

ON THE DISTRIBUTION OF ROY'S LARGEST ROOT TEST
IN MANOVA AND IN SIGNAL DETECTION IN NOISE

By

Boaz Nadler
Iain M. Johnstone

Technical Report No. 2011-04
May 2011

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065



ON THE DISTRIBUTION OF ROY'S LARGEST ROOT TEST
IN MANOVA AND IN SIGNAL DETECTION IN NOISE

By

Boaz Nadler
Weizmann Institute of Science, Israel

Iain M. Johnstone
Stanford University

Technical Report No. 2011-04
May 2011

**This research was supported in part by
National Science Foundation grant DMS 0906812 and
National Institutes of Health grant BIB R01 EB1988.**

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

<http://statistics.stanford.edu>

On the Distribution of Roy's Largest Root Test in MANOVA and in Signal Detection in Noise*

B. Nadler I.M. Johnstone

May 4, 2011

Abstract

Roy's largest root is a common test in multivariate analysis of variance (MANOVA), with applications in several other problems, such as signal detection in noise. In this paper, assuming multivariate Gaussian observations, we derive a simple yet accurate approximation for the distribution of Roy's largest root test, in the extreme case of concentrated non-centrality, where the signal or difference between groups is concentrated in a single direction. Our main result is that in the MANOVA setting, up to centering and scaling, Roy's largest root test approximately follows a non-central F distribution whereas in the signal detection application, it approximately follows a modified central F distribution (of the form $(s + \chi_a^2)/\chi_b^2$). Our results allow power calculations for Roy's test, as well as estimates of sample size required to detect a given (rank-one) group difference by this test, both of which are important quantities in hypothesis-driven research.

1 Introduction

In the analysis of multivariate data, hypothesis tests often play an important first and sometimes crucial step. Some classical examples of hypothesis testing problems include: i) testing the equality of p multivariate Gaussian distributions, and ii) testing independence between two sets of variables having joint multivariate Gaussian distribution with unknown mean. Of course, multivariate hypothesis testing problems are common in many other settings and applications. In this context, and of relevance to this paper, let us mention the signal processing literature whereby the fundamental task of detection of signals embedded in noise can also be cast as a hypothesis testing problem.

In his seminal work [24], S.N. Roy proposed the *union-intersection* principle to derive suitable test statistics for a variety of problems. For the above mentioned settings, the resulting statistic is the largest eigenvalue or characteristic root of a random matrix of the form $E^{-1}H$. This test statistic complements various other tests, such as Wilk's Lambda, Hotelling-Lawley trace and Pillai-Bartlett trace, all derived by different considerations, see [1].

Since the proposal of these various test statistics, significant efforts have been devoted to the study of their properties, in particular their distributions under the null and alternative hypotheses, and a comparison of their relative powers under different settings. Of course, accurate knowledge of a test's distribution under a given alternative is a valuable quantity as it enables the analytical computation of the test's power and thus of the minimal number of samples needed to detect a given alternative by this test.

For Wilk's Lambda, Hotelling-Lawley trace and Pillai-Bartlett trace tests, various accurate F approximations have been developed both for the null and for the non-null distributions. In

*This research was supported by the following grants: NSF DMS 0906812 and NIH BIB R01EB1988.

contrast, the derivation of a simple tractable approximation to the distribution of Roy's largest root test has remained a longstanding problem in multivariate analysis. To date, for dimension larger than two, no acceptable method has been developed for transforming Roy's largest root test statistic to an F or χ^2 statistic, and no straightforward method exists for computing powers for Roy's statistic itself [1, 13, 18].

In this paper, we aim to (partially) bridge this gap by presenting a simple yet accurate and easy-to-evaluate expression for the distribution of Roy's largest root test under a particular alternative, known as *concentrated non-centrality*. In the setting of MANOVA, this alternative hypothesis corresponds to an assumption that the mean responses of the different groups are concentrated along a *single* direction. Concentrated non-centrality can thus be viewed as a specific form of sparsity, indicating that the system under study can be described by relatively few parameters. This assumption leads to a rank one non-centrality matrix with a single non-zero eigenvalue. As we also discuss, our analysis equally applies to a similar and fundamental problem in signal processing, detection of a single signal embedded in additive Gaussian noise, when the noise covariance matrix is arbitrary and unknown.

In contrast to previous works that considered asymptotic expansions in the limit as sample size (or number of groups) tend to infinity, or in the joint limit as both sample size and dimension tend to infinity together, our approach keeps all of these parameters fixed. Instead, following our previous work [15], we study the behavior of the largest eigenvalue in the limit of large signal-to-noise ratio or a large non-centrality parameter, or equivalently as noise strength tends to zero. This analysis, using standard tools from matrix perturbation theory, yields an approximate stochastic representation for Roy's largest root test statistic, from which we can deduce its approximate distribution.

The main result of our analysis, summarized in Propositions I and II in Section 4 is that in the MANOVA setting, after suitable centering and scaling, the distribution of Roy's largest root test approximately follows a non-central F distribution. In contrast, in the signal processing application, Roy's test is approximately distributed as a *modified* central F , of the form $(s + \chi_a^2)/\chi_b^2$.

As a side result of our analysis, we also derive a novel approximate expression for the distribution of the largest root of a single covariance matrix H , when it contains a single significant one-dimensional structure, see Lemma (h) in Section 4. This result reveals that for small sample size the resulting distribution may significantly deviate from the classical asymptotic Gaussian approximation. Therefore, this result may be of independent interest to the statistical community.

The paper is organized as follows: In Section 2 two different hypothesis testing problems are described, the first is the problem of signal detection in noise and the second is the classical MANOVA problem of testing for equality of means of p multivariate Gaussian populations. Section 3 contains a description of Roy's largest root test and a review of previous results. The main results of the paper are stated in Section 4, with the proofs appearing in Section 5 and in the appendix. Section 6 contains simulation results. We conclude with a summary and discussion in Section 7.

Some preliminary results of this paper, mainly in the context of signal detection in noise, were presented in the 2011 IEEE Statistical Signal Processing conference [17].

2 Problem Setup

We consider two different settings, one from multivariate statistics and the other from signal processing. While the observed samples in these two settings have different distributions, statistical interpretations and uses, they both lend themselves to a similar analysis. We first describe the signal processing application and then the classical MANOVA setting.

Notation: We denote column vectors by boldface lowercase letters, as in \mathbf{v} or \mathbf{a} , whereas their row transpose is \mathbf{v}^T . The dot product between two vectors is then $\mathbf{v}^T \mathbf{a}$, whereas the Euclidean norm of \mathbf{v} is $\|\mathbf{v}\| = (\mathbf{v}^T \mathbf{v})^{1/2}$. Matrices are denoted by capital Latin letters as in E or H . Finally, $\|A\|$ denotes the spectral norm of A .

2.1 Signal Detection with an Arbitrary Noise Covariance Matrix

Consider a measurement device consisting of m sensors (antennas, microphones, etc). In the signal processing literature, see for example [10], a standard model for the observed samples in the presence of a *single* emitting signal is

$$\mathbf{x} = \sqrt{\rho_s} u \mathbf{h} + \sigma \boldsymbol{\xi} \quad (1)$$

where \mathbf{h} is an unknown channel vector, assumed to be fixed during the measurement time window, u is a random variable distributed $\mathcal{N}(0, 1)$, ρ_s is the signal strength, σ is the noise level and $\boldsymbol{\xi}$ is a random noise vector that follows a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. If $\rho_s = 0$ the observations contain only noise, whereas if $\rho_s > 0$ a signal is present. In this paper, for the sake of simplicity, we assume real valued signals and noise. The case of complex valued signals and noise can be handled in a similar manner.

Let $\mathbf{x}_i \in \mathbb{R}^m$, for $i = 1, \dots, n_H$, denote n_H i.i.d. observations from the assumed "signal plus noise" model, Eq. (1), and let H denote their sample covariance matrix,

$$H = \frac{1}{n_H} \sum_{i=1}^{n_H} \mathbf{x}_i \mathbf{x}_i^T. \quad (2)$$

A fundamental problem in statistical signal processing is to distinguish, given observed data, between the following two hypotheses,

$$\mathcal{H}_0 : \text{no signal present, } \rho_s = 0, \quad \text{vs.} \quad \mathcal{H}_1 : \text{signal present, } \rho_s > 0. \quad (3)$$

If the noise covariance matrix $\boldsymbol{\Sigma}$ is known, the observed data can be *whitened* by the transformation $\boldsymbol{\Sigma}^{-1/2} \mathbf{x}_i$. Various methods to detect the presence of a signal, based on the eigenvalues of the whitened matrix $\boldsymbol{\Sigma}^{-1} H$ can be employed.

In this paper we consider the case where the noise covariance matrix $\boldsymbol{\Sigma}$ is *arbitrary and unknown*, but we have at our disposal an additional independent set $\{\mathbf{z}_j\}_{j=1}^{n_E}$ of n_E noise-only i.i.d. observations of the form

$$\mathbf{z} = \sigma \boldsymbol{\xi}.$$

This setting is plausible in several communication systems whereby such noise-only observations can be collected in time slots at which it is a-priori known that no signals are transmitting.

Here, a standard method for detecting the presence of a signal is to form an estimate of the noise covariance matrix,

$$E = \frac{1}{n_E} \sum_{i=1}^{n_E} \mathbf{z}_i \mathbf{z}_i^T \quad (4)$$

and devise signal detection schemes based on the eigenvalues of $E^{-1} H$ (instead of the unknown $\boldsymbol{\Sigma}^{-1} H$).

Zhao *et. al.* [27] were amongst the first to study signal detection in this setting of noise-only and signal-plus-noise observations. Considering a more general scenario whereby several sources may be present, [27] proposed a source enumeration method based on all eigenvalues of $E^{-1} H$, with the number of signals determined by an information theoretic criteria. Subsequently, several other works considered this detection problem, typically making structural assumptions on the unknown covariance matrix $\boldsymbol{\Sigma}$, see for example [28, 26]. More recently, [21] suggested a much improved estimator for the number of sources, by sequentially testing the significance of the largest eigenvalues of $E^{-1} H$, e.g., by performing an iterative version of Roy's largest root test.

2.2 Multivariate Analysis of Variance (MANOVA)

In multivariate analysis of variance, there are several settings that give rise to a similar hypothesis testing problem. For the sake of completeness we briefly describe and latter analyze perhaps the most common of them. For each of p groups, $i = 1, \dots, p$, we are given n_i independent multivariate Gaussian observations, denoted as $\mathbf{y}_j^{(i)} \in \mathbb{R}^m$, $j = 1, \dots, n_i$. The model we consider is that inside class i , the observations are of the form

$$\mathbf{y} = \boldsymbol{\mu}^{(i)} + \sigma \boldsymbol{\xi}$$

where $\boldsymbol{\mu}^{(i)}$ is an unknown mean response vector characteristic of group i , and $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is an unknown and arbitrary covariance matrix common to all groups.

Given the observed data $\{\mathbf{y}_j^{(i)}\}$, the problem here is to distinguish between the following two hypotheses: The null hypothesis \mathcal{H}_0 of equality between groups,

$$\mathcal{H}_0 : \text{no differences between groups, } \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} = \dots = \boldsymbol{\mu}^{(p)} \quad (5)$$

against the alternative \mathcal{H}_1 , that some differences between the groups do exist,

$$\mathcal{H}_1 : \text{not all } \boldsymbol{\mu}^{(i)} \text{ are equal.} \quad (6)$$

To test these hypotheses, we form the between classes and the within classes matrices, both of size $m \times m$,

$$H = \frac{1}{p-1} \sum_{i=1}^p n_i (\bar{\mathbf{y}}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^T \quad (7)$$

and

$$E = \frac{1}{n-p} \sum_{i=1}^p \sum_{j=1}^{n_i} (\mathbf{y}_j^{(i)} - \bar{\mathbf{y}}_i)(\mathbf{y}_j^{(i)} - \bar{\mathbf{y}}_i)^T \quad (8)$$

where $n = \sum_i n_i$ is the total number of samples, $\bar{\mathbf{y}}^i$ is the sample mean inside group i ,

$$\bar{\mathbf{y}}^i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{y}_j^{(i)}$$

and $\bar{\mathbf{y}}$ is the overall mean.

This hypothesis testing problem is fundamental in multivariate statistics. Several test statistics to distinguish between \mathcal{H}_0 and \mathcal{H}_1 have been suggested, all based on the eigenvalues of the matrix $E^{-1}H$. The four most common ones are Wilk's Lambda test, Hotelling-Lawley trace, Pillai-Bartlett trace, and Roy's largest root test, see for example [1, 14, 11] or essentially any other book on multivariate statistics. The distributions of the first three test statistics, both under the null and under the alternative have been thoroughly studied, and relatively accurate expressions in terms of F or χ^2 distributions have been derived. However, as summarized in [12, 18], only limited distributional results for the fourth test, namely Roy's largest root test, have been published in the literature. Via extensive simulations [22], Olson concluded that in the case of "concentrated non-centrality", where the trace of the non-centrality matrix is concentrated in its largest eigenvalue, Roy's largest root test has the most power. The main goal of this paper is to derive approximate yet accurate distributional results for Roy's largest root test in this setting. In particular, we consider the most concentrated case possible, where the non-centrality matrix is of rank one. The study of the distribution of Roy's largest root test under less restrictive assumptions will be described in future work.

Remark: Our definition of the matrices H and E in the MANOVA case, Eqs. (7) and (8), differs slightly from the common practice in the statistical literature as it includes a division by the factors $(p-1)$ and $(n-p)$, respectively. The reason for doing so is to allow a unified treatment of the distribution of Roy's largest root test both in the MANOVA case and in the signal detection setting, where the relevant matrices do typically include such normalization factors, see Eqs. (2) and (4).

2.3 Equivalence of the two models under the null

Before we embark on our analysis, we note that under the null hypothesis, the two models are equivalent. In the MANOVA setting, under \mathcal{H}_0 , both $(n-p)E$ and $(p-1)H$ follow a Wishart distribution, with $(n-p)$ and $(p-1)$ degrees of freedom and with covariance matrix Σ , respectively. Similarly, in the signal detection case the two matrices $n_E E$ and $n_H H$ are also Wishart distributed with n_E and n_H degrees of freedom and covariance Σ . Therefore, under the null the two settings are equivalent, with the one-to-one correspondence

$$n_E = n - p \quad \text{and} \quad n_H = p - 1. \quad (9)$$

In our analysis we shall thus use the notation n_H and n_E , though in the statement of the main results we distinguish between the two cases, see Propositions I and II below.

2.4 Distributions Under the Alternative

The focus of this paper is on the distribution of the largest eigenvalue $\ell_1(E^{-1}H)$ under the alternative \mathcal{H}_1 . While the matrix $n_E E$ still follows a Wishart distribution with covariance matrix Σ , the matrix H has a different distribution in the two settings described above. As we describe explicitly below, this leads to a different distribution for $\ell_1(E^{-1}H)$, which is the quantity of interest.

In the signal detection setting, since only a single signal is present under \mathcal{H}_1 , the matrix $n_H H$ follows a Wishart distribution with a covariance matrix having a single spike, $W_m(n_H, \sigma^2 \Sigma + \rho_s \mathbf{h}\mathbf{h}^T)$. In the MANOVA case, in contrast, the matrix $(p-1)H$ follows a non-central Wishart distribution $W_m(p-1, \sigma^2 \Sigma, \Omega)$, where the non-centrality matrix Ω is given by

$$\Omega = \sigma^{-2} \Sigma^{-1} \sum_{i=1}^p (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})^T$$

with

$$\bar{\boldsymbol{\mu}} = \frac{1}{n} \sum_i n_i \boldsymbol{\mu}_i.$$

We refer the reader to Muirhead's book [11] for details regarding this distribution.

Rank One Assumption: The closest analogy to having a single signal present in the MANOVA setting is to assume that under the alternative, when a difference exists between the p groups, the mean responses of the different groups are all proportional to the *same* unknown vector $\boldsymbol{\mu}_0$, with each multiplied by a group dependent strength parameter. That is, we assume

$$\boldsymbol{\mu}_i = s_i \boldsymbol{\mu}_0.$$

where $\|\Sigma^{-1/2} \boldsymbol{\mu}_0\| = 1$. This yields a non-centrality matrix Ω which is of rank one.

In more detail, let $\bar{s} = \frac{1}{n} \sum_i n_i s_i$, and define

$$\delta = \sum_{i=1}^p n_i (s_i - \bar{s})^2 \quad (10)$$

then the non-centrality matrix is $\Omega = \frac{\delta}{\sigma^2} \Sigma^{-1} \boldsymbol{\mu}_0 \boldsymbol{\mu}_0^T$.

As mentioned above, in his simulation study [22], Olson termed this case as *concentrated non-centrality*. Given a constraint on the differences among the p groups, as captured by the trace of Ω , Olson found that the case of concentrated non-centrality is the most favorable scenario for detection. Furthermore, it can be shown that Roy's largest root test is asymptotically the optimal test in this setting. As such, our result allows to calculate, for any given false alarm rate, the lower bound on the number of samples needed to detect such differences. Of course, detection of a concentrated non-centrality effect by other test statistics, such as Hotelling-Lawley trace may require more samples.

3 Hypothesis Testing via Roy's Largest Root Test

Let $\ell_1 = \ell_1(E^{-1}H)$ denote the largest eigenvalue of $E^{-1}H$. To employ Roy's largest root test one typically sets a required false-alarm rate $\alpha \ll 1$, and then accepts the alternative hypothesis \mathcal{H}_1 if

$$\ell_1(E^{-1}H) > th(\alpha) \quad (11)$$

where $th(\alpha)$ is the corresponding threshold.

To set the threshold, the distribution of $\ell_1(E^{-1}H)$ under the null, or at least its right tail behavior, needs to be accurately known. To analyze the probability of detection or the power of the test,

$$P_D = \Pr [\ell_1(E^{-1}H) > th(\alpha) | \mathcal{H}_1]$$

the distribution of ℓ_1 under the alternative, and in particular its dependence on the various problem parameters needs to be understood.

Accurate and efficiently computable expressions for the distribution of the largest eigenvalue of $E^{-1}H$, under both the null and alternative hypotheses, have been an open problem in multivariate analysis for several decades. In principle, the distribution of the largest eigenvalue of $E^{-1}H$ has an exact representation in terms of a hypergeometric function of matrix argument. In certain cases this leads to a finite series of generalized Laguerre polynomials under the alternative or zonal polynomials under the null, see [8] for formulas and a discussion of the relevant references. However, unless all problem parameters m, n_E, n_H are small (say < 15), these formulas are difficult to evaluate numerically. Furthermore, a theoretical analysis of the distribution, and its dependence on the underlying parameters using this exact representation is also a challenging task. Under the null, fast and accurate expressions for the distribution of Roy's test have been recently devised, using the Pfaffian representations of the relevant hypergeometric functions, see [3].

A second line of research is to study the largest eigenvalue in the high dimensional setting. In [25], Silverstein proved that under the null hypothesis \mathcal{H}_0 , and in the joint limit as both $m, n_E, n_H \rightarrow \infty$ with their ratios converging to fixed constants, the largest eigenvalue of $E^{-1}H$ converges to a deterministic value, given by

$$b_2 = \left(\frac{1 + \sqrt{1 - (1 - \frac{m}{n_H})(1 - \frac{m}{n_E})}}{1 - \frac{m}{n_E}} \right)^2. \quad (12)$$

Note that if $n_E/m \rightarrow \infty$, which implies that $E \rightarrow I$, then as expected $b_2 \rightarrow (1 + \sqrt{m/n_H})^2$, which is the limit of the largest eigenvalue of H under the null.

Recently, using tools from random matrix theory, both Jiang [5] and Johnstone [8] proved that asymptotically in the limit as $n_H, n_E, m \rightarrow \infty$, the logarithm of the largest eigenvalue of $E^{-1}H$ follows a Tracy-Widom distribution, after appropriate scaling and centering.

Theorem: Let $W = \log(\frac{n_H}{n_E} \ell_1(E^{-1}H))$. Then under the null \mathcal{H}_0 and in the joint limit as $m, n_E, n_H \rightarrow \infty$ with their ratios converging to fixed constants,

$$\Pr \left[\frac{W - \mu_{TW}}{\sigma_{TW}} < s \right] \rightarrow F_1(s) \quad (13)$$

where $F_1(s)$ is the Tracy-Widom distribution of order one, and the centering and scaling constants are given by the following equations,

$$\mu_{TW} = 2 \log \tan \left(\frac{\varphi + \gamma}{2} \right) \quad (14)$$

$$\sigma_{TW}^3 = \frac{16}{(n_E + n_H - 1)^2} \frac{1}{\sin^2(\varphi + \gamma) \sin(\varphi) \sin(\gamma)} \quad (15)$$

with the angle parameters γ, φ given by

$$\sin^2\left(\frac{\gamma}{2}\right) = \frac{\min(m, n_H) - 1/2}{n_E + n_H - 1}, \quad \sin^2\left(\frac{\varphi}{2}\right) = \frac{\max(m, n_H) - 1/2}{n_E + n_H - 1}. \quad (16)$$

By inverting the Tracy-Widom distribution, Eq. (13) can be used to set the approximate threshold for Roy's test, for any required false alarm α . Specifically, let $s(\alpha) = F_1^{-1}(1 - \alpha)$, then the threshold for the largest eigenvalue is

$$th(\alpha) = \frac{n_E}{n_H} \exp(\mu_{TW} + \sigma_{TW}s(\alpha)). \quad (17)$$

4 On the Distribution of the Largest Root Test

The goal of this paper is to study, under the presence of a signal or of group differences (e.g., under the alternative hypothesis \mathcal{H}_1), the distribution of the largest eigenvalue of $E^{-1}H$, and its dependence on the largest eigenvalue of $\Sigma^{-1}H$, and the various parameters m , n_E and n_H .

As mentioned above, the derivation of a simple tractable approximation to the distribution of Roy's largest root test has been a longstanding problem in multivariate analysis. In particular, for dimension $m > 2$, no acceptable method has been developed for transforming Roy's test to an F or χ^2 statistic, and no straightforward method exists for computing powers for Roy's statistic itself [1, 18]. Our analysis sheds light on this failure, as we show that in the MANOVA setting, the largest eigenvalue indeed does not follow a standard F distribution, but rather a *non-central* one, only after proper scaling and centering.

Recently, Silverstein and Nadakuditi [21] studied the largest eigenvalue of $E^{-1}H$ in the joint limit as $m, n_E, n_H \rightarrow \infty$, under the alternative hypothesis of signals present, e.g., when the matrix H has a population covariance with a few spikes¹. As in the simpler case of principal component analysis with a spiked covariance matrix, for a signal to be detected by the largest eigenvalue, the underlying signal strength must be larger than some threshold. Otherwise, the signal is buried in noise and the largest eigenvalue converges to the limit of Eq. (12). In [21] the authors derived both an explicit expression for this threshold, as well as the deterministic limiting value for ℓ_1 when the underlying signal is sufficiently strong. The latter is given by

$$\ell_1 \rightarrow \frac{2c_H\mu_H}{2c_H + c_E \left(1 - c_H - \mu_H + \sqrt{f(\mu_H, c_H)}\right)} \quad (18)$$

where $c_E = m/n_E$, $c_H = m/n_H$,

$$f(\mu_H, c_H) = (\mu_H - (1 - \sqrt{c_H})^2)(\mu_H - (1 + \sqrt{c_H})^2)$$

and μ_H is the limit, as $m, n_H \rightarrow \infty$, of the largest eigenvalue of the whitened matrix $\sigma^{-2}\Sigma^{-1}H$, namely [2, 19, 15]

$$\mu_H = \frac{1}{\sigma^2}(\lambda_H + \sigma^2) \left(1 + \frac{m-1}{n_H} \frac{\sigma^2}{\lambda_H}\right). \quad (19)$$

In the above equation, $\lambda_H = \rho_s \|\Sigma^{-1}\mathbf{h}\|^2$ is the signal part of the spike in the whitened matrix.

It is instructive to consider the asymptotics of Eq. (18) when $\mu_H \gg (1 + \sqrt{c_H})^2$. We then have that

$$\ell_1 \rightarrow \mu_H \frac{1}{1 - c_E} + \frac{c_E}{(1 - c_E)^2} + O\left(\frac{1}{\mu_H}\right). \quad (20)$$

¹The analysis of [21] is in fact quite general and does not require Gaussian distributions. As described in their paper, due to universality and under appropriate regularity conditions, the deterministic limit of the largest eigenvalue holds for a large class of underlying noise and signal distributions.

We thus see that the largest eigenvalue of $E^{-1}H$ is *larger* than that of the matrix H itself, to leading order due to a multiplicative factor $1/(1 - c_E) > 1$ and to second order due to an additive constant $c_E/(1 - c_E)^2$ that is independent of μ_H .

Since the exact limiting expression in Eq. (18) was derived by analyzing the limiting Stieltjes transform of the spectral density of the matrix $E^{-1}H$ as both $m, n_E, n_H \rightarrow \infty$, its accuracy for finite values of m, n_E, n_H is unclear. Furthermore, although [21] suggested, by analogy to PCA, that asymptotically the largest eigenvalue may follow a Gaussian distribution, no expression for its asymptotic variance was derived. As such, while [21] is a major advancement, its results do not allow for calculations of power or probability of detection.

4.1 Main Results

In this paper we derive simple approximate expressions for the distribution of Roy's largest root test both for the classical MANOVA setting as well as for the signal detection problem described above. As a by-product, we provide a simple explanation for the emergence of the first two terms in Eq. (20), and an assessment of their accuracy for finite parameter values. Our analysis thus allows, in the case of concentrated non-centrality, a straightforward and simple computation of the approximate power of Roy's largest root test, as well as its analytic comparison to several alternative popular test statistics. Matlab code for the resulting approximate distributions and power of Roy's test is available at the author's website².

To present our results we shall first need to introduce the following distribution, which is a slight modification of the well known central F distribution.

Definition 1 *A random variable X follows a modified central F distribution, denoted $CF_{a,b,s}$, if it can be written as*

$$X = \frac{s + \chi_a^2/a}{\chi_b^2/b}. \quad (21)$$

where the two χ^2 random variables are independent.

Remarks: i) Note that when the shift parameter vanishes, $s = 0$, this random variable resorts to the classical central F . ii) We note that this modified F distribution has been considered previously in the signal processing literature, see Eqs. (8)-(9) in [23]. iii) The distribution of the modified F can be expressed as the following one-dimensional integral

$$\begin{aligned} \Pr[X < x] &= \Pr[s + \chi_a^2/a < x\chi_b^2/b] \\ &= 1 - \int_0^\infty P_b((bs + bt/a)/x) p_a(t) dt \end{aligned} \quad (22)$$

where $p_a(t)$ is the density of a χ_a^2 random variable, and P_b is the distribution of a χ_b^2 random variable. This integral and thus the distribution of X can be easily evaluated numerically. The density of X can be similarly evaluated numerically. It can also be written explicitly as a finite sum, see Eq. (9) in [23].

Our key results can be summarized by the following two propositions.

Proposition I: *Let $\ell_1(E^{-1}H)$ be the largest eigenvalue in the MANOVA setting with $\sigma = 1$, under the alternative hypothesis whereby $(p - 1)H$ follows a non-central Wishart distribution $W_m(p - 1, \mathbf{I}_m, \Omega)$, with a rank-one non-centrality matrix Ω whose non-zero root is δ . Then, in the asymptotic limit as $\delta \rightarrow \infty$, up to centering and scaling constants, Roy's largest root can be approximated by a non-central F distribution,*

$$\boxed{\ell_1 \approx c_1 F_{a,b}(\delta) + c_2}. \quad (23)$$

² <http://www.wisdom.weizmann.ac.il/~nadler>

Similarly,

Proposition II: *Let $\ell_1(E^{-1}H)$ be the largest eigenvalue in the signal detection setting, under the alternative hypothesis of a single Gaussian signal of strength λ_H and with $\sigma = 1$. Then, in the asymptotic limit as $\lambda_H \rightarrow \infty$, up to centering and scaling constants, Roy's largest root can be approximated by a modified central F distribution*

$$\boxed{\ell_1 \approx c_1 C F_{a,b,s} + c_2.} \quad (24)$$

The values of the parameters in the MANOVA setting are

$$a = p + m - 2, \quad b = n - p - m + 1, \quad (25)$$

and

$$c_1 = \frac{n-p}{n-p-m+1} \frac{p+m-2}{p-1}, \quad c_2 = \frac{m-1}{n-p} \frac{1}{(1-\frac{m}{n-p})(1-\frac{m+1}{n-p})}. \quad (26)$$

Similarly, in the signal processing case, the coefficients are given by

$$a = n_H, \quad b = n_E - m + 1, \quad s = \frac{m-1}{n_H(\lambda_H + 1)} \quad (27)$$

and

$$c_1 = \frac{n_E}{n_E - m + 1} (\lambda_H + 1), \quad c_2 = \frac{m-1}{n_E} \frac{1}{(1-\frac{m}{n_E})(1-\frac{m+1}{n_E})}. \quad (28)$$

Finally, let us say a few words about the accuracy of the approximation above. First recall that in the classical statistical literature the typical approach is to study the asymptotics of the random variable of interest as sample size $n \rightarrow \infty$. In this paper, in contrast, Eqs. (23) and (24) are derived based on a two term asymptotic expansion of the largest eigenvalue ℓ_1 , where the asymptotics are at *fixed* n_H, n_E, m but as $\lambda_H \rightarrow \infty$ (or equivalently as $\sigma \rightarrow 0$). The term $c_1 F$ is an approximation to the distribution of the first term in this expansion, whereas the constant c_2 is the approximation of the second term, which has a much smaller variance, by its mean. To leading order, the error incurred in the approximation of the first term is $O(\sqrt{m-1}/n_H)$. Since the variance of the second term is $O((m-1)/n_E^2)$, the overall error in the above expansions is $O(\frac{1}{\lambda_H}) + O(\sqrt{m-1}/\min(n_E, n_H))$. As shown in the simulation section, provided that the signal strength is sufficiently large, Eqs. (23) and (24) are quite accurate even for small dimension and sample size values.

4.2 A Matrix Perturbation Approach

Following our previous work [15], our approach to the analysis of Roy's largest root and to the proof of the two propositions is based on a matrix perturbation approach, considering the noise level σ as a small parameter. In our analysis, the dimension m of the observations, as well as the sample sizes n_E and n_H are all fixed. Therefore, rather than relying on asymptotic results from random matrix theory, we shall use well known results regarding the distribution of *finite* Wishart and inverse Wishart matrices.

First, note that since $E^{-1}H = (\Sigma^{-1}E)^{-1}(\Sigma^{-1}H)$, rather than analyzing the matrices E and H we can equivalently work with whitened matrices $\Sigma^{-1}E$ and $\Sigma^{-1}H$. In other words, for analysis purposes, we may assume that $n_E E$ follows a Wishart distribution $W_m(n_E, \mathbf{I}_m)$. Similarly, in the signal+noise setting, the matrix $n_H H$ is distributed as $W_m(n_H, \sigma^2 \mathbf{I}_m + \lambda_H \mathbf{v}\mathbf{v}^T)$, where $\|\mathbf{v}\| = 1$, namely a covariance matrix with a single spike. Similarly, in the MANOVA case the matrix $n_H H$ follows a noncentral Wishart distribution $W_m(n_H, \mathbf{I}_m, \Omega)$, where the non-centrality matrix Ω is of rank one and has unit spectral norm, that is $\Omega = \delta \mathbf{v}\mathbf{v}^T$ with $\|\mathbf{v}\| = 1$. For the two settings to have comparable large eigenvalue, we assume that $\delta = n_H \lambda_H$.

Next, rather than studying the non-symmetric matrix $E^{-1}H$, we work with the symmetric matrix $E^{-1/2}HE^{-1/2}$, which has the same eigenvalues as $E^{-1}H$. Let $h_1 \geq h_2 \geq \dots \geq h_m$ denote the eigenvalues of the matrix H , sorted in decreasing order of magnitude, and let $\{\mathbf{a}_i\}_{i=1}^m$, denote their corresponding unit-norm eigenvectors. Since the matrix H is symmetric, these constitute a (random) orthonormal basis of \mathbb{R}^m . We can thus write

$$H = \sum_{i=1}^m h_i \mathbf{a}_i \mathbf{a}_i^T \quad (29)$$

We consider the case where $\lambda_H \gg \sigma^2(1 + \sqrt{m/n_H})^2$. In this case, the largest eigenvalue $h_1 = O(\lambda_H)$ is substantially larger than the remaining eigenvalues h_2, \dots, h_m , which are all $O(\sigma^2)$. We thus write the matrix H as

$$H = H_0 + \sigma^2 H_1, \quad (30)$$

where $H_0 = h_1 \mathbf{a}_1 \mathbf{a}_1^T$ and $H_1 = \sum_{j=2}^m \tilde{h}_j \mathbf{a}_j \mathbf{a}_j^T$, where $\tilde{h}_j = h_j/\sigma^2 = O(1)$. In our study of the eigenvalues of $E^{-1/2}HE^{-1/2}$, we view the matrix $\sigma^2 E^{-1/2}H_1 E^{-1/2}$ as a perturbation of the matrix $E^{-1/2}H_0 E^{-1/2}$. Since both the unperturbed and the perturbed matrices are symmetric, standard results from matrix perturbation theory [9] imply that the largest eigenvalue $\ell_1(E^{-1/2}HE^{-1/2})$ is an analytic function of σ , for sufficiently small values of σ . Similar to the approach in [15], we expand the leading eigenvalue and corresponding eigenvector in a Taylor series,

$$\ell_1(\epsilon) = \lambda_0 + \epsilon \lambda_1 + \dots \quad (31)$$

$$\hat{\mathbf{v}}(\epsilon) = \mathbf{v}_0 + \epsilon \mathbf{v}_1 + \dots$$

where $\epsilon = \sigma^2$ is the small perturbation parameter. Our first result, proven in the appendix, is the following:

Theorem 1: The first two terms in Eq. (31) in the Taylor expansion as $\sigma \rightarrow 0$ of the largest eigenvalue of $E^{-1/2}HE^{-1/2}$ are given by

$$\lambda_0 = h_1 (\mathbf{a}_1^T E^{-1} \mathbf{a}_1) \quad (32)$$

and

$$\lambda_1 = \frac{\mathbf{a}_1^T E^{-1} H_1 E^{-1} \mathbf{a}_1}{\mathbf{a}_1^T E^{-1} \mathbf{a}_1} = \frac{1}{\mathbf{a}_1^T E^{-1} \mathbf{a}_1} \cdot \sum_{i=2}^m \tilde{h}_i (\mathbf{a}_1^T E^{-1} \mathbf{a}_i)^2 \quad (33)$$

The leading order term for the eigenvector is

$$\mathbf{v}_0 = E^{-1/2} \mathbf{a}_1. \quad (34)$$

Eqs. (32) and (33) provide an approximate stochastic representation for the largest eigenvalue $\ell_1(E^{-1}H)$, and reveal several interesting points. First, the leading order term in the largest eigenvalue of $E^{-1}H$ depends only on h_1 , the largest eigenvalue of H , whereas the next term depends on the remaining eigenvalues h_2, \dots, h_m but not on h_1 . Second, the leading order term depends also on an additional random variable $\mathbf{a}_1^T E^{-1} \mathbf{a}_1$, which, as we shall see below is *independent* of the random variable h_1 .

To prove the two propositions, we first introduce some further notations. We denote by

$$B = \mathbf{a}_1^T E^{-1} \mathbf{a}_1 \quad (35)$$

and by

$$C_i = \frac{(\mathbf{a}_1^T E^{-1} \mathbf{a}_i)^2}{B} \quad (36)$$

In this notation, the first and second terms in the expansion of the leading eigenvalue are simply

$$\lambda_0 = h_1 B, \quad \text{and} \quad \lambda_1 = \sum_{i=2}^m \tilde{h}_i C_i. \quad (37)$$

Next, we make use of the following auxiliary lemmas regarding the distributions of these random variables. The first lemma below provides the *exact* distribution of the random variable B :

Lemma (B): *Let $n_E E$ be a random Wishart matrix with identity covariance, and let $\{\mathbf{a}_i\}_{i=1}^m$ be either a deterministic orthonormal basis of \mathbb{R}^m , or a random orthonormal basis that is independent of E . Then for the random variable B defined in Eq.(35), we have*

$$1/B \sim \frac{1}{n_E} \chi_{n_E - m + 1}^2. \quad (38)$$

In particular, for $n_E > m + 3$

$$\mathbb{E}[B] = \frac{n_E}{n_E - m - 1} \quad (39)$$

and

$$\text{Var}[B] = \frac{2n_E^2}{(n_E - m - 1)^2(n_E - m - 3)}. \quad (40)$$

The following lemma characterizes the mean and variance of the random variables C_i . Obviously, they all have the same distribution.

Lemma (C): *Let $n_E E$ follow a Wishart distribution $W_m(n_E, \mathbf{I})$, and let $\{\mathbf{a}_i\}_{i=1}^m$ be either a deterministic orthonormal basis of \mathbb{R}^m , or a random orthonormal basis that is independent of E . Then,*

$$\mathbb{E}[C_i] = \frac{n_E}{(n_E - m)(n_E - m - 1)} = O\left(\frac{1}{n_E}\right). \quad (41)$$

Furthermore, asymptotically in n_E ,

$$\text{Var}[C_i] = O\left(\frac{1}{n_E^2}\right). \quad (42)$$

Our final ingredient is an approximate expression for the distribution of h_1 , the largest eigenvalue of the sample covariance matrix H . Recall that in the signal detection case the matrix H follows a Wishart distribution with a single spike of strength λ_H whereas in the MANOVA setting it follows a non-central Wishart distribution with a rank-one noncentrality matrix. In general, the distribution of roots of sample covariance matrices is a classical subject, see [1, 11]. For the case of non-central Wishart matrices, the asymptotic expansion of the various roots including the largest one has been studied in [4]. However, these results are asymptotic as sample size (or number of groups) tend to infinity. The following lemma, in contrast, provides an accurate approximation even for small sample sizes n_H , as it is asymptotic as the signal-to-noise ratio tends to infinity, instead. Although relatively simple to derive, to the best of our knowledge, the results of this lemma are new. Moreover, as shown in the simulation section, they provide much more accurate expressions for the distribution of the largest root as compared to the classical Gaussian approximation, in particular for small sample sizes.

Lemma (h_1): *Let h_1 denote the largest eigenvalue of the sample covariance matrix H . Then, in the MANOVA case, where $n_H H$ follows a non-central Wishart distribution $W_m(n_H, \sigma^2 \mathbf{I}, \Omega)$ with a rank one non-centrality matrix of the form $\Omega = \delta \mathbf{v} \mathbf{v}^T$, with $\|\mathbf{v}\| = 1$, asymptotically as $\sigma \rightarrow 0$,*

$$h_1 \sim \frac{\sigma^2}{n_H} \chi_{n_H + m - 1}^2 \left(\frac{\delta}{\sigma^2} \right) + o(\sigma^2) \quad (43)$$

In contrast, in the signal processing setting, with $H \sim W_m(n_H, \sigma^2 I + \lambda_H \mathbf{v}\mathbf{v}^T)$,

$$h_1 \sim \frac{1}{n_H} \cdot ((\lambda_H + \sigma^2)\chi_{n_H}^2 + \sigma^2\chi_{m-1}^2 + o(\sigma^2)) \quad (44)$$

where the two χ^2 random variables are independent in the equation above.

For our purposes, in particular when the dimension m is relatively small and $\lambda_H \gg \sigma^2$ it will suffice to approximate the χ_{m-1}^2 random variable in Eq. (44) by its mean value, equal to $m-1$. In general, one may instead approximate the weighted sum of the two independent χ^2 random variables by a single scaled χ^2 , though we do not explore this here. With this approximation, in the signal processing case the asymptotic distribution of the largest eigenvalue of H , up to scaling and centering, is distributed as a central χ^2 , whereas in the MANOVA setting it follows a non-central χ^2 distribution. When $\delta = \lambda_H n_H$, the largest eigenvalue has the same mean in both settings. The key difference is in the *variance* of h_1 . Whereas in the signal processing case, with $\sigma = 1$

$$\text{Var}[h_1] = \frac{2}{n_H}(\lambda_H + 1)^2 + o\left(\frac{1}{n_H}\right) \quad (45)$$

in the MANOVA setting the variance is

$$\text{Var}[h_1] = \frac{2}{n_H}(2\lambda_H + 1) + o\left(\frac{1}{n_H}\right). \quad (46)$$

Therefore, when $\lambda_H \gg 1$, the fluctuations of h_1 in the MANOVA setting are significantly smaller.

Given the above expansion, we can now relatively easily compute the leading order mean and variance of the largest eigenvalue, as described by the following theorem.

Theorem 2: As $\sigma \rightarrow 0$, the leading order terms for the mean and variance of the largest eigenvalue of $E^{-1}H$ are given by

$$\mathbb{E}[\ell_1(E^{-1}H)] = \mathbb{E}[\lambda_0] + \sigma^2\mathbb{E}[\lambda_1] + o(\sigma^2) + t.s.t. \quad (47)$$

where *t.s.t.* stands for *transcendentally small terms* in σ , and

$$\mathbb{E}[\lambda_0] = \mathbb{E}[h_1] \cdot \frac{1}{1 - \frac{m+1}{n_E}}, \quad (48)$$

$$\mathbb{E}[\lambda_1] = \frac{m-1}{n_E} \cdot \frac{1}{1 - \frac{m}{n_E}} \cdot \frac{1}{1 - \frac{m+1}{n_E}} \left(1 + O\left(\frac{1}{n_H}\right)\right). \quad (49)$$

As for the variance, to leading order

$$\begin{aligned} \text{Var}[\ell_1(E^{-1}H)] &= \frac{n_E^2}{(n_E - m - 1)(n_E - m - 3)} \left[\text{Var}[h_1] + \mathbb{E}[h_1]^2 \frac{2}{n_E - m - 1} \right] \\ &\quad + 2\sigma^2\mathbb{E}[h_1] \frac{(m-1)(m+3)n_E}{(n_E - m)(n_E - m - 1)(n_E - m - 3)} \end{aligned} \quad (50)$$

Remark: The above expressions, Eqs. (48)-(49) with fixed m, n_E and n_H , for the means of the first two terms in the Taylor expansion of the largest eigenvalue, shed new light on the limiting formula (18) derived in [21]. In particular, $\mathbb{E}[\lambda_0]$ is identical, up to a correction factor $O(1/n_E)$, to the first term in (20). Similarly, when all of m, n_E, n_H are large, and in particular in the limit as they all tend to infinity, $\mathbb{E}[\lambda_1] \rightarrow c_E/(1 - c_E)^2$ which is the second term in the asymptotic expansion (20).

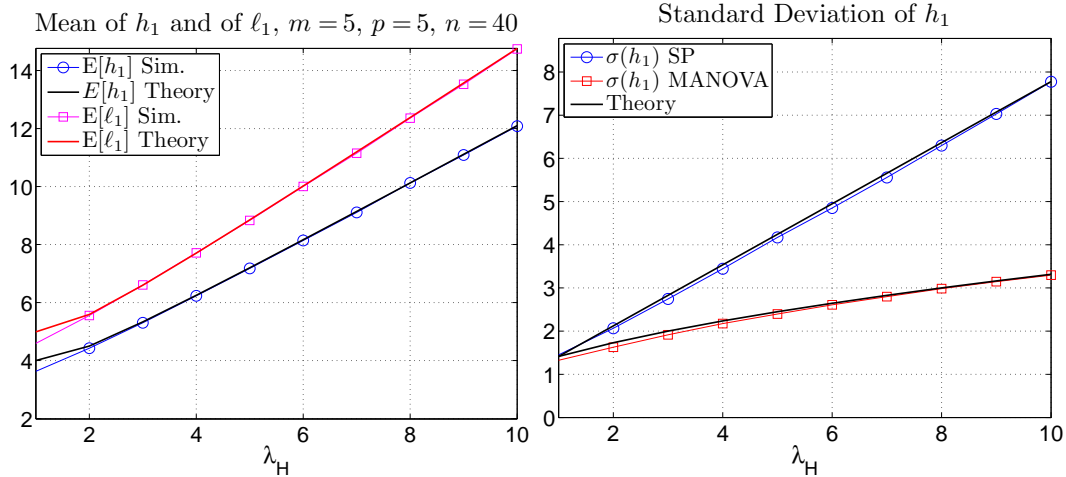


Figure 1: Mean and variance of the largest eigenvalue of H and of $E^{-1}H$ in both the signal processing setting (SP) and in MANOVA.

5 Proof of Propositions

Given the substantial preparations above, the two propositions 1 and 2 follow almost immediately. First, according to theorem 1, the largest eigenvalue $\ell_1(E^{-1}H)$ admits the form

$$\ell_1 = h_1 B + \sum_{j=2}^m h_j C_j + o(\sigma^2).$$

Next, note that since H and E are independent random matrices, and since the distribution of E is invariant to unitary transformations, it follows that h_1 and B are independent random variables. Furthermore, the distributions of h_1 and of B are characterized by the two lemmas above. Thus, the distribution of $h_1 B$ approximately follows a modified F distribution in the signal detection case and a non-central F in the MANOVA setting. Approximating the term $\sum_{j=2}^m h_j C_j$ by its mean value concludes the proof. \square

6 Simulations

We present a series of simulations that support our theoretical analysis and illustrate the accuracy of our approximations. For different signal strengths we make 150,000 independent random realizations of the two matrices E and H , and record the largest eigenvalue ℓ_1 . First, in the left panel of Fig. 1 we compare the empirical mean of both h_1 and of $\ell_1(E^{-1}H)$ to the theoretical formulas, Eqs. (10) and (47), as a function of λ_H . Next, in the right panel we compare the standard deviation $\sqrt{\text{Var}[h_1]}$ for both the MANOVA and the signal processing case to the theoretical formulas, Eqs (45) and (46), respectively. Finally, in Fig. 2 we compare the standard deviation of Roy's largest root test in the two settings to the leading order term in Eq. (50). Note that even though in this simulation all parameter values are small ($m = 5$ dimensions, $p = 5$ groups with $n_i = 8$ observations per group yielding a total of $n = 40$ samples), the fit between the simulations and theory is remarkably good.

Next, at the same parameter values, and with $\lambda_H = 10$, we compare the empirical density of h_1 both in the MANOVA and in the signal detection cases to the theoretical formulas, Eqs. (43) and (44), respectively. As shown in figure 3, the theoretical approximation is remarkably accurate, and far more accurate than the classical asymptotic Gaussian approximation.

Standard deviation of ℓ_1 , Roy's Largest Root

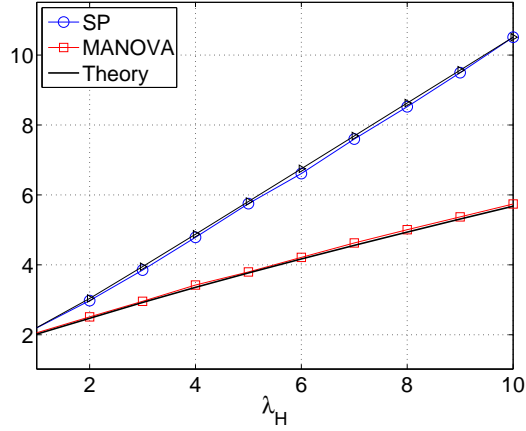


Figure 2: Standard deviation of the largest eigenvalue of $E^{-1}H$ in both the signal processing setting (SP) and in MANOVA. Comparison of simulations results to theoretical approximations.

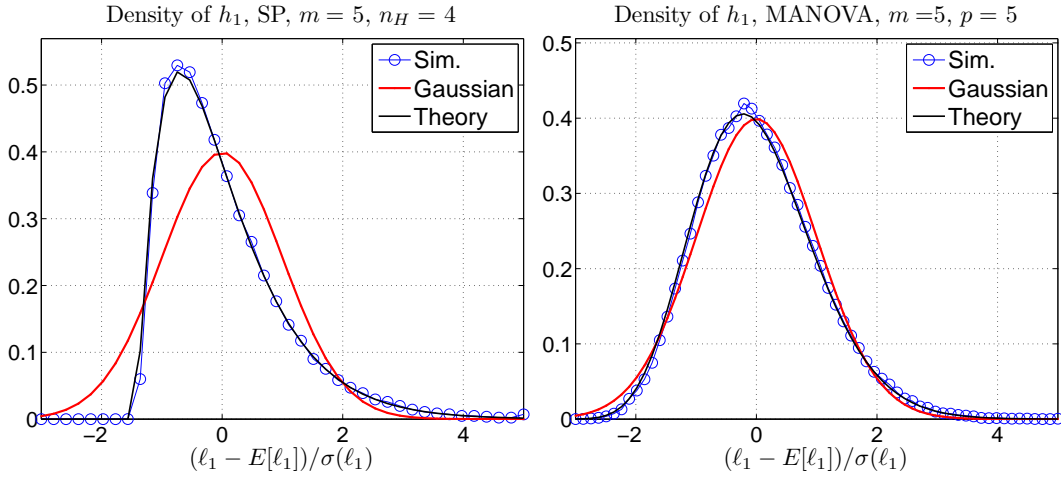


Figure 3: Density of largest eigenvalue of H in the signal processing setting (left) and in the MANOVA setting (right) with $\lambda_H = 10$. We compare the empirical density to the theoretical approximation from Lemma (h), Eqs. (44) and Eq.(43). For reference, the red curve is the density of a standard Gaussian.

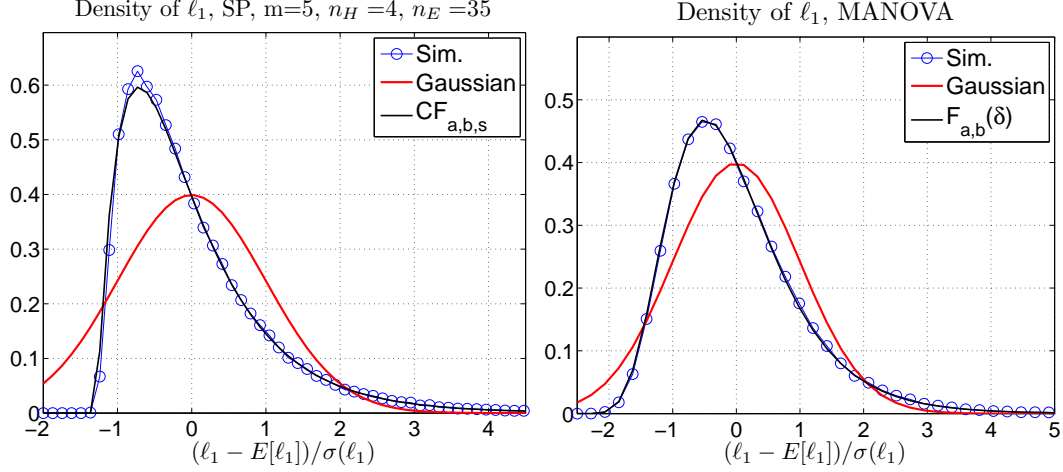


Figure 4: Density of Roy's Largest Root, $\ell_1(E^{-1}H)$, in the signal detection setting (left) and in the MANOVA setting (right) with $\lambda_H = 10$. We compare the empirical density to the theoretical approximation from Propositions 1 and 2, Eqs. (23) and (24), respectively. The red curve is the density of a standard normal.

Finally, we study the accuracy of the approximation to the full distribution of the largest eigenvalue $\ell_1(E^{-1}H)$. Recall that in the MANOVA setting, according to Eq. (54),

$$\Pr \left[\frac{\ell_1 - c_1}{c_2} < x \right] \approx F_{a,b}(\delta; x)$$

Making a change of variables, $\ell_1 = \mathbb{E}[\ell_1] + \sigma(\ell_1)\eta$, we have that

$$\Pr [\eta < t] = F_{a,b} \left(\delta; \frac{\mathbb{E}[\ell_1] - c_2}{c_1} + \frac{\sigma(\ell_1)}{c_1} t \right)$$

or, upon taking the derivative w.r.t. t ,

$$p(\eta = t) \approx \frac{\sigma(\ell_1)}{c_1} f_{a,b} \left(\delta; \frac{\mathbb{E}[\ell_1] - c_2}{c_1} + \frac{\sigma(\ell_1)}{c_1} t \right)$$

In Fig. 4, we compare the empirical density of $\eta = (\ell_1 - \mathbb{E}[\ell_1]) / \sigma(\ell_1)$ to the theoretical density of a non-central F variable in the MANOVA case and to a modified central F in the signal detection setting. For reference, we also compare to the density of a standard normal, $(2\pi)^{-1/2} e^{-t^2/2}$. Note that as expected from the analysis, the density of the largest eigenvalue is skewed, and that our theoretical distribution is quite accurate.

6.1 Power Calculations

We conclude this section with a comparison of the empirical detection power of Roy's largest root test to theoretical predictions based on Eq. (23). For a given rank-one non-centrality matrix with parameter δ , according to Eq. (23) we have that

$$\begin{aligned} P_D &= \Pr[\ell_1 > t(\alpha)] \\ &= 1 - F_{a,b} \left(\delta; \frac{t(\alpha) - c_2}{c_1} \right) \end{aligned} \quad (51)$$

where the parameters a, b, δ, c_1 and c_2 are given in Eqs. (25) and (26), and the threshold $t(\alpha)$ is given in Eq. (17).

dim. m	groups p	samples per group, n_i	non-centrality δ	P_d sim. ($\alpha = 1\%$)	P_d theory	P_d sim. $\alpha = 5\%$	P_d theory
3	3	10	5	0.053	0.049	0.227	0.212
3	3	10	10	0.174	0.171	0.483	0.475
3	3	10	20	0.529	0.530	0.847	0.846
3	3	10	40	0.938	0.939	0.995	0.995
6	3	10	5	0.032	0.027	0.148	0.128
6	3	10	10	0.098	0.093	0.320	0.304
6	3	10	20	0.336	0.333	0.671	0.668
6	3	10	40	0.806	0.808	0.964	0.966
6	6	10	5	0.024	0.011	0.111	0.061
6	6	10	10	0.075	0.055	0.244	0.183
6	6	10	20	0.294	0.264	0.581	0.533
6	6	10	40	0.802	0.789	0.946	0.940
10	6	20	5	0.021	0.007	0.101	0.038
10	6	20	10	0.060	0.037	0.208	0.127
10	6	20	20	0.256	0.213	0.520	0.442
10	6	20	40	0.785	0.759	0.932	0.914

Table 1: Comparison of empirical power of Roy’s largest root test to theoretical approximation at a false alarm rate of $\alpha = 1\%$.

Table 1 compares the theoretical expression (51) to the results of simulations. Each entry in the table is the result of 100,000 independent random realizations of matrices H and E . The parameters in the table are a subset of those studied by Olson [22]. We compare the empirical power with the predicted one both at the standard $\alpha = 5\%$ false alarm rate, as well as at the more stringent value $\alpha = 1\%$. As one can observe from the table, our approximations are quite accurate at high powers, say larger than 80%. In contrast, at low power, our estimate is lower than the true power, e.g. it is a bit conservative.

Let us relate these empirical observations to our analysis. Recall that our asymptotic expansion studies the behavior of the largest eigenvalue ℓ_1 , when it is indeed due to a signal and not due to noise, as it is based on a Taylor expansion as $\sigma \rightarrow 0$. It is thus valid only when the signal strength is sufficiently large, so no eigenvalue cross-over has occurred, meaning that the largest eigenvalue is not due to large fluctuations in the noise. Therefore, our theoretical predictions are indeed expected to be more accurate for larger values of δ and for smaller values of α . Fortunately, they are very accurate where it matters most to statistical applications, e.g. where the required power is large, say 80% or above. At the other extreme, when the signal strength is weak, our approximation of power is conservative since we do not model the case where the largest eigenvalue may arise due to large deviations of the noise. As expected, the discrepancy between true and estimated power is thus larger at larger values of α .

7 Discussion

In this paper, relatively accurate expressions for the distribution of Roy’s largest root test were derived in the extreme setting of a rank-one concentrated non-centrality matrix. Deriving such expressions, even in this restricted case, has been an open problem in multivariate analysis for several decades and has potentially limited the practical use of Roy’s test. The new distributions derived in this paper are simple and straightforward to compute. From the practical aspect, they allow for a simple prospective evaluation of the power of Roy’s largest root test in hypothesis driven research, for example in biomedical experiments and medical trials.

These approximate distributions were derived by a perturbation approach, in the limit of small noise or equivalently strong signal-to-noise ratio. This approach may be useful in a variety of other problems. First, Roy's largest root test for the generalized eigenvalues of sample covariance matrices is applicable in several additional settings. Examples include testing for independence of two sets of variates jointly distributed as Gaussian with unknown mean, and testing equality of covariance matrices of two normal distributions with unknown means. Furthermore, Roy's test can also be used for testing significance in canonical correlation and in multiple response linear regression. As mentioned in Section 4, even in the single matrix case, Roy's largest root may also be applied to test for interactions in two way tables and in a variety of signal detection problems. Each of these cases may yield a somewhat different distribution under the alternative, that hopefully one may still be able to analyze using our approach.

Second, in this paper, we studied the case of a single signal or a rank-one non-centrality matrix. It should be possible to relax this strict and somewhat unrealistic assumption and study the resulting distribution under say two strong signals, or perhaps one strong signal and many weak ones. Finally, our approach can be applied to study other test statistics, such as the Hotelling-Lawley trace. These and related issues, such as the sensitivity of the distributions to departures from normality, are interesting problems for further research.

Acknowledgments. It is a pleasure to thanks Prof. Donald Richards and Prof. David Banks for many useful discussions and suggestions.

A Proofs

A.1 Proofs of Auxiliary Lemmas

Proof of Lemma (B): Let A be the $m \times m$ matrix whose columns are the eigenvectors $\{\mathbf{a}_i\}$ of H . Then, by definition,

$$B = \mathbf{a}_1^T E^{-1} \mathbf{a}_1 = (A^T E^{-1} A)_{1,1} \quad (52)$$

That is, B is simply the upper left diagonal entry of the matrix E^{-1} , in the basis $\{\mathbf{a}_i\}$ that diagonalizes H . Next, recall that the matrices H and E are independent, and thus the random basis \mathbf{a}_i is independent of E . Since $n_E E$ is Wishart distributed with identity covariance and thus its distribution is invariant to unitary transformations, we may apply theorem 3.2.11 from Muirhead's book [11], to obtain Eq. (38). Equations (39) and (40) readily follow as the first two moments of an inverse χ^2 random variable. \square

Proof of Lemma (C): As in the proof of lemma B above, let A be the matrix whose columns are the vectors $\{\mathbf{a}_i\}$. Note that by definition, we can equivalently write

$$\mathbf{a}_1^T E^{-1} \mathbf{a}_i = (A^T E^{-1} A)_{1,i}.$$

Combining this with Eq. (52) gives that

$$C_i = \frac{((A^T E^{-1} A)_{1,i})^2}{(A^T E^{-1} A)_{1,1}}.$$

Again, since A is unitary and independent of E , and since the distribution of the matrix E is invariant to unitary transformations we may equivalently study the distribution of

$$C_i = \frac{((E^{-1})_{1,i})^2}{(E^{-1})_{1,1}}.$$

To the best of our knowledge, the exact distribution of this random variable is not explicitly known. However, its mean can be computed exactly by the following argument: First, w.l.g.

we consider the index $i = 2$ and the following 2×2 sub-matrix of E^{-1} ,

$$\begin{pmatrix} E_{11}^{-1} & E_{12}^{-1} \\ E_{12}^{-1} & E_{22}^{-1} \end{pmatrix} \quad (53)$$

According to Theorem 3.2.11 in [11], the inverse of this matrix, up to multiplication by a factor of n_E , follows a Wishart distribution $W_2(n_E - m + 2, \mathbf{I}_2)$. However, as this is a 2×2 matrix, its inverse can be computed explicitly. It is given by

$$\frac{1}{E_{11}^{-1}E_{22}^{-1} - (E_{12}^{-1})^2} \begin{pmatrix} E_{22}^{-1} & -E_{12}^{-1} \\ -E_{12}^{-1} & E_{11}^{-1} \end{pmatrix} \quad (54)$$

In particular, up to a multiplicative factor of $1/n_E$, its diagonal entries follow a χ^2 distribution with $(n_E - m + 2)$ degrees of freedom. Taking the inverse of the $(2, 2)$ entry gives that

$$E_{22}^{-1} - C_2 \sim \frac{n_E}{\chi_{n_E - m + 2}^2}.$$

Recall that according to Eq. (38), $E_{22}^{-1} \sim n_E/\chi_{n_E - m + 1}^2$. Taking expectations proves Eq. (41) of the lemma.

For the second part of the lemma, recall that E is the sample covariance matrix from n_E multivariate normal samples with an identity population covariance. Therefore, as $n_E \rightarrow \infty$, both $E \rightarrow \mathbf{I}_m$ and $E^{-1} \rightarrow \mathbf{I}_m$. In particular, $E_{i,i} = 1 + O_P(n_E^{-1/2})$ whereas for $i \neq j$, $E_{i,j} = O_P(n_E^{-1/2})$.

To study the value of $C_{1,j}$ and thus the off-diagonal entries of E^{-1} , can view the off-diagonal entries in the matrix E as a perturbation of the diagonal terms. That is, we write

$$E = E_0 + E_1 = E_0(I + E_0^{-1}E_1)$$

where E_0 contains the diagonal terms and E_1 the off-diagonal ones. Then

$$\begin{aligned} E^{-1} &= (I + E_0^{-1}E_1)^{-1}E_0^{-1} \\ &= (I - E_0^{-1}E_1 + (E_0^{-1}E_1)^2 + o_P(1/n_E))E_0^{-1} \end{aligned} \quad (55)$$

In particular, for the off-diagonal term we obtain

$$E_{1,2}^{-1} = -\frac{E_{1,2}}{E_{1,1}E_{2,2}} + \sum_{j>2} \frac{E_{1,j}E_{2,j}}{E_{1,1}E_{2,2}E_{j,j}} + o(1/n_E)$$

It thus follows that $E_{1,2}^{-1} = O_P(1/\sqrt{n_E})$ whereas $C_{1,2} = O(1/n_E)$, and as required $C_{1,2}^2 = O(1/n_E^2)$. □

Proof of Lemma (h_1): We decompose the matrix H into its signal part in the direction \mathbf{v} of the spike and its pure noise part, in the subspace of dimension $m - 1$, orthogonal to the spike. Let $\{\mathbf{v}_j\}_{j=2}^m$ be some arbitrary orthonormal basis for this subspace, that is independent of H . In the basis $\{\mathbf{v}, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ of \mathbb{R}^m , the matrix takes the form

$$H = \left(\begin{array}{c|ccc} Z_s & b_2 & \dots & b_m \\ \hline b_2 & z_{2,2} & & z_{2,m} \\ \vdots & & \ddots & \vdots \\ b_m & z_{m,2} & \dots & z_{m,m} \end{array} \right) \quad (56)$$

where Z_s denotes the variance in the direction of \mathbf{v} , and the random variables $z_{2,2}, \dots, z_{m,m}$ denote the variances of the observed data in the $(m - 1)$ directions \mathbf{v}_j orthogonal to \mathbf{v} . In

the MANOVA case, $Z_s \sim \sigma^2 \chi_{n_H}^2(\lambda_H/\sigma^2)/n_H$, whereas in the signal processing case, $Z_s \sim (\lambda_H + \sigma^2) \chi_{n_H}^2/n_H$. In both cases, as $\sigma \rightarrow 0$, we have that $Z_s = O(\lambda_H)$, and is the most significant entry in the matrix H , with all other entries converging to zero as $\sigma \rightarrow 0$.

Indeed, since the $(m-1)$ directions $\{\mathbf{v}_j\}_{j=2}^m$ are orthonormal and independent of the matrix H , all entries $z_{j,k}$ in the lower right $(m-1) \times (m-1)$ matrix above are $O(\sigma^2)$. In particular, the diagonal entries $z_{2,2}, \dots, z_{m,m}$ are all i.i.d. as $\sigma^2 \chi_{n_H}^2/n_H$.

Finally, the random variables b_j capture the sample covariance between the direction \mathbf{v} of the signal and the direction \mathbf{v}_j orthogonal to it. In terms of the n_H original observations \mathbf{x}_i used to construct the matrix H these are given by

$$b_j = \frac{1}{n_H} \sum_{i=1}^{n_H} x_{i,1} x_{i,j} \quad (57)$$

where $x_{i,1}$ is the projection of the i -th observation onto the direction \mathbf{v} , and $x_{i,j}$ is its projection onto \mathbf{v}_j . Note that by construction, since the directions \mathbf{v}_j are independent of H , for $j > 1$ all $x_{i,j}$ are i.i.d. $N(0, \sigma^2)$ random variables, and independent of $x_{i,1}$. Furthermore, $Z_s = \frac{1}{n_H} \sum_i x_{i,1}^2$. Hence, *conditional* on the n_H values $\{x_{i,1}\}_{i=1}^{n_H}$,

$$b_j = \frac{\sigma}{\sqrt{n_H}} \sqrt{\frac{1}{n_H} \sum_i X_{i,1}^2} \eta_j = \frac{\sigma}{\sqrt{n_H}} \sqrt{Z_s} \eta_j \quad (58)$$

where η_j are all i.i.d. $N(0, 1)$.

As in Eq. (3.6) in [15] we can thus write the matrix H as a sum of three symmetric matrices whose entries are $O(1)$, $O(\sigma)$ and $O(\sigma^2)$ respectively. It then follows that the largest eigenvalue h_1 of H is analytic in σ and we can expand it in a corresponding Taylor series. We then apply Eq. (2.14) from [15] (where in their notation $Z_s = \kappa^2 + 2\sigma\kappa\rho_1 + \sigma^2\beta_{1,1}$, and $\eta_j = \sqrt{n_H}\rho_j$)

$$h_1 = Z_s + \frac{\sigma^2}{n_H} \sum_{j=2}^m \eta_j^2 + o(\sigma^2) \quad (59)$$

To conclude the proof, we recall that since all η_j are independent $N(0, 1)$, this sum follows a χ^2 distribution with $(m-1)$ degrees of freedom. Finally, in the MANOVA setting we have that $Z_s \sim \frac{\sigma^2}{n_H} \chi_{n_H}^2(\delta/\sigma^2)$. Using the additivity property of sums of central and non-central χ^2 random variables gives Eq. (43). In the signal processing case, $Z_s \sim \frac{\lambda_H + \sigma^2}{n_H} \chi_{n_H}^2$ and hence Eq. (44) follows. \square

Remarks: i) In [15] the above result was proven only for the case of Gaussian signal and noise. Here we extended the proof to the case of a signal with a non-central χ^2 distribution. However, a closer inspection reveals that the technique is applicable under more general assumptions on the signal model. The key requirement is that the noise be multivariate Gaussian and independent of the signal. If the signal has a different distribution, possibly even be correlated in time, then that will affect the distribution of the first term Z_s but as $\sigma \rightarrow 0$ will not affect the distribution of the second term, which to leading order will still be $\sigma^2 \chi_{m-1}^2$.

ii) The results of this lemma are of independent interest, as they may be used to compute the power of tests that consider the largest eigenvalue of single covariance matrices. In particular, for the case of a non-central Wishart with rank one noncentrality matrix, this is relevant in testing the presence of interactions in two way tables, see [7, 4]. Similarly, in the case of complex valued signals and noise, the eigenvalues of non-central complex Wishart matrices are also of interest in communications systems, see for example [6].

A.2 Proofs of Theorems

Proof of Theorem 1: We insert the expansion (31) into the eigenvalue equation

$$E^{-1/2}(H_0 + \epsilon H_1)E^{-1/2} = \lambda(\epsilon)\mathbf{v}(\epsilon).$$

This yields to leading order

$$E^{-1/2}H_0E^{-1/2}\mathbf{v}_0 = \lambda_0\mathbf{v}_0.$$

Since the matrix H_0 is of rank one, so is $E^{-1/2}H_0E^{-1/2}$ and it thus has a single non-zero eigenvalue. Inserting the expression for $H_0 = h_1\mathbf{a}_1\mathbf{a}_1^T$ and simple linear algebra then gives Eqs. (32) and (34) for this eigenvalue and its corresponding eigenvector.

Next we consider the perturbation due to the remaining eigenvalues of H , stored in the matrix H_1 . The next equation of order $O(\epsilon)$ is

$$(E^{-1/2}H_1E^{-1/2})\mathbf{v}_0 + (E^{-1/2}H_0E^{-1/2})\mathbf{v}_1 = \lambda_0\mathbf{v}_1 + \lambda_1\mathbf{v}_0. \quad (60)$$

Next, recall that $\mathbf{v}_0 = E^{-1/2}\mathbf{a}_1$, and that $H_0 = h_1\mathbf{a}_1\mathbf{a}_1^T$. Thus

$$\mathbf{v}_0^T(E^{-1/2}H_0E^{-1/2})\mathbf{v}_1 = h_1(\mathbf{a}_1^TE^{-1}\mathbf{a}_1)\mathbf{a}_1^TE^{-1/2}\mathbf{v}_1 = \lambda_0\mathbf{v}_0^T\mathbf{v}_1$$

Hence, taking the dot product of Eq. (60) with \mathbf{v}_0 the term $\lambda_0\mathbf{v}_0^T\mathbf{v}_1$ cancels out, and we obtain

$$\mathbf{v}_0^T(E^{-1/2}H_1E^{-1/2})\mathbf{v}_0 = \lambda_1\mathbf{v}_0^T\mathbf{v}_0$$

Substituting the expression for \mathbf{v}_0 yields the required expression for λ_1 , Eq. (33). \square

Proof of Theorem 2: We first consider the mean of the largest eigenvalue of $E^{-1}H$. Using the decomposition (31), we thus compute the means of the two leading order terms λ_0 and λ_1 , Eqs. (32) and (33), respectively. Further recall that $\lambda_0 = h_1B$ and that h_1 depends only on the matrix H and that since the matrices H and E are independent, then the random variable B is independent of h_1 . Therefore,

$$\mathbb{E}[\lambda_0] = \mathbb{E}[h_1] \cdot \mathbb{E}[B].$$

Eq. (48) immediately follows from Eq. (39) for the mean of the random variable B .

Next, we consider the mean of λ_1 . To this end, recall that

$$\lambda_1 = \sum_{j=2}^m \tilde{h}_j C_j$$

and hence

$$\mathbb{E}[\lambda_1] = \sum_{j=2}^m \mathbb{E}[\tilde{h}_j C_j] = \sum_{j=2}^m \mathbb{E}_H[\tilde{h}_j \mathbb{E}[C_j | H]]$$

Next, from Lemma (C) it follows that $\mathbb{E}[C_j | H]$ is independent of the matrix H and is given by Eq. (41). Therefore,

$$\mathbb{E}[\lambda_1] = \mathbb{E}[C_2] \cdot \mathbb{E}\left[\sum_{j=2}^m \tilde{h}_j\right] = \mathbb{E}[C_2] \cdot \mathbb{E}[\text{Tr}(H_1)]$$

Similarly, since all the eigenvalues \tilde{h}_j are due to noise, the average of their sum is $(m-1)(1 + o(\sigma^2))$. Combining all of these results yields Eq. (49). As in [15], the reason for the additional transcendentally small terms in σ in Eq. (47) is due to small probability of a crossover between the eigenvalue due to the signal and the largest eigenvalue due to noise. When this crossing occurs, the Taylor expansion (31) does not capture anymore the behavior of the largest eigenvalue

of $E^{-1}H$. The probability of such an event is of the form e^{-c/σ^2} , and thus is transcendentally small in σ , as $\sigma \rightarrow 0$.

Next, we consider the variance of ℓ_1 . By definition,

$$\begin{aligned} \text{Var}[\ell_1] &= \mathbb{E}[(\lambda_0 + \sigma^2 \lambda_1)^2] - \mathbb{E}[\lambda_0 + \sigma^2 \lambda_1]^2 + o(\sigma^2) \\ &= \text{Var}[\lambda_0] + 2(\mathbb{E}[\lambda_0 \lambda_1] - \mathbb{E}[\lambda_0]\mathbb{E}[\lambda_1])\sigma^2 + o(\sigma^2) \end{aligned} \quad (61)$$

Again using the independence of the two random variables h_1 and B gives

$$\text{Var}(\lambda_0) = \text{Var}(h_1)\mathbb{E}[B^2] + \text{Var}(B)\mathbb{E}[h_1]^2.$$

where the variance of B is given by Eq. (40). As for the mixed term $\mathbb{E}[\lambda_0 \lambda_1]$, it can be written as

$$\mathbb{E} \left[h_1 \sum_j \tilde{h}_j ((A^T E^{-1} A)_{1j})^2 \right].$$

Since the matrix A is independent of the matrix E , we have that $\mathbb{E}[(A^T E^{-1} A)_{1j}]^2 = \mathbb{E}[(E_{1j}^{-1})^2]$. As is well known, see e.g. [20],

$$\mathbb{E}[(E_{ij}^{-1})^2] = \frac{n_E^2}{(n_E - m)(n_E - m - 1)(n_E - m - 3)}. \quad (62)$$

Furthermore, asymptotically as $\sigma \rightarrow 0$,

$$\mathbb{E} \left[h_1 \sum_{j=2}^m \tilde{h}_j \right] = (m-1)\mathbb{E}[h_1](1 + O(\sigma^2)).$$

Therefore,

$$\mathbb{E}[\lambda_0 \lambda_1] = \frac{(m-1)n_E^2}{(n_E - m)(n_E - m - 1)(n_E - m - 3)} \mathbb{E}[h_1] (1 + O(\sigma^2))$$

Combining all of the above expressions yields Eq. (50). \square

References

- [1] T.W. Anderson, *An introduction to multivariate statistical analysis*, Third Ed., Wiley, 2003.
- [2] J. Baik, J.W. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, *J. Mult. Anal.*, 97(6), pp. 1382–1408, 2006.
- [3] R. W. Butler and R.L. Paige, Exact distributional computations for Roy’s statistic and the largest eigenvalue of a Wishart distribution, *Stat. Comp.*, 2010.
- [4] E.M. Carter and M.S. Srivastava, Asymptotic distribution of the latent roots of the non-central Wishart distribution and the power of the likelihood ratio test for nonadditivity, *Can. J. Stat.*, 8(1), pp. 119–134, 1980.
- [5] T. Jiang, Approximation of Haar Distributed Matrices and Limiting Distributions of Eigenvalues of Jacobi Ensembles, *Prob. Th. and Related Fields*, 144(1), 221–246, 2007.
- [6] S. Jin, M.R. McKay, X. Gao, I.B. Collings, MIMO multichannel beamforming: SER and outage using new eigenvalue distributions of complex noncentral Wishart matrices, *IEEE Transactions on Communications*, 56(3), pp. 424–434, 2008.

- [7] D.E. Johnson, F. A. Graybill, An analysis of a two way table with interactions and no replications, *J. Am. Stat. Assoc.*, 67, pp. 862–868, 1972.
- [8] I.M. Johnstone, Multivariate Analysis and Jacobi Ensembles: Largest Eigenvalue, Tracy-Widom Limits and Rates of Convergence, *Ann. Stat.*, vol 36(6), pp. 2638–2716, 2008.
- [9] T. Kato, *Perturbation Theory for Linear Operators*, 2nd Edition, Springer, 1995.
- [10] S.M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, Prentice-Hall, 1998.
- [11] R.J. Muirhead, *Aspects of multivariate statistical theory*, Willey, New-York, 1982.
- [12] K.E. Muller, B.L. Peterson, Practical methods for computing power in testing the multivariate linear hypothesis, *Comp. Stat. Data Analysis*, vol. 2, pp. 143–158, 1984.
- [13] K.E. Muller, L.M. LaVange, S. Landesman Ramey and C.T. Ramey, Power calculations for general multivariate models including repeated measures applications, *J. Amer. Stat. Assoc.*, vol. 87, pp. 1209–1226, 1992.
- [14] K.E. Muller, P.W. Stewart, *Linear Model Theory, Univariate, Multivariate and Mixed Models*, Wiley, 2006.
- [15] B. Nadler, Finite sample approximation results for principal component analysis: A matrix perturbation approach, *Ann. Stat.*, vol. 36, no. 6, pp. 2791–2817, 2008.
- [16] B. Nadler, Non-parametric detection of signals by information theoretic criteria, performance analysis and an improved estimator, *IEEE Trans. Sig. Proc.*, 58(5), pp. 2746–2756, 2010.
- [17] B. Nadler, I.M. Johnstone, Detection Performance of Roy’s Largest Root Test when the noise covariance matrix is arbitrary, *IEEE Statistical Signal Processing Conference*, 2011.
- [18] R.G. O’Brien and G. Shieh, Pragmatic, Unifying algorithm gives power probabilities for common F tests of the multivariate general linear hypothesis, *unpublished*.
- [19] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, *Statistica Sinica*, vol. 17, No. 4, pp. 1617–1642, 2007.
- [20] S.J. Press, *Applied Multivariate Analysis*, Krieger Publishing Co., Malabar, FL, 2nd edition, 1982.
- [21] R. Rao Nadakuditi, J. Silverstein, Fundamental Limit of Sample Generalized Eigenvalue Based Detection of Signals in Noise using relatively few signal-bearing and noise-only samples, *IEEE J. Selected Topics in Signal Processing*, Vol. 4(3), pp. 468–480, 2010.
- [22] C.L. Olson, Comparative Robustness of Six Tests in Multivariate Analysis of Variance, *Journal of the American Statistical Association*, vol. 348, pp. 894–908, 1974.
- [23] B.D. Rao, M. Wengler and B. Judson, Performance analysis and comparison of MRC and optimal combining in antenna array systems, in *Proc. ICASSP*, pp. 2949–2952, 2001.
- [24] S.N. Roy, *Some aspects of multivariate analysis*, Wiley, New-York, 1957.
- [25] J.W. Silverstein, The limiting eigenvalue distribution of a multivariate F matrix, *SIAM J. Math. Anal.*, vol. 16(3), pp. 641–646, 1985.
- [26] P. Stoica and M. Cedervall, Detection tests for array processing in unknown correlated noise fields, *IEEE Trans. Sig. Proc.* vol. 45, no. 9, pp. 2351–2362, 1997.

- [27] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, On detection of the number of signals when the noise covariance matrix is arbitrary, *J. Multivariate Anal.*, vol. 20, no. 1, pp. 26-49, 1986.
- [28] Z. Zhu, S. Haykin and X. Huang, Estimating the number of signals using reference noise samples, *IEEE Trans. Aerosp. Electron. Syst.*, vol 27, no 3, pp. 575–579, 1991.