Submitted to the Annals of Statistics

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

By Zhou Fan and Iain M. Johnstone

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 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. Since the work of R. A. Fisher, random effects linear models have played a foundational role in quantitative genetics. Fisher described the decomposition of the variance of a quantitative trait in a population into components, which may be estimated by observing these traits in individuals of different relations [Fis18]. One important motivation for estimating these components is in predicting the evolutionary response of the population to natural or artificial selection. If an episode of selection changes the mean value of a trait in this generation by S, the change $\Delta\mu$ inherited by the next generation is predicted by the breeders' equation $\Delta\mu = \sigma_A^2(\sigma_z^2)^{-1}S$, where σ_z^2 is the total population variance and σ_A^2 is its additive genetic component. A common method of estimating σ_A^2 is using a random effects model with a suitable experimental design [LW98].

In reality, selection acting on a trait rarely only induces a response in that single trait, but instead also affects genetically correlated traits [LA83, PA89, Blo07]. 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 [Bar90, WB09, MCM$^+$14, BM15, BAC$^+$15]. Letting $S \in \mathbb{R}^p$ denote the changes in mean values of p traits in this generation due to selection, the changes inherited by the next generation are predicted by the multivariate breeders' equation $\Delta\mu = GP^{-1}S$, where $P \in \mathbb{R}^{p\times p}$ is the total phenotypic covariance of the traits and $G \in \mathbb{R}^{p\times p}$ is its additive genetic component [Lan79]. The response to selection may be understood via the principal eigenvectors of G and the alignment of the "selection gradient" $P^{-1}S$ with these eigenvectors. Hence, there is significant interest in understanding the spectral structure of G [Kir09, WB09, HMB14, BM15]. Analogously to the univariate setting, G may be estimated by variance components in multivariate random effects models.

Gene expression microarrays have enabled the measurements of thousands of quantitative phenotypic traits in a single experimental study, providing an opportunity to better understand the nature and extent of pleiotropy and the effective dimensionality of possible evolutionary response in the entire phenome of an organism [MCM+14, BAC+15]. However, the theory of large random matrices [PA14] as well as numerical simulations [BM15] both suggest that variance component matrices estimated in these settings may exhibit significant spectral noise due to their high dimensionality. In this work, we derive a characterization of the spectra of such estimates.

We study the general multivariate random effects model

$$Y = X\beta + \sum_{r=1}^{k} U_r\alpha_r, \qquad \alpha_r \sim \mathcal{N}(0, \mathrm{Id}_{I_r} \otimes \Sigma_r). \tag{1.1}$$

$Y \in \mathbb{R}^{n\times p}$ represents n observations of p traits, modeled as a sum of fixed effects $X\beta$ and k random effects $U_1\alpha_1, \ldots, U_k\alpha_k$. (It is common to add a residual error term ε; for notational convenience we incorporate ε by allowing $U_k = \mathrm{Id}$ and $\alpha_k = \varepsilon$.) $X \in \mathbb{R}^{n\times m}$ 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 r^{th} 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, i.e. estimators of the form

$$\hat{\Sigma}_r = Y^T B_r Y, \qquad (B_r X = 0) \tag{1.2}$$

for symmetric matrices $B_r \in \mathbb{R}^{n\times n}$. In particular, model (1.1) encompasses nested and crossed classification designs, and (1.2) encompasses MANOVA estimators and MINQUEs. We discuss examples in Section 2. We consider the asymptotic regime where $n, I_1, \ldots, I_k$ grow proportionally. For classification designs, this means that the number of groups at the highest level of division scales proportionally with n, and all further sub-divisions remain bounded in size. This is the relevant regime for experiments that estimate components of phenotypic covariance, from considerations of both experimental practicality and optimal design [Rob59a, Rob59b].

Our main result shows that when p is also comparable to n, the spectra of estimators (1.2) are accurately predicted by deterministic laws which depend on the true variance components $\Sigma_1, \ldots, \Sigma_k$. We characterize these laws by systems of fixed-point equations in their Stieltjes transforms, which generalize the Marcenko-Pastur equation for the usual sample covariance matrix [MP67]. These equations may be solved numerically to approximate the probability density functions of these laws.

For sample covariance matrices, the Marcenko-Pastur equation underpins many procedures for inferring the population spectrum [Kar08, Mes08, RMSE08, BCY10, LW12] and debiasing sample eigenvalues in "spiked" covariance models [BS06, Pau07, BGN11, BY12]. Similar inferential questions are of interest in variance components applications, and we hope that our result will enable the study of such questions.

1.1. *Main result.* To present an analogy, we review the Marcenko-Pastur equation for sample covariance matrices: Given $Y \in \mathbb{R}^{n \times p}$ consisting of n i.i.d. rows with distribution $\mathcal{N}(0, \Sigma)$, consider the sample covariance $\hat{\Sigma} = n^{-1} Y^T Y$. When $\Sigma = \mathrm{Id}$, the spectrum of $\hat{\Sigma}$ is well-approximated by the Marcenko-Pastur law [MP67, SB95]. More generally, for any Σ, the spectrum of $\hat{\Sigma}$ is predicted by the Marcenko-Pastur equation:

THEOREM 1.1 ([MP67, SB95]). *Let $\mu_{\hat{\Sigma}} = p^{-1} \sum_{i=1}^{p} \delta_{\lambda_i(\hat{\Sigma})}$ denote the empirical spectral measure of $\hat{\Sigma}$. 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 $m_0(z) \in \mathbb{C}^+$ satisfying*

$$m_0(z) = \frac{1}{p} \operatorname{Tr}\left[\left(\left(1 - \frac{p}{n} - \frac{p}{n} z m_0(z)\right)\Sigma - z \,\mathrm{Id}_p\right)^{-1}\right], \tag{1.3}$$

and $m_0 : \mathbb{C}^+ \to \mathbb{C}^+$ 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.

The Stieltjes transform m_0 determines μ_0 via the Stieltjes inversion formula.

Theorem 1.1 is usually stated in an alternative form, assuming convergence of p/n to $\gamma \in (0, \infty)$ and of the spectrum of Σ to a weak limit μ^*. In this case $\mu_{\hat{\Sigma}}$ converges to a fixed weak limit μ_0 depending on γ and μ^*. We have stated this theorem instead in a "deterministic equivalent" form [HLN07, CDS11], where μ_0 is defined by the finite-sample quantities p/n and Σ. This form is arguably more closely tied to applications, since one typically computes the analytic prediction for $\mu_{\hat{\Sigma}}$ directly from these finite-sample quantities, rather than first passing to an abstract limit. (See also the discussion in [Kar08].)

The main result of our paper is the following extension of Theorem 1.1 to the setting of model (1.1). 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

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

We state our result also in deterministic equivalent form, which avoids imposing "joint convergence" assumptions on $\Sigma_1, \ldots, \Sigma_k$:

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*

$$a_r = -\frac{1}{I_r} \operatorname{Tr}\left((z \operatorname{Id}_p + b \cdot \Sigma)^{-1}\Sigma_r\right), \tag{1.4}$$

$$b_r = -\frac{1}{I_r} \operatorname{Tr}_r\left([\operatorname{Id}_{I_+} + FD(a)]^{-1}F\right). \tag{1.5}$$

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

$$m_0(z) = -\frac{1}{p} \operatorname{Tr}\left((z \operatorname{Id}_p + b \cdot \Sigma)^{-1}\right) \tag{1.6}$$

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.

Note that μ_0 is a deterministic measure defined by $\Sigma_1, \ldots, \Sigma_k$ and the structure of the model, whereas $\mu_{\hat{\Sigma}}$ is random and depends on the data Y.

REMARK 1.3. When Y has n i.i.d. rows, the sample covariance $\hat{\Sigma} = n^{-1}Y^T Y$ corresponds to the special case of (1.1) with $k = 1$, $U_1 = \operatorname{Id}$, $\Sigma_1 = \Sigma$, and $B = n^{-1} \operatorname{Id}_n$. In this case, equations (1.4–1.6) reduce to

$$a_1 = -\frac{1}{n} \operatorname{Tr}\left((z \operatorname{Id}_p + b_1\Sigma)^{-1}\Sigma\right), \qquad b_1 = -1/(1 + a_1), \tag{1.7}$$

$$m_0(z) = -\tfrac{1}{p} \operatorname{Tr}\left((z \operatorname{Id}_p + b_1 \Sigma)^{-1}\right), \tag{1.8}$$

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

$$\begin{aligned} -1 - \frac{1}{b_1} = a_1 &= -\frac{z}{nb_1} \operatorname{Tr}\left((z \operatorname{Id}_p)^{-1} - (z \operatorname{Id}_p + b_1 \Sigma)^{-1}\right) \\ &= -\frac{p}{nb_1} + \frac{pzm_0(z)}{nb_1}. \end{aligned}$$

Hence $b_1 = -1 + (p/n) + (p/n)zm_0(z)$. Together with the above expression for $m_0(z)$, this recovers the Marcenko-Pastur equation (1.3).

In most cases, (1.4–1.6) 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 1.4. *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 (1.4) and $b_r^{(t+1)}$ from $a_r^{(t)}$ using (1.5).*

This yields a method for computing the density of μ_0 in Theorem 1.2: By the Stieltjes inversion formula, the density at $x \in \mathbb{R}$ is approximately $\pi^{-1}\Im m_0(x + i\varepsilon)$ for small ε, which we may compute from $b_1, \ldots, b_k$ using the above procedure. A software implementation is available upon request.

Theorems 1.2 and 1.4 are inspired by the study of similar models for wireless communication channels. In particular, [CDS11] 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 [Lix06, 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 [DL11] and the earlier work of [MS07] 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 [DL11] 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.

1.2. *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 [BG09] and the free deterministic equivalents approach of [SV12].

Let us write α_r in (1.1) 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

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

Hence we may equivalently consider the matrix

$$W = \sum_{r,s=1}^{k} H_r^T G_r^T F_{rs} G_s H_s \tag{1.9}$$

for $O_0, \ldots, O_{2k}$ 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 [BG09], 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 sub-algebra of projections in $\mathcal{A}$. We follow the deterministic equivalents approach of [SV12] 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 the equations (1.4–1.6).

We establish separately the existence and uniqueness of the fixed point to (1.4–1.5) using a contractive mapping argument and uniqueness of analytic continuation. This implies that the Stieltjes transform of w in step 2 is uniquely determined by (1.4–1.6), which implies by step 1 that (1.4–1.6) 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 arising in statistics and engineering, and it follows a line of work establishing asymptotic freeness of random matrices [Voi91, Dyk95, Voi98, HP00, Col03, CŚ06, BG09, SV12]. In the computation in step 2, the Stieltjes transform of w is exactly (rather than approximately) described by (1.4–1.6). The computation is thus entirely algebraic, using free cumulant tools of [NSS02, SV12], and it does not require analytic approximation arguments or bounds.

1.3. *Outline of paper.* Section 2 specializes Theorem 1.2 to several classification designs that arise in applications. 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.

1.4. *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. $\operatorname{diag}(A_1, \ldots, A_k)$ denotes the block-diagonal matrix with blocks $A_1, \ldots, A_k$. $\mathbb{C}^+ = \{z \in \mathbb{C} : \Im z > 0\}$ 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 sub-algebra, i.e. the ultraweak closure of $\langle\{a_i\}\rangle$, and $\|a_i\|$ denotes the C^*-norm.

Acknowledgments. We thank Mark Blows for introducing us to this problem and for much help in guiding us through understanding the quantitative genetics applications.

2. Specialization to classification designs. The form (1.2) encompasses MANOVA estimators, which solve for $\Sigma_1, \ldots, \Sigma_k$ in the system of equations $Y^T M_r Y = \mathbb{E}[Y^T M_r Y]$ for a certain choice of symmetric matrices $M_1, \ldots, M_k \in \mathbb{R}^{n\times n}$ [SCM09, Chapter 5.2]. From (1.1), the identity

$\mathbb{E}[\alpha_s^T M \alpha_s] = (\mathrm{Tr}\, M)\Sigma_s$ for any matrix M, and independence of α_r, we get

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

Hence each MANOVA estimate $\hat{\Sigma}_r$ takes the form (1.2), where B_r is a linear combination of $M_1, \ldots, M_k$.

In balanced or fully-nested classification designs, standard choices for $M_1, \ldots, M_k$ project onto subspaces of $\mathbb{R}^n$ such that each $Y^T M_r Y$ corresponds to a "sum-of-squares". We may simplify (1.5) in such settings by analytically computing the matrix inverse and block trace. We provide several examples below, deferring matrix algebra details and a more general procedure for obtaining such simplifications to Appendix A.

For more general designs and models, $M_1, \ldots, M_k$ may be ad-hoc, although Theorem 1.2 still applies to such estimators. The theorem also applies to MINQUEs [Rao72, LaM73] in these settings, which prescribe a specific form for $B \in \mathbb{R}^{n \times n}$ based on a variance minimization criterion.

2.1. *One-way classification.* $\{Y_{i,j} \in \mathbb{R}^p : 1 \le i \le I, 1 \le j \le J_i\}$ represent observations of p traits across $n = \sum_{i=1}^{I} J_i$ samples, belonging to I groups of sizes $J_1, \ldots, J_I$. The data are modeled as

$$Y_{i,j} = \mu + \alpha_i + \varepsilon_{i,j}, \tag{2.1}$$

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 [LW98].

Defining the sums-of-squares

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

where $\bar{Y}_i \in \mathbb{R}^p$ and $\bar{Y} \in \mathbb{R}^p$ denote the mean in the i^{th} group and of all samples, respectively, the standard MANOVA estimators are given [SCM09, Chapter 3.6] by

$$\hat{\Sigma}_1 = \frac{1}{K}\left(\frac{1}{I-1}\mathrm{SS}_1 - \frac{1}{n-I}\mathrm{SS}_2\right), \quad \hat{\Sigma}_2 = \frac{1}{n-I}\mathrm{SS}_2, \tag{2.2}$$

where $K = (n - \frac{1}{n}\sum_{i=1}^{I} J_i^2)/(I-1)$. The balanced case corresponds to $J_1 = \ldots = J_I = K$. Theorem 1.2 yields the following corollary:

COROLLARY 2.1. *Assume $p, n, I \to \infty$ such that $c < p/n < C$, $I/n > c$, $(n-I)/n > c$, $\max_{i=1}^I J_i < C$, $\|\Sigma_1\| < C$, 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

$$a_s = -\tfrac{1}{I_s} \operatorname{Tr}\left((z \operatorname{Id} + b_1\Sigma_1 + b_2\Sigma_2)^{-1}\Sigma_s\right) \qquad \text{for } s = 1, 2,$$
$$b_1 = -\textstyle\sum_{i=1}^I \frac{J_i}{KI+IJ_ia_1+na_2}, \qquad b_2 = \frac{n-I}{K(n-I)-na_2} - \sum_{i=1}^I \frac{1}{KI+IJ_ia_1+na_2},$$
$$m_0(z) = -\tfrac{1}{p} \operatorname{Tr}\left((z \operatorname{Id} + b_1\Sigma_1 + b_2\Sigma_2)^{-1}\right).$$

(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

$$a_2 = -\tfrac{1}{n} \operatorname{Tr}\left((z \operatorname{Id} + b_2\Sigma_2)^{-1}\Sigma_2\right), \qquad b_2 = -\tfrac{n-I}{n-I+na_2},$$
$$m_0(z) = -\tfrac{1}{p} \operatorname{Tr}\left((z \operatorname{Id} + b_2\Sigma_2)^{-1}\right).$$

"Determined by" is in the sense of Theorem 1.2, i.e. for each $z \in \mathbb{C}^+$ there exists a unique solution to these equations with $a_s \in \mathbb{C}^+ \cup \{0\}$, $b_s \in \overline{\mathbb{C}^+}$, and $m_0(z) \in \mathbb{C}^+$. This system may be solved by the procedure of Theorem 1.4.

Figure 1 displays the simulated spectrum of $\hat{\Sigma}_1$ in various settings. This spectrum depends on both Σ_1 and Σ_2. Overlaid on each histogram is the density of μ_0, approximated as $f(x) = \pi^{-1}\Im m_0(x + 0.0001i)$ and computed using the procedure of Theorem 1.4.

For $\hat{\Sigma}_2$ (but not $\hat{\Sigma}_1$), as in Remark 1.3, 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 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.

2.2. *Balanced nested classification.* $\{Y_{j_1,\ldots,j_k} \in \mathbb{R}^p : 1 \le j_1 \le J_1, \ldots, 1 \le j_k \le J_k\}$ are observations of p traits across $n = J_1 J_2 \ldots J_k$ samples. The samples are divided into $J_1 \ge 2$ groups of equal size $J_2 \ldots J_k$, the samples within each group are further divided into $J_2 \ge 2$ subgroups of equal size $J_3 \ldots J_k$, etc., and there are $J_k \ge 2$ samples in each subgroup at the finest level of division. The data are modeled as

$$Y_{j_1,\ldots,j_k} = \mu + \alpha^{(1)}_{j_1} + \alpha^{(2)}_{j_1,j_2} + \ldots + \alpha^{(k-1)}_{j_1,\ldots,j_{k-1}} + \varepsilon_{j_1,\ldots,j_k}, \tag{2.3}$$

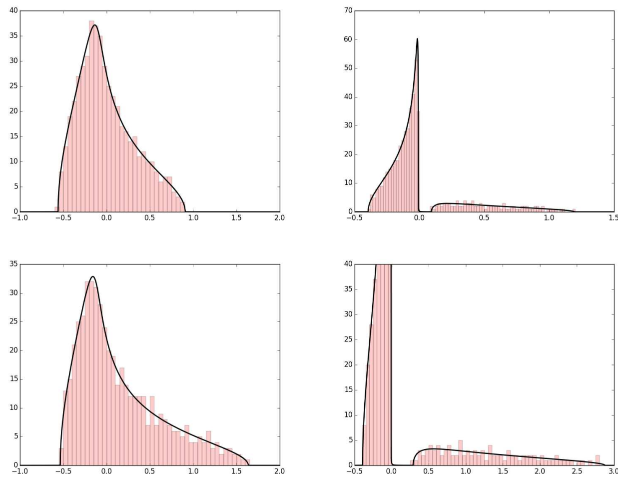

FIG 1. *Simulated spectrum of* $\hat{\Sigma}_1$ *for the balanced one-way classification model,* $p = 500$*, with theoretical predictions of Corollary 2.1 overlaid in black. Left: 400 groups of size 4. Right: 100 groups of size 8. Top:* $\Sigma_1 = 0$*,* $\Sigma_2 = \mathrm{Id}$*. Bottom:* Σ_1 *with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

where $\mu \in \mathbb{R}^p$ is the population mean, $\alpha^{(r)}_{j_1,\ldots,j_r} \sim \mathcal{N}(0, \Sigma_r)$ are i.i.d. group effects for the r^{th} level of grouping, and $\varepsilon_{j_1,\ldots,j_k} \sim \mathcal{N}(0, \Sigma_k)$ are i.i.d. residual errors. The case $k = 2$ is the one-way classification model of Section 2.1 when the design is balanced. The two-way model ($k = 3$) is the model for the full-sib half-sib design in which outer groups correspond to half-siblings and inner groups to full siblings. It is also the model for the monozygotic-twin half-sib design, in which outer groups correspond to offspring of one of two twins, and inner groups to offspring of one twin in the pair [LW98].

Sums-of-squares and MANOVA estimators $\hat{\Sigma}_r$ for Σ_r are defined analogously to the one-way model of Section 2.1; we review these definitions in Appendix A. Theorem 1.2 yields the following corollary for these estimators:

COROLLARY 2.2. *Fix* $J_2, \ldots, J_k \geq 2$*, let* $n = J_1 J_2 \ldots J_k$*, and assume* $p, n, J_1 \to \infty$ *such that* $c < p/n < C$ *and* $\|\Sigma_r\| < C$ *for all* $r = 1, \ldots, k$ *and some* $C, c > 0$*. Then for any* $r \in \{1, \ldots, k\}$*,* $\mu_{\hat{\Sigma}_r} - \mu_0 \to 0$ *weakly a.s. where* μ_0 *has Stieltjes transform* $m_0(z)$ *determined by*

$$a_s = -\tfrac{1}{J_1 \ldots J_s} \operatorname{Tr}\left((z\,\mathrm{Id} + b_r \Sigma_r + \ldots + b_k \Sigma_k)^{-1} \Sigma_s\right) \quad \textit{for } s = r, \ldots, k,$$

$$b_s = \begin{cases} -\frac{J_r-1}{J_r-1+J_r\sum_{j=r}^k a_j} & \textit{if } s=r, \\ -\frac{1}{J_{r+1}\dots J_s}\left(\frac{J_r-1}{J_r-1+J_r\sum_{j=r}^k a_j} - \frac{J_{r+1}-1}{J_{r+1}-1-\sum_{j=r+1}^k a_j}\right) & \textit{if } s\geq r+1, \end{cases}$$
$$m_0(z) = -\tfrac{1}{p}\operatorname{Tr}\left((z\operatorname{Id}+b_r\Sigma_r+\ldots+b_k\Sigma_k)^{-1}\right).$$

2.3. *Replicated crossed two-way classification.* $\{Y_{i,j,k,l}\in\mathbb{R}^p : 1\leq i\leq I, 1\leq j\leq J, 1\leq k\leq K, 1\leq l\leq L\}$ represent observations across $n=IJKL$ samples. The samples belong to I replicated experiments of a $J\times K$ crossed design with fixed numbers J and K of levels for two factors, and with L samples in each replicate i corresponding to each level cross $j\times k$. The data are modeled as

$$Y_{i,j,k,l} = \mu+\alpha_i+\beta_{i,j}+\gamma_{i,k}+\delta_{i,j,k}+\varepsilon_{i,j,k,l}, \tag{2.4}$$

where $\mu\in\mathbb{R}^p$ is the population mean, $\alpha_i\sim\mathcal{N}(0,\Sigma_1)$ are replicate effects, $\beta_{i,j}\sim\mathcal{N}(0,\Sigma_2)$ and $\gamma_{i,k}\sim\mathcal{N}(0,\Sigma_3)$ are effects for the two factors, $\delta_{i,j,k}\sim\mathcal{N}(0,\Sigma_4)$ are effects for the factor interactions, and $\varepsilon_{i,j,k,l}\sim\mathcal{N}(0,\Sigma_5)$ are residual errors. This crossed $J\times K$ design corresponds to the Comstock-Robinson model or North Carolina Design II commonly used in plant studies, in which each of J males is mated to each of K females. We consider the replicated setting with small J,K,L and large I, as is often done in practice for reasons of experimental design [LW98].

Definitions of MANOVA estimators $\hat{\Sigma}_1,\ldots,\hat{\Sigma}_5$ are reviewed in Appendix A. Theorem 1.2 yields the following for, e.g., the factor effect estimate $\hat{\Sigma}_2$.

Corollary 2.3. *Fix $J,K,L\geq 2$, let $n=IJKL$, and assume $p,n,I\to\infty$ such that $c<p/n<C$ and $\|\Sigma_r\|\leq C$ for each $r=1,\ldots,5$ and some $C,c>0$. Denote $I_2=IJ$, $I_4=IJK$, and $I_5=n$. Then $\mu_{\hat{\Sigma}_2}-\mu_0\to 0$ weakly a.s. where μ_0 has Stieltjes transform $m_0(z)$ determined by*

$$\begin{aligned} a_s &= -\tfrac{1}{I_s}\operatorname{Tr}\left((z\operatorname{Id}+b_2\Sigma_2+b_4\Sigma_4+b_5\Sigma_5)^{-1}\Sigma_s\right) \qquad \textit{for } s=2,4,5, \\ b_2 &= -\tfrac{J-1}{J-1+J(a_2+a_4+a_5)}, \\ b_4 &= -\tfrac{1}{K}\left(\tfrac{J-1}{J-1+J(a_2+a_4+a_5)} - \tfrac{(J-1)(K-1)}{(J-1)(K-1)-J(a_4+a_5)}\right), \\ b_5 &= \tfrac{1}{L}b_4, \\ m_0(z) &= -\tfrac{1}{p}\operatorname{Tr}\left((z\operatorname{Id}+b_2\Sigma_2+b_4\Sigma_4+b_5\Sigma_5)^{-1}\right). \end{aligned}$$

Appendix A discusses how to obtain analogous results for $\hat{\Sigma}_1,\hat{\Sigma}_3,\hat{\Sigma}_4,\hat{\Sigma}_5$.

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 [VDN92, Voi95, BG09].

Definition. A **non-commutative 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,

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

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 [BG09], we embed rectangular matrices into a larger square space according to the following structure.

Definition. Let $(\mathcal{A}, \tau)$ be a non-commutative 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 non-zero pairwise-orthogonal projections summing to 1, i.e. for all $r \neq s \in \{1, \ldots, d\}$,

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

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 $N = N_1 + \ldots + N_d$. 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 non-commutative 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 $a \in \mathcal{A}$,

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

We identify $\mathbb{C} \subset \mathcal{A}$ as a sub-algebra 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 non-commutative 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 non-commutative probability space and $\mathbf{F}^{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$ a conditional expectation onto a sub-algebra $\mathcal{B} \subset \mathcal{A}$. $\mathbf{F}^{\mathcal{B}}$ 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 sub-algebra 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 sub-algebras 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 sub-algebra 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 $b_1 y_1 b_2 y_2 \dots b_{l-1} y_{l-1} b_l$ where $l \geq 1$, $b_1, \dots, b_l \in \mathcal{B}$, and $y_1, \dots, y_{l-1} \in \{x_i, x_i^* : i \in \mathcal{I}\}$. 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. 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 $\mathcal{B}$

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

In the scalar setting where $\mathcal{B} = \mathbb{C}$ and $\mathbf{F}^{\mathcal{B}} = \tau$, a $*$-monomial takes the simpler form $z y_1 y_2 \dots y_{l-1}$ 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 [Voi95], 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:

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

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 3.2. *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 [Voi95, Proposition 1.3]. □

3.2. *Free deterministic equivalents and asymptotic freeness.* Free deterministic equivalents were introduced in [SV12]. 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 3.3. 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 $N \to \infty$,

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

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 3.3 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 3.1.) Freeness relations in the deterministic model will emerge from the following notion of rotational invariance of the random matrices.

Definition 3.4. 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 3.5. 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 semi-circle 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 3.3, we may verify $N_r^{-1}\operatorname{Tr}_r Q(G)-\tau_r(q(g))\to 0$ a.s. by the classical Wigner semi-circle theorem [Wig55]. Then G and g are asymptotically equal in $\mathcal{D}$-law a.s. Furthermore, G is block-orthogonally invariant.

Example 3.6. Fix $r_1\neq r_2\in\{1,\ldots,d\}$ and let $G\in\mathbb{C}^{N\times N}$ be a simple random matrix such that the block $G_{r_1r_2}$ has i.i.d. Gaussian or complex Gaussian entries with variance $1/N_{r_1}$. Let $(\mathcal{A},\tau,p_1,\ldots,p_d)$ 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:

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

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

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

For any corresponding $*$-polynomials Q and q as in Definition 3.3, 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 [MP67]. Then G and g are asymptotically equal in $\mathcal{D}$-law a.s., and G is block-orthogonally invariant.

EXAMPLE 3.7. Let $B_1, \ldots, B_k \in \mathbb{C}^{N \times N}$ be deterministic simple matrices, say with $P_{r_i} B_i P_{s_i} = B_i$ for each $i = 1, \ldots, k$ and $r_i, s_i \in \{1, \ldots, d\}$. Let $O_1 \in \mathbb{R}^{N_1 \times N_1}, \ldots, O_d \in \mathbb{R}^{N_d \times N_d}$ be independent Haar-distributed orthogonal matrices, define $O = \operatorname{diag}(O_1, \ldots, O_d) \in \mathbb{R}^{N \times N}$, and let $\check{B}_i = O^T B_i O$. Let $(\mathcal{A}, \tau, p_1, \ldots, p_d)$ satisfy $\tau(p_s) = N_s/N$ for each s, such that $\mathcal{A}$ contains simple elements $b_1, \ldots, b_k$ satisfying $p_{r_i} b_i p_{s_i} = b_i$ for each $i = 1, \ldots, k$, and

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

for any corresponding $*$-polynomials Q and q with coefficients in $\langle P_1, \ldots, P_d \rangle$ and $\langle p_1, \ldots, p_d \rangle$. As $\operatorname{Tr}_r Q(B_1, \ldots, B_k)$ 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 $(\check{B}_i)_{i \in \{1, \ldots, k\}}$ and $(b_i)_{i \in \{1, \ldots, k\}}$ are exactly (and hence also asymptotically) equal in $\mathcal{D}$-law, and $(\check{B}_i)_{i \in \{1, \ldots, k\}}$ 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 3.5, 3.6, and 3.7 above, and consider a combined model in which these families are $\mathcal{D}$-free:

DEFINITION 3.8. Consider $(\mathbb{C}^{N \times N}, N^{-1} \operatorname{Tr}, P_1, \ldots, P_d)$ as in Example 3.1. Suppose $(H_i)_{i \in \mathcal{I}_1}, \ldots, (H_i)_{i \in \mathcal{I}_J}$ are finite families of random matrices in $\mathbb{C}^{N \times N}$ such that:

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

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

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

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

THEOREM 3.9. *In the space $(\mathbb{C}^{N \times N}, N^{-1} \operatorname{Tr}, P_1, \ldots, P_d)$ of Example 3.1, suppose $(H_i)_{i \in \mathcal{I}_1}, \ldots, (H_i)_{i \in \mathcal{I}_J}$ are independent, block-orthogonally invariant* *families of random matrices, and let* $(h_i)_{i\in\mathcal{I}_1},\ldots,(h_i)_{i\in\mathcal{I}_J}$ *be any free deterministic equivalent in* $(\mathcal{A},\tau,p_1,\ldots,p_d)$*. 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* $(H_i)_{i\in\mathcal{I}_j,j\in\{1,\ldots,J\}}$ *and* $(h_i)_{i\in\mathcal{I}_j,j\in\{1,\ldots,J\}}$ *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 3.9 is analogous to [BG09, Theorem 1.6] and [SV12, 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 3.9 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

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

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

Corollary 3.10. *Under the assumptions of Theorem 3.9, let* Q *be a self-adjoint* $*$*-polynomial (with* $\mathbb{C}$*-valued coefficients) in* $(x_i)_{i\in\mathcal{I}_j,j\in\{1,\ldots,J\}}$*, and let*

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

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 3.9 and Corollary 3.10 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 [Spe98, NSS02, SV12]. 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 $l \geq 1$, the l^{th} 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 non-crossing partition lattice; we refer the reader to [SV12] and [Spe98, Chapters 2 and 3] for details. We will use the properties that $\kappa_l^{\mathcal{B}}$ is linear in each argument and satisfies the relations

$$\kappa_l^{\mathcal{B}}(ba_1, a_2, \ldots, a_{l-1}, a_l b') = b\kappa_l^{\mathcal{B}}(a_1, \ldots, a_l)b', \tag{3.9}$$

$$\kappa_l^{\mathcal{B}}(a_1, \ldots, a_{j-1}, a_j b, a_{j+1}, \ldots, a_l) = \kappa_l^{\mathcal{B}}(a_1, \ldots, a_j, ba_{j+1}, \ldots, a_l) \tag{3.10}$$

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

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

$$\mathcal{R}_a^{\mathcal{B}}(b) := \sum_{l \geq 1} \kappa_l^{\mathcal{B}}(ab, \ldots, ab, a). \tag{3.11}$$

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

$$G_a^{\mathcal{B}}(b) := \mathbf{F}^{\mathcal{B}}((b-a)^{-1}) = \sum_{l \geq 0} \mathbf{F}^{\mathcal{B}}(b^{-1}(ab^{-1})^l), \tag{3.12}$$

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

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

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

PROOF. See [Voi95, Theorem 4.9] and also [Spe98, Theorem 4.1.12]. □

REMARK. When $\mathcal{A}$ is a von Neumann algebra, the right 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

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

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

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

(see [NS06, Proposition 13.15]), so (3.11) defines a convergent series in $\mathcal{B}$ when $16\|b\| < 1/\|a\|$, with

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

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 sub-algebras of $\mathcal{A}$. Central to this approach is the following result from [NSS02] (see also [SV12]):

Proposition 3.12. *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 $l \geq 1$, $h_1, \ldots, h_l \in \mathcal{H}$, and $b_1, \ldots, b_{l-1} \in \mathcal{B}$,*

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

Proof. See [NSS02, Theorem 3.6]. $\square$

For sub-algebras $\mathcal{D} \subseteq \mathcal{B} \subseteq \mathcal{A}$ 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 1.4 for a slightly more general matrix model: Fix $k \geq 1$, let $p, n_1, \ldots, n_k, m_1, \ldots, m_k \in \mathbb{N}$, 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 $r = 1, \ldots, k$, 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 $(G_r)_{ij} \overset{iid}{\sim} \mathcal{N}(0, n_r^{-1})$ or $G_r \in \mathbb{C}^{n_r \times m_r}$ with $\Im(G_r)_{ij}, \Re(G_r)_{ij} \overset{iid}{\sim} \mathcal{N}(0, (2n_r)^{-1})$. 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 $p, n_1, \ldots, n_k, m_1, \ldots, m_k \to \infty$, such that $c < n_r/p < C$, $c < m_r/p < C$, $\|H_r\| < C$, and $\|F_{rs}\| < C$ for all $r, s = 1, \ldots, k$ and some constants $C, c > 0$. Then:*

(a) For each $z \in \mathbb{C}^+$, there exist unique 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

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

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

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

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

(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 1.4.

Theorems 1.2 and 1.4 follow by specializing this result to $F = U^T B U$ and $m_r = p$, $n_r = I_r$, and $H_r = \Sigma_r^{1/2}$ for each $r = 1, \ldots, k$.

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 1.2.

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 $O_0, \ldots, O_{2k}$ of the appropriate sizes. As in Section 1.2, 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			$F_{1,1}$	$\cdots$	$F_{1,k}$
		$\ddots$		$\vdots$	$\ddots$	$\vdots$
			G_k	$F_{k,1}$	$\cdots$	$F_{k,k}$

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

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

1. $p_0, \ldots, p_{2k}$ have traces
$$\tau(p_0) = p/N,\ \tau(p_r) = m_r/N,\ \tau(p_{k+r}) = n_r/N \qquad \text{for all } r = 1, \ldots, k.$$
2. f_{rs}, g_r, h_r are simple elements such that for all $r, s \in \{1, \ldots, k\}$,
$$p_{k+r} f_{rs} p_{k+s} = f_{rs}, \qquad p_{k+r} g_r p_r = g_r, \qquad p_r h_r p_0 = h_r.$$
3. $\{f_{rs} : 1 \le r, s \le k\}$ has the same joint $\mathcal{D}$-law as $\{\tilde{F}_{r,s} : 1 \le r, s \le k\}$, and $\{h_r : 1 \le r \le k\}$ has the same joint $\mathcal{D}$-law as $\{\tilde{H}_r : 1 \le r \le k\}$. I.e.,

for any $r \in \{0, \dots, 2k\}$ and any non-commutative $*$-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 $\langle p_0, \dots, p_{2k}\rangle$,

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

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

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

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

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 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 $\tilde{F}_{r,s}, \tilde{G}_r, \tilde{H}_r \in \mathbb{C}^{N \times N}$ in the sense of Definition 3.8.

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 sub-algebras $\langle \mathcal{D}, \{f_{rs}\}\rangle_{W^*}$, $\langle \mathcal{D}, \{h_r\}\rangle_{W^*}$, $\langle \mathcal{D}, g_1\rangle_{W^*}$, ..., $\langle \mathcal{D}, g_k\rangle_{W^*}$ 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:

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 sub-algebras 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*.*

PROOF. We use the computational idea of [SV12]: Denote by $\kappa_l^{\mathcal{C}}$ and $\kappa_l^{\mathcal{D}}$ the $\mathcal{C}$-valued and $\mathcal{D}$-valued free cumulants. For $l \ge 1$ and $e \in \mathcal{C}$,

$$\begin{aligned}
&\kappa_l^{\mathcal{C}}(ae, \dots, ae, a) \\
&= \kappa_l^{\mathcal{C}} \left(\sum_{r,s=1}^k c_r^* b_{rs} c_s e,\ \dots,\ \sum_{r,s=1}^k c_r^* b_{rs} c_s e, \sum_{r,s=1}^k c_r^* b_{rs} c_s \right) \\
&= \sum_{r_1, s_1, \dots, r_l, s_l = 1}^k \kappa_l^{\mathcal{C}}(c_{r_1}^* b_{r_1 s_1} c_{s_1} e,\ \dots,\ c_{r_{l-1}}^* b_{r_{l-1} s_{l-1}} c_{s_{l-1}} e,\ c_{r_l}^* b_{r_l s_l} c_{s_l}) \\
&= \sum_{r_1, s_1, \dots, r_l, s_l = 1}^k c_{r_1}^* \kappa_l^{\mathcal{C}}(b_{r_1 s_1} c_{s_1} e c_{r_2}^*,\ \dots,\ b_{r_{l-1} s_{l-1}} c_{s_{l-1}} e c_{r_l}^*,\ b_{r_l s_l})\, c_{s_l} \\
&= \sum_{r_1, s_1, \dots, r_l, s_l = 1}^k c_{r_1}^* \kappa_l^{\mathcal{D}}(b_{r_1 s_1} \mathbf{F}^{\mathcal{D}}(c_{s_1} e c_{r_2}^*),\ \dots,\ b_{r_{l-1} s_{l-1}} \mathbf{F}^{\mathcal{D}}(c_{s_{l-1}} e c_{r_l}^*),\ b_{r_l s_l})\, c_{s_l},
\end{aligned}$$

where we applied the definition of a, multi-linearity of $\kappa_l^{\mathcal{C}}$, the identities (3.9) and (3.10), and Proposition 3.12 using freeness of $\mathcal{B}$ and $\mathcal{C}$ over $\mathcal{D}$.

By the identity $c_r = q_r c_r$, each $c_s e c_r^*$ is simple, and we have from (3.2)

$$\mathbf{F}^{\mathcal{D}}(c_s e c_r^*) = \begin{cases} 0 & \text{if } s \ne r \\ \tau_s(c_s e c_s^*) q_s & \text{if } s = r. \end{cases}$$

Furthermore, for any $d \in \mathcal{D}$, as $d = \tau_0(d) q_0 + \ldots + \tau_k(d) q_k$, we have $c_r^* d c_s = c_r^* c_r \tau_r(d)$ if $r = s$ and 0 otherwise. Hence we may restrict the above sum to $s_1 = r_2,\ s_2 = r_3,\ \dots,\ s_{l-1} = r_l,\ s_l = r_1$. Then, setting

$$d = \sum_{r=1}^k \tau_r(c_r e c_r^*) q_r \tag{4.7}$$

and applying the identity $q_r b_{rs} q_s = b_{rs}$,
(4.8)
$$\kappa_l^{\mathcal{C}}(ae,\ldots,ae,a) = \sum_{r_1,\ldots,r_l=1}^{k} c_{r_1}^* c_{r_1} \tau_{r_1}\left(\kappa_l^{\mathcal{D}}(b_{r_1r_2}d,\ \ldots,\ b_{r_{l-1}r_l}d,\ b_{r_lr_1})\right).$$

On the other hand, similar arguments yield

$$\begin{aligned}
&\kappa_l^{\mathcal{D}}(bd,\ldots,bd,b)\\
&= \sum_{r_1,s_1,\ldots,r_l,s_l=1}^{k} \kappa_l^{\mathcal{D}}(b_{r_1s_1}d,\ \ldots,\ b_{r_{l-1}s_{l-1}}d,\ b_{r_ls_l})\\
&= \sum_{r_1,s_1,\ldots,r_l,s_l=1}^{k} q_{r_1}\kappa_l^{\mathcal{D}}(b_{r_1s_1}q_{s_1}dq_{r_2},\ \ldots,\ b_{r_{l-1}s_{l-1}}q_{s_{l-1}}dq_{r_l},\ b_{r_ls_l})\, q_{s_l}\\
&= \sum_{r_1,\ldots,r_l=1}^{k} q_{r_1}\kappa_l^{\mathcal{D}}(b_{r_1r_2}d,\ \ldots,\ b_{r_{l-1}r_l}d,\ b_{r_lr_1}).
\end{aligned}$$

Comparing with (4.8), $\kappa_l^{\mathcal{C}}(ae,\ldots,ae,a) = \sum_{r=1}^{k} c_r^* c_r \tau_r\left(\kappa_l^{\mathcal{D}}(bd,\ldots,bd,b)\right)$. Summing over l and recalling (3.11), for $\|e\|$ sufficiently small,

$$\mathcal{R}_a^{\mathcal{C}}(e) = \sum_{l\geq 1}\sum_{r=1}^{k} c_r^* c_r \tau_r\left(\kappa_l^{\mathcal{D}}(bd,\ldots,bd,b)\right).$$

Noting that $\|d\| \leq \sum_{s=1}^{k} \|c_s\|^2 \|e\|$ and applying (3.16), we may exchange the order of summations on the right and move the summation over l inside τ_r by linearity and norm-continuity of τ, yielding the desired result. □

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

LEMMA 4.4. *Under the conditions of Theorem 4.1, let* $(\mathcal{A},\tau,p_0,\ldots,p_{2k})$ *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* $a_1,\ldots,a_k : \mathbb{D} \to \mathbb{C}^+ \cup \{0\}$ *and* $b_1,\ldots,b_k : \mathbb{D} \to \mathbb{C}$ *that satisfy, for every* $z \in \mathbb{D}$ *and for* $m_0(z) = \tau_0((w-z)^{-1})$, *the 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 sub-algebras $\mathcal{D} = \langle p_r : 0 \leq r \leq 2k\rangle$, $\mathcal{F} = \langle \mathcal{D}, \{f_{rs}\}\rangle_{W^*}$, $\mathcal{G} = \langle \mathcal{D}, \{g_r\}\rangle_{W^*}$, 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 $r = 1, \ldots, k$, $\mathcal{C} = \mathcal{H}$, and $\mathcal{B} = \langle \mathcal{F}, \mathcal{G} \rangle$. Then for $c \in \mathcal{H}$,

$$\mathcal{R}_w^{\mathcal{H}}(c) = \sum_{r=1}^{k} h_r^* h_r \tau_r \Big(\mathcal{R}_v^{\mathcal{D}} \Big(\sum_{s=1}^{k} p_s \tau_s (h_s c h_s^*) \Big) \Big). \tag{4.9}$$

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

$$\alpha_r(d) := \tau_r \left(h_r G_w^{\mathcal{H}}(d) h_r^* \right), \tag{4.10}$$

$$\beta_r(d) := \tau_r \Big(\mathcal{R}_v^{\mathcal{D}} \Big(\sum_{s=1}^{k} p_s \alpha_s(d) \Big) \Big). \tag{4.11}$$

Then (3.14) and (4.9) with $c = G_w^{\mathcal{H}}(d)$ imply

$$G_w^{\mathcal{H}}(d) = \left(d - \mathcal{R}_w^{\mathcal{H}} \left(G_w^{\mathcal{H}}(d) \right) \right)^{-1} = \Big(d - \sum_{r=1}^{k} h_r^* h_r \beta_r(d) \Big)^{-1}. \tag{4.12}$$

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

$$G_w^{\mathcal{D}}(d) = \mathbf{F}^{\mathcal{D}} \Big(\Big(d - \sum_{r=1}^{k} h_r^* h_r \beta_r(d) \Big)^{-1} \Big). \tag{4.13}$$

Applying (4.12) to (4.10),

$$\alpha_r(d) = \tau_r \Big(h_r \Big(d - \sum_{s=1}^{k} h_s^* h_s \beta_s(d) \Big)^{-1} h_r^* \Big). \tag{4.14}$$

Noting that $(p_1 + \ldots + p_k) v (p_1 + \ldots + p_k) = v$, (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 (4.11) as

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

For $r = 0$ and $r \in \{k+1, \dots, 2k\}$, 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 $r \in \{k+1, \dots, 2k\}$, applying (3.11) and multi-linearity of $\kappa_l^{\mathcal{D}}$, we may rewrite the above as

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

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

$$G_v^{\mathcal{D}}\left(\sum_{r=0}^{2k} p_r\Big(\frac{1}{\alpha_r(d)} + \beta_r(d)\Big)\right) = \sum_{r=0}^{2k} p_r \alpha_r(d). \tag{4.15}$$

The relation between $G_w^{\mathcal{D}}$ and $G_v^{\mathcal{D}}$ is given by (4.13), (4.14), and (4.15).

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 $r = 1, \dots, k$, $\mathcal{C} = \mathcal{G}$, and $\mathcal{B} = \mathcal{F}$. Then for $c \in \mathcal{G}$,

$$\mathcal{R}_v^{\mathcal{G}}(c) = \sum_{r=1}^{k} g_r^* g_r \tau_{r+k}\left(\mathcal{R}_u^{\mathcal{D}}\Big(\sum_{s=1}^{k} p_{s+k}\tau_{s+k}(g_s c g_s^*)\Big)\right). \tag{4.16}$$

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

$$\gamma_{r+k}(d) = \tau_{r+k}(g_r G_v^{\mathcal{G}}(d) g_r^*), \tag{4.17}$$

$$\delta_{r+k}(d) = \tau_{r+k}\left(\mathcal{R}_u^{\mathcal{D}}\Big(\sum_{s=1}^{k} p_{s+k}\gamma_{s+k}(d)\Big)\right). \tag{4.18}$$

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 $(p_{k+1} + \dots + p_{2k})u(p_{k+1} + \dots + p_{2k}) = u$, the same arguments as in Step 1 yield the analogous identities

$$G_v^{\mathcal{D}}(d) = \mathbf{F}^{\mathcal{D}}\left(\Big(d - \sum_{s=1}^{k} g_s^* g_s \delta_{s+k}(d)\Big)^{-1}\right), \tag{4.19}$$

$$\gamma_{r+k}(d) = \tau_{r+k}\left(g_r\Big(d - \sum_{s=1}^{k} g_s^* g_s \delta_{s+k}(d)\Big)^{-1} g_r^*\right), \tag{4.20}$$

$$G_u^{\mathcal{D}}\left(\sum_{r=0}^{2k} p_r\Big(\frac{1}{\gamma_r(d)} + \delta_r(d)\Big)\right) = \sum_{r=0}^{2k} p_r \gamma_r(d). \tag{4.21}$$

As $g_r^* g_r$ has moments given by (4.6), we may write (4.19) and (4.20) explicitly: Denote $d = d_0 p_0 + \ldots + d_{2k} p_{2k}$ for $d_0, \ldots, d_{2k} \in \mathbb{C}$. As d is invertible, we have $d^{-1} = d_0^{-1} p_0 + \ldots + d_{2k}^{-1} p_{2k}$. For any $x \in \mathcal{A}$ that commutes with $\mathcal{D}$,

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

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

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

Noting that $g_s^* g_s$ commutes with $\mathcal{D}$, applying the above to (4.19) with $x = \sum_{s=1}^k g_s^* g_s$, and recalling (4.6),

$$\begin{aligned}\tau_r(G_v^{\mathcal{D}}(d)) &= \sum_{l \geq 0} \frac{\tau_r((g_r^* g_r)^l)\delta_{r+k}(d)^l}{d_r^{l+1}} \\ &= \int \sum_{l \geq 0} \frac{x^l \delta_{r+k}(d)^l}{d_r^{l+1}} \nu_{m_r/n_r}(x)dx \\ &= \int \frac{1}{d_r - x\delta_{r+k}(d)} \nu_{m_r/n_r}(x)dx \\ &= \frac{1}{\delta_{r+k}(d)} G_{\nu_{m_r/n_r}}^{\mathbb{C}}(d_r/\delta_{r+k}(d)),\end{aligned} \tag{4.22}$$

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

Similarly, we may write (4.20) as

$$\begin{aligned}\gamma_{r+k}(d) &= \frac{m_r}{n_r}\tau_r\left(\left(d - \sum_{s=1}^k g_s^* g_s \delta_{s+k}(d)\right)^{-1} g_r^* g_r\right) \\ &= \frac{m_r}{n_r} \int \frac{x}{d_r - x\delta_{r+k}(d)} \nu_{m_r/n_r}(x)dx \\ &= \frac{m_r}{n_r}\left(-\frac{1}{\delta_{r+k}(d)} + \frac{d_r}{\delta_{r+k}(d)^2} G_{\nu_{m_r/n_r}}^{\mathbb{C}}(d_r/\delta_{r+k}(d))\right) \\ &= \frac{m_r}{n_r}\left(-\frac{1}{\delta_{r+k}(d)} + \frac{d_r}{\delta_{r+k}(d)} \tau_r(G_v^{\mathcal{D}}(d))\right),\end{aligned} \tag{4.23}$$

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 (4.22). The relation between $G_v^{\mathcal{D}}$ and $G_u^{\mathcal{D}}$ is given by (4.22), (4.23), and (4.21).

Step 3: We compute $m_0(z)$ for $z \in \mathbb{C}^+$ using (4.13), (4.14), (4.15), (4.22), (4.23), and (4.21). Fixing $z \in \mathbb{C}^+$, let us write

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

Applying (4.13) and projecting down to $\mathbb{C}$,

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

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 $r = 1, \ldots, k$. Then, passing to a power series as in Step 2, and then applying (4.5) and the spectral calculus,

$$\begin{aligned} m_0(z) &= -\sum_{l \geq 0} z^{-(l+1)} \tau_0\bigg(\Big(\sum_{r=1}^{k} h_r^* h_r \beta_r\Big)^{l}\bigg) \\ &= -\sum_{l \geq 0} z^{-(l+1)} \frac{1}{p} \operatorname{Tr}\bigg(\Big(\sum_{r=1}^{k} \beta_r H_r^* H_r\Big)^{l}\bigg) \\ &= -\frac{1}{p} \operatorname{Tr}\Big(z \operatorname{Id}_p - \sum_{r=1}^{k} \beta_r H_r^* H_r\Big)^{-1}. \end{aligned} \tag{4.24}$$

Similarly, (4.14) implies for each $r = 1, \ldots, k$

$$\alpha_r = \frac{1}{m_r} \operatorname{Tr}\bigg(\Big(z \operatorname{Id}_p - \sum_{s=1}^{k} \beta_s H_s^* H_s\Big)^{-1} H_r^* H_r\bigg). \tag{4.25}$$

Now applying (4.22) and recalling (4.15) and the definition of d_r, for each $r = 1, \ldots, k$,

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

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

$$\frac{\beta_r}{\delta_{r+k}} = \mathcal{R}^{\mathbb{C}}_{\nu_{m_r/n_r}}(\alpha_r \delta_{r+k}) = \frac{n_r}{n_r - m_r \alpha_r \delta_{r+k}}. \tag{4.26}$$

By (4.23) and (4.15),

$$\gamma_{r+k} = \frac{m_r}{n_r} \frac{\alpha_r \beta_r}{\delta_{r+k}}. \tag{4.27}$$

We derive two consequences of (4.26) and (4.27). First, substituting for β_r in (4.27) using (4.26) and recalling the definition of e_{r+k} yields

$$e_{r+k} = \frac{n_r}{m_r \alpha_r}. \tag{4.28}$$

Second, rearranging (4.26), we get $\beta_r/\delta_{r+k} = 1 + m_r\alpha_r\beta_r/n_r$. Inserting into (4.27) yields this time

$$\beta_r = \frac{n_r}{m_r^2 \alpha_r^2}(n_r \gamma_{r+k} - m_r \alpha_r). \tag{4.29}$$

By (4.21), for each $r = 1, \ldots, k$,

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

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

$$\begin{aligned}
\gamma_{r+k} &= \frac{1}{n_r} \operatorname{Tr}_{r+k} \left(\operatorname{diag}(e_0 \operatorname{Id}_p, \ldots, e_{2k} \operatorname{Id}_{n_k}) - \tilde{F} \right)^{-1} \\
&= \frac{1}{n_r} \operatorname{Tr}_r \left(\operatorname{diag}(e_{k+1} \operatorname{Id}_{n_1}, \ldots, e_{2k} \operatorname{Id}_{n_k}) - F \right)^{-1} \\
&= \frac{1}{n_r} \operatorname{Tr}_r (D^{-1} - F)^{-1} \qquad (4.30)
\end{aligned}$$

where the last line applies (4.28) 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 $\operatorname{Tr}_r D = m_r \alpha_r$, (4.29) yields

$$\begin{aligned}
\beta_r &= \frac{1}{n_r D_r^2} \operatorname{Tr}_r[(D^{-1} - F)^{-1} - D] \\
&= \frac{1}{n_r} \operatorname{Tr}_r[(F^{-1} - D)^{-1}] = \frac{1}{n_r} \operatorname{Tr}_r((\operatorname{Id}_{n_+} - FD)^{-1} F) \qquad (4.31)
\end{aligned}$$

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

(4.24), (4.25), and (4.31).

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 $\|w\| \leq C$ 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 (4.9). $\alpha_r(z)$ as defined by (4.10) satisfies

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

for $z \in \mathbb{D}$. Since $|\tau(w^l h_r^* h_r)| \leq \|w\|^l \|h_r\|^2$, 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 non-zero 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), multi-linearity of κ_l, and (3.16), it is verified that $\beta_r(z)$ defined by (4.11) 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 (4.24) and (4.25) 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 (4.16). By the same arguments as above, $\gamma_{r+k}(d)$ as defined by (4.17) is non-zero and uniformly small, and $\delta_{r+k}(d)$ as defined by (4.18) is uniformly bounded. Then the formal series leading to (4.22) and (4.23) 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 (4.30) is convergent for $z \in \mathbb{D}$. This verifies the validity of the preceding calculations and concludes the proof. □

To finish the proof of Theorem 4.1, we show using a contractive mapping argument similar to [CDS11, DL11] that (4.1–4.2) have a unique solution in the stated domains, which is the limit of the procedure in Theorem 1.4. The result then follows from Lemma 4.4 and Corollary 3.10. These arguments are contained in Appendix C.

SUPPLEMENTARY MATERIAL

” **Supplement A: Supplementary Appendices**
(doi: 10.1214/00-AOASXXXXSUPP; .pdf). The supplementary appendices contain details for the examples in Section 2, proofs of Theorem 3.9 and Corollary 3.10, the conclusion of the proof of Theorem 4.1, the construction of the free deterministic equivalent models, and a separate exposition of the proof in Section 4 for the simpler setting of Theorem 1.1.

References.

[BAC+15] Mark W Blows, Scott L Allen, Julie M Collet, Stephen F Chenoweth, and Katrina McGuigan. The phenome-wide distribution of genetic variance. *The American Naturalist*, 186(1):15–30, 2015.

[Bar90] NH Barton. Pleiotropic models of quantitative variation. *Genetics*, 124(3):773–782, 1990.

[BCY10] Zhidong Bai, Jiaqi Chen, and Jianfeng Yao. On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Australian & New Zealand Journal of Statistics*, 52(4):423–437, 2010.

[BG09] Florent Benaych-Georges. Rectangular random matrices, related convolution. *Probability Theory and Related Fields*, 144(3-4):471–515, 2009.

[BGN11] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

[Blo07] Mark W Blows. A tale of two matrices: Multivariate approaches in evolutionary biology. *Journal of evolutionary biology*, 20(1):1–8, 2007.

[BM15] Mark W Blows and Katrina McGuigan. The distribution of genetic variance across phenotypic space and the response to selection. *Molecular ecology*, 24(9):2056–2072, 2015.

[BS06] Jinho Baik and Jack W Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.

[BY12] Zhidong Bai and Jianfeng Yao. On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106:167–177, 2012.

[CDS11] Romain Couillet, Mérouane Debbah, and Jack W Silverstein. A deterministic equivalent for the analysis of correlated MIMO multiple access channels. *IEEE Transactions on Information Theory*, 57(6):3493–3514, 2011.

[Col03] Benoit Collins. Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability. *International Mathematics Research Notices*, 2003(17):953–982, 2003.

[CŚ06] Benoit Collins and Piotr Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006.

[DL11] Florian Dupuy and Philippe Loubaton. On the capacity achieving covariance matrix for frequency selective MIMO channels using the asymptotic approach. *IEEE Transactions on Information Theory*, 57(9):5737–5753, 2011.

[Dyk95] Kenneth J Dykema. Amalgamated free products of multi-matrix algebras and a construction of subfactors of a free group factor. *American Journal of Mathematics*, 117(6):1555–1602, 1995.

[Fis18] Ronald A Fisher. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the royal society of Edinburgh*, 52(02):399–433, 1918.

[HLN07] Walid Hachem, Philippe Loubaton, and Jamal Najim. Deterministic equivalents for certain functionals of large random matrices. *The Annals of Applied Probability*, 17(3):875–930, 2007.

[HMB14] Emma Hine, Katrina McGuigan, and Mark W Blows. Evolutionary constraints in high-dimensional trait sets. *The American Naturalist*, 184(1):119–131, 2014.

[HP00] Fumio Hiai and Denes Petz. Asymptotic freeness almost everywhere for random matrices. *Acta Sci. Math. (Szeged)*, 66(3–4):809–834, 2000.

[Kar08] Noureddine El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *The Annals of Statistics*, 36(6):2757–2790, 2008.

[Kir09] Mark Kirkpatrick. Patterns of quantitative genetic variation in multiple dimensions. *Genetica*, 136(2):271–284, 2009.

[LA83] Russell Lande and Stevan J Arnold. The measurement of selection on correlated characters. *Evolution*, 37(6):1210–1226, 1983.

[LaM73] Lynn R LaMotte. Quadratic estimation of variance components. *Biometrics*, 29(2):311–330, 1973.

[Lan79] Russell Lande. Quantitative genetic analysis of multivariate evolution, applied to brain: Body size allometry. *Evolution*, 33(1):402–416, 1979.

[Lix06] Zhang Lixin. *Spectral analysis of large dimensional random matrices.* PhD thesis, National University of Singapore, 2006.

[LW98] Michael Lynch and Bruce Walsh. *Genetics and analysis of quantitative traits*, volume 1. Sinauer Sunderland, MA, 1998.

[LW12] Olivier Ledoit and Michael Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2), 2012.

[MCM+14] Katrina McGuigan, Julie M Collet, Elizabeth A McGraw, H Ye Yixin, Scott L Allen, Stephen F Chenoweth, and Mark W Blows. The nature and extent of mutational pleiotropy in gene expression of male drosophila serrata. *Genetics*, 196(3):911–921, 2014.

[Mes08] Xavier Mestre. Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory*, 54(11):5113–5129, 2008.

[MP67] Vladimir A Marčenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4):457–483, 1967.

[MS07] Aris L Moustakas and Steven H Simon. On the outage capacity of correlated multiple-path MIMO channels. *IEEE Transactions on Information Theory*, 53(11):3887–3903, 2007.

[NS06] A. Nica and R. Speicher. *Lectures on the Combinatorics of Free Probability.* Number v. 13 in Lectures on the combinatorics of free probability. Cambridge University Press, 2006.

[NSS02] Alexandru Nica, Dimitri Shlyakhtenko, and Roland Speicher. Operator-valued distributions. I. Characterizations of freeness. *International Mathematics Research Notices*, 2002(29):1509–1538, 2002.

[PA89] Patrick C Phillips and Stevan J Arnold. Visualizing multivariate selection. *Evolution*, pages 1209–1222, 1989.

[PA14] Debashis Paul and Alexander Aue. Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference*, 150:1–29, 2014.

[Pau07] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[Rao72] C Radhakrishna Rao. Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association*, 67(337):112–115, 1972.

[RMSE08] N Raj Rao, James A Mingo, Roland Speicher, and Alan Edelman. Statistical eigen-inference from large Wishart matrices. *The Annals of Statistics*, pages 2850–2885, 2008.

[Rob59a] Alan Robertson. Experimental design in the evaluation of genetic parameters. *Biometrics*, 15(2):219–226, 1959.

[Rob59b] Alan Robertson. The sampling variance of the genetic correlation coefficient. *Biometrics*, 15(3):469–485, 1959.

[SB95] Jack W Silverstein and Zhidong Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate analysis*, 54(2):175–192, 1995.

[SCM09] Shayle R Searle, George Casella, and Charles E McCulloch. *Variance components*, volume 391. John Wiley & Sons, 2009.

[Spe98] Roland Speicher. *Combinatorial theory of the free product with amalgamation and operator-valued free probability theory*, volume 627. American Mathematical Soc., 1998.

[SV12] Roland Speicher and Carlos Vargas. Free deterministic equivalents, rectangular random matrix models, and operator-valued free probability theory. *Random Matrices: Theory and Applications*, 1(02):1150008, 2012.

[VDN92] D V Voiculescu, K J Dykema, and A Nica. *Free Random Variables*. CRM Monograph Series. American Mathematical Soc., 1992.

[Voi91] Dan Voiculescu. Limit laws for random matrices and free products. *Inventiones mathematicae*, 104(1):201–220, 1991.

[Voi95] Dan Voiculescu. Operations on certain non-commutative operator-valued random variables. *Recent advances in operator algebras (Orléans, 1992)*, Astérisque(232):243–275, 1995.

[Voi98] Dan Voiculescu. A strengthened asymptotic freeness result for random matrices with applications to free entropy. *International Mathematics Research Notices*, 1998(1):41–63, 1998.

[WB09] Bruce Walsh and Mark W Blows. Abundant genetic variation+strong selection=multivariate genetic constraints: A geometric view of adaptation. *Annual Review of Ecology, Evolution, and Systematics*, 40:41–59, 2009.

[Wig55] Eugene P Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, 1955.

Department of Statistics
390 Serra Mall
Stanford, CA 94305
E-mail: zhoufan@stanford.edu; imj@stanford.edu

SUPPLEMENTARY APPENDICES TO "EIGENVALUE DISTRIBUTIONS OF VARIANCE COMPONENTS ESTIMATORS IN HIGH-DIMENSIONAL RANDOM EFFECTS MODELS"

BY ZHOU FAN AND IAIN M. JOHNSTONE

APPENDIX A: DETAILS FOR CLASSIFICATION DESIGNS

For an important class of classification designs, the system of equations (1.4–1.6) of our main result Theorem 1.2 can be simplified. In particular, for the examples of the unbalanced one-way design and balanced nested and crossed designs of Section 2, the matrix inverse and block trace in (1.5) can be carried out explicitly, leading to the expressions given in Corollaries 2.1, 2.2, and 2.3. We describe in this appendix the computations that lead to such simplifications.

We denote by $\mathbf{1}_l \in \mathbb{R}^l$ the vector of all 1's and $\mathbf{J}_l = \mathbf{1}_l\mathbf{1}_l^T \in \mathbb{R}^{l\times l}$ the matrix of all 1's. For $A \in \mathbb{R}^{l\times m}$ and $B \in \mathbb{R}^{l'\times m'}$, we use standard tensor product notation

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1m}B \\ \vdots & \ddots & \vdots \\ a_{l1}B & \cdots & a_{lm}B \end{pmatrix} \in \mathbb{R}^{ll'\times mm'}.$$

For $m = l_1 + \ldots + l_k$ and the direct sum decomposition $\mathbb{R}^m = \mathbb{R}^{l_1} \oplus \ldots \oplus \mathbb{R}^{l_k}$, we call

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & \ddots & \vdots \\ A_{k1} & \cdots & A_{kk} \end{pmatrix} \in \mathbb{R}^{m\times m}, \qquad A_{ij} \in \mathbb{R}^{l_i\times l_j} \text{ for all } i,j \in \{1,\ldots,k\}$$

the corresponding block matrix decomposition. For subspaces $S_1 \subset S_2$ of $\mathbb{R}^m$, we denote by $S_2 \ominus S_1$ the orthogonal complement of S_1 in S_2.

A.1. One-way classification. The model (2.1) may be written in the form (1.1) with $k = 2$ upon identifying $X\beta = \mathbf{1}_n\mu^T$, stacking the α_i's and $\varepsilon_{i,j}$'s as the rows of $\alpha_1 \in \mathbb{R}^{I\times p}$ and $\alpha_2 \in \mathbb{R}^{n\times p}$, and setting

$$U_1 = \begin{pmatrix} \mathbf{1}_{J_1} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{J_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_{J_I} \end{pmatrix},$$

$U_2 = \mathrm{Id}_n$, and $I_1 = I$ and $I_2 = n$.

Consider the nested subspaces $\mathrm{col}(\mathbf{1}_n) \subset \mathrm{col}(U_1) \subset \mathbb{R}_n$, where $\mathrm{col}(\mathbf{1}_n)$ is the 1-dimensional span of $\mathbf{1}_n$ and $\mathrm{col}(U_1)$ is the column span of U_1. Let π_0, π_1, π_2 be the orthogonal projections onto $\mathrm{col}(\mathbf{1}_n)$, $\mathrm{col}(U_1) \ominus \mathrm{col}(\mathbf{1}_n)$, and $\mathbb{R}^n \ominus \mathrm{col}(U_1)$. Then the quantities SS_1 and SS_2 are equivalently expressed as $\mathrm{SS}_1 = Y^T\pi_1Y$ and $\mathrm{SS}_2 = Y^T\pi_2Y$, so the MANOVA estimators (2.2) are given by $\hat{\Sigma}_1 = Y^TB_1Y$ and $\hat{\Sigma}_2 = Y^TB_2Y$ for

$$B_1 = \frac{1}{K}\left(\frac{1}{I-1}\pi_1 - \frac{1}{n-I}\pi_2\right), \qquad B_2 = \frac{1}{n-I}\pi_2.$$

To study $\hat{\Sigma}_1$, let us consider instead the matrix $\check{\Sigma}_1 = Y^T \check{B}_1 Y$ for

$$\check{B}_1 = \frac{1}{K}\left(\frac{1}{I}(\pi_0+\pi_1) - \frac{1}{n-I}\pi_2\right).$$

Note that the given assumptions imply $K \geq c$ and $\|U_1\| \leq C$ for constants $C, c > 0$. Then

$$\left\|\left(\frac{1}{K(I-1)} - \frac{1}{KI}\right) Y^T \pi_1 Y\right\| \leq \frac{1}{KI(I-1)}\|Y\|^2 \leq C/N$$

a.s. for large N, so $\hat{\Sigma}_1 - \check{\Sigma}_1$ is the sum of a rank-one matrix and a matrix of norm at most C/N. Then we have $\mu_{\hat{\Sigma}_1} - \mu_{\check{\Sigma}_1} \to 0$ a.s.

We apply Theorem 1.2 to $\check{\Sigma}_1$: Let us set

$$E = \operatorname{diag}(J_1, \ldots, J_I) = U_1^T U_1, \qquad V_1 = U_1 E^{-1/2}.$$

Then $\pi_0 + \pi_1 = V_1 V_1^T$. We may complete the basis and write $\pi_2 = V_2 V_2^T$ for V_2 of size $n \times (n-I)$, so that $[V_1 \mid V_2]$ is an orthogonal matrix of size n. Define the orthogonal change of basis matrix

$$Q = \begin{pmatrix} \mathrm{Id}_I & 0 \\ 0 & [V_1 \mid V_2] \end{pmatrix} \in \mathbb{R}^{(I+n)\times(I+n)}.$$

Recall $F = U^T \check{B}_1 U$ from Theorem 1.2, where $U = (\sqrt{I_1}U_1 \mid \sqrt{I_2}U_2) = (\sqrt{I}V_1E^{1/2} \mid \sqrt{n}\,\mathrm{Id}_n)$. Then $UQ = (\sqrt{I}V_1E^{1/2} \mid \sqrt{n}V_1 \mid \sqrt{n}V_2)$ in the decomposition $\mathbb{R}^{I+n} = \mathbb{R}^I \oplus \mathbb{R}^I \oplus \mathbb{R}^{n-I}$, so

$$M := Q^T F Q = \frac{1}{K}\begin{pmatrix} E & \sqrt{\frac{n}{I}}E^{1/2} & 0 \\ \sqrt{\frac{n}{I}}E^{1/2} & \frac{n}{I}\,\mathrm{Id}_I & 0 \\ 0 & 0 & -\frac{n}{n-I}\,\mathrm{Id}_{n-I} \end{pmatrix}.$$

We must compute the block traces of $(\mathrm{Id} + FD(a))^{-1}F$ in the decomposition $\mathbb{R}^{I+n} = \mathbb{R}^I \oplus \mathbb{R}^n$, where $D(a) = \operatorname{diag}(a_1\,\mathrm{Id}_I, a_2\,\mathrm{Id}_n)$. Note that Q preserves this decomposition as well as $D(a)$, so

$$S := Q^T(\mathrm{Id} + FD(a))^{-1}FQ = (\mathrm{Id} + MD(a))^{-1}M.$$

Moving now to the block decomposition $\mathbb{R}^{I+n} = \mathbb{R}^{2I} \oplus \mathbb{R}^{n-I}$, let us write $M = \operatorname{diag}(RR^T, -r_2^2\,\mathrm{Id}_{n-I})$ and $D(a) = \operatorname{diag}(\Delta, a_2\,\mathrm{Id}_{n-I})$, where we set $r_0^2 = 1/K$, $r_1^2 = n/(KI)$, $r_2^2 = n/(K(n-I))$, and

$$R = \begin{pmatrix} r_0 E^{1/2} \\ r_1\,\mathrm{Id}_I \end{pmatrix}, \qquad \Delta = \begin{pmatrix} a_1\,\mathrm{Id}_I & 0 \\ 0 & a_2\,\mathrm{Id}_I \end{pmatrix}.$$

To compute the upper-left $2I \times 2I$ block S_{11} in this decomposition, we use the Woodbury identity

$$(\Delta^{-1} + RR^T)^{-1} = \Delta - \Delta R(\mathrm{Id}_I + R^T\Delta R)^{-1}R^T\Delta$$

to obtain

$$S_{11} = (\mathrm{Id}_{2I} + RR^T\Delta)^{-1}RR^T = \Delta^{-1}(\Delta^{-1} + RR^T)^{-1}RR^T = R(\mathrm{Id}_I + R^T\Delta R)^{-1}R^T. \tag{A.1}$$

We compute $\mathrm{Id}_I + R^T\Delta R = \mathrm{Id}_I + a_1 r_0^2 E + a_2 r_1^2\,\mathrm{Id}_I$, which yields

$$S_{11} = \begin{pmatrix} \operatorname{diag}\left(\frac{IJ_i}{KI + IJ_i a_1 + n a_2}\right) & * \\ * & \operatorname{diag}\left(\frac{n}{KI + IJ_i a_1 + n a_2}\right) \end{pmatrix}$$

for values $*$ that we omit for brevity. The lower-right $(n-I) \times (n-I)$ block of S is given by

$$S_{22} = (\mathrm{Id} - r_2^2 a_2 \mathrm{Id})^{-1}(-r_2^2 \mathrm{Id}) = -\frac{n}{K(n-I) - na_2} \mathrm{Id}_{n-I}.$$

As Q preserves the decomposition $\mathbb{R}^{I+n} = \mathbb{R}^I \oplus \mathbb{R}^n$, the block traces of S in this decomposition are the same as those of $(\mathrm{Id} + FD(a))^{-1}F$. This yields the formulas for b_1 and b_2 in Corollary 2.1(a).

We next apply Theorem 1.2 for $\hat{\Sigma}_2$: The matrix $F = U^T B_2 U$ is now given by

$$F = \begin{pmatrix} 0 & 0 \\ 0 & \frac{n}{n-I}\pi_2 \end{pmatrix} \in \mathbb{R}^{(I+n)\times(I+n)}.$$

Then in the decomposition $\mathbb{R}^{I+n} = \mathbb{R}^I \oplus \mathbb{R}^I \oplus \mathbb{R}^{n-I}$, we have

$$Q^T F Q = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{n}{n-I} \mathrm{Id}_{n-I} \end{pmatrix}, \qquad Q^T \left(\mathrm{Id} + FD(a)\right)^{-1} FQ = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{n}{n-I+na_2} \mathrm{Id}_{n-I} \end{pmatrix}.$$

Taking block traces, a_1 is irrelevant, $b_1 \equiv 0$, and b_2 has the form of Corollary 2.1(b).

A.2. Balanced models. We consider models of the form (1.1) given by balanced crossed and nested classification designs. These satisfy the following "balanced model conditions":

1. For each r, let $c_r = n/I_r$. Then $U_r^T U_r = c_r \mathrm{Id}_{I_r}$, and $\Pi_r := c_r^{-1} U_r U_r^T$ is an orthogonal projection onto a subspace $S_r \subset \mathbb{R}^n$ of dimension I_r.
2. Define $S_0 = \mathrm{col}(X)$ as the column span of X. Then $S_0 \subset S_r \subset S_k = \mathbb{R}^n$ for each $r = 1, \ldots, k-1$.
3. Partially order the subspaces S_r by inclusion, $r' \preceq r$ if $S_r' \subseteq S_r$. Let $\mathring{S}_r$ denote the orthogonal complement in S_r of all $S_{r'}$ properly contained in S_r. Then for each r,

$$S_r = \bigoplus_{r' \preceq r} \mathring{S}_{r'}, \tag{A.2}$$

   where $\oplus$ denotes the orthogonal direct sum. In particular, $\mathbb{R}^n = S_k = \oplus_{r=0}^k \mathring{S}_r$.

We verify below that the models of Sections 2.2 and 2.3 are of this form, with the partial orderings of $S_0, S_1, \ldots, S_k$ depicted in Figure 1.

Let $d_r = \dim(\mathring{S}_r)$ and let π_r denote the orthogonal projection onto $\mathring{S}_r$. By (A.2), $\Pi_r = \sum_{r' \preceq r} \pi_{r'}$. The sum-of-squares and mean-squares corresponding to $\mathring{S}_r$ are defined respectively as

$$\mathrm{SS}_r = Y^T \pi_r Y, \qquad \mathrm{MS}_r = \mathrm{SS}_r / d_r.$$

To evaluate $\mathbb{E}[\mathrm{SS}_t]$ for $t \geq 1$, note that $\pi_t X = 0$ by the condition $S_0 \subset S_t$. Then

$$\mathbb{E}[\mathrm{SS}_t] = \sum_{r=1}^k \mathbb{E}[\alpha_r^T U_r^T \pi_t U_r \alpha_r] = \sum_{r=1}^k \mathrm{Tr}(U_r^T \pi_t U_r) \Sigma_r.$$

From the definitions, $\mathrm{Tr}(U_r^T \pi_t U_r) = c_r \mathrm{Tr}(\pi_t \Pi_r) = c_r d_t \mathbb{1}\{t \preceq r\}$, which yields

$$\mathbb{E}[\mathrm{MS}_t] = \sum_{r \succeq t} c_r \Sigma_r = \sum_r \zeta(t,r) c_r \Sigma_r,$$

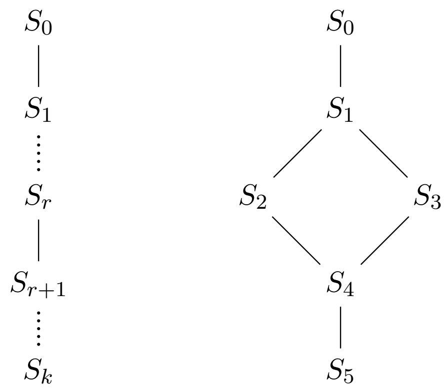

FIG 1. *Inclusion lattices for the subspaces $\{S_r\}$ determined by the nested (left) and crossed (right) examples.*

where $\zeta(t,z) := \mathbb{1}\{t \preceq r\}$ is the "zeta function" associated with the partial order. We can then use Möbius inversion to write Σ_r in terms of the expected mean squares,

$$c_r\Sigma_r = \sum_u \mu(r,u)\mathbb{E}[\mathrm{MS}_u],$$

where $\mu(r,u)$ is the associated "Möbius function". The MANOVA estimators are defined by substituting MS_u for $\mathbb{E}[\mathrm{MS}_u]$, which yields

$$\hat{\Sigma}_t = Y^T B_t Y, \qquad B_t = \sum_u \beta_{tu}\pi_u, \qquad \beta_{tu} = \frac{\mu(t,u)}{c_t d_u}. \tag{A.3}$$

For our examples, we may solve for $\mu(t,u)$ directly; more general discussions regarding the Möbius inversion may be found in [Spe83].

We apply Theorem 1.2 to each $\hat{\Sigma}_t$: To compute $F = U^T B_t U$, we change to a new basis for $\mathbb{R}^{I_+}$. Let V_r be an $n \times d_r$ matrix whose columns form an orthonormal basis for $\mathring{S}_r$. Let $\check{V}_r$ be the $n \times I_r$ matrix obtained by combining the columns of each $V_{r'}$ for $r' \preceq r$. In view of (A.2), the columns of $\check{V}_r$ are an orthonormal basis for S_r, and we have

$$U_r = \sqrt{c_r}\check{V}_r\check{W}_r^T$$

for some orthogonal matrix $\check{W}_r$ of size I_r. The block diagonal matrix $Q = \mathrm{diag}(\check{W}_r)$ of size I_+ is also orthogonal, and

$$U = (\sqrt{I_1}U_1 \mid \cdots \mid \sqrt{I_k}U_k) = \sqrt{n}(\check{V}_1 \mid \cdots \mid \check{V}_k)Q^T.$$

This yields $n^{-1/2}UQ = (\check{V}_1 \mid \cdots \mid \check{V}_k)$. The matrix V_r occurs in each $\check{V}_{r'}$ for which $r' \succeq r$. Let P be the permutation that collects all such occurrences of each V_r. Thus, if we set $s(r) = |\{r' : r' \succeq r\}|$ for $r \geq 1$ and $s(0) = k$, then we have

$$n^{-1/2}UQP = \left[\, \mathbf{1}_{s(0)}^T \otimes V_0 \mid \cdots \mid \mathbf{1}_{s(k)}^T \otimes V_k \,\right].$$

(This is now in the decomposition $\mathbb{R}^{I_+} = \mathbb{R}^{s(0)d_0} \oplus \cdots \oplus \mathbb{R}^{s(k)d_k}$.) Now write $O = QP$. Recalling that $F = U^T B_t U$ with $B_t = \sum_u \beta_{tu} V_u^T V_u$, and noting that $V_r^T V_u = \mathbb{1}\{r = u\}\,\mathrm{Id}_{d_u}$, we can write the (r,r') block of $O^T F O$, for $0 \leq r, r' \leq k$, in the form

$$(O^T F O)_{rr'} = n(\mathbf{1}_{s(r)} \otimes V_r^T)B_t(\mathbf{1}_{s(r')}^T \otimes V_{r'})$$

$$= \mathbb{1}\{r = r'\} n\beta_{tr} \mathbf{J}_{s(r)} \otimes \mathrm{Id}_{d_r} .$$

Thus we obtain the block diagonal representation

$$M := O^T F O = \mathrm{diag}(n\beta_{tr} \mathbf{J}_{s(r)} \otimes \mathrm{Id}_{d_r}).$$

We wish to compute the block trace of $(\mathrm{Id} + FD(a))^{-1}F$ in the original decomposition $\mathbb{R}^{I_+} = \mathbb{R}^{I_1} \oplus \cdots \oplus \mathbb{R}^{I_k}$. As the blocks of $Q = \mathrm{diag}(\check{W}_r)$ are orthogonal of size I_r, Q preserves this decomposition as well as $D(a)$, so we have

$$E(a) := O^T D(a) O = P^T D(a) P = \mathrm{diag}(D_0(a) \otimes \mathrm{Id}_{d_0}, \ldots, D_k(a) \otimes \mathrm{Id}_{d_k}),$$

where $D_r(a) = \mathrm{diag}(a_{r'} : r' \succeq r)$ is $s(r) \times s(r)$. Hence

$$\begin{aligned} S &:= O^T (\mathrm{Id} + FD(a))^{-1} F O \\ &= (\mathrm{Id} + ME(a))^{-1} M \\ &= \mathrm{diag}\left\{([\mathrm{Id}_{s(r)} + n\beta_{tr} \mathbf{J}_{s(r)} D_r(a)]^{-1} n\beta_{tr} \mathbf{J}_{s(r)}) \otimes \mathrm{Id}_{d_r}\right\}_{r=0}^{k} . \end{aligned}$$

We apply the Woodbury formula via (A.1) with $R = \sqrt{n\beta_{tr}} \mathbf{1}_{s(r)}$ and $\Delta = D_r(a)$, to obtain

$$\text{(A.4)} \qquad S = \mathrm{diag}\{\gamma_{tr} \mathbf{J}_{s(r)} \otimes \mathrm{Id}_{d_r}\}_{r=0}^{k}, \qquad \gamma_{tr} := \frac{n\beta_{tr}}{1 + n\beta_{tr} q_r}, \qquad q_r := \mathbf{1}_{s(r)}^T D_r(a) \mathbf{1}_{s(r)} = \sum_{r' \succeq r} a_{r'}.$$

Now undo the permutation P and observe that the (r, r) block corresponding to the decomposition $\mathbb{R}^{I_+} = \mathbb{R}^{I_1} \oplus \cdots \oplus \mathbb{R}^{I_k}$ is given by

$$((P^T)^{-1} S P^{-1})_{rr} = \mathrm{diag}(\gamma_{tu} \mathrm{Id}_{d_u} : u \preceq r).$$

Hence, we obtain the block trace

$$\mathrm{Tr}_r (\mathrm{Id} + FD(a))^{-1} F = \mathrm{Tr}_r[(P^T)^{-1} S P^{-1}] = \sum_{u \preceq r} \gamma_{tu} d_u.$$

Recall from (A.3) that $n\beta_{tu} d_u = I_t \mu(t, u)$. Then substituting for γ_{tu} as defined in (A.4), the equation (1.5) simplifies to

$$\text{(A.5)} \qquad b_r(z) = -\frac{I_t}{I_r} \sum_{u \preceq r} \frac{\mu(t, u)}{1 + (I_t/d_u)\mu(t, u) q_u}.$$

A.2.1. *Balanced nested classification.* The model (2.3) may be written in the form (1.1) upon identifying $I_r = \prod_{s=1}^{r} J_s$, $c_r = n/I_r$, $X\beta = \mathbf{1}_n \mu^T$, $U_r = \mathrm{Id}_{I_r} \otimes \mathbf{1}_{c_r}$, and stacking the vectors $\alpha_*^{(r)}$ as rows of $\alpha_r \in \mathbb{R}^{I_r \times p}$ and ε_* as rows of $\alpha_k \in \mathbb{R}^{n \times p}$. The balanced model conditions are easily verified, with (A.2) following from the linear structure of the inclusion lattice. Direct inversion yields

$$\mu(t, u) = \begin{cases} 1 & \text{if } u = t \\ -1 & \text{if } u = t + 1 \\ 0 & \text{otherwise.} \end{cases}$$

We also have $d_t = I_{t-1}(J_t - 1)$, so the form of the MANOVA estimator from (A.3) is

$$\hat{\Sigma}_t = Y^T \left(\frac{J_t}{n(J_t - 1)} \pi_t - \frac{1}{n(J_{t+1} - 1)} \pi_{t+1} \right) Y$$

for $t = 1, \ldots, k-1$, and the same without the π_{t+1} term for $t = k$. As $I_t/d_t = J_t/(J_t - 1)$ and $I_t/d_{t+1} = 1/(J_{t+1} - 1)$, (A.5) may be written as

$$b_r = \begin{cases} 0 & \text{if } r < t \\ -\frac{J_t - 1}{J_t - 1 + J_t q_t} & \text{if } r = t \\ -\frac{1}{J_{t+1} \ldots J_r}\left(\frac{J_t - 1}{J_t - 1 + J_t q_t} - \frac{J_{t+1} - 1}{J_{t+1} - 1 - q_{t+1}}\right) & \text{if } r > t. \end{cases}$$

Noting $q_t = \sum_{r \geq t} a_r$, this implies that $a_1, \ldots, a_{r-1}$ are irrelevant, and we obtain Corollary 2.2.

A.2.2. *Replicated crossed two-way classification.* The model (2.4) may be written in the form (1.1) with $k = 5$, upon identifying

$$I_1 = I,\ I_2 = IJ,\ I_3 = IK,\ I_4 = IJK,\ I_5 = n,$$

$X\beta = \mathbf{1}_n \mu^T$, $U_1 = \mathrm{Id}_I \otimes \mathbf{1}_{JKL}$, $U_2 = \mathrm{Id}_{IJ} \otimes \mathbf{1}_{KL}$, $U_3 = \mathrm{Id}_I \otimes \mathbf{1}_J \otimes \mathrm{Id}_K \otimes \mathbf{1}_L$, $U_4 = \mathrm{Id}_{IJK} \otimes \mathbf{1}_L$, $U_5 = \mathrm{Id}_n$, and stacking the vectors α_*, β_*, γ_*, δ_*, ε_* as the rows of matrices $\alpha_1, \ldots, \alpha_5$ respectively. The balanced model conditions are easily verified, where (A.2) uses the observation that $\mathring{S}_2 = S_2 \ominus S_1$ and $\mathring{S}_3 = S_3 \ominus S_1$ are orthogonal.

From the lattice structure of Figure 1 and direct evaluation, or by a general formula such as [Spe83, p. 380], we find that $M = (\mu(t,u))_{t,u=1,\ldots,5}$ has the upper triangular form

$$M = \begin{pmatrix} 1 & -1 & -1 & 1 & 0 \\ & 1 & 0 & -1 & 0 \\ & & 1 & -1 & 0 \\ & & & 1 & -1 \\ & & & & 1 \end{pmatrix}.$$

We also have the following values:

t	1	2	3	4	5
d_t	$I-1$	$I(J-1)$	$I(K-1)$	$I(J-1)(K-1)$	$IJKL$
q_t	$\sum_1^5 a_i$	$a_2 + a_4 + a_5$	$a_3 + a_4 + a_5$	$a_4 + a_5$	a_5

Then, for example, the MANOVA estimator $\hat{\Sigma}_2$ from (A.3) is given by

$$\hat{\Sigma}_2 = Y^T \left(\frac{1}{I(J-1)KL} \pi_2 - \frac{1}{I(J-1)(K-1)KL} \pi_4 \right) Y,$$

and the forms of the other estimators follow similarly.

To explicitly write (A.5), for $t > 1$, let $\sigma(t)$ be the successor of t in the partial order. (We do not need $\sigma(t)$ for $t = 5$.) Then $\mu(t,u)$ is only non-zero for $u = t$ and $u = \sigma(t)$, so we have

$$b_r = \begin{cases} -\frac{1}{1 + (I_t/d_t) q_t} & \text{if } r = t \\ -\frac{I_t}{I_r}\left(\frac{1}{1 + (I_t/d_t) q_t} - \frac{1}{1 - (I_t/d_{\sigma(t)}) q_{\sigma(t)}}\right) & \text{if } r \succ t \\ 0 & \text{otherwise.} \end{cases}$$

Specializing to $t = 2$ yields Corollary 2.3, and analogous expressions may be derived for $t = 3, 4, 5$.

For $t = 1$, we have the following values for $\check{\gamma}_u := (1 + (I_1/d_u)\mu(1,u) q_u)^{-1}$:

u	1	2	3	4
$\check{\gamma}_u$	$\dfrac{I-1}{I-1+I\sum_1^5 a_i}$	$\dfrac{J-1}{J-1-(a_2+a_4+a_5)}$	$\dfrac{K-1}{K-1-(a_3+a_4+a_5)}$	$\dfrac{(J-1)(K-1)}{(J-1)(K-1)+a_4+a_5}$

Then (A.5) simplifies to the equations

$$\begin{aligned} b_1 &= -\check{\gamma}_1 \\ b_2 &= -J^{-1}(\check{\gamma}_1 - \check{\gamma}_2) \\ b_3 &= -K^{-1}(\check{\gamma}_1 - \check{\gamma}_3) \\ b_4 &= -(JK)^{-1}(\check{\gamma}_1 - \check{\gamma}_2 - \check{\gamma}_3 + \check{\gamma}_4) \\ b_5 &= L^{-1} b_4. \end{aligned}$$

APPENDIX B: PROOF OF ASYMPTOTIC FREENESS

We prove Theorem 3.9 and Corollary 3.10. To ease subscript notation, throughout this section we denote by $M[i,j]$ the (i,j) entry of a matrix M.

Let Q be a $*$-polynomial in $(x_i)_{i\in\mathcal{I}_j, j\in\{1,\ldots,J\}}$ with coefficients in $\langle P_1,\ldots,P_d\rangle$, and let q denote the corresponding $*$-polynomial with coefficients in $\langle p_1,\ldots,p_d\rangle$. For Theorem 3.9, we wish to show for any r, almost surely as $N\to\infty$,

$$\left|N_r^{-1}\operatorname{Tr}_r Q\left(H_i : i\in\mathcal{I}_j, j\in\{1,\ldots,J\}\right) - \tau_r\left(q\left(h_i : i\in\mathcal{I}_j, j\in\{1,\ldots,J\}\right)\right)\right| \to 0. \tag{B.1}$$

The high-level strategy of the proof is the same as [BG09, Theorem 1.6], and follows these steps:

1. By applying linearity of Tr and τ, we may reduce to the case $Q=\prod_{k=1}^K Q_k$, where each Q_k is a simple-valued polynomial of a single family $(H_i : i\in\mathcal{I}_{j_k})$.
2. By "centering" each Q_k and inducting on K, it suffices to consider the case where $j_1\neq j_2$, $j_2\neq j_3,\ldots,j_K\neq j_1$ and each Q_k satisfies $\operatorname{Tr} Q_k(H_i : i\in\mathcal{I}_{j_k})=0$.
3. The main technical ingredient is Lemma B.2 below, which establishes the result for such Q. We use orthogonal invariance in law of $(H_i : i\in\mathcal{I}_{j_k})$ to introduce independently random block-orthogonal matrices, and then condition on the H_i's to reduce to a statement about Haar-orthogonal and deterministic matrices.

The last step above uses an explicit computation of the trace, together with basic properties of the joint moments of Haar-orthogonal matrices. We follow an approach inspired by [HP00, Theorem 2.1], but which (we believe) fills in an omission in the proof and also extends the combinatorial argument to deal with rectangular matrices and the orthogonal (rather than unitary) case.

Proof of Theorem 3.9. To show (B.1), by linearity of Tr and τ, it suffices to consider the case where Q is a $*$-monomial, which we may always write as a product of $Q_1,\ldots,Q_K$ where each Q_k depends only on the variables of a single family $\mathcal{I}_{j_k}$. Writing $Q_k=(P_1+\ldots+P_d)Q_k(P_1+\ldots+P_d)$ and again applying linearity of Tr and τ, it suffices to consider the case where each Q_k is simple-valued, i.e. $P_{r_k}Q_kP_{s_k}=Q_k$ for some $r_k,s_k\in\{1,\ldots,d\}$. If $s_k\neq r_{k+1}$ for any k (with the cyclic identification $r_{K+1}=r_1$), then (B.1) is trivial as both quantities on the left are 0. If $s_k=r_{k+1}$ for all k, then it suffices to consider $r=r_1$ and to replace $N_r^{-1}\operatorname{Tr}_r$ by $N^{-1}\operatorname{Tr}$ and τ_r by τ. The result then follows from Lemma B.1 below. $\square$

Lemma B.1. *Under the assumptions of Theorem 3.9, fix $K\geq 1$, $j_1,\ldots,j_K\in\{1,\ldots,J\}$, and $r_1,\ldots,r_K\in\{1,\ldots,d\}$. For each $k=1,\ldots,K$, let Q_k be a $*$-polynomial with coefficients in*

$\langle P_1,\ldots,P_d\rangle$ of the variables $(x_i)_{i\in\mathcal{I}_{j_k}}$ of the single family $\mathcal{I}_{j_k}$, such that $P_{r_k}Q_kP_{r_{k+1}}=Q_k$ (with the identification $r_{K+1}:=r_1$). Let $q_1,\ldots,q_K$ denote the corresponding $*$-polynomials with coefficients in $\langle p_1,\ldots,p_d\rangle$. Then, almost surely as $N\to\infty$,

$$\left|\frac{1}{N}\operatorname{Tr}\prod_{k=1}^{K}Q_k\left(H_i:i\in\mathcal{I}_{j_k}\right)-\tau\left(\prod_{k=1}^{K}q_k\left(h_i:i\in\mathcal{I}_{j_k}\right)\right)\right|\to 0. \tag{B.2}$$

Proof. We induct on K. For $K=1$, (B.2) holds by the assumption that $(h_i)_{i\in\mathcal{I}_{j_1}}$ and $(H_i)_{i\in\mathcal{I}_{j_1}}$ are asymptotically equal in $\mathcal{D}$-law a.s.

For $K\geq 2$, assume inductively that (B.2) holds for each value $1,\ldots,K-1$ in place of K. Let

$$t_k=\frac{1}{\tau(p_{r_k})}\tau\left(q_k\left(h_i:i\in\mathcal{I}_{j_k}\right)\right),$$

and define the "centered" $*$-polynomials

$$D_k=Q_k-t_kP_{r_k},\qquad d_k=q_k-t_kp_{r_k}.$$

We clarify that $t_k\in\mathbb{C}$ is a fixed constant (evaluated at the h_i's, not at the arguments x_i's of these $*$-polynomials), and thus D_k and d_k are still $*$-polynomials of $(x_i)_{i\in\mathcal{I}_{j_k}}$ with coefficients in $\langle P_1,\ldots,P_d\rangle$ and $\langle p_1,\ldots,p_d\rangle$. We have $t_k=0$ if $r_k\neq r_{k+1}$, because q_k is simple. Denoting by $\mathcal{S}_K$ the collection of all subsets of $\{k:r_k=r_{k+1}\}$ and applying a binomial expansion,

$$\frac{1}{N}\operatorname{Tr}\prod_{k=1}^{K}Q_k\left(H_i:i\in\mathcal{I}_{j_k}\right)=\sum_{S\in\mathcal{S}_K}Q(S)$$

where

$$Q(S):=\prod_{k\in S}t_k\cdot\frac{1}{N}\operatorname{Tr}\prod_{k\in\{1,\ldots,K\}\setminus S}D_k\left(H_i:i\in\mathcal{I}_{j_k}\right).$$

Each D_k still satisfies $P_{r_k}D_kP_{r_{k+1}}=D_k$. Hence, for every $S\neq\emptyset$, applying the induction hypothesis,

$$\left|Q(S)-\prod_{k\in S}t_k\cdot\tau\left(\prod_{k\in\{1,\ldots,K\}\setminus S}d_k\left(h_i:i\in\mathcal{I}_{j_k}\right)\right)\right|\to 0. \tag{B.3}$$

For $S=\emptyset$, if $j_k=j_{k+1}$ for some $k\in\{1,\ldots,K\}$ (or $j_K=j_1$), then combining D_kD_{k+1} into a single polynomial (and applying cyclic invariance of Tr and τ if $j_K=j_1$), the induction hypothesis still yields (B.3).

The remaining case is when $S=\emptyset$ and $j_k\neq j_{k+1}$ for each $k=1,\ldots,K$. Note, by definition of d_k, that

$$\tau\left(p_rd_k\left(h_i:i\in\mathcal{I}_{j_k}\right)p_r\right)=0$$

for each r and k, so by freeness of $(h_i)_{i\in\mathcal{I}_1},\ldots,(h_i)_{i\in\mathcal{I}_k}$ with amalgamation over $\langle p_1,\ldots,p_d\rangle$,

$$\tau\left(\prod_{k=1}^{K}d_k\left(h_i:i\in\mathcal{I}_{j_k}\right)\right)=0.$$

Thus, it remains to show that $Q(\emptyset)\to 0$. Note first that the definition of the free deterministic equivalent and the condition $N_r/N>c$ imply, almost surely as $N\to\infty$,

$$\left|\frac{N}{N_{r_k}}-\frac{1}{\tau(p_{r_k})}\right|\to 0,\qquad\left|\frac{1}{N}\operatorname{Tr}\left(Q_k\left(H_i:i\in\mathcal{I}_{j_k}\right)\right)-\tau\left(q_k\left(h_i:i\in\mathcal{I}_{j_k}\right)\right)\right|\to 0.$$

Hence $|t_k - T_k| \to 0$ a.s. for

$$T_k = \frac{1}{N_{r_k}} \operatorname{Tr} Q_k \left(H_i : i \in \mathcal{I}_{j_k}\right).$$

Then it suffices to show

$$M(\emptyset) := \frac{1}{N} \operatorname{Tr} \prod_{k=1}^{K} M_k \to 0$$

for the matrices

$$M_k = Q_k \left(H_i : i \in \mathcal{I}_{j_k}\right) - T_k P_{r_k},$$

as we may replace in $Q(\emptyset)$ each t_k by T_k and bound the remainders using the operator norm.

Finally, let us introduce random matrices $(O_{j,r})_{j \in \mathbb{N}, r \in \{1,\ldots,d\}}$ that are independent of each other and of the H_i's, such that each $O_{j,r}$ is orthogonal and Haar-distributed in $\mathbb{R}^{N_r \times N_r}$. For each $j \in \mathbb{N}$, define the block diagonal matrix $O_j = \operatorname{diag}(O_{j,1}, \ldots, O_{j,d})$. By orthogonal invariance in law of $(H_i)_{i \in \mathcal{I}_{j_k}}$, we have the equality in law

$$M(\emptyset) \stackrel{L}{=} \frac{1}{N} \operatorname{Tr} \prod_{k=1}^{K} O_{j_k} M_k O_{j_k}^{-1}.$$

Write $\check{M}_k \in \mathbb{R}^{N_{r_k} \times N_{r_{k+1}}}$ as the non-zero block of M_k. Then the above may be written as

$$M(\emptyset) \stackrel{L}{=} \frac{1}{N} \operatorname{Tr} \prod_{k=1}^{K} O_{j_k, r_k} \check{M}_k O_{j_k, r_{k+1}}^{-1} \operatorname{Id}_{N_{r_{k+1}}}. \tag{B.4}$$

Conditional on the H_i's, $\check{M}_k$ are deterministic matrices satisfying $\|\check{M}_k\| \le C$ for some constant $C > 0$ and all large N a.s., and if $r_k = r_{k+1}$ then $\operatorname{Tr} \check{M}_k = \operatorname{Tr} M_k = 0$ by definition of T_k. Furthermore, recall that we are in the case $j_k \neq j_{k+1}$ for each k.

The claim $M(\emptyset) \to 0$ follows from the following lemma:

Lemma B.2. *Fix $d, K \ge 1$, $l_1, \ldots, l_K \in \mathbb{N}$, $r_1, \ldots, r_K \in \{1, \ldots, d\}$, and $e_1, \ldots, e_K \in \{-1, 1\}$. For $N_1, \ldots, N_d \ge 1$, let $\{O_{l,r}\}_{l \in \mathbb{N}, r \in \{1,\ldots,d\}}$ be independent random matrices such that each $O_{l,r}$ is a Haar-distributed orthogonal matrix in $\mathbb{R}^{N_r \times N_r}$. Let $D_1 \in \mathbb{C}^{N_{r_1} \times N_{r_2}}, D_2 \in \mathbb{C}^{N_{r_2} \times N_{r_3}}, \ldots, D_K \in \mathbb{C}^{N_{r_K} \times N_{r_1}}$ be deterministic matrices such that, for each $k = 1, \ldots, K$ (and cyclically identifying $l_{K+1} := l_1$, etc.), if $(l_k, r_k, e_k) = (l_{k+1}, r_{k+1}, -e_{k+1})$, then $\operatorname{Tr} D_k = 0$.*

Let $N = N_1 + \ldots + N_d$, and suppose there exist constants $C, c > 0$ such that, as $N \to \infty$, $N_r/N > c$ for each $r = 1, \ldots, d$ and $\|D_k\| < C$ for each $k = 1, \ldots, K$. Then, almost surely,

$$N^{-1} \operatorname{Tr} \left(O_{l_1, r_1}^{e_1} D_1 O_{l_2, r_2}^{e_2} D_2 \ldots O_{l_K, r_K}^{e_K} D_K \right) \to 0.$$

(We emphasize that the matrices $O_{l,r}$ and D_k are N-dependent, while $(l_k, r_k, e_k, k = 1, \ldots, K)$ remain fixed as N grows.)

Assuming this lemma for now, write the right side of (B.4) in the form

$$N^{-1} \operatorname{Tr} \left(O_{l_1, r_1}^{e_1} D_1 O_{l_2, r_2}^{e_2} D_2 \ldots O_{l_{2K}, r_{2K}}^{e_{2K}} D_{2K} \right),$$

by making the identifications

$$(l_{2k-1}, r_{2k-1}, e_{2k-1}, D_{2k-1}) \leftarrow (j_k, r_k, 1, \check{M}_k)$$

$$(l_{2k}, r_{2k}, e_{2k}, D_{2k}) \leftarrow (j_k, r_{k+1}, -1, \mathrm{Id}_{N_{r_{k+1}}}).$$

Then Lemma B.2 implies $M(\emptyset) \to 0$ a.s. conditional on the H_i's, and hence unconditionally as well. Thus (B.3) holds for all $S \in \mathcal{S}_K$.

Finally, reversing the binomial expansion,

$$\sum_{S \in \mathcal{S}_K} \prod_{k \in S} t_k \cdot \tau \left(\prod_{k \in \{1,\ldots,K\} \setminus S} d_k \left(h_i : i \in \mathcal{I}_{j_k} \right) \right) = \tau \left(\prod_{k=1}^{K} q_k \left(h_i : i \in \mathcal{I}_{j_k} \right) \right).$$

This establishes (B.2), completing the induction. $\square$

To conclude the proof of Theorem 3.9, it remains to establish the above Lemma B.2. We require the following fact about joint moments of entries of Haar-orthogonal matrices:

Lemma B.3. *Let $O \in \mathbb{R}^{N \times N}$ be a random Haar-distributed real orthogonal matrix, let $K \geq 1$ be any positive integer, and let $i_1, j_1, \ldots, i_K, j_K \in \{1, \ldots, N\}$. Then:*

(a) There exists a constant $C := C_K > 0$ such that

$$\mathbb{E}\big[\big|O[i_1, j_1]O[i_2, j_2] \ldots O[i_K, j_K]\big|\big] \leq CN^{-K/2}.$$

(b) If there exists $i \in \{1, \ldots, N\}$ such that $i_k = i$ for an odd number of indices $k \in \{1, \ldots, K\}$ or $j_k = i$ for an odd number of indices $k \in \{1, \ldots, K\}$, then $\mathbb{E}[O[i_1, j_1] \ldots O[i_K, j_K]] = 0$.

Proof. [CŚ06, Eq. (21) and Theorem 3.13] imply $\mathbb{E}[O[i_1, j_1]^2 \ldots O[i_K, j_K]^2] \leq CN^{-K}$ for a constant $C := C_K > 0$. Part (a) then follows by Cauchy-Schwarz. Part (b) follows from the fact that the distribution of O is invariant to multiplication of row i or column i by -1, hence if $i_k = i$ or $j_k = i$ for an odd number of indices k, then $\mathbb{E}[O[i_1, j_1] \ldots O[i_K, j_K]] = -\mathbb{E}[O[i_1, j_1] \ldots O[i_K, j_K]]$. $\square$

Proof of Lemma B.2. Define $V_k = O_{l_k, r_k}^{e_k}$ (which is O_{l_k, r_k}^T if $e_k = -1$). Expanding the trace,

$$\text{(B.5)} \qquad \mathrm{Tr}\left[\prod_{k=1}^{K} V_k D_k\right] = \sum_{\mathbf{i}, \mathbf{j}} V(\mathbf{i}, \mathbf{j}) D(\mathbf{i}, \mathbf{j}),$$

where the summation is over all tuples $(\mathbf{i}, \mathbf{j}) := (i_1, j_1, i_2, j_2, \ldots, i_K, j_K)$ satisfying

$$\text{(B.6)} \qquad 1 \leq i_k, j_k \leq N_{r_k}$$

for each $k = 1, \ldots, K$, and where we have defined (with the identification $i_{K+1} := i_1$)

$$V(\mathbf{i}, \mathbf{j}) = \prod_{k=1}^{K} V_k[i_k, j_k], \quad D(\mathbf{i}, \mathbf{j}) = \prod_{k=1}^{K} D_k[j_k, i_{k+1}].$$

Denote

$$\text{(B.7)} \qquad \mathcal{E} = \mathbb{E}\left[\left|N^{-1} \mathrm{Tr}\left(\prod_{k=1}^{K} V_k D_k\right)\right|^2\right] = N^{-2} \sum_{\mathbf{i}, \mathbf{j}} \sum_{\mathbf{i}', \mathbf{j}'} D(\mathbf{i}, \mathbf{j}) \overline{D(\mathbf{i}', \mathbf{j}')} \mathbb{E}[V(\mathbf{i}, \mathbf{j}) V(\mathbf{i}', \mathbf{j}')],$$

where the second equality uses that each V_k is real and each D_k is deterministic. By the Borel-Cantelli lemma, it suffices to show $\mathcal{E} \leq CN^{-2}$ for some constant $C := C_K > 0$.

Let $\mathcal{R}$ be the set of distinct pairs among (l_k, r_k) for $k = 1, \dots, K$, corresponding to the set of distinct matrices $O_{l,r}$ that appear in (B.5). By independence of the matrices $O_{l,r}$,

$$\mathbb{E}[V(\mathbf{i},\mathbf{j})V(\mathbf{i}',\mathbf{j}')] = \prod_{(l,r)\in\mathcal{R}} \mathbb{E}\left[\prod_{k:(l_k,r_k)=(l,r)} V_k[i_k, j_k]V_k[i'_k, j'_k]\right]. \tag{B.8}$$

Since $O_{l,r}$ is invariant in law under permutations of rows and columns, each expectation on the right side above depends only on which indices are equal, and not on the actual index values. (For example, denoting $O := O_{l,r}$,

$$O[1,2]O^{-1}[2,3]O[1,4]O^{-1}[3,3] \overset{L}{=} O[8,7]O^{-1}[7,6]O[8,5]O^{-1}[6,6] \tag{B.9}$$

where the equality in law holds by permutation of both the rows and the columns of O.) We therefore analyse $\mathcal{E}$ by decomposing the sum in (B.7) over the different relevant partitions of $(\mathbf{i},\mathbf{j},\mathbf{i}',\mathbf{j}')$ specifying which indices are equal.

More precisely, let

$$\mathcal{I} = (i_k, j_k, i'_k, j'_k : k = 1, \dots, K)$$

be the collection of all indices, with cardinality $|\mathcal{I}| = 4K$. For each $(l,r) \in \mathcal{R}$, let

$$\mathcal{I}(l,r) = (i_k, j_k, i'_k, j'_k : k \text{ such that } l_k = l, r_k = r).$$

These sets $\mathcal{I}(l,r)$ form a fixed partition of $\mathcal{I}$. For each (l,r), denote by $\mathcal{Q}(l,r)$ any further partition of the indices in $\mathcal{I}(l,r)$, and let

$$\mathcal{Q} = \bigsqcup_{(l,r)\in\mathcal{R}} \mathcal{Q}(l,r) \tag{B.10}$$

be their combined partition of $\mathcal{I}$. Denoting by $Q_{l,r} = |\mathcal{Q}(l,r)|$ the number of elements of $\mathcal{Q}$ that partition $\mathcal{I}(l,r)$, we may identify

$$\mathcal{Q} \equiv \{(l,r,q) : (l,r) \in \mathcal{R},\ q \in \{1,\dots,Q_{l,r}\}\}.$$

We say that $(\mathbf{i},\mathbf{j},\mathbf{i}',\mathbf{j}')$ *induces* $\mathcal{Q}$ if, for every two indices belonging to the same set $\mathcal{I}(l,r)$, they are equal in value if and only if they belong to the same element of $\mathcal{Q}$.[1] Then $\mathbb{E}[V(\mathbf{i},\mathbf{j})V(\mathbf{i}',\mathbf{j}')]$ is the same for all $(\mathbf{i},\mathbf{j},\mathbf{i}',\mathbf{j}')$ that induce the same partition $\mathcal{Q}$. Thus we may define $E(\mathcal{Q}) = \mathbb{E}[V(\mathbf{i},\mathbf{j})V(\mathbf{i}',\mathbf{j}')]$ for any such $(\mathbf{i},\mathbf{j},\mathbf{i}',\mathbf{j}')$ and write

$$\mathcal{E} = N^{-2}\sum_{\mathcal{Q}} E(\mathcal{Q}) \sum_{\mathbf{i},\mathbf{j},\mathbf{i}',\mathbf{j}'|\mathcal{Q}} D(\mathbf{i},\mathbf{j})\overline{D(\mathbf{i}',\mathbf{j}')},$$

where the first sum is over all partitions $\mathcal{Q}$ of the form (B.10), and the second is over all $(\mathbf{i},\mathbf{j},\mathbf{i}',\mathbf{j}')$ satisfying (B.6) and inducing $\mathcal{Q}$.

Applying Lemma B.3(a) and the bound $N_r/N > c$ to (B.8), we have $|E(\mathcal{Q})| \le CN^{-K}$ for a constant $C := C_K > 0$ and all partitions $\mathcal{Q}$. Thus

$$\mathcal{E} \le CN^{-2-K} \sum_{\mathcal{Q}:E(\mathcal{Q})\neq 0} |D(\mathcal{Q})| \tag{B.11}$$

[1] For example, if $K = 2$, in display (B.9), both $(i_1, j_1, i_2, j_2, i'_1, j'_1, i'_2, j'_2) = (1,2,2,3,1,4,3,3)$ and $(8,7,7,6,8,5,6,6)$ induce

$$\mathcal{Q}(l,r) = \{\{i_1, i'_1\}, \{j_1, i_2\}, \{j_2, i'_2, j'_2\}, \{j'_1\}\} \quad \text{with} \quad Q_{l,r} = 4.$$

.

where

$$D(\mathcal{Q}) := \sum_{\mathbf{i},\mathbf{j},\mathbf{i}',\mathbf{j}'|\mathcal{Q}} D(\mathbf{i},\mathbf{j})\overline{D(\mathbf{i}',\mathbf{j}')} = \sum_{\mathbf{i},\mathbf{j},\mathbf{i}',\mathbf{j}'|\mathcal{Q}} \prod_{k=1}^{K} D_k[j_k, i_{k+1}] \prod_{k=1}^{K} \overline{D_k}[j'_k, i'_{k+1}].$$

For fixed $\mathcal{Q}$, we may rewrite $D(\mathcal{Q})$ as follows: Denote $L = 2K$, $M_k = D_k$, and $M_{K+k} = \overline{D_k}$. Let $\mathfrak{q}, \mathfrak{q}' : \{1, \ldots, L\} \to \mathcal{Q}$ be the maps such that $\mathfrak{q}(k), \mathfrak{q}'(k), \mathfrak{q}(K+k), \mathfrak{q}'(K+k)$ are the elements of $\mathcal{Q}$ containing $j_k, i_{k+1}, j'_k, i'_{k+1}$, respectively. Then

$$D(\mathcal{Q}) = \sum_{\alpha} \prod_{\ell=1}^{L} M_\ell[\alpha_{\mathfrak{q}(\ell)}, \alpha_{\mathfrak{q}'(\ell)}],$$

where $\sum_\alpha$ denotes the summation over all maps $\alpha : \mathcal{Q} \to \mathbb{N}$ such that $\alpha(l, r, q) \in \{1, \ldots, N_r\}$ for each $(l, r, q) \in \mathcal{Q}$ and $\alpha(l, r, q) \neq \alpha(l, r, q')$ whenever $q \neq q'$. (So α gives the index values, which must be distinct for elements of $\mathcal{Q}$ corresponding to the same $(l, r) \in \mathcal{R}$.)

We may simplify this condition on α by considering the following embedding: Let

$$\tilde{N} = \sum_{(l,r)\in\mathcal{R}} N_r,$$

and consider the corresponding block decomposition of $\mathbb{C}^{\tilde{N}}$ with blocks indexed by $\mathcal{R}$. (So the (l, r) block has size N_r.) For each $\ell = 1, \ldots, L$, if $\mathfrak{q}(\ell) = (l, r, q)$ and $\mathfrak{q}'(\ell) = (l', r', q')$, then note that M_ℓ is of size $N_r \times N_{r'}$. Let $\tilde{M}_\ell \in \mathbb{C}^{\tilde{N}\times\tilde{N}}$ be its embedding whose $(l, r) \times (l', r')$ block equals M_ℓ and whose remaining blocks equal 0. Then

$$D(\mathcal{Q}) = \sum_{\alpha} \prod_{\ell=1}^{L} \tilde{M}_\ell[\alpha_{\mathfrak{q}(\ell)}, \alpha_{\mathfrak{q}'(\ell)}],$$

where $\sum_\alpha$ now denotes the summation over all maps $\alpha : \mathcal{Q} \to \{1, \ldots, \tilde{N}\}$ such that each $\alpha(l, r, q)$ belongs to the (l, r) block of $\{1, \ldots, \tilde{N}\}$, and the values $\alpha(l, r, q)$ are distinct across all $(l, r, q) \in \mathcal{Q}$. Extending the range of summation of each $\alpha(l, r, q)$ to all of $\{1, \ldots, \tilde{N}\}$ simply adds 0 by the definition of $\tilde{M}_\ell$, so we finally obtain

$$D(\mathcal{Q}) = \sum_{\alpha_1,\ldots,\alpha_Q}^{*} \prod_{\ell=1}^{L} \tilde{M}_\ell[\alpha_{\mathfrak{q}(\ell)}, \alpha_{\mathfrak{q}'(\ell)}] \tag{B.12}$$

where $Q = |\mathcal{Q}|$ and the sum is over all tuples of Q distinct indices in $\{1, \ldots, \tilde{N}\}$.

We must bound $|D(\mathcal{Q})|$ for any $\mathcal{Q}$ such that $E(\mathcal{Q}) \neq 0$. By Lemma B.3(b) and the expression (B.8) for $E(\mathcal{Q})$, if $E(\mathcal{Q}) \neq 0$, then for each $(l, r) \in \mathcal{R}$ and each index value $i \in \{1, \ldots, N_r\}$, there must be an even number of indices in $\mathcal{I}(l, r)$ equal in value to i, i.e. each element $S \in \mathcal{Q}$ must have even cardinality. Furthermore, if exactly two indices in $\mathcal{I}(l, r)$ equal i, then they must both be row indices or both be column indices for $O_{l,r}$. In particular, if $S \in \mathcal{Q}$ has cardinality $|S| = 2$, and if $S = \{j_k, i_{k+1}\}$ or $S = \{j'_k, i'_{k+1}\}$, then this implies $(l_k, r_k, e_k) = (l_{k+1}, r_{k+1}, -e_{k+1})$. The condition of the lemma ensures in this case that $\operatorname{Tr} D_k = 0$, so also $\operatorname{Tr} \tilde{M}_k = \operatorname{Tr} \tilde{M}_{K+k} = 0$.

We pause to formulate a lemma which provides the bound for $|D(\mathcal{Q})|$ that we need.

Lemma B.4. *Fix integers $L, Q \geq 1$ and a constant $B > 0$. Let $\mathfrak{q}, \mathfrak{q}' : \{1, \ldots, L\} \to \{1, \ldots, Q\}$ be two fixed maps. Let $M_1, \ldots, M_L \in \mathbb{C}^{N\times N}$ be such that $\|M_l\| \leq B$ for all l. Call an index $q \in \{1, \ldots, Q\}$ "good" if both of the following hold:*

- *Exactly two of* $\mathfrak{q}(1), \ldots, \mathfrak{q}(L), \mathfrak{q}'(1), \ldots, \mathfrak{q}'(L)$ *are equal to* q.
- *If* $\mathfrak{q}(\ell) = \mathfrak{q}'(\ell) = q$ *for some* ℓ, *then* $\operatorname{Tr} M_\ell = 0$.

Let T *be the number of good indices* $q \in Q$.

Denote by $\sum^*_{\alpha_1,\ldots,\alpha_Q}$ *the sum over all tuples of* Q *indices* $\alpha_1, \ldots, \alpha_Q \in \{1, \ldots, N\}$ *with all values distinct. Then, for some constant* $C := C(L, Q, B) > 0$,

$$\left| \sum^*_{\alpha_1,\ldots,\alpha_Q} \prod_{\ell=1}^L M_\ell[\alpha_{\mathfrak{q}(\ell)}, \alpha_{\mathfrak{q}'(\ell)}] \right| \leq C N^{Q-T/2}. \tag{B.13}$$

Assuming this lemma for now, we can complete the proof of Lemma B.2. We saw that any $S \in \mathcal{Q}$ of cardinality $|S| = 2$ is good, for if $S = \{\mathfrak{q}(\ell), \mathfrak{q}'(\ell)\}$, then either $S = \{j_k, i_{k+1}\}$ or $S = \{j'_k, i'_{k+1}\}$ and so $\operatorname{Tr} \tilde{M}_\ell = 0$. Letting T be the number of elements of $\mathcal{Q}$ with cardinality 2, we have $2T + 4(Q - T) \leq 4K$. But T is also the number of good indices q, so Lemma (B.13) implies

$$|D(\mathcal{Q})| \leq C \tilde{N}^{Q-T/2} \leq C \tilde{N}^K. \tag{B.14}$$

Noting that $\tilde{N}/N$ and the number of distinct partitions $\mathcal{Q}$ are also both bounded by a K-dependent constant, and combining with (B.11), we obtain $\mathcal{E} \leq C N^{-2}$ as desired, and hence Lemma B.2. $\square$

Proof of Lemma B.4. Denote $[L] = \{1, \ldots, L\}$ and $[Q] = \{1, \ldots, Q\}$. We will show the following claim by induction on t: For any $L, Q \geq 1$ and $B > 0$, if the number of good indices T satisfies $T \geq t$, then there exists a constant $C := C(L, Q, B, t) > 0$ for which

$$\left| \sum^*_{\alpha_1,\ldots,\alpha_Q} \prod_{l=1}^L M_l[\alpha_{\mathfrak{q}(l)}, \alpha_{\mathfrak{q}'(l)}] \right| \leq C N^{Q-t/2}. \tag{B.15}$$

The desired result follows from this claim applied with $t = T$ and $C = \max_{t=0}^Q C(L, Q, B, t)$.

For the base case $t = 0$, the left side of (B.15) is bounded by CN^Q for $C = B^L$, regardless of T, as each entry of M_l is bounded by B.

For the inductive step, let $t \geq 1$, suppose the number T of good indices satisfies $T \geq t$, and suppose the inductive claim holds for $t-1, t-2, \ldots, 0$. We consider two cases corresponding to the two possibilities for goodness of an index q:

Case 1: There exists a good index q and some $l \in [L]$ such that $\mathfrak{q}(l) = \mathfrak{q}'(l) = q$ and $\operatorname{Tr} M_l = 0$. For notational convenience, assume without loss of generality that $q = Q$ and $l = L$. Summing first over $\alpha_1, \ldots, \alpha_{Q-1}$ and then over α_Q, and noting that no other $\mathfrak{q}(l)$ or $\mathfrak{q}'(l)$ equals Q for $l \leq L-1$ because Q is good, the left side of (B.15) may be written as

$$\mathrm{LS} := \left| \sum^*_{\alpha_1,\ldots,\alpha_{Q-1}} \left(\prod_{l=1}^{L-1} M_l[\alpha_{\mathfrak{q}(l)}, \alpha_{\mathfrak{q}'(l)}] \right) \sum_{\substack{\alpha_Q=1 \\ \alpha_Q \notin \{\alpha_1,\ldots,\alpha_{Q-1}\}}}^N M_L[\alpha_Q, \alpha_Q] \right|.$$

Then applying $\operatorname{Tr} M_L = 0$, if $Q = 1$, then LS vanishes and there is nothing further to do. If $Q > 1$, we get

$$\mathrm{LS} = \left| \sum^*_{\alpha_1,\ldots,\alpha_{Q-1}} \left(\prod_{l=1}^{L-1} M_l[\alpha_{\mathfrak{q}(l)}, \alpha_{\mathfrak{q}'(l)}] \right) \sum_{\alpha_Q \in \{\alpha_1,\ldots,\alpha_{Q-1}\}} M_L[\alpha_Q, \alpha_Q] \right|$$

$$\leq \sum_{k=1}^{Q-1} \left| \sum_{\alpha_1,\ldots,\alpha_{Q-1}}^{*} \left(\prod_{l=1}^{L-1} M_l[\alpha_{\mathfrak{q}(l)}, \alpha_{\mathfrak{q}'(l)}] \right) M_L[\alpha_k, \alpha_k] \right|.$$

We may apply the induction hypothesis to each of the $Q-1$ terms of the above sum: Define $\tilde{\mathfrak{q}}, \tilde{\mathfrak{q}}' : [L] \to [Q-1]$ by $\tilde{\mathfrak{q}}(l) = \mathfrak{q}(l)$ and $\tilde{\mathfrak{q}}'(l) = \mathfrak{q}'(l)$ for $l \in [L-1]$ and $\tilde{\mathfrak{q}}(L) = \tilde{\mathfrak{q}}'(L) = k$. Each $q \in [Q-1]$ that was good for i, j remains good for $\tilde{i}, \tilde{j}$, except possibly $q = k$. Thus the number of good indices for $\tilde{\mathfrak{q}}, \tilde{\mathfrak{q}}'$ is at least $\check{t} := \max(t-2, 0)$. The induction hypothesis implies

$$\text{LS} \leq (Q-1) \cdot C(L, Q-1, B, \check{t}) N^{Q-1-\check{t}/2} \leq (Q-1) \cdot C(L, Q-1, B, \check{t}) N^{Q-t/2}.$$

Case 2: There exists a good index q and distinct $l \neq l' \in [L]$ such that one of $\mathfrak{q}(l), \mathfrak{q}'(l)$ and one of $\mathfrak{q}(l'), \mathfrak{q}'(l')$ equal q. For notational convenience, assume without loss of generality that $q = Q$, $l = L-1$, and $l' = L$. By possibly replacing M_{L-1} and/or M_L by M_{L-1}^T and/or M_L^T, we may further assume $\mathfrak{q}'(L-1) = \mathfrak{q}(L) = Q$.

Summing first over $\alpha_1, \ldots, \alpha_{Q-1}$ and then over α_Q as in Case 1, and noting that no $\mathfrak{q}(l)$ or $\mathfrak{q}'(l)$ equals Q for $l \leq L-2$ because Q is good, the left side of (B.15) may be written as

$$\text{LS} := \left| \sum_{\alpha_1,\ldots,\alpha_{Q-1}}^{*} \left(\prod_{l=1}^{L-2} M_l[\alpha_{\mathfrak{q}(l)}, \alpha_{\mathfrak{q}'(l)}] \right) \sum_{\substack{\alpha_Q=1 \\ \alpha_Q \notin \{\alpha_1,\ldots,\alpha_{Q-1}\}}}^{N} M_{L-1}[\alpha_{\mathfrak{q}(L-1)}, \alpha_Q] M_L[\alpha_Q, \alpha_{\mathfrak{q}'(L)}] \right|.$$

Define $M = M_{L-1} M_L$. Then $\|M\| \leq B^2$, and

$$\begin{aligned}
\text{LS} = \Bigg| \sum_{\alpha_1,\ldots,\alpha_{Q-1}}^{*} \left(\prod_{l=1}^{L-2} M_l[\alpha_{\mathfrak{q}(l)}, \alpha_{\mathfrak{q}'(l)}] \right) \Bigg(& M[\alpha_{\mathfrak{q}(L-1)}, \alpha_{\mathfrak{q}'(L)}] \\
& - \sum_{\alpha_Q \in \{\alpha_1,\ldots,\alpha_{Q-1}\}} M_{L-1}[\alpha_{\mathfrak{q}(L-1)}, \alpha_Q] M_L[\alpha_Q, \alpha_{\mathfrak{q}'(L)}] \Bigg) \Bigg| \\
\leq \left| \sum_{\alpha_1,\ldots,\alpha_{Q-1}}^{*} \left(\prod_{l=1}^{L-2} M_l[\alpha_{\mathfrak{q}(l)}, \alpha_{\mathfrak{q}'(l)}] \right) M[\alpha_{\mathfrak{q}(L-1)}, \alpha_{\mathfrak{q}'(L)}] \right| & \\
+ \sum_{k=1}^{Q-1} \left| \sum_{\alpha_1,\ldots,\alpha_{Q-1}}^{*} \left(\prod_{l=1}^{L-2} M_l[\alpha_{\mathfrak{q}(l)}, \alpha_{\mathfrak{q}'(l)}] \right) M_{L-1}[\alpha_{\mathfrak{q}(L-1)}, \alpha_k] M_L[\alpha_k, \alpha_{\mathfrak{q}'(L)}] \right| &.
\end{aligned}$$

We may again apply the induction hypothesis to each term of the above sum: For the first term, each original good index $q \in [Q-1]$ remains good, except possibly $k := \mathfrak{q}(L-1) = \mathfrak{q}'(L)$ if k was originally good but now $\operatorname{Tr} M \neq 0$. Hence for this first term there are still at least $\check{t} := \max(t-2, 0)$ good indices. The other $Q-1$ terms are present only if $Q > 1$. For each of these terms, each original good index $q \in [Q-1]$ remains good, except possibly $q = k$—hence there are also at least $\check{t}$ good indices. Then the induction hypothesis yields, similarly to Case 1,

$$\text{LS} \leq \Big(C(L-1, Q-1, B^2, \check{t}) + (Q-1) \cdot C(L, Q-1, B, \check{t}) \Big) N^{Q-t/2}.$$

This concludes the induction in both cases, upon setting $C(L, Q, B, t) = C(L-1, Q-1, B^2, \check{t}) + (Q-1) \cdot C(L, Q-1, B, \check{t})$. $\square$

This concludes the proof of Theorem 3.9. Finally, we prove Corollary 3.10 which establishes the approximation at the level of Stieltjes transforms.

PROOF OF COROLLARY 3.10. Under the given conditions, there exists a constant $C_0 > 0$ such that $|\tau(w^l)| \leq C_0^l$ for all N and $l \geq 0$, and also $|N^{-1} \operatorname{Tr} W^l| \leq \|W\|^l \leq C_0^l$ a.s. for all $l \geq 0$ and all sufficiently large N. Fix $z \in \mathbb{C}^+$ with $|z| > C_0$. Then $m_w(z) = -\sum_{l=0}^{\infty} z^{-(l+1)}\tau(w^l)$ and $m_W(z) = -N^{-1} \operatorname{Tr}(z - W)^{-1} = -\sum_{l=0}^{\infty} z^{-(l+1)} N^{-1} \operatorname{Tr} W^l$ define convergent series for all large N. For any $\varepsilon > 0$, there exists L such that

$$\left|\sum_{l=L+1}^{\infty} z^{-(l+1)} N^{-1} \operatorname{Tr} W^l\right| < \varepsilon, \qquad \left|\sum_{l=L+1}^{\infty} z^{-(l+1)} \tau(w^l)\right| < \varepsilon$$

for all large N, while by Theorem 3.9, as $N \to \infty$

$$\left|\sum_{l=0}^{L} z^{-(l+1)} N^{-1} \operatorname{Tr} W^l - z^{-(l+1)}\tau(w^l)\right| \to 0.$$

Hence $\limsup_{N\to\infty} |m_W(z) - m_w(z)| \leq 2\varepsilon$ a.s., and the result follows by taking $\varepsilon \to 0$. $\square$

APPENDIX C: ANALYSIS OF FIXED-POINT EQUATIONS

We analyze the fixed-point equations (4.1–4.2) and conclude the proof of the main result, Theorem 4.1. The analysis follows arguments similar to those in [CDS11] and [DL11].

LEMMA C.1 ([CL11]). *Let $\Omega \subseteq \mathbb{C}$ be a connected open set, let $E \subseteq \Omega$ be any set with an accumulation point in Ω, let $a, b \in \mathbb{C}$ be any two distinct fixed values, and let $\{f_n\}_{n=1}^{\infty}$ be a sequence of analytic functions $f_n : \Omega \to \mathbb{C}$. If $f_n(z) \notin \{a, b\}$ for all $z \in \Omega$ and $n \geq 1$, and if $\lim_{n\to\infty} f_n(z)$ exists (and is finite) for each $z \in E$, then $\{f_n\}_{n=1}^{\infty}$ converges uniformly on compact subsets of Ω to an analytic function.*

PROOF. The result is originally due to [CL11]. It also follows by the theory of normal families: $\{f_n\}_{n=1}^{\infty}$ is a normal family by Montel's fundamental normality test, see e.g. [Sch13, Section 2.7]. Hence every subsequence has a further subsequence that converges uniformly on compact sets to an analytic function. All such analytic functions must coincide on E, hence they coincide on all of Ω by uniqueness of analytic extensions, implying the desired result. $\square$

In the notation of Theorem 4.1, denote $a = (a_1, \ldots, a_k)$, $b = (b_1, \ldots, b_k)$,

$$f_r(z, b) = -\frac{1}{n_r} \operatorname{Tr}\left((z \operatorname{Id}_p + b \cdot H^*H)^{-1} H_r^* H_r\right),$$
$$g_r(a) = -\frac{1}{n_r} \operatorname{Tr}_r\left([\operatorname{Id}_{n_+} + FD(a)]^{-1} F\right).$$

LEMMA C.2. *Under the conditions of Theorem 4.1:*

(a) For all $z \in \mathbb{C}^+$ and $b \in (\overline{\mathbb{C}^+})^k$, $z \operatorname{Id}_p + b \cdot H^*H$ is invertible, $f_r(z, b) \in \mathbb{C}^+ \cup \{0\}$, and $m_0(z) \in \mathbb{C}^+$ for m_0 as defined by (4.3).

(b) For all $a \in (\mathbb{C}^+ \cup \{0\})^k$, $\operatorname{Id}_{n_+} + FD(a)$ is invertible and $g_r(a) \in \overline{\mathbb{C}^+}$.

Proof of Lemma C.2. For any $v \in \mathbb{C}^p$,

$$\Im\left[v^*(z\operatorname{Id}_p + b \cdot H^*H)v\right] = (\Im z)v^*v + \sum_s (\Im b_s)v^*H_s^*H_s v > 0.$$

Hence $z\operatorname{Id}_p + b \cdot H^*H$ is invertible. Letting $T = (z\operatorname{Id}_p + b \cdot H^*H)^{-1}$,

$$\begin{aligned} n_r f_r(z,b) &= -\operatorname{Tr} TH_r^*H_r = -\operatorname{Tr} TH_r^*H_rT^*\left(z\operatorname{Id}_p + b \cdot H^*H\right)^* \\ &= -\overline{z}\operatorname{Tr} TH_r^*H_rT^* - \sum_{s=1}^k \overline{b_s}\operatorname{Tr} TH_r^*H_rT^*H_s^*H_s. \end{aligned}$$

As $\operatorname{Tr} TRT^*S$ is real and nonnegative for any Hermitian positive-semidefinite matrices R and S, the above implies $\Im f_r(z,b) \geq 0$. In fact, as $\operatorname{Tr} TH_r^*H_rT^* > 0$ unless $H_r = 0$, either $\Im f_r(z,b) > 0$ or $f_r(z,b) = 0$. Similarly,

$$pm_0(z) = -\operatorname{Tr} T = -\overline{z}\operatorname{Tr} TT^* - \sum_{s=1}^k \overline{b_s}\operatorname{Tr} TT^*H_s^*H_s,$$

and as $\operatorname{Tr} TT^* > 0$, $\Im m_0(z) > 0$. This establishes (a).

For (b), let us first show $\operatorname{Id}_{n_+} + FD(a)$ is invertible. Note if $a_1 = 0$, then by the fact that a block matrix

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

is invertible if and only if A and C are invertible, it suffices to show invertibility of the lower-right $(n_2 + \ldots + n_k) \times (n_2 + \ldots + n_k)$ submatrix. Hence we may reduce to the case where $a_s \neq 0$, i.e. $a_s \in \mathbb{C}^+$, for all s. Suppose $\operatorname{rank}(F) = m$ and let $F^\dagger$ denote the pseudo-inverse of F, so that $FF^\dagger$ is a projection matrix of rank m onto the column span of F. $F^\dagger$ is Hermitian, since F is. Let Q denote the projection orthogonal to $FF^\dagger$, of rank $n_+ - m$. Then

$$\operatorname{Id}_{n_+} + FD(a) = Q + F(F^\dagger + D(a)).$$

For each $s = 1, \ldots, k$, let P_s be the projection of rank n_s such that $D(a) = \sum_{s=1}^k a_s P_s$. Then for any $v \in \mathbb{C}^{n_+}$,

$$\Im[v^*(F^\dagger + D(a))v] = \Im[v^*D(a)v] = \sum_s (\Im a_s)v^*P_s v > 0,$$

as $v^*F^\dagger v$ and $v^*P_s v$ are real and $\Im a_s > 0$ for each s. Hence $F^\dagger + D(a)$ is invertible, so $\operatorname{Id}_{n_+} + FD(a)$ is of full column rank and thus also invertible.

For the second claim, supposing momentarily that F is invertible and letting $M = (F^{-1} + D(a))^{-1}$,

$$\begin{aligned} n_r g_r(a) &= -\operatorname{Tr}_r M = -\operatorname{Tr}_r\left(M\left(F^{-1} + \sum_{s=1}^k a_s P_s\right)^* M^*\right) \\ &= -\operatorname{Tr} P_r MF^{-1}M^* - \sum_{s=1}^k \overline{a_s}\operatorname{Tr} P_r MP_s M^*. \end{aligned}$$

As $\operatorname{Tr} P_r MF^{-1}M^*$ is real and $\operatorname{Tr} P_r MP_s M^*$ is real and nonnegative, this implies $\Im g_r(a) \geq 0$. By continuity in F, this must hold also when F is not invertible, establishing (b). $\square$

Lemma C.3. *Let $C, M > 0$ and let $\mathcal{S}$ denote the space of k-tuples $b = (b_1, \ldots, b_k)$ such that each b_r is an analytic function $b_r : \mathbb{C}^+ \to \overline{\mathbb{C}^+}$ and $\sup_{z \in \mathbb{C}^+ : \Im z > M} \|b(z)\| \leq C$. For sufficiently large C and M (depending on p, n_r, m_r and the matrices H_r and $F_{r,s}$ in Theorem 4.1):*

(a) $\rho : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ defined by

$$\rho(b, \tilde{b}) := \sup_{z \in \mathbb{C}^+ : \Im z > M} \|b(z) - \tilde{b}(z)\|$$

is a complete metric on $\mathcal{S}$, and

(b) Letting $g = (g_1, \ldots, g_k)$ and $f = (f_1, \ldots, f_k)$ where g_r and f_r are as above, $b \mapsto g(f(z, b))$ defines a map from $\mathcal{S}$ to itself, and there exists $c \in (0, 1)$ such that for all $b, \tilde{b} \in \mathcal{S}$,

$$\rho(g(f(z, b)), g(f(z, \tilde{b}))) \leq c\rho(b, \tilde{b}).$$

Proof. For part (a), ρ is clearly nonnegative, symmetric, and satisfies the triangle inequality. By definition of $\mathcal{S}$, $\rho(b, \tilde{b}) < \infty$ for all $b, \tilde{b} \in \mathcal{S}$. By uniqueness of analytic extensions, $\rho(b, \tilde{b}) = 0 \Leftrightarrow b = \tilde{b}$, hence ρ is a metric. If $\{b^{(l)}\}_{l=1}^\infty$ is a Cauchy sequence in $(\mathcal{S}, \rho)$, then for each $z \in \mathbb{C}^+$ with $\Im z > M$, $\{b^{(l)}(z)\}_{l=1}^\infty$ is Cauchy in $(\overline{\mathbb{C}^+})^k$ and hence converges to some $b(z) = (b_1(z), \ldots, b_k(z)) \in (\overline{\mathbb{C}^+})^k$. Then Lemma C.1 implies each $b_r(z)$ has an analytic extension to all of $\mathbb{C}^+$, and $b_r^{(l)} \to b_r$ uniformly over compact subsets of $\mathbb{C}^+$. This implies $b_r(z) \in \overline{\mathbb{C}^+}$ for all $z \in \mathbb{C}^+$ and $\sup_{z \in \mathbb{C}^+ : \Im z > M} \|b(z)\| \leq C$, so $b \in \mathcal{S}$. Furthermore $\rho(b^{(l)}, b) \to 0$, hence $(\mathcal{S}, \rho)$ is complete.

For part (b), clearly if $b = (b_1, \ldots, b_k)$ is a k-tuple of analytic functions on $\mathbb{C}^+$, then $g(f(z, b))$ is as well. Now consider $z \in \mathbb{C}^+$ with $\Im z > M$ and fixed values $b \in (\overline{\mathbb{C}^+})^k$ with $\|b\| \leq C$, and define

$$T = (z \operatorname{Id}_p + b \cdot H^* H)^{-1}, \qquad R = \left(\operatorname{Id}_{n_+} + FD(f(z, b))\right)^{-1}, \tag{C.1}$$

where invertibility of these quantities follows from Lemma C.2. Since $H_s^* H_s$ is positive-semidefinite, [CDS11, Lemma 8] implies $\|T\| \leq (\Im z)^{-1}$. Then if $C, M > 0$ (depending on $p, n_r, m_r, H_r, F_{r,s}$) are sufficiently large, we have $|f_r(z, b)| \leq C(\Im z)^{-1}$, $\|FD(f(z, b))\| < 1/2$, $\|R\| < 2$, and $\|g(f(z, b))\| \leq C$. This establishes that for sufficiently large $C, M > 0$, if $b \in \mathcal{S}$, then $g(f(z, b)) \in \mathcal{S}$.

Next, consider also $\tilde{b} \in (\overline{\mathbb{C}^+})^k$ with $\|\tilde{b}\| \leq C$, and define $\tilde{T}$ and $\tilde{R}$ by (C.1) with $\tilde{b}$ in place of b. For each $s = 1, \ldots, k$, let P_s be the projection such that $D(a) = \sum_{s=1}^k a_s P_s$. Then by the matrix identity $A^{-1} - (A + E)^{-1} = A^{-1} E (A + E)^{-1}$,

$$\begin{aligned}
f_r(z, b) - f_r(z, \tilde{b}) &= \frac{1}{n_r} \operatorname{Tr}\left(\tilde{T}(T^{-1} - \tilde{T}^{-1}) T H_r^* H_r\right) \\
&= \frac{1}{n_r} \sum_{s=1}^k (b_s - \tilde{b}_s) \operatorname{Tr}\left(\tilde{T} H_s^* H_s T H_r^* H_r\right), \\
g_r(f(z, b)) - g_r(f(z, \tilde{b})) &= \frac{1}{n_r} \operatorname{Tr} P_r \tilde{R}(R^{-1} - \tilde{R}^{-1}) R F \\
&= \frac{1}{n_r} \sum_{s=1}^k (f_s(z, b) - f_s(z, \tilde{b})) \operatorname{Tr} P_r \tilde{R} F P_s R F.
\end{aligned}$$

Then $g(f(z, b)) - g(f(z, \tilde{b})) = M^{(2)} M^{(1)} (b - \tilde{b})$ for the matrices $M^{(1)}, M^{(2)} \in \mathbb{C}^{k \times k}$ having entries

$$M_{rs}^{(1)} = \frac{1}{n_r} \operatorname{Tr}\left(\tilde{T} H_s^* H_s T H_r^* H_r\right), \quad M_{rs}^{(2)} = \frac{1}{n_r} \operatorname{Tr} P_r \tilde{R} F P_s R F.$$

For sufficiently large $C, M > 0$, we have $\|T\| \leq (\Im z)^{-1}$, $\|\tilde{T}\| \leq (\Im z)^{-1}$, $\|M^{(1)}\| \leq C(\Im z)^{-2}$, $\|R\| < 2$, $\|\tilde{R}\| < 2$, and $\|M^{(2)}\| \leq C$, hence $\|M^{(2)} M^{(1)}\| \leq C^2 (\Im z)^{-2} \leq C^2 M^{-2}$. Increasing M if necessary so that $C^2 M^{-2} < 1$, this yields part (b). $\square$

We conclude the proof of Theorem 4.1 using these lemmas, Corollary 3.10, and Lemma 4.4.

Proof of Theorem 4.1. Let $C, M > 0$ be (p, n_r, m_r-dependent values) such that the conclusions of Lemma C.3 hold. Increasing C if necessary, assume $\|b^{(0)}\| < C$ where $b^{(0)} = (b_1^{(0)}, \ldots, b_k^{(0)})$ are the initial values for the iterative procedure of part (c). Lemma C.3 and the Banach fixed point theorem imply the existence of a unique point $b \in \mathcal{S}$ such that $g(f(z, b)) = b$. Defining $a = f(z, b)$, Lemma C.2 implies $a \in (\mathbb{C}^+ \cup \{0\})^k$ for each $z \in \mathbb{C}^+$. Then a_r, b_r satisfy (4.1) and (4.2) for each $z \in \mathbb{C}^+$ by construction, which verifies existence in part (a). For part (c), define the constant functions $\tilde{b}_r^{(0)}(z) \equiv b_r^{(0)}$ over $z \in \mathbb{C}^+$. Then $\tilde{b}^{(0)} := (\tilde{b}_1^{(0)}, \ldots, \tilde{b}_r^{(0)}) \in \mathcal{S}$. Define iteratively $\tilde{b}^{(t+1)} = g(f(z, \tilde{b}^{(t)}))$. Then Lemma C.3 implies

$$c\rho(b, \tilde{b}^{(t)}) \geq \rho(g(f(z, b)), g(f(z, \tilde{b}^{(t)}))) = \rho(b, \tilde{b}^{(t+1)}),$$

for b the above fixed point and some $c \in (0, 1)$. Hence $\rho(b, \tilde{b}^{(t)}) \to 0$ as $t \to \infty$. This implies by Lemma C.1 that $\tilde{b}^{(t)}(z) \to b(z)$ for all $z \in \mathbb{C}^+$, which establishes part (c) upon noting that $\tilde{b}_r^{(t)}(z)$ is exactly the value $b_r^{(t)}$ of the iterative procedure applied at z. Part (c) implies uniqueness in part (a), since $(b_1^{(t)}, \ldots, b_k^{(t)})$ would not converge to $(b_1, \ldots, b_k)$ if this iterative procedure were initialized to a different fixed point. For part (b), Lemma C.2 verifies that $m_0(z) \in \mathbb{C}^+$ for $z \in \mathbb{C}^+$. As $b_1(z), \ldots, b_k(z)$ are analytic, $m_0(z)$ is also analytic. Furthermore, as $b \in \mathcal{S}$, $b_1(z), \ldots, b_k(z)$ remain bounded as $\Im z \to \infty$, so $m_0(z) \sim -z^{-1}$ as $\Im z \to \infty$. Then m_0 defines the Stieltjes transform of a probability measure μ_0 by [GH03, Lemma 2].

It remains to verify that μ_0 approximates μ_W. Let $f_{rs}, g_r, h_r \in \mathcal{A}$ be the free deterministic equivalent constructed by Lemma 4.2, and let $N = p + \sum_r m_r + \sum_r n_r$. Uniqueness of the solution a_r, b_r to (4.1) and (4.2) in the stated domains implies that the analytic functions $a_1, \ldots, a_k, b_1, \ldots, b_k$ in Lemma 4.4 must coincide with this solution for $z \in \mathbb{D}$. Then Lemma 4.4 implies, for $z \in \mathbb{D}$,

$$m_w(z) := \tau((w - z)^{-1}) = \frac{p}{N} m_0(z) - \frac{N - p}{Nz}.$$

The conditions of Corollary 3.10 are satisfied by Lemma 4.2, so Corollary 3.10 implies $m_{\tilde{W}}(z) - m_w(z) \to 0$ as $p, n_r, m_r \to \infty$, pointwise a.s. over $\mathbb{D}$, where $\tilde{W} \in \mathbb{C}^{N \times N}$ is the embedding of W and $m_{\tilde{W}}$ is its empirical spectral measure. As

$$m_{\tilde{W}}(z) = \frac{p}{N} m_W(z) - \frac{N - p}{Nz},$$

we have $m_W(z) - m_0(z) \to 0$ pointwise a.s. over $\mathbb{D}$. As $m_W - m_0$ is uniformly bounded over $\{z \in \mathbb{C}^+ : \Im z > \varepsilon\}$ for any $\varepsilon > 0$, Lemma C.1 implies $m_W(z) - m_0(z) \to 0$ pointwise a.s. for $z \in \mathbb{C}^+$. Hence $\mu_W - \mu_0 \to 0$ vaguely a.s. (see, e.g., [BS10, Theorem B.9]). By the conditions of the theorem and [YBK88], $\|W\|$ is almost surely bounded by a constant for all large p, n_r, m_r. Furthermore, by Lemma 4.2, we have $\tau(w^l) \leq \|w\|^l \leq C^l$ for some constant $C > 0$ and all $l \geq 0$, so m_w and m_0 are Stieltjes transforms of probability measures with bounded support. Then the convergence $\mu_W - \mu_0 \to 0$ holds weakly a.s., concluding the proof of the theorem. ☐

APPENDIX D: FREE PROBABILITY CONSTRUCTIONS

We construct the spaces $(\mathcal{A}, \tau, p_1, \ldots, p_d)$ in Examples 3.5, 3.6, 3.7, and point the reader to the relevant references that establish Lemma 4.2.

Recall that a von Neumann algebra $\mathcal{A}$ is a sub-$*$-algebra of the space of bounded linear operators $B(H)$ acting on a Hilbert space H, such that $\mathcal{A}$ is σ-weakly closed and contains the identity.

The trace τ is positive, faithful, and normal if $\tau(a^*a) \geq 0$ for all $a \in \mathcal{A}$, $\tau(a^*a) = 0$ only if $a = 0$, and τ is σ-weakly continuous. (See I.9.1.2 and III.2.1.4 of [Bla06] for equivalent topological characterizations.) $\mathcal{B}$ is a von Neumann sub-algebra of $\mathcal{A}$ if it is algebraically and σ-weakly closed.

If $\mathcal{A}$ is a von Neumann algebra with positive, faithful, and normal trace τ, and $\mathcal{B}$ is a von Neumann sub-algebra of $\mathcal{A}$, then there is a unique τ-invariant conditional expectation $\mathbf{F}^{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$; furthermore $\mathbf{F}^{\mathcal{B}}$ is both σ-weakly and norm-continuous, satisfying $\|\mathbf{F}^{\mathcal{B}}(a)\| \leq \|a\|$ (see [Kad04, Theorem 7 and Proposition 1]). The tower property (3.3) follows from this uniqueness, as both $\mathbf{F}^{\mathcal{D}}$ and $\mathbf{F}^{\mathcal{D}} \circ \mathbf{F}^{\mathcal{B}}$ are τ-invariant conditional expectations onto $\mathcal{D}$.

Lemma D.1. *Rectangular probability spaces $(\mathcal{A}, \tau, p_1, \ldots, p_d)$ satisfying the properties of Examples 3.5, 3.6, and 3.7 exist, such that in each example, $\mathcal{A}$ is a von Neumann algebra and τ is a positive, normal, and faithful trace.*

Proof. In Examples 3.5 and 3.6, let $(\Omega, \mathbb{P})$ be a (classical) probability space and let $\mathcal{A}$ be the von Neumann algebra of $d \times d$ random matrices with entries in $L^\infty(\Omega, \mathbb{P})$, the bounded complex-valued random variables on Ω. ($\mathcal{A}$ acts on the Hilbert space H of length-d random vectors with elements in $L^2(\Omega, \mathbb{P})$, endowed with inner-product $v, w \mapsto \mathbb{E}\langle v, w\rangle$.) Defining $\tau(a) = N^{-1}\mathbb{E}[\sum_{r=1}^d N_r a_{rr}]$, τ is a positive and faithful trace. As $a \mapsto \mathbb{E}[a_{rr}]$ is weakly continuous and hence σ-weakly continuous for each $r = 1, \ldots, d$, τ is normal. Letting $p_r \in \mathcal{A}$ be the (deterministic) matrix with (r, r) entry 1 and remaining entries 0, $(\mathcal{A}, \tau, p_1, \ldots, p_d)$ is a rectangular probability space, and $\tau(p_r) = N_r/N$ for each $r = 1, \ldots, d$. For Example 3.5, the element $g \in \mathcal{A}$ may be realized as the random matrix with (r, r) entry equal to X and all other entries 0, where $X \in L^\infty(\Omega, \mathbb{P})$ is a random variable with standard semi-circle distribution on $[-2, 2]$. For Example 3.6, the element $g \in \mathcal{A}$ may be realized as the matrix with (r_1, r_2) entry equal to X and all other entries 0, where $X \in L^\infty(\Omega, \mathbb{P})$ is the square root of a random variable having the Marcenko-Pastur distribution (3.6) with $\lambda = N_{r_2}/N_{r_1}$.

For Example 3.7, we may simply take $(\mathcal{A}, \tau, p_1, \ldots, p_d)$ to be the rectangular probability space of deterministic $N \times N$ matrices from Example 3.1. ($\mathcal{A}$ is the space $B(H)$ for $H = \mathbb{C}^N$, and τ is clearly positive, faithful, and normal as H is finite-dimensional.) We may take the elements $b_1, \ldots, b_k \in \mathcal{A}$ to be the original matrices $B_1, \ldots, B_k$. ☐

The sub-$*$-algebras $\mathcal{D}$ in the three examples above are isomorphic. They are also finite-dimensional, hence σ-weakly closed, so each is a von Neumann sub-algebra of $\mathcal{A}$.

Proof of Lemma 4.2. For each $r = 1, \ldots, k$, let $(\mathcal{A}^{(r)}, \tau^{(r)}, p_0, \ldots, p_{2k})$ be the space constructed as in Lemma D.1 corresponding to Example 3.6 and containing the element g_r, satisfying conditions 1, 2, and 4. Let $(\mathcal{A}^{(k+1)}, \tau^{(k+1)}, p_0, \ldots, p_{2k})$ and $(\mathcal{A}^{(k+2)}, \tau^{(k+2)}, p_0, \ldots, p_{2k})$ be the spaces constructed as in Lemma D.1 corresponding to Example 3.7 and containing the families $\{h_r\}$ and $\{f_{rs}\}$, respectively, satisfying conditions 1, 2, and 3. $\mathcal{D} = \langle p_0, \ldots, p_{2k}\rangle$ is a common (up to isomorphism) $(2k+1)$-dimensional von Neumann sub-algebra of each $\mathcal{A}^{(r)}$, and each $\tau^{(r)}$ restricts to the same trace on $\mathcal{D}$. Then the construction of the finite von Neumann amalgamated free product of $(\mathcal{A}^{(1)}, \tau^{(1)}), \ldots, (\mathcal{A}^{(k+2)}, \tau^{(k+2)})$ with amalgamation over $\mathcal{D}$ [Voi85, Pop93] yields a von Neumann algebra $\mathcal{A}$ with a positive, faithful, and normal trace τ such that:

- $\mathcal{A}$ contains (as an isomorphically embedded von Neumann sub-algebra) each $\mathcal{A}^{(r)}$, where $\mathcal{A}^{(r)}$ contains the common sub-algebra $\mathcal{D}$.
- Letting $\mathbf{F} : \mathcal{A} \to \mathcal{D}$ and $\mathbf{F}^{(r)} : \mathcal{A}^{(r)} \to \mathcal{D}$ denote the τ-invariant and $\tau^{(r)}$-invariant conditional expectations, $\mathbf{F}|_{\mathcal{A}^{(r)}} \equiv \mathbf{F}^{(r)}$.
- $\tau = \tau^{(r)} \circ \mathbf{F}$ for any r, so in particular, $\tau|_{\mathcal{A}^{(r)}} = \tau^{(r)}$.

- The sub-algebras $\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(k+2)}$ of $\mathcal{A}$ are free with amalgamation over $\mathcal{D}$ in the $\mathcal{D}$-valued probability space $(\mathcal{A}, \mathcal{D}, \mathbf{F})$.

(For more details about the amalgamated free product construction, see the Introduction of [Dyk95] and also Section 3.8 of [VDN92].) Since τ restricts to $\tau^{(r)}$ on each $\mathcal{A}^{(r)}$, conditions 1–4 continue to hold for the elements p_r, f_{rs}, g_r, h_r in $\mathcal{A}$. The generated von Neumann algebra $\langle D, g_r \rangle_{W^*}$ is contained in $\mathcal{A}^{(r)}$ and similarly for $\langle D, h_1, \ldots, h_k \rangle_{W^*}$ and $\langle D, f_{11}, f_{12}, \ldots, f_{kk} \rangle_{W^*}$, so $\mathcal{D}$-freeness of these algebras is implied by the $\mathcal{D}$-freeness of the sub-algebras $\mathcal{A}^{(r)}$. The elements f_{rs}, g_r, h_r have bounded norms in the original algebras $\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(k+2)}$ and hence also in the free product. $\square$

APPENDIX E: MARCENKO-PASTUR CASE

We illustrate the proof of Theorem 4.1 in the special setting of Remark 1.3: Y has n i.i.d. rows distributed as $\mathcal{N}(0,\Sigma)$ and we consider the sample covariance matrix, so that $k=1$, $B=n^{-1}\operatorname{Id}_n$,

$$W = n^{-1}Y^TY = \Sigma^{1/2}G^TG\Sigma^{1/2},$$

and G is an $n\times p$ matrix with i.i.d. $\mathcal{N}(0,1/n)$ entries.

Let O_l and O_r be Haar distributed $p\times p$ orthogonal matrices, independent of each other and of G, and let $H = O_l^T\Sigma^{1/2}O_r$ be a randomized version of $\Sigma^{1/2}$. With slight abuse of notation, we rewrite

$$W = H^TG^TGH,$$

as the spectrum of W is unchanged in law by the replacement.

Approximation. The matrices W, H, and G are embedded into larger block matrices, in the following regions corresponding to the decomposition $\mathbb{C}^N = \mathbb{C}^p \oplus \mathbb{C}^p \oplus \mathbb{C}^n$

$$\begin{bmatrix} W & H^* & \cdot \\ H & \cdot & G^* \\ \cdot & G & \operatorname{Id}_n \end{bmatrix}$$

(we use conjugate transpose notation even though – at this point – all matrix entries are real).

More formally, let $N_0 = N_1 = p$, $N_2 = n$ and $N = 2p+n$. We define diagonal projection matrices P_i having Id_{N_i} in the ith diagonal block and zeros elsewhere. Let $\tilde{H}$ and $\tilde{G}$ denote the embeddings of H and G into $N\times N$ matrices, as above. Then $(\mathbb{C}^{N\times N}, N^{-1}\operatorname{Tr}, P_0, P_1, P_2)$ is a rectangular probability space as in Example 3.1, and $\tilde{H},\tilde{G}$ are independent simple matrices in $\mathbb{C}^{N\times N}$, with each being block-orthogonally invariant. That is, for any orthogonal matrices $O_r \in \mathbb{R}^{N_r\times N_r}$, $O_1^THO_0$ has the same law as H and $O_2^TGO_1$ has the same law as G.

For the approximating free model, consider a rectangular probability space $(\mathcal{A},\tau,p_0,p_1,p_2)$ with sub-$*$-algebra $\mathcal{D} = \langle p_0,p_1,p_2\rangle$ and with deterministic elements $g,h\in\mathcal{A}$ satisfying the following conditions:

1. $\tau(p_0) = \tau(p_1) = p/N$, $\tau(p_2) = n/N$.
2. g and h are simple: $\quad p_2gp_1 = g, \quad p_1hp_0 = h.$
3. For each $l \geq 0$,

$$\tau_0((h^*h)^l) = p^{-1}\operatorname{Tr}((H^*H)^l). \tag{E.1}$$

4. g^*g has Marcenko-Pastur law: For each $l\geq 0$,

$$\tau_1((g^*g)^l) = \int x^l\nu_\lambda(x)dx \tag{E.2}$$

   where ν_λ is as in (3.6) with $\lambda = p/n$.
5. $\langle\mathcal{D},g\rangle_{W^*}$ and $\langle\mathcal{D},h\rangle^*_W$ are $\mathcal{D}$-free.

Since h is $(1,0)$-simple, (E.1) is enough to specify the full $\mathcal{D}$-law of h, and it implies that h and $\tilde{H}$ are equal in $\mathcal{D}$-law. Similarly, (E.2) is enough to specify the full $\mathcal{D}$-law of g, and g and $\tilde{G}$ are asymptotically equal in $\mathcal{D}$-law as argued in Example 3.6. Finally, by definition, $\tau(p_r) = N^{-1}\operatorname{Tr}(P_r)$. Therefore $(\mathcal{A},\tau,p_0,p_1,p_2)$ along with g,h forms a free deterministic equivalent for $(\mathbb{C}^{N\times N}, N^{-1}\operatorname{Tr}, P_0, P_1, P_2)$ along with $\tilde{G},\tilde{H}$.

For constants $C, c > 0$, suppose that $n, p \to \infty$ in such a way that $c < p/n < C$ and $\|\Sigma\| < C$. Theorem 3.9 asserts that the pairs $(\tilde{H}, \tilde{G})$ and (h, g) are *jointly* asymptotically equal in $\mathcal{D}$-law a.s. In particular

$$N_r^{-1} \operatorname{Tr}[Q(\tilde{G}, \tilde{H})] - \tau_r[Q(g,h)] \stackrel{\text{a.s.}}{\to} 0$$

for any $*$-polynomial Q. Corollary 3.10 applies this to $Q(g,h) = (h^*g^*gh)^l$ for each positive integer l, and arrives at a conclusion about approximation of Stieltjes transforms of $\tilde{W}$ and $w = h^*g^*gh$, namely that for all large $z \in \mathbb{C}^+$,

$$m_{\tilde{W}}(z) - m_w(z) \stackrel{\text{a.s.}}{\to} 0.$$

In terms of the non-embedded matrix W, denoting $m_0(z) = \tau_0((w-z)^{-1})$, we deduce

$$m_W(z) - m_0(z) \stackrel{\text{a.s.}}{\to} 0.$$

Computation. We develop equations for $m_0(z) = \tau_0((w-z)^{-1})$ in the approximating free model, proving the special case of Lemma 4.4: For $C_0 > 0$ large, there exist analytic functions $a_1 : \mathbb{D}(C_0) \to \mathbb{C}^+ \cup \{0\}$ and $b_1 : \mathbb{D}(C_0) \to \mathbb{C}$ so that equations (1.7–1.8) of Remark 1.3 hold.

A suitably specialized form of Proposition 3.12 shows the role of $\mathcal{D}$-freeness of g and h: Let $b = g^*g$. If $\mathcal{H} := \langle \mathcal{D}, h\rangle_{W^*}$ and $\langle \mathcal{D}, b\rangle_{W^*}$ are $\mathcal{D}$-free, then for all $l \geq 1$,

$$\kappa_l^{\mathcal{H}}(bh, \ldots, bh, b) = \kappa_l^{\mathcal{D}}(b\mathbf{F}^{\mathcal{D}}(h), \ldots, b\mathbf{F}^{\mathcal{D}}(h), b). \tag{E.3}$$

Remark E.1. Classical cumulants of a random variable X are derived from the log moment generating function $\log \mathbb{E}e^{bX} = \sum_{l\geq 1} \kappa_l(X)b^l/l!$. In the notation of classical multivariate conditional cumulants (e.g. [Bri69, Spe83]), the conditional distribution of X given a σ-field $\mathcal{H}$ is described by

$$\log \mathbb{E}(e^{bX}|\mathcal{H}) = \sum_{l\geq 1} \kappa_l(bX, \ldots, bX \mid \mathcal{H})/l!.$$

If classical variables X and Y are conditionally independent given a σ-field $\mathcal{D}$, and $\mathcal{H}$ is the σ-field generated by $\mathcal{D}$ and Y, then $\mathbb{E}(e^{bX}|\mathcal{H}) = \mathbb{E}(e^{bX}|\mathcal{D})$. Proposition 3.12 may be seen as a non-commutative version of this identity, written in terms of cumulants.

Using (E.3), we may express a possibly complicated transform $\mathcal{R}_w^{\mathcal{H}}$ in terms of a simpler one, namely $\mathcal{R}_{g^*g}^{\mathcal{D}}$. Indeed, the simpler version of Lemma 4.3 needed here is

Lemma E.2. *Let $\langle \mathcal{D}, g\rangle_{W^*}$ and $\mathcal{H} := \langle \mathcal{D}, h\rangle_{W^*}$ be $\mathcal{D}$-free, and let $w = h^*g^*gh$. For $c \in \mathcal{H}$ with $\|c\|$ sufficiently small,*

$$\mathcal{R}_w^{\mathcal{H}}(c) = h^*h\, \tau_1(\mathcal{R}_{g^*g}^{\mathcal{D}}(p_1\tau_1(hch^*))). \tag{E.4}$$

Proof. We use expression (3.11) for $\mathcal{R}_w^{\mathcal{H}}(c)$ in terms of cumulants. We have

$$\begin{aligned}\kappa_l^{\mathcal{H}}(wc, \ldots, wc, w) &= h^*\kappa_l^{\mathcal{H}}(g^*g\, hch^*, \ldots, g^*g\, hch^*, g^*g)h \\ &= h^*\kappa_l^{\mathcal{D}}(g^*g\, \mathbf{F}^{\mathcal{D}}(hch^*), \ldots, g^*g\, \mathbf{F}^{\mathcal{D}}(hch^*), g^*g)h.\end{aligned}$$

Here the first equality uses properties (3.9–3.10) of $\kappa^{\mathcal{H}}$, while the second equality relies on $\mathcal{D}$-freeness of g and h through (E.3). Since hch^* is $(1,1)$-simple, we have from (3.2) that $\mathbf{F}^{\mathcal{D}}(hch^*) = \sum p_r\tau_r(hch^*) = p_1\tau_1(hch^*)$. Summing over l in the previous display, we obtain

$$\mathcal{R}_w^{\mathcal{H}}(c) = h^*\mathcal{R}_{g^*g}^{\mathcal{D}}(p_1\tau_1(hch^*))h.$$

Since h is $(1,0)$-simple, for any $a \in \mathcal{D}$ we have $h^*ah = \sum_r h^*p_rh\tau_r(a) = h^*h\tau_1(a)$. Noting that $\mathcal{R}_{g^*g}^{\mathcal{D}}$ is $\mathcal{D}$-valued, we obtain (E.4). $\square$

To deduce (1.7–1.8), note first that since $\tau_0(\mathbf{F}^{\mathcal{D}}(a)) = \tau_0(a)$, we have

$$-m_0(z) = \tau_0((z-w)^{-1}) = \tau_0 \circ \mathbf{F}^{\mathcal{D}}((z-w)^{-1}) = \tau_0(G_w^{\mathcal{D}}(z)) = \tau_0 \circ \mathbf{F}^{\mathcal{D}}(G_w^{\mathcal{H}}(z)), \tag{E.5}$$

the last step applying (3.17). For fixed $z \in \mathbb{D}(C_0)$, we can rewrite the inversion formula (3.14) using (E.4), with $c = G_w^{\mathcal{H}}(z)$, as

$$G_w^{\mathcal{H}}(z) = (z - \mathcal{R}_w^{\mathcal{H}}(G_w^{\mathcal{H}}(z)))^{-1} = (z - h^* h \beta)^{-1}, \tag{E.6}$$

where we define

$$\alpha = \tau_1(h G_w^{\mathcal{H}}(z) h^*), \qquad \beta = \tau_1(\mathcal{R}_{g^*g}^{\mathcal{D}}(p_1 \alpha)). \tag{E.7}$$

Lemma E.3 below computes $\mathcal{R}_{g^*g}^{\mathcal{D}}(p_1\alpha)$ using the $\mathcal{R}$-transform of the standard Marcenko-Pastur law, yielding

$$\mathcal{R}_{g^*g}^{\mathcal{D}}(p_1\alpha) = p_1(1-\lambda\alpha)^{-1}$$

for $\lambda = p/n$. Applying this and (E.6) to (E.5) and (E.7), we obtain the equations

$$\alpha = \tau_1(h(z - h^*h\beta)^{-1}h^*), \quad \beta = (1-\lambda\alpha)^{-1}, \quad -m_0(z) = \tau_0((z - h^*h\beta)^{-1}).$$

Now passing to a power series, then applying (E.1) and the spectral calculus, we obtain

$$-m_0(z) = \sum_{l \geq 0} z^{-(l+1)} \tau_0((h^*h)^l)\beta^l = \sum_{l\geq 0} z^{-(l+1)} \frac{1}{p} \operatorname{Tr}((H^*H)^l)\beta^l = \frac{1}{p}\operatorname{Tr}(z \operatorname{Id}_p - \beta H^*H)^{-1}. \tag{E.8}$$

Using the definition of τ_r and cyclic property of τ, a similar calculation shows that

$$\alpha = \frac{\tau(p_0)}{\tau(p_1)} \tau_0((z - h^*h\beta)^{-1} h^*h) = \frac{1}{p}\operatorname{Tr}[(z\operatorname{Id}_p - \beta H^*H)^{-1} H^*H]. \tag{E.9}$$

Setting $a_1 = -(p/n)\alpha = -\lambda\alpha$ and $b_1 = -\beta$ and recalling that $\Sigma = O_r H^* H O_r^*$, we recover (1.7–1.8).

We check a few analytic details of the above argument: For $z \in \mathbb{D}(C_0)$ and C_0 sufficiently large, α is defined by the series expansion (3.12) and we have

$$\alpha = \tau_1\left(h \sum_{l=0}^{\infty} \mathbf{F}^{\mathcal{H}}(z^{-1}(wz^{-1})^l)h^*\right) = \sum_{l=0}^{\infty} z^{-(l+1)}\tau_1(hw^l h^*).$$

For C_0 sufficiently large, boundedness of τ implies that α is analytic in z, with $\alpha \sim z^{-1}\tau_1(hh^*)$ as $z \to \infty$. Then either $h = 0$ in which case $\alpha = 0$ for all z, or positivity and faithfulness of τ yields $\Im\alpha < 0$ and $\Im a_1 > 0$ for all $z \in \mathbb{D}(C_0)$. Furthermore, this implies $\beta = (1-\lambda\alpha)^{-1}$ is also analytic in z and bounded over $\mathbb{D}(C_0)$, which justifies the use of formal series and spectral calculus in (E.8) and (E.9) for large C_0. This establishes Lemma 4.4 in this special case.

Lemma E.3. *For any $z \in \mathbb{C}$ with $|z|$ sufficiently small, $\mathcal{R}_{g^*g}^{\mathcal{D}}(p_1 z) = p_1(1-\lambda z)^{-1}$.*

Proof. We first verify that the Marcenko-Pastur law ν_λ given in (3.6) has $\mathcal{R}$-transform $\mathcal{R}(z) = (1-\lambda z)^{-1}$: Indeed, its Stieltjes transform $m(z)$ satisfies the functional equation

$$m(z) = (1 - \lambda - \lambda z m(z) - z)^{-1} \tag{E.10}$$

for each $z \in \mathbb{C}^+$ [Sil95, eq.(1.4)]—this is the limiting version of (1.3) for $\Sigma = \mathrm{Id}$ as $p/n \to \lambda$. The Cauchy transform $w = -m(z)$ has a functional inverse which we write as $z = K(w)$. Rewriting (E.10) in terms of w and $K(w)$ yields

$$1 - \lambda + (\lambda w - 1)K(w) = -w^{-1}.$$

The $\mathcal{R}$-transform is then $\mathcal{R}(w) = K(w) - 1/w$, for example from (3.13). Inserting $K(w) = R(w) + 1/w$ into the previous display and rearranging yields $\mathcal{R}(w) = (1 - \lambda w)^{-1}$.

Now denote by $\mathcal{R}^{\mathbb{C},1}_{g^*g}$ the scalar $\mathcal{R}$-transform of g^*g with respect to trace τ_1. By (E.2), the above implies $\mathcal{R}^{\mathbb{C},1}_{g^*g}(\alpha) = (1 - \lambda\alpha)^{-1}$. The lemma follows from relating $\mathcal{R}^{\mathcal{D}}_{g^*g}(p_1\alpha)$ to $\mathcal{R}^{\mathbb{C},1}_{g^*g}(\alpha)$. In the full proof of Lemma 4.4, we do the analogous step by relating $\mathcal{R}^{\mathcal{D}}_{g^*g}$ to $G^{\mathcal{D}}_{g^*g}$ using (3.13), projecting down to $G^{\mathbb{C}}_{g^*g}$ using (3.17), and relating this back to $\mathcal{R}^{\mathbb{C}}_{g^*g}$. Here, we use a simpler direct argument to show that if a is $(1,1)$-simple and $b = \sum_r p_r b_r \in \mathcal{D}$, then

$$\mathcal{R}^{\mathcal{D}}_a(b) = p_1 \mathcal{R}^{\mathbb{C},1}_a(b_1).$$

Noting that $b_r \in \mathbb{C} \subset \mathcal{D}$ and $ab = \sum_r a p_r b_r = b_1 a$ since a is $(1,1)$-simple, we have by (3.11) and (3.9–3.10) that

$$\mathcal{R}^{\mathcal{D}}_a(b) = \sum_{l \geq 1} \kappa^{\mathcal{D}}_l(ab, \ldots, ab, a) = \sum_{l \geq 1} b_1^{l-1} \kappa^{\mathcal{D}}_l(a, \ldots, a). \tag{E.11}$$

Since a is $(1,1)$-simple, the $\mathcal{D}$-valued moments of a are given by $\mathbf{F}^{\mathcal{D}}(a^l) = p_1\tau_1(a^l)$. The cumulants are defined by the moment-cumulant relations

$$\kappa^{\mathcal{D}}_l(a, \ldots, a) = \sum_{\pi \in \mathrm{NC}(l)} \mu(\pi, \{1, \ldots, l\}) \prod_{S \in \pi} \mathbf{F}^{\mathcal{D}}(a^{|S|}),$$

where $\mathrm{NC}(l)$ is the lattice of non-crossing partitions on $\{1, \ldots, l\}$ and μ is the Mobius inversion function on this lattice, see e.g. [NS06, Eq. (11.5)]. Then

$$\kappa^{\mathcal{D}}_l(a, \ldots, a) = p_1 \sum_{\pi \in \mathrm{NC}(l)} \mu(\pi, \{1, \ldots, l\}) \prod_{S \in \pi} \tau_1(a^{|S|}) = p_1 \kappa^{\mathbb{C},1}_l(a, \ldots, a),$$

where $\kappa^{\mathbb{C},1}_l$ are the scalar-valued free cumulants for trace τ_1. Recalling (E.11), we obtain $\mathcal{R}^{\mathcal{D}}_a(b) = p_1\mathcal{R}^{\mathbb{C},1}_a(b_1)$, which concludes the proof. $\square$

REFERENCES

[BG09] Florent Benaych-Georges. Rectangular random matrices, related convolution. *Probability Theory and Related Fields*, 144(3-4):471–515, 2009.

[Bla06] Bruce Blackadar. *Operator algebras: Theory of C*-algebras and von Neumann algebras*, volume 122. Springer Science & Business Media, 2006.

[Bri69] David R. Brillinger. The calculation of cumulants via conditioning. *Annals of the Institute of Statistical Mathematics*, 21(1):215–218, Dec 1969.

[BS10] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.

[CDS11] Romain Couillet, Mérouane Debbah, and Jack W Silverstein. A deterministic equivalent for the analysis of correlated MIMO multiple access channels. *IEEE Transactions on Information Theory*, 57(6):3493–3514, 2011.

[CL11] C Caratheéodory and E Landau. Beiträge zur konvergenz von funktionenfolgen. *Sitz. Ber. Königl. Preuss. Akad. Wiss., Phys.-math.*, 26:587–613, 1911.

[CŚ06] Benoit Collins and Piotr Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006.

[DL11] Florian Dupuy and Philippe Loubaton. On the capacity achieving covariance matrix for frequency selective MIMO channels using the asymptotic approach. *IEEE Transactions on Information Theory*, 57(9):5737–5753, 2011.

[Dyk95] Kenneth J Dykema. Amalgamated free products of multi-matrix algebras and a construction of subfactors of a free group factor. *American Journal of Mathematics*, 117(6):1555–1602, 1995.

[GH03] Jeffrey S Geronimo and Theodore P Hill. Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform. *Journal of Approximation Theory*, 121(1):54–60, 2003.

[HP00] Fumio Hiai and Denes Petz. Asymptotic freeness almost everywhere for random matrices. *Acta Sci. Math. (Szeged)*, 66(3–4):809–834, 2000.

[Kad04] Richard V Kadison. Non-commutative conditional expectations and their applications. In *Operator Algebras, Quantization, and Noncommutative Geometry*, pages 143–179. American Mathematical Society, 2004.

[NS06] A. Nica and R. Speicher. *Lectures on the Combinatorics of Free Probability*. Number v. 13 in Lectures on the combinatorics of free probability. Cambridge University Press, 2006.

[Pop93] Sorin Popa. Markov traces on universal Jones algebras and subfactors of finite index. *Inventiones mathematicae*, 111(1):375–405, 1993.

[Sch13] Joel L Schiff. *Normal families*. Springer Science & Business Media, 2013.

[Sil95] Jack W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, 55(2):331–339, 1995.

[Spe83] TP Speed. Cumulants and partition lattices. *Australian & New Zealand Journal of Statistics*, 25(2):378–388, 1983.

[VDN92] D V Voiculescu, K J Dykema, and A Nica. *Free Random Variables*. CRM Monograph Series. American Mathematical Soc., 1992.

[Voi85] Dan Voiculescu. Symmetries of some reduced free product C^*-algebras. In *Operator algebras and their connections with topology and ergodic theory*, pages 556–588. Springer, 1985.

[YBK88] Yong-Quan Yin, Zhi-Dong Bai, and Pathak R Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability theory and related fields*, 78(4):509–521, 1988.