Journal of Multivariate Analysis 174 (2019) 104524

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Roy's largest root under rank-one perturbations: The complex valued case and applications

Prathapasinghe Dharmawansa [a], Boaz Nadler [b,*], Ofer Shwartz [b]

[a] Department of Electronic and Telecommunication Engineering, University of Moratuwa, Sri Lanka
[b] Department of Computer Science, Weizmann Institute of Science, Rehovot, 76100, Israel

ARTICLE INFO

Article history:
Received 3 June 2018
Received in revised form 19 March 2019
Accepted 20 May 2019
Available online 2 July 2019

AMS 2010 subject classifications:
primary 60B20
secondary 62H10
33C15

Keywords:
Complex Wishart distribution
Rank-one perturbation
Roy's largest root
Signal detection in noise

ABSTRACT

The largest eigenvalue of a single or a double Wishart matrix, both known as Roy's largest root, plays an important role in a variety of applications. Recently, via a small noise perturbation approach with fixed dimension and degrees of freedom, Johnstone and Nadler derived simple yet accurate approximations to its distribution in the real valued case, under a rank-one alternative. In this paper, we extend their results to the complex valued case for five common single matrix and double matrix settings. In addition, we study the finite sample distribution of the leading eigenvector. We present the utility of our results in several signal detection and communication applications, and illustrate their accuracy via simulations.

1. Introduction

Wishart matrices, both real and complex valued, play a central role in statistics, with numerous engineering applications, specifically signal processing and communications. Of particular interest are the roots of a single Wishart matrix H, and of a double Wishart matrix $E^{-1}H$, with H and E independent [1]. The latter can be viewed as the multivariate analogue of the univariate F distribution and is also closely related to the multivariate beta distribution [32, Section 3.3]. Here we consider the largest eigenvalue ℓ_1 of either the matrix H or the matrix $E^{-1}H$, a test statistic proposed by Roy [38,39], known as Roy's largest root [32, Section 10.6]. Specifically, we focus on the complex-valued case where H, E are independent complex-valued Wishart matrices. Throughout this paper, we consider $m \times m$ matrices, where E follows a complex valued central Wishart distribution with n_E degrees of freedom and identity covariance matrix $\Sigma_E = I$, denoted $E \sim \mathcal{CW}_m(n_E, I)$. The distribution of the matrix H will either be central $H \sim \mathcal{CW}_m(n_H, \Sigma_H)$, or non-central $H \sim \mathcal{CW}_m(n_H, \Sigma_H, \Omega)$. For the definition of central and non-central complex valued Wishart matrices, see for example [15] and [19, Section 8].

Obtaining simple expressions, exact or approximate, for the distribution of this top eigenvalue, denoted by ℓ_1, in the single or double matrix case has been a subject of intense research for more than 50 years. Khatri [27] derived an exact expression for the distribution of ℓ_1 in the single central matrix case with an identity covariance matrix ($\Sigma_H = I$). His result was generalized to several other settings, such as an arbitrary covariance matrix or a non-centrality matrix [24,28,36,37,41]. The resulting expressions are, in general, challenging to evaluate numerically. More recently,

* Corresponding author.
E-mail address: boaz.nadler@weizmann.ac.il (B. Nadler).

https://doi.org/10.1016/j.jmva.2019.05.009

Table 1
Five common single-matrix and double-matrix cases. The middle column describes the distribution of the covariance matrices of the observed data. In the first two cases only one sample covariance matrix is computed. The right column describes several relevant applications.

Case	General Form of Distribution	Application
1	$H \sim \mathcal{CW}_m(n_H, \Sigma + \lambda \mathbf{v}\mathbf{v}^\dagger)$ Σ is known	Signal detection in noise, known noise covariance matrix.
2	$H \sim \mathcal{CW}_m(n_H, \Sigma, \omega \mathbf{v}\mathbf{v}^\dagger)$ Σ is known	Constant modulus signal detection in noise, known noise covariance matrix.
3	$H \sim \mathcal{CW}_m(n_H, \Sigma + \lambda \mathbf{v}\mathbf{v}^\dagger)$ $E \sim \mathcal{CW}_m(n_E, \Sigma)$	Signal detection in noise, estimated noise covariance matrix.
4	$H \sim \mathcal{CW}_m(n_H, \Sigma, \omega \mathbf{v}\mathbf{v}^\dagger)$ $E \sim \mathcal{CW}_m(n_E, \Sigma)$	Constant modulus signal detection in noise, estimated noise covariance matrix.
5	$H \sim \mathcal{CW}_p(q, \Phi, \Omega)$ $E \sim \mathcal{CW}_p(n - q, \Phi)$ Ω is a rank-one matrix	Canonical correlation analysis between two groups of sizes $p \le q$

Zanella et al. [44] derived simpler exact, yet recursive expressions, both for the central case with arbitrary Σ_H and for the non-central case but with $\Sigma_H = I$. Alternative recursive formulas in the real-valued case and in the complex-valued case were derived by Chiani [4–6].

A different approach to derive approximate distributions for the largest eigenvalue when $\Sigma_E = \Sigma_H = I$, is based on random matrix theory. Considering the limit as n_H and m (and in the double matrix case also n_E) tend to infinity, with their ratios converging to constants, ℓ_1 in the single matrix case and $\ln(\ell_1)$ in the double matrix case, asymptotically follow a Tracy–Widom distribution [20–22]. Furthermore, with suitable centering and scaling, the convergence to these limiting distributions is quite fast [10,31].

In this paper, motivated by statistical signal detection and communication applications, we consider complex valued Wishart matrices H whose population covariance is a rank-one perturbation of a base covariance matrix. Specifically, in the central case we assume $\Sigma_H = I + \lambda \mathbf{v}\mathbf{v}^\dagger$, where λ is a measure of signal strength, the unit norm vector $\mathbf{v} \in \mathbb{C}^m$ is its direction, and $\mathbf{v}^\dagger$ denotes the conjugate transpose of $\mathbf{v}$. Similarly, in the non-central case, we assume that $H \sim \mathcal{CW}_m(n_H, I, \Omega)$, with a rank-one non-centrality matrix $\Omega = \lambda \mathbf{v}\mathbf{v}^\dagger$. Our goal is to study the distribution of ℓ_1 and its dependence on λ, which as discussed below is a central quantity of interest in various applications. A classical result in the single-matrix case is that with dimension m fixed, as $n_H \to \infty$, the largest eigenvalue of H converges to a Gaussian distribution [1]. In the random matrix setting, as both n_H and m tend to infinity with their ratio tending to a constant, Baik et al. [3] and Paul [34] proved that if $\lambda > \sqrt{m/n_H}$ then ℓ_1 still converges to a Gaussian distribution, but with a different variance. In the two-matrix case, the location of the phase transition and the limiting value of the largest eigenvalue of $E^{-1}H$ were recently studied by Nadakuditi and Silverstein [33]. Dharmawansa et al. [8] proved that above the phase transition, ℓ_1 converges to a Gaussian distribution and provided an explicit expression for its asymptotic variance.

Whereas the above results assume that dimension and degrees of freedom tend to infinity, in various common applications these quantities are relatively small. In such settings, the above mentioned asymptotic results may provide a poor approximation to the distribution of the largest eigenvalue ℓ_1, which can be quite far from Gaussian, see Fig. 1 (left) for an illustrative example. Accurate expressions for the distribution of ℓ_1, for small dimension and degrees of freedom, were recently derived for single and double real-valued Wishart matrices by Johnstone and Nadler [23], via a small noise perturbation approach. In this paper, we build upon their work and extend their results to the *complex valued case* and to the study of the distribution of the leading sample eigenvector, not considered in their work. As discussed below, both are important quantities in various applications.

Propositions 1–5 in Section 2 provide approximate expressions for the distribution of ℓ_1 under the five single-matrix and double-matrix cases outlined in Table 1. In Section 3 we study the finite sample fluctuations of the leading eigenvector and its overlap with the population eigenvector. Next, in Section 4 we illustrate the utility of these approximations in signal detection and communication applications. Specifically, Section 4.1 considers the power of Roy's largest root test under two common signal models, whereas Section 4.2 considers the outage probability in a specific multiple-input and multiple-output (MIMO) communication system [24]. For a rank-one Rician fading channel, we show analytically that to minimize the outage probability it is preferable to have an equal number of transmitting and receiving antennas. This important design property was previously observed via simulations [24].

2. On the distribution of Roy's largest root

Table 1 outlines five common single matrix and double matrix complex Wishart cases, along with some representative applications. Propositions 1–5, are the complex analogues of those in [23], and provide simple approximations to the distribution of Roy's largest root in these cases. As outlined in the appendix, their proof follows those of [23], with some notable differences. In particular, we present complex valued analogues of some well known results for real valued Wishart matrices. In what follows we denote by $\mathbb{E}$ the expectation operator. We also denote by χ^2_k the chi-squared distribution with k degrees of freedom and by $\chi_k^2(\eta)$ the non-central chi-squared distribution with non-centrality parameter η. Throughout the manuscript we follow the standard definition of complex valued multivariate Gaussian random variables, see [15]. Specifically, if $X \sim \mathcal{CN}(0, \sigma^2)$ then it can be written as $(A + \iota B)/\sqrt{2}$ where $A, B \in \mathbb{R}$ are independent $\mathcal{N}(0, \sigma^2)$ random variables and $\iota = \sqrt{-1}$.

We start with the simplest Case 1 in Table 1, involving a single central Wishart matrix, $H \sim \mathcal{CW}_m(n_H, \Sigma + \lambda \mathbf{v}\mathbf{v}^\dagger)$. In various engineering applications the matrix Σ denotes the covariance of the noise measured at m sensors and is often assumed to be known, whereas λ is a measure of the signal strength and the unit norm vector $\mathbf{v}$ denotes its direction. Without loss of generality, we thus assume $\Sigma = \sigma^2 I$, where σ^2 then denotes the noise variance. In contrast to previous asymptotic approaches, whereby the number of samples $n_H \to \infty$ and possibly also the dimension $m \to \infty$, in the following we keep n_H and m fixed, and study the distribution of the largest eigenvalue in the limit of small noise, namely as $\sigma \to 0$. To emphasize that we study the dependence of the largest eigenvalue of H on the parameter σ, we shall denote it by $\ell_1(\sigma)$.

Proposition 1. *Let $H \sim \mathcal{CW}_m(n_H, \lambda \mathbf{v}\mathbf{v}^\dagger + \sigma^2 I)$, with $\|\mathbf{v}\| = 1$, $\lambda > 0$ and let $\ell_1(\sigma)$ be its largest eigenvalue. Then, with (m, n_H, λ) fixed, as $\sigma \to 0$*

$$\ell_1(\sigma) = \frac{\lambda + \sigma^2}{2} A + \frac{\sigma^2}{2} B + \frac{\sigma^4}{2(\lambda + \sigma^2)} \frac{BC}{A} + o_P(\sigma^4) \tag{1}$$

where A, B, C are independent random variables, distributed as $A \sim \chi^2_{2n_H}$, $B \sim \chi^2_{2m-2}$, and $C \sim \chi^2_{2n_H-2}$.

Remark 1. Given that ℓ_1 is the largest eigenvalue of a Wishart matrix, it has finite mean and variance. Approximate formulas for these quantities follow directly from (1). Since $\mathbb{E}\{\chi_k^2\} = k$ $\mathrm{Var}\{\chi_k^2\} = 2k$, and $\mathbb{E}\{1/\chi_k^2\} = 1/(k-2)$ for $k > 2$ then for $n_H > 1$

$$\mathbb{E}\{\ell_1(\sigma)\} = \lambda n_H + (n_H + m - 1)\sigma^2 + \frac{\sigma^4}{\lambda + \sigma^2}(m-1) + o(\sigma^4),$$

and similarly,

$$\mathrm{Var}\{\ell_1(\sigma)\} = \lambda^2 n_H + 2\lambda n_H \sigma^2 + (n_H + m - 1)\sigma^4 + o(\sigma^4).$$

Remark 2. The exact distribution of the largest eigenvalue ℓ_1 in the setting of Proposition 1, with number of samples larger than the dimension, has been recently derived by Chiani [6, Theorem 4, part 3]. The result is given in terms of the determinant of an $m \times m$ matrix whose entries depend on the generalized incomplete gamma function, with parameters that depend on λ and on σ. In contrast, while (1) is approximate, the dependence on the values of λ and σ is more explicit.

The next proposition considers a non-central single Wishart, Case 2 in Table 1.

Proposition 2. *Let $H \sim \mathcal{CW}_m(n_H, \sigma^2 I, (\omega/\sigma^2)\mathbf{v}\mathbf{v}^\dagger)$, with $\|\mathbf{v}\| = 1$, $\omega > 0$ and let $\ell_1(\sigma)$ be its largest eigenvalue. Then, with (m, n_H, ω) fixed, as $\sigma \to 0$*

$$\ell_1(\sigma) = \frac{\sigma^2}{2}\left(A + B + \frac{BC}{A}\right) + o_P(\sigma^4) \tag{2}$$

where A, B, C are all independent and distributed as $A \sim \chi^2_{2n_H}(2\omega/\sigma^2)$, $B \sim \chi^2_{2m-2}$ and $C \sim \chi^2_{2n_H-2}$.

Remark 3. By definition, $\mathbb{E}\{\chi_k^2(\eta)\} = k + \eta$. Furthermore, it is easy to show that as $\eta \to \infty$, $\mathbb{E}\{(\chi_k^2(\eta))^{-1}\} = (k - 2 + \eta)^{-1}\{1 + O(\eta^{-1})\}$ and $\mathrm{Var}\{(\chi_k^2(\eta))^{-1}\} = 2/\{(k + \eta - 2)^2(k + \eta - 4)\} \cdot \{1 + O(\eta^{-1})\}$. Note that as $\sigma \to 0$ the non-centrality parameter $2\omega/\sigma^2$ which appears in the random variable A in (2) tends to infinity. Hence, for small σ, we can approximate the mean and variance of $\ell_1(\sigma)$ in (2) by

$$\mathbb{E}\{\ell_1(\sigma)\} \approx \omega + (n_H + m - 1)\sigma^2 + \frac{(n_H - 1)(m - 1)}{\sigma^2(n_H - 1) + \omega}\sigma^4$$

and

$$\mathrm{Var}\{\ell_1(\sigma)\} \approx 8\omega + 4\sigma^2 \left\{ n_H + m - 1 + \frac{(n_H - 1)(m - 1)}{2(n_H + \frac{\sigma^2}{\omega} - 1)^2 (n_H + \frac{\sigma^2}{\omega} - 2)} \right\}.$$

The next two propositions provide approximations to the distribution of Roy's largest root in the central and non-central double matrix settings, which correspond to Cases 3 and 4 in Table 1. For Case 3, for example, in principle we need to study $\ell_1(E^{-1}H)$ where $E \sim \mathcal{CW}_m(n_E, \Sigma)$ and $H \sim \mathcal{CW}_m(n_H, \Sigma + \tilde{\lambda}\mathbf{w}\mathbf{w}^\dagger)$. However, a simplification can be made based on the following observations: (i) The matrix $E^{-1}H$ has the same eigenvalues as $\Sigma^{1/2}E^{-1}H\Sigma^{-1/2}$, which is equal to $(\Sigma^{-1/2}E\Sigma^{-1/2})^{-1}(\Sigma^{-1/2}H\Sigma^{-1/2})$, (ii) the matrix $\Sigma^{-1/2}E\Sigma^{-1/2} \sim \mathcal{CW}_m(n_E, I)$ and (iii) the matrix $\Sigma^{-1/2}H\Sigma^{-1/2} \sim \mathcal{CW}_m(n_H, I + \lambda \mathbf{v}\mathbf{v}^\dagger)$, where $\mathbf{v} = \Sigma^{-1/2}\mathbf{w}/\|\Sigma^{-1/2}\mathbf{w}\|$ has unit norm, and $\lambda = \|\Sigma^{-1/2}\mathbf{w}\|^2\,\tilde{\lambda}$. Hence, in the following propositions we assume without loss of generality that the covariance matrix of E is $\Sigma = I$.

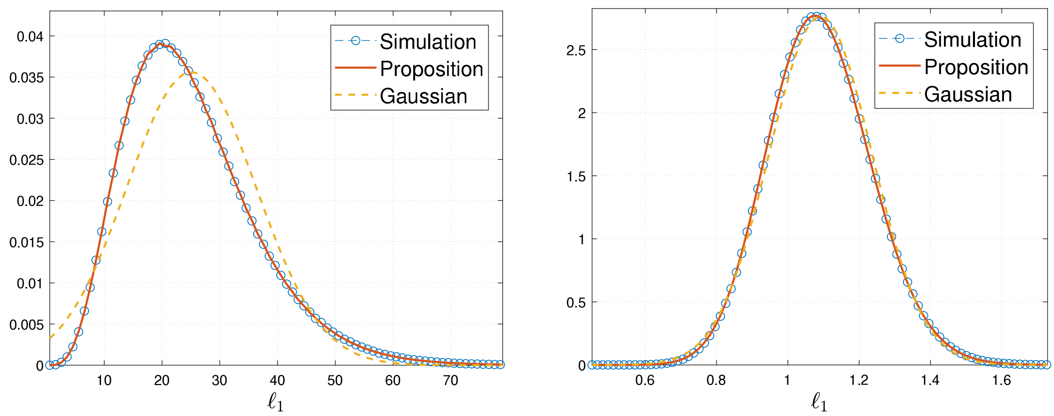

Fig. 1. Density of the largest eigenvalue in Case 1 (left) and Case 2 (right). The parameters for Case 1 are $n_H = m = 5, \lambda = 5$ and $\sigma^2 = 0.01$. For Case 2 they are $n_H = m = 5$, $\omega = 1$ and $\sigma^2 = 0.01$. The red solid line corresponds to Propositions 1 and 2.

Proposition 3. *Let $H \sim \mathcal{CW}_m(n_H, I + \lambda \mathbf{v}\mathbf{v}^\dagger)$ and $E \sim \mathcal{CW}_m(n_E, I)$ be independent, with $n_E > m + 1$ and $\|\mathbf{v}\| = 1$. Let ℓ_1 be the largest eigenvalue of $E^{-1}H$. Then, with (m, n_H, n_E) fixed, as λ becomes large*

$$\ell_1(\lambda) \approx (1 + \lambda) a_1 F_{b_1,c_1} + a_2 F_{b_2,c_2} + a_3 \tag{3}$$

where the two F distributed random variates are independent and

$$\begin{aligned} a_1 &= \frac{n_H}{n_E - m + 1}, \qquad a_2 = \frac{m-1}{n_E - m + 2}, \qquad a_3 = \frac{m-1}{(n_E - m)(n_E - m - 1)}, \\ b_1 &= 2n_H, \quad b_2 = 2m - 2, \quad c_1 = 2n_E - 2m + 2, \quad c_2 = 2n_E - 2m + 4. \end{aligned} \tag{4}$$

Proposition 4. *Suppose that $H \sim \mathcal{CW}_m(n_H, I, \omega \mathbf{v}\mathbf{v}^\dagger)$ and $E \sim \mathcal{CW}_m(n_E, I)$ are independent, with $n_E > m + 1$, $\omega > 0$, and $\|\mathbf{v}\| = 1$. Let ℓ_1 be the largest eigenvalue of $E^{-1}H$. Then, with (m, n_H, n_E) fixed, as ω becomes large*

$$\ell_1(\omega) \approx a_1 F_{b_1,c_1}(2\omega) + a_2 F_{b_2,c_2} + a_3 \tag{5}$$

where the two F distributed random variates are independent and the parameters a_i, b_i, c_i are given in (4).

Remark 4. In the limit as $n_E \to \infty$, the two F-distributed random variables in (3) and (5) converge to χ^2 distributed random variables, thus recovering the leading order terms in (1) and (2), respectively.

Let us illustrate the accuracy of our approximations via several simulations. Fig. 1 compares the empirical density of the largest eigenvalue, computed from 10^5 independent Monte Carlo realizations, in Cases 1 and 2 defined in Table 1, to the two corresponding propositions. For reference, we also plot the standard Gaussian density. The accuracy of our proposition for computing tail probabilities of the form $\Pr(\ell_1 > t)$ is illustrated in Fig. 2 for Case 1. Similar results (not shown) hold for other cases. Results for Cases 3 and 4 of Table 1 are shown in Fig. 3. As can be seen, in all cases, due to the small sample size and dimension, the distribution of the largest root deviates significantly from the asymptotic Gaussian one, with our propositions being significantly more accurate.

2.1. On the leading canonical correlation coefficient

We now consider the fifth Case of Table 1 and study the largest sample canonical correlation coefficient between a first group of p variables and a second group of q variables, in the presence of a single large canonical correlation coefficient in the population. Canonical correlation analysis is widely used in a variety of applications, for example in medical image processing [7,26,30], signal processing [2,35,40], and array processing [11].

Since the canonical correlation is invariant under unitary transformations within each of the two groups of variables, in the presence of a single large correlation coefficient, without loss of generality we can choose the following form for the matrix Σ,

$$\Sigma = \begin{pmatrix} I_p & \tilde{P} \\ \tilde{P}^\top & I_q \end{pmatrix}.$$

Here $\tilde{P} = \begin{pmatrix} P & \mathbf{0}_{p\times(q-p)} \end{pmatrix}$ with $P = \mathrm{diag}\,(\rho, 0, \ldots, 0) \in \mathbb{R}^{p\times p}$ and ρ is the value of the correlation coefficient.

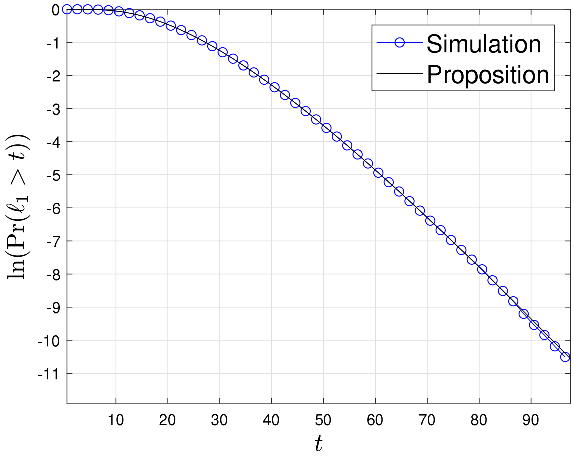

Fig. 2. Tail probabilities for largest eigenvalue in Case 1, same parameters as in Fig. 1.

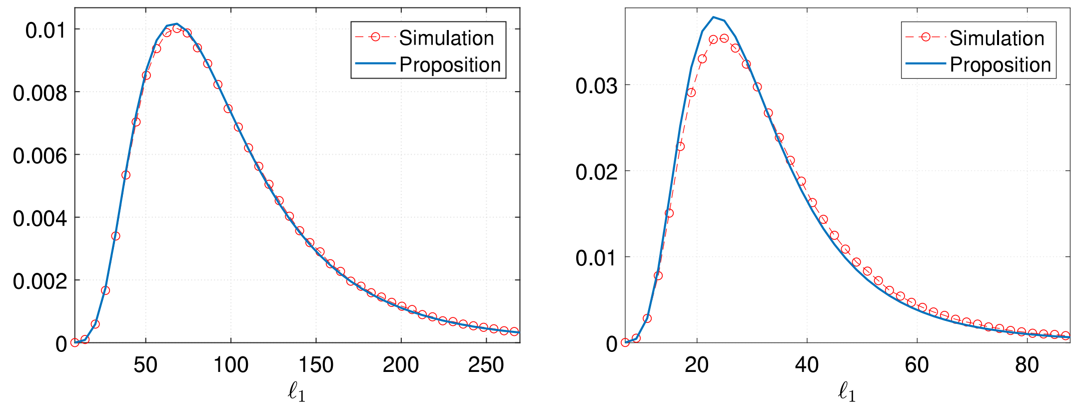

Fig. 3. Density of the largest eigenvalue in Case 3 (left) and Case 4 (right). In both plots $n_E = n_H = 10$, $m = 5$. In Case 3, $\lambda = 50$ and in Case 4 $\omega = 150$. The blue solid line corresponds to Propositions 3 and 4.

To study the sample canonical correlation, consider $n + 1$ complex-valued m-dimensional multivariate Gaussian observations $\mathbf{x}_i \sim \mathcal{CN}(0, \Sigma)$, $i \in \{1, \ldots, n+1\}$ on $m = p + q$ variables, where without loss of generality $p \le q$. The corresponding sample covariance matrix S decomposes as

$$nS = \begin{pmatrix} Y^\dagger Y & Y^\dagger X \\ X^\dagger Y & X^\dagger X \end{pmatrix},$$

where $Y \in \mathbb{C}^{n\times p}$ and $X \in \mathbb{C}^{n\times q}$ represent the first p variables and the remaining q variables, respectively.

Our interest is in the largest sample canonical correlation coefficient, denoted by r_1. Similar to the real valued case [32, Chapter 10], its square r_1^2 is the largest root of the following characteristic equation

$$\det\left(r^2 Y^\dagger Y - Y^\dagger Q Y\right) = 0, \tag{6}$$

where $Q = X\left(X^\dagger X\right)^{-1} X^\dagger$. Introducing the notation $H = Y^\dagger Q Y$ and $E = Y^\dagger (I_p - Q) Y$, (6) can be rewritten as

$$\det\left(r^2(H + E) - H\right) = 0.$$

Hence, we may equivalently study the largest root of $E^{-1}H$, since it is related to r_1^2 by $\ell_1 = r_1^2/(1 - r_1^2)$.

Similar to [23], it can be shown that with $\Phi = I_p - P^2$, conditional on X, the two matrices H and E are independent and distributed as

$$H|X \sim \mathcal{CW}_p(q, \Phi, \Omega) \quad \text{and} \quad E|X \sim \mathcal{CW}_p(n - q, \Phi) \tag{7}$$

with the non-centrality matrix given by

$$\Omega = \Phi^{-1}\tilde{P}X^\dagger X\tilde{P}^\top = \frac{\rho^2}{1-\rho^2}\left(X^\dagger X\right)_{11}\mathbf{e}_1\mathbf{e}_1^\dagger = \omega\,\mathbf{e}_1\mathbf{e}_1^\dagger \quad \text{where } \omega = \frac{\rho^2}{1-\rho^2}\left(X^\dagger X\right)_{11}. \tag{8}$$

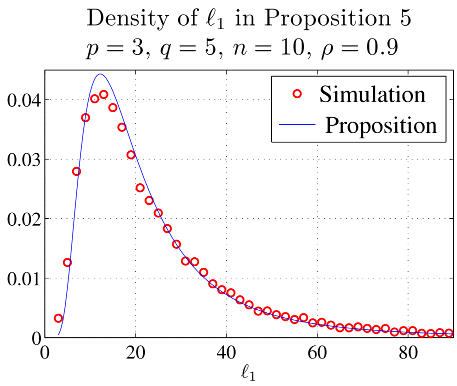

Fig. 4. Density function of $\ell_1(E^{-1}H)$ in canonical correlation analysis.

Since $X^\dagger X \sim \mathcal{CW}_q\left(n, I_q\right)$, then all diagonal entries of $X^\dagger X$ follow a chi-square distribution. In particular, $\left(X^\dagger X\right)_{11} \sim \chi^2_{2n}/2$. The next proposition provides an approximation to the distribution of the largest sample canonical correlation in the presence of a single population canonical correlation. To this end, we introduce the following notation. We denote by $F^{\chi}_{a,b}(c, n)$ a random variable, which is defined as a function of three other random variables as follows: First, generate a random variable $Z \sim c\chi^2_n$. Next, generate two independent random variables, one distributed as $\chi^2_a(Z)$ and the other as χ^2_b. Finally, compute their ratio

$$\frac{\chi^2_a(Z)/a}{\chi^2_b/b} \sim F^{\chi}_{a,b}(c, n). \tag{9}$$

Proposition 5. *Let $\ell_1 = r_1^2/(1 - r_1^2)$, where r_1 is the largest sample canonical correlation between two groups of size $p \le q$ computed from $n + 1$ i.i.d. observations with $\nu = n - p - q > 1$. Then in the presence of a single large population correlation coefficient ρ between the two groups, asymptotically as $\rho \to 1$,*

$$\ell_1 \approx a_1 F^{\chi}_{b_1,c_1}\left(\frac{\rho^2}{1-\rho^2}, 2n\right) + a_2 F_{b_2,c_2} + a_3$$

where

$$\begin{aligned}
&a_1 = q/(\nu + 1), \quad a_2 = (p - 1)/(\nu + 2), \quad a_3 = (p - 1)/\nu(\nu - 1),\\
&b_1 = 2q, \quad b_2 = 2p - 2, \quad c_1 = 2(\nu + 1), \quad c_2 = 2(\nu + 2).
\end{aligned}$$

Remark 5. It can be shown that the probability density of $F^{\chi}_{a,b}(c, n)$ is

$$f_X(x) = \frac{1}{\mathcal{B}\left(\frac{c_1}{2}, \frac{b_1}{2}\right)(1 + c)^{n/2}} \left(\frac{b}{a}\right)^{\frac{b}{2}} \frac{x^{\frac{a}{2}-1}}{\left(x + \frac{b}{a}\right)^{\frac{1}{2}(a+b)}} \cdot {}_2F_1\left(\frac{n}{2}, \frac{1}{2}(a + b); \frac{a}{2}; \frac{xc}{(c + 1)\left(x + \frac{b}{a}\right)}\right)$$

where ${}_2F_1(a, b; c; z)$ is the Gauss hypergeometric function and $\mathcal{B}(p, q)$ is the beta function. This formula is useful for numerical evaluation for small parameter values.

Fig. 4 illustrates the accuracy of Proposition 5. A good match between the theoretical approximation formula and simulation results is clearly visible, particularly at the right tail of the distribution.

3. Distribution of the leading sample eigenvector

Another key quantity of both theoretical and practical importance is the squared dot product between the leading sample eigenvector, denoted $\hat{\mathbf{v}}$, and its corresponding population eigenvector $\mathbf{v}$. Assuming $\|\mathbf{v}\| = \|\hat{\mathbf{v}}\| = 1$,

$$R = |\hat{\mathbf{v}}^\dagger \mathbf{v}|^2. \tag{10}$$

A practical application where it is important to understand the behavior of R under a rank one spike, involves the design of dominant mode rejection (DMR) adaptive beamformers in array processing [42]. The main purpose of this beamformer is to eliminate interferences from undesired directions other than the steering direction. As shown in [43], an important parameter which determines the performance of the DMR scheme is the correlation between the random sample eigenvectors and the unknown population eigenvectors. Specifically, in the presence of a single dominant interferer, the population covariance matrix takes the form of a rank one spiked model [43, Eq. 17], and the effectiveness of the DMR depends on the quantity R. Another application where the quantity R plays a key role is passive radar detection with digital illuminators having several periodic identical pulses [12]. In a sequence of papers [12–14], the authors developed a new framework for passive radar detection based on the leading eigenvector of the sample covariance matrix. This detection scheme outperforms traditional detectors [12]. Motivated by these and other applications, we now develop stochastic approximations to R. For Case 1 of Table 1, we have:

Proposition 6. *Let $H \sim \mathcal{CW}_m(n_H, \lambda \mathbf{v}\mathbf{v}^\dagger + \sigma^2 I)$, with $\|\mathbf{v}\| = 1$ and $\lambda > 0$. Let $\hat{\mathbf{v}}$ be the eigenvector corresponding to the largest eigenvalue of H. Then, with (m, n_H, λ) fixed, for small σ*

$$R \approx \frac{1}{1 + \dfrac{\sigma^2}{\lambda+\sigma^2}\dfrac{B}{A} + \dfrac{2\sigma^4}{\left(\lambda+\sigma^2\right)^2}\dfrac{BC}{A^2}},$$

where $A \sim \chi^2_{2n_H}$, $B \sim \chi^2_{2m-2}$ and $C \sim \chi^2_{2n_H-2}$ are all independent.

The distribution of R in Case 2 of Table 1 is given by the following proposition.

Proposition 7. *Let $H \sim \mathcal{CW}_m(n_H, \sigma^2 I, (\omega/\sigma^2)\mathbf{v}\mathbf{v}^\dagger)$, with $\|\mathbf{v}\| = 1$ and $\omega > 0$. Let $\hat{\mathbf{v}}$ be the eigenvector corresponding to the largest eigenvalue of H. Then, with (m, n_H, ω) fixed, for small σ*

$$R \approx \frac{1}{1 + \dfrac{B}{A_\sigma} + 2\dfrac{BC}{A_\sigma^2}},$$

where $A_\sigma \sim \chi^2_{2n_H}(2\omega/\sigma^2)$, $B \sim \chi^2_{2m-2}$ and $C \sim \chi^2_{2n_H-2}$ are all independent.

Propositions 6 and 7 can be useful to analyze theoretically various DMR and radar detections schemes, and shed light on their dependence on the relevant system parameters.

For the double-matrix Case 3 in Table 1, we have

Proposition 8. *Let $H \sim \mathcal{CW}_m(n_H, \lambda \mathbf{v}\mathbf{v}^\dagger + I)$ and $E \sim \mathcal{CW}_m(n_E, I)$ be independent, with $n_E > m + 1$ and $\|\mathbf{v}\| = 1$. Let $\hat{\mathbf{v}}$ be the eigenvector corresponding to the largest eigenvalue of $E^{-1}H$. Then, with (m, n_H, n_E) fixed, for large λ*

$$R \approx \frac{1}{1 + \dfrac{B}{D}},$$

where $B \sim \chi^2_{2m-2}$ and $D \sim \chi^2_{2n_E+4-2m}$ are independent.

In the context of array processing, the double matrix Case 3 of Table 1 corresponds to a setting where the noise characteristics of the m sensors are not perfectly known, but rather their covariance matrix is estimated from n_E samples that do not contain any signal. Comparing Proposition 8 with Proposition 6 sheds light on the effect of estimating the covariance matrix of the noise. Whereas in Case 1, as signal strength $\lambda \to \infty$ the quantity R converges to one, in Case 3, the random variable R does not converge to one, but rather to a Beta distribution.

Figs. 5 and 6 illustrate the accuracy of our approximate distributions of the squared inner product between the leading sample and population eigenvectors.

4. Applications

We now demonstrate the utility of our approximations to Roy's largest root distribution under a rank-one perturbation in three different engineering applications. The first two are concerned with common problems in signal detection, whereas the third with the outage probability of a rank-one Rician fading MIMO channel.

4.1. Signal detection in noise

Detecting the presence of a signal in a noisy environment is a fundamental problem in detection theory. Specific examples include spectrum sensing in cognitive radio [17] and target detection in sonar and radar [42]. Assuming additive Gaussian noise, the observed vector $\mathbf{y}(t) \in \mathbb{C}^m$ at time t is of the form

$$\mathbf{y}(t) = \sqrt{\lambda} s(t)\mathbf{u} + \mathbf{n}(t), \tag{11}$$

where $s(t) \in \mathbb{C}$ is the time dependent signal, $\mathbf{u} \in \mathbb{C}^m$ is normalized such that $\|\mathbf{u}\| = 1$ is its direction, $\lambda \geq 0$ is a measure of the signal strength and the vector $\mathbf{n} \in \mathbb{C}^m$ is a zero mean complex valued random noise, assumed to be independent of

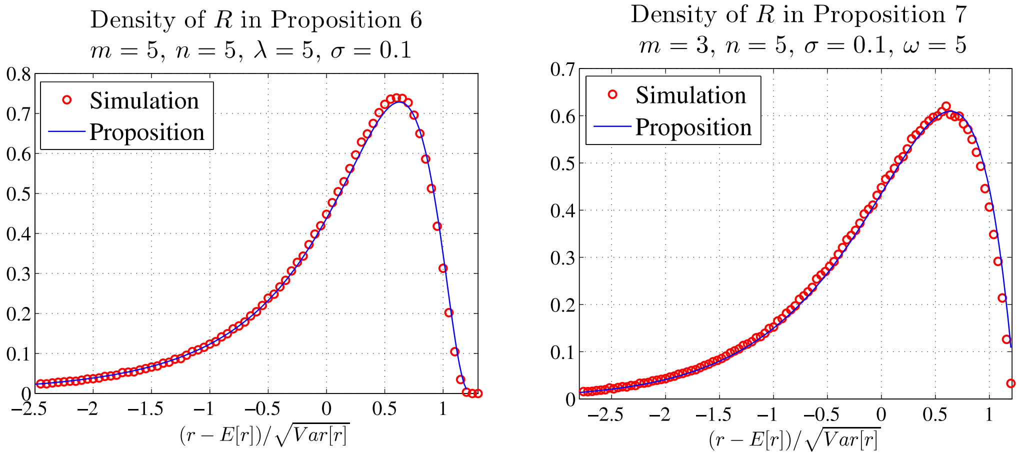

Fig. 5. Empirical versus theoretical density of R in Case 1 (left) and Case 2 (right).

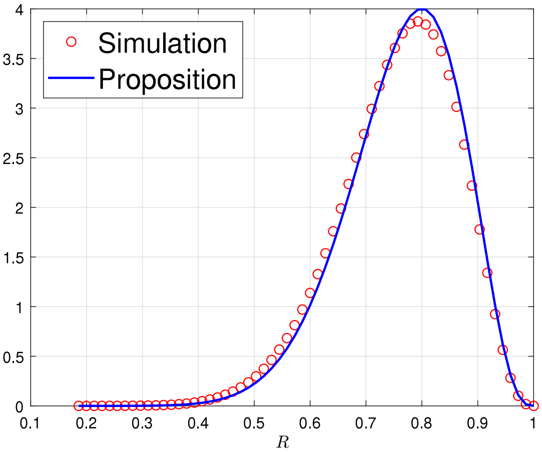

Fig. 6. Comparison of empirical density of R in Case 3 of Table 1 with Proposition 8, for $n_H = 10$, $n_E = 16$, $m = 5$ and $\lambda = 100$.

the signal and distributed as $\mathbf{n} \sim \mathcal{CN}(0, \Sigma)$. The positive definite Hermitian matrix Σ is thus the population covariance of the additive random noise. In some cases it is assumed to be explicitly known, whereas in others it needs to be estimated. The signal $s(t)$ is often modeled as a random quantity with $\mathbb{E}\{|s(t)|^2\} = 1$. For example, in multiple antenna spectrum sensing for cognitive radio a common model is that $s(t) \sim \mathcal{CN}(0, 1)$, namely $s(t) = s_1(t) + \imath s_2(t)$ where $s_1(t)$ and $s_2(t)$ are real valued and independent random variables distributed $\mathcal{N}(0, 1)$ [45,46]. Similarly, in detection of constant modulus signals (e.g., FM signals [18]), $s(t) = \exp(\imath\phi(t))$, where $\phi(t)$ is random.

When the covariance matrix Σ of the noise vector $\mathbf{n}$ is assumed known, the observed data used to detect if a signal is present are often n_H i.i.d. observations $\mathbf{y}_1, \ldots, \mathbf{y}_{n_H}$, from (11). A popular approach is to compute the sample covariance matrix $H = \sum_{j=1}^{n_H} \mathbf{y}_j \mathbf{y}_j^\dagger$, and declare that a signal is present if some function of its eigenvalues is larger than a suitable threshold. Several such detection tests have been proposed [18,45,46], including Roy's largest root [29]. As discussed below, depending on the model of the signal, this leads precisely to Cases 1 and 2 in Table 1.

In other situations, Σ is unknown, but it is possible to observe both the n_H samples $\mathbf{y}_i$ of (11) as well as an additional set of n_E independent realizations $\mathbf{n}_1, \ldots, \mathbf{n}_{n_E}$ of the noise vector $\mathbf{n}$. The latter are measured, for example, in time slots at which it is a-priori known that no signals are emitted. Here, a typical approach is to form both the matrix H as above and the matrix $E = \Sigma_{j=1}^{n_E} \mathbf{n}_j \mathbf{n}_j^\dagger$ and detect the presence of a signal via some function of the eigenvalues of $E^{-1}H$. Signal detection based on the largest eigenvalue of $E^{-1}H$ leads to Cases 3 and 4 in Table 1.

As discussed in Section 2, one may assume without loss of generality that $\Sigma = \sigma^2 I$. Thus, when $s \sim \mathcal{CN}(0, 1)$,

$$H \sim \mathcal{CW}_m(n_H, \lambda \mathbf{u}\mathbf{u}^\dagger + \sigma^2 I).$$

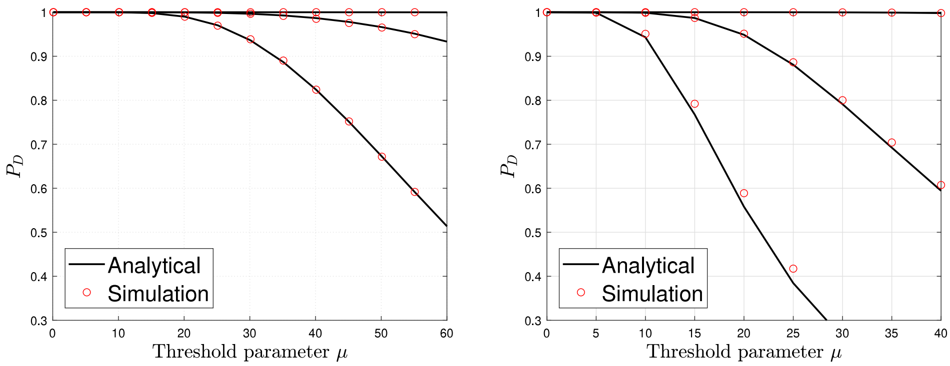

Fig. 7. Detection power profile for several signal to noise ratios as a function of threshold μ for a known covariance matrix or an unknown covariable. In both cases $\lambda = 1$, $n_H = 5$, $m = 5$. In the right panel $n_E = 10$. From top to bottom, $\sigma = 1/10, 2/10$ and $3/10$.

In contrast, if $s = \exp(\iota\phi)$, conditional on $\phi_1, \ldots, \phi_{n_H}$,

$$H \sim \mathcal{CW}_m\left(n_H, \sigma^2 I, \frac{\lambda n_H}{\sigma^2}\mathbf{u}\mathbf{u}^\dagger\right).$$

Propositions 1–4 can thus be used to approximate the detection power of Roy's largest root test as a function of signal strength λ in both the single matrix cases and the double matrix cases,

$$P_D = \Pr\left\{\ell_1 > \mu \mid \text{signal present with strength } \lambda\right\}, \tag{12}$$

where μ is a given threshold parameter. The accuracy of (12) is illustrated in Fig. 7.

4.2. Rank-one Rician-fading MIMO channel

As a last application, consider the outage probability of a MIMO communication channel with n_T transmitters and n_R receivers. Here, the transmitted signals $\mathbf{x} \in \mathbb{C}^{n_T}$ and received signals $\mathbf{y} \in \mathbb{C}^{n_R}$ are related as

$$\mathbf{y} = H\mathbf{x} + \mathbf{n}$$

where H is the $n_R \times n_T$ channel matrix and $\mathbf{n}$ is additive random complex valued noise, assumed to be distributed as $\mathbf{n} \sim \mathcal{CN}\left(n, \sigma_n^2 I\right)$, where σ_n^2 is its (real-valued) variance. Due to fluctuations in the environment, the channel matrix H is modeled as a random quantity. In particular, under a common Rician fading model [16], H has the form

$$H = \sqrt{\frac{K}{K+1}}H_1 + \sqrt{\frac{1}{K+1}}H_2 \tag{13}$$

where H_1 represents the specular (Rician) component from a direct line-of-sight between transmitter and receiver antennas and H_2 represents the scattered Rayleigh-fading component. With fixed sender and receiver locations, the matrix H_1 is constant whereas H_2 is random with entries modeled as i.i.d. complex Gaussians, $\mathcal{CN}(0, \sigma_H^2)$. Under the normalization $\operatorname{tr}(H_1 H_1^\dagger) = n_R n_T$, the factor K represents the ratio of deterministic-to-scattered power of the environment.

Under the maximal ratio transmission strategy, where the transmitter sends information along the leading eigenvector of $HH^\dagger$, the channel signal to noise ratio is given by

$$\mu = \frac{\Omega_D}{\sigma_n^2}\ell_1\left(HH^\dagger\right) \tag{14}$$

where $\Omega_D = \mathbb{E}[\|\mathbf{x}\|^2]$ is the power of the transmitted signal vectors [24]. An important quantity is the channel's outage probability, defined as the probability of failing to achieve a specified minimal SNR $\mu_{\min}$ required for satisfactory reception. Based on (14), the outage probability P_{out} can be written as

$$P_{\text{out}} = \Pr\left(\frac{\Omega_D}{\sigma_n^2}\ell_1 \le \mu_{\min}\right). \tag{15}$$

One particularly interesting case is when the Rician component H_1 is assumed to be of rank one, $H_1 = \mathbf{u}\mathbf{v}^\dagger$, where $\mathbf{u} \in \mathbb{C}^{n_R}$, $\mathbf{v} \in \mathbb{C}^{n_T}$. An important design question is which configuration of antennas minimizes (15), under the constraint that the total number of transmitting and receiving antennas is fixed. Via simulations, [24] showed it is best to have an equal number of transmitting and receiving antennas. Here we analytically prove this result asymptotically in the limit of small scattering variance (i.e., $\sigma_H \ll 1$).

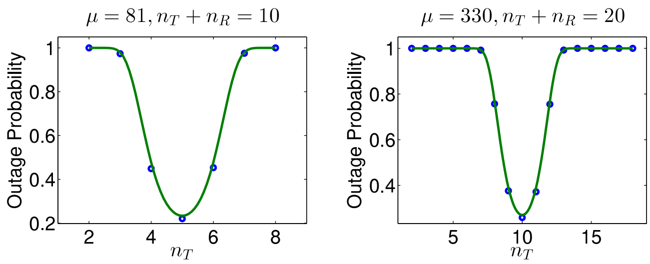

Fig. 8. Outage probability as a function of n_T, with n_T+n_R fixed. Circles represent a Monte-Carlo simulation whereas the solid line is our approximation (which can be computed for any non-integer $n_T \in \mathbb{R}^+$). These graphs support Proposition 9 and demonstrate the accuracy of our approximations. In both graphs, $K = 2, \sigma_H = 0.3, \sigma_n = 1$ and $\Omega_D = 5$.

Proposition 9. *Consider a rank-one Rician fading channel with a fixed number of antennas, $n_T + n_R = N$. Then, for $\sigma_H \ll 1$, the outage probability is minimized at $n_T = n_R = N/2$ for N even (or say $n_T = \lfloor N/2 \rfloor, n_R = \lceil N/2 \rceil$ for N odd).*

Proof. Under the model in (13) and the assumption that $H_1 = \mathbf{u}\mathbf{v}^\dagger$ is rank one, the jth column of H, of dimension n_R, is distributed as $\mathcal{CN}\left(\sqrt{K/(K+1)}\mathbf{u}v_j, \sigma_H^2/(K+1)I_{n_R}\right)$. Therefore,

$$HH^\dagger \sim \mathcal{CW}_{n_R}(n_T, \alpha^2 I_{n_R}, \beta^2/\alpha^2 \mathbf{w}\mathbf{w}^\dagger)$$

is non-central Wishart, with

$$\mathbf{w} = \mathbf{v}/\|\mathbf{v}\|, \quad \alpha^2 = \frac{1}{K+1}\sigma_H^2 \quad \text{and} \quad \beta^2 = \frac{K}{K+1}\|\mathbf{u}\|^2\|\mathbf{v}\|^2 = \frac{K}{K+1}n_R n_T.$$

Thus, Proposition 2 implies that for fixed (n_T, n_R, K),

$$\mu = \frac{\Omega_D}{\sigma_n^2}\ell_1 = c_1\left(A + B + \frac{BC}{A}\right) + o_P(\sigma_H^4) \tag{16}$$

where A, B, C are independent random variables distributed as

$$A \sim \chi^2_{2n_T}(c_2), \qquad B \sim \chi^2_{2n_R-2}, \qquad C \sim \chi^2_{2n_T-2}$$

and

$$c_1 = \frac{\Omega_D \sigma_H^2}{2(K+1)\sigma_n^2}, \qquad c_2 = \frac{2\beta^2}{\alpha^2} = \frac{2K}{\sigma_H^2}n_R n_T. \tag{17}$$

Since $\mathbb{E}(A) = 2n_T + c_2 \gg 1$, and $c_2 \to \infty$ as $\sigma_H \to 0$, we may neglect the third term in (16). Furthermore, since A and B are independent,

$$\mu \approx c_1(A+B) = c_1(\chi^2_{2n_T}(c_2) + \chi^2_{2n_R-2}) = c_1\chi^2_{2N-2}(c_2).$$

Clearly P_{out} of (15) is minimal when the largest eigenvalue ℓ_1 is stochastically as large as possible, or in turn, when its non-centrality parameter c_2 is maximal. Since by (17), $c_2 \propto n_T n_R$, the proposition follows (see Fig. 8). $\square$

Acknowledgments

We thank the editor in chief, the associate editor and the referees for their constructive comments and suggestions. This work was supported in part by grants NIH BIB R01EB1988 (PD) and BSF 2012-159 (PD, BN, OS). B.N. is incumbent of the William Petschek professorial chair of mathematics.

Appendix A. Proofs of main propositions

We prove our main results using the analytical framework developed in [23]. For a complex-valued number $z \in \mathbb{C}$, its real and imaginary parts are denoted $\Re(z)$ and $\Im(z)$, respectively, whereas $\bar{z}$ is its complex conjugate. We begin with the following auxiliary lemma, which describes the analytic structure of the leading eigenvalue and eigenvector of a covariance matrix constructed from vectors all in the same direction, which without loss of generality we choose as the standard vector $\mathbf{e}_1 = (1, 0, \ldots, 0)^\top$, but corrupted by small perturbations. Its proof is in Appendix B.

Lemma 1. *Let $\{\mathbf{x}_j\}_{j=1}^n$ be n vectors in $\mathbb{C}^m$ of the form*

$$\mathbf{x}_j = u_j \mathbf{e}_1 + \epsilon \xi_j^{\perp} \tag{A.1}$$

where u_j are complex valued scalars, $\xi_j^{\perp} = \begin{pmatrix} 0 \\ \xi_j \end{pmatrix}$ with $\xi_j \in \mathbb{C}^{m-1}$ are the perturbations in orthogonal directions to $\mathbf{e}_1$ and $\epsilon \in \mathbb{R}$ is a small parameter. Define $z \in \mathbb{R}$, $b \in \mathbb{C}^{m-1}$ and $Z \in C^{(m-1)\times(m-1)}$

$$z = \sum_{j=1}^n u_j \bar{u}_j, \quad b = z^{-\frac{1}{2}} \sum_{j=1}^n \bar{u}_j \xi_j, \quad Z = \sum_{j=1}^n \xi_j \xi_j^{\dagger}. \tag{A.2}$$

Let $\ell_1(\epsilon)$ be the largest eigenvalue of $H(\epsilon) = \sum_{j=1}^n \mathbf{x}_j \mathbf{x}_j^{\dagger}$ with corresponding leading eigenvector $v_1(\epsilon)$ normalized such that $\mathbf{e}_1^{\dagger} v_1(\epsilon) = 1$. Then $\ell_1(\epsilon)$ is an even analytic function of ϵ, whereas $v_1(\epsilon) - \mathbf{e}_1$ is an odd function of ϵ. In particular, the Taylor expansions of $\ell_1(\epsilon)$ and $v_1(\epsilon)$ around $\epsilon = 0$ are given by

$$\ell_1(\epsilon) = z + \|b\|^2 \epsilon^2 + z^{-1} b^{\dagger}(Z - bb^{\dagger}) b \epsilon^4 + \cdots \tag{A.3}$$

$$v_1(\epsilon) = \mathbf{e}_1 + z^{-1/2} \begin{pmatrix} 0 \\ b \end{pmatrix} \epsilon + z^{-3/2} \begin{pmatrix} 0 \\ Zb - \|b\|^2 b \end{pmatrix} \epsilon^3 + \cdots$$

Proof of Propositions 1 and 2. Since the eigenvalues of H do not depend on the direction of the vector $\mathbf{v}$, without loss of generality we thus assume that $\mathbf{v} = \mathbf{e}_1$. Then, H may be realized from n_H i.i.d. observations of the form (A.1) with ϵ replaced by σ,

$$\xi_j \sim \mathcal{CN}(0, I_{m-1}), \qquad u_j \sim \begin{cases} \mathcal{CN}(0, \sigma^2 + \lambda), & \text{Proposition 1}, \\ \mathcal{CN}(\mu_j, \sigma^2), & \text{Proposition 2}, \end{cases} \tag{A.4}$$

and μ_j are arbitrary complex numbers satisfying $\sum_j |\mu_j|^2 = \omega$.

For each realization of $u = (u_k)$ and $\Xi = [\xi_1, \ldots, \xi_{n_H}] \in \mathbb{C}^{(m-1)\times n_H}$, Lemma 1 yields the approximation (A.3) for $\ell_1(\sigma)$. To derive the distributions of the various terms in (A.3) we proceed as follows. Define $o_1 = \bar{u}/\|u\| \in \mathbb{C}^{n_H}$, choose columns $o_2, \ldots, o_n$ so that $O = [o_1, \ldots, o_{n_H}]$ is an $n_H \times n_H$ unitary matrix, and consider the following $(m-1) \times n_H$ matrix $V = \Xi O$. Its first column is $v_1 = \Xi \bar{u}/\|u\| = b$, and thus the $O(\epsilon^2)$ term in (A.3) is $b^{\dagger} b = \|v_1\|^2$. For the fourth order term, observe that $Z = \Xi \Xi^{\dagger} = VV^{\dagger}$ and so the quantity $D = b^{\dagger}(Z - bb^{\dagger})b$ may be written as

$$D = v_1^{\dagger}(VV^{\dagger} - v_1 v_1^{\dagger}) v_1 = (v_1^{\dagger} V)(v_1^{\dagger} V)^{\dagger} - (v_1^{\dagger} v_1)(v_1^{\dagger} v_1)^{\dagger} = \sum_{j=2}^{n_H} |v_1^{\dagger} v_j|^2.$$

Hence, (A.3) becomes

$$\ell_1(\epsilon) = V_0 + V_2 \epsilon^2 + V_4 \epsilon^4 + \cdots$$

where $V_0 = \|u\|^2$, $V_2 = \|v_1\|^2$ and $V_4 = V_0^{-1} D$. To study the distributions of V_0, V_2, V_4, note that by assumption in (A.4), $u_j = (a_j + \imath b_j)/\sqrt{2}$ with

$$a_j \sim \begin{cases} \mathcal{N}(0, \lambda + \sigma^2) & \text{Proposition 1} \\ \mathcal{N}(\sqrt{2}\Re(\mu_j), \sigma^2) & \text{Proposition 2} \end{cases} \qquad b_j \sim \begin{cases} \mathcal{N}(0, \lambda + \sigma^2) & \text{Proposition 1} \\ \mathcal{N}(\sqrt{2}\Im(\mu_j), \sigma^2) & \text{Proposition 2}. \end{cases}$$

Therefore, $\|u\|^2 = \frac{1}{2} \sum_{j=1}^{n_H} (a_j^2 + b_j^2)$ is a sum of $2n_H$ independent squares of either mean centered or non-centered Gaussian random variables. This in turn gives

$$V_0 = \|u\|^2 \sim \begin{cases} \frac{\sigma^2 + \lambda}{2} \chi^2_{2n_H}, & \text{Proposition 1}, \\ \frac{\sigma^2}{2} \chi^2_{2n_H}\left(\frac{2\omega}{\sigma^2}\right), & \text{Proposition 2}. \end{cases}$$

Since given u, O is unitary and fixed, then $v_j | u \sim \mathcal{CN}(0, I_{m-1})$. Since this distribution is independent of u, $v_j \sim \mathcal{CN}(0, I_{m-1})$. By similar arguments

$$V_2 = \|v_1\|^2 \sim \frac{1}{2} \chi^2_{2m-2}$$

which is independent of $\|u\|^2$. Finally, conditioned on (u, v_1), we have $v_1^\dagger v_j \sim \mathcal{CN}(0, \|v_1\|^2)$ and $|v_1^\dagger v_j|^2 \sim \|v_1\|^2\chi_2^2/2$. Thus,

$$D|(u, v_1) = \sum_{j=2}^{n_H} |v_1^\dagger v_j|^2 |(u, v_1) \sim \frac{\|v_1\|^2}{2}\chi^2_{2n_H-2},$$

where the $\chi^2_{2n_H-2}$ variate is independent of (u, v_1). We conclude that

$$V_4 \sim \begin{cases} \frac{1}{2\sigma^2+2\lambda}(\chi^2_{2n_H})^{-1}\chi^2_{2m-2}\chi^2_{2n_H-2}, & \text{Proposition 1},\\ \frac{1}{2\sigma^2}(\chi^2_{2n_H}(\frac{2\omega}{\sigma^2}))^{-1}\chi^2_{2m-2}\chi^2_{2n_H-2}, & \text{Proposition 2}.\end{cases}$$

Since the random variables V_0, V_2, V_4 are independent, then so are A, B, C in either (1) or (2). This completes the proof of Propositions 1 and 2. □

To prove Propositions 3 and 4, we first introduce some additional notation and two auxiliary lemmas, whose proofs are deferred to Appendix B. For a matrix S, denote by S_{jk} and S^{jk} the (j, k)th entries of S and S^{-1}, respectively.

Lemma 2. *Let $E \sim \mathcal{CW}_m(n_E, I)$ and $M = [\mathbf{e}_1, b] \in \mathbb{C}^{m\times 2}$, with the vector b fixed and orthogonal to $\mathbf{e}_1$. Define a 2×2 diagonal matrix $D = diag\left(1, 1/\|b\|^2\right)$. Then*

$$S = (M^\dagger E^{-1}M)^{-1} \sim \mathcal{CW}_2(n_E - m + 2, D),$$

and the two random variables S^{11} and S_{22} are independent with

$$S^{11} \sim \frac{2}{\chi^2_{2n_E-2m+2}}, \qquad S_{22} \sim \frac{\chi^2_{2n_E-2m+4}}{2\|b\|^2}.$$

Lemma 3. *Let $E \sim \mathcal{CW}_m(n_E, I)$ and let $A_2 = \begin{pmatrix} 0 & 0\\ 0 & Z\end{pmatrix}$ where Z is an $(m-1)\times(m-1)$ random matrix independent of E, with $\mathbb{E}(Z) = I_{m-1}$. Then*

$$\mathbb{E}\left(\frac{\mathbf{e}_1^\dagger E^{-1}A_2E^{-1}\mathbf{e}_1}{E^{11}}\right) = \frac{m-1}{(n_E-m)(n_E-m+1)}.$$

Proof of Propositions 3 and 4. Without loss of generality we may assume that the signal direction is $\mathbf{v} = \mathbf{e}_1$. Hence

$$H \sim \begin{cases}\mathcal{CW}_m(n_H, I + \lambda\mathbf{e}_1\mathbf{e}_1^\dagger), & \text{Proposition 3},\\ \mathcal{CW}_m(n_H, I, \omega\mathbf{e}_1\mathbf{e}_1^\dagger), & \text{Proposition 4}.\end{cases}$$

Next, we apply a perturbation approach similar to the one used in the previous proof. To introduce a small parameter, set

$$\epsilon^2 = \begin{cases} 1/(1+\lambda), & \text{Proposition 3},\\ 1/\omega, & \text{Proposition 4}.\end{cases}$$

The matrix $H_\epsilon = \epsilon^2 H$ has a representation of the form $X^\dagger X$ with $X = [\mathbf{x}_1, \ldots, \mathbf{x}_{n_H}]$ where each $\mathbf{x}_j$ follows (A.1) but now with

$$\xi_j \sim \mathcal{CN}(0, I_{m-1}), \qquad u_j \sim \begin{cases}\mathcal{CN}(0, 1), & \text{Proposition 3},\\ \mathcal{CN}(\mu_j/\sqrt{\omega}, 1/\omega), & \text{Proposition 4},\end{cases}$$

where $\sum |\mu_j|^2 = \omega$. In particular,

$$z = \sum_{j=1}^{n_H} |u_j|^2 \sim \begin{cases}\frac{1}{2}\chi^2_{2n_H}, & \text{Proposition 3},\\ \frac{1}{2\omega}\chi^2_{2n_H}(2\omega), & \text{Proposition 4}.\end{cases}$$

With b as in (A.2), using the same arguments as in the previous proof, we have that $b \sim \mathcal{CN}(0, I_{m-1})$, independently of u.

The matrix H_ϵ may be written as $H_\epsilon = A_0 + \epsilon A_1 + \epsilon^2 A_2$, where

$$A_0 = \begin{pmatrix} z & 0\\ 0 & 0_{m-1}\end{pmatrix}, \qquad A_1 = \sqrt{z}\begin{pmatrix} 0 & b^\dagger\\ b & 0_{m-1}\end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & 0\\ 0 & Z\end{pmatrix} \tag{A.5}$$

with Z as in (A.2). For future use we define the following quantities

$$E^{11} = \mathbf{e}_1^\dagger E^{-1}\mathbf{e}_1, \quad \hat{b} = \begin{pmatrix}0\\ b\end{pmatrix}, \quad E^{b1} = \hat{b}^\dagger E^{-1}\mathbf{e}_1, \quad E^{bb} = \hat{b}^\dagger E^{-1}\hat{b}.$$

Note that the condition $n_E \geq m$ ensures that E is invertible with probability 1. This follows for example from Theorem 3.2 in [9].

The matrix $E^{-1}H_\epsilon$ is similar to the Hermitian matrix $E^{-1/2}H_\epsilon E^{-1/2}$. Therefore, all its eigenvalues are real-valued for any value of ϵ. Furthermore, since $E^{-1/2}H_\epsilon E^{-1/2}$ is a holomorphic symmetric function of ϵ, it follows from Kato ([25], Theorem 6.1 page 120) that the largest eigenvalue ℓ_1 and its eigenprojection $\tilde{P}(\epsilon)$ are analytic functions of ϵ in some neighborhood of zero, where the largest eigenvalue has multiplicity one. The projection to the corresponding eigenspace of $E^{-1}H_\epsilon$ is $P(\epsilon) = E^{-1/2}\tilde{P}(\epsilon)$. As the matrix E does not depend on ϵ, this projection is also an analytic function in some neighborhood of $\epsilon = 0$.

At $\epsilon = 0$, $E^{-1}\mathbf{e}_1$ is an eigenvector with eigenvalue $E^{11}z$, that is,

$$E^{-1}H_0E^{-1}\mathbf{e}_1 = zE^{-1}\mathbf{e}_1\mathbf{e}_1^\dagger E^{-1}\mathbf{e}_1 = zE^{11}E^{-1}\mathbf{e}_1,$$

from which we obtain

$$\mathbf{e}_1^\dagger \tilde{P}(0)E^{-1}\mathbf{e}_1 = \mathbf{e}_1^\dagger E^{-1}\mathbf{e}_1 = E^{11}. \tag{A.6}$$

Since $\tilde{P}(\epsilon)$ is an analytic function of ϵ and the inner product is a smooth function, then there exists a neighborhood of $\epsilon = 0$ where $\mathbf{e}_1^\dagger \tilde{P}(\epsilon)E^{-1}\mathbf{e}_1, \mathbf{e}_1$ is both analytic in ϵ and strictly positive. In this neighborhood, we may define

$$v_1(\epsilon) = \frac{E^{11}}{\mathbf{e}_1^\dagger \tilde{P}(\epsilon)E^{-1}\mathbf{e}_1}\tilde{P}(\epsilon)E^{-1}\mathbf{e}_1. \tag{A.7}$$

Clearly $v_1(\epsilon)$ is the eigenvector corresponding to the eigenvalue $\ell_1(\epsilon)$ and it is also analytic. We thus expand

$$\ell_1(\epsilon) = \sum_{j=0}^{\infty} \lambda_j \epsilon^j, \qquad v_1(\epsilon) = \sum_{j=0}^{\infty} w_j \epsilon^j. \tag{A.8}$$

Inserting these expansions into the eigenvalue–eigenvector equations $E^{-1}H_\epsilon v_1 = \ell_1 v_1$ gives the following equations: at the $O(1)$ level,

$$E^{-1}A_0w_0 = \lambda_0 w_0$$

whose solution is

$$\lambda_0 = zE^{11}, \qquad w_0 = \text{const} \cdot E^{-1}\mathbf{e}_1. \tag{A.9}$$

By (A.6)–(A.7), $w_0 = v_1(0) = E^{-1}\mathbf{e}_1$, so the above constant is one.

By (A.7), $\mathbf{e}_1^\dagger v_1(\epsilon) = E^{11} = \mathbf{e}_1^\dagger w_0$. Hence $\mathbf{e}_1^\dagger w_j = 0$ for all $j \geq 1$. Furthermore, since $A_0 = z\mathbf{e}_1\mathbf{e}_1^\dagger$, then $A_0 w_j = 0$ for all $j \geq 1$. The $O(\epsilon)$ equation is thus

$$E^{-1}A_1w_0 + E^{-1}A_0w_1 = \lambda_1 w_0 + \lambda_0 w_1. \tag{A.10}$$

However, $A_0w_1 = 0$. Multiplying this equation by $\mathbf{e}_1^\dagger$ gives that

$$\begin{aligned}\lambda_1 &= \frac{\mathbf{e}_1^\dagger E^{-1}w_0}{E^{11}} = \frac{\sqrt{z}}{E^{11}}\left\{\mathbf{e}_1^\dagger E^{-1}\begin{pmatrix}0 & b^\dagger\\ b & 0\end{pmatrix}E^{-1}\mathbf{e}_1\right\}\\ &= \frac{\sqrt{z}}{E^{11}}\left\{\mathbf{e}_1^\dagger E^{-1}\begin{pmatrix}0 & b^\dagger\\ 0 & 0\end{pmatrix}E^{-1}\mathbf{e}_1 + \mathbf{e}_1^\dagger E^{-1}\begin{pmatrix}0 & 0\\ b & 0\end{pmatrix}E^{-1}\mathbf{e}_1\right\} = 2\sqrt{z}\Re(E^{b1}).\end{aligned} \tag{A.11}$$

Inserting the expression for λ_1 into (A.10) gives that

$$\begin{aligned}w_1 &= \frac{1}{\sqrt{z}E^{11}}\left\{E^{-1}\begin{pmatrix}0 & b^\dagger\\ b & 0\end{pmatrix}E^{-1}\mathbf{e}_1 - 2\Re(E^{b1})E^{-1}\mathbf{e}_1\right\}\\ &= \frac{1}{\sqrt{z}E^{11}}\left(E^{b1}E^{-1}\mathbf{e}_1 + E^{11}E^{-1}\hat{b} - 2\Re(E^{b1})E^{-1}\mathbf{e}_1\right) = \frac{1}{\sqrt{z}}\left(E^{-1}\hat{b} - \frac{\overline{E^{b1}}}{E^{11}}E^{-1}\mathbf{e}_1\right).\end{aligned}$$

The next $O(\epsilon^2)$ equation is

$$E^{-1}A_2w_0 + E^{-1}A_1w_1 + E^{-1}A_0w_2 = \lambda_2 w_0 + \lambda_1 w_1 + \lambda_0 w_2.$$

Multiply this equation by $\mathbf{e}_1^\dagger$ and recall that $A_0w_2 = 0$ and $\mathbf{e}_1^\dagger w_0 = E^{11}$ gives

$$\begin{aligned}\lambda_2 &= \frac{\mathbf{e}_1^\dagger E^{-1}A_2E^{-1}\mathbf{e}_1}{E^{11}} + \frac{\mathbf{e}_1^\dagger E^{-1}A_1\frac{1}{\sqrt{z}}(E^{-1}\hat{b} - \frac{\overline{E^{b1}}}{E^{11}}E^{-1}\mathbf{e}_1)}{E^{11}}\\ &= \frac{\mathbf{e}_1^\dagger E^{-1}A_2E^{-1}\mathbf{e}_1}{E^{11}} + \frac{E^{11}E^{bb} + (\overline{E^{b1}})^2 - 2\overline{E^{b1}}\Re(E^{b1})}{E^{11}} = \frac{\mathbf{e}_1^\dagger E^{-1}A_2E^{-1}\mathbf{e}_1}{E^{11}} + \frac{E^{11}E^{bb} - E^{b1}\overline{E^{b1}}}{E^{11}}.\end{aligned} \tag{A.12}$$

Combining (A.9)–(A.12), we obtain the following approximate stochastic representation for the largest eigenvalue ℓ_1 of $E^{-1}H_\epsilon$

$$\ell_1(\epsilon)=zE^{11}+2\epsilon\sqrt{z}\Re(E^{b1})+\epsilon^2\frac{\mathbf{e}_1^\dagger E^{-1}A_2E^{-1}\mathbf{e}_1}{E^{11}}+\epsilon^2\frac{E^{11}E^{bb}-E^{b1}\overline{E^{b1}}}{E^{11}}+O_P(\epsilon^3). \tag{A.13}$$

Next, to derive the approximate distribution of ℓ_1 corresponding to the above equation, we study a 2×2 Hermitian matrix S, whose inverse is defined by

$$S^{-1}=\begin{pmatrix} E^{11} & \overline{E^{b1}} \\ E^{b1} & E^{bb}\end{pmatrix}=M^\dagger E^{-1}M,$$

where $M=[\mathbf{e}_1,\hat{b}]$ is a 2×2 matrix. Inverting this matrix gives

$$S=\frac{1}{E^{11}E^{bb}-E^{b1}\overline{E^{b1}}}\begin{pmatrix} E^{bb} & -E^{b1} \\ -\overline{E^{b1}} & E^{11}\end{pmatrix}.$$

Hence in terms of the matrices S and S^{-1}, (A.13) can be written as

$$\ell_1(\epsilon)=zS^{11}+2\epsilon\sqrt{z}\Re(E^{b1})+\frac{\epsilon^2}{S_{22}}+\epsilon^2\frac{\mathbf{e}_1^\dagger E^{-1}A_2E^{-1}\mathbf{e}_1}{E^{11}}+O_P(\epsilon^3). \tag{A.14}$$

To establish Propositions 3 and 4, we start from (A.14). We neglect the second term $T_1=2\epsilon\sqrt{z}\Re(E^{b1})$ which is symmetric with mean zero, and whose variance is much smaller than that of the first term. We also approximate the last term, denoted by T_2, by its mean value, using Lemma 3. We now have

$$\ell_1(\epsilon)\approx zS^{11}+\epsilon^2\left\{\frac{1}{S_{22}}+c(m,n_E)\right\},$$

where $c(m,n)$ is the expectation from Lemma 3. Since $\ell(\epsilon)$ is the largest eigenvalue of $E^{-1}H_\epsilon=\epsilon^2E^{-1}H$, (A.14) should be divided by ϵ^2 to obtain the largest eigenvalue of $E^{-1}H$. By doing so, and inserting the distributions of S^{11} and combining this with the S_{22} from Lemma 2 gives

$$\ell_1\approx\frac{2z}{\epsilon^2\chi^2_{2n_E-2m+2}}+\frac{2\|b\|^2}{\chi^2_{2n_E-2m+4}}+\frac{m-1}{(n_E-m)(n_E-m-1)}.$$

Next, by inserting the distributions of $\|b\|^2$, z and the relevant value of ϵ, we get that for Proposition 3

$$\ell_1(\lambda)\approx(1+\lambda)\frac{\chi^2_{2n_H}}{\chi^2_{2n_E-2m+2}}+\frac{\chi^2_{2m-2}}{\chi^2_{2n_E-2m+4}}+\frac{m-1}{(n_E-m)(n_E-m-1)}$$

and for Proposition 4

$$\ell_1(\omega)\approx\frac{\chi^2_{2n_H}(2\omega)}{\chi^2_{2n_E-2m+2}}+\frac{\chi^2_{2m-2}}{\chi^2_{2n_E-2m+4}}+\frac{m-1}{(n_E-m)(n_E-m-1)}.$$

From Lemma 2 and the independency of u and z, all of the above χ^2 random variables are independent. Finally, since ratios of independent χ^2 random variables follow an F distribution, the two propositions follow. □

Proof of Proposition 5. By (8), the non-centrality parameter ω depends on the data only through $X^\dagger X$. Conditioning on $X^\dagger X$, following (7), we invoke Proposition 4 with the parameters $m=p$, $n_H=q$, and $n_E=n-q$ to obtain

$$\ell_1\left(E^{-1}H\right)|X\approx a_1F_{b_1,c_1}\left(\frac{2\rho^2}{1-\rho^2}\left(X^\dagger X\right)_{11}\right)+a_2F_{b_2,c_2}+a_3.$$

Now the final result follows by integrating over the distribution of $\left(X^\dagger X\right)_{11}\sim\frac{1}{2}\chi^2_{2n}$, and using the definition of $F^{\chi}_{a,b}(c,n)$ given in (9). □

Proof of Propositions 6 and 7. Let us assume without loss of generality that $\mathbf{v}=\mathbf{e}_1$. If $\hat{\mathbf{v}}$ is not normalized, then we can write (10) as $R=|\hat{\mathbf{v}}^\dagger\mathbf{e}_1|^2/\|\hat{\mathbf{v}}\|^2$. From Lemma 1, we have

$$\hat{\mathbf{v}}=w_0+\sigma w_1+\sigma^3w_3+\cdots$$

where

$$w_0=\mathbf{e}_1,\quad w_1=\frac{1}{\|u\|}\begin{pmatrix}0\\ v_1\end{pmatrix},\quad w_3=\frac{1}{\|u\|^3}\begin{pmatrix}0\\ \sum_{j=2}^{n_H}v_jv_j^\dagger v_1\end{pmatrix},$$

with i.i.d. variables $v_j \sim \mathcal{CN}(0, I_{m-1})$ all independent of $u \in \mathbb{C}^n$,

$$u \sim \begin{cases} \mathcal{CN}(0, (\lambda+\sigma^2)I_n) & \text{for } H \sim \mathcal{CW}_m(n_H, \lambda \mathbf{e}_1\mathbf{e}_1^\dagger + \sigma^2 I_m) \\ \mathcal{CN}(\mu, \sigma^2 I_n) & \text{for } H \sim \mathcal{CW}_m(n_H, \sigma^2 I_m, (\omega/\sigma^2)\mathbf{e}_1\mathbf{e}_1^\dagger) \end{cases}$$

and $\|\mu\|^2 = \omega$. Therefore,

$$R = \frac{1}{1 + \sigma^2 \frac{\|v_1\|^2}{\|u\|^2} + 2\sigma^4 \frac{\sum_{j=2}^{n_H} |v_1^\dagger v_j|^2}{\|u\|^4} + O_P(\sigma^6)}.$$

The result follows from the distribution of these quantities. $\square$

Proof of Proposition 8. Let us rewrite (A.8) as follows

$$\hat{v}(\epsilon) = w_0 + w_1\epsilon + O_P(\epsilon^2),$$

where $w_0 = E^{-1}\mathbf{e}_1$, $w_1 = \frac{1}{\sqrt{z}}\left(E^{-1}\hat{b} - \frac{\overline{E^{b1}}}{E^{11}}E^{-1}\mathbf{e}_1\right)$. For convenience, decompose the matrices E and E^{-1} as

$$E = \begin{pmatrix} E_{11} & E_{12}^\dagger \\ E_{12} & E_{22} \end{pmatrix}, \qquad E^{-1} = \begin{pmatrix} E^{11} & E^{12\dagger} \\ E^{12} & E^{22} \end{pmatrix} \tag{A.15}$$

where $E_{11} \in \mathbb{R}$, $E_{12} \in \mathbb{C}^{(m-1)\times 1}$, and $E_{22} \in \mathbb{C}^{(m-1)\times(m-1)}$. Consequently, $E^{11} = 1/(E_{11} - E_{12}{}^\dagger E_{22}{}^{-1}E_{12}) \in \mathbb{R}$ and $E^{12} = -E^{11}E_{22}{}^{-1}E_{12} \in \mathbb{C}^{(m-1)\times 1}$. The exact form of E^{22} is unimportant as it does not affect our calculations.

Let us now focus on the numerator of R. Since $\mathbf{e}^\dagger w_1 = 0$, we have

$$\hat{v}^\dagger e_1 = E^{11} + O_P(\epsilon^2),$$

from which we obtain

$$|\hat{v}^\dagger e_1|^2 = \left(E^{11}\right)^2 + O_P(\epsilon^2).$$

The denominator of R can be written as

$$\|\hat{v}\|^2 = \mathbf{e}_1^\dagger \left(E^{-1}\right)^2 \mathbf{e}_1 + \frac{2}{\sqrt{z}}\Re\left\{\mathbf{e}_1^\dagger \left(E^{-1}\right)^2 \hat{b}\right\}\epsilon - \frac{2}{\sqrt{z}E^{11}}\Re\left\{E^{b1}\right\}\mathbf{e}_1^\dagger \left(E^{-1}\right)^2 \mathbf{e}_1\epsilon + O_P(\epsilon^2)$$

Using the decomposition of E^{-1} given in (A.15), we get

$$\|\hat{v}\|^2 = \left(E^{11}\right)^2 + \|E^{12}\|^2 + \frac{2}{\sqrt{z}}\Re\left\{E^{11}E^{12\dagger}b + E^{21\dagger}E^{22}b\right\}\epsilon - \frac{2}{\sqrt{z}E^{11}}\Re\left\{E^{b1}\right\}\left\{\left(E^{11}\right)^2 + \|E^{12}\|^2\right\}\epsilon + O_P(\epsilon^2).$$

Now we can conveniently express R as

$$R = \frac{\left(E^{11}\right)^2}{\left(E^{11}\right)^2 + \|E^{12}\|^2} + \frac{\left(E^{11}\right)^2}{\left\{\left(E^{11}\right)^2 + \|E^{12}\|^2\right\}^2}\left(P_E - Q_E\right)\epsilon + O_P(\epsilon^2),$$

where

$$P_E = \frac{2}{\sqrt{z}E^{11}}\left\{\left(E^{11}\right)^2 + \|E^{12}\|^2\right\}\Re\left\{E^{b1}\right\}, \qquad Q_E = \frac{2}{\sqrt{z}}\Re\left\{E^{12\dagger}\left(E^{11}I_{m-1} + E^{22}\right)b\right\}.$$

Since P_E and Q_E are zero mean random variables, we neglect them to obtain

$$R \approx \frac{1}{1 + \|E_{22}^{-1}E_{12}\|^2},$$

where we have used the relation $E^{12} = -E^{11}E_{22}^{-1}E_{12} \in \mathbb{C}^{(m-1)\times 1}$. Noting that $E_{22}^{-1}E_{12}|E_{22} \sim \mathcal{CN}(0, E_{22}^{-1})$ with $E_{22} \sim \mathcal{CW}_{m-1}(n_E, I_{m-1})$, we can show that $1/\left(1 + \|E_{22}^{-1}E_{12}\|^2\right)$ is beta distributed with parameters $n_E - m + 2$ and $m - 1$. Now the final result follows from the observation that, for $X \sim \chi_p^2$ and $Y \sim \chi_q^2$, $X/(X+Y)$ is beta distributed with parameters $p/2$ and $q/2$. $\square$

Appendix B. Proof of auxiliary lemmas

Proof of Lemma 1. Write the $m \times n$ matrix $X(\epsilon) = [x_1, \ldots, x_n]$ and observe that $X(-\epsilon) = UX(\epsilon)$, where $U = \mathrm{diag}(1, -1, \ldots, -1)$, is an orthogonal matrix. Thus, $H(-\epsilon) = U^T H(\epsilon) U$ has the same eigenvalues as $H(\epsilon)$. In particular, the largest eigenvalue ℓ_1 and its corresponding eigenvector v_1 satisfy

$$\ell_1(-\epsilon) = \ell_1(\epsilon), \qquad v_1(-\epsilon) = Uv_1(\epsilon). \tag{B.1}$$

Hence ℓ_1 and the first component of v_1 are even functions of ϵ whereas the remaining components of v_1 are odd.

We decompose the matrix $H(\epsilon)=\sum_{j=1}^{n}\mathbf{x}_j\mathbf{x}_j^\dagger$ as

$$\begin{aligned}H(\epsilon)&=\sum_{j=1}^{n}(u_j\mathbf{e}_1+\epsilon\xi_j^\perp)(u_j\mathbf{e}_1+\epsilon\xi_j^\perp)^\dagger=\sum_{j=1}^{n}|u_j|^2\mathbf{e}_1\mathbf{e}_1^\dagger+\epsilon\sum_{j=1}^{n}[\xi_j^\perp\cdot\overline{u_j}\mathbf{e}_1^\dagger+u_j\mathbf{e}_1\cdot\xi_j^{\perp\dagger}]+\epsilon^2\sum_{j=1}^{n}\xi_j^\perp\cdot\xi_j^{\perp\dagger}\\&=\begin{pmatrix}z&0\\0&0_{m-1}\end{pmatrix}+\epsilon\sqrt{z}\begin{pmatrix}0&b^\dagger\\b&0_{m-1}\end{pmatrix}+\epsilon^2\begin{pmatrix}0&0\\0&Z\end{pmatrix}=A_0+\epsilon A_1+\epsilon^2A_2,\end{aligned}$$

with the matrices A_0,A_1 and A_2 given in (A.5). Following similar arguments which lead to (A.7) and (A.8) with $E=I$, we can establish that $\ell_1(\epsilon)$ and $v_1(\epsilon)$ are analytic in some neighborhood of zero. Therefore, we have the following Taylor series expansions:

$$\ell_1(\epsilon)=\lambda_0+\epsilon^2\lambda_2+\epsilon^4\lambda_4+\cdots\quad\text{and}\quad v_1(\epsilon)=w_0+\epsilon w_1+\epsilon^2w_2+\epsilon^3w_3+\epsilon^4w_4+\cdots\tag{B.2}$$

Also, the eigenprojection $P(\epsilon)$ of ℓ_1 satisfies

$$v_1(\epsilon)=\frac{1}{\mathbf{e}_1^\dagger P(\epsilon)\mathbf{e}_1}P(\epsilon)\mathbf{e}_1.\tag{B.3}$$

Inserting the expansions (B.2) into the eigenvalue equation $Hv_1=\ell_1v_1$ gives the following set of equations for $r\geq 0$

$$A_0w_r+A_1w_{r-1}+A_2w_{r-2}=\lambda_0w_r+\lambda_2w_{r-2}+\lambda_4w_{r-4}+\cdots\tag{B.4}$$

with the convention that vectors with negative subscripts are zero. From the $r=0$ equation, $A_0w_0=\lambda_0w_0$, we readily find that

$$\lambda_0=z,\qquad w_0=const\cdot\mathbf{e}_1.$$

Eq. (B.3) implies that $\mathbf{e}_1^\dagger v_1=1$ and $w_0=v_1(0)=\mathbf{e}_1$. This implies that w_j, for $j\geq 1$, is orthogonal to $\mathbf{e}_1$, that is orthogonal to w_0.

From the eigenvector remarks following (B.1) it follows that $w_{2j}=0$ for $j\geq 1$. These remarks allow considerable simplification of (B.4); we use those for $r=1$ and $r=3$

$$A_1w_0=\lambda_0w_1,\qquad A_2w_1=\lambda_0w_3+\lambda_2w_1\tag{B.5}$$

from which we obtain

$$w_1=z^{-1/2}\hat{b},\qquad w_3=\lambda_0^{-1}(A_2-\lambda_2I)w_1.\tag{B.6}$$

Multiply (B.4) on the left by $w_0^\dagger$ and use the first equation of (B.5) to obtain, for r even,

$$\lambda_r=(A_1w_0)^\dagger w_{r-1}=\lambda_0w_1^\dagger w_{r-1}$$

and hence

$$\lambda_2=\lambda_0w_1^\dagger w_1=b^\dagger b\quad\text{and}\quad\lambda_4=w_1^\dagger(A_2-\lambda_2I)w_1=z^{-1}b^\dagger(Z-bb^\dagger)b.$$

Therefore, we can further simplify (B.6) to yield

$$w_1=z^{-1/2}\hat{b},\qquad w_2=z^{-3/2}\left(A_2-\|b\|^2I_{m-1}\right)\hat{b}=z^{-3/2}\begin{pmatrix}0\\Zb-\|b\|^2b\end{pmatrix}.\quad\square$$

To prove Lemmas 2 and 3, we shall use the following two claims, which are the complex analogues of Theorems 3.2.10 and 3.2.11 in Muirhead [32]. While their proofs are similar to those in the real valued case, for completeness we present them below.

Claim 1. *Suppose $A\sim\mathcal{CW}_m(n,\Sigma)$ with $n>m-1$ where A and Σ are partitioned as follows*

$$A=\begin{pmatrix}A_{11}&A_{12}\\A_{21}&A_{22}\end{pmatrix}\qquad\Sigma=\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix}$$

and let $A_{11\cdot2}=A_{11}-A_{12}A_{22}^{-1}A_{21}$, and $\Sigma_{11\cdot2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Then, $A_{11\cdot2}$ is distributed as $\mathcal{CW}_k(n-m+k,\Sigma_{11\cdot2})$ and is independent of A_{12},A_{21} and A_{22}.

Claim 2. *Let $A\sim\mathcal{CW}_m(n,\Sigma)$ and let M be a $k\times m$ matrix of rank k, where M is independent of A. Then $(MA^{-1}M^\dagger)^{-1}\sim\mathcal{CW}_k(n-m+k,(M\Sigma^{-1}M^\dagger)^{-1})$.*

Proof of Claim 1. Let $C=\Sigma^{-1}$. We partition it as follows,

$$C=\begin{pmatrix}C_{11}&C_{12}\\C_{21}&C_{22}\end{pmatrix},\tag{B.7}$$

where $C_{11}\in\mathbb{C}^{k\times k}$, $C_{22}\in\mathbb{C}^{(m-k)\times(m-k)}$, and $C_{12}\in\mathbb{C}^{k\times(m-k)}$ with $C_{12}^\dagger=C_{21}$. Consequently, $\Sigma_{11\cdot2}^{-1}=C_{11}$.

Following [15,19], the density of A is given by

$$f(A)=\frac{\det^{n-m}(A)}{\Gamma_m(n)\det^n(\Sigma)}e^{-\mathrm{tr}\left(\Sigma^{-1}A\right)} \tag{B.8}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator and

$$\Gamma_m(n)=\pi^{\frac{m}{2}(m-1)}\prod_{j=1}^{m}\Gamma(n-j+1)$$

with $\Gamma(\cdot)$ denoting the classical gamma function.

To prove the claim we shall study the form of $\det(A)$ and of $\mathrm{tr}(\Sigma^{-1}A)$. First of all, we have that

$$\det(A)=\det\left(A_{22}\right)\det\left(A_{11\cdot 2}\right).$$

Next, we introduce a change of variables from the entries of the matrix A, to $A_{11\cdot 2}=A_{11}-A_{12}A_{22}^{-1}A_{21}$, $B_{12}=A_{12}$, $B_{22}=A_{22}$. The Jacobian of this transformation is an upper triangular matrix, with all diagonal entries equal to one. Hence, the volume element in (B.8) is $\mathrm{d}A=\mathrm{d}A_{11}\mathrm{d}A_{12}\mathrm{d}A_{22}=\mathrm{d}A_{11\cdot 2}\mathrm{d}B_{12}\mathrm{d}B_{22}$. Furthermore, using the expansion

$$\begin{aligned}\mathrm{tr}\left(\Sigma^{-1}A\right)&=\mathrm{tr}\left(\begin{pmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{pmatrix}\begin{pmatrix}A_{11\cdot 2}+B_{12}B_{22}^{-1}B_{21} & B_{12}\\ B_{21} & B_{22}\end{pmatrix}\right)\\ &=\mathrm{tr}\left(C_{11}A_{11\cdot 2}\right)+\mathrm{tr}\left(C_{11}B_{12}B_{22}^{-1}B_{21}\right)+\mathrm{tr}\left(C_{12}B_{21}\right)+\mathrm{tr}\left(C_{21}B_{12}\right)+\mathrm{tr}\left(C_{22}B_{22}\right)\end{aligned}$$

along with the fact that $B_{21}=B_{12}^{\dagger}$ yields that

$$\begin{aligned}f\left(A_{11\cdot 2},B_{12},B_{22}\right)=&\frac{\det^{n-m}\left(B_{22}\right)\det^{n-m}\left(A_{11\cdot 2}\right)}{\det^{n}\left(\Sigma_{22}\right)\det^{n}\left(\Sigma_{11\cdot 2}\right)\Gamma_m(n)}\times\\ &e^{-\mathrm{tr}\left(\Sigma_{11\cdot 2}^{-1}A_{11\cdot 2}\right)-\mathrm{tr}\left(\Sigma_{11\cdot 2}^{-1}B_{12}B_{22}^{-1}B_{12}^{\dagger}\right)}\times e^{-\mathrm{tr}(C_{21}B_{12})-\mathrm{tr}(C_{21}B_{12})^{\dagger}-\mathrm{tr}(C_{22}B_{22})}.\end{aligned} \tag{B.9}$$

Now we may use the decomposition

$$\begin{aligned}\Gamma_m(n)&=\pi^{\frac{k}{2}(k-1)}\prod_{j=1}^{k}\Gamma(n-m+k-j+1)\times\pi^{\frac{(m-k)}{2}(m+k-1)}\prod_{j=1}^{m-k}\Gamma(n-j+1)\\ &=\Gamma_k(n-m+k)\times\pi^{\frac{(m-k)}{2}(m+k-1)}\prod_{j=1}^{m-k}\Gamma(n-j+1)\end{aligned}$$

to rewrite (B.9) as

$$f\left(A_{11\cdot 2},B_{12},B_{22}\right)=f_1\left(A_{11\cdot 2}\right)\times f_2\left(B_{12},B_{22}\right), \tag{B.10}$$

where

$$f_1\left(A_{11\cdot 2}\right)=\frac{\det^{n-m+k-k}\left(A_{11\cdot 2}\right)}{\det^{n-m+k}\left(\Sigma_{11\cdot 2}\right)\Gamma_k(n-m+k)}e^{-\mathrm{tr}\left(\Sigma_{11\cdot 2}^{-1}A_{11\cdot 2}\right)}, \tag{B.11}$$

and

$$\begin{aligned}f_2\left(B_{12},B_{22}\right)=&\frac{\det^{n-m}\left(B_{22}\right)}{\pi^{\frac{(m-k)}{2}(m+k-1)}\prod_{j=1}^{m-k}\Gamma(n-j+1)\det^{m-k}\left(\Sigma_{11\cdot 2}\right)\det^{n}\left(\Sigma_{22}\right)}\\ &\times e^{-\mathrm{tr}\left(\Sigma_{11\cdot 2}^{-1}B_{12}B_{22}^{-1}B_{12}^{\dagger}\right)-\mathrm{tr}(C_{21}B_{12})-\mathrm{tr}(C_{21}B_{12})^{\dagger}-\mathrm{tr}(C_{22}B_{22})}.\end{aligned}$$

The factorization in (B.10) establishes that $A_{11\cdot 2}$ is independent of A_{12} and A_{22}. Finally, (B.11) implies that $A_{11\cdot 2}\sim\mathcal{CW}_k\left(n-m+k,\Sigma_{11\cdot 2}\right)$ which concludes the proof. □

Proof of Claim 2. Set $B=\Sigma^{-1/2}A\Sigma^{-1/2}$. Now $B\sim\mathcal{CW}_m(n,I)$. For $R=M\Sigma^{-1/2}$, $(MA^{-1}M^{\dagger})^{-1}=(RB^{-1}R^{\dagger})^{-1}$ and $(M\Sigma^{-1}M^{\dagger})^{-1}=(RR^{\dagger})^{-1}$. Thus, it is sufficient to prove that $(RB^{-1}R^{\dagger})^{-1}\sim\mathcal{CW}_k(n-m+k,(RR^{\dagger})^{-1})$. Let $R=L[I_k:0]H$ be the SVD decomposition of R, where L is $k\times k$ and nonsingular and H is $m\times m$ unitary. Now,

$$\begin{aligned}(RB^{-1}R^{\dagger})^{-1}&=\left(L[I_k:0]HB^{-1}H^{\dagger}[I_k:0]'L^{\dagger}\right)^{-1}=(L^{-1})^{\dagger}\left([I_k:0](HBH^{\dagger})^{-1}[I_k:0]'\right)^{-1}L^{-1}\\ &=(L^{-1})^{\dagger}\left([I_k:0]C^{-1}[I_k:0]'\right)^{-1}L^{-1}\end{aligned}$$

where $C=HBH^{\dagger}\sim\mathcal{CW}_m(n,I)$. Let

$$F=C^{-1}=\begin{pmatrix}F_{11} & F_{12}\\ F_{21} & F_{22}\end{pmatrix},\quad C=\begin{pmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{pmatrix}$$

where F_{11} and C_{11} are $k \times k$. Then $(RB^{-1}R^{\dagger})^{-1} = (L^{-1})^{\dagger}F_{11}^{-1}L^{-1}$, and since $F_{11}^{-1} = C_{11} - C_{12}C_{22}^{-1}C_{21}$, it follows from Claim 1 that $F_{11}^{-1} \sim \mathcal{CW}_k(n-m+k, I_k)$. Hence $(L^{-1})^{\dagger}F_{11}^{-1}L^{-1} \sim \mathcal{CW}_k(n-m+k, (LL^{\dagger})^{-1})$, and since $(LL^{\dagger})^{-1} = (RR^{\dagger})^{-1}$, the proof is complete. □

Proof of Lemma 2. Note that $S^{11} = E^{11} = \mathbf{e}_1^T E^{-1}\mathbf{e}_1$. Then, by Claim 2, $(S^{11})^{-1} \sim \mathcal{CW}_1(n_E - m + 1, I_1) = \chi^2_{2n_E-2m+2}/2$, meaning $S^{11} \sim 2/\chi^2_{2n_E-2m+2}$. Next, by definition $S = (M^{\dagger}E^{-1}M)^{-1}$, with fixed M. Thus, by the same claim, $S \sim \mathcal{CW}_2(n_E - m + 2, D)$ from which we obtain $S_{22} \sim \chi^2_{2n_E-2m+4}/(2\|b\|^2)$. Finally, since $(S^{11})^{-1} = S_{11} - S_{12}S_{22}^{-1}S_{21}$, by Claim 1, $(S^{11})^{-1}$ is independent of S_{22}. □

Proof of Lemma 3. First we decompose the expectation as follows:

$$\mathbb{E}\left(\frac{\mathbf{e}_1^T E^{-1}A_2E^{-1}\mathbf{e}_1}{E^{11}}\right) = \mathbb{E}_E\left\{\mathbb{E}_{A|E}\left(\frac{\mathbf{e}_1^T E^{-1}A_2E^{-1}\mathbf{e}_1}{E^{11}}\right)\right\}.$$

Next, since A_2 is independent of E,

$$\mathbb{E}(A_2|E) = \mathbb{E}(A_2) = \begin{pmatrix} 0 & 0 \\ 0 & I_{m-1} \end{pmatrix}.$$

Combining the above two equations gives that

$$\mathbb{E}\left(\frac{\mathbf{e}_1^t opE^{-1}A_2E^{-1}\mathbf{e}_1}{E^{11}}\right) = \mathbb{E}\left(\sum_{j=2}^{m}\frac{\|E^{1j}\|^2}{E^{11}}\right) = (m-1)\mathbb{E}\left(\frac{\|E^{12}\|^2}{E^{11}}\right).$$

To compute this expectation, consider the matrix $S^{-1} = [\mathbf{e}_1\ \ \mathbf{e}_2]^{\top}E^{-1}[\mathbf{e}_1\ \ \mathbf{e}_2] = \begin{pmatrix} E^{11} & E^{21} \\ \bar{E}^{21} & E^{22} \end{pmatrix}$. Since $S^{22} = E^{22}$ and $S_{22} = E^{11}/(E^{11}E^{22} - \|E^{12}\|^2)$, we have

$$\frac{1}{S_{22}} = E^{22} - \frac{\|E^{12}\|^2}{E^{11}}. \tag{B.12}$$

Noting that $S_{22} \sim \frac{1}{2}\chi^2_{2n_E-2m+4}$ and $E^{22} \sim 2/\chi^2_{2n_E-2m+2}$, we take the expectation of both sides of (B.12) to obtain

$$\mathbb{E}\left\{\frac{(E^{12})^2}{E^{11}}\right\} = \mathbb{E}\left(E^{22}\right) - \mathbb{E}\left(\frac{2}{\chi^2_{2n_E-2m+4}}\right) = \frac{1}{(n_E-m)(n_E-m+1)}$$

which completes the proof. □

References

[1] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, third ed., Wiley, New York, 2003.

[2] N. Asendorf, R.R. Nadakuditi, Improved detection of correlated signals in low-rank-plus-noise type data sets using informative canonical correlation analysis (ICCA), IEEE Trans. Inform. Theory 63 (6) (2017) 3451–3467.

[3] J. Baik, G. Ben Arous, S. Péché, Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices, Ann. Probab. (2005) 1643–1697.

[4] M. Chiani, Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy-Widom distribution, J. Multivariate Anal. 129 (2014) 68–81.

[5] M. Chiani, Distribution of the largest root of a matrix for Roy's test in multivariate analysis of variance, J. Multivariate Anal. 143 (2016) 467–471.

[6] M. Chiani, On the probability that all eigenvalues of Gaussian, Wishart, and double Wishart random matrices Lie within an interval, IEEE Trans. Inform. Theory 63 (7) (2017) 4521–4531, http://dx.doi.org/10.1109/TIT.2017.2694846.

[7] N.M. Correa, T. Adali, Y.O. Li, V.D. Calhoun, Canonical correlation analysis for data fusion and group inferences, IEEE Sig. Proc. Mag. 27 (4) (2010) 39–50.

[8] P. Dharmawansa, I.M. Johnstone, A. Onatski, Local asymptotic normality of the spectrum of high-dimensional spiked F-ratios, 2014, arXiv preprint arXiv:1411.3875.

[9] A. Edelman, Eigenvalues and condition numbers of random matrices, SIAM J. Matrix Anal. Appl. 9 (4) (1988) 543–560.

[10] N. El Karoui, A rate of convergence result for the largest eigenvalue of complex white Wishart matrices, Ann. Probab. 34 (6) (2006) 2077–2117.

[11] H. Ge, I.P. Kirsteins, X. Wang, Does canonical correlation analysis provide reliable information on data correlation in array processing? in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 2113–2116.

[12] S. Gogineni, P. Setlur, M. Rangaswamy, R.R. Nadakuditi, Passive radar detection with noisy reference signal using measured data, in: IEEE Radar Conference, 2017, pp. 858–861.

[13] S. Gogineni, P. Setlur, M. Rangaswamy, R.R. Nadakuditi, Comparison of passive radar detectors with noisy reference signal, in: IEEE Statistical Signal Processing Workshop (SSP), 2016, pp. 1–5.

[14] S. Gogineni, P. Setlur, M. Rangaswamy, R.R. Nadakuditi, Random matrix theory inspired passive bistatic radar detection of low-rank signals, in: IEEE Radar Conference, 2015, pp. 1656–1659.

[15] N.R. Goodman, Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction), Ann. Math. Stat. 34 (1) (1963) 152–177.

[16] J. Hansen, H. Bolcskei, A geometrical investigation of the rank-1 Ricean MIMO channel at high SNR, in: Intl. Symp. on Inform. Theory, IEEE, 2004, p. 64.

[17] S. Haykin, Cognitive radio: brain-empowered wireless communications, IEEE J. Sel. Areas Commun. 23 (2) (2005) 201–220.
[18] S. Haykin, M. Moher, Communication Systems, fifth ed., Wiley, New York, 2009.
[19] A.T. James, Distributions of matrix variates and latent roots derived from normal samples, Ann. Math. Stat. 35 (1964) 475–501.
[20] K. Johansson, Shape fluctuations and random matrices, Comm. Math. Phys. 209 (2) (2000) 437–476.
[21] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2001) 295–327.
[22] I.M. Johnstone, Approximate null distribution of the largest root in multivariate analysis, Ann. Appl. Statist. 3 (4) (2009) 1616–1633.
[23] I.M. Johnstone, B. Nadler, Roy's largest root test under rank-one alternatives, Biometrika 104 (2017) 181–193.
[24] M. Kang, M.-S. Alouini, Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO MRC systems, IEEE J. Sel. Areas Commun. 21 (3) (2003) 418–426.
[25] T. Kato, Perturbation theory of linear operators, second ed., Springer, Berlin, 1995.
[26] M.U. Khalid, A.K. Seghouane, Improving functional connectivity detection in FMRI by combining sparse dictionary learning and canonical correlation analysis, in: IEEE 10th Intl. Symp. on Biomedical Imaging, 2013, pp. 286–289.
[27] C. Khatri, Distribution of the largest or the smallest characteristic root under null hypothesis concerning complex multivariate normal populations, Ann. Math. Stat. 35 (1964) 1807–1810.
[28] C. Khatri, Non-central distributions of the i-th largest characteristic roots of three matrices concerning complex multivariate normal populations, Ann. I. Stat. Math. 21 (1) (1969) 23–32.
[29] S. Kritchman, B. Nadler, Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory, IEEE Trans. Signal Process. 57 (10) (2009) 3930–3941.
[30] D. Lin, J. Zhang, J. Li, V. Calhoun, Y.P. Wang, Identifying genetic connections with brain functions in schizophrenia using group sparse canonical correlation analysis, in: IEEE 10th Intl. Symp. Biomedical Imaging, 2013, pp. 278–281.
[31] Z. Ma, Accuracy of the Tracy-Widom limits for the extreme eigenvalues in white Wishart matrices, Bernoulli 18 (2012) 322–359.
[32] R.J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.
[33] R.R. Nadakuditi, J.W. Silverstein, Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples, IEEE J. Sel. Topics Sig. Proc. 4 (3) (2010) 468–480.
[34] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, Statist. Sinica 17 (2007) 1617–1642.
[35] A. Pezeshki, L.L. Scharf, M.R. Azimi-Sadjadi, M. Lundberg, Empirical canonical correlation analysis in subspaces, in: Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, vol. 1, 2004, pp. 994–997.
[36] T. Ratnarajah, R. Vaillancourt, M. Alvo, Eigenvalues and condition numbers of complex random matrices, SIAM J. Matrix Anal. Appl. 26 (2) (2004) 441–456.
[37] T. Ratnarajah, R. Vaillancourt, M. Alvo, Complex random matrices and Rician channel capacity, Probl. Inf. Transm. 41 (1) (2005) 1–22.
[38] S.N. Roy, On a heuristic method of test construction and its use in multivariate analysis, Ann. Math. Stat. 24 (2) (1953) 220–238.
[39] S.N. Roy, Some Aspects of Multivariate Analysis, Wiley, New York, 1957.
[40] L. Scharf, J.K. Thomas, Wiener filters in canonical coordinates for transform coding, filtering, and quantizing, IEEE Trans. Signal Process. 46 (3) (1998) 647–654.
[41] T. Sugiyama, Distributions of the largest latent root of the multivariate complex Gaussian distribution, Ann. I. Stat. Math. 24 (1) (1972) 87–94.
[42] H.L. Van Trees, Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory, John Wiley & Sons, New York, 2002.
[43] K.E. Wage, J.R. Buck, Snapshot performance of the dominant mode rejection beamformer, IEEE J. Ocean. Eng. 39 (2) (2014) 212–225.
[44] A. Zanella, M. Chiani, M.Z. Win, On the marginal distribution of the eigenvalues of Wishart matrices, IEEE Trans. Commun. 57 (4) (2009) 1050–1060.
[45] Y. Zeng, Y.-C. Liang, Eigenvalue-based spectrum sensing algorithms for cognitive radio, IEEE Trans. Commun. 57 (6) (2009) 1784–1793.
[46] Y. Zeng, Y.-C. Liang, A.T. Hoang, R. Zhang, A review on spectrum sensing for cognitive radio: challenges and solutions, EURASIP J. Adv. Signal Process. 2010 (2010) 381465.