

CANONICAL CORRELATION COEFFICIENTS OF HIGH-DIMENSIONAL GAUSSIAN VECTORS: FINITE RANK CASE

BY ZHIGANG BAO^{1,*}, JIANG HU^{2,†}, GUANGMING PAN^{3,‡} AND
WANG ZHOU^{4,§}

*Hong Kong University of Science and Technology**, *Northeast Normal
University†*, *Nanyang Technological University‡* and *National University of
Singapore§*

Consider a Gaussian vector $\mathbf{z} = (\mathbf{x}', \mathbf{y}')'$, consisting of two sub-vectors \mathbf{x} and \mathbf{y} with dimensions p and q , respectively. With n independent observations of \mathbf{z} , we study the correlation between \mathbf{x} and \mathbf{y} , from the perspective of the canonical correlation analysis. We investigate the high-dimensional case: both p and q are proportional to the sample size n . Denote by Σ_{uv} the population cross-covariance matrix of random vectors \mathbf{u} and \mathbf{v} , and denote by S_{uv} the sample counterpart. The canonical correlation coefficients between \mathbf{x} and \mathbf{y} are known as the square roots of the nonzero eigenvalues of the canonical correlation matrix $\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx}$. In this paper, we focus on the case that Σ_{xy} is of finite rank k , that is, there are k nonzero canonical correlation coefficients, whose squares are denoted by $r_1 \geq \dots \geq r_k > 0$. We study the sample counterparts of r_i , $i = 1, \dots, k$, that is, the largest k eigenvalues of the sample canonical correlation matrix $S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{yx}$, denoted by $\lambda_1 \geq \dots \geq \lambda_k$. We show that there exists a threshold $r_c \in (0, 1)$, such that for each $i \in \{1, \dots, k\}$, when $r_i \leq r_c$, λ_i converges almost surely to the right edge of the limiting spectral distribution of the sample canonical correlation matrix, denoted by d_+ . When $r_i > r_c$, λ_i possesses an almost sure limit in $(d_+, 1]$, from which we can recover r_i 's in turn, thus provide an estimate of the latter in the high-dimensional scenario. We also obtain the limiting distribution of λ_i 's under appropriate normalization. Specifically, λ_i possesses Gaussian type fluctuation if $r_i > r_c$, and follows Tracy–Widom distribution if $r_i < r_c$. Some applications of our results are also discussed.

Received June 2017; revised March 2018.

¹Supported in part by Hong Kong RGC Grant ECS 26301517.

²Supported in part by NSFC (No. 11771073) and the Fundamental Research Funds for the Central Universities.

³Supported in part by the MOE Tier 2 Grant 2014-T2-2-060 and MOE Tier 1 Grant RG 25/14 at Nanyang Technological University.

⁴Supported in part by the MOE Tier 2 Grant MOE2015-T2-2-039 (R-155-000-171-112) at the National University of Singapore.

MSC2010 subject classifications. 62H20, 60B20, 60F99.

Key words and phrases. Canonical correlation analysis, random matrices, MANOVA ensemble, high-dimensional data, finite rank perturbation, largest eigenvalues.

1. Introduction. In multivariate analysis, the most general and favorable method to investigate the relationship between two random vectors \mathbf{x} and \mathbf{y} , is the Canonical Correlation Analysis (CCA), which was raised in the seminal work of Hotelling [28]. CCA aims at finding two sets of basis vectors, such that the correlations between the projections of the variables \mathbf{x} and \mathbf{y} onto these basis vectors are mutually maximized, namely, seeking vectors $\mathbf{a} = \mathbf{a}_1$ and $\mathbf{b} = \mathbf{b}_1$ to maximize the correlation coefficient

$$(1.1) \quad \rho \equiv \rho(\mathbf{a}, \mathbf{b}) := \frac{\text{Cov}(\mathbf{a}'\mathbf{x}, \mathbf{b}'\mathbf{y})}{\sqrt{\text{Var}(\mathbf{a}'\mathbf{x})} \cdot \sqrt{\text{Var}(\mathbf{b}'\mathbf{y})}}.$$

Conventionally, $\rho_1 := \rho(\mathbf{a}_1, \mathbf{b}_1)$ is called the *first canonical correlation coefficient*. Having obtained the first m canonical correlation coefficients $\rho_i, i = 1, \dots, m$ and the corresponding vector pairs $(\mathbf{a}_i, \mathbf{b}_i), i = 1, \dots, m$, one can proceed to seek vectors $(\mathbf{a}_{m+1}, \mathbf{b}_{m+1})$ maximizing ρ subject to the constraint that $(\mathbf{a}'_{m+1}\mathbf{x}, \mathbf{b}'_{m+1}\mathbf{y})$ is uncorrelated with $(\mathbf{a}'_i\mathbf{x}, \mathbf{b}'_i\mathbf{y})$ for all $i = 1, \dots, m$. Analogously, we call ρ_i the *i th canonical correlation coefficient* if it is nonzero. Denoting by Σ_{uv} the population cross-covariance matrix of arbitrary two random vectors \mathbf{u} and \mathbf{v} , it is well known that $r_i := \rho_i^2$ is the i th largest eigenvalue of the (*population*) *canonical correlation matrix* $\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$. Let $\mathbf{z}_i = (\mathbf{x}'_i, \mathbf{y}'_i)'$, $i = 1, \dots, n$ be n independent observations of the vector $\mathbf{z} := (\mathbf{x}', \mathbf{y}')' \sim N(\boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu}$ and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}.$$

We can study the canonical correlation coefficients via their sample counterparts. To be specific, we employ the notation S_{uv} to represent the sample cross-covariance matrix for arbitrary two random vectors \mathbf{u} and \mathbf{v} , where the implicit sample size of $(\mathbf{u}', \mathbf{v}')'$ is assumed to be n , henceforth. Then the square of the i th sample canonical correlation coefficient is defined as the i th largest eigenvalue of the *sample canonical correlation matrix* (CCA matrix for short) $S_{xx}^{-1}S_{xy}S_{yy}^{-1}S_{yx}$, denoted by λ_i in the sequel.

Let p and q be the dimensions of the sub-vectors \mathbf{x} and \mathbf{y} , respectively. In the classical low-dimensional setting, that is, both p and q are fixed but n is large, one can safely use λ_i to estimate r_i , considering the convergence of the sample cross-covariance matrices toward their population counterparts. However, nowadays, due to the increasing demand in the analysis of high-dimensional data springing up in various fields such as genomics, signal processing, microarray, finance and proteomics, putting forward a theory on high-dimensional CCA is much needed. So far, there are only a handful of works devoted to this topic. Fujikoshi in [21] derived the asymptotic distributions of the canonical correlation coefficients when q is fixed while p is proportional to n . Oda et al. in [40] considered the problem of testing for redundancy in high-dimensional canonical correlation analysis.

Recently, with certain sparsity assumption, the theoretical results and potential applications of high-dimensional sparse CCA have been discussed in [23, 24]. In the null case, that is, \mathbf{x} and \mathbf{y} are independent, the Tracy–Widom law for the largest canonical correlation coefficients has been studied in [26, 27, 33], when p, q, n are proportional. Recently, in [35] Johnstone and Onatski derived the asymptotics of the likelihood ratio processes of CCA corresponding to the null hypothesis of no spikes and the alternative of a single spike.

In this paper, we will work with the following high-dimensional setting.

ASSUMPTION 1.1. We assume $p := p(n), q := q(n)$, and when $n \rightarrow \infty$

$$p/n = c_1 \rightarrow y_1 \in (0, 1), \quad q/n = c_2 \rightarrow y_2 \in (0, 1), \quad \text{s.t. } y_1 + y_2 \in (0, 1).$$

Without loss of generality, we always work with the additional assumption $p \geq q$ thus $c_1 \geq c_2$.

Observe that here y_1 and y_2 are asymptotic parameters, while c_1 and c_2 are nonasymptotic parameters. Hence, in general, a c_1, c_2 -dependent random variable $X(c_1, c_2)$ cannot serve as a limiting target of a random sequence $X_n(c_1, c_2)$, $n \geq 1$. Nevertheless, to ease the presentation, from time to time, we still write $X_n(c_1, c_2) \rightarrow X(c_1, c_2)$, if $X_n(c_1, c_2) - X(c_1, c_2) \rightarrow 0$, where the convergence could be in distribution, in probability or a.s., etc.

Let $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ be the sample means of n samples $\{\mathbf{x}_i\}_{i=1}^n$ and $\{\mathbf{y}_i\}_{i=1}^n$ respectively, and use the notation $\hat{\mathbf{x}}_i := \mathbf{x}_i - \bar{\mathbf{x}}$ and $\hat{\mathbf{y}}_i := \mathbf{y}_i - \bar{\mathbf{y}}$ for $i = 1, \dots, n$. We can then write $S_{ab} = \frac{1}{n-1} \sum_{i=1}^n \hat{\mathbf{a}}_i \hat{\mathbf{b}}_i', \mathbf{a}, \mathbf{b} = \mathbf{x}$ or \mathbf{y} . It is well known that there exist $n - 1$ i.i.d. Gaussian vectors $\tilde{\mathbf{z}}_i = (\tilde{\mathbf{x}}_i', \tilde{\mathbf{y}}_i')' \sim N(\mathbf{0}, \Sigma)$, such that $S_{ab} = \frac{1}{n-1} \sum_{i=1}^{n-1} \tilde{\mathbf{a}}_i \tilde{\mathbf{b}}_i', \mathbf{a}, \mathbf{b} = \mathbf{x}$ or \mathbf{y} . For simplicity, we recycle the notation \mathbf{x}_i and \mathbf{y}_i to replace $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{y}}_i$, and work with n instead of $n - 1$, noticing that such a replacement on sample size is harmless to Assumption 1.1. Hence, we can and do assume that \mathbf{z} is centered in the sequel and denote $S_{ab} = \frac{1}{n} \sum_{i=1}^n \mathbf{a}_i \mathbf{b}_i', \mathbf{a}, \mathbf{b} = \mathbf{x}$ or \mathbf{y} . For brevity, we introduce the notation

$$(1.2) \quad C_{xy} := S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{yx},$$

and C_{yx} can be analogously defined via switching the roles of x and y in (1.2). Notice that C_{xy} and C_{yx} possess the same nonzero eigenvalues.

By our assumption $p \geq q$, there are at most q nonzero canonical correlations, either population ones or sample ones. An elementary fact is that $\lambda_i, r_i \in [0, 1]$ for all $i = 1, \dots, q$. Note that $\lambda_i, i = 1, \dots, q$ are also eigenvalues of the $q \times q$ matrix C_{yx} , whose empirical spectral distribution (ESD) will be denoted by

$$(1.3) \quad F_n(x) := \frac{1}{q} \sum_{i=1}^q \mathbf{1}(\lambda_i \leq x).$$

Here and in the following, $\mathbf{1}(\cdot)$ is the indicator function. We denote by

$$\mathcal{X} := \mathcal{X}_n = (\mathbf{x}_1, \dots, \mathbf{x}_n), \quad \mathcal{Y} := \mathcal{Y}_n = (\mathbf{y}_1, \dots, \mathbf{y}_n).$$

Let $\text{vec}(\mathcal{X}) = (\mathbf{x}'_1, \dots, \mathbf{x}'_n)'$ be the vectorization of the matrix \mathcal{X} , and define $\text{vec}(\mathcal{Y})$ be the analogously. We see that $\text{vec}(\mathcal{X}) \sim N(\mathbf{0}, I_n \otimes \Sigma_{xx})$ and $\text{vec}(\mathcal{Y}) \sim N(\mathbf{0}, I_n \otimes \Sigma_{yy})$. Our aim, in this work, is to study the asymptotic behavior of a few largest sample canonical correlation coefficients $\sqrt{\lambda_i}$'s, and try to get the information about the population ones $\rho_i = \sqrt{r_i}$ from the sample ones. We will focus on the case of finite rank, that is, there is some fixed nonnegative integer k , such that $r_1 \geq \dots \geq r_k \geq r_{k+1} = \dots = r_q = 0$. Specifically, we make the following assumption throughout the work.

ASSUMPTION 1.2 (On the rank of the population matrix). We assume that $\text{rank}(\Sigma_{xy}) \leq k$ for some fixed positive integer k . Furthermore, setting $r_0 = 1$, we denote by k_0 the nonnegative integer satisfying

$$(1.4) \quad 1 = r_0 \geq \dots \geq r_{k_0} > r_c \geq r_{k_0+1} \geq \dots \geq r_k > r_{k+1} = 0,$$

where

$$(1.5) \quad r_c \equiv r_c(c_1, c_2) := \sqrt{\frac{c_1 c_2}{(1 - c_1)(1 - c_2)}}.$$

In Section 1.2, we will state our main results. Before that, we introduce in Section 1.1 some known results in the *null case*, that is, $k = 0$, which will be the starting point of our discussion.

1.1. The null case: MANOVA ensemble. At first, we introduce some known results on the limiting behavior of $\{\lambda_i\}_{i=1}^q$ in the null case, that is, \mathbf{x} and \mathbf{y} are independent, or else, $r_i = 0$ for all $i = 1, \dots, q$. It is elementary to see that the canonical correlation coefficients are invariant under the block diagonal transformation $(\mathbf{x}_i, \mathbf{y}_i) \rightarrow (\mathbf{A}\mathbf{x}_i, \mathbf{B}\mathbf{y}_i)$, for any $p \times p$ matrix \mathbf{A} and $q \times q$ matrix \mathbf{B} , as long as both of them are nonsingular. Hence, without loss of generality, in this section, we tentatively assume that $\Sigma_{xx} = I_p$ and $\Sigma_{yy} = I_q$. Under our high-dimensional setting, that is, Assumption 1.1, it is known that λ_i 's do not converge to 0 even in the null case; instead, they typically spread out over an interval contained in $[0, 1]$. Specifically, we have the following theorem on $F_n(x)$ [cf. (1.3)], which is due to Wachter [45].

THEOREM 1.3. *When \mathbf{x} and \mathbf{y} are independent Gaussian and Assumption 1.1 holds, almost surely, F_n converges weakly to a deterministic probability distribution $F(x)$ with density*

$$(1.6) \quad f(x) = \frac{1}{2\pi c_2} \frac{\sqrt{(d_+ - x)(x - d_-)}}{x(1 - x)} \mathbf{1}(d_- \leq x \leq d_+),$$

where $d_{\pm} = (\sqrt{c_1(1 - c_2)} \pm \sqrt{c_2(1 - c_1)})^2$.

REMARK 1.4. In the null case, the convergence of the ESD of the CCA matrix actually holds under a more general distribution assumption; see [48].

Conventionally, we call $F(x)$ in Theorem 1.3 the limiting spectral distribution (LSD) of C_{yx} . One might note that $F(x)$ is also the LSD of the so-called MANOVA ensemble with appropriately chosen parameters, which is widely studied in the Random Matrix Theory (RMT). Actually, when \mathbf{x} and \mathbf{y} are Gaussian and independent, the CCA matrix C_{yx} is exactly a MANOVA matrix. To see this, we note that $P_x := \mathcal{X}'(\mathcal{X}\mathcal{X}')^{-1}\mathcal{X}$ is a projection matrix independent of Y . Hence, we can write $C_{yx} = (\mathcal{Y}(I - P_x)\mathcal{Y}' + \mathcal{Y}P_x\mathcal{Y}')^{-1}\mathcal{Y}P_x\mathcal{Y}'$. Using Cochran's theorem, we see that $\mathcal{Y}(I - P_x)\mathcal{Y}'$ and $\mathcal{Y}P_x\mathcal{Y}'$ are independent Wishart, that is, $\mathcal{Y}(I - P_x)\mathcal{Y}' \sim \mathcal{W}_q(I_q, n - p)$, and $\mathcal{Y}P_x\mathcal{Y}' \sim \mathcal{W}_q(I_q, p)$. Hereafter, we use the notation $\mathcal{W}_m(\cdot, \cdot)$ to denote the Wishart distribution $\text{Wishart}_m(\cdot, \cdot)$ for short. Consequently, $\lambda_i, i = 1, \dots, q$ are known to possess the following joint density function:

$$p_n(\lambda_1, \dots, \lambda_q) = C_n \prod_{i < j}^q |\lambda_i - \lambda_j| \prod_{i=1}^q (1 - \lambda_i)^{\frac{n-p-q-1}{2}} \lambda_i^{\frac{p-q-1}{2}} \mathbf{1}(\lambda_i \in [0, 1]),$$

for some constant C_n (see Muirhead [39], page 112). Or else, one can refer to [33], for more related discussions. In the context of RMT, the point process possessing the above joint density is also called Jacobi ensemble.

Throughout the paper, we will say that an n -dependent event $A \equiv A(n)$ holds with *overwhelming probability*, if for any given positive number ℓ , there exists a constant C_ℓ such that $\mathbb{P}(A) \geq 1 - C_\ell n^{-\ell}$. Especially, for any fixed integer $K \geq 0$, we have $\bigcap_{i=1}^K A_i$ holds with overwhelming probability if A_i holds with overwhelming probability individually with the common C_ℓ 's. The next known result concerns the convergence of the largest eigenvalues.

THEOREM 1.5. *When \mathbf{x} and \mathbf{y} are independent and Assumption 1.1 holds, we have*

$$(1.7) \quad \lambda_i - d_+ \xrightarrow{a.s.} 0,$$

for any fixed positive integer i . Moreover, for any small constant $\varepsilon > 0$,

$$(1.8) \quad \lambda_1 \leq d_+ + \varepsilon$$

holds with overwhelming probability.

REMARK 1.6. The estimate (1.8) can be derived from the small deviation estimate of the largest eigenvalue of the Jacobi ensemble in [37], and (1.7) is a direct consequence of (1.8) and Theorem 1.3.

1.2. *Finite rank case.* We now turn to the case of finite rank. To wit, Assumption 1.2 holds. It will be clear that the CCA matrix in such a finite rank case can be viewed as a finite rank perturbation of that in the null case. Consequently, the global behavior (LSD), turns out to coincide with the null case. However, finite rank perturbation may significantly alter the behavior of the extreme eigenvalues, when the perturbation is strong enough. Similar problems have been studied widely for various random matrix models, not trying to be comprehensive, we refer to the spiked sample covariance matrices [4–6, 20, 32, 42], the deformed Wigner matrices [14, 15, 19, 38, 43], the deformed unitarily invariant matrices [10, 36] and some other deformed models [11, 12, 46]. In this work, for our CCA matrix C_{xy} , we study the limits and the fluctuations of its largest eigenvalues, that is, squares of the largest sample canonical correlation coefficients, under Assumption 1.2.

Our main results are the following three theorems. Let

$$(1.9) \quad \gamma_i := r_i(1 - c_1 + c_1 r_i^{-1})(1 - c_2 + c_2 r_i^{-1}).$$

Recall r_c and d_+ defined in (1.5) and Theorem 1.3, respectively. It is easy to check that $\gamma_i \geq d_+$ if $r_i \geq r_c$. For brevity, we further introduce the notation

$$(1.10) \quad \varpi := c_1 c_2, \quad \vartheta := (1 - c_1)(1 - c_2).$$

THEOREM 1.7 (Limits). *Under Assumptions 1.1 and 1.2, the squares of the canonical correlation coefficients exhibit the following convergence:*

- (i) (*Outliers*) For $1 \leq i \leq k_0$, as $n \rightarrow \infty$, we have $\lambda_i - \gamma_i \xrightarrow{a.s.} 0$.
- (ii) (*Sticking eigenvalues*) For each fixed $i \geq k_0 + 1$, as $n \rightarrow \infty$, we have $\lambda_i - d_+ \xrightarrow{a.s.} 0$.

To state the results on fluctuations, we need the following definition.

DEFINITION 1.8. For two (possibly) n -dependent numbers $a(n), b(n) \in \mathbb{C}$, we say $a(n)$ is well separated from $b(n)$, if there exists a small positive constant ε such that $|a(n) - b(n)| \geq \varepsilon$ for sufficiently large n .

THEOREM 1.9 (Fluctuations of the outliers). *Suppose that Assumptions 1.1 and 1.2 hold. Let l_0 be the cardinality of $\Gamma := \{r_1, \dots, r_{k_0}\}$ (not counting multiplicity), and denote by $r_1 = r_{(1)} > \dots > r_{(l_0)} = r_{k_0}$ the l_0 different values in Γ . Set $n_0 = 0$ and denote by n_l the multiplicity of $r_{(l)}$ for $1 \leq l \leq l_0$. Let $J_l = [\sum_{i=0}^{l-1} n_i + 1, \sum_{i=0}^l n_i] \cap \mathbb{Z}$ for $1 \leq l \leq l_0$. If $r_{(l)}$ is well separated from $r_c, 1, r_{(l-1)}$ and $r_{(l+1)}$, the random vector $\{\sqrt{n}(\lambda_j - \gamma_j)/\xi(r_j), j \in J_l\}$ converges weakly to the distribution of the ordered eigenvalues of an n_l -dimensional symmetric Gaussian random matrix $G = (g_{ij})$ with independent (up to symmetry) entries $g_{ij} \sim N(0, 1 + \delta_{ij})$. Here,*

$$(1.11) \quad \xi^2(r_j) := \frac{(1 - r_j)^2(2\vartheta r_j + c_1 + c_2 - 2\varpi)(\vartheta r_j^2 - \varpi)}{r_j^2}.$$

REMARK 1.10. The assumption that r_ℓ is away from r_c , 1, $r_{(\ell-1)}$ and $r_{(\ell+1)}$ by a constant distance $\varepsilon > 0$ (well separated) may not be optimal. It is possible to reduce ε to some n -dependent distance $n^{-\alpha}$ for small α .

THEOREM 1.11 (Fluctuations of the sticking eigenvalues). *Suppose that Assumptions 1.1 and 1.2 hold. In addition, we assume that r_c is well separated from r_{k_0} and r_{k_0+1} . There exists a CCA matrix in the null case with the same parameters p, q, n , whose nonzero eigenvalues are denoted by $\mathring{\lambda}_1 > \mathring{\lambda}_2 > \cdots > \mathring{\lambda}_q$, such that for any fixed positive integer m and any small constant $\varepsilon > 0$, we have*

$$(1.12) \quad \max_{1 \leq i \leq m} |\lambda_{k_0+i} - \mathring{\lambda}_i| \leq n^{-1+\varepsilon}$$

in probability. This implies

$$(1.13) \quad n^{\frac{2}{3}}(\lambda_{k_0+1} - d_+)/\xi_{tw} \Longrightarrow F_1,$$

where F_1 is the Type 1 TW law and $\xi_{tw}^3 = d_+^2(1 - d_+)^2/\sqrt{\varpi\vartheta}$.

REMARK 1.12. From (1.12), it is easy to conclude (1.13), by using the Tracy–Widom limit of $\mathring{\lambda}_1$ derived in [26, 27, 33]. Observe that in these references, the Tracy–Widom law is stated for the logit transform of $\mathring{\lambda}_1$, that is, $\log(\mathring{\lambda}_1/(1 - \mathring{\lambda}_1))$. Using a Taylor expansion, it is elementary to check that that $n^{\frac{2}{3}}(\mathring{\lambda}_1 - d_+)/\xi_{tw} \Longrightarrow F_1$ from the Tracy–Widom law for the logit transform of $\mathring{\lambda}_1$ (see Theorem 1 and Section 2.1.1 in [33] for more details).

REMARK 1.13. Similar to Remark 1.10, the assumption that r_c is away from r_{k_0} and r_{k_0+1} by a constant distance can be weakened. But we do not pursue this direction here.

To illustrate the result in Theorems 1.7, 1.9 and 1.11, we did some numerical simulations, whose results can be found in the Supplementary Material [9]. The different limiting behavior of λ_i in (i) and (ii) of Theorem 1.7 can be observed in Figure S1.1. The fluctuation for λ_i in Theorem 1.9 and Theorem 1.11 can be seen from Figures S1.2 and S1.1.

We remark here that we focus on the Gaussian population in this paper, due to the technical reason. A very natural and important further problem is the extension of the above results to the general distribution case. We refer to Section 8 for more discussion.

1.3. *Organization and notation.* Our paper is organized as follows. In Section 2, we display some applications of our results. We introduce in Section 3 some necessary preliminaries. In Section 4, we will reformulate the sample CCA matrix in the finite rank case as a perturbation of that in the null case, thereby obtaining

a determinant equation for the largest eigenvalues. Then we derive the limits of the largest eigenvalues in Section 5, that is, prove Theorem 1.7. In Section 6, we derive the fluctuations of the outliers (Theorem 1.9), and in Section 7, we study the fluctuations of the sticking eigenvalues (Theorem 1.11).

Throughout the paper, the notation C represents some generic constants which may vary from line to line. The notation $\mathbf{0}_{k \times \ell}$ is used to denote the k by ℓ null matrix, which will be abbreviated to $\mathbf{0}_k$ if $k = \ell$, and from time to time we also use the abbreviation $\mathbf{0}$ if the dimension is clear from the context. For any matrix \mathbf{A} , its (i, j) th entry will be written as A_{ij} . When \mathbf{A} is square, we denote by $\text{Spec}(\mathbf{A})$ its spectrum. For a function $f : \mathbb{C} \rightarrow \mathbb{C}$ and a Hermitian matrix \mathbf{A} with spectral decomposition $\mathbf{U}_\mathbf{A} \Lambda_\mathbf{A} \mathbf{U}_\mathbf{A}^*$, we define $f(\mathbf{A})$ as usual, in the sense of functional calculus, to wit, $f(\mathbf{A}) = \mathbf{U}_\mathbf{A} f(\Lambda_\mathbf{A}) \mathbf{U}_\mathbf{A}^*$, where $f(\Lambda_\mathbf{A})$ is the diagonal matrix obtained via mapping the eigenvalues of \mathbf{A} to their images under f . We will use the notation $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_{\text{HS}}$ to represent the operator norm and Hilbert–Schmidt norm of a matrix \mathbf{A} , respectively, while $\|\mathbf{b}\|$ stands for the Euclidean norm of a vector \mathbf{b} . Throughout this paper, we use $o_p(1)$ to denote a scalar negligible (in probability) or a fixed-dimensional random matrix with negligible (in probability) entries. And the notation $o_{\text{a.s.}}(1)$, $O_p(1)$ and $O_{\text{a.s.}}(1)$ are used in a similar way.

2. Applications. In this section, we discuss three applications of our results in hypothesis testing, estimation of the number and the values of the population canonical correlation coefficients (CCC for short). At the end, we present an experiment on a real limestone grassland community data.

2.1. *Application 1: Power of testing for the independence of two high-dimensional normal vectors.* For two normal random vectors \mathbf{x} and \mathbf{y} with dimensions p and q , respectively, we consider the test

$$H_0 : \Sigma_{\mathbf{xy}} = \mathbf{0}_{p \times q} \quad \text{v.s. } H_1 : \text{not } H_0.$$

Currently, for high-dimensional cases, there are three kinds of widely discussed test procedures: (i) corrected likelihood ratio tests (see [30, 31]); (ii) trace tests (see [7, 29, 30, 47, 49]); (iii) largest eigenvalue tests (see [27, 33, 34]). It has been shown by numerical results in [7, 27] that if $\Sigma_{\mathbf{xy}}$ is sparse, the corrected likelihood ratio tests and trace tests fail and the largest eigenvalue tests works well. In the following, we propose a statistic based on the CCC, and show that it is powerful against finite rank case. It is well known that testing for the independence of two high-dimensional normal vectors equals to testing their first CCC being zero or not, that is,

$$H_0 : r_1 = 0 \quad \text{v.s. } H_1 : r_1 > 0.$$

Therefore, a natural test statistic is the largest eigenvalue λ_1 of the sample CCA matrix $C_{\mathbf{xy}}$. Then according to Section 2.1.1 in [33] and Theorem 2.1 in [27], under the null hypothesis and Assumption 1.1 we have

$$(2.1) \quad n^{2/3}(\lambda_1 - d_+) / \xi_{tw} \implies F_1.$$

Therefore, we reject H_0 if

$$(2.2) \quad \lambda_1 > n^{-2/3}q_\alpha\xi_{tw} + d_+,$$

where q_α is the $1 - \alpha$ quantile of TW law F_1 . If the sample is under the Assumption 1.2 with $k_0 > 0$ and r_1 is well separated from r_c and 1, then this test will be able to detect the alternative hypothesis with a power tending to one as the dimension tends to infinity.

THEOREM 2.1 (Power function). *Suppose that the assumptions in Theorems 1.9 and 1.11 hold with r_1 being well separated from r_c and 1. Then as $n \rightarrow \infty$, the power function of the test procedure (2.2)*

$$\text{Power} = \mathbb{P}\left(\frac{\sqrt{n}(\lambda_1 - \gamma_1)}{\xi(r_1)} > \frac{n^{-1/6}q_\alpha\xi_{tw}}{\xi(r_1)} + \frac{\sqrt{n}(d_+ - \gamma_1)}{\xi(r_1)}\right) \rightarrow 1.$$

The proof of this theorem is stated in the Supplementary Material [9].

REMARK 2.2. We conjecture that under the same assumptions of Theorem 2.1, the corrected likelihood ratio tests and trace tests mentioned above should be inconsistent.

2.2. Application 2: Estimating the number of the population CCC. As fundamental problems in CCA, the estimation of the number and the values of the population CCC are widely investigated. In this subsection, we first apply our results to determine the number of the outliers of high-dimensional population CCC (counting multiplicity). Actually, due to the threshold in our results, we cannot detect the population CCC which are smaller than $\sqrt{r_c}$. To the best of our knowledge, there is no effective method in general to successfully detect the population spikes below the threshold, even for the simpler spike PCA problem (see [2] for instance).

Our estimator of the number of the outliers of population CCC k_0 is the number of eigenvalues of the sample CCA matrix which are larger than d_+ :

$$(2.3) \quad \hat{k}_0 := \max\{i : \lambda_i \geq d_+ + \epsilon_n\},$$

where ϵ_n is a sequence of positive number only depending on n and satisfying $\epsilon_n\sqrt{n} \rightarrow 0$ and $\epsilon_n n^{2/3} \rightarrow \infty$. Then the estimator \hat{k}_0 is weakly consistent according to the following theorem, whose proof is stated in the Supplementary Material [9].

THEOREM 2.3 (Weak consistency of \hat{k}_0). *Suppose that the assumptions in Theorems 1.9 and 1.11 hold. As $n \rightarrow \infty$, the estimator \hat{k}_0 in (2.3) is weakly consistent for the number of the outliers of population CCC k_0 , that is,*

$$(2.4) \quad \mathbb{P}(\hat{k}_0 = k_0) \rightarrow 1.$$

REMARK 2.4. Although in theory, any sequence ϵ_n which satisfies the conditions $\epsilon_n\sqrt{n} \rightarrow 0$ and $\epsilon_n n^{2/3} \rightarrow \infty$ is applicable, we have to choose one in practice. Thus, it is worth to notice that the smaller ϵ_n one chooses, the easier one overestimates k_0 and vice versa. In the simulation, based on the idea in the paper [41, 46], we choose ϵ_n to be $\log \log(n)/n^{2/3}$, which makes our estimator conservative.

In the following, we report a short simulation result to illustrate the performance of our estimator. For comparison, we also present the performance of three other estimators which are introduced in [22] and [13]. To be specific, we first use the classical PCA technique to project the original data \mathcal{X}_n and \mathcal{Y}_n onto the space spanned by their first few eigenvectors with the cumulative contribution rate $\geq 80\%$. The projections are denoted by $\widehat{\mathcal{X}}_n$ and $\widehat{\mathcal{Y}}_n$, respectively. Then we estimate k_0 by the following model selection criteria: AIC, BIC and C_p , respectively, which are

$$\hat{k}_A = \arg \min_{0 \leq j \leq \min\{\hat{p}, \hat{q}\}} \text{AIC}_j, \quad \hat{k}_B = \arg \min_{0 \leq j \leq \min\{\hat{p}, \hat{q}\}} \text{BIC}_j, \quad \hat{k}_C = \arg \min_{0 \leq j \leq \min\{\hat{p}, \hat{q}\}} CP_j.$$

Here, \hat{p} and \hat{q} are the dimensions of the projections $\widehat{\mathcal{X}}_n$ and $\widehat{\mathcal{Y}}_n$, respectively. Moreover, $\text{AIC}_0 = \text{BIC}_0 = CP_0 = 0$ and if $j \geq 1$,

$$\begin{aligned} \text{AIC}_j &= -n \log \left[\prod_{i=j+1}^{\min\{\hat{p}, \hat{q}\}} (1 - \hat{\lambda}_i) \right] - 2(\hat{p} - j)(\hat{q} - j), \\ \text{BIC}_j &= -n \log \left[\prod_{i=j+1}^{\min\{\hat{p}, \hat{q}\}} (1 - \hat{\lambda}_i) \right] - \log(n)(\hat{p} - j)(\hat{q} - j), \\ CP_j &= n \sum_{i=j+1}^{\min\{\hat{p}, \hat{q}\}} \frac{\hat{\lambda}_i}{1 - \hat{\lambda}_i} - 2(\hat{p} - j)(\hat{q} - j), \end{aligned}$$

where $\hat{\lambda}_i$ is the i th largest eigenvalue of CCA matrix of $\widehat{\mathcal{X}}_n$ and $\widehat{\mathcal{Y}}_n$.

The simulation results are stated in the Supplementary Material (see Table S1.1 and Table S1.2 in [9]). According to these results, we can see that the performance of our estimator is much better than AIC, BIC and C_p estimators when the dimensions are large. If the dimensions are small and the sample size is large, then apart from our estimator, the AIC and C_p estimators also perform well, but BIC estimator performs bad.

2.3. *Application 3: Estimating the values of the population CCC.* As a measure to detect the correlation level between two random vectors, the values of population CCC are also important. Traditionally, one uses the sample CCC to estimate the population ones directly, since under the traditional assumption that p, q are fixed and $n \rightarrow \infty$, the sample CCC tend to the population ones almost surely (see

Chapter 12 in [1]). However, it has been noticed that sample CCC suffer from inflation, which occurs when the number of observations is not sufficient (see [17, 44, 50]). But the good news is that now we can easily explain the reason of the inflation properties of the sample CCC by our Theorem 1.7. In this subsection, we will give an approach to estimate the population CCC which are bigger than $\sqrt{r_c}$.

Solving the equation (1.9) via replacing γ_i by λ_i , we have two solutions

$$(2.5) \quad \hat{r}_i := \frac{2c_1c_2 - c_1 - c_2 + \lambda_i \pm \sqrt{(\lambda_i - d_-)(\lambda_i - d_+)}}{2(c_1c_2 - c_1 - c_2 + 1)}.$$

Notice that the product of two solutions equals $c_1c_2/[(1 - c_1)(1 - c_2)] = r_c^2$. Hence, by the fact that $\lambda_i > d_+$ and $\hat{r}_i \rightarrow r_i > r_c$ when $n \rightarrow \infty$, we should choose a plus sign in (2.5). Hence, the estimator of r_i is chosen to be

$$(2.6) \quad \hat{r}_i := \phi(\lambda_i) = \frac{2c_1c_2 - c_1 - c_2 + \lambda_i + \sqrt{(\lambda_i - d_-)(\lambda_i - d_+)}}{2(c_1c_2 - c_1 - c_2 + 1)}.$$

Then according to Theorem 1.7, we have the following theorem.

THEOREM 2.5. *Under the same conditions of Theorem 1.7, for any $1 \leq i \leq k_0$, we have almost surely $\hat{r}_i \rightarrow r_i$.*

Next, we show a short simulation result for the estimators \hat{r}_i 's. In the simulation, we first assume that the number k_0 of the supercritical population CCCs is known. Hence, if $\lambda_i > d_+$, $i = 1, \dots, k_0$, we can obtain \hat{r}_i by (2.6) directly. And if there is some $\lambda_i \leq d_+$, $i = 1, \dots, k_0$, we then set $\hat{r}_i = r_c$. This result is presented in Figure S1.3 in Supplementary Material [9] in order to show the convergence of \hat{r}_i as $\{p, q, n\} = \{0.1n, 0.05n, n\}$ increases. From this figure, we can see that our estimators converge to the true values very fast especially when the true values are far away from r_c . If the number k_0 of the supercritical population CCC is unknown, we first estimate k_0 by (2.3) with the tuning parameter $\epsilon_n = \log \log(n)/n^{2/3}$, and then for $i \leq \hat{k}_0$, we obtain the estimators \hat{r}_i by (2.6). The results are displayed in Table S1.3 and Table S1.4 in Supplementary Material [9], which are the sample means and standard deviations (s.d.) of the estimators \hat{r}_i . According to the results, we find that our estimator is excellent, especially when the population CCC are not equal to each other. If the multiplicity of some CCC is bigger than one, as the results shown in Table S1.4, there should be some \hat{r}_i 's close to each other. In this case, although $r_i = r_{i+1}$ for certain multiple population CCC, the estimator \hat{r}_i may differ from \hat{r}_{i+1} by a certain amount since our λ_i 's are ordered. Suppose now we can determine $r_i = r_{i+1}$ from the information in the CCA matrix, we may then use the average of \hat{r}_i and \hat{r}_{i+1} to get more precise estimate of both r_i and r_{i+1} .

Hence, determining if there are multiple r_i 's would be important for the estimate of r_i as well, although sometimes directly using \hat{r}_i is sufficient. In the following, we propose a statistic to test the hypothesis:

$$(2.7) \quad H_0 : r_{j_0-1} > r_{j_0} = r_{j_0+1} = \dots = r_{j_0+j_1-1} > r_{j_0+j_1} \quad \text{v.s.} \quad H_1 : \text{not } H_0,$$

where $j_0 \geq 1$, $j_0 + j_1 - 1 \leq k_0$ and $r_0 = 1$. If this test is not rejected, we then estimate $r_{j_0} = r_{j_0+1} = \dots = r_{j_0+j_1-1}$ by $j_1^{-1} \sum_{i=j_0}^{j_0+j_1-1} \hat{r}_i$. Under H_0 in (2.7), using Theorem 1.9, we know that the j_1 -dimensional random vector $\{\sqrt{n}(\lambda_j - \gamma_j)/\xi(r_{j_0}), j \in \{j_0, \dots, j_1 - 1\}\}$ converges weakly to the distribution of the ordered eigenvalues of j_1 -dimensional symmetric Gaussian random matrix G . Thus we can naturally use the testing statistics

$$T_n := \sqrt{n}(\lambda_{j_0} - \lambda_{j_0+j_1-1})/\xi(\hat{r}_{j_0}),$$

where $\hat{r}_{j_0} = j_1^{-1} \sum_{i=j_0}^{j_0+j_1-1} \hat{r}_i$. We have the following theorem.

THEOREM 2.6. *Suppose that the assumptions in Theorem 1.9 hold. Under the null hypothesis of (2.7), we have for any $x \in \mathbb{R}$,*

$$|\mathbb{P}(T_n \leq x) - \mathbb{P}(\lambda_1^G - \lambda_{j_1}^G \leq x)| \rightarrow 0,$$

where λ_1^G and $\lambda_{j_1}^G$ are the largest and smallest eigenvalues of j_1 -dimensional symmetric Gaussian random matrix G (defined in Theorem 1.9), respectively. In addition, if the alternative hypothesis H_1 is: there exists some $j_2 \in [0, j_1 - 2] \cap \mathbb{Z}$ such that $r_{j_0+j_2}$ and $r_{j_0+j_2+1}$ are well separated (cf. Definition 1.8), we have the power function

$$\text{Power} = \mathbb{P}(T_n > q_\alpha(j_1)) \rightarrow 1,$$

where $q_\alpha(j_1)$ is the $1 - \alpha$ quantile of the distribution $\lambda_1^G - \lambda_{j_1}^G$.

This theorem can be easily obtained by Skorohod strong representation theorem and Theorem 1.9. We omit the proof. To ease the application, under the null hypothesis the 95% empirical quantile of the distribution $\lambda_1^G - \lambda_{j_1}^G$ is shown in Table 1 with some different j_1 by 10^8 bootstrap replicates.

Finally, we give a procedure of estimating the population CCC. The algorithm is shown in Table 2.

2.4. The limestone grassland community data. To illustrate the application of canonical correlation, we apply our result to a limestone grassland community data which can be obtained from Table A-2 of [25]. This data records eight species (\mathbf{x} , $p = 8$) and six soil variables (\mathbf{y} , $q = 6$) from forty-five ($n = 44$) $10 \text{ m} \times 10 \text{ m}$

TABLE 1
The 5% empirical quantile of the distribution $\lambda_1^G - \lambda_{j_1}^G$ with 4 digits

j_1	2	3	4	5	6	7	8	9	10
$q_{5\%}(j_1)$	4.895	6.497	7.720	8.755	9.670	10.499	11.263	11.975	12.646

TABLE 2
Algorithm for estimating the population CCC

Step 1: $\{\lambda_1 \geq \dots \geq \lambda_q\} = \text{Eigenvalue}(C_{yx})$;
 Step 2: Test whether $r_1 = 0$.
 If not reject, then continue;
 Step 3: $\hat{k}_0 := \max\{i : \lambda_i \geq d_+ + \epsilon_n\}$;
 Step 4: $\hat{r}_i = \phi(\lambda_i), i = 1, \dots, \hat{k}_0$;
 Step 5: If there exists some $\{j_0, \dots, j_0 + j_1 - 1\}$,
 such that $\frac{\sqrt{n}(\lambda_{j_0} - \lambda_{j_0+j_1-1})}{\xi(\hat{r}_{j_0})} \leq q5\%(j_1)$,
 then reset $\hat{r}_{j_0} = \dots = \hat{r}_{j_0+j_1-1} = j_1^{-1} \sum_{i=j_0}^{j_0+j_1-1} \hat{r}_i$.
 Step 6: $\hat{\rho}_i = \sqrt{\hat{r}_i}$.

stands in a limestone grassland community in Anglesey, North Wales. This experiment concerns relationships between the abundances of several plant species and associated soil characteristics. More explanations can be found in Chapter 7 of [25]. Notice that although the dimensions p, q and the sample size n are not so “large,” from the simulations (some results are not reported in this manuscript), we can already observe the high-dimensional effect when the dimensions are bigger than 5 and sample sizes are bigger than 20, that is, the classical estimates are already not reliable. The eigenvalues of C_{yx} are as follows: $\lambda_1 = 0.829, \lambda_2 = 0.520, \lambda_3 = 0.359, \lambda_4 = 0.107, \lambda_5 = 0.094, \lambda_6 = 0.038$. In addition, we have $d_+ = 0.533$ and $\xi_{tw}^3 = 0.468$. Then we use (2.1) to test the hypothesis $r_1 = 0$, and obtain the p-value to be 3.71×10^{-5} . Thus we have strong evidence to reject the null hypothesis, which suggests the existence of a linear relationship between \mathbf{x} and \mathbf{y} . Since there is only one eigenvalue of C_{yx} which is bigger than d_+ , we can only determine the existence of the first CCC $\hat{\rho}_1 = \sqrt{\hat{r}_1} = 0.864$, which is estimated by (2.6). For the rest CCC, we have no enough evidence to ensure their existence.

3. Preliminaries. In this section, we introduce some basic notions and known technical results escorting our proofs and calculations in the subsequent sections.

For any given probability distribution $\sigma(\lambda)$, its Stieltjes transform is

$$s_\sigma(z) := \int (\lambda - z)^{-1} d\sigma(\lambda), \quad z \in \mathbb{C}^+ := \{\omega \in \mathbb{C} : \Im\omega > 0\}.$$

From the definition, we can immediately get the fact that $\Im s_\sigma(z) > 0$ for $z \in \mathbb{C}^+$. Actually, the definition of $s_\sigma(z)$ can be extended to the domain $\mathbb{C} \setminus \text{supp}(\sigma)$, by setting $s_\sigma(\bar{z}) = \overline{s_\sigma(z)}$, where $\text{supp}(\sigma)$ represents the support of $\sigma(\lambda)$. Then $s_\sigma(z)$ is holomorphic on $\mathbb{C} \setminus \text{supp}(\sigma)$.

In the sequel, we will also need the Stieltjes transform of the MANOVA matrices $A(A + B)^{-1}$ and $A(A + B)^{-1}$, where

$$A \sim \mathcal{W}_p(I_p, q), \quad B \sim \mathcal{W}_p(I_p, n - q),$$

$$A \sim \mathcal{W}_q(I_q, p), \quad B \sim \mathcal{W}_q(I_q, n - p).$$

From [3], the Stieltjes transform of $A(A + B)^{-1}$ converges (a.s.) to

$$(3.1) \quad \check{s}(z) := \frac{z - c_1 - c_2 - \sqrt{(z - d_-)(z - d_+)}}{2c_1z(z - 1)} - \frac{1}{z},$$

and the Stieltjes transform of $\mathcal{A}(\mathcal{A} + \mathcal{B})^{-1}$ converges (a.s.) to

$$(3.2) \quad \tilde{s}(z) := \frac{z - c_1 - c_2 - \sqrt{(z - d_-)(z - d_+)}}{2c_2z(z - 1)} - \frac{1}{z}.$$

Note that $\tilde{s}(z)$ is the Stieltjes transform of the distribution given in (1.6).

Throughout the paper, we will often use the well-known large deviation result of the extreme eigenvalues of Wishart matrices. Assume that $S \sim \mathcal{W}_{n_1}(I_{n_1}, n_2)$ for some $n_1 := n_1(n)$ and $n_2 := n_2(n)$ satisfying $n_1/n \rightarrow a_1 \in (0, 1)$, $n_2/n \rightarrow a_2 \in (0, 1)$ as n tends to infinity and $a := a_1/a_2 \in (0, 1)$. Denoting $\lambda_1(n_2^{-1}S)$ and $\lambda_{n_1}(n_2^{-1}S)$ the largest and the smallest eigenvalues of $n_2^{-1}S$, respectively, it is known that for any given positive number $\varepsilon > 0$,

$$(3.3) \quad (1 + \sqrt{a})^2 + \varepsilon \geq \lambda_1(n_2^{-1}S) \geq \lambda_{n_1}(n_2^{-1}S) \geq (1 - \sqrt{a})^2 - \varepsilon$$

holds with overwhelming probability (see Theorem 2.13 of [16]).

4. Determinant equation. In this section, we derive a determinant equation for the eigenvalues of C_{xy} which are not in the spectrum of the null case. We will see that the CCA matrix in the finite rank case can be viewed as a finite rank perturbation of that in the null case.

As mentioned above, the CCC are invariant under the block diagonal transformation $(\mathbf{x}_i, \mathbf{y}_i) \rightarrow (\mathbf{A}\mathbf{x}_i, \mathbf{B}\mathbf{y}_i)$, for any $p \times p$ matrix \mathbf{A} and $q \times q$ matrix \mathbf{B} , as long as both of them are nonsingular. Hence, to study λ_i , we can start with the following setting:

$$\begin{pmatrix} \mathcal{X} \\ \mathcal{Y} \end{pmatrix} = \Sigma^{1/2}V, \quad \Sigma = \begin{pmatrix} I_p & T \\ T' & I_q \end{pmatrix},$$

where V is a $(p + q) \times n$ matrix with i.i.d. $N(0, 1)$ entries, and

$$(4.1) \quad T = \text{diag}(\sqrt{r_1}, \dots, \sqrt{r_k}) \oplus \mathbf{0}_{(p-k) \times (q-k)},$$

for example, see Muirhead [39], page 530, formula (7). Apparently, there exists a Gaussian matrix \mathcal{W} independent of \mathcal{Y} such that

$$(4.2) \quad \begin{pmatrix} \mathcal{X} \\ \mathcal{Y} \end{pmatrix} = \begin{pmatrix} \mathcal{W} + T\mathcal{Y} \\ \mathcal{Y} \end{pmatrix}.$$

Here, $\text{vec}(\mathcal{W}) \sim N(\mathbf{0}, I_n \otimes (I_p - TT'))$. Hence, \mathcal{W} is a $p \times n$ matrix with independent entries. More specifically, $w_{ij} \sim N(0, 1 - r_i)$. According to the above

definitions, we have $\Sigma_{xx} = I_p$, $\Sigma_{yy} = I_q$, $\Sigma_{xy} = T$ and $\Sigma_{yx} = T'$. For brevity, from now on, we use symbols W, X, Y to denote

$$(4.3) \quad W := \frac{1}{\sqrt{n}} \mathscr{W}, \quad X := \frac{1}{\sqrt{n}} \mathscr{X}, \quad Y := \frac{1}{\sqrt{n}} \mathscr{Y}.$$

Correspondingly, we introduce the notation

$$(4.4) \quad S_{ww} = WW', \quad S_{wy} = WY', \quad S_{yw} = YW', \quad S_{yy} = YY'.$$

In light of (4.2), we have

$$(4.5) \quad \begin{aligned} S_{xx} &= S_{ww} + TS_{yw} + S_{wy}T' + TS_{yy}T', \\ S_{xy} &= S_{wy} + TS_{yy}, \quad S_{yx} = S_{yw} + S_{yy}T'. \end{aligned}$$

Hence, due to the assumption that $\text{rank}(T) = k$, C_{xy} can be regarded as a finite rank perturbation of C_{wy} . Note that, with probability 1, the eigenvalues of C_{xy} are the solutions for λ of the characteristic equation

$$(4.6) \quad \det(D) := \det(S_{xy}S_{yy}^{-1}S_{yx} - \lambda S_{xx}) = 0.$$

In light of (4.5), it is equivalent to

$$\det(S_{wy}S_{yy}^{-1}S_{yw} - \lambda S_{ww} + (1 - \lambda)(TS_{yw} + S_{wy}T' + TS_{yy}T')) = 0.$$

Denote by

$$(4.7) \quad P_y = Y'(YY')^{-1}Y.$$

Then both P_y and $I_n - P_y$ are projections with $\text{rank}(P_y) = q$ and $\text{rank}(I_n - P_y) = n - q$ almost surely. Further, we decompose the matrix W and T as

$$(4.8) \quad W = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}, \quad T = \begin{pmatrix} T_1 \\ \mathbf{0} \end{pmatrix},$$

where W_1 (resp., T_1) is a $k \times n$ (resp., $k \times q$) matrix composed by the first k rows of W (resp., T). Correspondingly, we can introduce the notation like $S_{yw_2} = YW_2'$, $S_{w_2w_2} = W_2W_2'$, analogously to (4.4). Applying the decompositions in (4.8), we can rewrite (4.6) in the following form:

$$(4.9) \quad \det(D) = \det \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} = 0$$

where

$$\begin{aligned} D_{11} &\equiv D_{11}(\lambda) \\ &= W_1P_yW_1' - \lambda W_1W_1' + (1 - \lambda)(T_1YW_1' + W_1Y'T_1' + T_1YY'T_1'), \end{aligned}$$

$$D_{12} \equiv D_{12}(\lambda) = W_1P_yW_2' - \lambda W_1W_2' + (1 - \lambda)T_1YW_2',$$

$$(4.10) \quad D_{21} \equiv D_{21}(\lambda) = W_2P_yW_1' - \lambda W_2W_1' + (1 - \lambda)W_2Y'T_1',$$

$$(4.11) \quad D_{22} \equiv D_{22}(\lambda) = W_2P_yW_2' - \lambda W_2W_2' = S_{w_2w_2}(C_{w_2y} - \lambda I_p),$$

where $C_{w_{2y}}$ is defined analogously to (1.2). In case D_{22} is invertible, we have the equation $\det(D) = \det(D_{22}) \det(D_{11} - D_{12}D_{22}^{-1}D_{21}) = 0$, which implies

$$(4.12) \quad \det(D_{11}(\lambda) - D_{12}(\lambda)D_{22}^{-1}(\lambda)D_{21}(\lambda)) = 0.$$

Apparently, the sufficient condition of $D_{22}(\lambda)$ being invertible is that λ is not an eigenvalue of $C_{w_{2y}}$. Notice that since k is fixed and W_2 is independent of Y , we can apply the results in Theorems 1.3 and 1.5 to the matrix $C_{w_{2y}}$ as well. Hence, if we want to investigate the eigenvalues which are not in $\text{Spec}(C_{w_{2y}})$, it suffices to solve (4.12) and find the properties of the limits of its solutions. For brevity, we introduce the notation

$$(4.13) \quad M_n(z) := D_{11}(z) - D_{12}(z)D_{22}^{-1}(z)D_{21}(z)$$

which is a well-defined $k \times k$ matrix-valued function for all $z \in \mathbb{C} \setminus \text{Spec}(C_{w_{2y}})$. Using (4.10), (4.11) and the fact $Y P_y = Y$, we can write

$$(4.14) \quad M_n(z) = \mathcal{Z}_1 \mathcal{A}(z) \mathcal{Z}'_1,$$

where $\mathcal{Z}_1 = W_1 + T_1 Y$ and

$$(4.15) \quad \mathcal{A}(z) = (P_y - z) - (P_y - z)W'_2(W_2(P_y - z)W'_2)^{-1}W_2(P_y - z).$$

Hereafter, we often use z to represent zI for short.

5. Limits. In this section, we provide the proof of Theorem 1.7 based on several lemmas, whose proofs will be postponed. Our discussion consists of two parts, aimed at (i) and (ii) in Theorem 1.7, respectively: (1) For the outliers, we locate them by deriving the limits of the solutions to the equation (4.12); (2) For the eigenvalues sticking to d_+ , we simply use Cauchy’s interlacing property to get the conclusion.

- *The outliers*

To locate the outliers, we start with the equation (4.12), that is, $\det M_n(z) = 0$. Intuitively, if $M_n(z)$ is close to some deterministic matrix-valued function $M(z)$, it is reasonable to expect that the solutions of (4.12) are close to those of the equation $\det[M(z)] = 0$. Such an implication can be explicitly formulated in the *location lemma* below; see Lemma 5.2. To state the result, we introduce more notation. Set for any positive constant δ the domain

$$(5.1) \quad \mathcal{D} \equiv \mathcal{D}(\delta) := \{z \in \mathbb{C} : d_+ + \delta < \Re z \leq 2, |\Im z| \leq 1\}.$$

Define the functions $s(z), m_i(z) : \mathcal{D} \rightarrow \mathbb{C}$ as follows:

$$(5.2) \quad s(z) = \frac{z - c_1 - c_2 - \sqrt{(z - d_-)(z - d_+)}}{2(z - 1)},$$

$$(5.3) \quad m_i(z) = (c_2 - (1 + c_1)z - (1 - z)s(z))(1 - r_i) + (1 - z) \left(1 - \frac{1 - z}{c_2} s(z) \right) r_i,$$

where the square root is specified as the one with positive imaginary part in case $z \in \mathbb{C}^+$, and $s(\bar{z}) = \overline{s(z)}$. It is elementary to see that $s(z)$ and $m_i(z)$'s are all holomorphic on $\mathbb{C} \setminus [d_-, d_+]$, and thus also on \mathcal{D} . In addition, from the definition (5.2), $s(z)$ satisfies the following equation:

$$(5.4) \quad (z - 1)s^2(z) + (c_1 + c_2 - z)s(z) - c_1c_2 = 0.$$

Recall $\check{s}(z)$ and $\tilde{s}(z)$ from (3.1) and (3.2). We see that

$$(5.5) \quad \check{s}(z) = \frac{1}{c_1z}s(z) - \frac{1}{z}, \quad \tilde{s}(z) = \frac{1}{c_2z}s(z) - \frac{1}{z}.$$

Further, we define the diagonal matrix

$$M(z) := \text{diag}(m_1(z), \dots, m_k(z)).$$

Recall δ in the definition of \mathcal{D} in (5.1). We introduce the following event:

$$(5.6) \quad \Xi_1 \equiv \Xi_1(n, \delta) := \left\{ \|C_{w_{2y}}\| \leq d_+ + \frac{\delta}{2} \right\}.$$

Note that, Theorem 1.5 tells us that Ξ_1 holds with overwhelming probability. Hence, we have $\|(C_{w_{2y}} - z)^{-1}\| = O(1)$ for all $z \in \mathcal{D}$ with overwhelming probability. This implies, in Ξ_1 , $M_n(z)$ is holomorphic on \mathcal{D} almost surely.

Our main technical task for the limits part is the following lemma.

LEMMA 5.1. *For any given $\delta > 0$, and any sufficiently small $\varepsilon > 0$,*

$$\sup_{z \in \mathcal{D}} \sup_{i, j=1, \dots, k} |(M_n)_{ij}(z) - M_{ij}(z)| \leq n^{-\varepsilon}$$

holds almost surely.

The proof of Lemma 5.1 is postponed to the Supplementary Material [9]. The following lemma is a consequence of Lemma 5.1. We remark here, it will be clear that the solutions of the equation $\det(M(z)) = 0$ can only be real.

LEMMA 5.2 (The location lemma). *For any given (but small) $\delta > 0$, let $z_1 > \dots > z_\kappa$ be the solutions in $(d_+ + \delta, 1)$ of the equation $\det(M(z)) = 0$, with multiplicities n_1, \dots, n_κ , respectively. Then for any fixed $\eta > 0$ and each $i \in \{1, \dots, \kappa\}$, almost surely, there exists $z_{n,i,1} > \dots > z_{n,i,\kappa_i}$ with multiplicities $m_{i,1}, \dots, m_{i,\kappa_i}$, respectively, satisfying $\sum_{j=1}^{\kappa_i} m_{i,j} = n_i$, s.t.,*

$$(5.7) \quad \sup_{j=1, \dots, \kappa_i} |z_{n,i,j} - z_i| \leq \eta,$$

and $\bigcup_{i=1}^s \{z_{n,i,j}, j = 1, \dots, \kappa_i\}$ is the collection of all solutions (multiplicities counted) of the equation $\det M_n(z) = 0$ in $(d_+ + \delta, 1)$.

PROOF OF LEMMA 5.2 WITH LEMMA 5.1 GRANTED. At first, as mentioned above, in Ξ_1 , $M_n(z)$ is holomorphic on \mathcal{D} . Moreover, according to (3.3), both Y and W are bounded by some constant in operator norm with overwhelming probability. In addition, we know that Ξ_1 holds with overwhelming probability. Hence, according to the definition of $M_n(z)$ in (4.13), one sees that $\|M_n(z)\|$ is bounded uniformly on \mathcal{D} with overwhelming probability. Therefore, the entries of $M_n(z)$ are all bounded in magnitude with overwhelming probability as well. In addition, it is clear that $M(z)$ is holomorphic and its entries are also bounded in magnitude on \mathcal{D} . Hence, Lemma 5.1 implies

$$\sup_{z \in \mathcal{D}} |\det(M_n(z)) - \det(M(z))| \leq n^{-\varepsilon} \quad \text{almost surely}$$

taking into account the fact that the determinant is a multivariate polynomial of the matrix entries. It is obvious that $\det[M_n(z)]$ only has real roots since the equation (4.6) does. Then by Rouché’s theorem we get (5.7). \square

Now, with the aid of Lemma 5.2, we prove (i) of Theorem 1.7.

PROOF OF (I) OF THEOREM 1.7. According to Lemma 5.2, it suffices to solve $\det(M(z)) = 0$ in $(d_+ + \delta, 1]$ to get z_i , for sufficiently small $\delta > 0$. By the definition of $M(z)$, we shall solve the equation

$$(5.8) \quad m_i(z) = 0.$$

It will be clear that there is a unique simple solution for the above equation. We denote it by γ_i in the sequel.

First, if $r_i = 1$, we get from the definition (5.3) that

$$m_i(\gamma_i) = (1 - \gamma_i) \left(1 - \frac{1 - \gamma_i}{c_2} s(\gamma_i) \right) = 0.$$

It is elementary to check that the only solution is $\gamma_i = 1$. If $r_i < 1$, we introduce the notation

$$(5.9) \quad t_i = r_i / (1 - r_i).$$

According to the definition in (5.3), we see that (5.8) is equivalent to

$$(5.10) \quad \begin{aligned} & \sqrt{\gamma_i^2 + 2(2c_1c_2 - c_1 - c_2)\gamma_i + (c_1 - c_2)^2(c_2 - t_i\gamma_i + t_i)} \\ & = t_i\gamma_i^2 + (t_i(c_2 - c_1 - 1) + c_2 - 2c_1c_2)\gamma_i \\ & \quad + (c_1 - c_2)t_i + c_1c_2 - c_2^2. \end{aligned}$$

Notice that if $\gamma_i \in (d_+, 1]$, we have

$$\gamma_i^2 + 2(2c_1c_2 - c_1 - c_2)\gamma_i + (c_1 - c_2)^2 = (\gamma_i - d_-)(\gamma_i - d_+) > 0$$

and $c_2 - t_i \gamma_i + t_i > 0$. Thus under the restriction that

$$(5.11) \quad t_i \gamma_i^2 + (t_i(c_2 - c_1 - 1) + c_2 - 2c_1 c_2) \gamma_i + (c_1 - c_2)t_i + c_1 c_2 - c_2^2 \geq 0,$$

solving (5.10), we get

$$(5.12) \quad \gamma_i = \frac{(1 + t_i^{-1} c_1)(1 + t_i^{-1} c_2)}{1 + t_i^{-1}} = r_i(1 - c_1 + c_1 r_i^{-1})(1 - c_2 + c_2 r_i^{-1}).$$

Applying (5.9) and (5.12), it is easy to check that (5.11) is equivalent to

$$(5.13) \quad r_i \geq r_c = \sqrt{\frac{c_1 c_2}{(1 - c_1)(1 - c_2)}}.$$

Note that

$$(5.14) \quad \begin{aligned} \gamma_i &= c_1(1 - c_2) + c_2(1 - c_1) + r_i(1 - c_1)(1 - c_2) + r_i^{-1} c_1 c_2 \\ &\geq c_1(1 - c_2) + c_2(1 - c_1) + 2\sqrt{c_1 c_2(1 - c_1)(1 - c_2)} = d_+. \end{aligned}$$

Moreover, equality holds in the second step of (5.14) only if $r_i = r_c$. Hence, (5.8) has solution in (d_+, ∞) only if $r_i > r_c$, with the solution γ_i given by (5.12). Now, what remains is to check $\gamma_i < 1$. It again follows from (5.13) by elementary calculation. Then by (5.7) in Lemma 5.1, we get that

$$\lambda_i - \gamma_i \xrightarrow{\text{a.s.}} 0, \quad \text{for } i = 1, \dots, k_0.$$

Hence, we conclude the proof of (i) of Theorem 1.7. \square

Now, we proceed to the proof of (ii).

• *The Sticking eigenvalues*

PROOF OF THEOREM 1.7(ii). At first, according to the proof of (i) of Theorem 1.7, we see that with overwhelming probability, there are exactly k_0 largest eigenvalues (multiplicities counted) of $C_{x,y}$ in the interval $(d_+ + \delta, 1]$, for any sufficiently small $\delta > 0$. Hence, we see that $\limsup_{n \rightarrow \infty} \lambda_i \leq d_+ + \delta$ for all $i \geq k_0 + 1$ almost surely. Moreover, by the Cauchy's interlacing property and (1.7), we always have $\liminf_{n \rightarrow \infty} \lambda_i \geq d_+ - \delta$ almost surely for any fixed i . Since δ can be arbitrarily small, we get the conclusion that $\lambda_i - d_+ \xrightarrow{\text{a.s.}} 0$ for any fixed $i \geq k_0 + 1$. Hence, we complete the proof. \square

6. Fluctuations of the outliers. For the fluctuations of the outliers, we first derive a CLT for the entries of M_n , which together with (4.12) will imply a CLT for the outliers. Recall from Lemma 5.2 the notation n_1, \dots, n_κ as the multiplicities of the solutions of $\det[M(z)] = 0$. Set

$$J_l = \left\{ \sum_{i=1}^{l-1} n_i + 1, \dots, \sum_{i=1}^l n_i \right\}.$$

We derive the central limit theorem for the n_l -packed sample eigenvalues $\{\sqrt{n}(\lambda_j - \gamma_j), j \in J_l\}$. Note that γ_j 's are all the same for $j \in J_l$, and they are l th largest solution of equation $\det[M(z)] = 0$ with multiplicity n_l . Actually, from Lemma 5.1 and Lemma 5.2, for $j \in J_l$, it would not be difficult to conclude that for any sufficiently small $\varepsilon > 0$,

$$(6.1) \quad \sup_{1 \leq \alpha, \beta \leq k} |(M_n)_{\alpha\beta}(\lambda_j) - M_{\alpha\beta}(\gamma_j)| \leq n^{-\varepsilon}$$

almost surely. However, for Theorem 1.9, we need to take a step further to consider the fluctuation of the difference in (6.1). We split the task into three steps, and the results are collected in Lemmas 6.1–6.3. The first lemma is to expand $M_n(\lambda_j)$ around $M_n(\gamma_j)$. Recall the notation ϑ, ϖ defined in (1.10).

LEMMA 6.1. *Suppose that the assumptions in Theorem 1.9 hold. As $n \rightarrow \infty$, we have*

$$\sqrt{n}[M_n(\lambda_j) - M_n(\gamma_j)] = -\sqrt{n}(\lambda_j - \gamma_j)(\Delta + o_p(1)),$$

where Δ is a diagonal matrix defined as

$$\begin{aligned} \Delta = & \frac{(1 - c_1)\vartheta r_j^2 + c_1^2 c_2}{\vartheta r_j^2 - \varpi} (I_k - T_1 T_1') \\ & + \frac{(1 - c_1)\vartheta r_j^3 + c_1(3c_1 c_2 - c_1 - 2c_2 + 1)r_j - 2c_1^2 c_2}{r_j(\vartheta r_j^2 - \varpi)} T_1 T_1', \end{aligned}$$

and $o_p(1)$ is a $k \times k$ matrix with negligible (in probability) entries.

Recall $\mathcal{A}(z)$ defined in (4.15). We further introduce the following matrices:

$$(6.2) \quad \begin{aligned} \mathcal{M}_1 &\equiv \mathcal{M}_1(z) := W_1 \mathcal{A}(z) W_1', & \mathcal{M}_2 &\equiv \mathcal{M}_2(z) := T_1 Y \mathcal{A}(z) W_1', \\ \mathcal{M}_3 &\equiv \mathcal{M}_3(z) := T_1 Y \mathcal{A}(z) Y' T_1'. \end{aligned}$$

According to the definitions in (4.13), (4.10) and (4.11), it is easy to check

$$(6.3) \quad M_n(z) = \mathcal{M}_1 + \mathcal{M}_2 + \mathcal{M}_2' + \mathcal{M}_3.$$

We denote the SVD of Y by

$$(6.4) \quad Y = U_y \Lambda_y V_y.$$

It is well known that U_y and V_y are Haar distributed, and U_y, Λ_y, V_y are mutually independent (see (3.9) of [18] for instance). Then we see that

$$(6.5) \quad P_y = V_y' \Lambda_y' (\Lambda_y \Lambda_y')^{-1} \Lambda_y V_y$$

is independent of U_y . With the above notation, we introduce

$$(6.6) \quad \mathcal{B} \equiv \mathcal{B}(z) := \Lambda_y V_y \mathcal{A}(z), \quad \mathcal{D} \equiv \mathcal{D}(z) := \Lambda_y V_y \mathcal{A}(z) V_y' \Lambda_y'.$$

Therefore, we can write

$$(6.7) \quad \mathcal{M}_1 = W_1 \mathcal{A} W_1', \quad \mathcal{M}_2 = T_1 U_y \mathcal{B} W_1', \quad \mathcal{M}_3 = T_1 U_y \mathcal{D} U_y' T_1'.$$

Set

$$\mathcal{M}(\gamma_j) := \frac{1}{n} \text{tr} \mathcal{A}(\gamma_j) \cdot (I_k - T_1 T_1') + \frac{1}{q} \text{tr} \mathcal{D}(\gamma_j) \cdot T_1 T_1'.$$

The next lemma depicts the fluctuation of $M_n(\gamma_j)$ around $\mathcal{M}(\gamma_j)$.

LEMMA 6.2. *Suppose that the assumptions in Theorem 1.9 hold. For any $j \in \{1, \dots, k_0\}$ such that r_j is well separated from r_c and 1, we have that $\sqrt{n}(M_n(\gamma_j) - \mathcal{M}(\gamma_j))$ converges weakly to a $k \times k$ symmetric Gaussian matrix \mathcal{R} with independent (up to symmetry) mean zero entries and the variance of $\mathcal{R}_{\alpha\beta}$ is*

$$(6.8) \quad \begin{aligned} & \frac{1 + \delta_{\alpha\beta}}{r_j^2(\vartheta r_j^2 - \varpi)} \left(((1 - c_1)r_j + c_1)^2 (\vartheta r_j - \varpi)^2 \right. \\ & \quad \times ((1 - c_2)r_j^2 + c_2)(1 - r_\alpha)(1 - r_\beta) \\ & \quad + (1 - c_2)((1 - c_1)r_j + c_1)^2 (1 - r_j)^2 (\vartheta r_j - \varpi)^2 \\ & \quad \times (r_\alpha(1 - r_\beta) + r_\beta(1 - r_\alpha)) \\ & \quad + ((1 - c_1)r_j + c_1)^2 (\vartheta r_j - \varpi)^2 \\ & \quad \left. \times ((1 - c_2)r_j^2 + 2c_1(1 - c_2)r_j + c_1(1 - 2c_2))r_\alpha r_\beta \right). \end{aligned}$$

We then claim that $\mathcal{M}(\gamma_j) - M(\gamma_j)$ is of order $o_p(\frac{1}{\sqrt{n}})$.

LEMMA 6.3. *Suppose that the assumptions in Theorem 1.9 hold. For any $j \in \{1, \dots, k_0\}$ such that r_j is well separated from r_c and 1, we have that*

$$\sqrt{n}(\mathcal{M}(\gamma_j) - M(\gamma_j)) = o_p(1),$$

where $o_p(1)$ is a $k \times k$ matrix with negligible (in probability) entries.

The proofs of Lemmas 6.1–6.3 will be stated in the Supplementary Material [9]. With the aid of these lemmas, we can now prove Theorem 1.9.

PROOF OF THEOREM 1.9. Note that $\det M_n(\lambda_j) = 0$, and $M(\gamma_j)$ is a diagonal matrix with diagonal elements $m_i(\gamma_j)$, $i = 1, \dots, k$ [cf. (5.3)]. In addition, for any $l = 1, \dots, l_0$, $m_i(\gamma_j)$ is zero if $i \in J_l$ and nonzero otherwise. Therefore by

Lemmas 6.1–6.3, the definition of T_1 and Skorohod strong representation theorem, we conclude that there exists a random matrix \mathcal{R} defined in Lemma 6.2, such that for any $j, \alpha, \beta \in J_l$,

$$(6.9) \quad \begin{aligned} \sqrt{n}((M_n)_{\alpha\beta}(\lambda_j) - M_{\alpha\beta}(\gamma_j)) &= -\delta_{\alpha\beta} \Delta_{jj} \sqrt{n}(\lambda_j - \gamma_j)(1 + o_p(1)) \\ &\quad + \mathcal{R}_{\alpha\beta} + o_p(1). \end{aligned}$$

Let $\mathcal{N}^{(l)}$ be a $k \times k$ diagonal matrix with $\mathcal{N}_{jj}^{(l)} = n^{1/4}$ if $j \in J_l$ and $\mathcal{N}_{jj}^{(l)} = 1$ otherwise. Apparently, we have $\det[\mathcal{N}^{(l)}] = n^{n_l/4} \neq 0$. Combining Theorem 1.7 with Lemmas 6.1–6.3, we conclude that

$$\mathcal{N}^{(l)} M_n(\lambda_j) \mathcal{N}^{(l)} = M(\gamma_j) + (\sqrt{n}(M_n)_{\alpha\beta}(\lambda_j) \mathbf{1}(j, \alpha, \beta \in J_l))_{k \times k} + o_p(1).$$

Note that $M(\gamma_j)$ is diagonal with J_l diagonal elements being zeros and

$$(\sqrt{n}(M_n)_{\alpha\beta}(\lambda_j) \mathbf{1}(j, \alpha, \beta \in J_l))_{k \times k}$$

is a null matrix except the $J_l \times J_l$ block. According to $\det(\mathcal{N}^{(l)} M_n(\lambda_j) \mathcal{N}^{(l)}) = 0$ and (6.9), we obtain that

$$(6.10) \quad \det\left(-\lim_{n \rightarrow \infty} (\sqrt{n}(\lambda_j - \gamma_j)(I_{n_l} + o_p(1)) + \mathcal{R}(J_l \times J_l) / \Delta(j, j))\right) = 0,$$

where $\mathcal{R}(J_l \times J_l)$ represents the $J_l \times J_l$ block of \mathcal{R} . From (6.10), we see that $\lim_{n \rightarrow \infty} \sqrt{n}(\lambda_j - \gamma_j)$ is the eigenvalue of $\mathcal{R}(J_l \times J_l) / \Delta_{jj}$. Since the eigenvalues of C_{yx} are simple almost surely, we see that the n_l random variables $\{\sqrt{n}(\lambda_j - \gamma_j), j \in J_l\}$ converge in probability to the set of eigenvalues of the matrix $\mathcal{R}(J_l \times J_l) / \Delta_{jj}$. The variance in (1.11) can be obtained from the definition of Δ in Lemma 6.1, (6.8) and also (5.9), after some elementary calculation. Therefore, we complete the proof of Theorem 1.9. \square

7. Fluctuations of the sticking eigenvalues. In this section, we prove Theorem 1.11. When $k = 1$, the main idea is to compare the eigenvalues of C_{xy} with those of C_{w_2y} directly. Then we use the Tracy–Widom law for the largest eigenvalues of C_{w_2y} to conclude the fluctuation of the sticking eigenvalues of C_{xy} . For more general k , we use the argument for the case $k = 1$ recursively. Our proof strategy is inspired by the work [11]. For the comparison, we denote by $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \dots \geq \tilde{\lambda}_q$ the ordered eigenvalues of C_{yw_2} , while use $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q$ to denote the ordered eigenvalues of C_{yx} as before. Our aim is to show the following proposition.

PROPOSITION 7.1. *Suppose that the assumptions in Theorem 1.11 hold. Then for any small positive constant ε and any fixed integer $i \geq 1$, the following holds in probability:*

$$(7.1) \quad |\lambda_{k_0+i} - \tilde{\lambda}_i| \leq n^{-1+\varepsilon}.$$

The full version of the proof of Proposition 7.1 will be presented in the Supplementary Material [9]. Later on, in this section, we will record a special case of Proposition 7.1 and its detailed proof; see Proposition 7.2.

With the aid of Proposition 7.1, we can prove Theorem 1.11 below.

PROOF OF THEOREM 1.11. Let

$$(7.2) \quad \hat{X} := W + T\hat{Y},$$

where \hat{Y} is a i.i.d. copy of Y . Denote by $C_{y\hat{x}}$ via replacing X by \hat{X} in C_{yx} . Then $C_{y\hat{x}}$ is a CCA matrix in the null case. We first compare the eigenvalues of $C_{y\hat{x}}$, denoted by $\hat{\lambda}_i$'s, with the eigenvalues of C_{yw_2} , that is, $\tilde{\lambda}_i$. Applying Proposition 7.1 with $r_1 = \dots = r_k = 0$, we get $\max_{1 \leq i \leq m} |\hat{\lambda}_i - \tilde{\lambda}_i| \leq n^{-1+\varepsilon}$ in probability. This together with (7.1) implies (1.12), which together with Remark 1.12 completes the proof of Theorem 1.11. \square

In the sequel, we show the proof of Proposition 7.1 in the case of $k = 1$. The general k case will be proved in the Supplementary Material [9] via a similar argument. Note that when $k = 1$, We have $k_0 = 1$ or $k_0 = 0$, which means $r_1 > r_c$ or $r_1 < r_c$, respectively. We restate Proposition 7.1 for $k = 1$ as the following proposition.

PROPOSITION 7.2. *Suppose that $k = 1$. We have the following:*

- (i) *If $r_1 > r_c$, then for any fixed integer $m \geq 1$, and any $\varepsilon > 0$, we have $|\lambda_{m+1} - \tilde{\lambda}_m| \leq n^{-1+\varepsilon}$ in probability.*
- (ii) *If $r_1 < r_c$, then for any fixed integer $m \geq 1$, and any $\varepsilon > 0$, we have $|\lambda_m - \tilde{\lambda}_m| \leq n^{-1+\varepsilon}$ in probability.*

To prove Proposition 7.2, we will need the following two lemmas, whose proofs will also be postponed to the Supplementary Material [9].

LEMMA 7.3. *For any $i = 0, \dots, q$, and $j = 1, \dots, k$, the diagonal entry $(M_n)_{jj}(z)$ is a decreasing function in $z \in (\tilde{\lambda}_{i+1}, \tilde{\lambda}_i)$, where $\tilde{\lambda}_0 = \infty$ and $\tilde{\lambda}_{q+1} = -\infty$.*

Let $m \geq 1$ be any fixed integer. For any small constants $\varepsilon, \delta > 0$, we introduce the following random domain:

$$(7.3) \quad \Omega := (\tilde{\lambda}_1 + n^{-1+\varepsilon}, d_+ + \delta) \cup \left\{ \bigcup_{i=1}^m (\tilde{\lambda}_{i+1} + n^{-1+\varepsilon}, \tilde{\lambda}_i - n^{-1+\varepsilon}) \right\}.$$

We also need the following lemma.

LEMMA 7.4. *With the above notation, for any small $\varepsilon > 0$ and $\varepsilon' > 0$:*

- (i) *if $i \neq j$, for any small $\delta > 0$, there exists some positive constant $c > 0$, such that $\sup_{z \in \Omega} |(M_n)_{ij}(z)| \leq n^{-c}$ holds with overwhelming probability;*
- (ii) *there exists a sufficiently small δ , such that $\sup_{z \in \Omega} |(M_n)_{ii}(z) - m_i(d_+)| \leq \varepsilon'$ holds with overwhelming probability.*

With Lemma 7.3 and Lemma 7.4, we prove Proposition 7.2.

PROOF OF PROPOSITION 7.2. Since $k = 1$, we have $\det M_n = M_n(z) = M_{11}(z)$. Applying Lemma 7.4 with $k = 1$, we see that for any small ε' , there is a sufficiently small $\delta > 0$ such that

$$(7.4) \quad \sup_{z \in \Omega} |\det M_n(z) - m_1(d_+)| \leq \varepsilon'$$

with overwhelming probability. Since we assume that r_1 is well separated from r_c , it is easy to check from the definition (5.3) that $|m_1(d_+)| \geq c$ for some positive constant c . Hence, choosing ε' in (7.4) sufficiently small, we see that $|\det M_n(z)| \neq 0$ uniformly on Ω , with overwhelming probability. Hence, with overwhelming probability, there is no eigenvalue of C_{xy} in Ω .

Let $\mathbf{w}_i, i = 1, \dots, p$ be the rows of W . Note that, in case of $k = 1$, we have $\mathbf{w}_1 = W_1$. Moreover, P_w is the projection onto the subspace spanned by $\{\mathbf{w}_1 + T_1 Y, \mathbf{w}_2, \dots, \mathbf{w}_p\}$, and P_{w_2} is the projection onto the subspace spanned by $\{\mathbf{w}_2, \dots, \mathbf{w}_p\}$. Then, by Cauchy interlacing, we know that $\text{Spec}(P_w P_y P_w)$ and $\text{Spec}(P_{w_2} P_y P_{w_2})$ are interlacing. This implies that

$$(7.5) \quad \lambda_1 \geq \tilde{\lambda}_1 \geq \lambda_2 \geq \dots \geq \tilde{\lambda}_{q-1} \geq \lambda_q \geq \tilde{\lambda}_q,$$

since the nonzero eigenvalues of C_{wy} (resp., $C_{w_2 y}$) are the same as those of $P_w P_y P_w$ (resp., $P_{w_2} P_y P_{w_2}$).

In case (i): $r_1 > r_c$, from Theorem 1.7 we know that $\lambda_1 \rightarrow \gamma_1 > d_+ + \delta$ almost surely, for any sufficiently small $\delta > 0$. Since λ_i 's are solutions to $\det M_n(z) = 0$, and $m_1(d_+) \neq 0$, we see from (7.4) and (7.5) that in probability $\lambda_i \in [\tilde{\lambda}_{i+1}, \tilde{\lambda}_{i+1} + n^{-1+\varepsilon}] \cup [\tilde{\lambda}_i - n^{-1+\varepsilon}, \tilde{\lambda}_i]$ for $i = 2, \dots, m$.

In case (ii): $r_1 < r_c$, from Theorem 1.7 we know that $\lambda_1 \rightarrow d_+$ in probability. Hence, we have $\lambda_1 \in [\tilde{\lambda}_1, \tilde{\lambda}_1 + n^{-1+\varepsilon}]$, and for all $i = 2, \dots, m$ $\lambda_i \in [\tilde{\lambda}_{i+1}, \tilde{\lambda}_{i+1} + n^{-1+\varepsilon}] \cup [\tilde{\lambda}_i - n^{-1+\varepsilon}, \tilde{\lambda}_i]$. Therefore, to prove Proposition 7.2, it suffices to check that for all $i = 2, \dots, m$, λ_i is in $[\tilde{\lambda}_{i+1}, \tilde{\lambda}_{i+1} + n^{-1+\varepsilon}]$ or $[\tilde{\lambda}_i - n^{-1+\varepsilon}, \tilde{\lambda}_i]$. Note that $\det M_n(z) = (M_n)_{11}(z)$ is decreasing in the interval $(\tilde{\lambda}_{i+1}, \tilde{\lambda}_i)$, according to Lemma 7.3. Furthermore, since $s(d_+) = \frac{d_+ - c_1 - c_2}{2(d_+ - 1)}$, it is elementary to check $m_j(d_+) > 0$, if $r_j > r_c$, while $m_j(d_+) < 0$ if $r_j < r_c$. Therefore, if $r_1 > r_c$, according to (7.4), we have $\det M_n(z) > 0$ on $(\tilde{\lambda}_{i+1} + n^{-1+\varepsilon}, \tilde{\lambda}_i - n^{-1+\varepsilon})$. By the monotonicity of $\det M_n(z)$, we see that $\det M_n(z) = 0$ can only be achieved if $z \in [\tilde{\lambda}_i - n^{-1+\varepsilon}, \tilde{\lambda}_i]$. Hence, in case of $r_1 > r_c$, we have

$\lambda_i \in [\tilde{\lambda}_i - n^{-1+\varepsilon}, \tilde{\lambda}_i]$ for all $i = 2, \dots, m$. In contrast, when $r_1 < r_c$, we have $\lambda_i \in [\tilde{\lambda}_{i+1}, \tilde{\lambda}_{i+1} + n^{-1+\varepsilon}]$ for all $i = 2, \dots, m$. This concludes the proof of Proposition 7.2. \square

8. Conclusions and discussion. In this work, we study the sample CCCs under the high-dimensional setting stated in Assumption 1.1, for the Gaussian random vector, with a finite rank assumption on the population CCA matrix. Similar to the spiked covariance matrices studied in [4–6, 20, 32, 42], a phase transition phenomenon is found for the largest sample CCCs. To be specific, when the population CCC is larger than the threshold r_c [cf. (1.5)], the corresponding sample CCC will jump out of the continuous part of the spectrum of the CCA matrix and possesses a Gaussian-type fluctuation; when the population CCC is smaller than r_c , the corresponding sample CCC will still stick to the right edge of the continuous part of the spectrum and possesses a Tracy–Widom fluctuation.

In contrast to the well investigated high-dimensional PCA, the study of the high-dimensional CCA is still in its infancy. It is then natural to ask if many other results established for the sample covariance matrices have their analogues for the sample CCA matrix, apart from those obtained in the current paper. The following questions are particularly interesting.

General distribution. It would be natural to ask if our results in the current work can be extended from Gaussian to more general distribution of the matrix entries. Similar to high-dimensional PCA, within the framework of the random matrix theory, it would be reasonable to assume certain linear structure of the random vector, under the assumption of general distribution. More specifically, one can consider the linear model

$$(8.1) \quad \mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \Theta \mathbf{w},$$

where Θ is a $(p + q) \times M$ rectangle matrix with some $M \equiv M(n)$ proportional to n , and \mathbf{w} is an M -dimensional random vector with independent (generally distributed) components. In addition, the matrix $\Sigma = \Theta \Theta'$ has a finite-rank off-diagonal block Σ_{xy} . In general, the linear model (8.1) cannot be further simplified to (4.2) with diagonal T [cf. (4.1)], under the general distribution assumption. Moreover, the proofs of many technical results heavily rely on the Gaussian assumption in the current paper. Therefore, we leave the extension of our results to the general distribution to the future work. We predict that most of the results in the current work would be valid under general distribution assumption.

Full rank case. In the current work, only the finite-rank case is discussed. However, in practice, it can happen that the population CCA matrix itself is of large rank, or even full rank (by full rank we mean rank q). Although most of the r_i 's may be small, they may not be exactly 0. In this case, we have a general nonnull

population. The sample CCA matrix with general population does not correspond to Jacobi ensemble directly, even in the Gaussian case. Even the global behavior of the spectrum is not known yet. We aim at deriving the limiting spectral distribution and the limiting behavior of the largest CCCs under the general population assumption in a future work [8].

Moderately large rank case. Between the finite-rank and full-rank cases, the moderately large rank case deserves a separate discussion. Here by *moderately large rank*, we mean that $\text{rank}(\Sigma_{xy}) \rightarrow \infty$ while $\text{rank}(\Sigma_{xy})/n \rightarrow 0$ as $n \rightarrow \infty$. For instance, $\text{rank}(\Sigma_{xy}) = n^\alpha$ for some constant $\alpha \in (0, 1)$. We predict that the results in Theorems 1.7 and 1.11 will be still true for sufficiently small α . The Gaussian-type fluctuation in Theorem 1.9 may still be valid for the case $k_0 \sim n^\alpha$ with sufficiently small α , if the multiplicity of each distinct r_i , $i \leq k_0$ is still fixed and the distance between the two distinct neighboring r_i 's is not too small. Otherwise, we predict that there would be some interaction between the fluctuations of two neighboring λ_i 's. Another interesting case is that the multiplicity of r_i for some $i \leq k_0$ is large, say, of order n^α . Especially, when the multiplicity of $1 > r_1 > r_c$ is moderately large, we predict that the largest eigenvalue λ_1 will still follow the TW_1 law, but with a different scaling. A similar phenomenon was discovered in [43] for the deformed complex Wigner matrix.

Acknowledgments. The authors would like to thank the anonymous referees, the Associate Editor and the Editor for their invaluable and constructive comments.

SUPPLEMENTARY MATERIAL

Supplement to “Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case”. (DOI: [10.1214/18-AOS1704SUPP](https://doi.org/10.1214/18-AOS1704SUPP); .pdf). In this supplementary material, we present some simulation results and prove Theorem 2.1 and 2.3, Lemmas 6.1–6.3, 7.3–7.4, and also Proposition 7.1.

REFERENCES

- [1] ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ. MR1990662
- [2] BAI, Z., CHOI, K. P. and FUJIKOSHI, Y. (2018). Consistency of AIC and BIC in estimating the number of significant components in high-dimensional principal component analysis. *Ann. Statist.* **46** 1050–1076. MR3797996
- [3] BAI, Z., HU, J., PAN, G. and ZHOU, W. (2015). Convergence of the empirical spectral distribution function of Beta matrices. *Bernoulli* **21** 1538–1574. MR3352053
- [4] BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053
- [5] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575
- [6] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680

- [7] BAO, Z., HU, J., PAN, G. and ZHOU, W. (2017). Test of independence for high-dimensional random vectors based on freeness in block correlation matrices. *Electron. J. Stat.* **11** 1527–1548. [MR3635921](#)
- [8] BAO, Z. and HU, J. (2018) High-dimensional CCA with general population. (In preparation).
- [9] BAO, Z., HU, J., PAN, G. and ZHOU, W. (2019). Supplement to “Canonical correlation coefficients of high-dimensional Gaussian vectors: finite rank case.” DOI:10.1214/18-AOS1704SUPP.
- [10] BELINSCHI, S. T., BERCOVICI, H., CAPITAINE, M. and FÉVRIER, M. (2017). Outliers in the spectrum of large deformed unitarily invariant models. *Ann. Probab.* **45** 3571–3625. [MR3729610](#)
- [11] BENAYCH-GEORGES, F., GUIONNET, A. and MAIDA, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.* **16** 1621–1662. [MR2835249](#)
- [12] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. [MR2782201](#)
- [13] BRETHERTON, C. S., SMITH, C. and WALLACE, J. M. (1992). An intercomparison of methods for finding coupled patterns in climate data. *J. Climate* **5** 541–560.
- [14] CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. [MR2489158](#)
- [15] CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2012). Central limit theorems for eigenvalues of deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 107–133. [MR2919200](#)
- [16] DAVIDSON, K. R. and SZAREK, S. J. (2001). Local operator theory, random matrices and Banach spaces. In *Handbook of the Geometry of Banach Spaces, Vol. 1* 317–366. North-Holland, Amsterdam. [MR1863696](#)
- [17] DUTILLEUL, P., PELLETIER, B. and ALPARGU, G. (2008). Modified F tests for assessing the multiple correlation between one spatial process and several others. *J. Statist. Plann. Inference* **138** 1402–1415. [MR2388019](#)
- [18] EDELMAN, A. and RAO, N. R. (2005). Random matrix theory. *Acta Numer.* **14** 233–297. [MR2168344](#)
- [19] FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. [MR2291807](#)
- [20] FÉRAL, D. and PÉCHÉ, S. (2009). The largest eigenvalues of sample covariance matrices for a spiked population: Diagonal case. *J. Math. Phys.* **50** 073302, 33. [MR2548630](#)
- [21] FUJIKOSHI, Y. (2016). High-Dimensional Asymptotic Distributions of Characteristic Roots in multivariate linear models and canonical correlation analysis. Technical report.
- [22] FUJIKOSHI, Y. and SAKURAI, T. (2016). High-dimensional consistency of rank estimation criteria in multivariate linear model. *J. Multivariate Anal.* **149** 199–212. [MR3507324](#)
- [23] GAO, C., MA, Z., REN, Z. and ZHOU, H. H. (2015). Minimax estimation in sparse canonical correlation analysis. *Ann. Statist.* **43** 2168–2197. [MR3396982](#)
- [24] GAO, C., MA, Z. and ZHOU, H. H. (2017). Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist.* **45** 2074–2101. [MR3718162](#)
- [25] GITTINS, R. (1985). *Canonical Analysis. Biomathematics* **12**. Springer, Berlin, Heidelberg.
- [26] HAN, X., PAN, G. and YANG, Q. (2018). A unified matrix model including both CCA and F matrices in multivariate analysis: The largest eigenvalue and its applications. *Bernoulli* **24** 3447–3468. [MR3788178](#)
- [27] 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](#)

- [28] HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28** 321–377.
- [29] HYODO, M., SHUTOH, N., NISHIYAMA, T. and PAVLENKO, T. (2015). Testing block-diagonal covariance structure for high-dimensional data. *Stat. Neerl.* **69** 460–482. [MR3414707](#)
- [30] JIANG, D., BAI, Z. and ZHENG, S. (2013). Testing the independence of sets of large-dimensional variables. *Sci. China Math.* **56** 135–147. [MR3016588](#)
- [31] JIANG, T. and YANG, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074. [MR3127857](#)
- [32] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. [MR1863961](#)
- [33] JOHNSTONE, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. [MR2485010](#)
- [34] JOHNSTONE, I. M. (2009). Approximate null distribution of the largest root in multivariate analysis. *Ann. Appl. Stat.* **3** 1616–1633. [MR2752150](#)
- [35] JOHNSTONE, I. M. and ONATSKI, A. (2015). Testing in high-dimensional spiked models. [arXiv:1509.07269](#).
- [36] KARGIN, V. (2015). Subordination for the sum of two random matrices. *Ann. Probab.* **43** 2119–2150. [MR3353823](#)
- [37] KATZ-MOSES, B. (2012). *Small Deviations for the Beta–Jacobi Ensemble*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—University of Colorado at Boulder. [MR3034656](#)
- [38] KNOWLES, A. and YIN, J. (2013). The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.* **66** 1663–1750. [MR3103909](#)
- [39] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. [MR0652932](#)
- [40] ODA, R., YANAGIHARA, H. and FUJIKOSHI, Y. (2016). Asymptotic non-null distributions of test statistics for redundancy in the high-dimensional canonical correlation analysis Technical report.
- [41] PASSEMIER, D. and YAO, J.-F. (2012). On determining the number of spikes in a high-dimensional spiked population model. *Random Matrices Theory Appl.* **1** 1150002, 19. [MR2930380](#)
- [42] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. [MR2399865](#)
- [43] PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. [MR2221787](#)
- [44] RENCHER, A. C. and PUN, F. C. (1980). Inflation of R^2 in best subset regression. *Technometrics* **22** 49.
- [45] WACHTER, K. W. (1980). The limiting empirical measure of multiple discriminant ratios. *Ann. Statist.* **8** 937–957. [MR0585695](#)
- [46] WANG, Q. and YAO, J. (2017). Extreme eigenvalues of large-dimensional spiked Fisher matrices with application. *Ann. Statist.* **45** 415–460. [MR3611497](#)
- [47] YAMADA, Y., HYODO, M. and NISHIYAMA, T. (2017). Testing block-diagonal covariance structure for high-dimensional data under non-normality. *J. Multivariate Anal.* **155** 305–316. [MR3607897](#)
- [48] YANG, Y. and PAN, G. (2012). The convergence of the empirical distribution of canonical correlation coefficients. *Electron. J. Probab.* **17** no. 64, 13. [MR2968671](#)
- [49] YANG, Y. and PAN, G. (2015). Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann. Statist.* **43** 467–500. [MR3316187](#)
- [50] ZHENG, S., JIANG, D., BAI, Z. and HE, X. (2014). Inference on multiple correlation coefficients with moderately high dimensional data. *Biometrika* **101** 748–754. [MR3254916](#)

Z. BAO
DEPARTMENT OF MATHEMATICS
HONG KONG UNIVERSITY OF SCIENCE
AND TECHNOLOGY
HONG KONG
E-MAIL: mazbao@ust.hk

G. PAN
DIVISION OF MATHEMATICAL SCIENCES
NANYANG TECHNOLOGICAL UNIVERSITY
637371
SINGAPORE
E-MAIL: gmpan@ntu.edu.sg

J. HU
SCHOOL OF MATHEMATICS AND STATISTICS
NORTHEAST NORMAL UNIVERSITY
130024 CHANG CHUN
P. R. CHINA
E-MAIL: huj156@nenu.edu.cn

W. ZHOU
DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
117546
SINGAPORE
E-MAIL: stazw@nus.edu.sg