DETECTION PERFORMANCE OF ROY'S LARGEST ROOT TEST WHEN THE NOISE COVARIANCE MATRIX IS ARBITRARY

B. Nadler[*]

Department of CS and Applied Math.
Weizmann Institute of Science
Rehovot, 76100, Israel

I. M. Johnstone

Department of Statistics
Stanford University
Stanford, 94305, CA.

ABSTRACT

Detecting the presence of a signal embedded in noise from a multi-sensor system is a fundamental problem in signal and array processing. In this paper we consider the case where the noise covariance matrix is arbitrary and unknown but we are given both signal bearing and noise-only samples. Using a matrix perturbation approach, combined with known results on the eigenvalues of inverse Wishart matrices, we study the behavior of the largest eigenvalue of the relevant covariance matrix, and derive an approximate expression for the detection probability of Roy's largest root test. The accuracy of our expressions is confirmed by simulations.

Index Terms— signal detection, Roy's largest root test, matrix perturbation, inverse Wishart distribution.

1. INTRODUCTION AND PROBLEM SETUP

Detecting the presence of a signal corrupted by additive Gaussian noise from a multi-sensor system is a fundamental problem in signal and array processing. When the noise covariance matrix is the identity, the performance of signal detection schemes is well understood [1]. In this paper we focus on the case of an unknown and arbitrary noise covariance matrix, estimated from noise-only samples.

We consider the following setup. Let $\mathbf{x}_i \in \mathbb{R}^m$, $i = 1, \ldots, n_H$, denote n_H i.i.d. observations from the following "signal plus noise" model

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

where $\mathbf{h}$ is an unknown channel vector, 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 with a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

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

$$\mathcal{H}_0 : \text{no signal}, \rho_s = 0 \;\text{ vs. }\; \mathcal{H}_1 : \text{signal present}, \rho_s > 0. \tag{2}$$

[*]Work performed while on Sabbatical at UC-Berkeley and Stanford departments of statistics.

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

$$H = \frac{1}{n_H} \sum_{i=1}^{n_H} \mathbf{x}_i \mathbf{x}_i^T$$

is the sample covariance of the original observations.

In this paper we consider the case where $\boldsymbol{\Sigma}$ is arbitrary and unknown, but we have at our disposal another independent set $\{\mathbf{z}_j\}_{j=1}^{n_E}$ of n_E noise-only i.i.d. observations of the form $\mathbf{z} = \sigma\boldsymbol{\xi}$. In this setting, it is common to estimate the noise covariance matrix via

$$E = \hat{\boldsymbol{\Sigma}} = \frac{1}{n_E} \sum_{i=1}^{n_E} \mathbf{z}_i \mathbf{z}_i^T$$

and consider a function of the eigenvalues of $E^{-1}H$ (instead of the unknown $\boldsymbol{\Sigma}^{-1}H$) as a test statistic for signal detection.

Zhao *et. al.* [2] were among the first to consider this setting, and derived an iterative sphericity test combined with an information theoretic criteria. Whereas [2] used all eigenvalues for signal detection, recently Rao and Silverstein [3] developed an improved detection method that sequentially tests the significance of each of the largest eigenvalues of $E^{-1}H$.

In this paper, using a matrix perturbation approach, we study the distribution of the largest eigenvalue of $E^{-1}H$ when a signal is present, and derive simple approximate expressions for the detection probability of this test. Our analysis allows, for a given false alarm rate and the various system parameters (number of sensors, samples, etc.), to determine the required signal strength for detection with high probability, which is an important quantity for system design. While beyond the scope of this paper our analysis has applications to several other hypothesis testing problems in multivariate statistics.

2. DETECTION BY ROY'S LARGEST ROOT TEST

Let $\ell_1 = \ell_1(E^{-1}H)$ denote the largest eigenvalue of $E^{-1}H$. The estimator suggested in [3] is a sequential version of the well known Roy's largest root test in multivariate analysis of variance (MANOVA), see [4]. In our single signal setting, given a false alarm rate $\alpha \ll 1$, a signal is detected if

$$\ell_1 > th(\alpha) \tag{3}$$

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

To set the threshold, the right tail distribution of $\ell_1(E^{-1}H)$ under the null is needed. For the probability of detection,

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

the distribution of ℓ_1 under the alternative and 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, and a subject of several works. In principle, the exact distribution can be represented in terms of a hypergeometric function of matrix argument, see [4] for references. However, unless all problem parameters m, n_H, n_E are small, this representation is difficult to evaluate numerically. Similarly, it is also difficult to analyze theoretically.

Recently, using tools from random matrix theory, Johnstone [4] proved that under the null hypothesis, asymptotically as $m, n_E, n_H \to \infty$, with their ratios converging to fixed constants, $\log(\ell_1(E^{-1}H))$ 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 in the joint limit as* $m, n_E, n_H \to \infty$

$$\Pr\left[\tfrac{W-\mu_{TW}}{\sigma_{TW}} < s\right] \to F_1(s)$$

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

$$\mu_{TW} = 2\log\tan\left(\tfrac{\varphi+\gamma}{2}\right) \tag{4}$$

$$\sigma_{TW}^3 = \frac{16}{(n_E+n_H-1)^2}\frac{1}{\sin^2(\varphi+\gamma)\sin(\varphi)\sin(\gamma)} \tag{5}$$

where the angle parameters γ, φ *are*

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

3. ON THE DISTRIBUTION OF THE LARGEST ROOT TEST

The goal of this paper is to study, under the presence of a single signal (e.g., under $\mathcal{H}_1$), the distribution of the largest eigenvalue of $E^{-1}H$, in particular its mean and variance.

As mentioned above, while the exact distribution can be written in terms of a hypergeometric function of matrix argument. this representation is difficult to analyze.

Recently, Rao and Silverstein [3] studied the largest eigenvalue of $E^{-1}H$ in the joint limit as $m, n_E, n_H \to \infty$, under the alternative hypothesis of signals present. As in principal component analysis with a spiked covariance matrix, there is a phase transition phenomenon whereby to be detected by the largest eigenvalue, the signal strength must be larger than some threshold. In [3] the authors derived both this limiting threshold, as well as the deterministic limit for $\ell_1(E^{-1}H)$ when the signal is sufficiently strong,

$$\ell_1 \to \frac{2c_H\lambda_H}{2c_H + c_E\left(1 - c_H - \lambda_H + \sqrt{f(\lambda_H, c_H)}\right)} \tag{6}$$

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

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

and λ_H is the limiting value of the largest eigenvalue of the whitened matrix $(\sigma^2\mathbf{\Sigma})^{-1}H$ [5, 6, 7]

$$\lambda_H = \frac{1}{\sigma^2}(\lambda_s + \sigma^2)\left(1 + \frac{m-1}{n_H}\frac{\sigma^2}{\lambda_s}\right) \tag{7}$$

with $\lambda_s = \rho_s\|\mathbf{\Sigma}^{-1}\mathbf{h}\|^2$ the (whitened) signal strength.

It is instructive to consider the asymptotics of Eq. (6) when $\lambda_H \gg (1+\sqrt{c_H})^2$. We then have that

$$\ell_1 \to \lambda_H\frac{1}{1-c_E} + \frac{c_E}{(1-c_E)^2} + O\left(\frac{1}{\lambda_H}\right). \tag{8}$$

Eq. (8) shows 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 λ_H.

Since the exact limit in Eq. (6) was derived by analyzing the limiting Stieltjes transform of the spectral density of the matrix $E^{-1}H$ as both $m, n_E, n_H \to \infty$, its accuracy for finite values of m, n_E, n_H is unclear. Furthermore, although [3] suggested, by analogy to PCA, that asymptotically the largest eigenvalue may follow a Gaussian distribution, no expression for its asymptotic variance was derived.

3.1. A Matrix Perturbation Approach

In this paper we present a simple explanation for the emergence of the first two terms in Eq. (8), and also derive an approximate expression for the *variance* of the largest eigenvalue of $E^{-1}H$. This allows computation of the approximate power of Roy's largest root test, as well as its analytic comparison to several alternative popular test statistics.

Following [7], our technique is based on a matrix perturbation approach, considering the noise level σ as a small parameter. In our analysis, the dimension m as well as the sample sizes n_E and n_H are all fixed. Therefore, rather than relying on random matrix theory, we use well known results regarding the eigenvalues of *finite* inverse Wishart matrices.

Our point of departure is the following observation: Since $E^{-1}H = (\Sigma^{-1}E)^{-1}(\Sigma^{-1}H)$, rather than working with E and H we can instead consider the whitened matrices $\Sigma^{-1}E$ and $\Sigma^{-1}H$. For analysis purposes, we thus assume that E follows a Wishart distribution $W_m(n_E, \mathbf{I}_m)$, and similarly that H is distributed as $W_m(n_H, \sigma^2\mathbf{I}_m + \lambda_s\mathbf{h}\mathbf{h}^T)$, namely a covariance matrix with a single spike.

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 $E^{-1}H$. We denote by $\{\mathbf{a}_i\}_i$ the (random) orthonormal basis that diagonalizes the matrix H,

$$H = \sum_{i=1}^{m} h_i \mathbf{a}_i \mathbf{a}_i^T \tag{9}$$

where $h_1 \geq h_2 \geq \ldots \geq h_m$ are the sample eigenvalues of H, sorted in decreasing order of magnitude.

We also denote by μ_i and $\boldsymbol{\psi}_i$ the real-valued eigenvalues and orthonormal eigenvectors of the symmetric matrix $E^{-1/2}$

$$E^{-1/2} = \sum_{i=1}^{m} \mu_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T.$$

Two important properties of the eigenvalues and eigenvectors of $E^{-1/2}$, that follow from the invariance property of Wishart matrices to unitary transformations, are as follows: i) The eigenvalues μ_i are *independent* of the eigenvectors $\boldsymbol{\psi}_i$, and ii) The eigenvectors $\boldsymbol{\psi}_i$ are uniformly distributed, with Haar measure, on the unit sphere of $\mathbb{R}^m$.

We study $\ell_1(E^{-1}H)$ in the case where $\lambda_s \gg \sigma^2\sqrt{m/n_H}$. In this case $h_2, \ldots, h_m$ are all $O(\sigma^2)$, whereas h_1 is substantially larger. We thus write the matrix H as

$$H = H_0 + \sigma^2 H_1, \tag{10}$$

where $H_0 = h_1\mathbf{a}_1\mathbf{a}_1^T$ and $H_1 = \sigma^2\sum_{j=2}^m \tilde{h}_j \mathbf{a}_j\mathbf{a}_j^T$, with $\tilde{h}_j = h_j/\sigma^2 = O(1)$. Similarly, we view the matrix $\sigma^2 E^{-1/2}H_1E^{-1/2}$ as a perturbation of $E^{-1/2}H_0E^{-1/2}$. We thus expand the leading eigenvalue and corresponding eigenvector of $E^{-1/2}HE^{-1/2}$ in a Taylor series,

$$\begin{aligned} \ell_1(E^{-1/2}HE^{-1/2}) &= \lambda_0 + \epsilon\lambda_1 + \ldots \\ \mathbf{v}(E^{-1/2}HE^{-1/2}) &= \mathbf{v}_0 + \epsilon\mathbf{v}_1 + \ldots \end{aligned} \tag{11}$$

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

Theorem: *The first two terms in Eq. (11) are given by*

$$\lambda_0 = \quad h_1 \textstyle\sum_i \mu_i^2 \langle \mathbf{a}_1, \boldsymbol{\psi}_i\rangle^2 = h_1 \mathbf{a}_1^T E^{-1}\mathbf{a}_1 \tag{12}$$

$$\begin{aligned} \lambda_1 &= \frac{1}{\sum_j \mu_j^2 \langle \mathbf{a}_1, \boldsymbol{\psi}_j\rangle^2} \sum_{i=2}^{m} \tilde{h}_i \big(\sum_j \mu_j^2 \langle \mathbf{a}_1, \boldsymbol{\psi}_j\rangle\langle \mathbf{a}_i, \boldsymbol{\psi}_j\rangle \big)^2 \\ &= \mathbf{a}_1^T E^{-1} H_1 E^{-1}\mathbf{a}_1 / (\mathbf{a}_1^T E^{-1}\mathbf{a}_1) \end{aligned} \tag{13}$$

Remark: Eqs. (12) and (13) reveal several interesting points. First, to leading order the largest eigenvalue depends on three independent factors. One factor is the largest eigenvalue h_1 of the matrix H, the second is the eigenvalues μ_i^2 of the inverse of a Wishart matrix, and the third factor, as captured by $\langle \mathbf{a}_1, \boldsymbol{\psi}_i\rangle^2$, is purely *geometric*, as it measures the incoherence between two random orthonormal bases in $\mathbb{R}^m$. The second remark is that to leading order there are no interactions between the eigenvalues h_i of H, in the sense that the largest eigenvalue h_1 of H affects only the leading order term whereas the remaining ones affect the next order term.

Next, we compute the leading order mean and variance of the largest eigenvalue, as described by the following theorem.

Theorem: *As $\sigma \to 0$, the leading order terms for the mean and variance of $\ell_1(E^{-1}H)$ are given by*

$$\mathbb{E}[\ell_1(E^{-1}H)] = \mathbb{E}[\lambda_0] + \sigma^2\mathbb{E}[\lambda_1] + o(\sigma^2) + t.s.t. \tag{14}$$

where $t.s.t.$ stands for transcendentally small terms, and

$$\mathbb{E}[\lambda_0] = \mathbb{E}[h_1] \cdot \frac{1}{1 - \frac{m+1}{n_E}} \tag{15}$$

$$\mathbb{E}[\lambda_1] = \frac{m-1}{n_E} \cdot \frac{1}{1 - \frac{m}{n_E}} \cdot \frac{1}{1 - \frac{m+1}{n_E}} \left(1 + O(\tfrac{1}{n_H}, \tfrac{1}{n_E})\right) \tag{16}$$

As for the variance, to leading order and up to terms $o(1/n_E)$ and $o(1/n_H)$,

$$\begin{aligned} Var[\ell_1(E^{-1}H)] \quad &= \quad \frac{2(\lambda_s+1)^2}{\left(1 - \frac{m+1}{n_E}\right)\left(1 - \frac{m+3}{n_E}\right)} \times \\ & \quad \left\{ \frac{1}{n_H} + \frac{1}{n_E}\frac{1}{1 - \frac{m+1}{n_E}} \right\}. \end{aligned} \tag{17}$$

Remarks: i) The additional transcendentally small terms of the form Ce^{-A/σ^2} arise from the small probability of an eigenvalue swap, see [7].

ii) Eqs. (15)-(16) with fixed m, n_E and n_H, for the means of the first two terms in the Taylor expansion, shed light on the asymptotic formula (8). In particular, up to a correction factor $O(1/n_E)$, $\mathbb{E}[\lambda_0]$ is identical to the first term in (8). Similarly, when all of m, n_E, n_H are large, $\mathbb{E}[\lambda_1] \approx c_E/(1-c_E)^2$ which is the second term in (8). Our analysis thus shows that the limiting Eq. (6) of [3] is quite accurate for the mean of $\ell_1(E^{-1}H)$ also for finite and small values of m, n_E, n_H.

3.2. Power Calculations

We consider the case where both $n_E, n_H \gg 1$. Then, both h_1 and the random variable $B = \sum_j \mu_j^2 \langle \mathbf{a}_1, \boldsymbol{\psi}_j\rangle^2$ approximately Gaussian distributed, and by the Delta method, so is their product. We thus write

$$\ell_1(E^{-1}H) \approx \mathbb{E}[\ell_1] + \sigma(\ell_1)\eta$$

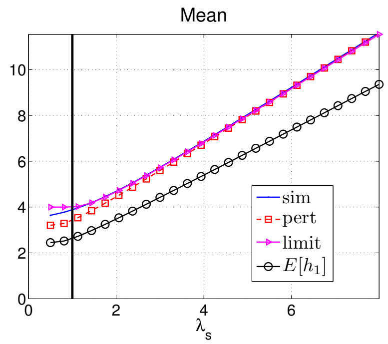

Fig. 1. Mean of $\ell_1(E^{-1}H)$ vs. signal strength. The black vertical line is the limiting threshold for signal detection

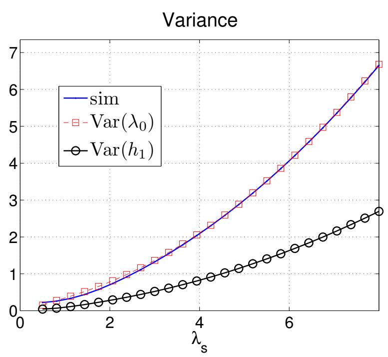

Fig. 2. Variance of $\ell_1(E^{-1}H)$ vs. signal strength. Top two curves are comparison of simulation with Eq. (17).

where $\eta \sim N(0,1)$. Then an approximate expression for the detection probability of Roy's largest root test is

$$\begin{aligned} P_D &= \Pr[\ell_1(E^{-1}H) > th(\alpha)\,|\,\mathcal{H}_1] \qquad (18) \\ &= \Pr\left[\eta > \frac{th(\alpha) - \mathbb{E}[\ell_1]}{\sigma(\ell_1)}\right] = \Phi^c\left(\frac{th(\alpha) - \mathbb{E}[\ell_1]}{\sigma(\ell_1)}\right) \end{aligned}$$

where $\Phi(z)$ is the cdf of a standard Gaussian r.v.

4. SIMULATIONS

To illustrate the accuracy of our expressions, we performed simulations with real-valued signals with an array of $m = 20$ sensors, and with $n_H = 60$ signal-bearing samples and $n_E = 120$ noise-only samples. Figs. 1 and 2 show the mean and variance of ℓ_1 vs. signal strength λ_s. Fig. 3 compares the empirical detection performance to Eq. (18).

5. DISCUSSION

In this paper we presented a perturbation approach for the distribution of the largest eigenvalue (Roy's largest root), in the presence of a single signal. The simulation results show that when $n_E, n_H \gg 1$ our theoretical formula for the detection performance is quite accurate. When n_E, n_H and m are small Eq. (18) is less accurate due to the non-Gaussian distribution of $\ell_1(E^{-1}H)$. More precise approximations are a subject of future study. Finally, while here we considered the case of a single real-valued signal, the analysis may be extended both to several signals as well as to the complex valued case.

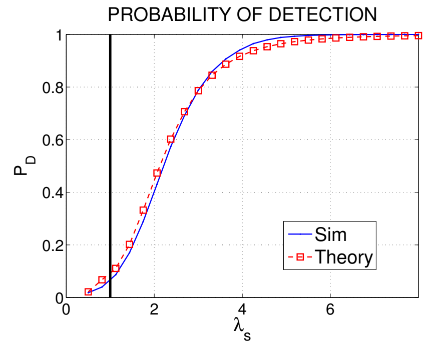

Fig. 3. Probability of detection at $\alpha = 0.01$.

6. REFERENCES

[1] S. Kritchman and B. Nadler, "Nonparametric detection of the number of signals, hypothesis testing and random matrix theory," *IEEE Tran. Sig. Proc.*, vol. 57, pp. 3930–41, 2009.

[2] 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, pp. 26–49, 1986.

[3] R. Rao Nadakuditi and J.W. Silverstein, "Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples," *IEEE Selected Topics in Sig. Proc.*, vol. 4, no. 3, pp. 468–480, 2010.

[4] I.M. Johnstone, "Multivariate analysis and Jacobi ensembles," *Ann. Stat.*, vol. 36, no. 6, pp. 2638–2716, 2008.

[5] J. Baik and J.W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *J. Multivariate Anal.*, vol. 97, pp. 1382–1408, 2006.

[6] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Stat. Sinica*, vol. 17, pp. 1617–1642, 2007.

[7] B. Nadler, "Finite sample approximation results for principal component analysis: A matrix perturbation approach," *Ann. Stat.*, vol. 36, no. 6, pp. 2791–2817, 2008.