U_r^T B U_s \mapsto F_{rs} := O_{k+r}^T U_r^T B U_s O_{k+s}.$$

Hence we may equivalently consider the matrix

$$W = \sum_{r,s=1}^{k} H_r^T G_r^T F_{rs} G_s H_s \tag{1.11}$$

for $O_0, \dots, O_{2k}$ independent and Haar-distributed. The families $\{F_{rs}\}$, $\{G_r\}$, $\{H_r\}$ are independent of each other, with each family satisfying a certain joint orthogonal invariance in law (formalized in Section 3).

Following Benaych-Georges (2009), we embed the matrices $\{F_{rs}\}$, $\{G_r\}$, $\{H_r\}$ into a square matrix space $\mathbb{C}^{N\times N}$. We then consider deterministic elements $\{f_{rs}\}$, $\{g_r\}$, $\{h_r\}$ in a von Neumann algebra $\mathcal{A}$ with tracial state τ, such that these elements model the embedded matrices, and $\{f_{rs}\}$, $\{g_r\}$ and $\{h_r\}$ are free with amalgamation over a diagonal subalgebra of projections in $\mathcal{A}$. We follow the deterministic equivalents approach of Speicher and Vargas (2012) and allow $(\mathcal{A}, \tau)$ and $\{f_{rs}\}$, $\{g_r\}$, $\{h_r\}$ to also depend on n and p.

Our proof of Theorem 1.2 consists of two steps:

1. For independent, jointly orthogonally invariant families of random matrices, we formalize the notion of a free deterministic equivalent and prove an asymptotic freeness result establishing validity of this approximation.
2. For our specific model of interest, we show that the Stieltjes transform of $w := \sum_{r,s} h_r^* g_r^* f_{rs} g_s h_s$ in the free model satisfies equations (1.6)–(1.8).

We establish separately the existence and uniqueness of the fixed point to (1.6)–(1.7) using a contractive mapping argument. Then the Stieltjes transform of w in step 2 is uniquely determined by (1.6)–(1.8), which implies by step 1 that (1.6)–(1.8) asymptotically determine the Stieltjes transform of W.

An advantage of this approach is that the approximation is separated from the computation of the approximating measure μ_0. The approximation in step 1 is general—it may be applied to other matrix models such as the above, and it follows a line of work establishing asymptotic freeness of random matrices (Benaych-Georges (2009), Collins (2003), Collins and Śniady (2006), Dykema (1993), Hiai and Petz (2000), Speicher and Vargas (2012), Voiculescu (1991, 1998)). In the computation in step 2, the Stieltjes transform of w is exactly (rather than approximately) described by (1.6)–(1.8). The computation is thus entirely algebraic, using free cumulant tools of Nica, Shlyakhtenko and Speicher (2002), Speicher and Vargas (2012), and it does not require analytic approximation arguments or bounds.

REMARK 1.6. Our proof uses rotational invariance of $\{G_r\}$, which follows from our Gaussian assumption on $\{\alpha_r\}$. Rotational invariance is a natural setting that leads to asymptotic freeness (Collins (2003), Collins and Śniady (2006), Hiai and Petz (2000)), but freeness may arise in other contexts; see, for example, Dykema (1993) for an early example in non-Gaussian–Wigner models. We believe that with additional work, our main result may be extended to general distributions of entries of $\{G_r\}$ under mild moment assumptions, but we will not pursue this in the current paper.

1.5. *Outline of paper.* Section 2 specializes Theorem 1.2 to the one-way design; other specializations are discussed in Appendix A. Section 3 reviews free probability theory and states the asymptotic freeness result. Section 4 performs the computation in the free model. The remainder of the proof and other details are deferred to the supplementary Appendices (Fan and Johnstone (2019)).

Notation. $\|\cdot\|$ denotes the l_2 norm for vectors and the $l_2 \to l_2$ operator norm for matrices. M^T, M^* and $\operatorname{Tr} M = \sum_i M_{ii}$ denote the transpose, conjugate-transpose and trace of M. Id_n denotes the identity matrix of size n. $\operatorname{diag}(A_1, \dots, A_k)$ denotes the block-diagonal matrix with blocks $A_1, \dots, A_k$. $\mathbb{C}^+ = \{z \in \mathbb{C} : \Im z > 0\}$ and $\overline{\mathbb{C}^+} = \{z \in \mathbb{C} : \Im z \geq 0\}$ denote the open and closed half-planes.

For a $*$-algebra $\mathcal{A}$ and elements $(a_i)_{i \in \mathcal{I}}$ of $\mathcal{A}$, $\langle a_i : i \in \mathcal{I} \rangle$ denotes the sub-$*$-algebra generated by $(a_i)_{i \in \mathcal{I}}$. We write $\langle \{a_i\} \rangle$ if the index set $\mathcal{I}$ is clear from context. If $\mathcal{A}$ is a von Neumann algebra, $\langle \{a_i\} \rangle_{W^*}$ denotes the generated von Neumann subalgebra, that is, the ultraweak closure of $\langle \{a_i\} \rangle$, and $\|a_i\|$ denotes the C^*-norm.

2. Specialization to one-way classification. The form (1.3) encompasses MANOVA estimators, which solve for $\Sigma_1, \dots, \Sigma_k$ in the system of equations $Y^T M_r Y = \mathbb{E}[Y^T M_r Y]$ for a certain choice of symmetric matrices $M_1, \dots, M_k \in \mathbb{R}^{n \times n}$ (Searle, Casella and McCulloch (2006), Chapter 5.2). From (1.2), the identity $\mathbb{E}[\alpha_s^T M \alpha_s] = (\operatorname{Tr} M)\Sigma_s$ for any matrix M, and independence of α_r, we get

$$\mathbb{E}[Y^T M_r Y] = \sum_{s=1}^{k} \mathbb{E}[\alpha_s^T U_s^T M_r U_s \alpha_s] = \sum_{s=1}^{k} \operatorname{Tr}(U_s^T M_r U_s)\Sigma_s.$$

Hence each MANOVA estimate $\hat{\Sigma}_r$ takes the form (1.3), where B_r is a linear combination of $M_1, \dots, M_k$.

In classification designs, standard choices for $M_1, \dots, M_k$ project onto subspaces of $\mathbb{R}^n$ such that each $Y^T M_r Y$ corresponds to a "sum-of-squares." We may simplify (1.7) in such settings by analytically computing the matrix inverse and block trace. We discuss here the one-way (balanced) design as an example. Appendix A provides details in the context of a more general discussion, first of the unbalanced one-way design, and second of balanced crossed and nested designs. As specific examples of the second class, formulas are given for nested models, Section A.2.1 and for the replicated crossed two-way layout, Section A.2.2.

For more general designs and models, $M_1, \dots, M_k$ may be ad hoc, although Theorem 1.2 still applies to such estimators. The theorem also applies to MINQUEs (LaMotte (1973), Rao (1972)) in these settings, which prescribe a specific form for $B \in \mathbb{R}^{n \times n}$ based on a variance minimization criterion.

In the one-way design, $\{Y_{i,j} \in \mathbb{R}^p : 1 \le i \le I, 1 \le j \le J_i\}$ represent observations of p traits across $n = \sum_{i=1}^I J_i$ samples, belonging to I groups of sizes $J_1, \dots, J_I$. The balanced case corresponds to $J_1 = \cdots = J_I = J$. The data are modeled as (1.1) where $\mu \in \mathbb{R}^p$ is a vector of population mean values, $\alpha_i \sim \mathcal{N}(0, \Sigma_1)$ are i.i.d. random group effects, and $\varepsilon_{i,j} \sim \mathcal{N}(0, \Sigma_2)$ are i.i.d. residual errors. In quantitative genetics, this is the model for the half-sib experimental design and also for the standard twin study, where groups correspond to half-siblings or twin pairs (Lynch and Walsh (1998)).

Define the sums-of-squares

$$\text{(2.1)}\quad \mathrm{SS}_1 = J\sum_{i=1}^{I}(\bar{Y}_i - \bar{Y})(\bar{Y}_i - \bar{Y})^T, \qquad \mathrm{SS}_2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{i,j} - \bar{Y}_i)(Y_{i,j} - \bar{Y}_i)^T,$$

where $\bar{Y}_i \in \mathbb{R}^p$ and $\bar{Y} \in \mathbb{R}^p$ denote the mean in the ith group and of all samples, respectively. The standard MANOVA estimators are given (Searle, Casella and McCulloch (2006), Chapter 3.6) by

$$\text{(2.2)}\qquad \hat{\Sigma}_1 = \frac{1}{J}\left(\frac{1}{I-1}\mathrm{SS}_1 - \frac{1}{n-I}\mathrm{SS}_2\right), \qquad \hat{\Sigma}_2 = \frac{1}{n-I}\mathrm{SS}_2.$$

Theorem 1.2 yields the following corollary.

COROLLARY 2.1. *Assume $p, n, I \to \infty$ such that $c < p/n < C$, $c < J < C$, $\|\Sigma_1\| < C$ and $\|\Sigma_2\| < C$ for some $C, c > 0$. Denote $I_1 = I$ and $I_2 = n$. Then*:

(a) *For $\hat{\Sigma}_1$, $\mu_{\hat{\Sigma}_1} - \mu_0 \to 0$ weakly a.s. where μ_0 has Stieltjes transform $m_0(z)$ determined by*

$$a_s = -I_s^{-1}\operatorname{Tr}\big((z\,\mathrm{Id} + b_1\Sigma_1 + b_2\Sigma_2)^{-1}\Sigma_s\big) \qquad \text{for } s = 1, 2,$$

$$b_1 = -(1 + a_1 + a_2)^{-1}, \qquad b_2 = J^{-1}(J-1)(J-1-a_2)^{-1} + J^{-1}b_1,$$

$$m_0(z) = -p^{-1}\operatorname{Tr}\big((z\,\mathrm{Id} + b_1\Sigma_1 + b_2\Sigma_2)^{-1}\big).$$

(b) *For $\hat{\Sigma}_2$, $\mu_{\hat{\Sigma}_2} - \mu_0 \to 0$ weakly a.s. where μ_0 has Stieltjes transform $m_0(z)$ determined by*

$$a_2 = -n^{-1}\operatorname{Tr}\big((z\,\mathrm{Id} + b_2\Sigma_2)^{-1}\Sigma_2\big), \qquad b_2 = -(J-1)(J-1+Ja_2)^{-1},$$

$$m_0(z) = -p^{-1}\operatorname{Tr}\big((z\,\mathrm{Id} + b_2\Sigma_2)^{-1}\big).$$

For each $z \in \mathbb{C}^+$, these equations have a unique solution with $a_s \in \mathbb{C}^+ \cup \{0\}$, $b_s \in \overline{\mathbb{C}^+}$ and $m_0(z) \in \mathbb{C}^+$, which may be computed as in Theorem 1.5. Figure 1 displays the simulated spectrum of $\hat{\Sigma}_1$ and the result of this computation (for the density of $\mu_0 \star \mathrm{Cauchy}(0, 10^{-4})$) in various settings.

For $\hat{\Sigma}_2$ (but not $\hat{\Sigma}_1$), as in Remark 1.4, the three equations of Corollary 2.1(b) may be simplified to the single Marcenko–Pastur equation for population covariance Σ_2. This also follows directly from the observation that $\hat{\Sigma}_2$ is equal in law

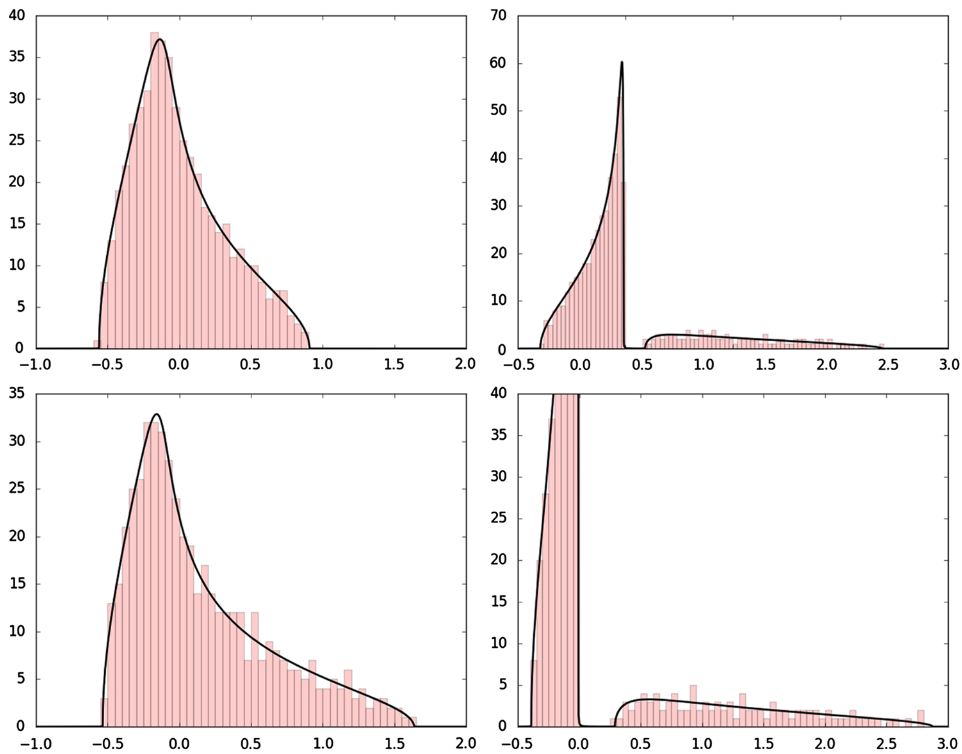

FIG. 1. *Simulated spectrum of* $\hat{\Sigma}_1$ *for the balanced one-way classification model*, $p = 500$, *with theoretical predictions of Corollary* 2.1 *overlaid in black*. *Left*: 400 *groups of size* 4. *Right*: 100 *groups of size* 8. *Top*: $\Sigma_1 = 0$, $\Sigma_2 = \mathrm{Id}$. *Bottom*: Σ_1 *with equally spaced eigenvalues in* $[0, 0.3]$, $\Sigma_2 = \mathrm{Id}$.

to $\varepsilon^T \pi \varepsilon$ where $\varepsilon \in \mathbb{R}^{n \times p}$ is the matrix of residual errors and π is a normalized projection onto a space of dimensionality $n - I$. This phenomenon holds generally for the MANOVA estimate of the residual error covariance in usual classification designs.

3. Operator-valued free probability.

3.1. *Background*. We review definitions from operator-valued free probability theory and its application to rectangular random matrices, drawn from Benaych-Georges (2009), Voiculescu (1995), Voiculescu, Dykema and Nica (1992).

DEFINITION. A *noncommutative probability space* $(\mathcal{A}, \tau)$ is a unital $*$-algebra $\mathcal{A}$ over $\mathbb{C}$ and a $*$-linear functional $\tau : \mathcal{A} \to \mathbb{C}$ called the *trace* that satisfies, for all $a, b \in \mathcal{A}$ and for $1_{\mathcal{A}} \in \mathcal{A}$ the multiplicative unit

$$\tau(1_{\mathcal{A}}) = 1, \qquad \tau(ab) = \tau(ba).$$

In this paper, $\mathcal{A}$ will always be a von Neumann algebra having norm $\|\cdot\|$, and τ a positive, faithful and normal trace. (These definitions are reviewed in Appendix D.) In particular, τ will be norm-continuous with $|\tau(a)| \leq \|a\|$.

Following Benaych-Georges (2009), we embed rectangular matrices into a larger square space according to the following structure.

DEFINITION. Let $(\mathcal{A}, \tau)$ be a noncommutative probability space and $d \geq 1$ a positive integer. For $p_1, \ldots, p_d \in \mathcal{A}$, $(\mathcal{A}, \tau, p_1, \ldots, p_d)$ is a *rectangular probability space* if $p_1, \ldots, p_d$ are nonzero pairwise-orthogonal projections summing to 1, that is, for all $r \neq s \in \{1, \ldots, d\}$,

$$p_r \neq 0, \qquad p_r = p_r^* = p_r^2, \qquad p_r p_s = 0, \qquad p_1 + \cdots + p_d = 1.$$

An element $a \in \mathcal{A}$ is *simple*, or (r, s)-simple, if $p_r a p_s = a$ for some $r, s \in \{1, \ldots, d\}$ (possibly $r = s$).

EXAMPLE 3.1. Let $N_1, \ldots, N_d \geq 1$ be positive integers and denote $N = N_1 + \cdots + N_d$. Consider the $*$-algebra $\mathcal{A} = \mathbb{C}^{N \times N}$, with the involution $*$ given by the conjugate transpose map $A \mapsto A^*$. For $A \in \mathbb{C}^{N \times N}$, let $\tau(A) = N^{-1} \operatorname{Tr} A$. Then $(\mathcal{A}, \tau) = (\mathbb{C}^{N \times N}, N^{-1} \operatorname{Tr})$ is a noncommutative probability space. Any $A \in \mathbb{C}^{N \times N}$ may be written in block form as

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1d} \\ A_{21} & A_{22} & \cdots & A_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ A_{d1} & A_{d2} & \cdots & A_{dd} \end{pmatrix},$$

where $A_{st} \in \mathbb{C}^{N_s \times N_t}$. For each $r = 1, \ldots, d$, denote by P_r the matrix with (r, r) block equal to Id_{N_r} and (s, t) block equal to 0 for all other s, t. Then P_r is a projection, and $(\mathbb{C}^{N \times N}, N^{-1} \operatorname{Tr}, P_1, \ldots, P_d)$ is a rectangular probability space. $A \in \mathbb{C}^{N \times N}$ is simple if $A_{st} \neq 0$ for at most one block (s, t).

In a rectangular probability space, the projections $p_1, \ldots, p_d$ generate a sub-$*$-algebra

$$\mathcal{D} := \langle p_1, \ldots, p_d \rangle = \left\{ \sum_{r=1}^{d} z_r p_r : z_r \in \mathbb{C} \right\}. \tag{3.1}$$

We may define a $*$-linear map $\mathbf{F}^{\mathcal{D}} : \mathcal{A} \to \mathcal{D}$ by

$$\mathbf{F}^{\mathcal{D}}(a) = \sum_{r=1}^{d} p_r \tau_r(a), \qquad \tau_r(a) = \tau(p_r a p_r)/\tau(p_r), \tag{3.2}$$

which is a projection onto $\mathcal{D}$ in the sense $\mathbf{F}^{\mathcal{D}}(d) = d$ for all $d \in \mathcal{D}$. In Example 3.1, $\mathcal{D}$ consists of matrices $A \in \mathbb{C}^{N \times N}$ for which A_{rr} is a multiple of the identity for each r and $A_{rs} = 0$ for each $r \neq s$. In this example, $\tau_r(A) = N_r^{-1} \operatorname{Tr}_r A$ where $\operatorname{Tr}_r A = \operatorname{Tr} A_{rr}$, so $\mathbf{F}^{\mathcal{D}}$ encodes the trace of each diagonal block.

The tuple $(\mathcal{A}, \mathcal{D}, \mathbf{F}^{\mathcal{D}})$ is an example of the following definition for an operator-valued probability space.

DEFINITION. A *$\mathcal{B}$-valued probability space* $(\mathcal{A}, \mathcal{B}, \mathbf{F}^{\mathcal{B}})$ is a $*$-algebra $\mathcal{A}$, a sub-$*$-algebra $\mathcal{B} \subseteq \mathcal{A}$ containing $1_{\mathcal{A}}$ and a $*$-linear map $\mathbf{F}^{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$ called the *conditional expectation* satisfying, for all $b, b' \in \mathcal{B}$ and $a \in \mathcal{A}$,

$$\mathbf{F}^{\mathcal{B}}(bab') = b\mathbf{F}^{\mathcal{B}}(a)b', \qquad \mathbf{F}^{\mathcal{B}}(b) = b.$$

We identify $\mathbb{C} \subset \mathcal{A}$ as a subalgebra via the inclusion map $z \mapsto z1_{\mathcal{A}}$, and we write 1 for $1_{\mathcal{A}}$ and z for $z1_{\mathcal{A}}$. Then a noncommutative probability space $(\mathcal{A}, \tau)$ is also a $\mathbb{C}$-valued probability space with $\mathcal{B} = \mathbb{C}$ and $\mathbf{F}^{\mathcal{B}} = \tau$.

DEFINITION. Let $(\mathcal{A}, \tau)$ be a noncommutative probability space and $\mathbf{F}^{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$ a conditional expectation onto a subalgebra $\mathcal{B} \subset \mathcal{A}$. $\mathbf{F}^{\mathcal{B}}$ is *τ-invariant* if $\tau \circ \mathbf{F}^{B} = \tau$.

It is verified that $\mathbf{F}^{\mathcal{D}} : \mathcal{A} \to \mathcal{D}$ defined by (3.2) is τ-invariant. When $\mathcal{B}$ is a von Neumann subalgebra of (a von Neumann algebra) $\mathcal{A}$, there exists a unique τ-invariant conditional expectation $\mathbf{F}^{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$, which is norm-continuous and satisfies $\|\mathbf{F}^{\mathcal{B}}(a)\| \leq \|a\|$. If $\mathcal{D} \subseteq \mathcal{B} \subseteq \mathcal{A}$ are nested von Neumann subalgebras with τ-invariant conditional expectations $\mathbf{F}^{\mathcal{D}} : \mathcal{A} \to \mathcal{D}$, $\mathbf{F}^{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$, then we have the analogue of the classical tower property,

$$\mathbf{F}^{\mathcal{D}} = \mathbf{F}^{\mathcal{D}} \circ \mathbf{F}^{\mathcal{B}}. \tag{3.3}$$

We note that $\mathcal{D}$ in (3.1) is a von Neumann subalgebra of $\mathcal{A}$, as it is finite-dimensional.

In the space $(\mathcal{A}, \tau)$, $a \in \mathcal{A}$ may be thought of as an analogue of a bounded random variable, $\tau(a)$ its expectation, and $\mathbf{F}^{\mathcal{B}}(a)$ its conditional expectation with respect to a sub-sigma-field. The following definitions then provide an analogue of the conditional distribution of a, and more generally of the conditional joint distribution of a collection $(a_i)_{i \in \mathcal{I}}$.

DEFINITION. Let $\mathcal{B}$ be a $*$-algebra and $\mathcal{I}$ be any set. A *$*$-monomial* in the variables $\{x_i : i \in \mathcal{I}\}$ with coefficients in $\mathcal{B}$ is an expression of the form $b_1 y_1 b_2 y_2 \ldots b_{l-1} y_{l-1} b_l$ where $l \geq 1$, $b_1, \ldots, b_l \in \mathcal{B}$, and $y_1, \ldots, y_{l-1} \in \{x_i, x_i^* : i \in \mathcal{I}\}$. A *$*$-polynomial* in $\{x_i : i \in \mathcal{I}\}$ with coefficients in $\mathcal{B}$ is any finite sum of such monomials.

We write $Q(a_i : i \in \mathcal{I})$ as the evaluation of a $*$-polynomial Q at $x_i = a_i$.

DEFINITION 3.2. Let $(\mathcal{A}, \mathcal{B}, \mathbf{F}^{\mathcal{B}})$ be a $\mathcal{B}$-valued probability space, let $(a_i)_{i\in\mathcal{I}}$ be elements of $\mathcal{A}$ and let $\mathcal{Q}$ denote the set of all $*$-polynomials in variables $\{x_i : i \in \mathcal{I}\}$ with coefficients in $\mathcal{B}$. The (joint) $\mathcal{B}$-*law* of $(a_i)_{i\in\mathcal{I}}$ is the collection of values in $\mathcal{B}$

$$\{\mathbf{F}^{\mathcal{B}}[Q(a_i : i \in I)]\}_{Q\in\mathcal{Q}}. \tag{3.4}$$

In the scalar setting where $\mathcal{B} = \mathbb{C}$ and $\mathbf{F}^{\mathcal{B}} = \tau$, a $*$-monomial takes the simpler form $zy_1y_2 \dots y_{l-1}$ for $z \in \mathbb{C}$ and $y_1, \dots, y_{l-1} \in \{x_i, x_i^* : i \in \mathcal{I}\}$ (because $\mathbb{C}$ commutes with $\mathcal{A}$). Then the collection of values (3.4) is determined by the scalar-valued moments $\tau(w)$ for all words w in the letters $\{x_i, x_i^* : i \in \mathcal{I}\}$. This is the analogue of the unconditional joint distribution of a family of bounded random variables, as specified by the joint moments.

Finally, the following definition of operator-valued freeness, introduced in Voiculescu (1995), has similarities to the notion of conditional independence of sub-sigma-fields in the classical setting.

DEFINITION. Let $(\mathcal{A}, \mathcal{B}, \mathbf{F}^{\mathcal{B}})$ be a $\mathcal{B}$-valued probability space and $(\mathcal{A}_i)_{i\in\mathcal{I}}$ a collection of sub-$*$-algebras of $\mathcal{A}$ which contain $\mathcal{B}$. $(\mathcal{A}_i)_{i\in\mathcal{I}}$ are $\mathcal{B}$-*free*, or free with amalgamation over $\mathcal{B}$, if for all $m \geq 1$, for all $i_1, \dots, i_m \in \mathcal{I}$ with $i_1 \neq i_2, i_2 \neq i_3, \dots, i_{m-1} \neq i_m$ and for all $a_1 \in \mathcal{A}_{i_1}, \dots, a_m \in \mathcal{A}_{i_m}$, the following implication holds:

$$\mathbf{F}^{\mathcal{B}}(a_1) = \mathbf{F}^{\mathcal{B}}(a_2) = \dots = \mathbf{F}^{\mathcal{B}}(a_m) = 0 \quad \Rightarrow \quad \mathbf{F}^{\mathcal{B}}(a_1a_2 \dots a_m) = 0.$$

Subsets $(S_i)_{i\in\mathcal{I}}$ of $\mathcal{A}$ are $\mathcal{B}$-free if the sub-$*$-algebras $(\langle S_i, \mathcal{B}\rangle)_{i\in\mathcal{I}}$ are.

In the classical setting, the joint law of (conditionally) independent random variables is determined by their marginal (conditional) laws. A similar statement holds for freeness.

PROPOSITION 3.3. *Suppose $(\mathcal{A}, \mathcal{B}, \mathbf{F}^{\mathcal{B}})$ is a $\mathcal{B}$-valued probability space, and subsets $(S_i)_{i\in\mathcal{I}}$ of $\mathcal{A}$ are $\mathcal{B}$-free. Then the $\mathcal{B}$-law of $\bigcup_{i\in\mathcal{I}} S_i$ is determined by the individual $\mathcal{B}$-laws of the S_i's.*

PROOF. See Voiculescu (1995), Proposition 1.3. □

3.2. *Free deterministic equivalents and asymptotic freeness.* Free deterministic equivalents were introduced in Speicher and Vargas (2012). Here, we formalize a bit this definition for independent jointly orthogonally invariant families of matrices, and we establish closeness of the random matrices and the free approximation in a general setting.

DEFINITION 3.4. For fixed $d \geq 1$, consider two sequences of N-dependent rectangular probability spaces $(\mathcal{A}, \tau, p_1, \ldots, p_d)$ and $(\mathcal{A}', \tau', p'_1, \ldots, p'_d)$ such that for each $r \in \{1, \ldots, d\}$, as $N \to \infty$,

$$|\tau(p_r) - \tau'(p'_r)| \to 0.$$

For a common index set $\mathcal{I}$, consider elements $(a_i)_{i \in \mathcal{I}}$ of $\mathcal{A}$ and $(a'_i)_{i \in \mathcal{I}}$ of $\mathcal{A}'$. Then $(a_i)_{i \in \mathcal{I}}$ and $(a'_i)_{i \in \mathcal{I}}$ are *asymptotically equal in $\mathcal{D}$-law* if the following holds: For any $r \in \{1, \ldots, d\}$ and any $*$-polynomial Q in the variables $\{x_i : i \in \mathcal{I}\}$ with coefficients in $\mathcal{D} = \langle p_1, \ldots, p_d \rangle$, denoting by Q' the corresponding $*$-polynomial with coefficients in $\mathcal{D}' = \langle p_1, \ldots, p_d \rangle$, as $N \to \infty$,

$$|\tau_r[Q(a_i : i \in \mathcal{I})] - \tau'_r[Q'(a'_i : i \in \mathcal{I})]| \to 0. \tag{3.5}$$

If $(a_i)_{i \in \mathcal{I}}$ and/or $(a'_i)_{i \in \mathcal{I}}$ are random elements of $\mathcal{A}$ and/or $\mathcal{A}'$, then they are *asymptotically equal in $\mathcal{D}$-law a.s.* if the above holds almost surely for each individual $*$-polynomial Q.

In the above, τ_r and τ'_r are defined by (3.2). "Corresponding" means that Q' is obtained by expressing each coefficient $d \in \mathcal{D}$ of Q in the form (3.1) and replacing $p_1, \ldots, p_d$ by $p'_1, \ldots, p'_d$.

We will apply Definition 3.4 by taking one of the two rectangular spaces to be $(\mathbb{C}^{N \times N}, N^{-1}\operatorname{Tr})$ as in Example 3.1, containing random elements, and the other to be an approximating deterministic model. (We will use "distribution" for random matrices to mean their distribution as random elements of $\mathbb{C}^{N \times N}$ in the usual sense, reserving the term "$\mathcal{B}$-law" for Definition 3.2.) Freeness relations in the deterministic model will emerge from the following notion of rotational invariance of the random matrices.

DEFINITION 3.5. Consider $(\mathbb{C}^{N \times N}, N^{-1}\operatorname{Tr}, P_1, \ldots, P_d)$ as in Example 3.1. A family of random matrices $(H_i)_{i \in \mathcal{I}}$ in $\mathbb{C}^{N \times N}$ is *block-orthogonally invariant* if, for any orthogonal matrices $O_r \in \mathbb{R}^{N_r \times N_r}$ for $r = 1, \ldots, d$, denoting $O = \operatorname{diag}(O_1, \ldots, O_d) \in \mathbb{R}^{N \times N}$, the joint distribution of $(H_i)_{i \in \mathcal{I}}$ is equal to that of $(O^T H_i O)_{i \in \mathcal{I}}$.

Let us provide several examples. We discuss the constructions of the spaces $(\mathcal{A}, \tau, p_1, \ldots, p_d)$ for these examples in Appendix D.

EXAMPLE 3.6. Fix $r \in \{1, \ldots, d\}$ and let $G \in \mathbb{C}^{N \times N}$ be a simple random matrix such that the diagonal block $G_{rr} \in \mathbb{C}^{N_r \times N_r}$ is distributed as the GUE or GOE, scaled to have entries of variance $1/N_r$. (Simple means $G_{st} = 0$ for all other blocks (s, t).) Let $(\mathcal{A}, \tau, p_1, \ldots, p_d)$ be a rectangular space with $\tau(p_s) = N_s/N$ for each $s = 1, \ldots, d$, such that $\mathcal{A}$ contains a self-adjoint simple element g satisfying $g = g^*$ and $p_r g p_r = g$, with moments given by the semicircle law:

$$\tau_r(g^l) = \int_{-2}^{2} \frac{x^l}{2\pi} \sqrt{4 - x^2}\, dx \qquad \text{for all } l \geq 0.$$

For any corresponding $*$-polynomials Q and q as in Definition 3.4, we may verify $N_r^{-1} \operatorname{Tr}_r Q(G) - \tau_r(q(g)) \to 0$ a.s. by the classical Wigner semicircle theorem (Wigner (1955)). Then G and g are asymptotically equal in $\mathcal{D}$-law a.s. Furthermore, G is block-orthogonally invariant.

EXAMPLE 3.7. Fix $r_1 \neq r_2 \in \{1, \dots, d\}$ and let $G \in \mathbb{C}^{N \times N}$ be a simple random matrix such that the block $G_{r_1 r_2}$ has i.i.d. Gaussian or complex Gaussian entries with variance $1/N_{r_1}$. Let $(\mathcal{A}, \tau, p_1, \dots, p_d)$ satisfy $\tau(p_s) = N_s/N$ for each s, such that $\mathcal{A}$ contains a simple element g satisfying $p_{r_1} g p_{r_2} = g$, where $g^* g$ has moments given by the Marcenko–Pastur law:

$$\tau_{r_2}\big((g^* g)^l\big) = \int x^l \nu_{N_{r_2}/N_{r_1}}(x)\, dx \qquad \text{for all } l \geq 0,$$

where ν_λ is the standard Marcenko–Pastur density

$$\nu_\lambda(x) = \frac{1}{2\pi} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{\lambda x} \mathbb{1}_{[\lambda_-, \lambda_+]}(x), \qquad \lambda_\pm = (1 \pm \sqrt{\lambda})^2. \tag{3.6}$$

By definition of τ_r and the cyclic property of τ, we also have

$$\tau_{r_1}\big((g g^*)^l\big) = (N_{r_2}/N_{r_1}) \tau_{r_2}\big((g^* g)^l\big).$$

For any corresponding $*$-polynomials Q and q as in Definition 3.4, we may verify $N_{r_2}^{-1} \operatorname{Tr}_{r_2} Q(G) - \tau_{r_2}(q(g)) \to 0$ and $N_{r_1}^{-1} \operatorname{Tr}_{r_1} Q(G) - \tau_{r_1}(q(g)) \to 0$ a.s. by the classical Marcenko–Pastur theorem (Marčenko and Pastur (1967)). Then G and g are asymptotically equal in $\mathcal{D}$-law a.s., and G is block-orthogonally invariant.

EXAMPLE 3.8. Let $B_1, \dots, B_k \in \mathbb{C}^{N \times N}$ be deterministic simple matrices, say with $P_{r_i} B_i P_{s_i} = B_i$ for each $i = 1, \dots, k$ and $r_i, s_i \in \{1, \dots, d\}$. Let $O_1 \in \mathbb{R}^{N_1 \times N_1}, \dots, O_d \in \mathbb{R}^{N_d \times N_d}$ be independent Haar-distributed orthogonal matrices, define $O = \operatorname{diag}(O_1, \dots, O_d) \in \mathbb{R}^{N \times N}$ and let $\check{B}_i = O^T B_i O$. Let $(\mathcal{A}, \tau, p_1, \dots, p_d)$ satisfy $\tau(p_s) = N_s/N$ for each s, such that $\mathcal{A}$ contains simple elements $b_1, \dots, b_k$ satisfying $p_{r_i} b_i p_{s_i} = b_i$ for each $i = 1, \dots, k$ and

$$N_r^{-1} \operatorname{Tr}_r Q(B_1, \dots, B_k) = \tau_r\big(q(b_1, \dots, b_k)\big) \tag{3.7}$$

for any corresponding $*$-polynomials Q and q with coefficients in $\langle P_1, \dots, P_d \rangle$ and $\langle p_1, \dots, p_d \rangle$. As $\operatorname{Tr}_r Q(B_1, \dots, B_k)$ is invariant under $B_i \mapsto O^T B_i O$, (3.7) holds also with $\check{B}_i$ in place of B_i. Then $(\check{B}_i)_{i \in \{1,\dots,k\}}$ and $(b_i)_{i \in \{1,\dots,k\}}$ are exactly (and hence also asymptotically) equal in $\mathcal{D}$-law, and $(\check{B}_i)_{i \in \{1,\dots,k\}}$ is block-orthogonally invariant by construction.

To study the interaction of several independent and block-orthogonally invariant matrix families, we will take a deterministic model for each family, as in Examples 3.6, 3.7 and 3.8 above, and consider a combined model in which these families are $\mathcal{D}$-free.

DEFINITION 3.9. Consider $(\mathbb{C}^{N\times N}, N^{-1}\operatorname{Tr}, P_1, \dots, P_d)$ as in Example 3.1. Suppose $(H_i)_{i\in\mathcal{I}_1}, \dots, (H_i)_{i\in\mathcal{I}_J}$ are finite families of random matrices in $\mathbb{C}^{N\times N}$ such that:

- These families are independent from each other, and
- For each $j = 1, \dots, J$, $(H_i)_{i\in\mathcal{I}_j}$ is block-orthogonally invariant.

Then a *free deterministic equivalent* for $(H_i)_{i\in\mathcal{I}_1}, \dots, (H_i)_{i\in\mathcal{I}_J}$ is any (N-dependent) rectangular probability space $(\mathcal{A}, \tau, p_1, \dots, p_d)$ and families $(h_i)_{i\in\mathcal{I}_1}, \dots, (h_i)_{i\in\mathcal{I}_J}$ of deterministic elements in $\mathcal{A}$ such that, as $N \to \infty$:

- For each $r = 1, \dots, d$, $|N^{-1}\operatorname{Tr} P_r - \tau(p_r)| \to 0$,
- For each $j = 1, \dots, J$, $(H_i)_{i\in\mathcal{I}_j}$ and $(h_i)_{i\in\mathcal{I}_j}$ are asymptotically equal in $\mathcal{D}$-law a.s., and
- $(h_i)_{i\in\mathcal{I}_1}, \dots, (h_i)_{i\in\mathcal{I}_J}$ are free with amalgamation over $\mathcal{D} = \langle p_1, \dots, p_d\rangle$.

The main result of this section is the following asymptotic freeness theorem, which establishes the validity of this approximation.

THEOREM 3.10. *In the space $(\mathbb{C}^{N\times N}, N^{-1}\operatorname{Tr}, P_1, \dots, P_d)$ of Example 3.1, suppose $(H_i)_{i\in\mathcal{I}_1}, \dots, (H_i)_{i\in\mathcal{I}_J}$ are independent, block-orthogonally invariant families of random matrices, and let $(h_i)_{i\in\mathcal{I}_1}, \dots, (h_i)_{i\in\mathcal{I}_J}$ be any free deterministic equivalent in $(\mathcal{A}, \tau, p_1, \dots, p_d)$. If there exist constants $C, c > 0$ (independent of N) such that $c < N_r/N$ for all r and $\|H_i\| < C$ a.s. for all $i \in \mathcal{I}_j$, all $\mathcal{I}_j$, and all large N, then $(H_i)_{i\in\mathcal{I}_j, j\in\{1,\dots,J\}}$ and $(h_i)_{i\in\mathcal{I}_j, j\in\{1,\dots,J\}}$ are asymptotically equal in $\mathcal{D}$-law a.s.*

More informally, if $(h_i)_{i\in\mathcal{I}_j}$ asymptotically models the family $(H_i)_{i\in\mathcal{I}_j}$ for each j, and these matrix families are independent and block-orthogonally invariant, then a system in which $(h_i)_{i\in\mathcal{I}_j}$ are $\mathcal{D}$-free asymptotically models the matrices jointly over j.

Theorem 3.10 is analogous to Benaych-Georges ((2009), Theorem 1.6) and Speicher and Vargas ((2012), Theorem 2.7), which establish similar results for complex unitary invariance. It permits multiple matrix families (where matrices within each family are not independent), uses the almost-sure trace $N^{-1}\operatorname{Tr}$ rather than $\mathbb{E} \circ N^{-1}\operatorname{Tr}$, and imposes boundedness rather than joint convergence assumptions. This last point fully embraces the deterministic equivalents approach.

We will apply Theorem 3.10 in the form of the following corollary. Suppose that $w \in \mathcal{A}$ satisfies $|\tau(w^l)| \le C^l$ for a constant $C > 0$ and all $l \ge 1$. We may define its Stieltjes transform by the convergent series

$$m_w(z) = \tau\big((w - z)^{-1}\big) = -\sum_{l\ge 0}^{\infty} z^{-(l+1)}\tau\big(w^l\big) \tag{3.8}$$

for $z \in \mathbb{C}^+$ with $|z| > C$, where we use the convention $w^0 = 1$ for all $w \in \mathcal{A}$.

COROLLARY 3.11. *Under the assumptions of Theorem 3.10, let Q be a self-adjoint $*$-polynomial* (*with $\mathbb{C}$-valued coefficients*) *in* $(x_i)_{i\in\mathcal{I}_j, j\in\{1,\dots,J\}}$, *and let*

$$W = Q\big(H_i : i \in \mathcal{I}_j, j \in \{1,\dots,J\}\big) \in \mathbb{C}^{N\times N},$$
$$w = Q\big(h_i : i \in \mathcal{I}_j, j \in \{1,\dots,J\}\big) \in \mathcal{A}.$$

Suppose $|\tau(w^l)| \le C^l$ for all $N, l \ge 1$ and some $C > 0$. Then for a sufficiently large constant $C_0 > 0$, letting $\mathbb{D} = \{z \in \mathbb{C}^+ : |z| > C_0\}$ and defining $m_W(z) = N^{-1}\operatorname{Tr}(W - z\operatorname{Id}_N)^{-1}$ and $m_w(z) = \tau((w-z)^{-1})$,

$$m_W(z) - m_w(z) \to 0$$

pointwise almost surely over $z \in \mathbb{D}$.

Proofs of Theorem 3.10 and Corollary 3.11 are contained in Appendix B.

3.3. *Computational tools.* Our computations in the free model will use the tools of free cumulants, $\mathcal{R}$-transforms, and Cauchy transforms discussed in Nica, Shlyakhtenko and Speicher (2002), Speicher (1998), Speicher and Vargas (2012). We review some relevant concepts here.

Let $(\mathcal{A}, \mathcal{B}, \mathbf{F}^{\mathcal{B}})$ be a $\mathcal{B}$-valued probability space and $\mathbf{F}^{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$ a conditional expectation. For $l \ge 1$, the lth order *free cumulant* of $\mathbf{F}^{\mathcal{B}}$ is a map $\kappa_l^{\mathcal{B}} : \mathcal{A}^l \to \mathcal{B}$ defined by $\mathbf{F}^{\mathcal{B}}$ and certain moment-cumulant relations over the noncrossing partition lattice; we refer the reader to Speicher and Vargas (2012) and Speicher ((1998), Chapters 2 and 3) for details. We will use the properties that $\kappa_l^{\mathcal{B}}$ is linear in each argument and satisfy the relations

$$\kappa_l^{\mathcal{B}}\big(ba_1, a_2, \dots, a_{l-1}, a_l b'\big) = b\kappa_l^{\mathcal{B}}(a_1,\dots,a_l)b', \tag{3.9}$$
$$\kappa_l^{\mathcal{B}}(a_1,\dots,a_{j-1}, a_j b, a_{j+1},\dots,a_l) = \kappa_l^{\mathcal{B}}(a_1,\dots,a_j, ba_{j+1},\dots,a_l) \tag{3.10}$$

for any $b, b' \in \mathcal{B}$ and $a_1,\dots,a_l \in \mathcal{A}$.

For $a \in \mathcal{A}$, the *$\mathcal{B}$-valued $\mathcal{R}$-transform* of a is defined, for $b \in \mathcal{B}$, as

$$\mathcal{R}_a^{\mathcal{B}}(b) := \sum_{l\ge1} \kappa_l^{\mathcal{B}}(ab,\dots,ab,a). \tag{3.11}$$

The *$\mathcal{B}$-valued Cauchy transform* of a is defined, for invertible $b \in \mathcal{B}$, as

$$G_a^{\mathcal{B}}(b) := \mathbf{F}^{\mathcal{B}}\big((b-a)^{-1}\big) = \sum_{l\ge0} \mathbf{F}^{\mathcal{B}}\big(b^{-1}(ab^{-1})^l\big), \tag{3.12}$$

with the convention $a^0 = 1$ for all $a \in \mathcal{A}$. The moment-cumulant relations imply that $G_a^{\mathcal{B}}(b)$ and $\mathcal{R}_a^{\mathcal{B}}(b) + b^{-1}$ are inverses with respect to composition.

PROPOSITION 3.12. *Let $(\mathcal{A}, \mathcal{B}, \mathbf{F}^{\mathcal{B}})$ be a $\mathcal{B}$-valued probability space. For $a \in \mathcal{A}$ and invertible $b \in \mathcal{B}$,*

$$G_a^{\mathcal{B}}\big(b^{-1} + \mathcal{R}_a^{\mathcal{B}}(b)\big) = b, \tag{3.13}$$

$$G_a^{\mathcal{B}}(b) = \big(b - \mathcal{R}_a^{\mathcal{B}}\big(G_a^{\mathcal{B}}(b)\big)\big)^{-1}. \tag{3.14}$$

PROOF. See Voiculescu ((1995), Theorem 4.9) and also Speicher ((1998), Theorem 4.1.12). $\square$

REMARK. When $\mathcal{A}$ is a von Neumann algebra, the right-hand sides of (3.11) and (3.12) may be understood as convergent series in $\mathcal{A}$ with respect to the norm $\|\cdot\|$, for sufficiently small $\|b\|$ and $\|b^{-1}\|$, respectively. Indeed, (3.12) defines a convergent series in $\mathcal{B}$ when $\|b^{-1}\| < 1/\|a\|$, with

$$\big\|G_a^{\mathcal{B}}(b)\big\| \le \sum_{l\ge 0} \big\|b^{-1}\big\|^{l+1} \|a\|^l = \frac{\|b^{-1}\|}{1 - \|a\| \|b^{-1}\|}. \tag{3.15}$$

Also, explicit inversion of the moment-cumulant relations for the noncrossing partition lattice yields the cumulant bound

$$\big\|\kappa_l^{\mathcal{B}}(a_1, \dots, a_l)\big\| \le 16^l \prod_{i=1}^{l} \|a_i\| \tag{3.16}$$

(see Nica and Speicher (2006), Proposition 13.15), so (3.11) defines a convergent series in $\mathcal{B}$ when $16\|b\| < 1/\|a\|$, with

$$\big\|\mathcal{R}_a^{\mathcal{B}}(b)\big\| \le \sum_{l\ge 1} 16^l \|a\|^l \|b\|^{l-1} = \frac{16\|a\|}{1 - 16\|a\| \|b\|}.$$

The identities (3.13) and (3.14) hold as equalities of elements in $\mathcal{B}$ when $\|b\|$ and $\|b^{-1}\|$ are sufficiently small, respectively.

Our computation will pass between $\mathcal{R}$-transforms and Cauchy transforms with respect to nested subalgebras of $\mathcal{A}$. Central to this approach is the following result from Nica, Shlyakhtenko and Speicher (2002) (see also Speicher and Vargas (2012)).

PROPOSITION 3.13. *Let $(\mathcal{A}, \mathcal{D}, \mathbf{F}^{\mathcal{D}})$ be a $\mathcal{D}$-valued probability space, let $\mathcal{B}, \mathcal{H} \subseteq \mathcal{A}$ be sub-$*$-algebras containing $\mathcal{D}$ and let $\mathbf{F}^{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$ be a conditional expectation such that $\mathbf{F}^{\mathcal{D}} \circ \mathbf{F}^{\mathcal{B}} = \mathbf{F}^{\mathcal{D}}$. Let $\kappa_l^{\mathcal{B}}$ and $\kappa_l^{\mathcal{D}}$ denote the free cumulants for $\mathbf{F}^{\mathcal{B}}$ and $\mathbf{F}^{\mathcal{D}}$. If $\mathcal{B}$ and $\mathcal{H}$ are $\mathcal{D}$-free, then for all $l \ge 1$, $h_1, \dots, h_l \in \mathcal{H}$ and $b_1, \dots, b_{l-1} \in \mathcal{B}$,*

$$\kappa_l^{\mathcal{B}}(h_1 b_1, \dots, h_{l-1} b_{l-1}, h_l) = \kappa_l^{\mathcal{D}}\big(h_1 \mathbf{F}^{\mathcal{D}}(b_1), \dots, h_{l-1} \mathbf{F}^{\mathcal{D}}(b_{l-1}), h_l\big).$$

PROOF. See Nica, Shlyakhtenko and Speicher (2002), Theorem 3.6. □

For subalgebras $\mathcal{D} \subseteq \mathcal{B} \subseteq \mathcal{A}$ and conditional expectations $\mathbf{F}^{\mathcal{D}} : \mathcal{A} \to \mathcal{D}$ and $\mathbf{F}^{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$ satisfying (3.3), we also have for any $a \in \mathcal{A}$ and invertible $d \in \mathcal{D}$ (with sufficiently small $\|d^{-1}\|$), by (3.12),

$$G_a^{\mathcal{D}}(d) = \mathbf{F}^{\mathcal{D}} \circ G_a^{\mathcal{B}}(d). \tag{3.17}$$

Finally, note that for $\mathcal{B} = \mathbb{C}$ and $\mathbf{F}^{\mathcal{B}} = \tau$, the scalar-valued Cauchy transform $G_a^{\mathbb{C}}(z)$ is simply $-m_a(z)$ from (3.8). (The minus sign is a difference in sign convention for the Cauchy–Stieltjes transform.)

4. Computation in the free model. We will prove analogues of Theorems 1.2 and 1.5 for a slightly more general matrix model: Fix $k \geq 1$, let $p, n_1, \dots, n_k, m_1, \dots, m_k \in \mathbb{N}$ and denote $n_+ = \sum_{r=1}^k n_r$. Let $F \in \mathbb{C}^{n_+ \times n_+}$ be deterministic with $F^* = F$, and denote by $F_{rs} \in \mathbb{C}^{n_r \times n_s}$ its (r, s) submatrix. For $r = 1, \dots, k$, let $H_r \in \mathbb{C}^{m_r \times p}$ be deterministic, and let G_r be independent random matrices such that either $G_r \in \mathbb{R}^{n_r \times m_r}$ with $(G_r)_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, n_r^{-1})$ or $G_r \in \mathbb{C}^{n_r \times m_r}$ with $\Im(G_r)_{ij}, \Re(G_r)_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, (2n_r)^{-1})$. Define

$$W := \sum_{r,s=1}^k H_r^* G_r^* F_{rs} G_s H_s \in \mathbb{C}^{p \times p},$$

with empirical spectral measure μ_W. Denote $b \cdot H^* H = \sum_{s=1}^k b_s H_s^* H_s$, and let $D(a)$ and Tr_r be as in Theorem 1.2.

THEOREM 4.1. *Suppose $p, n_1, \dots, n_k, m_1, \dots, m_k \to \infty$, such that $c < n_r/p < C$, $c < m_r/p < C$, $\|H_r\| < C$, and $\|F_{rs}\| < C$ for all $r, s = 1, \dots, k$ and some constants $C, c > 0$. Then*:

(a) *For each $z \in \mathbb{C}^+$, there exist unique values $a_1, \dots, a_k \in \mathbb{C}^+ \cup \{0\}$ and $b_1, \dots, b_k \in \overline{\mathbb{C}^+}$ that satisfy, for $r = 1, \dots, k$, the equations*

$$a_r = -\frac{1}{n_r} \mathrm{Tr}\big((z \,\mathrm{Id}_p + b \cdot H^* H)^{-1} H_r^* H_r\big), \tag{4.1}$$

$$b_r = -\frac{1}{n_r} \mathrm{Tr}_r\big([\mathrm{Id}_{n_+} + F D(a)]^{-1} F\big). \tag{4.2}$$

(b) *$\mu_W - \mu_0 \to 0$ weakly a.s. for a probability measure μ_0 on $\mathbb{R}$ with Stieltjes transform*

$$m_0(z) := -\frac{1}{p} \mathrm{Tr}\big((z \,\mathrm{Id}_p + b \cdot H^* H)^{-1}\big). \tag{4.3}$$

(c) *For each $z \in \mathbb{C}^+$, the values a_r, b_r in (a) are the limits, as $t \to \infty$, of $a_r^{(t)}$, $b_r^{(t)}$ computed by iterating (4.1)–(4.2) in the manner of Theorem 1.5.*

Theorems 1.2 and 1.5 follow by specializing this result to $F = U^T BU$ and $m_r = p$, $n_r = I_r$ and $H_r = \Sigma_r^{1/2}$ for each $r = 1, \dots, k$.

In this section, we carry out the bulk of the proof of Theorem 4.1 by:

1. Defining a free deterministic equivalent for this matrix model, and
2. Showing that the Stieltjes transform of the element w (modeling W) satisfies (4.1)–(4.3).

These steps correspond to the separation of approximation and computation discussed in Section 1.4.

For the reader's convenience, in Appendix E we provide a simplified version of these steps for the special case of Theorem 4.1 corresponding to Theorem 1.1 for sample covariance matrices, which illustrates the main ideas.

4.1. *Defining a free deterministic equivalent.* Consider the transformations

$$H_r \mapsto O_r^T H_r O_0, \qquad F_{rs} \mapsto O_{k+r}^T F_{rs} O_{k+s}$$

for independent Haar-distributed orthogonal matrices $O_0, \dots, O_{2k}$ of the appropriate sizes. As in Section 1.4, μ_W remains invariant in law under these transformations. Hence it suffices to prove Theorem 4.1 with H_r and F_{rs} replaced by these randomly rotated matrices, which (with a slight abuse of notation) we continue to denote by H_r and F_{rs}.

Let $N = p + \sum_{r=1}^k m_r + \sum_{r=1}^k n_r$, and embed the matrices W, H_r, G_r, F_{rs} as simple elements of $\mathbb{C}^{N \times N}$ in the following regions of the block-matrix decomposition corresponding to $\mathbb{C}^N = \mathbb{C}^p \oplus \mathbb{C}^{m_1} \oplus \cdots \oplus \mathbb{C}^{m_k} \oplus \mathbb{C}^{n_1} \oplus \cdots \oplus \mathbb{C}^{n_k}$:

W	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}

Denote by $P_0, \dots, P_{2k}$ the diagonal projections corresponding to the above decomposition, and by $\tilde{W}, \tilde{F}_{rs}, \tilde{G}_r, \tilde{H}_r \in \mathbb{C}^{N \times N}$ the embedded matrices (i.e., $P_0 = \operatorname{diag}(\operatorname{Id}_p, 0, \dots, 0)$, $P_1 = \operatorname{diag}(0, \operatorname{Id}_{m_1}, \dots, 0)$, etc. $\tilde{W}$ has upper-left block equal to W and remaining blocks 0, etc.). Then $\tilde{W}$, $\tilde{F}_{rs}$, $\tilde{G}_r$, $\tilde{H}_r$ are simple elements of the rectangular space $(\mathbb{C}^{N \times N}, N^{-1} \operatorname{Tr}, P_0, \dots, P_{2k})$, and the $k+2$ families $\{\tilde{F}_{rs}\}$, $\{\tilde{H}_r\}$, $\tilde{G}_1, \dots, \tilde{G}_k$ are independent of each other and are block-orthogonally invariant.

For the approximating free model, consider a second (N-dependent) rectangular space $(\mathcal{A}, \tau, p_0, \dots, p_{2k})$ with deterministic elements $f_{rs}, g_r, h_r \in \mathcal{A}$, such that the following hold:

1. $p_0, \dots, p_{2k}$ have traces

$$\tau(p_0) = p/N, \qquad \tau(p_r) = m_r/N,$$
$$\tau(p_{k+r}) = n_r/N \qquad \text{for all } r = 1, \dots, k.$$

2. f_{rs}, g_r, h_r are simple elements such that for all $r, s \in \{1, \dots, k\}$,

$$p_{k+r} f_{rs} p_{k+s} = f_{rs}, \qquad p_{k+r} g_r p_r = g_r, \qquad p_r h_r p_0 = h_r.$$

3. $\{f_{rs} : 1 \leq r, s \leq k\}$ has the same joint $\mathcal{D}$-law as $\{\tilde{F}_{rs} : 1 \leq r, s \leq k\}$, and $\{h_r : 1 \leq r \leq k\}$ has the same joint $\mathcal{D}$-law as $\{\tilde{H}_r : 1 \leq r \leq k\}$. That is, for any $r \in \{0, \dots, 2k\}$ and any noncommutative $*$-polynomials Q_1, Q_2 with coefficients in $\langle P_0, \dots, P_{2k}\rangle$, letting q_1, q_2 denote the corresponding $*$-polynomials with coefficients in $\langle p_0, \dots, p_{2k}\rangle$,

$$\tau_r\big[q_1\big(f_{st} : s, t \in \{1, \dots, k\}\big)\big] = N_r^{-1} \operatorname{Tr}_r Q_1\big(\tilde{F}_{s,t} : s, t \in \{1, \dots, k\}\big), \tag{4.4}$$

$$\tau_r\big[q_2\big(h_s : s \in \{1, \dots, k\}\big)\big] = N_r^{-1} \operatorname{Tr}_r Q_2\big(\tilde{H}_s : s \in \{1, \dots, k\}\big). \tag{4.5}$$

4. For each r, $g_r^* g_r$ has Marcenko–Pastur law with parameter $\lambda = m_r/n_r$. That is, for ν_λ as in (3.6),

$$\tau_r\big((g_r^* g_r)^l\big) = \int x^l \nu_{m_r/n_r}(x)\, dx \qquad \text{for all } l \geq 0. \tag{4.6}$$

5. The $k+2$ families $\{f_{rs}\}$, $\{h_r\}$, $g_1, \dots, g_k$ are free with amalgamation over $\mathcal{D} = \langle p_0, \dots, p_{2k}\rangle$.

The right-hand sides of (4.4) and (4.5) are deterministic, as they are invariant to the random rotations of F_{rs} and H_r. Also, (4.6) completely specifies $\tau(q(g_r))$ for any $*$-polynomial q with coefficients in $\mathcal{D}$. Then these conditions 1–5 fully specify the joint $\mathcal{D}$-law of all elements $f_{rs}, g_r, h_r \in \mathcal{A}$. These elements are a free deterministic equivalent for $\tilde{F}_{rs}, \tilde{G}_r, \tilde{H}_r \in \mathbb{C}^{N \times N}$ in the sense of Definition 3.9.

The following lemma establishes existence of this model as a von Neumann algebra; its proof is deferred to Appendix D.

LEMMA 4.2. *Under the conditions of Theorem 4.1, there exists a (N-dependent) rectangular probability space $(\mathcal{A}, \tau, p_0, \dots, p_{2k})$ such that*:

(a) *$\mathcal{A}$ is a von Neumann algebra and τ is a positive, faithful, normal trace.*

(b) *$\mathcal{A}$ contains elements f_{rs}, g_r, h_r for $r, s \in \{1, \dots, k\}$ that satisfy the above conditions. Furthermore, the von Neumann subalgebras $\langle \mathcal{D}, \{f_{rs}\}\rangle_{W^*}$, $\langle \mathcal{D}, \{h_r\}\rangle_{W^*}$, $\langle \mathcal{D}, g_1\rangle_{W^*}, \dots, \langle \mathcal{D}, g_k\rangle_{W^*}$ are free over $\mathcal{D}$.*

(c) *There exists a constant $C > 0$ such that $\|f_{rs}\|, \|h_r\|, \|g_r\| \leq C$ for all N and all r, s.*

4.2. *Computing the Stieltjes transform of* w. We will use twice the following intermediary lemma, whose proof follows ideas of Speicher and Vargas (2012) and which we defer to Appendix D.

LEMMA 4.3. *Let* $(\mathcal{A}, \tau, q_0, q_1, \dots, q_k)$ *be a rectangular probability space, where* $\mathcal{A}$ *is von Neumann and* τ *is positive, faithful and normal. Let* $\mathcal{D} = \langle q_0, \dots, q_k \rangle$, *let* $\mathcal{B}, \mathcal{C} \subset \mathcal{A}$ *be von Neumann subalgebras containing* $\mathcal{D}$ *that are free over* $\mathcal{D}$ *and let* $\mathbf{F}^{\mathcal{D}} : \mathcal{A} \to \mathcal{D}$ *and* $\mathbf{F}^{\mathcal{C}} : \mathcal{A} \to \mathcal{C}$ *be the* τ*-invariant conditional expectations.*

Let $b_{rs} \in \mathcal{B}$ *and* $c_r \in \mathcal{C}$ *for* $1 \le r, s \le k$ *be such that* $q_r b_{rs} q_s = b_{rs}$, $q_r c_r = c_r$, $\|b_{rs}\| \le C$, *and* $\|c_r\| \le C$ *for some constant* $C > 0$. *Define* $a = \sum_{r,s=1}^{k} c_r^* b_{rs} c_s$ *and* $b = \sum_{r,s=1}^{k} b_{rs}$. *Then, for* $e \in \mathcal{C}$ *with* $\|e\|$ *sufficiently small,*

$$\mathcal{R}_a^{\mathcal{C}}(e) = \sum_{r=1}^{k} c_r^* c_r \tau_r \left(\mathcal{R}_b^{\mathcal{D}} \left(\sum_{s=1}^{k} \tau_s (c_s e c_s^*) q_s \right) \right),$$

where $\mathcal{R}_a^{\mathcal{C}}$ *and* $\mathcal{R}_b^{\mathcal{D}}$ *are the* $\mathcal{C}$*-valued and* $\mathcal{D}$*-valued* $\mathcal{R}$*-transforms of* a *and* b.

We now perform the desired computation of the Stieltjes transform of w.

LEMMA 4.4. *Under the conditions of Theorem* 4.1, *let* $(\mathcal{A}, \tau, p_0, \dots, p_{2k})$ *and* f_{rs}, g_r, h_r *be as in Lemma* 4.2, *and let* $w = \sum_{r,s=1}^{k} h_r^* g_r^* f_{rs} g_s h_s$. *Then for a constant* $C_0 > 0$, *defining* $\mathbb{D} := \{z \in \mathbb{C}^+ : |z| > C_0\}$, *there exist analytic functions* $a_1, \dots, a_k : \mathbb{D} \to \mathbb{C}^+ \cup \{0\}$ *and* $b_1, \dots, b_k : \mathbb{D} \to \mathbb{C}$ *that satisfy, for every* $z \in \mathbb{D}$ *and for* $m_0(z) = \tau_0((w - z)^{-1})$, *equations* (4.1)–(4.3).

PROOF. If $H_r = 0$ for some r, then we may set $a_r \equiv 0$, define b_r by (4.2) and reduce to the case $k - 1$. Hence, it suffices to consider $H_r \neq 0$ for all r.

Define the von Neumann subalgebras $\mathcal{D} = \langle p_r : 0 \le r \le 2k \rangle$, $\mathcal{F} = \langle \mathcal{D}, \{f_{rs}\} \rangle_{W^*}$, $\mathcal{G} = \langle \mathcal{D}, \{g_r\} \rangle_{W^*}$, and $\mathcal{H} = \langle \mathcal{D}, \{h_r\} \rangle_{W^*}$. Denote by $\mathbf{F}^{\mathcal{D}}$, $\mathcal{R}^{\mathcal{D}}$, and $G^{\mathcal{D}}$ the τ-invariant conditional expectation onto $\mathcal{D}$ and the $\mathcal{D}$-valued $\mathcal{R}$-transform and Cauchy transform, and similarly for $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$.

We first work algebraically (Steps 1–3), assuming that arguments b to Cauchy transforms are invertible with $\|b^{-1}\|$ sufficiently small, arguments b to $\mathcal{R}$-transforms have $\|b\|$ sufficiently small, and applying series expansions for $(b - a)^{-1}$. We will check that these assumptions hold and also establish the desired analyticity properties in Step 4.

Step 1: We first relate the $\mathcal{D}$-valued Cauchy transform of w to that of $v := \sum_{r,s=1}^{k} g_r^* f_{rs} g_s$. We apply Lemma 4.3 with $q_0 = p_0 + \sum_{r=k+1}^{2k} p_r$, $q_r = p_r$ for $r = 1, \dots, k$, $\mathcal{C} = \mathcal{H}$ and $\mathcal{B} = \langle \mathcal{F}, \mathcal{G} \rangle$. Then for $c \in \mathcal{H}$,

$$\mathcal{R}_w^{\mathcal{H}}(c) = \sum_{r=1}^{k} h_r^* h_r \tau_r \left(\mathcal{R}_v^{\mathcal{D}} \left(\sum_{s=1}^{k} p_s \tau_s (h_s c h_s^*) \right) \right). \tag{4.7}$$

To rewrite this using Cauchy transforms, for invertible $d \in \mathcal{D}$ and each $r = 1, \ldots, k$, define

$$\alpha_r(d) := \tau_r\big(h_r G_w^{\mathcal{H}}(d) h_r^*\big), \tag{4.8}$$

$$\beta_r(d) := \tau_r\left(\mathcal{R}_v^{\mathcal{D}}\left(\sum_{s=1}^k p_s \alpha_s(d)\right)\right). \tag{4.9}$$

Then (3.14) and (4.7) with $c = G_w^{\mathcal{H}}(d)$ imply

$$G_w^{\mathcal{H}}(d) = \big(d - \mathcal{R}_w^{\mathcal{H}}\big(G_w^{\mathcal{H}}(d)\big)\big)^{-1} = \left(d - \sum_{r=1}^k h_r^* h_r \beta_r(d)\right)^{-1}. \tag{4.10}$$

Projecting down to $\mathcal{D}$ using (3.17) yields

$$G_w^{\mathcal{D}}(d) = \mathbf{F}^{\mathcal{D}}\left(\left(d - \sum_{r=1}^k h_r^* h_r \beta_r(d)\right)^{-1}\right). \tag{4.11}$$

Applying (4.10) to (4.8),

$$\alpha_r(d) = \tau_r\left(h_r\left(d - \sum_{s=1}^k h_s^* h_s \beta_s(d)\right)^{-1} h_r^*\right). \tag{4.12}$$

Noting that $(p_1 + \cdots + p_k)v(p_1 + \cdots + p_k) = v$, (3.11) and (3.9) imply $\mathcal{R}_v^{\mathcal{D}}(d) \in \langle p_1, \ldots, p_k\rangle$ for any $d \in \mathcal{D}$, so we may write (4.9) as

$$\mathcal{R}_v^{\mathcal{D}}\left(\sum_{r=1}^k p_r \alpha_r(d)\right) = \sum_{r=1}^k p_r \beta_r(d).$$

For $r = 0$ and $r \in \{k+1, \ldots, 2k\}$, set $\beta_r(d) = 0$ and define $\alpha_r(d)$ arbitrarily, say by $\alpha_r(d) = \|d^{-1}\|$. Since $vp_r = p_r v = 0$ if $r = 0$ or $r \in \{k+1, \ldots, 2k\}$, applying (3.11) and multilinearity of $\kappa_l^{\mathcal{D}}$, we may rewrite the above as

$$\mathcal{R}_v^{\mathcal{D}}\left(\sum_{r=0}^{2k} p_r \alpha_r(d)\right) = \sum_{r=0}^{2k} p_r \beta_r(d).$$

Applying (3.13) with $b = \sum_{r=0}^{2k} p_r \alpha_r(d)$, we get

$$G_v^{\mathcal{D}}\left(\sum_{r=0}^{2k} p_r\left(\frac{1}{\alpha_r(d)} + \beta_r(d)\right)\right) = \sum_{r=0}^{2k} p_r \alpha_r(d). \tag{4.13}$$

The relation between $G_w^{\mathcal{D}}$ and $G_v^{\mathcal{D}}$ is given by (4.11), (4.12) and (4.13).

Step 2: Next, we relate the $\mathcal{D}$-valued Cauchy transforms of v and $u := \sum_{r,s=1}^k f_{rs}$. We apply Lemma 4.3 with $q_0 = \sum_{r=0}^k p_r$, $q_r = p_{r+k}$ for $r = 1, \ldots, k$,

$\mathcal{C} = \mathcal{G}$ and $\mathcal{B} = \mathcal{F}$. Then for $c \in \mathcal{G}$,

$$\mathcal{R}_v^{\mathcal{G}}(c) = \sum_{r=1}^{k} g_r^* g_r \tau_{r+k}\left(\mathcal{R}_u^{\mathcal{D}}\left(\sum_{s=1}^{k} p_{s+k}\tau_{s+k}\big(g_s c g_s^*\big)\right)\right). \tag{4.14}$$

To rewrite this using Cauchy transforms, for invertible $d \in \mathcal{D}$ and all $r = 1, \dots, k$, define

$$\gamma_{r+k}(d) = \tau_{r+k}\big(g_r G_v^{\mathcal{G}}(d) g_r^*\big), \tag{4.15}$$

$$\delta_{r+k}(d) = \tau_{r+k}\left(\mathcal{R}_u^{\mathcal{D}}\left(\sum_{s=1}^{k} p_{s+k}\gamma_{s+k}(d)\right)\right). \tag{4.16}$$

As in Step 1, for $r = 0, \dots, k$ let us also define $\delta_r(d) = 0$ and $\gamma_r(d) = \|d^{-1}\|$. Then, noting $(p_{k+1} + \cdots + p_{2k})u(p_{k+1} + \cdots + p_{2k}) = u$, the same arguments as in Step 1 yield the analogous identities

$$G_v^{\mathcal{D}}(d) = \mathbf{F}^{\mathcal{D}}\left(\left(d - \sum_{s=1}^{k} g_s^* g_s \delta_{s+k}(d)\right)^{-1}\right), \tag{4.17}$$

$$\gamma_{r+k}(d) = \tau_{r+k}\left(g_r\left(d - \sum_{s=1}^{k} g_s^* g_s \delta_{s+k}(d)\right)^{-1} g_r^*\right), \tag{4.18}$$

$$G_u^{\mathcal{D}}\left(\sum_{r=0}^{2k} p_r\left(\frac{1}{\gamma_r(d)} + \delta_r(d)\right)\right) = \sum_{r=0}^{2k} p_r \gamma_r(d). \tag{4.19}$$

As $g_r^* g_r$ has moments given by (4.6), we may write (4.17) and (4.18) explicitly: Denote $d = d_0 p_0 + \cdots + d_{2k} p_{2k}$ for $d_0, \dots, d_{2k} \in \mathbb{C}$. As d is invertible, we have $d^{-1} = d_0^{-1} p_0 + \cdots + d_{2k}^{-1} p_{2k}$. For any $x \in \mathcal{A}$ that commutes with $\mathcal{D}$,

$$(d - x)^{-1} = \sum_{l \geq 0} d^{-1}\big(x d^{-1}\big)^l = \sum_{l \geq 0} x^l d^{-l-1}.$$

So for $r = 1, \dots, k$, noting that $p_r = p_r^2$ and that $\mathcal{D}$ commutes with itself,

$$\begin{aligned}\tau_r\big((d - x)^{-1}\big) &= \frac{N}{m_r} \sum_{l \geq 0} \tau\big(p_r x^l d^{-l-1} p_r\big) \\ &= \frac{N}{m_r} \sum_{l \geq 0} \tau\big((p_r x^l p_r)(p_r d^{-1} p_r)^{l+1}\big) = \sum_{l \geq 0} \frac{\tau_r(x^l)}{d_r^{l+1}}.\end{aligned}$$

Noting that $g_s^* g_s$ commutes with $\mathcal{D}$, applying the above to (4.17) with $x = \sum_{s=1}^k g_s^* g_s \delta_{s+k}(d)$, and recalling (4.6),

$$\tau_r\big(G_v^{\mathcal{D}}(d)\big) = \sum_{l \ge 0} \frac{\tau_r((g_r^* g_r)^l)\delta_{r+k}(d)^l}{d_r^{l+1}}$$

$$= \int \sum_{l\ge 0} \frac{x^l \delta_{r+k}(d)^l}{d_r^{l+1}} \nu_{m_r/n_r}(x)\,dx$$

$$= \int \frac{1}{d_r - x\delta_{r+k}(d)} \nu_{m_r/n_r}(x)\,dx$$

$$= \frac{1}{\delta_{r+k}(d)} G^{\mathbb{C}}_{\nu_{m_r/n_r}}\big(d_r/\delta_{r+k}(d)\big), \tag{4.20}$$

where $G^{\mathbb{C}}_{\nu_{m_r/n_r}}$ is the Cauchy transform of the Marcenko–Pastur law ν_{m_r/n_r}.

Similarly, we may write (4.18) as

$$\gamma_{r+k}(d) = \frac{m_r}{n_r}\tau_r\left(\left(d - \sum_{s=1}^k g_s^* g_s \delta_{s+k}(d)\right)^{-1} g_r^* g_r\right)$$

$$= \frac{m_r}{n_r}\int \frac{x}{d_r - x\delta_{r+k}(d)} \nu_{m_r/n_r}(x)\,dx$$

$$= \frac{m_r}{n_r}\left(-\frac{1}{\delta_{r+k}(d)} + \frac{d_r}{\delta_{r+k}(d)^2} G^{\mathbb{C}}_{\nu_{m_r/n_r}}\big(d_r/\delta_{r+k}(d)\big)\right)$$

$$= \frac{m_r}{n_r}\left(-\frac{1}{\delta_{r+k}(d)} + \frac{d_r}{\delta_{r+k}(d)} \tau_r\big(G_v^{\mathcal{D}}(d)\big)\right), \tag{4.21}$$

where the first equality applies the cyclic property of τ and the definitions of τ_{r+k} and τ_r, the second applies (4.6) upon passing to a power series and back as above, the third applies the definition of the Cauchy transform and the last applies (4.20). The relation between $G_v^{\mathcal{D}}$ and $G_u^{\mathcal{D}}$ is given by (4.20), (4.21) and (4.19).

Step 3: We compute $m_0(z)$ for $z \in \mathbb{C}^+$ using (4.11), (4.12), (4.13), (4.20), (4.21) and (4.19). Fixing $z \in \mathbb{C}^+$, let us write

$$\alpha_r = \alpha_r(z), \qquad \beta_r = \beta_r(z), \qquad d_r = \frac{1}{\alpha_r} + \beta_r, \qquad d = \sum_{r=0}^{2k} d_r p_r,$$

$$\gamma_r = \gamma_r(d), \qquad \delta_r = \delta_r(d), \qquad e_r = \frac{1}{\gamma_r} + \delta_r, \qquad e = \sum_{r=0}^{2k} e_r p_r.$$

Applying (4.11) and projecting down to $\mathbb{C}$,

$$m_0(z) = -\tau_0\left(\left(z - \sum_{r=1}^k h_r^* h_r \beta_r\right)^{-1}\right).$$

Note that $h_r^* h_r$ commutes with $\mathcal{D}$ and $p_0 h_r^* h_r p_0 = h_r^* h_r$ for each $r = 1, \dots, k$. Then, passing to a power series as in Step 2, and then applying (4.5) and the spectral calculus,

$$
\begin{aligned}
m_0(z) &= -\sum_{l\geq 0} z^{-(l+1)} \tau_0\left(\left(\sum_{r=1}^k h_r^* h_r \beta_r\right)^l\right) \\
&= -\sum_{l\geq 0} z^{-(l+1)} \frac{1}{p} \operatorname{Tr}\left(\left(\sum_{r=1}^k \beta_r H_r^* H_r\right)^l\right) \\
&= -\frac{1}{p} \operatorname{Tr}\left(z \operatorname{Id}_p - \sum_{r=1}^k \beta_r H_r^* H_r\right)^{-1}. \qquad (4.22)
\end{aligned}
$$

Similarly, (4.12) implies for each $r = 1, \dots, k$

$$
\alpha_r = \frac{1}{m_r} \operatorname{Tr}\left(\left(z \operatorname{Id}_p - \sum_{s=1}^k \beta_s H_s^* H_s\right)^{-1} H_r^* H_r\right). \qquad (4.23)
$$

Now applying (4.20) and recalling (4.13) and the definition of d_r, for each $r = 1, \dots, k$,

$$
\alpha_r = \tau_r\big(G_v^{\mathcal{D}}(d)\big) = \frac{1}{\delta_{r+k}} G_{\nu_{m_r/n_r}}^{\mathbb{C}}\left(\frac{1}{\alpha_r \delta_{r+k}} + \frac{\beta_r}{\delta_{r+k}}\right).
$$

Applying (3.14) and the Marcenko–Pastur $\mathcal{R}$-transform $\mathcal{R}_{\nu_\lambda}^{\mathbb{C}}(z) = (1 - \lambda z)^{-1}$, this is rewritten as

$$
\frac{\beta_r}{\delta_{r+k}} = \mathcal{R}_{\nu_{m_r/n_r}}^{\mathbb{C}}(\alpha_r \delta_{r+k}) = \frac{n_r}{n_r - m_r \alpha_r \delta_{r+k}}. \qquad (4.24)
$$

By (4.21) and (4.13),

$$
\gamma_{r+k} = \frac{m_r}{n_r} \frac{\alpha_r \beta_r}{\delta_{r+k}}. \qquad (4.25)
$$

We derive two consequences of (4.24) and (4.25). First, substituting for β_r in (4.25) using (4.24) and recalling the definition of e_{r+k} yields

$$
e_{r+k} = \frac{n_r}{m_r \alpha_r}. \qquad (4.26)
$$

Second, rearranging (4.24), we get $\beta_r / \delta_{r+k} = 1 + m_r \alpha_r \beta_r / n_r$. Inserting into (4.25) yields this time

$$
\beta_r = \frac{n_r}{m_r^2 \alpha_r^2} (n_r \gamma_{r+k} - m_r \alpha_r). \qquad (4.27)
$$

By (4.19), for each $r = 1, \dots, k$,

$$
\gamma_{r+k} = \tau_{r+k}\big(G_u^{\mathcal{D}}(e)\big) = \tau_{r+k}\big((e - u)^{-1}\big).
$$

Passing to a power series for $(e-u)^{-1}$, applying (4.4) and passing back,

$$\begin{aligned}\gamma_{r+k} &= \frac{1}{n_r}\operatorname{Tr}_{r+k}\big(\operatorname{diag}(e_0 \operatorname{Id}_p, \ldots, e_{2k} \operatorname{Id}_{n_k}) - \tilde{F}\big)^{-1} \\ &= \frac{1}{n_r}\operatorname{Tr}_r\big(\operatorname{diag}(e_{k+1} \operatorname{Id}_{n_1}, \ldots, e_{2k} \operatorname{Id}_{n_k}) - F\big)^{-1} \\ &= \frac{1}{n_r}\operatorname{Tr}_r\big(D^{-1} - F\big)^{-1}, \end{aligned} \tag{4.28}$$

where the last line applies (4.26) and sets $D = \operatorname{diag}(D_1 \operatorname{Id}_{n_1}, \ldots, D_k \operatorname{Id}_{n_k})$ for $D_r = m_r\alpha_r/n_r$. Noting $\operatorname{Tr}_r D = m_r\alpha_r$, (4.27) yields

$$\begin{aligned}\beta_r &= \frac{1}{n_r D_r^2}\operatorname{Tr}_r\big[\big(D^{-1} - F\big)^{-1} - D\big] \\ &= \frac{1}{n_r}\operatorname{Tr}_r\big[\big(F^{-1} - D\big)^{-1}\big] = \frac{1}{n_r}\operatorname{Tr}_r\big((\operatorname{Id}_{n_+} - FD)^{-1}F\big), \end{aligned} \tag{4.29}$$

where we used the Woodbury identity and $\operatorname{Tr}_r DAD = D_r^2 \operatorname{Tr} A$. (These equalities hold when F is invertible, and hence for all F by continuity.) Setting $a_r = -m_r\alpha_r/n_r$ and $b_r = -\beta_r$, we obtain (4.1), (4.2) and (4.3) from (4.22), (4.23) and (4.29).

Step 4: Finally, we verify the validity of the preceding calculations when $z \in \mathbb{D} := \{z \in \mathbb{C}^+ : |z| > C_0\}$ and $C_0 > 0$ is sufficiently large. Call a scalar quantity $u := u(N, z)$ "uniformly bounded" if $|u| < C$ for all $z \in \mathbb{D}$, all N and some constants $C_0, C > 0$. Call u "uniformly small" if for any constant $c > 0$ there exists $C_0 > 0$ such that $|u| < c$ for all $z \in \mathbb{D}$ and all N.

As $\|w\| \le C$ by Lemma 4.2(c), $c = G_w^{\mathcal{H}}(z)$ is well defined by the convergent series (3.12) for $z \in \mathbb{D}$. Furthermore, by (3.15), $\|c\|$ is uniformly small, so we may apply (4.7). $\alpha_r(z)$ as defined by (4.8) satisfies

$$\begin{aligned}\alpha_r(z) &= \tau_r\left(h_r \sum_{l=0}^{\infty} \mathbf{F}^{\mathcal{H}}\big(z^{-1}\big(wz^{-1}\big)^l\big)h_r^*\right) \\ &= \sum_{l=0}^{\infty} z^{-(l+1)}\tau(p_r)^{-1}\tau\big(h_r\mathbf{F}^{\mathcal{H}}\big(w^l\big)h_r^*\big) = \sum_{l=0}^{\infty} z^{-(l+1)}\frac{N}{m_r}\tau\big(w^l h_r^* h_r\big)\end{aligned}$$

for $z \in \mathbb{D}$. Since $|\tau(w^l h_r^* h_r)| \le \|w\|^l \|h_r\|^2$, α_r defines an analytic function on $\mathbb{D}$ such that $\alpha_r(z) \sim (zm_r)^{-1}\operatorname{Tr}(H_r^* H_r)$ as $|z| \to \infty$. In particular, since H_r is nonzero by our initial assumption, $\alpha_r(z) \neq 0$ and $\Im\alpha_r(z) < 0$ for $z \in \mathbb{D}$. This verifies that $a_r(z) = -m_r\alpha_r(z)/n_r \in \mathbb{C}^+$ and a_r is analytic on $\mathbb{D}$. Furthermore, α_r is uniformly small for each r. Then applying (3.11), multilinearity of κ_l and (3.16), it is verified that $\beta_r(z)$ defined by (4.9) is uniformly bounded and analytic on $\mathbb{D}$. So $b_r(z) = -\beta_r(z)$ is analytic on $\mathbb{D}$.

As β_r is uniformly bounded, the formal series leading to (4.22) and (4.23) are convergent for $z \in \mathbb{D}$. Furthermore, $d_r = 1/\alpha_r + \beta_r$ is well defined as $\alpha_r \neq 0$, and $\|d^{-1}\|$ is uniformly small. Then $c = G_v^{\mathcal{G}}(d)$ is well defined by (3.12) and also uniformly small, so we may apply (4.14). By the same arguments as above, $\gamma_{r+k}(d)$ as defined by (4.15) is nonzero and uniformly small and $\delta_{r+k}(d)$ as defined by (4.16) is uniformly bounded. Then the formal series leading to (4.20) and (4.21) are convergent for $z \in \mathbb{D}$. Furthermore, $e_r = 1/\gamma_r + \delta_r$ is well defined and $\|e^{-1}\|$ is uniformly small, so the formal series leading to (4.28) is convergent for $z \in \mathbb{D}$. This verifies the validity of the preceding calculations and concludes the proof. □

To complete the proof of Theorem 4.1, we show using a contractive mapping argument similar to Couillet, Debbah and Silverstein (2011), Dupuy and Loubaton (2011) that (4.1)–(4.2) have a unique solution in the stated domains, which is the limit of the procedure in Theorem 1.5. The result then follows from Lemma 4.4 and Corollary 3.11. These arguments are contained in Appendix C.

Acknowledgments. We thank Mark Blows for introducing us to this problem and for much help in guiding us through understanding the quantitative genetics applications. Thanks also to the Associate Editor and referees for suggestions that have improved the presentation of the results.

SUPPLEMENTARY MATERIAL

Supplementary Appendices (DOI: 10.1214/18-AOS1767SUPP; .pdf). The Appendices contain a discussion of more general classification designs, proofs of Theorem 3.10 and Corollary 3.11, the proof of Lemma 4.3 and the conclusion of the proof of Theorem 4.1 and a separate exposition of the proof in Section 4 for the simpler setting of Theorem 1.1.

REFERENCES

BAI, Z., CHEN, J. and YAO, J. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Aust. N. Z. J. Stat.* **52** 423–437. MR2791528

BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792

BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686

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

BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680

BARTON, N. H. (1990). Pleiotropic models of quantitative variation. *Genetics* **124** 773–782.

BENAYCH-GEORGES, F. (2009). Rectangular random matrices, related convolution. *Probab. Theory Related Fields* **144** 471–515. MR2496440

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

BLOWS, M. W. (2007). A tale of two matrices: Multivariate approaches in evolutionary biology. *J. Evol. Biol.* **20** 1–8.

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.

BLOWS, M. W., ALLEN, S. L., COLLET, J. M., CHENOWETH, S. F. and MCGUIGAN, K. (2015). The phenome-wide distribution of genetic variance. *Amer. Nat.* **186** 15–30.

COLLINS, B. (2003). Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability. *Int. Math. Res. Not.* **2003** 953–982. MR1959915

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

COUILLET, R., DEBBAH, M. and SILVERSTEIN, J. W. (2011). A deterministic equivalent for the analysis of correlated MIMO multiple access channels. *IEEE Trans. Inform. Theory* **57** 3493–3514. MR2817033

DOBRIBAN, E. (2015). Efficient computation of limit spectra of sample covariance matrices. *Random Matrices Theory Appl.* **4** 1550019, 36. MR3418848

DOBRIBAN, E. (2017). Sharp detection in PCA under correlations: All eigenvalues matter. *Ann. Statist.* **45** 1810–1833. MR3670197

DUPUY, F. and LOUBATON, P. (2011). On the capacity achieving covariance matrix for frequency selective MIMO channels using the asymptotic approach. *IEEE Trans. Inform. Theory* **57** 5737–5753. MR2857932

DYKEMA, K. (1993). On certain free product factors via an extended matrix model. *J. Funct. Anal.* **112** 31–60. MR1207936

EL KAROUI, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** 2757–2790. MR2485012

FAN, Z. and JOHNSTONE, I. M. (2017). Tracy–Widom at each edge of real covariance estimators. Unpublished manuscript.

FAN, Z. and JOHNSTONE, I. M. (2019). Supplement to "Eigenvalue distributions of variance components estimators in high-dimensional random effects models." DOI:10.1214/18-AOS1767SUPP.

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

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

HACHEM, W., LOUBATON, P. and NAJIM, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. MR2326235

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

HINE, E., MCGUIGAN, K. and BLOWS, M. W. (2014). Evolutionary constraints in high-dimensional trait sets. *Amer. Nat.* **184** 119–131.

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

KIRKPATRICK, M. (2009). Patterns of quantitative genetic variation in multiple dimensions. *Genetica* **136** 271–284.

LAMOTTE, L. R. (1973). Quadratic estimation of variance components. *Biometrics* **29** 311–330. MR0329142

LANDE, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: Body size allometry. *Evolution* **33** 402–416.

LANDE, R. and ARNOLD, S. J. (1983). The measurement of selection on correlated characters. *Evolution* **37** 1210–1226.

LEDOIT, O. and PÉCHÉ, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields* **151** 233–264. MR2834718

LEDOIT, O. and WOLF, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* **40** 1024–1060. MR2985942

LOH, P.-R., TUCKER, G., BULIK-SULLIVAN, B. K., VILHJÁLMSSON, B. J., FINUCANE, H. K., SALEM, R. M., CHASMAN, D. I., RIDKER, P. M., NEALE, B. M. et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47** 284–290.

LUSH, J. L. (1937). *Animal Breeding Plans*. Iowa State College Press, Ames, IA.

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

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

MCGUIGAN, K., COLLET, J. M., MCGRAW, E. A., YE YIXIN, H., ALLEN, S. L., CHENOWETH, S. F. and BLOWS, M. W. (2014). The nature and extent of mutational pleiotropy in gene expression of male drosophila serrata. *Genetics* **196** 911–921.

MESTRE, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans. Inform. Theory* **54** 5113–5129. MR2589886

MOUSTAKAS, A. L. and SIMON, S. H. (2007). On the outage capacity of correlated multiple-path MIMO channels. *IEEE Trans. Inform. Theory* **53** 3887–3903. MR2446543

NICA, A., SHLYAKHTENKO, D. and SPEICHER, R. (2002). Operator-valued distributions. I. Characterizations of freeness. *Int. Math. Res. Not.* **2002** 1509–1538. MR1907203

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

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multispiked case. *Ann. Statist.* **42** 225–254. MR3189485

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

PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718

PHILLIPS, P. C. and ARNOLD, S. J. (1989). Visualizing multivariate selection. *Evolution* 1209–1222.

RAO, C. R. (1971). Minimum variance quadratic unbiased estimation of variance components. *J. Multivariate Anal.* **1** 445–456. MR0301870

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

RAO, N. R., MINGO, J. A., SPEICHER, R. and EDELMAN, A. (2008). Statistical eigen-inference from large Wishart matrices. *Ann. Statist.* **36** 2850–2885. MR2485015

ROBERTSON, A. (1959a). The sampling variance of the genetic correlation coefficient. *Biometrics* **15** 469–485. MR0107568

ROBERTSON, A. (1959b). The sampling variance of the genetic correlation coefficient. *Biometrics* **15** 469–485. MR0107568

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

SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408

SOSHNIKOV, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108** 1033–1056. MR1933444

SPEICHER, R. (1998). Combinatorial theory of the free product with amalgamation and operator-valued free probability theory. *Mem. Amer. Math. Soc.* **132** x+88. MR1407898

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

VOICULESCU, D. (1991). Limit laws for random matrices and free products. *Invent*. *Math*. **104** 201–220. MR1094052

VOICULESCU, D. (1995). Operations on certain non-commutative operator-valued random variables. Recent advances in operator algebras (Orléans, 1992). *Astérisque* **232** 243–275. MR1372537

VOICULESCU, D. (1998). A strengthened asymptotic freeness result for random matrices with applications to free entropy. *Int*. *Math*. *Res*. *Not*. **1998** 41–63. MR1601878

VOICULESCU, D. V., DYKEMA, K. J. and NICA, A. (1992). *Free Random Variables*: *A Noncommutative Probability Approach to Free Products with Applications to Random Matrices*, *Operator Algebras and Harmonic Analysis on Free Groups*. *CRM Monograph Series* **1**. Amer. Math. Soc., Providence, RI. MR1217253

WALSH, B. and BLOWS, M. W. (2009). Abundant genetic variation + strong selection = multivariate genetic constraints: A geometric view of adaptation. *Annu*. *Rev*. *Ecol*. *Evol*. *Syst*. **40** 41–59.

WIGNER, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Ann*. *of Math*. (2) **62** 548–564. MR0077805

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

YANG, M., GOLDSTEIN, H., BROWNE, W. and WOODHOUSE, G. (2002). Multivariate multilevel analyses of examination results. *J*. *Roy*. *Statist*. *Soc*. *Ser*. *A* **165** 137–153. MR1909740

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.

YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. *Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554

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

DEPARTMENT OF STATISTICS AND DATA SCIENCE
YALE UNIVERSITY
24 HILLHOUSE AVENUE
NEW HAVEN, CONNECTICUT 06511
USA
E-MAIL: zhou.fan@yale.edu

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
390 SERRA MALL
STANFORD, CALIFORNIA 94305
USA
E-MAIL: imj@stanford.edu