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

A R T I C L E   I N F O

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

A B S T R A C T

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.

© 2019 Elsevier Inc. All rights reserved.

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 $\ell_1$ 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 $E \sim CW_m(n_E, I)$. The distribution of the matrix $H$ will either be central $H \sim CW_m(n_H, \Sigma_H)$, or non-central $H \sim CW_m(n_H, \Sigma_H, \Omega)$. 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 $\ell_1$, 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 $\ell_1$ 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,
Zanella et al. [44] derived simpler exact, yet recursive expressions, both for the central case with arbitrary \( \Sigma_I \) and for the non-central case but with \( \Sigma_I = I \). 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 \( \Sigma_E = \Sigma_I = I \), is based on random matrix theory. Considering the limit as \( n_I \) and \( m \) (and in the double matrix case also \( n_E \)) tend to infinity, with their ratios converging to constants, \( \ell_1 \) 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 \( \Sigma_I = I + \lambda vv^\dagger \), where \( \lambda \) is a measure of signal strength, the unit norm vector \( v \in \mathbb{C}^m \) is its direction, and \( vv^\dagger \) denotes the conjugate transpose of \( v \). Similarly, in the non-central case, we assume that \( H \sim \mathcal{C}W_{m}(n_I, I, \Omega) \), with a rank-one non-centrality matrix \( \Omega = \lambda vv^\dagger \). Our goal is to study the distribution of \( \ell_1 \) and its dependence on \( \lambda \), 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 \( n_I \rightarrow \infty \), the largest eigenvalue of \( H \) converges to a Gaussian distribution [1]. In the random matrix setting, as both \( n_I \) and \( m \) tend to infinity with their ratio tending to a constant, Baik et al. [3] and Paul [34] proved that if \( \lambda > \sqrt{m/n_I} \) then \( \ell_1 \) 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, \( \ell_1 \) 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 \( \ell_1 \), which can be quite far from Gaussian, see Fig. 1 (left) for an illustrative example. Accurate expressions for the distribution of \( \ell_1 \), 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 \( \ell_1 \) 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 \( \chi^2_k \) the
chi-squared distribution with $k$ degrees of freedom and by $\chi^2_k(\eta)$ the non-central chi-squared distribution with non-centrality parameter $\eta$. Throughout the manuscript we follow the standard definition of complex valued multivariate Gaussian random variables, see [15]. Specifically, if $X \sim CN(0, \sigma^2)$ then it can be written as $(A + iB)/\sqrt{2}$ where $A, B \in \mathbb{R}$ are independent $N(0, \sigma^2)$ random variables and $i = \sqrt{-1}$.

We start with the simplest Case 1 in Table 1, involving a single central Wishart matrix, $H \sim CVW_m(n_H, \Sigma)$, in various engineering applications the matrix $\Sigma$ denotes the covariance of the noise measured at $m$ sensors and is often assumed to be known, whereas $\lambda$ is a measure of the signal strength and the unit norm vector $v$ denotes its direction. Without loss of generality, we thus assume $\Sigma = \sigma^2 I$, where $\sigma^2$ 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 $\sigma$, we shall denote it by $\ell_1(\sigma)$.

**Proposition 1.** Let $H \sim CVW_m(n_H, \lambda vv^T + \sigma^2 I)$, with $\|v\| = 1$, $\lambda > 0$ and let $\ell_1(\sigma)$ be its largest eigenvalue. Then, with $(m, n_H, \lambda)$ fixed, as $\sigma \to 0$

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

(1)

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

**Remark 1.** Given that $\ell_1$ is the largest eigenvalue of a Wishart matrix, it has finite mean and variance. Approximate formulas for these quantities follow directly from (1). Since $\mathbb{E}\{\chi^2_k\} = k \mathbb{V}ar\{\chi^2_k\} = 2k$, and $\mathbb{E}\{1/\chi^2_k\} = 1/(k - 2)$ 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,

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

**Proposition 2.** Let $H \sim CVW_m(n_H, \sigma^2 I, (\omega/\sigma^2) vv^T)$, with $\|v\| = 1$, $\omega > 0$ and let $\ell_1(\sigma)$ be its largest eigenvalue. Then, with $(m, n_H, \omega)$ fixed, as $\sigma \to 0$

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

(2)

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

**Remark 2.** The exact distribution of the largest eigenvalue $\ell_1$ 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 $\lambda$ and $\sigma$. In contrast, while (1) is approximate, the dependence on the values of $\lambda$ and $\sigma$ is more explicit.

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

**Proposition 3.** Let $H \sim CVW_m(n_H, \sigma^2 I, (\omega/\sigma^2) vv^T)$, with $\|v\| = 1$, $\omega > 0$ and let $\ell_1(\sigma)$ be its largest eigenvalue. Then, with $(m, n_H, \omega)$ fixed, as $\sigma \to 0$

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

and

$$\mathbb{V}ar\{\ell_1(\sigma)\} \approx 8\omega + 4\sigma^2 \left\{ n_H + m - 1 + \frac{(n_H - 1)(m - 1)}{2(n_H + \sigma^2/\omega - 1)^2(n_H + \sigma^2/\omega - 2)} \right\}.$$
Proposition 3. Let $H \sim CV_m(n_H, I + \lambda vv^\top)$ and $E \sim CV_m(n_E, I)$ be independent, with $n_E > m + 1$ and $\|v\| = 1$. Let $\ell_1$ be the largest eigenvalue of $E^{-1}H$. Then, with $(m, n_H, n_E)$ fixed, as $\lambda$ becomes large

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

where the two $F$ distributed random variates are independent and

$$a_1 = \frac{n_H}{n_E - m + 1}, \quad a_2 = \frac{m - 1}{n_E - m + 2}, \quad a_3 = \frac{m - 1}{(n_E - m)(n_E - m - 1)},$$

$$b_1 = 2n_H, \quad b_2 = 2m - 2, \quad c_1 = 2n_E - 2m + 2, \quad c_2 = 2n_E - 2m + 4.$$  (4)

Proposition 4. Suppose that $H \sim CV_m(n_H, I, \omega vv^\top)$ and $E \sim CV_m(n_E, I)$ are independent, with $n_E > m + 1$, $\omega > 0$, and $\|v\| = 1$. Let $\ell_1$ be the largest eigenvalue of $E^{-1}H$. Then, with $(m, n_H, n_E)$ fixed, as $\omega$ becomes large

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

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 $\chi^2$ 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 $\Sigma$,

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

Here $\tilde{P} = (P - 0_{p \times (q-p)})$ with $P = \text{diag}(\rho, 0, \ldots, 0) \in \mathbb{R}^{p \times p}$ and $\rho$ is the value of the correlation coefficient.
To study the sample canonical correlation, consider \( n + 1 \) complex-valued \( m \)-dimensional multivariate Gaussian observations \( x_i \sim \mathcal{CN}(0, \Sigma), \ i \in \{1, \ldots, n + 1\} \) on \( m = p + q \) variables, where without loss of generality \( p \leq q \).

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 \cite[Chapter 10]{Huber1981}, its square \( r_1^2 \) is the largest root of the following characteristic equation
\[
\det \left( r^2 Y^\dagger Y - Y^\dagger QY \right) = 0, \tag{6}
\]
where \( Q = X (X^\dagger X)^{-1} X^\dagger \). Introducing the notation \( H = Y^\dagger QY \) 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 \cite{Li1992}, 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
\[
H|X \sim \mathcal{CW}_p(q, \Phi, \Omega) \quad \text{and} \quad E|X \sim \mathcal{CW}_p(n - q, \Phi)
\]
with the non-centrality matrix given by
\[
\Omega = \Phi^{-1} P X^\dagger X P^\dagger = \frac{\rho_2^2}{1 - \rho_2^2} (X^\dagger X)_{11} e_1 e_1^\dagger = \omega \ e_1 e_1^\dagger \quad \text{where} \ \omega = \frac{\rho_2^2}{1 - \rho_2^2} (X^\dagger X)_{11}. \tag{8}
\]
Fig. 4. Density function of $\ell_1(E^{-1}H)$ in canonical correlation analysis.

Since $X^\intercal X \sim \mathcal{C}_V(n, I_q)$, then all diagonal entries of $X^\intercal X$ follow a chi-square distribution. In particular, $(X^\intercal X)_{11} \sim \chi^2_{2n}/2$. 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 $F_{a,b}^{\chi^2}(c, n)$ a random variable, which is defined as a function of three other random variables as follows: First, generate a random variable $Z \sim \chi^2_{n}$. Next, generate two independent random variables, one distributed as $\chi^2_a(Z)$ and the other as $\chi^2_b$. Finally, compute their ratio

$$ \frac{\chi^2_a(Z)}{\chi^2_b} \sim F_{a,b}^{\chi^2}(c, n). \quad (9) $$

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 $p \leq q$ computed from $n + 1$ i.i.d. observations with $n = n - p - q > 1$. Then in the presence of a single large population correlation coefficient $\rho$ between the two groups, asymptotically as $\rho \to 1$,

$$ \ell_1 \approx a_1 F_{b_1,c_1}^{\chi^2}(\rho^2, 2n) + a_2 F_{b_2,c_2} + a_3 $$

where

$$ a_1 = q/(\nu + 1), \quad a_2 = (p - 1)/(\nu + 2), \quad a_3 = (p - 1)/(\nu(\nu - 1)), $$

$$ b_1 = 2q, \quad b_2 = 2p - 2, \quad c_1 = 2(\nu + 1), \quad c_2 = 2(\nu + 2). $$

Remark 5. It can be shown that the probability density of $F_{a,b}^{\chi^2}(c, n)$ is

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

where ${}_2F_1(a, b; c; z)$ is the Gauss hypergeometric function and $B(p, q)$ 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 $\hat{v}$, and its corresponding population eigenvector $v$. Assuming $\|v\| = \|\hat{v}\| = 1$,

$$ R = |\hat{v}^\intercal v|^2. \quad (10) $$

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 $H \sim CV_m(n_h, \sigma^2 I)$, with $\|v\| = 1$ and $\lambda > 0$. Let $\hat{v}_m$ be the eigenvector corresponding to the largest eigenvalue of $H$. Then, with $\sigma$, $n_h$, $\lambda$ fixed, for small $\sigma$

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

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

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

**Proposition 7.** Let $H \sim CV_m(n_h, \sigma^2 I, (\omega/\sigma^2)vv^t)$, with $\|v\| = 1$ and $\omega > 0$. Let $\hat{v}_m$ be the eigenvector corresponding to the largest eigenvalue of $H$. Then, with $\sigma$, $n_h$, $\omega$ fixed, for small $\sigma$

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

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

Propositions 6 and 7 can be useful to analyze theoretically various DMR and radar detection 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 $H \sim CV_m(n_h, \lambda vv^t + I)$ and $E \sim CV_m(n_E, I)$ be independent, with $n_E > m + 1$ and $\|v\| = 1$. Let $\hat{v}$ be the eigenvector corresponding to the largest eigenvalue of $E^{-1}H$. Then, with $\sigma$, $n_h$, $n_E$ fixed, for large $\lambda$

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

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 $y(t) \in \mathbb{C}^m$ at time $t$ is of the form

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

where $s(t) \in \mathbb{C}$ is the time dependent signal, $u \in \mathbb{C}^m$ is normalized such that $\|u\| = 1$ is its direction, $\lambda \geq 0$ is a measure of the signal strength and the vector $n \in \mathbb{C}^m$ is a zero mean complex valued random noise, assumed to be independent of
the signal and distributed as $\mathbf{n} \sim \mathcal{CN}(0, \Sigma)$. The positive definite Hermitian matrix $\Sigma$ 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 $s(t) = s_1(t) + is_2(t)$ 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]), $s(t) = \exp(i\phi(t))$, where $\phi(t)$ is random.

When the covariance matrix $\Sigma$ 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, $\Sigma$ 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 $E = \sum_{j=1}^{n_E} \mathbf{n}_j \mathbf{n}_j^\dagger$ 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 $s \sim \mathcal{CN}(0, 1)$,

$$H \sim \mathcal{CW}_m(n_H, \lambda uu^\dagger + \sigma^2 I).$$
In contrast, if $s = \exp(i\phi)$, conditional on $\phi_1, \ldots, \phi_{nH}$,

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

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

$$P_D = \Pr \{ \ell_1 > \mu | \text{signal present with strength } \lambda \},$$

where $\mu$ 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 $x \in \mathbb{C}^{n_T}$ and received signals $y \in \mathbb{C}^{n_R}$ are related as

$$y = Hx + n$$

where $H$ is the $n_R \times n_T$ channel matrix and $n$ is additive random complex valued noise, assumed to be distributed as $n \sim \mathcal{CN} \left( n, \sigma_n^2 I \right)$, where $\sigma_n^2$ 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

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

(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 $\text{tr}(H_1 H_1^\dagger) = n_R n_T$, 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

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

(14)

where $\Omega_D = \mathbb{E}[\|x\|^2]$ 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 $P_{\text{out}}$ can be written as

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

(15)

One particularly interesting case is when the Rician component $H_1$ is assumed to be of rank one, $H_1 = uv^\dagger$, where $u \in \mathbb{C}^{n_R}$, $v \in \mathbb{C}^{n_T}$. 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., $\sigma_H \ll 1$).
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 \( H_1 = u w^\dagger \) is rank one, the \( j \)th column of \( H \), of dimension \( n_R \), is distributed as \( CN(\sqrt{K/(K + 1)} u, \sigma_H^2/(K + 1) I_{n_R}) \). Therefore,

\[
HH^\dagger \sim CW_{n_R}(n_T, \alpha^2 I_{n_R}, \beta^2/\alpha^2 w w^\dagger)
\]

is non-central Wishart, with

\[
w = \nu/\|\nu\|, \quad \alpha^2 = \frac{1}{K + 1} \sigma_H^2 \quad \text{and} \quad \beta^2 = \frac{K}{K + 1} \||\nu||^2\|
\]

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

\[
\mu = \frac{\Omega_D}{\sigma^2_H} \ell_1 = c_1 \left( A + B + \frac{BC}{A} \right) + o_p(\sigma_H^2)
\]

where \( A, B, C \) are independent random variables distributed as

\[
A \sim \chi^2_{2n_T}(c_2), \quad B \sim \chi^2_{2n_R-2}, \quad C \sim \chi^2_{2n_R-2}
\]

and

\[
c_1 = \frac{\Omega_D \sigma_H^2}{2(K + 1) \sigma^2_H}, \quad c_2 = \frac{2 \beta^2}{\alpha^2} = \frac{2K}{\sigma^2_H} n_R n_T.
\]

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 \( P_{\text{out}} \) of (15) is minimal when the largest eigenvalue \( \ell_1 \) 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). \( \Box \)

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 \( e_1 = (1, 0, \ldots, 0)^T \), but corrupted by small perturbations. Its proof is in Appendix B.
Lemma 1. Let \( \{x_i\}_{j=1}^n \) be \( n \) vectors in \( \mathbb{C}^m \) of the form
\[
x_i = u_i e_1 + \epsilon \xi_i^1 \quad (A.1)
\]
where \( u_i \) are complex valued scalars, \( \xi_i^1 = \begin{pmatrix} 0 \\ \xi_i \end{pmatrix} \) with \( \xi_i \in \mathbb{C}^{m-1} \) are the perturbations in orthogonal directions to \( e_1 \) and \( \epsilon \in \mathbb{R} \) is a small parameter. Define \( Z \in \mathbb{R}^{m} \), \( b \in \mathbb{C}^{m-1} \) and \( Z \in \mathbb{C}^{(m-1) \times (m-1)} \)
\[
z = \sum_{j=1}^{n} u_j \overline{\eta}_j, \quad b = z^{-\frac{1}{2}} \sum_{j=1}^{n} \overline{\eta}_j \xi_j, \quad Z = \sum_{j=1}^{n} \xi_j \xi_j^1. \quad (A.2)
\]
Let \( \ell_1(\epsilon) \) be the largest eigenvalue of \( H(\epsilon) = \sum_{j=1}^{n} x_i x_i^\dagger \) with corresponding leading eigenvector \( v_1(\epsilon) \) normalized such that \( v_1^\dagger v_1(\epsilon) = 1 \). Then \( \ell_1(\epsilon) \) is an even analytic function of \( \epsilon \), whereas \( v_1(\epsilon) - e_1 \) is an odd function of \( \epsilon \). In particular, the Taylor expansions of \( \ell_1(\epsilon) \) and \( v_1(\epsilon) \) around \( \epsilon = 0 \) are given by
\[
\ell_1(\epsilon) = z + \|b\|^2 \epsilon^2 + z^{-1} b^\dagger (Z - bb^\dagger) b \epsilon^4 + \cdots \\
v_1(\epsilon) = e_1 + z^{-1/2} \begin{pmatrix} 0 \\ b \end{pmatrix} \epsilon + z^{-3/2} \begin{pmatrix} 0 \\ Z b - \|b\|^2 b \end{pmatrix} \epsilon^3 + \cdots \quad (A.3)
\]

Proof of Propositions 1 and 2. Since the eigenvalues of \( H \) do not depend on the direction of the vector \( v \), without loss of generality we thus assume that \( v = e_1 \). Then, \( H \) may be realized from \( n_H \) i.i.d. observations of the form \( (A.1) \) with \( \epsilon \) replaced by \( \sigma \).
\[
\xi_i \sim \mathcal{CN}(0, I_{m-1}), \quad u_j \sim \begin{cases} \mathcal{CN}(0, \sigma^2 + \lambda), & \text{Proposition 1} \\ \mathcal{CN}(\mu_j, \sigma^2), & \text{Proposition 2} \end{cases} \quad \text{Proposition 1}
\]
and \( \mu_j \) are arbitrary complex numbers satisfying \( \sum_j |\mu_j|^2 = \omega \).

For each realization of \( u = (u_k) \) and \( \mathcal{S} = [\xi_1, \ldots, \xi_n] \in \mathbb{C}^{(m-1) \times nh} \), 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 \( o_1 = \overline{u}/\|u\| \in \mathbb{C}^m \), choose columns \( o_2, \ldots, o_n \) so that \( O = [o_1, \ldots, o_n] \) is an \( nh \times nh \) unitary matrix, and consider the following \( (m-1) \times nh \) matrix \( V = \mathcal{S} O \). Its first column is \( v_1 = \mathcal{S} \overline{u}/\|u\| = b \), and thus the \( O(\epsilon^2) \) term in \( (A.3) \) is \( b^\dagger b = \|v_1\|^2 \). For the fourth order term, observe that \( Z = \mathcal{S} \mathcal{S}^\dagger = V V^\dagger \) and so the quantity \( D = b^\dagger (Z - bb^\dagger) b \) may be written as
\[
D = v_1^\dagger (VV^\dagger - v_1 v_1^\dagger) v_1 = (v_1^\dagger V)(v_1^\dagger V)^\dagger - (v_1^\dagger v_1)(v_1^\dagger v_1)^\dagger = \sum_{j=2}^{nh} |v_1^\dagger u_j|^2.
\]
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 \( (A.4) \),
\[
u_j = (a_j + ib_j) / \sqrt{2} \quad \text{with}
\]
\[
a_j \sim \begin{cases} \mathcal{N}(0, \lambda + \sigma^2) & \text{Proposition 1} \\ \mathcal{N}(\sqrt{2} \mathcal{N}(\mu_j), \sigma^2) & \text{Proposition 2} \end{cases} \quad b_j \sim \begin{cases} \mathcal{N}(0, \lambda + \sigma^2) & \text{Proposition 1} \\ \mathcal{N}(\sqrt{2} \mathcal{N}(\mu_j), \sigma^2) & \text{Proposition 2} \end{cases}
\]
Therefore, \( \|u\|^2 = \frac{1}{2} \sum_{j=1}^{nh} (a_j^2 + b_j^2) \) is a sum of \( 2nh \) independent squares of either mean centered or non-centered Gaussian random variables. This in turn gives
\[
V_0 = \|u\|^2 \sim \begin{cases} \frac{\sigma^2 + \lambda}{2} \chi_2^{2nh} & \text{Proposition 1} \\ \frac{\sigma^2 + \lambda}{2} \chi_2^{2nh} & \text{Proposition 2} \end{cases}
\]
Since given \( u, O \) is unitary and fixed, then \( u_j \sim \mathcal{CN}(0, I_{m-1}) \). Since this distribution is independent of \( u, v_j \sim \mathcal{CN}(0, I_{m-1}) \). By similar arguments
\[
V_2 = \|v_1\|^2 \sim \frac{1}{2} \chi_{2m-2}^2
\]
which is independent of $||u||^2$. Finally, conditioned on $(u, v_1)$, we have $v_1^T v_j \sim \mathcal{CN}(0, ||v_1||^2)$ and $|v_1^T v_j|^2 \sim ||v_1||^2 \chi_{2n_{H-2}}^2$. Thus,

$$D|(u, v_1) = \sum_{j=2}^{n_H} |v_1^T v_j|^2 ||u, v_1|| \sim \frac{||v_1||^2}{2} \chi_{2n_{H-2}}^2,$$

where the $\chi_{2n_{H-2}}^2$ variate is independent of $(u, v_1)$. We conclude that

$$V_4 \sim \left\{ \begin{array}{ll} \frac{1}{2} \left( x_{2n_{H}} \right)^{-1} \chi_{2n_{H-2}}^2 \chi_{2n_{H-2}}^2, & \text{Proposition 1}, \\ \frac{1}{2} \left( x_{2n_{H}} \right)^{-1} \chi_{2n_{H-2}}^2 \chi_{2n_{H-2}}^2, & \text{Proposition 2}. \end{array} \right.$$  

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 $E \sim \mathcal{CWN}(n_E, I)$ and $M = \begin{bmatrix} b \mid e \end{bmatrix} \in \mathbb{C}^{m \times 2}$, with the vector $b$ fixed and orthogonal to $e_1$. Define a $2 \times 2$ diagonal matrix $D = \text{diag}(1, 1/||b||^2)$. Then

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

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

$$S_{11} \sim \frac{2}{\chi_{2n_{E}-2m+2}^2}, \quad S_{22} \sim \frac{\chi_{2n_{E}-2m+4}^2}{2||b||^2}.$$  

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

$$\mathbb{E}\left( e_1^T E^{-1} A_2 E^{-1} e_1 \right) = \frac{m - 1}{(n_E - m)(n_E - m + 1)}.$$  

**Proof of Propositions 3 and 4.** Without loss of generality we may assume that the signal direction is $v = e_1$. Hence

$$H \sim \begin{cases} \mathcal{CWN}(n_H, I + \lambda e_1 e_1^T), & \text{Proposition 3}, \\ \mathcal{CWN}(n_H, I, \omega e_1 e_1^T), & \text{Proposition 4}. \end{cases}$$

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 $X^T X$ with $X = [x_1, \ldots, x_{n_H}]$ where each $x_j$ follows (A.1) but now with

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

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

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

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

$$A_0 = \begin{pmatrix} z & 0 \\ 0 & 0 \end{pmatrix}, \quad A_1 = \sqrt{z} \begin{pmatrix} 0 & b^T \\ b & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 \\ 0 & z \end{pmatrix},$$

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

$$E_{11} = e_1^T E^{-1} e_1, \quad \hat{b} = \begin{pmatrix} 0 \\ b \end{pmatrix}, \quad E^{b^T} = \hat{b}^T E^{-1} e_1, \quad E^{b\hat{b}} = \hat{b}^T E^{-1} \hat{b}.$$
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 \(E^{-1}H\) is similar to the Hermitian matrix \(E^{-1/2}H_eE^{-1/2}\). Therefore, all its eigenvalues are real-valued for any value of \(\epsilon\). Furthermore, since \(E^{-1/2}H_eE^{-1/2}\) is a holomorphic symmetric function of \(\epsilon\), it follows from Kato ([25], Theorem 6.1 page 120) that the largest eigenvalue \(\ell_1\) and its eigenprojection \(P(\epsilon)\) are analytic functions of \(\epsilon\) in some neighborhood of zero, where the largest eigenvalue has multiplicity one. The projection to the corresponding eigenspace of \(E^{-1}H\) is \(P(\epsilon) = E^{-1/2}P(\epsilon)E\). As the matrix \(E\) does not depend on \(\epsilon\), this projection is also an analytic function in some neighborhood of \(\epsilon = 0\).

At \(\epsilon = 0\), \(E^{-1}e_1\) is an eigenvector with eigenvalue \(E^{1/2}z\), that is,
\[
E^{-1}H_0E^{-1}e_1 = zE^{-1}e_1e_1^\dagger E^{-1}e_1 = zE^{1/2}E^{-1}e_1, 
\]
from which we obtain
\[
e_1^\dagger P(0)E^{-1}e_1 = e_1^\dagger E^{-1}e_1 = E^{11}. \quad (A.6)
\]
Since \(P(\epsilon)\) is an analytic function of \(\epsilon\) and the inner product is a smooth function, then there exists a neighborhood of \(\epsilon = 0\) where \(e_1^\dagger P(\epsilon)E^{-1}e_1, e_1\) is both analytic in \(\epsilon\) and strictly positive. In this neighborhood, we may define
\[
v_1(\epsilon) = \frac{E^{11}}{e_1^\dagger P(\epsilon)E^{-1}e_1}P(\epsilon)E^{-1}e_1. \quad (A.7)
\]
Clearly \(v_1(\epsilon)\) is the eigenvector corresponding to the eigenvalue \(\ell_1(\epsilon)\) and it is also analytic. We thus expand
\[
\ell_1(\epsilon) = \sum_{j=0}^\infty \lambda_je^j, \quad v_1(\epsilon) = \sum_{j=0}^\infty w_je^j. \quad (A.8)
\]
Inserting these expansions into the eigenvalue–eigenvector equations \(E^{-1}H_v = \ell_1v_1\) gives the following equations: at the \(O(1)\) level,
\[
E^{-1}A_0w_0 = \lambda_0w_0
\]
whose solution is
\[
\lambda_0 = zE^{11}, \quad w_0 = \text{const} \cdot E^{-1}e_1. \quad (A.9)
\]
By (A.6)–(A.7), \(w_0 = v_1(0) = E^{-1}e_1\), so the above constant is one.

By (A.7), \(e_1^\dagger v_1(\epsilon) = E^{11} = e_1^\dagger w_0\). Hence \(e_1^\dagger w_j = 0\) for all \(j \geq 1\). Furthermore, since \(A_0 = ze_1e_1^\dagger\), then \(A_0w_j = 0\) for all \(j \geq 1\). The \(O(\epsilon)\) equation is thus
\[
E^{-1}A_1w_0 + E^{-1}A_0w_1 = \lambda_1w_0 + \lambda_0w_1. \quad (A.10)
\]
However, \(A_0w_1 = 0\). Multiplying this equation by \(e_1^\dagger\) gives that
\[
\lambda_1 = \frac{e_1^\dagger E^{-1}w_0}{E^{11}} = \frac{\sqrt{z}}{E^{11}} \left\{ e_1^\dagger E^{-1} \left( \begin{array}{c} 0 \\ b^\dagger \\ 0 \end{array} \right) E^{-1}e_1 \right\} 
\]
\[
= \frac{\sqrt{z}}{E^{11}} \left\{ e_1^\dagger E^{-1} \left( \begin{array}{c} 0 \\ b^\dagger \\ 0 \end{array} \right) E^{-1}e_1 + e_1^\dagger E^{-1} \left( \begin{array}{c} 0 \\ b^\dagger \\ 0 \end{array} \right) E^{-1}e_1 \right\} = 2\sqrt{z}\|\{E^{b1}\}. \quad (A.11)
\]
Inserting the expression for \(\lambda_1\) into (A.10) gives that
\[
w_1 = \frac{1}{\sqrt{2}E^{11}} \left\{ E^{-1} \left( \begin{array}{c} 0 \\ b^\dagger \\ 0 \end{array} \right) E^{-1}e_1 - 2\|\{E^{b1}\}E^{-1}e_1 \right\} 
\]
\[
= \frac{1}{\sqrt{2}E^{11}} \left( E^{b1}E^{-1}e_1 + E^{11}E^{-1}b - 2\|\{E^{b1}\}E^{-1}e_1 \right) = \frac{1}{\sqrt{2}} \left( E^{-1}b - \frac{E^{b1}}{E^{11}}E^{-1}e_1 \right). 
\]
The next \(O(\epsilon^2)\) equation is
\[
E^{-1}A_2w_0 + E^{-1}A_1w_1 + E^{-1}A_0w_2 = \lambda_2w_0 + \lambda_1w_1 + \lambda_0w_2.
\]
Multiply this equation by \(e_1^\dagger\) and recall that \(A_0w_2 = 0\) and \(e_1^\dagger w_0 = E^{11}\) gives
\[
\lambda_2 = \frac{e_1^\dagger E^{-1}A_2E^{-1}e_1}{E^{11}} + \frac{e_1^\dagger E^{-1}A_1}{\sqrt{z}} (E^{-1}b - \frac{E^{b1}}{E^{11}}E^{-1}e_1) 
\]
\[
= \frac{e_1^\dagger E^{-1}A_2E^{-1}e_1}{E^{11}} + \frac{E^{11}E^{bb} + (E^{b1})^2 - 2E^{b1}\|\{E^{b1}\}E^{b1}}{E^{11}} = \frac{e_1^\dagger E^{-1}A_2E^{-1}e_1}{E^{11}} + \frac{E^{11}E^{bb} - E^{b1}E^{b1}}{E^{11}}. \quad (A.12)
\]
Combining (A.9)–(A.12), we obtain the following approximate stochastic representation for the largest eigenvalue $\ell_1$ of $E^{-1}H_e$

$$\ell_1(\epsilon) = z\ell_{11} + 2\epsilon \sqrt{2\mathbb{N}(E^{b1})} + \epsilon^2 \frac{e^1E^{-1}A_2E^{-1}e_1}{E^{11}} + \epsilon^2 \frac{E^{11}E^{bb} - E^{b1}E^{bt}}{E^{11}} + O_p(\epsilon^3). \quad (A.13)$$

Next, to derive the approximate distribution of $\ell_1$ corresponding to the above equation, we study a $2 \times 2$ Hermitian matrix $S$, whose inverse is defined by

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

where $M = [e_1, \hat{b}]$ is a $2 \times 2$ matrix. Inverting this matrix gives

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

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

$$\ell_1(\epsilon) = z\ell_{11} + 2\epsilon \sqrt{2\mathbb{N}(E^{b1})} + \epsilon^2 \frac{e^1E^{-1}A_2E^{-1}e_1}{E^{11}} + O_p(\epsilon^3). \quad (A.14)$$

To establish Propositions 3 and 4, we start from (A.14). We neglect the second term $T_1 = 2\epsilon \sqrt{2\mathbb{N}(E^{b1})}$ 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

$$\ell_1(\epsilon) \approx z\ell_{11} + \epsilon^2 \left\{ \frac{1}{\sqrt{2\epsilon}} + c(m, n_E) \right\},$$

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

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

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

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

and for Proposition 4

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

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

**Proof of Proposition 5.** By (8), the non-centrality parameter $\omega$ depends on the data only through $X^1X$. Conditioning on $X^1X$, following (7), we invoke Proposition 4 with the parameters $m = p, n_H = q$, and $n_E = n - q$ to obtain

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

Now the final result follows by integrating over the distribution of $(X^1X)_{11} \sim \frac{1}{2} \chi_{2n}^2$, and using the definition of $F_{a,b}(\epsilon, n)$ given in (9). \qed

**Proof of Propositions 6 and 7.** Let us assume without loss of generality that $v = e_1$. If $\hat{v}$ is not normalized, then we can write (10) as $R = |\hat{v}^i e_1|^2 / ||\hat{v}||^2$. From Lemma 1, we have

$$\hat{v} = w_0 + \sigma w_1 + \sigma^2 w_3 + \ldots$$

where

$$w_0 = e_1, \quad w_1 = \frac{1}{||u||} \begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \quad w_3 = \frac{1}{||u||^3} \begin{pmatrix} 0 \\ \sum_{j=2}^{n_H} v_j v_j^* v_1 \end{pmatrix},$$

with $\sigma^2 = \epsilon^2$. The last term in (10) then gives

$$\frac{\sigma^2}{||u||^2} \left( \sum_{j=2}^{n_H} v_j v_j^* v_1 \right) \rightarrow \frac{\epsilon^2}{2}.$$ 

The next four terms in (10) yield

$$\frac{\epsilon^2}{2} \left( \sum_{j=2}^{n_H} v_j v_j^* v_1 \right) \rightarrow \frac{\epsilon^2}{2}.$$ 

The following terms in (10) yield

$$\frac{\epsilon^2}{2} \left( \sum_{j=2}^{n_H} v_j v_j^* v_1 \right) \rightarrow \frac{\epsilon^2}{2}.$$ 

The remaining term gives

$$\frac{\epsilon^2}{2} \left( \sum_{j=2}^{n_H} v_j v_j^* v_1 \right) \rightarrow \frac{\epsilon^2}{2}.$$ 

The final term in (10) gives

$$\frac{\epsilon^2}{2} \left( \sum_{j=2}^{n_H} v_j v_j^* v_1 \right) \rightarrow \frac{\epsilon^2}{2}.$$ 

The final result follows by integrating over the distribution of $(X^1X)_{11} \sim \frac{1}{2} \chi_{2n}^2$, and using the definition of $F_{a,b}(\epsilon, n)$ given in (9). \qed
with i.i.d. variables $v_j \sim \mathcal{CN}(0, I_{m-1})$ all independent of $u \in \mathbb{C}^n$,
\[
  u \sim \begin{cases} 
    \mathcal{CN}(0, (\lambda + \sigma^2)I_m) & \text{for } H \sim \mathcal{CN}(\mu, \sigma^2I_m), \\
    \mathcal{CN}(\mu, \sigma^2I_m) & \text{for } H \sim \mathcal{CN}(\mu, \sigma^2I_m, (\omega/\sigma^2)e_1^2) 
  \end{cases}
\]
and $\|\mu\|^2 = \omega$. Therefore,
\[
  R = \frac{1}{1 + \sigma^2 \frac{\|v_1\|^2}{\|u\|^2} + 2\sigma^4 \sum_{j=2}^{n} \frac{|v_j'|^2}{\|u\|^4} + O_p(\sigma^6)}.
\]
The result follows from the distribution of these quantities. □

**Proof of Proposition 8.** Let us rewrite (A.8) as follows
\[
  \hat{v}(\epsilon) = w_0 + w_1 \epsilon + O_p(\epsilon^2),
\]
where $w_0 = E^{-1}e_1$, $w_1 = \frac{1}{\sqrt{2}} \left( E^{-1}b - \frac{\text{E}^*}{E^{-1}} E^{-1}e_1 \right)$. For convenience, decompose the matrices $E$ and $E^{-1}$ as
\[
  E = \begin{pmatrix} E_{11} & E_{12} \\ E_{12} & E_{22} \end{pmatrix}, \quad E^{-1} = \begin{pmatrix} E_{11}^{-1} & E_{12}^{*} \\ E_{12} & E_{22}^{*} \end{pmatrix},
\]
where $E_{11} \in \mathbb{R}$, $E_{12} \in \mathbb{C}^{(m-1)\times1}$, and $E_{22} \in \mathbb{C}^{(m-1)\times(m-1)}$. Consequently, $E_{11} = 1/(E_{11} - E_{12}^{*}E_{22}^{-1}E_{12}) \in \mathbb{R}$ and $E_{12} = -E_{12}^{*}E_{22}^{-1}E_{12} \in \mathbb{C}^{(m-1)\times1}$. 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 $e_1w_1 = 0$, we have
\[
  \hat{v}_1 e_1 = E_{11}^{*} + O_p(\epsilon^2),
\]
from which we obtain
\[
  \|\hat{v}_1 e_1\|^2 = (E_{11}^{*})^2 + O_p(\epsilon^2).
\]
The denominator of $R$ can be written as
\[
  \|\hat{v}\|^2 = \|e_1 (E_{11}^{*})^2 e_1 + \frac{2}{\sqrt{2}} \Re \{ e_1 \Re (E_{11}^{*})^2 e_1 \} \epsilon - \frac{2}{\sqrt{2}} \Re (E_{11} \epsilon b + E_{11}^{*} E_{22} b) \epsilon - \frac{2}{\sqrt{2}} \Re (E_{11}^{*} + \|E_{12}\|^2) \epsilon + O_p(\epsilon^2).
\]
Using the decomposition of $E^{-1}$ given in (A.15), we get
\[
  \|\hat{v}\|^2 = (E_{11}^{*})^2 + \|E_{12}\|^2 + \frac{2}{\sqrt{2}} \Re (E_{11} E_{12}^{*} b + E_{11}^{*} E_{22} b) \epsilon - \frac{2}{\sqrt{2}} \Re (E_{11}^{*} + \|E_{12}\|^2) \epsilon + O_p(\epsilon^2).
\]
Now we can conveniently express $R$ as
\[
  R = \frac{(E_{11}^{*})^2}{(E_{11}^{*})^2 + \|E_{12}\|^2} + \frac{(E_{11}^{*})^2}{(E_{11}^{*})^2 + \|E_{12}\|^2} (P_E - Q_E) \epsilon + O_p(\epsilon^2),
\]
where
\[
  P_E = \frac{2}{\sqrt{2} \Re (E_{11}^{*} + \|E_{12}\|^2) \Re (E_{11}^{*} + \|E_{12}\|^2) \epsilon, \quad Q_E = \frac{2}{\sqrt{2} \Re (E_{11}^{*} + \|E_{12}\|^2) \Re (E_{11}^{*} + \|E_{12}\|^2) \epsilon.
\]

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 $U = \text{diag}(1, -1, \ldots, -1)$, is an orthogonal matrix. Thus, $H(-\epsilon) = U^T H(\epsilon) U$ has the same eigenvalues as $H(\epsilon)$. In particular, the largest eigenvalue $\ell_1$ and its corresponding eigenvector $v_1$ satisfy
\[
  \ell_1(-\epsilon) = \ell_1(\epsilon), \quad v_1(-\epsilon) = UV_1(\epsilon).
\]
Hence $\ell_1$ and the first component of $v_1$ are even functions of $\epsilon$ whereas the remaining components of $v_1$ are odd.
We decompose the matrix $H(\epsilon) = \sum_{j=1}^{n} x_j x_j^\dagger$ as

$$H(\epsilon) = \sum_{j=1}^{n} (u_j e_1 + \epsilon \xi_j^\dagger)(u_j e_1 + \epsilon \xi_j^\dagger)^\dagger = \sum_{j=1}^{n} |u_j|^2 e_j e_j^\dagger + \epsilon \sum_{j=1}^{n} [\xi_j^\dagger \cdot \mathbb{I} e_j^\dagger + u_j e_1 \cdot \xi_j^\dagger] + \epsilon^2 \sum_{j=1}^{n} [\xi_j^\dagger \cdot \xi_j^\dagger]\,$$

$$= \begin{pmatrix} z & 0 \\ 0 & 0 \end{pmatrix} + \epsilon \sqrt{z} \begin{pmatrix} 0 & b^\dagger \\ b & 0 \end{pmatrix} + \epsilon^2 \begin{pmatrix} 0 & 0 \\ 0 & Z \end{pmatrix} = A_0 + \epsilon A_1 + \epsilon^2 A_2,$$

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:

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

Also, the eigenprojection $p(\epsilon)$ of $\ell_1$ satisfies

$$v_1(\epsilon) = \frac{1}{e_1 p(\epsilon) e_1} p(\epsilon) e_1. \quad (B.3)$$

Inserting the expansions (B.2) into the eigenvalue equation $Hv_1 = \ell_1 v_1$ gives the following set of equations for $r \geq 0$

$$A_0 w_r + A_1 w_{r-1} + A_2 w_{r-2} = \lambda_0 w_r + \lambda_2 w_{r-2} + \lambda_4 w_{r-4} + \cdots \quad (B.4)$$

with the convention that vectors with negative subscripts are zero. From the $r = 0$ equation, $A_0 w_0 = \lambda_0 w_0$, we readily find that

$$\lambda_0 = z, \quad w_0 = \text{const} \cdot e_1.$$  

Eq. (B.3) implies that $e_1 v_1 = 1$ and $w_0 = v_1(0) = e_1$. This implies that $w_j$, for $j \geq 1$, is orthogonal to $e_1$, 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 $r = 3$

$$A_1 w_0 = \lambda_0 w_1, \quad A_2 w_1 = \lambda_0 w_3 + \lambda_2 w_1 \quad (B.5)$$

from which we obtain

$$w_1 = z^{-1/2} b, \quad w_3 = \lambda_0^{-1} (A_2 - \lambda_0 l) w_1. \quad (B.6)$$

Multiply (B.4) on the left by $w_{0}^\dagger$ and use the first equation of (B.5) to obtain, for $r$ even, 

$$\lambda_2 = (A_1 w_0)^\dagger w_{r-1} = \lambda_0 w_{r}^\dagger w_{r-1}$$

and hence

$$\lambda_2 = \lambda_0 w_{r}^\dagger w_{r-1} = b^\dagger b \quad \text{and} \quad \lambda_4 = w_{r}^\dagger (A_2 - \lambda_0 l) w_1 = z^{-1/2} b^\dagger (Z - bb^\dagger) b.$$

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

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

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 $A \sim CW_m(n, \Sigma)$ with $n > m - 1$ where $A$ and $\Sigma$ are partitioned as follows

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

and let $A_{11,2} = A_{11} - A_{12} A_{22}^{-1} A_{21}$, and $\Sigma_{11,2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$. Then, $A_{11,2}$ is distributed as $CW_k(n - m + k, \Sigma_{11,2})$ and is independent of $A_{12}, A_{21}$ and $A_{22}$.

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

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

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

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, $\Sigma_{11,2}^{-1} = C_{11}$. 
Following [15,19], the density of $A$ is given by
\[ f(A) = \frac{\det^{n-m}(A)}{\Gamma_m(n)\det^n(\Sigma)} e^{-\text{tr}(\Sigma^{-1}A)} \]  
(B.8)
where $\text{tr}(\cdot)$ 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 $\text{tr}(\Sigma^{-1}A)$. First of all, we have that $\det(A) = \det(A_{11,2})\det(A_{11,2})$.

Next, we introduce a change of variables from the entries of the matrix $A$, to $A_{11,2} = A_{11} - A_{12}A_{22}^{-1}A_{21}, B_{12} = A_{12}, B_{22} = A_{22}$. 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 $dA = dA_{11}dA_{12}dA_{22} = dA_{11}dA_{12}dA_{22}$. Furthermore, using the expansion
\[ \text{tr}(\Sigma^{-1}A) = \text{tr} \left( \left( \begin{array}{cc} C_{11} & C_{12} \\ C_{21} & C_{22} \end{array} \right) \left( \begin{array}{cc} A_{11,2} + B_{12}B_{22}^{-1}B_{21} & B_{12} \\ B_{21} & B_{22} \end{array} \right) \right) \]
along with the fact that $B_{21} = B_{12}^t$ yields that
\[ f(A_{11,2}, B_{12}, B_{22}) = \frac{\det^{n-m}(B_{22})}{\det^n(\Sigma_{11,2})} \frac{\det^{n-m}(A_{11,2})}{\Gamma_m(n)} \times e^{-\text{tr}(\Sigma^{-1}_{11,2}A_{11,2})} \times e^{-\text{tr}(\Sigma_{11,2}^{-1}B_{12}^tB_{22}^{-1}B_{21})} \times e^{-\text{tr}(\Sigma_{21}B_{12}) - \text{tr}(\Sigma_{11,2}B_{12}B_{22}^{-1}B_{21})}. \]

Now we may use the decomposition
\[ \Gamma_m(n) = \pi^{\frac{1}{2}(k-1)} \prod_{j=1}^{k} \Gamma(n - m + k - j + 1) \times \pi^{\frac{(m-k)(m-k-1)}{2}} \prod_{j=1}^{m-k} \Gamma(n - j + 1) \]

and rewrite (B.9) as
\[ f(A_{11,2}, B_{12}, B_{22}) = f_1(A_{11,2}) \times f_2(B_{12}, B_{22}), \]

where
\[ f_1(A_{11,2}) = \frac{\det^{n-m+k-k}(A_{11,2})}{\det^{n-m+k}(\Sigma_{11,2})} \frac{1}{\Gamma_k(n - m + k)} e^{-\text{tr}(\Sigma^{-1}_{11,2}A_{11,2})}, \]
and
\[ f_2(B_{12}, B_{22}) = \frac{\det^{n-m}(B_{22})}{\pi^{\frac{(m-k)(m-k-1)}{2}}} \prod_{j=1}^{m-k} \Gamma(n - j + 1) \times \det^{m-k}(\Sigma_{11,2}) \det^n(\Sigma_{22}) \times e^{-\text{tr}(\Sigma^{-1}_{11,2}B_{12}B_{22}^{-1}B_{21})} - \text{tr}(\Sigma_{11,2}B_{12}) - \text{tr}(\Sigma_{11,2}B_{12}B_{22}^{-1}B_{21}). \]

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

Proof of Claim 2. Set $B = \Sigma^{-1/2}A\Sigma^{-1/2}$. Now $B \sim CV_m(n, I)$. For $R = M\Sigma^{-1/2}(MA^{-1}M^t)^{-1} = (RB^{-1}R^t)^{-1}$ and $(M\Sigma^{-1/2}M^t)^{-1} = (RB^{-1}R^t)^{-1}$. Thus, it is sufficient to prove that $(RB^{-1}R^t)^{-1} \sim CV_k(n - m + k, (RB^{-1}R^t)^{-1})$. 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,
\[ (RB^{-1}R^t)^{-1} = ([I_k : 0]HB^{-1}H^t[I_k : 0]L^t)^{-1} = (L^{-1})^t([I_k : 0](HH^{-1})^{-1}[I_k : 0])^{-1}L^{-1}, \]
where $C = HBB^t \sim CV_m(n, I)$. Let
\[ F = C^{-1} = \left( \begin{array}{cc} F_{11} & F_{12} \\ F_{21} & F_{22} \end{array} \right), \quad C = \left( \begin{array}{cc} C_{11} & C_{12} \\ C_{21} & C_{22} \end{array} \right), \]
and
\[ F_{11} = [I_k : 0]HB^{-1}H^t[I_k : 0]L^t. \]
where \( F_{11} \) and \( C_{11} \) are \( k \times k \). Then \((RB^{-1}R^T)^{-1} = (L^{-1})F_{11}^{-1}L^{-1}\), and since \( F_{11}^{-1} = C_{11} - C_{12}C_{21}^{-1}C_{21} \), it follows from Claim 1 that \( F_{11}^{-1} \sim C\mathcal{V}_k(n - m + k, k) \). Hence \((L^{-1})F_{11}^{-1}L^{-1} \sim C\mathcal{V}_k(n - m + k, (LL^{-1})^{-1})\), and since \((LL^{-1})^{-1} = (RR^T)^{-1}\), the proof is complete. \( \square \)

**Proof of Lemma 2.** Note that \( S^{11} = E^{11} = e_1^Te^{-1}e_1 \). Then, by Claim 2, \((S^{11})^{-1} \sim C\mathcal{V}_1(n_F - m + 1, I_1) = \chi^2_{2n_F - 2m + 2}/2\), meaning \( S^{11} \sim 2/\chi^2_{2n_F - 2m + 2} \). Next, by definition \( S = (M^*E^{-1}M)^{-1}\), with fixed \( M \). Thus, by the same claim, \( S \sim C\mathcal{V}_2(n_F - m + 2, D) \) from which we obtain \( S_{22} \sim \chi^2_{2n_F - 2m + 4}/(2\|b\|^2) \). Finally, since \((S^{11})^{-1} = S_{11} - S_{12}S_{22}^{-1}S_{21}\), by Claim 1, \((S^{11})^{-1}\) is independent of \( S_{22}\). \( \square \)

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

\[
E\left(\frac{e_1^Te^{12}A_{1E}^{-1}e_1}{E^{11}}\right) = E_{A|E}\left[E_{A|E}\left(\frac{e_1^Te^{12}A_{1E}^{-1}e_1}{E^{11}}\right)\right].
\]

Next, since \( A_2 \) is independent of \( E\),

\[
E(A_2|E) = E(A_2) = \begin{pmatrix} 0 \\ 0 \\ I_{m-1} \end{pmatrix}.
\]

Combining the above two equations gives that

\[
E\left(\frac{e_1^Te^{12}A_{1E}^{-1}e_1}{E^{11}}\right) = E\left(\sum_{j=2}^{m} \frac{\|E^j\|^2}{E^{11}}\right) = (m - 1)E\left(\frac{\|E^{12}\|^2}{E^{11}}\right).
\]

To compute this expectation, consider the matrix \( S^{-1} = [e_1, e_2]^TE^{-1}[e_1, e_2] = \left(\begin{array}{cc} E^{11} & E^{21} \\ E^{21} & E^{22} \end{array}\right)\). Since \( S^{22} = E^{22} \) and \( S_{22} = E^{11}/(E^{11}E^{22} - \|E^{12}\|^2), \) we have

\[
S_{22} = \frac{E^{11}/E^{22} - \|E^{12}\|^2}{E^{11}}.
\]

(B.12)

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

\[
E\left(\frac{\|E^{12}\|^2}{E^{11}}\right) = E\left(E^{22}\right) - E\left(\frac{2}{\chi^2_{2n_F - 2m + 4}}\right) = \frac{1}{(n_F - m)(n_F - m + 1)}
\]

which completes the proof. \( \square \)

**References**


