Bernoulli **24**(4B), 2018, 3447–3468
https://doi.org/10.3150/17-BEJ965

A unified matrix model including both CCA and F matrices in multivariate analysis: The largest eigenvalue and its applications

XIAO HAN*, GUANGMING PAN** and QING YANG†

*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore. E-mail: *xhan011@e.ntu.edu.sg; **gmpan@ntu.edu.sg; †qyang1@e.ntu.edu.sg*

Let $\mathbf{Z}_{M_1\times N}=\mathbf{T}^{\frac{1}{2}}\mathbf{X}$ where $(\mathbf{T}^{\frac{1}{2}})^2=\mathbf{T}$ is a positive definite matrix and $\mathbf{X}$ consists of independent random variables with mean zero and variance one. This paper proposes a unified matrix model

$$\Omega=\left(\mathbf{Z}\mathbf{U}_2\mathbf{U}_2^T\mathbf{Z}^T\right)^{-1}\mathbf{Z}\mathbf{U}_1\mathbf{U}_1^T\mathbf{Z}^T,$$

where $\mathbf{U}_1$ and $\mathbf{U}_2$ are isometric with dimensions $N\times N_1$ and $N\times(N-N_2)$ respectively such that $\mathbf{U}_1^T\mathbf{U}_1=\mathbf{I}_{N_1}$, $\mathbf{U}_2^T\mathbf{U}_2=\mathbf{I}_{N-N_2}$ and $\mathbf{U}_1^T\mathbf{U}_2=0$. Moreover, $\mathbf{U}_1$ and $\mathbf{U}_2$ (random or non-random) are independent of $\mathbf{Z}_{M_1\times N}$ and with probability tending to one, $\operatorname{rank}(\mathbf{U}_1)=N_1$ and $\operatorname{rank}(\mathbf{U}_2)=N-N_2$. We establish the asymptotic Tracy–Widom distribution for its largest eigenvalue under moment assumptions on $\mathbf{X}$ when N_1, N_2 and M_1 are comparable.

The asymptotic distributions of the maximum eigenvalues of the matrices used in Canonical Correlation Analysis (CCA) and of F matrices (including centered and non-centered versions) can be both obtained from that of Ω by selecting appropriate matrices $\mathbf{U}_1$ and $\mathbf{U}_2$. Moreover, via appropriate matrices $\mathbf{U}_1$ and $\mathbf{U}_2$, this matrix Ω can be applied to some multivariate testing problems that cannot be done by both types of matrices. To see this, we explore two more applications. One is in the MANOVA approach for testing the equivalence of several high-dimensional mean vectors, where $\mathbf{U}_1$ and $\mathbf{U}_2$ are chosen to be two nonrandom matrices. The other one is in the multivariate linear model for testing the unknown parameter matrix, where $\mathbf{U}_1$ and $\mathbf{U}_2$ are random. For each application, theoretical results are developed and various numerical studies are conducted to investigate the empirical performance.

Keywords: canonical correlation analysis; F matrix; largest eigenvalue; MANOVA; multivariate linear model; random matrix theory; Tracy–Widom distribution

1. Introduction

Rapid development of modern technology nowadays necessitates statistical inference on high-dimensional data in many scientific fields such as image processing, genetic engineering, machine learning and so on. This raises a boom in pursuing methodologies to remedy classical theories which are designed for the fixed dimensions. For such a purpose one popular tool is the spectral analysis of high-dimensional matrices in random matrix theory. The readers may refer to the monograph [2] and the references therein for a comprehensive reading.

This paper focuses on the largest eigenvalues. Ever since the pioneer work discovering the limiting distribution of the largest eigenvalue for the large Gaussian Wigner ensemble by Tracy

and Widom in [15,16], the largest eigenvalues of large random matrices have been widely studied. To name a few we mention [3,5,6] and [14]. The largest eigenvalues prove to be fruitful objects of study, playing an important role in multivariate statistical analysis such as principle component analysis (PCA), multivariate analysis of variance (MANOVA), canonical correlation analysis (CCA) and discriminant analysis. Among the vast literature, we refer the readers to a seminal work [12], as well as a recent work [10]. Johnstone in [12] considered a double Wishart setting and developed the Tracy–Widom law of its largest root when the dimension of the data matrix $\mathbf{X}$ and the sample size are comparable with the dimension being even. This limiting distribution can be applied to conduct various statistical inferences in his companion paper [13]. Considering that the results in [12] work for the Gaussian distribution only, the authors in [10] investigated an F type matrix for the general distributions without even dimension restriction. However, one may notice that the Tracy–Widom result in [10] is only verified for zero mean data.

We now set a stage to present our matrix model. The most initial motivation is the matrix frequently used in CCA. Suppose that we are given two sets of random variables, organized into two random vectors $\mathbf{x}$ and $\mathbf{y}$ with dimensions M_1 and M_2, respectively. Without loss of generality, we may assume that $M_1 \leq M_2$. In multivariate analysis, CCA is the favorite method to investigate the correlation structure between two random vectors, which was introduced by Hotelling [11] first. The aim of CCA is to seek two vectors $\mathbf{a}$ and $\mathbf{b}$ such that the linear combination of $\mathbf{a}^T\mathbf{x}$ and $\mathbf{b}^T\mathbf{y}$ can get the highest correlation coefficient that is,

$$\rho(\mathbf{a},\mathbf{b}) := \frac{\operatorname{Cov}(\mathbf{a}^T\mathbf{x},\mathbf{b}^T\mathbf{y})}{\sqrt{\operatorname{Var}(\mathbf{a}^T\mathbf{x})}\sqrt{\operatorname{Var}(\mathbf{b}^T\mathbf{y})}}. \tag{1.1}$$

If $\rho_1 = \rho_1(\mathbf{a}_1,\mathbf{b}_1) := \max_{\mathbf{a},\mathbf{b}} \rho(\mathbf{a},\mathbf{b})$, then ρ_1 is called the first canonical correlation coefficient. Given the first canonical correlation coefficient, one can continue to seek the second canonical correlation coefficient which is the maximum correlation coefficient of $\mathbf{a}_2^T\mathbf{x}$ and $\mathbf{b}_2^T\mathbf{y}$, uncorrelated to $\mathbf{a}_1^T\mathbf{x}$ and $\mathbf{b}_1^T\mathbf{y}$. Iterating this procedure to the end, we can get the canonical correlation coefficients $\rho_1,\rho_2,\ldots,\rho_{M_1}$. Denote the population covariance matrix of any two random vectors $\mathbf{u}$ and $\mathbf{v}$ by $\Sigma_{\mathbf{uv}}$. By (1.1), it is not hard to conclude that in order to find the population canonical correlation coefficients $\rho_1,\rho_2,\ldots,\rho_{M_1}$, one only need to solve the determinant equation

$$\det\bigl(\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{xy}}^T - \rho^2\Sigma_{\mathbf{xx}}\bigr) = 0. \tag{1.2}$$

If $\mathbf{x}$ and $\mathbf{y}$ are independent, then $\rho_1^2 = \cdots = \rho_{M_1}^2 = 0$ or equivalently the largest eigenvalue of $\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{xy}}^T$, $\rho_1^2 = 0$. For the moment, we assume that $\mathbb{E}\mathbf{x} = \mathbb{E}\mathbf{y} = 0$ for ease of illustration, but bearing in mind that such conditions are not needed in this work. Then under the classical low-dimensional setting, that is, both M_1 and M_2 are fixed but N is large, one can safely use γ_1, the largest eigenvalue of $\mathbf{A}_{\mathbf{xx}}^{-1}\mathbf{A}_{\mathbf{xy}}\mathbf{A}_{\mathbf{yy}}^{-1}\mathbf{A}_{\mathbf{xy}}^T$, to estimate ρ_1^2 since the sample covariance matrices converge to their population counterparts as N tends to infinity, where

$$\mathbf{A}_{\mathbf{xx}} = \mathbf{X}\mathbf{X}^T, \qquad \mathbf{A}_{\mathbf{yy}} = \mathbf{Y}\mathbf{Y}^T, \qquad \mathbf{A}_{\mathbf{xy}} = \mathbf{X}\mathbf{Y}^T.$$

However, when M_1 and M_2 are comparable with the sample size N, the consistency will no longer hold for the sample covariance matrices and accordingly the largest sample canonical correlation coefficient γ_1. Putting forward a theory on high-dimensional CCA is then much needed.

If $\mathbf{x}$ or $\mathbf{y}$ is Gaussian distributed, it is not difficult to derive that the largest eigenvalue of

$$\mathbf{S}_{xy} = \mathbf{A}_{\mathbf{xx}}^{-1}\mathbf{A}_{\mathbf{xy}}\mathbf{A}_{\mathbf{yy}}^{-1}\mathbf{A}_{\mathbf{xy}}^{T}$$

reduces to that of the double Wishart matrices in [12], see the matrix (1.3) below. Thus after centralizing and re-scaling, it converges to the Type-1 Tracy–Widom distribution as proved in [12] and [10]. However, to our best knowledge, corresponding results are not yet available for non-gaussian distributions, which is the starting point of this paper. Here we would also remark some other existing works about CCA in the high-dimensional case. Central limit theorems of linear spectral statistics of CCA have been established in [17], which are for zero mean data, while spiked eigenvalues are investigated for CCA in [4]. There are also a lot of existing works about sparse CCA and we mention [8] among others.

Denote the largest eigenvalue of $\mathbf{S}_{xy}$ by γ_1. Then γ_1 is also the largest eigenvalue of $\mathbf{T}_{xy} := \mathbf{P}_y\mathbf{P}_x\mathbf{P}_y$, where

$$\mathbf{P}_x = \mathbf{X}^T\left(\mathbf{X}\mathbf{X}^T\right)^{-1}\mathbf{X}, \qquad \mathbf{P}_y = \mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}\mathbf{Y}.$$

Equivalently, it is the largest solution to $\det(\mathbf{X}\mathbf{P}_y\mathbf{X}^T - \gamma_1\mathbf{X}\mathbf{X}^T) = 0$. Define $\lambda_1 = \frac{\gamma_1}{1-\gamma_1}$. Then under the condition that $\liminf_{N\to\infty}\frac{N}{M_1+M_2} > 1$, λ_1 is also the largest solution of

$$\det\left(\mathbf{X}\mathbf{P}_y\mathbf{X}^T - \lambda_1\mathbf{X}(\mathbf{I}-\mathbf{P}_y)\mathbf{X}^T\right) = 0.$$

The matrix of interest now becomes

$$\left(\mathbf{X}(I-\mathbf{P}_y)\mathbf{X}^T\right)^{-1}\mathbf{X}\mathbf{P}_y\mathbf{X}^T. \tag{1.3}$$

Inspired by (1.3), we propose a unified matrix model

$$\Omega = \left(\mathbf{Z}\mathbf{U}_2\mathbf{U}_2^T\mathbf{Z}^T\right)^{-1}\mathbf{Z}\mathbf{U}_1\mathbf{U}_1^T\mathbf{Z}^T, \tag{1.4}$$

where $\mathbf{U}_1^T\mathbf{U}_1 = \mathbf{I}_{N_1}$, $\mathbf{U}_2^T\mathbf{U}_2 = \mathbf{I}_{N-N_2}$ and $\mathbf{U}_1^T\mathbf{U}_2 = 0$ (see (2.1) below for more details). We establish the asymptotic Tracy–Widom law for its largest eigenvalue in this work. An intriguing observation is that although our Tracy–Widom approximation is theoretically established for diverging dimensions, it keeps accurate for small ones (the dimension M_1 can be as small as 5 in Table 1).

The motivations behind the construction of such a matrix model Ω are illustrated as follows. First, the matrix (1.3) used in CCA is a special case of Ω by noticing that $\mathbf{P}_y$ and $\mathbf{I}-\mathbf{P}_y$ are orthogonal projection matrices. In addition, the non-zero mean data can be accommodated by writing $\mathbf{U}_2\mathbf{U}_2^T = \mathbf{P}_N(I-\mathbf{P}_{Ny})\mathbf{P}_N$, $\mathbf{U}_1\mathbf{U}_1^T = \mathbf{P}_N\mathbf{P}_{Ny}\mathbf{P}_N$ and observing that the mean vectors can be absorbed into the matrix $\mathbf{P}_N = \mathbf{I}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$, see Remark 6 below (the definition of $\mathbf{P}_{Ny}$ is given there). Further illustrations are given in Section 3, where we deal with the independence testing via CCA in detail.

Second, the matrix Ω generalizes the models in [12] and [10]. We would like to point out that if the matrix $\mathbf{Z}$ is generated from Gaussian distribution, then the two terms $(\mathbf{Z}\mathbf{U}_2\mathbf{U}_2^T\mathbf{Z}^T)$ and $(\mathbf{Z}\mathbf{U}_1\mathbf{U}_1^T\mathbf{Z}^T)$ in Ω are independent with normal entries, which simply reduces to the one studied

Table 1. Simulated quantiles for rescaled λ_1, i.e. the values $\frac{\#\{\text{rescaled } \lambda_1 \le \text{"Percentile"}\}}{10\,000}$ based on 10 000 replications under different data distributions and different dimensions

		$M^{(1)} = (M_1, N_1, N_2, N) = (5, 8, 10, 30)$					$M^{(2)} = (M_1, N_1, N_2, N) = (15, 8, 10, 50)$				
Percentile	TW	$M^{(1)}$	$2M^{(1)}$	$8M^{(1)}$	$16M^{(1)}$	$20M^{(1)}$	$M^{(2)}$	$2M^{(2)}$	$8M^{(2)}$	$16M^{(2)}$	$20M^{(2)}$
					Standard Normal						
−3.90	0.01	0.0132	0.0083	0.0099	0.0080	0.0108	0.0109	0.0102	0.0085	0.0091	0.0090
−3.18	0.05	0.0546	0.0501	0.0497	0.0502	0.0491	0.0514	0.0495	0.0467	0.0450	0.0476
−2.78	0.10	0.1041	0.1011	0.0995	0.1030	0.0992	0.1028	0.0974	0.0981	0.0956	0.0975
−1.91	0.30	0.2941	0.2948	0.3024	0.3026	0.3028	0.3047	0.3049	0.2944	0.2908	0.3004
−1.27	0.50	0.5031	0.5007	0.5026	0.5114	0.5048	0.5072	0.5077	0.4987	0.4971	0.5009
−0.59	0.70	0.7101	0.7057	0.7116	0.7116	0.7081	0.7074	0.7040	0.7075	0.7037	0.7051
0.45	0.90	0.9138	0.9027	0.9050	0.9062	0.9014	0.9055	0.9019	0.9019	0.9048	0.9038
0.98	0.95	0.9610	0.9507	0.9552	0.9538	0.9519	0.9569	0.9525	0.9502	0.9560	0.9560
2.02	0.99	0.9933	0.9896	0.9898	0.9909	0.9912	0.9916	0.9906	0.9900	0.9910	0.9912
					Discrete						
−3.90	0.01	0.0116	0.0093	0.0094	0.0099	0.0082	0.0099	0.0091	0.0098	0.0104	0.0080
−3.18	0.05	0.0523	0.0464	0.0503	0.0514	0.0477	0.0496	0.0480	0.0529	0.0495	0.0460
−2.78	0.10	0.0996	0.0943	0.1034	0.0983	0.0998	0.0951	0.0986	0.1037	0.0978	0.0974
−1.91	0.30	0.3049	0.2954	0.3054	0.2974	0.3024	0.2933	0.2915	0.3069	0.2968	0.3050
−1.27	0.50	0.5068	0.5002	0.5114	0.4961	0.4984	0.4989	0.4965	0.5069	0.5015	0.4964
−0.59	0.70	0.7124	0.7080	0.7065	0.6986	0.7045	0.7065	0.6976	0.7062	0.7042	0.6946
0.45	0.90	0.9102	0.9098	0.9035	0.9014	0.9021	0.9065	0.9031	0.9058	0.9067	0.8966
0.98	0.95	0.9583	0.9565	0.9537	0.9508	0.9512	0.9559	0.9540	0.9515	0.9546	0.9494
2.02	0.99	0.9931	0.9917	0.9905	0.9911	0.9894	0.9921	0.9903	0.9915	0.9912	0.9894
					Gamma(4, 0.5)						
−3.90	0.01	0.0109	0.0091	0.0099	0.0093	0.0104	0.0098	0.0069	0.0107	0.0096	0.0104
−3.18	0.05	0.0507	0.0502	0.0500	0.0501	0.0507	0.0494	0.0452	0.0503	0.0486	0.0495
−2.78	0.10	0.1025	0.1011	0.0996	0.1013	0.0991	0.1006	0.0957	0.1021	0.0965	0.1002
−1.91	0.30	0.3024	0.2953	0.3008	0.2993	0.2934	0.2985	0.2965	0.2970	0.2972	0.2983
−1.27	0.50	0.4992	0.4994	0.5013	0.4967	0.4865	0.5009	0.4890	0.5028	0.4995	0.5010
−0.59	0.70	0.7033	0.7097	0.6994	0.6935	0.6923	0.7100	0.7006	0.7015	0.7040	0.7080
0.45	0.90	0.9065	0.9062	0.9018	0.9005	0.9023	0.9052	0.9045	0.8970	0.9027	0.9037
0.98	0.95	0.9546	0.9531	0.9503	0.9509	0.9503	0.9515	0.9531	0.9499	0.9503	0.9523
2.02	0.99	0.9908	0.9912	0.9906	0.9895	0.9888	0.9910	0.9901	0.9904	0.9904	0.9900

in [12] without even dimension restriction. Without this Gaussian assumption, we indeed investigate a more general case–the two terms can only be considered as uncorrelated with each other. We would also like to highlight that Ω not only covers the F-matrix in [10], but also generalizes it to any non-zero mean vectors by choosing some special $\mathbf{U}_2$ and $\mathbf{U}_1$. Detailed explanations will be given in Section 4.

Third, by selecting appropriate matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ (*random* or *nonrandom*), the Tracy–Widom distribution for the largest eigenvalue of this unified matrix Ω can be applied to the other multivariate testing problems, which cannot be done by the traditional CCA matrix (1.3).

To see this, we explore two more applications. One is the MANOVA approach in testing the equivalence of g groups' mean vectors. It is well known that classical MANOVA relies on the eigenvalues of the matrix $\mathbf{V} = \mathbf{W}^{-1}\mathbf{B}$, where $\mathbf{W}$ is the within sum of squares and cross-product matrix (SSCP) and $\mathbf{B}$ is the between SSCP, see [1]. The matrix $\mathbf{V}$ can be written in terms of Ω by choosing *nonrandom* matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ as in equations (5.2)–(5.3) below, with the derivation details postponed to Section 5. The other one is in the multivariate linear regression model $\mathbf{Y} = \mathbf{XB} + \mathbf{Z}$ for testing the unknown parameter matrix $\mathbf{B}$. We consider both the linear hypothesis testing $H_0 : \mathbf{C}_1\mathbf{B} = \mathbf{\Gamma}_1$ and the general intra-subject hypothesis testing $H_0 : \mathbf{CBD} = \mathbf{\Gamma}$. Taking the linear one as an example, we can rewrite its testing matrix $\mathbf{M}_1 = \mathbf{E}_1^{-1}\mathbf{H}_1$ in the form of Ω by selecting *random* matrices $\mathbf{U}_1\mathbf{U}_1^T = \mathbf{P}_{\widetilde{\mathbf{X}}}$ and $\mathbf{U}_2\mathbf{U}_2^T = \mathbf{I} - \mathbf{P}_{\mathbf{X}}$ in (6.3), where $\mathbf{E}_1$ is the error SSCP and $\mathbf{H}_1$ the hypothesis SSCP described in Section 6. Simulation results in Sections 7.3–7.4 show that the largest eigenvalue performs well in these two applications for both dense but weak alternative (DWA) and sparse but strong alternative (SSA).

This paper is organized as follows. In Section 2, the main theorem about the Tracy–Widom distribution for the largest eigenvalue λ_1 of the unified matrix Ω is presented. Three applications are introduced in Sections 3, 5 and 6, regarding the high-dimensional independence testing via CCA, MANOVA and multivariate linear regression, respectively. We remark that all these applications here cannot be done by either [12] or [10] because we neither assume Gaussian distribution for $\mathbf{Z}$ nor impose an independent structure on $(\mathbf{ZU}_2\mathbf{U}_2^T\mathbf{Z}^T)$ and $(\mathbf{ZU}_1\mathbf{U}_1^T\mathbf{Z}^T)$. The relation between Ω and F-matrix is illustrated in Section 4. Except for theoretical results developed in previous sections, we also conduct a series of simulations in Section 7 to investigate the accuracy of the proposed asymptotic Tracy–Widom distribution (Section 7.1) as well as its numerical performance in our three applications (Sections 7.2–7.4). All detailed proofs are relegated to the supplementary material [9].

2. Main result on Ω

We investigate the largest eigenvalue of the unified matrix

$$\Omega = \left(\mathbf{ZU}_2\mathbf{U}_2^T\mathbf{Z}^T\right)^{-1}\mathbf{ZU}_1\mathbf{U}_1^T\mathbf{Z}^T \tag{2.1}$$

in this section and develop its Tracy–Widom distribution. Here $\mathbf{Z}_{M_1\times N} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, $\mathbf{T}_{M_1\times M_1}$ can be any positive definite matrix and $\mathbf{X} = (X_{ij})_{M_1\times N}$ satisfies the following Condition 1. Assume that $\mathbf{U}_1$ and $\mathbf{U}_2$ are two isometries with dimensions $N \times N_1$ and $N \times (N - N_2)$, respectively such that $N_1 \le N_2$, $\mathbf{U}_1^T\mathbf{U}_1 = \mathbf{I}_{N_1}$, $\mathbf{U}_2^T\mathbf{U}_2 = \mathbf{I}_{N-N_2}$ and $\mathbf{U}_1^T\mathbf{U}_2 = 0$. Moreover, $\mathbf{U}_1$ and $\mathbf{U}_2$ (random or non-random) are independent of $\mathbf{X}$ and with probability tending to one, $\operatorname{rank}(\mathbf{U}_1) = N_1$ and $\operatorname{rank}(\mathbf{U}_2) = N - N_2$. The notation "0" may indicate a zero value, a zero vector or a zero matrix in this paper, changing from line to line.

Condition 1. A matrix $\mathbf{X} = (X_{ij})_{M_1\times N}$ satisfies Condition 1 if its entries X_{ij} are independent (but not necessarily identically distributed) with all moments being finite and

$$\mathbb{E}X_{ij} = 0, \qquad \mathbb{E}X_{ij}^2 = \mathbb{E}X_{it}^2, \qquad 1 \le i \le M_1, 1 \le j, t \le N. \tag{2.2}$$

Remark 1. Note that the matrix $\mathbf{T}$ does not influence the largest eigenvalue of Ω and it can be any positive definite matrix. Indeed, let $\Omega_x = (\mathbf{X}\mathbf{U}_2\mathbf{U}_2^T\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{U}_1\mathbf{U}_1^T\mathbf{X}^T$. One can easily observe that Ω and Ω_x share the same largest eigenvalue by the fact that AB and BA share the same nonzero eigenvalues.

We now state the limiting distribution for the largest eigenvalue of the unified matrix Ω.

Theorem 2.1. *Consider the matrix Ω defined in (2.1). Suppose that $\mathbf{T}$ is any positive definite matrix and $\mathbf{X}$ satisfies Condition 1. Suppose that $\liminf_{N\to\infty}\frac{N}{M_1+N_2} > 1$, $N_1 \le N_2$, $\frac{N_1}{N_2}$ and $\frac{M_1}{N-N_2}$ are both bounded away from 0, and $\frac{N_1}{M_1}$ is bounded away from 0 and ∞. Denote the largest eigenvalue of Ω by λ_1. Then there exist μ_N and σ_N such that*

$$\lim_{N\to\infty} P\big(\sigma_N N_1^{2/3}(\lambda_1 - \mu_N) \le s\big) = F_1(s), \tag{2.3}$$

where $F_1(s)$ is the Type-1 Tracy–Widom distribution. Moreover, the centering and scaling parameters μ_N and σ_N can be determined as follows. Suppose that $c_N \in [0, (1 - \sqrt{\frac{M_1}{N-N_2}})^2]$ satisfies the equation

$$\int_{-\infty}^{+\infty}\left(\frac{c_N}{\lambda - c_N}\right)^2 dF(\lambda) = \frac{N_1}{M_1}, \tag{2.4}$$

where $F(\lambda)$ is the limit spectral density (LSD) of $(\mathbf{X}\mathbf{U}_2\mathbf{U}_2^T\mathbf{X}^T)^{-1}$. Then

$$\mu_N = \frac{1}{c_N}\left(1 + \frac{M_1}{N_1}\int_{-\infty}^{+\infty}\left(\frac{c_N}{\lambda - c_N}\right) dF(\lambda)\right) \tag{2.5}$$

and

$$\frac{1}{\sigma_N^3} = \frac{1}{c_N^3}\left(1 + \frac{M_1}{N_1}\int_{-\infty}^{+\infty}\left(\frac{c_N}{\lambda - c_N}\right)^3 dF(\lambda)\right). \tag{2.6}$$

Remark 2. The LSD of the empirical spectral distribution of $(\mathbf{X}\mathbf{U}_2\mathbf{U}_2^T\mathbf{X}^T)$ (equivalent to the sample covariance matrix in the Gaussian case) is the famous Marcenko Pastur distribution. From there one can easily find $F(\lambda)$.

Remark 3. When $\mathbf{X}$ is a complex random matrix, Theorem 2.1 still holds but the Tracy–Widom distribution $F_1(s)$ should be replaced by $F_2(s)$. One may refer to [16] for the definitions of $F_i(s), i = 1, 2$.

Remark 4. The condition $\mathbf{U}_1^T\mathbf{U}_2 = 0$ imposed on the matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ can be relaxed to $\mathbf{U}_1^T\mathbf{U}_2 = (\mathbf{I}_{N_1}, 0)$. In fact, if $\mathbf{U}_1^T\mathbf{U}_2 = (\mathbf{I}_{N_1}, 0)$, then we can write $\mathbf{U}_2$ as $\mathbf{U}_2 = (\mathbf{U}_1, \mathbf{U}_4)$ such that $\mathbf{U}_1^T\mathbf{U}_4 = 0$. This is because if we denote $\mathbf{U}_2 = (\mathbf{U}_3, \mathbf{U}_4)$, then the relation $\mathbf{U}_1^T\mathbf{U}_2 = (\mathbf{I}_{N_1}, 0)$ suggests that $\mathbf{U}_1^T\mathbf{U}_3 = \mathbf{I}_{N_1}$, $\mathbf{U}_1^T\mathbf{U}_4 = 0$. Denoting the ith columns of $\mathbf{U}_1$ and $\mathbf{U}_3$ by $\mathbf{u}_{1i}$ and $\mathbf{u}_{3i}$

respectively, we have $\mathbf{u}_{1i}^T\mathbf{u}_{3i} = 1$. By the Cauchy–Schwarz inequality, we see that

$$1 = \mathbf{u}_{1i}^T\mathbf{u}_{3i} \le \|\mathbf{u}_{1i}\|\|\mathbf{u}_{3i}\| = 1,$$

which forces $\mathbf{u}_{1i} = \mathbf{u}_{3i}$ and consequently $\mathbf{U}_1 = \mathbf{U}_3$, $\mathbf{U}_2 = (\mathbf{U}_1, \mathbf{U}_4)$ with $\mathbf{U}_1^T\mathbf{U}_4 = 0$. By the arguments above (1.3), the largest eigenvalue of

$$\big(\mathbf{X}\mathbf{U}_2\mathbf{U}_2^T\mathbf{X}^T\big)^{-1}\mathbf{X}\mathbf{U}_1\mathbf{U}_1^T\mathbf{X}^T = \big(\mathbf{X}\mathbf{U}_1\mathbf{U}_1^T\mathbf{X}^T + \mathbf{X}\mathbf{U}_4\mathbf{U}_4^T\mathbf{X}^T\big)^{-1}\mathbf{X}\mathbf{U}_1\mathbf{U}_1^T\mathbf{X}^T$$

can be transferred to a function of the largest eigenvalue of $(\mathbf{X}\mathbf{U}_4\mathbf{U}_4^T\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{U}_1\mathbf{U}_1^T\mathbf{X}^T$ so that Theorem 2.1 is applicable. Therefore, one can also work out the asymptotic distribution for the largest eigenvalue of the matrix $(\mathbf{X}\mathbf{U}_2\mathbf{U}_2^T\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{U}_1\mathbf{U}_1^T\mathbf{X}^T$ under the condition $\mathbf{U}_1^T\mathbf{U}_2 = (\mathbf{I}_{N_1}, 0)$.

Remark 5. Theorem 2.1 can be extended to the joint distribution of the first k largest eigenvalues, that is,

$$\begin{aligned}&\lim_{N\to\infty} P\big(\sigma_N N_1^{2/3}(\lambda_1 - \mu_N) \le s_1, \ldots, \sigma_N N_1^{2/3}(\lambda_k - \mu_N) \le s_k\big)\\ &\quad= \lim_{N\to\infty} P\big(N_1^{2/3}\big(\lambda_1^{\mathrm{GOE}} - 2\big) \le s_1, \ldots, N_1^{2/3}\big(\lambda_k^{\mathrm{GOE}} - 2\big) \le s_k\big),\end{aligned} \tag{2.7}$$

where $\lambda_1^{\mathrm{GOE}} \ge \cdots \ge \lambda_k^{\mathrm{GOE}}$ are the first k largest eigenvalues of $N_1 \times N_1$ GOE matrix and k is a finite number independent of N. In fact, such an extension can be accomplished by a discussion parallel to Corollary 3.19 of [14] since we show the local behavior of the Stieltjes transform near the edge (such as Theorem 8.1). Here we omit the proof.

In the simulations of Section 7.1 to study the Tracy–Widom approximation accuracy of Theorem 2.1, we set two initial choices for the dimensions (M_1, N_1, N_2, N), $M^{(1)} = (M_1, N_1, N_2, N) = (5, 8, 10, 30)$ and $M^{(2)} = (M_1, N_1, N_2, N) = (15, 8, 10, 50)$. The magnification factor k attached to the initial choices changed from 1 to 20, so in Table 1, one can see the results under dimensions $kM^{(i)}$. Although our Tracy–Widom approximation is theoretically developed for large dimension sets, that is, when k is large, we observe that it keeps accurate for small dimensions regardless of the data distribution, see the case when $k = 1$ and correspondingly $M_1 = 5$ in Table 1, the estimated quantiles are well matched with theoretical ones. In the following applications, M_1 corresponds to the data dimension. That is to say, our Tracy–Widom approximation can be applied even when the data dimension is as small as 5. We guess the reason is that in our numerical study, $\tilde{\mu}$ and $\tilde{\sigma}$ (see Section 7) are used to replace μ_N and σ_N respectively, which are calculated explicitly by orthogonal polynomial (see [12]). Such replacement may make the approximation more accurate and thus improves the empirical performance. Moreover, to get a rule of thumb how small the dimensions could be, we also try other values. According to our empirical experience, when the dimension M_1 is 2, (N_1, N_2, N) can be as small as $(2, 3, 8)$. When the dimension is M_1 is 4, (N_1, N_2, N) can be as small as $(2, 2, 11)$. And when the dimension M_1 is 5, (N_1, N_2, N) can be as small as $(2, 3, 14)$.

In the following Sections 3, 5 and 6, we propose three applications of this limiting Tracy–Widom distribution for λ_1. The first one is our motivation of studying Ω as stated in the Introduction, the high-dimensional independence testing by using canonical correlation analysis. The second one is the MANOVA approach in testing the equivalence of g groups' mean vectors. And the last one is the unknown parameter matrix testing in the multivariate linear model. In Section 4, we briefly illustrate how to simplify the unified matrix Ω to the F-matrix.

2.1. Outline of the proof for Theorem 2.1

Before proceeding to the applications, we first give a sketch of the strategy of proof here. Note that the matrix $\mathbf{T}$ does not influence the largest eigenvalue of Ω in (2.1) and hence we can directly work on the matrix $(\mathbf{X}\mathbf{U}_2\mathbf{U}_2^T\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{U}_1\mathbf{U}_1^T\mathbf{X}^T$. However, unlike sample covariance matrices, it involves four $\mathbf{X}$. Moreover $\mathbf{X}\mathbf{U}_1\mathbf{U}_1^T\mathbf{X}^T$ is not independent of $\mathbf{X}\mathbf{U}_2\mathbf{U}_2^T\mathbf{X}^T$ for general $\mathbf{X}$ (not necessarily consisting of Gaussian entries), which makes it even harder to work on this matrix directly. In view of this, we construct a Wigner-type linearization matrix

$$\mathbf{H} = \mathbf{H}(\mathbf{X}) := \begin{pmatrix} -zI & \mathbf{U}_1^T\mathbf{X}^T & 0 \\ \mathbf{X}\mathbf{U}_1 & 0 & \mathbf{X}\mathbf{U}_2 \\ 0 & \mathbf{U}_2^T\mathbf{X}^T & I \end{pmatrix}. \tag{2.8}$$

As will be seen, the linearization matrix is much more convenient when taking derivative with respect to the entries of $\mathbf{X}$ than Ω. By the Schur complement formula (8.4) in the supplement it turns out that the upper-left block of the 3×3 block matrix $\mathbf{H}^{-1}$ is the Stieltjes transform of $\mathbf{U}_1^T\mathbf{X}^T(\mathbf{X}\mathbf{U}_2\mathbf{U}_2^T\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{U}_1$ (one can also refer to (9.9) in the supplement). It then suffices to consider the linearization matrix $\mathbf{H}$ instead. First the strong local law of $\mathbf{H}^{-1}$ around μ_N (Theorem 8.1 in the supplement) is developed which is the main body of the proof. The overall strategy of proving Theorem 8.1 is similar to that used in [14] and it consists of two main parts. Part one is to prove Theorem 8.1 by applying a new Linderberg's comparison approach raised by [14] under the first three moments of the entries of $\mathbf{X}$ matching those of standard Gaussian entries. This part is similar to [10]. However, in order to make this paper more self-consistent and clear, we also repeat the necessary steps but omit some parts done in [10]. Building on part one, part two further proves Theorem 8.1 when the first two moments of the entries of $\mathbf{X}$ match those of standard Gaussian entries (by dropping the 3rd moment matching condition). After that, we use this local law to prove the edge universality (i.e., (2.3) is not affected by the distribution of $\mathbf{X}$) by adopting the strategy stated in [3] and [6]. The proof of Theorem 2.1 is complete by the fact that (2.3) holds because Ω becomes a F matrix when $\mathbf{X}$ consists of the Gaussian random variable (one can refer to Theorem 1 of [12] and Theorem 2.1 of [10]).

We would highlight the difference between the proof of this paper and that of [10]. The result about the edge university for F matrices (corresponding to Ω in the Gaussian case) in [10] is our starting point because we need to use Linderberg's comparison approach to link the edge universality of Ω in the general case to that of F matrices. However, in order to prove the strong local law, a main difficulty is that our main result about Ω doesn't assume $\mathbb{E}\mathbf{Z}_{ij}^3 = 0$ (matching the Gaussian third moment), which is much different from the paper [10] when handling the

dimension is bigger than the sample size there. As a consequence, the expectation of the higher moments of the variable of interest has to be evaluated by a much more complicated method. For example, in order to calculate the higher moments, we need to extract the ith row of $\mathbf{X}$ from $\Pi(z)$ defined at (8.5) in the supplement. However $\Pi(z)$ is a complex function of $\mathbf{X}$, which is not easy to deal with. To handle this, we introduce a transition matrix $\Pi_{\mathbf{1}}(z)$ (defined at (9.55)) to find out a compact and manageable expansion of $\Pi(z)$.

3. Unified matrix in CCA

Suppose that we have two sets of random variables, organized into two random vectors $\mathbf{z} = (z_1, \ldots, z_{M_1})^T$ and $\mathbf{y} = (y_1, \ldots, y_{M_2})^T$, with mean vectors and covariance matrices $(\boldsymbol{\mu}_z, \Sigma_{\mathbf{zz}})$ and $(\boldsymbol{\mu}_y, \Sigma_{\mathbf{yy}})$, respectively. For each of them, N observations are measured and the data matrices are denoted as $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_N)_{M_1 \times N}$ and $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)_{M_2 \times N}$. We want to test

$$H_0 : \mathbf{z} \text{ and } \mathbf{y} \text{ are independent.} \tag{3.1}$$

As illustrated in the Introduction, if $\mathbf{z}$ and $\mathbf{y}$ are independent, the largest eigenvalue ρ_1^2 of the matrix $\Sigma_{\mathbf{zz}}^{-1}\Sigma_{\mathbf{zy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{zy}}^T$ should be zero. The corresponding sample version is

$$\begin{aligned} \mathbf{S}_{zy} &= \left(\sum_{i=1}^N (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T\right)^{-1} \left(\sum_{i=1}^N (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{y}_i - \bar{\mathbf{y}})^T\right) \\ &\quad \times \left(\sum_{i=1}^N (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T\right)^{-1} \left(\sum_{i=1}^N (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{y}_i - \bar{\mathbf{y}})^T\right)^T \end{aligned} \tag{3.2}$$

$$= \left(\mathbf{Z}\mathbf{P}_N\mathbf{Z}^T\right)^{-1}\left(\mathbf{Z}\mathbf{P}_N\mathbf{Y}^T\right)\left(\mathbf{Y}\mathbf{P}_N\mathbf{Y}^T\right)^{-1}\left(\mathbf{Z}\mathbf{P}_N\mathbf{Y}^T\right)^T, \tag{3.3}$$

where $\mathbf{P}_N = \mathbf{I}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$ and $\mathbf{1}_N$ indicates an N-dimensional column vector with all entries being one. Denote the largest eigenvalue of $\mathbf{S}_{zy}$ by $\gamma_1^{\mathbf{S}}$ and let $\lambda_1^{\mathbf{S}} = \frac{\gamma_1^{\mathbf{S}}}{1-\gamma_1^{\mathbf{S}}}$. Note that $\mathbf{P}_N$ is a projection matrix. Then the property of $\lambda_1^{\mathbf{S}}$ is a special case of λ_1 in Theorem 2.1 by observing that we can equivalently consider $\lambda_1^{\mathbf{S}}$ as the largest eigenvalue of the matrix

$$\left(\mathbf{Z}\mathbf{P}_N(I - \mathbf{P}_{Ny})\mathbf{P}_N\mathbf{Z}^T\right)^{-1}\mathbf{Z}\mathbf{P}_N\mathbf{P}_{Ny}\mathbf{P}_N\mathbf{Z}^T,$$

where $\mathbf{P}_{Ny} = (\mathbf{Y}\mathbf{P}_N)^T(\mathbf{Y}\mathbf{P}_N\mathbf{Y}^T)^{-1}(\mathbf{Y}\mathbf{P}_N)$. This equivalence has been specified in the Introduction, see the derivation of (1.3). It is easy to check that $(\mathbf{P}_N(I - \mathbf{P}_{Ny})\mathbf{P}_N)(\mathbf{P}_N\mathbf{P}_{Ny}\mathbf{P}_N) = 0$. Since both $\mathbf{P}_N(I - \mathbf{P}_{Ny})\mathbf{P}_N$ and $\mathbf{P}_N\mathbf{P}_{Ny}\mathbf{P}_N$ are projection matrices such that $\operatorname{rank}(\mathbf{P}_N(I - \mathbf{P}_{Ny})\mathbf{P}_N) = N - M_2$ and $\operatorname{rank}(\mathbf{P}_N\mathbf{P}_{Ny}\mathbf{P}_N) = M_2$ with high probability by Lemma 2 in the supplement, we can take $N_1 = N_2 = M_2$ in Theorem 2.1 to obtain the following Corollary 1.

Corollary 1. *Suppose that the data matrix* $\mathbf{Z}$ *can be written as* $\mathbf{Z} = \mathbf{T}^{\frac{1}{2}}\mathbf{X} + \boldsymbol{\mu}_z\mathbf{1}_N^T$ *for some positive definite matrix* $\mathbf{T}$ *and the matrix* $\mathbf{X}_{M_1 \times N}$ *satisfies Condition* 1. *We do not impose any*

condition on the random vector $\mathbf{y}$. *Here* $\boldsymbol{\mu}_z$ *is the mean vector of* $\mathbf{z}$ *and can be any possible value. Assume that* $\liminf_{N\to\infty} \frac{N}{M_1+M_2} > 1$, $\frac{M_1}{N-M_2}$ *is bounded away from* 0 *and* $\frac{M_2}{M_1}$ *is bounded away from* 0 *and* ∞. *Denote the largest eigenvalue of* $\mathbf{S}_{zy}$ *by* $\gamma_1^{\mathbf{S}}$ *and let* $\lambda_1^{\mathbf{S}} = \frac{\gamma_1^{\mathbf{S}}}{1-\gamma_1^{\mathbf{S}}}$. *Then under the null hypothesis* (3.1), *there exist* μ_N *and* σ_N *such that*

$$\lim_{N\to\infty} P\big(\sigma_N M_2^{2/3}\big(\lambda_1^{\mathbf{S}} - \mu_N\big) \le s\big) = F_1(s),$$

where $F_1(s)$ *is the Type-*1 *Tracy–Widom distribution. Denote the LSD of* $(\mathbf{X}\mathbf{P}_N(I - \mathbf{P}_{N_y}) \times \mathbf{P}_N\mathbf{X}^T)^{-1}$ *by* $F(\lambda)$ *and suppose that* $c_N \in [0, (1 - \sqrt{\frac{M_1}{N-N_1}})^2)$. *Then the centering and scaling parameters* μ_N *and* σ_N *can be decided in the same way as in Theorem* 1 *by replacing* N_1 *and* N_2 *with* M_2.

According to Corollary 1, we suggest to use $\lambda_1^{\mathbf{S}}$ for the hypothesis testing (3.1) by comparing the rescaled $\lambda_1^{\mathbf{S}}$ value with the theoretical critical point obtained from the Type-1 Tracy–Widom distribution. One can also refer to the numerical studies in Section 7.2.

Remark 6. One may notice that there is an additional term $\boldsymbol{\mu}_z\mathbf{1}_N^T$ in the expression of $\mathbf{Z}$ in Corollary 1 compared with the one in Theorem 2.1. This allows the mean vectors to be any possible values. We would like to point out that this mean vector does not influence the analysis of $\lambda_1^{\mathbf{S}}$ due to the observation that $\boldsymbol{\mu}_z\mathbf{1}_N^T\mathbf{P}_N = 0$.

Remark 7. For the Tracy–Widom distribution in Corollary 1, a similar result can be concluded if we assume that the data matrix $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X} + \boldsymbol{\mu}_y\mathbf{1}_N^T$ for some positive definite matrix $\mathbf{T}$ instead and $\boldsymbol{\mu}_y$ is the mean vector of $\mathbf{y}$. In this case, no condition is imposed on the random vector $\mathbf{z}$. And we only need to exchange the roles of M_1 and M_2 in the conclusions of Corollary 1. This is easy to see according to the fact that the largest eigenvalue of $\mathbf{S}_{zy}$ does not change if the roles of $\mathbf{Z}$ and $\mathbf{Y}$ are exchanged in (3.2).

Remark 8. For the case $N < M_1 + M_2$, it is trivial that $\gamma_1^{\mathbf{S}} \equiv 1$ and $\lambda_1^{\mathbf{S}} = +\infty$.

4. Unified matrix in F-matrix model

We would like to point out that the unified matrix Ω not only covers the F-matrix model studied in [10], but also generalizes it to the nonzero mean value case. To see this, choose

$$\mathbf{Z} = (\mathbf{Y}_{M_1\times n_1}, \mathbf{W}_{M_1\times n_2}), \qquad \mathbf{U}_2 = \begin{pmatrix} 0 \\ \mathcal{P}_2 \end{pmatrix}, \qquad \mathbf{U}_1 = \begin{pmatrix} \mathcal{P}_1 \\ 0 \end{pmatrix}$$

with appropriate dimensions, respectively. Let

$$\mathcal{P}_2\mathcal{P}_2^T = \mathbf{I}_{n_2} - \frac{1}{n_2}\mathbf{1}_{n_2}\mathbf{1}_{n_2}^T, \qquad \mathcal{P}_1\mathcal{P}_1^T = \mathbf{I}_{n_1} - \frac{1}{n_1}\mathbf{1}_{n_1}\mathbf{1}_{n_1}^T,$$
$$\mathcal{P}_2^T\mathcal{P}_2 = \mathbf{U}_2^T\mathbf{U}_2 = \mathbf{I}_{N-N_2}, \qquad \mathcal{P}_1^T\mathcal{P}_1 = \mathbf{U}_1^T\mathbf{U}_1 = \mathbf{I}_{N_1},$$

where $\mathbf{1}_{n_i}$ indicates an n_i-dimensional column vector with all entries being one ($i = 1, 2$). Then

$$\Omega = \left(\mathbf{Z}\mathbf{U}_2\mathbf{U}_2^T\mathbf{Z}^T\right)^{-1}\mathbf{Z}\mathbf{U}_1\mathbf{U}_1^T\mathbf{Z}^T = \left(\mathbf{W}\mathcal{P}_2\mathcal{P}_2^T\mathbf{W}^T\right)^{-1}\mathbf{Y}\mathcal{P}_1\mathcal{P}_1^T\mathbf{Y}^T$$
$$= \left[\mathbf{W}\left(\mathbf{I}_{n_2} - \frac{1}{n_2}\mathbf{1}_{n_2}\mathbf{1}_{n_2}^T\right)\mathbf{W}^T\right]^{-1}\mathbf{Y}\left(\mathbf{I}_{n_1} - \frac{1}{n_1}\mathbf{1}_{n_1}\mathbf{1}_{n_1}^T\right)\mathbf{Y}^T.$$

Noticing that the data matrices $\mathbf{W}$ and $\mathbf{Y}$ are centralized in Ω, we thus extend the results of F-matrix under the assumption of zero mean values in [10] to the nonzero mean vectors.

5. Unified matrix in multivariate analysis of variance (MANOVA)

Suppose that we have g populations. Let n_i samples $(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})$ be available from the ith population with mean vector $\boldsymbol{\mu}_i$ (p-dimensional) and common covariance matrix Σ ($i = 1, \ldots, g$). The total sample size is denoted by $n = \sum_{i=1}^{g} n_i$. One frequently discussed problem in multivariate analysis is to investigate whether the g groups have the same mean vector that is,

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_g. \tag{5.1}$$

The MANOVA approach is well-known for this testing problem. Two main SSCPs, the between SSCP $\mathbf{B}$ and the within SSCP $\mathbf{W}$ are constructed as

$$\mathbf{B} = \sum_{i=1}^{g} n_i(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T, \qquad \mathbf{W} = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T,$$

where $\bar{\mathbf{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{x}_{ij}$ is the ith group sample mean and $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{n_i}\mathbf{x}_{ij} = \sum_{i=1}^{g}\frac{n_i}{n}\bar{\mathbf{x}}_i$ is the overall sample mean. The classical testing methods for (5.1) are based on the eigenvalues of the matrix $\mathbf{V} = \mathbf{W}^{-1}\mathbf{B}$. We can show that under the null hypothesis (5.1), the matrix $\mathbf{V}$ can be written as a special form of Ω in Section 2 and thus the limiting distribution of its largest eigenvalue $\lambda_1^{\mathbf{V}}$ follows from Theorem 1.

To see this, denote $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})^T$ (of size $n_i \times p$). Note that under the null hypothesis (5.1), the common mean vector does not influence the matrix $\mathbf{V}$. Then without loss of generality, we can simply assume that $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_g = 0$ under H_0. In this section, we use i to denote the ith group ($i = 1, \ldots, g$) and use j to denote the jth observation from the ith group ($j = 1, \ldots, n_i$). For each $\mathbf{X}_i$, let $\mathbf{H}_i$ be an $n_i \times n_i$ orthogonal matrix with the first column being $\frac{1}{\sqrt{n_i}}\mathbf{1}_{n_i}$. The matrix $\mathbf{I}_{n_i}$ indicates an $n_i \times n_i$ identity matrix, $\mathbf{U}_{i1}$ indicates the first column of $\mathbf{I}_{n_i}$ and $\mathbf{U}_{i2}$ indicates the remaining $n_i \times (n_i - 1)$ block of $\mathbf{I}_{n_i}$. An intuitive example for easy understanding when $n_1 = 3$ is

$$\mathbf{I}_{n_1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{U}_{11} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{U}_{12} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Arrange these $\mathbf{U}_{i1}$'s as blocks placed on the diagonal of a block matrix $\mathbf{U}_1$ and $\mathbf{U}_{i2}$'s as blocks placed on the diagonal of another block matrix $\mathbf{U}_2$, that is,

$$\begin{aligned}
\mathbf{U}_1 &= \begin{pmatrix} \underset{(n_1\times 1)}{\mathbf{U}_{11}} & & & \\ & \underset{(n_2\times 1)}{\mathbf{U}_{21}} & & \\ & & \ddots & \\ & & & \underset{(n_g\times 1)}{\mathbf{U}_{g1}} \end{pmatrix}_{n\times g}, \\
\mathbf{U}_2 &= \begin{pmatrix} \underset{(n_1\times(n_1-1))}{\mathbf{U}_{12}} & & & \\ & \underset{(n_2\times(n_2-1))}{\mathbf{U}_{22}} & & \\ & & \ddots & \\ & & & \underset{(n_g\times(n_g-1))}{\mathbf{U}_{g2}} \end{pmatrix}_{n\times(n-g)}.
\end{aligned} \tag{5.2}$$

Consider the orthogonal transformations $\mathbf{Z}_i = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{in_i})^T = \mathbf{H}_i^T \mathbf{X}_i$ (of size $n_i \times p$). It is easy to find that $\mathbf{z}_{i1} = \sqrt{n_i}\bar{\mathbf{x}}_i$. Furthermore, denote $\mathbf{a}_g = (\sqrt{\frac{n_1}{n}}, \ldots, \sqrt{\frac{n_g}{n}})^T$, $\mathbf{P}_g = \mathbf{I}_g - \mathbf{a}_g\mathbf{a}_g^T$ and $\mathbf{Z} = (\mathbf{Z}_1^T, \mathbf{Z}_2^T, \ldots, \mathbf{Z}_g^T)_{p\times n}$. Considering the relationship $\sqrt{n}\bar{\mathbf{x}} = (\mathbf{z}_{11}, \ldots, \mathbf{z}_{g1})\mathbf{a}_g$, we can obtain

$$\begin{aligned}
\mathbf{B} &= \sum_{i=1}^{g} n_i(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T = \sum_{i=1}^{g} n_i\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i^T - \sqrt{n}\bar{\mathbf{x}}\cdot\sqrt{n}\bar{\mathbf{x}} \\
&= (\mathbf{z}_{11}, \ldots, \mathbf{z}_{g1})\big(\mathbf{I}_g - \mathbf{a}_g\mathbf{a}_g^T\big)(\mathbf{z}_{11}, \ldots, \mathbf{z}_{g1})^T = \mathbf{Z}\mathbf{U}_1\mathbf{P}_g\mathbf{U}_1^T\mathbf{Z}^T = \mathbf{Z}\widetilde{\mathbf{U}}_1\widetilde{\mathbf{U}}_1^T\mathbf{Z}^T, \\
\mathbf{W} &= \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T = \sum_{i=1}^{g}\left(\sum_{j=1}^{n_i}\mathbf{x}_{ij}\mathbf{x}_{ij}^T - n_i\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i^T\right) = \sum_{i=1}^{g}\big(\mathbf{X}_i^T\mathbf{X}_i - \mathbf{z}_{i1}\mathbf{z}_{i1}^T\big) \\
&= \sum_{j=2}^{n_1}\mathbf{z}_{1j}\mathbf{z}_{1j}^T + \sum_{j=2}^{n_2}\mathbf{z}_{2j}\mathbf{z}_{2j}^T + \cdots + \sum_{j=2}^{n_g}\mathbf{z}_{gj}\mathbf{z}_{gj}^T = \mathbf{Z}\mathbf{U}_2\mathbf{U}_2^T\mathbf{Z}^T,
\end{aligned} \tag{5.3}$$

where $\widetilde{\mathbf{U}}_1 = \mathbf{U}_1\mathbf{P}_g$ and $\mathbb{E}(\mathbf{Z}) = 0$ under H_0. According to the construction of $\mathbf{U}_1$ and $\mathbf{U}_2$ in (5.2), we can easily conclude that $\widetilde{\mathbf{U}}_1^T\mathbf{U}_2 = 0$. Then the limiting distribution of the largest eigenvalue $\lambda_1^{\mathbf{V}}$ of

$$\mathbf{V} = \mathbf{W}^{-1}\mathbf{B} = \big(\mathbf{Z}\mathbf{U}_2\mathbf{U}_2^T\mathbf{Z}^T\big)^{-1}\mathbf{Z}\widetilde{\mathbf{U}}_1\widetilde{\mathbf{U}}_1^T\mathbf{Z}^T$$

can follow from Theorem 2.1 by assigning $M_1 = p$, $N_1 = g - 1$ and $N_2 = g$ since $\operatorname{rank}(\widetilde{\mathbf{U}}_1) = g - 1, \operatorname{rank}(\mathbf{U}_2) = n - g$. See the following Corollary 2.

Table 2. Simulated values for $\frac{\#\{\text{rescaled } \lambda_1^{\mathbf{V}} > c_\alpha\}}{10\,000}$ based on 10 000 replications. The "H_0" columns record estimated sizes and other columns record estimated powers. The significance level is $\alpha = 0.05$

	$M^{(1)} = (p, n_0) = (5, 8)$					$M^{(2)} = (p, n_0) = (8, 5)$			
	H_0	$H_1^{(1)}$	$H_1^{(1)'}$	$H_1^{(2)}$		H_0	$H_1^{(1)}$	$H_1^{(1)'}$	$H_1^{(2)}$
$M^{(1)}$	0.0375	0.0831	0.5317	0.5589	$M^{(2)}$	0.0374	0.0511	0.1502	0.1098
$2M^{(1)}$	0.0392	0.2454	0.9955	0.8693	$2M^{(2)}$	0.0399	0.1099	0.7505	0.2449
$4M^{(1)}$	0.0405	0.8535	1.0000	0.9907	$4M^{(2)}$	0.0386	0.4395	1.0000	0.5020
$8M^{(1)}$	0.0414	1.0000	1.0000	1.0000	$8M^{(2)}$	0.0375	0.9956	1.0000	0.8341
$16M^{(1)}$	0.0445	1.0000	1.0000	1.0000	$16M^{(2)}$	0.0424	1.0000	1.0000	0.9897
$32M^{(1)}$	0.0429	1.0000	1.0000	1.0000	$32M^{(2)}$	0.0432	1.0000	1.0000	0.9999
$64M^{(1)}$	0.0396	1.0000	1.0000	1.0000	$64M^{(2)}$	0.0390	1.0000	1.0000	1.0000
$100M^{(1)}$	0.0442	1.0000	1.0000	1.0000	$100M^{(2)}$	0.0452	1.0000	1.0000	1.0000

Corollary 2. *Consider the multivariate mean vectors' hypothesis testing problem in* (5.1). *We use the largest eigenvalue* $\lambda_1^{\mathbf{V}}$ *of the matrix* $\mathbf{V} = \mathbf{W}^{-1}\mathbf{B}$ *as the test criterion. Under the null hypothesis, suppose that* $\mathbf{Z}$ *can be written as* $\mathbf{Z} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$ *for some positive definite matrix* $\mathbf{T}_{p\times p}$ *and the matrix* $\mathbf{X}_{p\times n}$ *satisfies Condition* 1. *Assume that* $\liminf_{n\to\infty} \frac{n}{p+g} > 1$ *and* $\frac{g-1}{p}$ *is bounded away from* 0 *and* ∞. *Then there exist* μ_n *and* σ_n *such that*

$$\lim_{n\to\infty} P\big(\sigma_n(g-1)^{2/3}\big(\lambda_1^{\mathbf{V}} - \mu_n\big) \le s\big) = F_1(s),$$

where $F_1(s)$ *is the Type-*1 *Tracy–Widom distribution. The centering and scaling parameters* μ_n *and* σ_n *can be decided in the same way as in Theorem* 2.1 *by replacing* M_1 *with* p *and* N_1 *with* $(g-1)$.

According to Corollary 2, if the rescaled $\lambda_1^{\mathbf{V}}$ value is smaller than the theoretical critical point obtained from Type-1 Tracy–Widom distribution, we fail to reject the null hypothesis (5.1), that is, we accept that the g groups share the same mean vector. Otherwise, reject H_0. In the simulation studies of Section 7.3, regarding the pattern of different mean vectors under the alternative, we consider two cases. One is the dense but weak alternative (DWA), which means that there are many different entries among the mean vectors, but these differences are faint, see the setting $H_1^{(1)}$ and $H_1^{(1)'}$ in Section 7.3(1). The other one is the sparse but strong alternative (SSA), which means that the differences are rare, but significant where they appear, see the alternative $H_1^{(2)}$, where the differences only appear in one out of p components. The numerical results in Table 2 indicate that this $\lambda_1^{\mathbf{V}}$ shows satisfactory performance for both alternatives.

Remark 9. If we assume that all the observations come from multivariate normal distribution as in the classical setting, then the positive definite matrix $\mathbf{T}$ in Corollary 2 obviously exists by choosing $\mathbf{T} = \Sigma$. This is due to the fact that we can write each $\mathbf{X}_i$ as $\mathbf{X}_i^T = \Sigma^{\frac{1}{2}}\widetilde{\mathbf{X}}_i = \mathbf{T}^{\frac{1}{2}}\widetilde{\mathbf{X}}_i$ and

the entries of $\widetilde{\mathbf{X}}_i$ are i.i.d. $N(0,1)$. Then

$$\begin{aligned}\mathbf{Z} &= \left(\mathbf{Z}_1^T, \mathbf{Z}_2^T, \ldots, \mathbf{Z}_g^T\right) = \left(\mathbf{X}_1^T\mathbf{H}_1, \mathbf{X}_2^T\mathbf{H}_2, \ldots, \mathbf{X}_g^T\mathbf{H}_g\right) = \mathbf{T}^{\frac{1}{2}}(\widetilde{\mathbf{X}}_1\mathbf{H}_1, \widetilde{\mathbf{X}}_2\mathbf{H}_2, \ldots, \widetilde{\mathbf{X}}_g\mathbf{H}_g)\\ &:= \mathbf{T}^{\frac{1}{2}}\mathbf{X},\end{aligned}$$

where $\mathbf{X} = (\widetilde{\mathbf{X}}_1\mathbf{H}_1, \widetilde{\mathbf{X}}_2\mathbf{H}_2, \ldots, \widetilde{\mathbf{X}}_g\mathbf{H}_g)_{p\times n}$ satisfies Condition 1, taking into account the orthogonality of each $\mathbf{H}_i$ and the independence among each $\widetilde{\mathbf{X}}_i$.

6. Unified matrix in high-dimensional multivariate linear model

In this section, we investigate one more application of the unified matrix Ω in the multivariate linear model. Let us consider a linear relationship between p_2 response variables $y_1, \ldots, y_{p_2}$ and p_1 explanatory variables $x_1, \ldots, x_{p_1}$. Suppose that there are N observations, organized into two data matrices:

$$\mathbf{Y} = \begin{pmatrix}\mathbf{Y}_1^T\\ \vdots\\ \mathbf{Y}_N^T\end{pmatrix}_{N\times p_2}, \qquad \mathbf{X} = \begin{pmatrix}\mathbf{X}_1^T\\ \vdots\\ \mathbf{X}_N^T\end{pmatrix}_{N\times p_1}.$$

Then the multivariate linear model assumes that

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{Z}, \tag{6.1}$$

where $\mathbf{B}$ is a $p_1 \times p_2$ unknown parameter matrix and $\mathbf{Z}$ is a $N \times p_2$ error matrix with the assumption that the rows of $\mathbf{Z}$ are independent having mean zero and common covariance matrix Σ. We first consider the linear hypothesis testing of the form

$$H_0 : \mathbf{C}_1\mathbf{B} = \mathbf{\Gamma}_1, \tag{6.2}$$

where $\mathbf{C}_1$ is a $g_1 \times p_1$ known matrix of rank g_1 and $\mathbf{\Gamma}_1$ is a $g_1 \times p_2$ known matrix of rank $\min\{g_1, p_2\}$. As an example, in the simulation studies of Section 7.4, if we select $\mathbf{C}_1 = \mathbf{C}_1^{(b)} = [\mathbf{I}_{g_1}, 0]$ and $\mathbf{\Gamma}_1 = \mathbf{\Gamma}_1^{(a)} = 0$, then the testing problem (6.2) reduces to analyzing whether the first g_1 rows of $\mathbf{B}$ equal to zeros.

The initial step in conducting the linear hypothesis testing (6.2) is to estimate the unknown parameter matrix $\mathbf{B}$. As stated in Section 2, our proposed Tracy–Widom distribution performs well when the dimensions are small so that we can simply apply the classic least square estimator for $\mathbf{B}$, which is well known to be $\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$. The hypothesis SSCP for testing (6.2) is given by $\mathbf{H}_1 = (\mathbf{C}_1\hat{\mathbf{B}} - \mathbf{\Gamma}_1)^T[\mathbf{C}_1(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}_1^T]^{-1}(\mathbf{C}_1\hat{\mathbf{B}} - \mathbf{\Gamma}_1)$ and the error SSCP is $\mathbf{E}_1 = \mathbf{Y}^T[\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\mathbf{Y}$. One can refer to Chapter 7 of [7] for detailed derivations. Under the null hypothesis (6.2), $\mathbf{H}_1$ and $\mathbf{E}_1$ can be further rewritten as

$$\begin{aligned}\mathbf{H}_1 &= \left[\mathbf{C}_1\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Z}\right]^T\left[\mathbf{C}_1\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{C}_1^T\right]^{-1}\left[\mathbf{C}_1\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Z}\right] = \mathbf{Z}^T\mathbf{P}_{\widetilde{\mathbf{X}}}\mathbf{Z},\\ \mathbf{E}_1 &= (\mathbf{X}\mathbf{B} + \mathbf{Z})^T\left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right](\mathbf{X}\mathbf{B} + \mathbf{Z}) = \mathbf{Z}^T[\mathbf{I} - \mathbf{P}_{\mathbf{X}}]\mathbf{Z},\end{aligned} \tag{6.3}$$

where $\widetilde{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}_1^T$, $\mathbf{P}_{\widetilde{\mathbf{X}}} = \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^T$ and $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. It is easy to check that $\mathbf{P}_{\widetilde{\mathbf{X}}}(\mathbf{I} - \mathbf{P}_{\mathbf{X}}) = 0$. Denote the largest eigenvalue of

$$\mathbf{M}_1 = \mathbf{E}_1^{-1}\mathbf{H}_1 = \big(\mathbf{Z}^T(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Z}\big)^{-1}\mathbf{Z}^T\mathbf{P}_{\widetilde{\mathbf{X}}}\mathbf{Z} \tag{6.4}$$

by $\lambda_1^{\mathbf{M}_1}$. As stated in Section 3, both $\mathbf{I} - \mathbf{P}_{\mathbf{X}}$ and $\mathbf{P}_{\widetilde{\mathbf{X}}}$ are projection matrices with $\operatorname{rank}(\mathbf{I} - \mathbf{P}_{\mathbf{X}}) = N - p_1$ and $\operatorname{rank}(\mathbf{P}_{\widetilde{\mathbf{X}}}) = g_1$ with high probability. Assuming $N_2 = p_1$, $N_1 = g_1$ and $M_1 = p_2$ in Theorem 2.1, we can develop the following corollary for $\lambda_1^{\mathbf{M}_1}$.

Corollary 3. *Assume that* $\mathbf{Z}$ *in the multivariate linear model* (6.1) *can be written as* $\mathbf{Z} = \mathbf{W}\mathbf{T}^{\frac{1}{2}}$ *for some positive definite matrix* $\mathbf{T}_{p_2\times p_2}$ *and the matrix* $\mathbf{W}_{N\times p_2}$ *satisfies Condition* 1. *Suppose that* $\liminf_{N\to\infty}\frac{N}{p_2+p_1} > 1$, $\frac{g_1}{p_1}$ *and* $\frac{p_2}{N-p_1}$ *are both bounded away from* 0 *and* $\frac{g_1}{p_2}$ *is bounded away from* 0 *and* ∞. *Denote the largest eigenvalue of* $\mathbf{M}_1 = \mathbf{E}_1^{-1}\mathbf{H}_1$ *by* $\lambda_1^{\mathbf{M}_1}$. *Then under the null hypothesis* (6.2), *there exist* μ_N *and* σ_N *such that*

$$\lim_{N\to\infty} P\big(\sigma_N g_1^{2/3}\big(\lambda_1^{\mathbf{M}_1} - \mu_N\big) \le s\big) = F_1(s),$$

where $F_1(s)$ *is the Type-*1 *Tracy–Widom distribution. Denote the LSD of* $(\mathbf{W}^T(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{W})^{-1}$ *by* $F(\lambda)$ *and suppose that* $c_N \in [0, (1 - \sqrt{\frac{p_2}{N-g_1}})^2]$. *Then the centering and scaling parameters* μ_N *and* σ_N *can be decided in the same way as in Theorem* 2.1 *by replacing* N_2 *with* p_1, N_1 *with* g_1 *and* M_1 *with* p_2.

Remark 10. One should notice that $\mathbf{Z}$ in this Corollary and Corollary 4 corresponds to $\mathbf{Z}^T$ in Theorem 2.1. To see this, one may compare (6.4) with (2.1).

By Corollary 3, we can use $\lambda_1^{\mathbf{M}_1}$ for the linear hypothesis testing (6.2) and reject H_0 if the rescaled $\lambda_1^{\mathbf{M}_1}$ is larger than the theoretical critical point obtained from Type-1 Tracy–Widom distribution. In Section 7.4, we consider the special testing of whether a certain part of $\mathbf{B}$, say $\mathbf{B}_2$, equals a zero matrix. And as in MANOVA, with regard to the pattern under the alternative, both DWA and SSA are applied, i.e. when many entries of $\mathbf{B}_2$ are nonzero but the values are small, see the third combination $(\mathbf{C}_1^{(a)}, \mathbf{B}_2^{(d)}, \mathbf{\Gamma}_1^{(a)})$, as well as when only two entries of $\mathbf{B}_2$ are nonzero but the values are significant, see the last combination $(\mathbf{C}_1^{(a)}, \mathbf{B}_2^{(s)}, \mathbf{\Gamma}_1^{(a)})$. The numerical results in Table 3 show that $\lambda_1^{\mathbf{M}_1}$ performs well under both alternatives.

We next consider the intra-subject hypothesis testing of the form

$$H_0 : \mathbf{CBD} = \mathbf{\Gamma}, \tag{6.5}$$

where $\mathbf{C}$ is a $g_1 \times p_1$ known matrix of rank g_1, $\mathbf{D}$ is a $p_2 \times g_2$ known matrix of rank g_2 and $\mathbf{\Gamma}$ is a $g_1 \times g_2$ known matrix of rank $\min\{g_1, g_2\}$. The hypothesis and error SSCPs for (6.5) can be obtained from $\mathbf{H}_1$ and $\mathbf{E}_1$ by modifying the multivariate linear model (6.1) to the following expression

$$\mathbf{YD} = \mathbf{XBD} + \mathbf{ZD}.$$

Table 3. Simulated values for $\frac{\#\{\text{rescaled } \lambda_1^{\mathbf{M}_1} > c_\alpha\}}{10\,000}$ based on 10 000 replications. The first two combinations record estimated sizes and the last two record estimated powers. The significance level is $\alpha = 0.05$

$(\mathbf{C}_1, \mathbf{B}_2, \mathbf{\Gamma}_1)$	$M^{(0)}$	$2M^{(0)}$	$3M^{(0)}$	$4M^{(0)}$	$6M^{(0)}$	$8M^{(0)}$	$10M^{(0)}$	$20M^{(0)}$
				$M^{(0)} = (p_1, p_2, N) = (10, 6, 25)$				
$(\mathbf{C}_1^{(b)}, \mathbf{B}_2^{(d)}, \mathbf{\Gamma}_1^{(a)})$	0.0400	0.0447	0.0453	0.0469	0.0487	0.0460	0.0466	0.0468
$(\mathbf{C}_1^{(a)}, \mathbf{B}_2^{(d)}, \mathbf{\Gamma}_1^{(b)})$	0.0397	0.0467	0.0450	0.0490	0.0466	0.0470	0.0501	0.0481
$(\mathbf{C}_1^{(a)}, \mathbf{B}_2^{(d)}, \mathbf{\Gamma}_1^{(a)})$	0.2298	0.8923	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
$(\mathbf{C}_1^{(a)}, \mathbf{B}_2^{(s)}, \mathbf{\Gamma}_1^{(a)})$	0.8337	0.9451	0.9821	0.9940	0.9992	1.0000	0.9999	1.0000

Replacing $\mathbf{Y}$, $\mathbf{B}$ and $\mathbf{Z}$ by $\mathbf{YD}$, $\mathbf{BD}$ and $\mathbf{ZD}$ respectively, we can then conclude that the SSCPs for conducting the hypothesis testing (6.5) are

$$\mathbf{H} = (\mathbf{ZD})^T \mathbf{P}_{\widetilde{\mathbf{X}}} (\mathbf{ZD}), \qquad \mathbf{E} = (\mathbf{ZD})^T [\mathbf{I} - \mathbf{P}_{\mathbf{X}}] (\mathbf{ZD}), \tag{6.6}$$

where $\widetilde{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T$, $\mathbf{P}_{\widetilde{\mathbf{X}}} = \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^T$ and $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. It is easy to check that $\mathbf{P}_{\widetilde{\mathbf{X}}}(\mathbf{I} - \mathbf{P}_{\mathbf{X}}) = 0$. Denote the largest eigenvalue of $\mathbf{M} = \mathbf{E}^{-1}\mathbf{H}$ by $\lambda_1^{\mathbf{M}}$. The only difference between the analysis of $\lambda_1^{\mathbf{M}_1}$ and $\lambda_1^{\mathbf{M}}$ is that $\mathbf{Z}_{N\times p_2}$ in (6.3) is replaced by $(\mathbf{ZD})_{N\times g_2}$ in (6.6). So assigning $p_2 = g_2$ in Corollary 3, we can obviously obtain the following conclusion for $\lambda_1^{\mathbf{M}}$.

Corollary 4. *For the known matrix* $\mathbf{D}$ *and the error matrix* $\mathbf{Z}$ *in the multivariate linear model* (6.1), *assume that* $\mathbf{ZD}$ *can be written as* $\mathbf{ZD} = \mathbf{WT}^{\frac{1}{2}}$ *for some positive definite matrix* $\mathbf{T}_{g_2\times g_2}$ *and the matrix* $\mathbf{W}_{N\times g_2}$ *satisfies Condition* 1. *Suppose that* $\liminf_{N\to\infty} \frac{N}{g_2+p_1} > 1$, $\frac{g_1}{p_1}$ *and* $\frac{g_2}{N-p_1}$ *are both bounded away from* 0 *and* $\frac{g_1}{g_2}$ *is bounded away from* 0 *and* ∞. *Denote the largest eigenvalue of* $\mathbf{M} = \mathbf{E}^{-1}\mathbf{H}$ *by* $\lambda_1^{\mathbf{M}}$. *Then under the null hypothesis* (6.2), *there exist* μ_N *and* σ_N *such that*

$$\lim_{N\to\infty} P\big(\sigma_N g_1^{2/3}\big(\lambda_1^{\mathbf{M}} - \mu_N\big) \le s\big) = F_1(s),$$

where $F_1(s)$ *is the Type*-1 *Tracy–Widom distribution. The centering and scaling parameters* μ_N *and* σ_N *can be decided in the same way as in Corollary* 3 *by replacing* p_2 *with* g_2.

7. Numerical studies

This section is to investigate the accuracy of our proposed asymptotic Tracy–Widom distribution (Section 7.1) as well as its numerical performance in various applications (Sections 7.2–7.4). Before proceeding to the simulation results, we first introduce an asymptotic substitution of the limiting distribution for the largest eigenvalue in Theorem 2.1. The formulae for calculating μ_N and σ_N in (2.4)–(2.6) are difficult to work with. Referring to [12] and [10], we facilitate the

computation by using an approximation in terms of the log transform of λ_1 in Theorem 2.1 as

$$\lim_{N\to\infty} P\left(\frac{\ln\lambda_1 - \widetilde{\mu}}{\widetilde{\sigma}} \le s\right) = F_1(s), \tag{7.1}$$

where $F_1(s)$ still indicates the Type-1 Tracy–Widom distribution and $\widetilde{\mu}$ and $\widetilde{\sigma}$ are defined by

$$\widetilde{\mu} = 2\ln\tan\left(\frac{\phi+\varphi}{2}\right), \qquad \widetilde{\sigma}^3 = \frac{16}{(N-N_2+N_1-1)^2}\frac{1}{\sin^2(\phi+\varphi)\sin\phi\sin\varphi}.$$

The angle parameters ϕ and φ are defined by

$$\sin^2\left(\frac{\varphi}{2}\right) = \frac{\min(M_1,N_1)-1/2}{N-N_2+N_1-1}, \qquad \sin^2\left(\frac{\phi}{2}\right) = \frac{\max(M_1,N_1)-1/2}{N-N_2+N_1-1}.$$

The asymptotic equivalence between the approximation (7.1) and the one in Theorem 2.1 have been proved in [12] and [10]. All simulations in this section are conducted by adopting this $\ln\lambda_1$'s asymptotic expression. In the sequel, we also use the word "rescaled λ_1" to denote the term $\frac{\ln\lambda_1-\widetilde{\mu}}{\widetilde{\sigma}}$ in (7.1). The values of $\widetilde{\mu}$ and $\widetilde{\sigma}$ in the applications can be obtained simply by replacing N, N_1, N_2, M_1 with their corresponding notations in Sections 3–6. All simulated results below are recorded based on 10 000 replications of such a re-scaled largest eigenvalue.

7.1. Approximation accuracy

This subsection is to investigate the Tracy–Widom approximation accuracy for the unified matrix Ω in Section 2. Since the positive definite matrix $\mathbf{T}$ does not influence λ_1, we simply let $\mathbf{T} = \mathbf{I}_{M_1}$. Other settings to be used in the simulation are summarized below.

(1) *Data distribution*: Three data distributions will be used to generate the entries of $\mathbf{X}$ in the model (2.1).

- Data 1: Standard Normal distribution $N(0,1)$.
- Data 2: Discrete distribution with probability mass function $P(x=-\sqrt{3}) = P(x=\sqrt{3}) = 1/6$ and $P(x=0) = 2/3$.
- Data 3: Standardized Gamma distribution Gamma$(4, 0.5)$.

The three distributions are used to verify Condition 1, that is, for the data distribution, we do not need other restrictions except for the first two moments match and all moments are finite. Data 2 supports that the distribution can be a discrete one, while Data 3 is a skewed one with the third and fourth moments different from those of the standard normal distribution.

(2) *Dimensions* (M_1, N_1, N_2, N): Considering the restrictions on the dimensions, we set two initial choices: $M^{(1)} = (M_1, N_1, N_2, N) = (5, 8, 10, 30)$ and $M^{(2)} = (M_1, N_1, N_2, N) = (15, 8, 10, 50)$, with M_1 being smaller than (N_1, N_2) and larger than (N_1, N_2), respectively. Then we change the magnification factor attached to the initial choices to investigate the performance when the dimensions increase. See the second row of Table 1.

(3) *Matrices* $\mathbf{U}_1$ *and* $\mathbf{U}_2$: We randomly generate two matrices $\mathbf{L}_{N\times N_2}$ and $\mathbf{D}_{N_2\times N_1}$ with entries from standard normal distribution. Let $\mathbf{U}_1\mathbf{U}_1^T = (\mathbf{LD})(\mathbf{D}^T\mathbf{L}^T\mathbf{LD})^{-1}(\mathbf{LD})^T$ and $\mathbf{U}_2\mathbf{U}_2^T = \mathbf{I}_N - \mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T$ in the model (2.1). It is easy to check that such settings satisfy the conditions on $\mathbf{U}_1$ and $\mathbf{U}_2$, taking into account the properties of projection matrices.

Simulated results based on above settings are recorded in Table 1. The column titled "Percentile" lists the percentiles of Tracy–Widom distribution corresponding to quantiles in the column "TW". The next ten columns record our estimated cumulative probabilities (i.e., estimated quantiles) for the rescaled λ_1 under various settings stated above, that is, repeating 10 000 times and finding 10 000 rescaled λ_1's, then the proportion of values that are less than corresponding percentiles are recorded in Table 1 that is, $\frac{\#\{\text{rescaled } \lambda_1 \le \text{"Percentile"}\}}{10\,000}$. Comparing the empirical results (the last ten columns) with the theoretical ones (the "TW" column), we can see that the rescaled λ_1 matches with the Tracy–Widom law quite well, which supports the accuracy of approximation in Theorem 2.1. Moreover, although our theoretical result is developed for large dimensions, Table 1 indicates that such approximation also works well even when the dimensions are small.

7.2. Performance in the independence testing

This subsection is to investigate the performance of our proposed largest eigenvalue $\lambda_1^{\mathbf{S}}$ in the independence testing of Section 3. For ease of construction, we let $M_1 = M_2$ and consider a series of settings for the two random vectors $\mathbf{z}$ and $\mathbf{y}$ in the following way:

$$\mathbf{z} = \sqrt{1-\tau}\mathbf{x} + \sqrt{\tau}\mathbf{y}, \qquad 0 \le \tau \le 1,$$

where two $(M_1 \times 1)$ random vectors $\mathbf{x}$ and $\mathbf{y}$ are independent and τ is a parameter determining the level of dependence between $\mathbf{z}$ and $\mathbf{y}$. When $\tau = 0$, $\mathbf{z}$ and $\mathbf{y}$ are independent, which is the null hypothesis (3.1) in Section 3. Otherwise, as $\tau > 0$ becomes larger, the dependence between $\mathbf{z}$ and $\mathbf{y}$ increases.

Considering the conditions on the dimensions, as in Section 7.1, we also set an initial choice for (M_1, M_2, N) as $M^{(0)} = (M_1, M_2, N) = (10, 10, 40)$ and then change the magnification factor to check the influence of dimensionality. The nominal significance level is set to be $\alpha = 0.05$. According to Table 1, the corresponding theoretical quantile value is $c_\alpha = 0.98$. That is to say, we compare the rescaled $\lambda_1^{\mathbf{S}}$ introduced in Section 3 with c_α. If it is smaller than c_α, then the null hypothesis (3.1) is accepted, i.e. $\mathbf{z}$ and $\mathbf{y}$ are independent. Otherwise, we conclude that they are dependent. We use discrete distribution or Gamma distribution, stated in above Section 7.1(1), to generate N samples for $\mathbf{x}$ and $\mathbf{y}$. Repeating 10 000 times, we can find 10 000 rescaled $\lambda_1^{\mathbf{S}}$'s and the proportion of values that are larger than c_α are recorded in Table 4 that is, $\frac{\#\{\text{rescaled } \lambda_1^{\mathbf{S}} > c_\alpha\}}{10\,000}$.

So when $\tau = 0$, the fourth row of Table 4 records the estimated sizes, which are close to 0.05. When τ changes from 0.1 to 0.4, the corresponding rows give the estimated powers. We can observe that as the dependence between $\mathbf{z}$ and $\mathbf{y}$ becomes stronger and as the dimensions become larger, the power values increase. We do not attach the results when $\tau > 0.4$ here because the powers are always around 1. One can also expect such a phenomenon according to the trend in Table 4.

Table 4. Simulated values for $\frac{\#\{\text{rescaled } \lambda_1^{\mathbf{S}} > c_\alpha\}}{10\,000}$ based on 10 000 replications. So "$\tau = 0$" row records estimated sizes and other rows record estimated powers. The significance level is $\alpha = 0.05$

	Discrete distribution					Gamma distribution				
τ	$M^{(0)}$	$2M^{(0)}$	$4M^{(0)}$	$8M^{(0)}$	$10M^{(0)}$	$M^{(0)}$	$2M^{(0)}$	$4M^{(0)}$	$8M^{(0)}$	$10M^{(0)}$
					$M^{(0)} = (M_1, M_2, N) = (10, 10, 40)$					
0	0.0663	0.0618	0.0622	0.0608	0.0559	0.0672	0.0663	0.0591	0.0589	0.0563
0.1	0.2766	0.5049	0.8428	0.9978	0.9998	0.2932	0.5117	0.8540	0.9981	1.0000
0.15	0.4533	0.7754	0.9915	1.0000	1.0000	0.4641	0.7887	0.9909	1.0000	1.0000
0.2	0.6280	0.9396	0.9999	1.0000	1.0000	0.6483	0.9463	1.0000	1.0000	1.0000
0.25	0.7828	0.9911	1.0000	1.0000	1.0000	0.7959	0.9934	1.0000	1.0000	1.0000
0.3	0.8959	0.9997	1.0000	1.0000	1.0000	0.9113	0.9997	1.0000	1.0000	1.0000
0.4	0.9908	1.0000	1.0000	1.0000	1.0000	0.9920	1.0000	1.0000	1.0000	1.0000

7.3. Performance in MANOVA

This subsection is to investigate the performance of our proposed largest eigenvalue $\lambda_1^{\mathbf{V}}$ in the MANOVA approach of Section 5. The nominal significance level is set to be $\alpha = 0.05$. Consider $g = 3$ groups with mean vectors $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3$ and common covariance matrix Σ. We select Σ as the covariance matrix of MA(1) model with the parameter $\theta_1 = 0.2$ and use Gamma distribution stated in Section 7.1(1) to generate the data. Other settings that will be used in the simulation are summarized below.

(1) *Mean vectors*: Let $\boldsymbol{\mu}_1 = 0_p$, a p-dimensional zero vector, $\mathbf{a}_1 = (\tau_1, \ldots, \tau_1)^T$, a p-dimensional vector with all entries being τ_1 and $\mathbf{a}_2 = (\tau_2, 0, \ldots, 0)^T$, a p-dimensional vector with only the first entry having a nonzero value τ_2. Three different settings on the mean vectors are considered.

- H_0: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = 0_p$. This setting corresponds to the null hypothesis (5.1) in Section 5. It is used to check the empirical size performance when the null hypothesis is true. Both of the following two settings are under the alternative hypothesis, that is, the three groups do not share the same mean vector.
- $H_1^{(1)}$ and $H_1^{(1)'}$: $\boldsymbol{\mu}_1 = 0_p$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 + \mathbf{a}_1$ and $\boldsymbol{\mu}_3 = \boldsymbol{\mu}_2 + \mathbf{a}_1$. This setting reflects the dense but weak alternative (DWA), which means that there are many different entries, but these differences are faint. We choose $\tau_1 = 0.2$ for $H_1^{(1)}$ and a larger $\tau_1 = 0.5$ for $H_1^{(1)'}$. The magnitude of the difference vector $\mathbf{a}_1$ is $\|\mathbf{a}_1\|^2 = \tau_1^2 p = 0.04p$ or $0.25p$.
- $H_1^{(2)}$: $\boldsymbol{\mu}_1 = 0_p$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 + \mathbf{a}_2$ and $\boldsymbol{\mu}_3 = \boldsymbol{\mu}_2 + \mathbf{a}_2$. This setting reflects the sparse but strong alternative (SSA), which means that the differences are rare, but significant where they appear. We choose $\tau_2 = 1$. Then the magnitude of the difference vector $\mathbf{a}_2$ is always 1.

(2) *Dimensions* (n_0, p): For simplicity, let $n_1 = n_2 = n_3 := n_0$. Then $n = 3n_0$. We select two initial choices for (n_0, p) as $M^{(1)} = (p, n_0) = (5, 8)$ and $M^{(2)} = (p, n_0) = (8, 5)$,

with $p < n_0$ and $p > n_0$, respectively. Then we change the magnification factor for the initial choices from 1 to 100 (see the first and sixth columns of Table 2) to investigate the influence of dimensions on the numerical performance.

As in the above Section 7.2, by repeating 10 000 times, we can find 10 000 rescaled $\lambda_1^{\mathbf{V}}$'s and the proportion of values that are larger than c_α are recorded in Table 2 that is, $\frac{\#\{\text{rescaled } \lambda_1^{\mathbf{V}} > c_\alpha\}}{10\,000}$. The two columns titled "H_0" record estimated sizes, from which we can see that the size performance becomes better as the dimensions become larger. This matches with our theoretical conclusion, which relies on $n \to \infty$. Other columns report estimated powers under different mean vectors' settings. Generally speaking, the powers increase fast as the dimensions become larger, say the power values of the $8M^{(i)}$ row already all exceed 0.8. And for small dimensions, the $M^{(1)}$ domain shows better performance than $M^{(2)}$, which indicates that $\lambda_1^{\mathbf{V}}$ prefers $p < n_0$ when both p and n_0 are small. However, for moderate and large dimensions, such preference will be weakened since all the power values are close to 1.

7.4. Performance in multivariate linear model

This subsection is to investigate the performance of our proposed largest eigenvalue $\lambda_1^{\mathbf{M}_1}$ in the multivariate linear model of Section 6. The nominal significance level is set to be $\alpha = 0.05$. The covariance matrix Σ of the error matrix $\mathbf{Z}$ is selected to be a Toeplize matrix with first row $(1, 0.5, 0.5^2, 0.5^3, \ldots, 0.5^{p-1})$, that is, the covariance matrix for the AR(1) model with the parameter $\sigma_1 = 0.5$. And we use Gamma distribution stated in Section 7.1(1) to generate the data $\mathbf{Z}$. According to Section 6, the distribution of $\mathbf{X}$ does not influence the result. So we simply obtain the entries $\mathbf{X}$ from a uniform distribution $U(-2, 2)$. Considering the conditions on the dimensions, we set an initial choice for (p_1, p_2, N) as $M^{(0)} = (p_1, p_2, N) = (10, 6, 25)$ and then change the magnification factor from 1 to 20 to check the influence of dimensionality. Other settings for the model (6.1) that will be used in the simulation are summarized below.

(1) *Parameter matrix* $\mathbf{B}$: Set $\mathbf{B} = \begin{pmatrix} (\mathbf{B}_1)_{g_1 \times p_2} \\ (\mathbf{B}_2)_{(p_1 - g_1) \times p_2} \end{pmatrix}_{p_1 \times p_2}$. For ease of matrix construction, we let $g_1 = \frac{1}{2} p_1$ in the simulation. $\mathbf{B}_1$ is chosen to be a $(g_1 \times p_2)$ zero matrix, i.e. $\mathbf{B}_1 = 0_{g_1 \times p_2}$. $(\mathbf{B}_2)_{g_1 \times p_2}$ has two different settings.

- $\mathbf{B}_2^{(d)}$: All entries of $\mathbf{B}_2^{(d)}$ are generated from a discrete distribution with probability mass function $P(x = 0.1) = P(x = 0.2) = P(x = 0.3) = 1/3$. Then this $\mathbf{B}_2^{(d)}$ consists of nonzero small components. This corresponds to the DWA (dense but weak alternative) stated in the mean vectors' setting of Section 7.3.
- $\mathbf{B}_2^{(s)}$: The entries of $\mathbf{B}_2^{(s)}$ are all zeros except for the first 2 diagonal elements being ones, that is, $\mathbf{B}_2^{(s)} = \begin{pmatrix} \mathbf{I}_2 \\ 0 \end{pmatrix}$. This corresponds to the SSA (sparse but strong alternative) stated in the mean vectors' setting of Section 7.3.

The two different settings of $\mathbf{B}_2$ are to investigate the power performance of $\lambda_1^{\mathbf{M}_1}$ in testing (6.2) under different alternatives.

(2) *Matrix* $\mathbf{C}_1$: We consider two special cases: $\mathbf{C}_1^{(a)} = [0, \mathbf{I}_{g_1}]$ and $\mathbf{C}_1^{(b)} = [\mathbf{I}_{g_1}, 0]$.

(3) *Matrix* $\mathbf{\Gamma}_1$: $\mathbf{\Gamma}_1$ is selected to be $\mathbf{\Gamma}_1^{(a)} = 0$ or $\mathbf{\Gamma}_1^{(b)} = \mathbf{B}_2$.

Four combinations of $(\mathbf{C}_1, \mathbf{B}_2, \mathbf{\Gamma}_1)$ are used in Table 3. For each combination, as in previous sections, by repeating 10 000 times, we can find 10 000 rescaled $\lambda_1^{\mathbf{M}_1}$'s and the proportion of values that are larger than c_α are recorded in Table 3 that is, $\frac{\#\{\text{rescaled } \lambda_1^{\mathbf{M}_1} > c_\alpha\}}{10\,000}$.

The first two combinations are used for size testing. Since the two settings of $\mathbf{B}_2$ are constructed to investigate power performance under different alternatives, for size purpose, we just adopt one of them – $\mathbf{B}_2^{(d)}$. The first combination $(\mathbf{C}_1^{(b)}, \mathbf{B}_2^{(d)}, \mathbf{\Gamma}_1^{(a)})$ is to test whether the first $(g_1 \times p_2)$ block of $\mathbf{B}$ is a zero block, that is, $H_0 : \mathbf{B}_1 = 0$. The second combination $(\mathbf{C}_1^{(a)}, \mathbf{B}_2^{(d)}, \mathbf{\Gamma}_1^{(b)})$ is to test whether the second $((p_1 - g_1) \times p_2)$ block of $\mathbf{B}$ equals to a given matrix, that is, $H_0 : \mathbf{B}_2 = \mathbf{\Gamma}_1^{(b)}$. One can observe that the sizes are always close to 0.05, confirming the asymptotic distribution developed for $\lambda_1^{\mathbf{M}_1}$ in Section 6.

The last two combinations are used for power testing, that is, testing whether $\mathbf{B}_2 = 0$. Two alternatives are considered. The third combination $(\mathbf{C}_1^{(a)}, \mathbf{B}_2^{(d)}, \mathbf{\Gamma}_1^{(a)})$ is for DWA (dense but weak alternative) and the last one $(\mathbf{C}_1^{(a)}, \mathbf{B}_2^{(s)}, \mathbf{\Gamma}_1^{(a)})$ is for SSA (sparse but strong alternative). We can see that for small dimensions, SSA works better than DWA, while as the dimensions increase, a reversal takes place. This is reasonable because the magnitude of difference for DWA is much involved by values of dimensions. And for appropriate large dimensions, all power values are close to 1.

Acknowledgements

G. M. Pan (corresponding author) was partially supported by a MOE Tier 2 Grant 2014-T2-2-060 and by a MOE Tier 1 Grant RG25/14 at the Nanyang Technological University, Singapore.

Supplementary Material

Supplement to "A unified matrix model including both CCA and F matrices in multivariate analysis: The largest eigenvalue and its applications" (DOI: 10.3150/17-BEJ965SUPP; .pdf). We provide the detailed proof of Theorem 2.1 in the supplementary file.

References

[1] Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. New York: Wiley. MR0771294

[2] Bai, Z. and Silverstein, J.W. (2006). *Spectral Analysis of Large Dimensional Random Matrices*, 1st ed. New York: Springer.

[3] Bao, Z., Pan, G. and Zhou, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43** 382–421. MR3311864

[4] Bao, Z.G., Hu, J., Pan, G.M. and Zhou, W. (2015). Canonical correlation coefficients of high-dimensional normal vectors: Finite rank case. Available at http://arxiv.org/abs/1407.7194.
[5] El Karoui, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. MR2308592
[6] Erdős, L., Yau, H.-T. and Yin, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** 1435–1515. MR2871147
[7] Fujikoshi, Y., Ulyanov, V.V. and Shimizu, R. (2010). *Multivariate Statistics*: *High-Dimensional and Large-Sample Approximations*. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley. MR2640807
[8] Gao, C., Ma, Z., Ren, Z. and Zhou, H.H. (2015). Minimax estimation in sparse canonical correlation analysis. *Ann. Statist.* **43** 2168–2197. MR3396982
[9] Han, X., Pan, G. and Yang, Q. (2017). Supplement to "A unified matrix model including both CCA and F matrices in multivariate analysis: The largest eigenvalue and its applications." DOI:10.3150/17-BEJ965SUPP.
[10] Han, X., Pan, G. and Zhang, B. (2016). The Tracy–Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.* **44** 1564–1592. MR3519933
[11] Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* 321–377.
[12] Johnstone, I.M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. MR2485010
[13] Johnstone, I.M. (2009). Approximate null distribution of the largest root in multivariate analysis. *Ann. Appl. Stat.* **3** 1616–1633. MR2752150
[14] Knowles, A. and Yin, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169** 257–352. MR3704770
[15] Tracy, C.A. and Widom, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. MR1257246
[16] Tracy, C.A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083
[17] Yang, Y. and Pan, G. (2015). Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann. Statist.* **43** 467–500. MR3316187

Received January 2017 and revised June 2017