The Performance of a Matched Subspace Detector that Uses Subspaces Estimated from Finite, Noisy, Training Data

Nicholas Asendorf* and Raj Rao Nadakuditi
EDICS: SSP-DETC, SSP-PERF

***Abstract*—We analyze the performance of a matched subspace detector (MSD) where the test signal vector is assumed to reside in an unknown, low-rank k subspace which must be estimated from finite, noisy, signal-bearing training data. We consider stochastic and deterministic models for the test vector and characterize the associated ROC performance curves of the resulting detectors. Our analysis utilizes recent results from random matrix theory (RMT) that precisely quantify the quality of the subspace estimate as a function of the eigen-SNR, dimensionality of the system, and the number of training samples. We exploit this knowledge of the subspace estimation accuracy to derive a new RMT detector whose performance matches that of a plug-in detector that uses only the $k_{\text{eff}} \leq k$ 'informative' eigenvectors of the sample covariance matrix. Detectors using more than k_{eff} components will realize a performance loss that our analysis quantifies in the large system limit. We validate our asymptotic predictions with simulations on moderately sized systems.**

***Index Terms*—Matched subspace detector, deterministic, stochastic, random matrix theory, ROC analysis**

I. Introduction

THE matched subspace detector (MSD) is a powerful technique, used widely in signal processing [1] and machine learning [2], to detect a signal, which is assumed to lie in a low-rank subspace, buried in noise. The performance of a MSD when the signal subspace is known a priori has been extensively studied (see, for example, [3], [4], [5], [6], [7]).

In contrast, this paper considers the previously unstudied setting where the signal subspace is unknown and must be estimated from finite, noisy, signal-bearing training data. We are motivated by real-world applications (such as, for example, detection of targets using wavefronts that have propagated through dynamic random media) where an analytical model for the signal subspace is either not known or is inaccurate. In these applications, noise-free training data is unavailable but the practitioner has access to a dataset consisting of noisy, signal-bearing observations collected in a variety of representative experimental conditions.

In such a setting, an estimate of the unknown low-rank signal subspace may be formed from an eigen-decomposition of the sample covariance matrix of the training data. This subspace estimate may then be 'plugged-in' to the detectors derived assuming a known subspace. Because of additive noise and finite training data, the estimated subspace is noisy and we expect a degradation in the performance of these plug-in detectors. Analytically quantifying this performance degradation as a function of the system dimensionality, number of samples, and eigen-SNR is the focus and main contribution of this paper.

This performance characterization relies on recent results from random matrix theory (RMT) [8], [9] that precisely quantify the accuracy of the subspace estimate. We characterize the ROC performance of a MSD assuming deterministic and stochastic models for the test vector and also derive a new RMT detector that incorporates this knowledge of the accuracy of the estimated subspace. This work builds on [10] by also considering the deterministic test vector setting and unifying the performance analysis of the stochastic and deterministic MSD's.

Our analysis highlights the importance of using the k_{eff} *informative* signal subspace components, where k_{eff} is a function of the system dimensionality, number of samples, and eigen-SNR. By only utilizing the k_{eff} *informative* signal subspace components, the new RMT detector avoids the possible performance degradation suffered by the plug-in detector, which may use uninformative components for detection. When the plug-in detector uses exactly k_{eff} subspace components, it achieves the same performance as the RMT detector.

The paper is organized as follows. We describe the generative models for the training data and test vector and formally pose the questions addressed herein in Section II. Section III contains pertinent results from RMT and justifies our definition in (5) of k_{eff}, the (effective) number of informative subspace components. Of particular importance is Corollary 3.1, which justifies a diagonal approximation for a certain type of matrix that appears throughout our derivations. In sections IV and V we derive the plug-in and the new RMT detectors for the stochastic and deterministic test vector models, respectively. Aided by RMT, we derive the (asymptotic) distribution of the test statistics in Section VI; Tables I and II summarize the resulting conditional distributions for the plug-in and RMT detector statistics. These distributions are a weighted sum of chi-square random variables from which we can compute ROC curves via a saddlepoint approximation of the distributions obtained. We validate our asymptotic ROC predictions in

N. Asendorf is with the Department of Electrical Engineering and Computer Science, 1301 Beal Avenue, Room 4313, University of Michigan, Ann Arbor, MI, 48109 USA (e-mail: asendorf@umich.edu).

R.R. Nadakuditi is with the Department of Electrical Engineering and Computer Science, 1301 Beal Avenue, Room 4118, University of Michigan, Ann Arbor, MI, 48109 USA (e-mail: rajnrao@umich.edu).

Manuscript received ??

Section VII and demonstrate the optimality of using the k_{eff} informative subspace components. We provide concluding remarks in Section VIII.

II. Problem Statement

A. Training Data Model

We are given m signal-bearing training vectors $y_i \in \mathbb{C}^{n\times 1}$, $i = 1,\ldots,m$, modeled[1] as $y_i = Ux_i + z_i$ where $z_i \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, I_n)$, U is an unknown $n \times k$ complex matrix with orthonormal columns, and $x_i \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, \Sigma)$ where $\Sigma = \mathbf{diag}(\sigma_1^2, \ldots, \sigma_k^2)$ with $\sigma_1 > \sigma_2 > \cdots > \sigma_k > 0$ unknown. For each observation, x_i and z_i are independent. The dimension, k, of our subspace is unknown and we assume throughout that $k \ll n$ so that we have a low-rank signal embedded in a high-dimensional observation vector.

Given a dimension estimate, $\widehat{k}$, and the signal bearing training data $Y = \begin{bmatrix} y_1 & \ldots & y_m \end{bmatrix}$, we form a subspace estimate $\widehat{U} \in \mathbb{C}^{n\times\widehat{k}}$ by taking the leading $\widehat{k}$ eigenvectors of YY^H/m and a signal covariance estimate $\widehat{\Sigma} \in \mathbb{R}^{\widehat{k}\times\widehat{k}}$ (in a manner to be specified).

B. Testing Data Model

We will consider two models for the test vectors. In the stochastic setting, the test vector $y \in \mathbb{C}^{n\times 1}$ is modeled as

$$\text{Stochastic Model: } y = \begin{cases} z & y \in H_0: \text{ Noise only} \\ Ux + z & y \in H_1: \text{ Signal-plus noise} \end{cases}, \tag{1}$$

where U, z, and x are modeled as described in Section II-A. This assumes that the signal, Ux, may lie anywhere in the subspace and whose position in the subspace is governed by the signal covariance matrix Σ.

In the deterministic setting, the test vector $y \in \mathbb{C}^{n\times 1}$ is modeled as

$$\text{Deterministic Model: } y = \begin{cases} z & y \in H_0: \text{ Noise only} \\ U\Sigma^{1/2}x + z & y \in H_1: \text{ Signal-plus noise} \end{cases}, \tag{2}$$

where U, Σ, and z are modeled as before. Here, in contrast to the stochastic setting, x is a non-random deterministic vector. Thus the signal, $U\Sigma^{1/2}x$, lies at a fixed point in the unknown subspace. Note that placing a mean zero, identity covariance Gaussian prior on x in (2) yields the stochastic model described in (1).

C. Problem 1: Characterize the ROC Performance Curves

Given an independent test observation from (1) or (2), we first use $\widehat{U}$ to generate a $\widehat{k} \times 1$ test vector $w = \widehat{U}^H y$. The vector w is a sufficient statistic [1] when $\widehat{U} = U$. We focus on detectors of the form

$$w^H D w \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{3}$$

where D is a diagonal matrix and the test statistic $\Lambda(w) := w^H D w$ is compared against a threshold, η, set to achieve a prescribed false alarm rate α. For detectors of this form and for test vectors modeled as (1) or (2), our goal is to

$$\text{Predict } P_D =: \mathbb{P}(\text{Detection}), \text{ for every } P_F := \alpha \in (0,1) \text{ given } n,\ m,\ \widehat{k},\ D \text{ and } \Sigma.$$

This performance prediction relies on RMT results quantifying the accuracy of the subspace estimate $\widehat{U}$ as a function of the parameters listed.

D. Problem 2: Derive and Predict the Performance of a Detector that Exploits Predictions of Subspace Accuracy

In the Neyman-Pearson setting (see [11]), a MSD is a likelihood ratio test (LRT) taking the form

$$\Lambda(w) := \frac{f(w|H_1)}{f(w|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

where $\Lambda(w)$ is the test statistic, η is the threshold set to achieve a given false alarm rate, and $w = \widehat{U}^H y$. Here, $\widehat{U}$ is a noisy estimate of the underlying U; its accuracy, relative to U, can be quantified using RMT. Therefore, $f(w|H_0)$ and $f(w|H_1)$, and consequently $\Lambda(w)$, depend on the 'noisiness' of the estimated subspaces. Our goal is thus to

> Design & analyze the performance of a detector that exploits RMT predictions of subspace estimation accuracy.

The design and performance prediction aspect of this problem will provide insights on when, if, and how the performance of plug-in detectors that do not exploit the knowledge of subspace estimation accuracy can be improved.

III. Pertinent Results from Random Matrix Theory

We are given signal-bearing training data $\{y_1, \ldots, y_m\}$ where $y_i \in H_1$ for $i = 1,\ldots,m$ from which we obtain estimates $\widehat{U}$ and $\widehat{\Sigma}$ of the unknown U and Σ. Consider the $n \times m$ data matrix $Y = \begin{bmatrix} y_1 & \ldots & y_m \end{bmatrix}$. The eigenvalue decomposition of the sample covariance matrix, $S = \frac{1}{m}YY^H$ yields the maximum likelihood estimates of U and Σ.

A. Eigenvector Aspects

Specifically, assuming k is known, we set $\widehat{U}$ to equal the $n \times k$ matrix whose columns are the k eigenvectors associated with the k largest eigenvalues of S. We now quantify the accuracy of the estimated eigenvectors.

Proposition 3.1: Assume that the columns of the training data matrix Y were generated as described in Section II-A. Let $\widehat{u}_i$ denote the eigenvector associated with the i-th largest eigenvalue of S. Then for $i = 1,\ldots,k$ and $n, m \longrightarrow \infty$ with $n/m \to c$, we have that

$$|\langle u_i, \widehat{u}_i\rangle|^2 \xrightarrow{\text{a.s.}} \begin{cases} \dfrac{\sigma_i^4 - c}{\sigma_i^4 + \sigma_i^2 c} & \text{if } \sigma_i^2 > \sqrt{c} \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

[1] For expositional simplicity, we have assumed that all our matrices and vectors are complex-valued; our results also hold for real-valued matrices and vectors.

Proof: This result appears in [8], [9]. ∎

The key insight from Proposition 3.1 is that only the eigenvectors corresponding to the signal eigenvalues lying above the phase transition $\sqrt{c}$ are *informative*. When a signal eigenvalue drops below this critical threshold, the corresponding eigenvector estimate is essentially noise-like (i.e. $|\langle u_i, \widehat{u}_i\rangle|^2 = o_p(1)$) and thus *uninformative*.

The term $|\langle u_i, \widehat{u}_i\rangle|^2$ quantifies mismatch between the estimated and underlying eigenvectors and will play an important role in deriving a new detector and in characterizing detector performance; a similar term also comes up in the analysis of the resolving power of arrays due to model mismatch such as in [12].

Following [13], we define the effective number of identifiable subspace components k_{eff} as:

$$\boxed{k_{\text{eff}} = \text{Number of } \sigma_i^2 > \sqrt{c}}. \tag{5}$$

We can estimate k_{eff} using, for example, 'Algorithm 2' of [14]. We now state an additional theorem that will prove useful in our derivations.

Theorem 3.1: Assume the same hypothesis as in Proposition 3.1. Then for fixed k and $i, j = 1, \ldots, k$ and $n, m \longrightarrow \infty$ with $n/m \to c$, we have that for $i \neq j$,

$$\langle u_j, \widehat{u}_i\rangle \xrightarrow{\text{a.s.}} 0. \tag{6}$$

Proof: In the Appendix we prove the statement for $i, j = 1, \ldots, k_{\text{eff}}$. The proof for the setting where $i, j > k_{\text{eff}}$ is considerably more involved and we omit it here for space considerations. ∎

Matrices of the form $\widehat{U}^H U D U^H \widehat{U}$ will appear throughout our derivations in Sections IV, V, and VI. We apply Proposition 3.1 and Theorem 3.1 to state a result that permits approximation, in the large matrix limit, of $\widehat{U}^H U D U^H \widehat{U}$ by a suitable diagonal matrix.

Corollary 3.1: Suppose $\widehat{k} \leq k$ and let D be a $k \times k$ (non-random) diagonal matrix such that $D = \mathbf{diag}(d_1, \ldots, d_k)$, independent of $\widehat{U}$. Then as $n, m \longrightarrow \infty$ with $n/m \to c$, we have that

$$\widehat{U}^H U D U^H \widehat{U} \xrightarrow{\text{a.s.}} \mathbf{diag}(d_1 |\langle u_1, \widehat{u}_1\rangle|^2, \ldots, d_{\widehat{k}} |\langle u_{\widehat{k}}, \widehat{u}_{\widehat{k}}\rangle|^2)$$

where for $i = 1, \ldots, \widehat{k}$ the quantity $|\langle u_i, \widehat{u}_i\rangle|^2$ is given in Proposition 3.1.

B. Eigenvalue Aspects

Random matrix theory also provides insights about the accuracy of the eigenvalues of S, as described next.

Proposition 3.2: As $n, m \longrightarrow \infty$ with $n/m \to c$, when $\sigma_i^2 > \sqrt{c}$, we have that:

$$\widehat{\sigma}_i^2 \xrightarrow{\text{a.s.}} \sigma_i^2 + c + \frac{c}{\sigma_i^2}.$$

Proof: See Theorem 2 in [8] for the real setting when $c < 1$ and [15] for the complete result. ∎

This estimate is clearly biased and we now characterize the fluctuations on the signal variance estimate.

Proposition 3.3: As $n, m \longrightarrow \infty$ with $n/m \to c$, we have that for $i = 1, \ldots, k_{\text{eff}}$

$$\sqrt{n}\left(\widehat{\sigma}_i^2 - \left(\sigma_i^2 + c + \frac{c}{\sigma_i^2}\right)\right) \Rightarrow \mathcal{N}\left(0, \frac{2\left(\sigma_i^2+1\right)^2}{\beta}\left(1 - \frac{c}{\sigma_i^4}\right)\right),$$

where $\beta = 1$ when the data is real-valued and $\beta = 2$ when the data is complex-valued.

Proof: See Theorem 3 in [8] for the real setting when $c < 1$ and [15] for the complete result. ∎

We form an improved estimate of the unknown signal variance, σ_i^2, by employing maximum-likelihood (ML) estimation on the distribution in Proposition 3.3. Specifically, for only the k_{eff} signal eigenvalues, we form the estimate:

$$\widehat{\sigma}_{i_{\text{rmt}}}^2 = \underset{\sigma_i^2}{\text{argmax}} \log\left(f_{\widehat{\sigma}_i^2}(\sigma_i^2)\right) \tag{7}$$

where

$$f_{\widehat{\sigma}_i^2}(\sigma_i^2) := \mathcal{N}\left(\left(\sigma_i^2 + c + \frac{c}{\sigma_i^2}\right), \frac{2\left(\sigma_i^2+1\right)^2}{n\beta}\left(1 - \frac{c}{\sigma_i^4}\right)\right).$$

We may then estimate $|\langle u_i, \widehat{u}_i\rangle|^2$ by substituting the improved signal variance estimates, $\widehat{\sigma}_{i_{\text{rmt}}}^2$, for the unknown σ_i^2 in Proposition 3.1. We refer to this estimate as $|\langle u_i, \widehat{u}_i\rangle|^2_{\text{rmt}}$. For the $\max(0, \widehat{k} - k_{\text{eff}})$ uninformative subspace components, we set $|\langle u_i, \widehat{u}_i\rangle|^2_{\text{rmt}} = 0$. We estimate the signal variances under the phase transition by invoking the following result.

Proposition 3.4: As $n, m \longrightarrow \infty$ with $n/m \to c$, when $\sigma_i^2 \leq \sqrt{c}$, we have that:

$$\widehat{\sigma}_i^2 \xrightarrow{\text{a.s.}} c + 2\sqrt{c}.$$

Proof: See Theorem 1 in [8] for the real setting when $c < 1$ and [15] for the complete result. ∎

We use (7) and Propositions 3.1, 3.2, and 3.4 to derive theoretical ROC curves in Section VI.

IV. Family of Stochastic Matched Subspace Detectors

We now derive a family of detectors for the stochastic observation vector model in (1). Recall that we form a test vector $w = \widehat{U}^H y$; the elements of w are the 'co-ordinates' of y in the subspace spanned by the columns of $\widehat{U}$. In the Neyman-Pearson setting, the oracle detector is a LRT which relies on the conditional distributions of our test vector w under each hypothesis. By properties of Gaussian random variables these distributions are simply

$$\begin{aligned} w|H_0 &\sim \mathcal{N}\left(0, I_{\widehat{k}}\right) \\ w|H_1 &\sim \mathcal{N}\left(0, \widehat{U}^H U \Sigma U^H \widehat{U} + I_{\widehat{k}}\right). \end{aligned} \tag{8}$$

We obtain a family of detectors by placing various assumptions on the covariance of $w|H_1$, as described next.

A. Oracle Detector

The oracle detector assumes that k, Σ, and $\widehat{U}^H U$ are all known in (8). The LRT statistic is

$$\Lambda(w) = \frac{\mathcal{N}(0, \widehat{U}^H U \Sigma U^H \widehat{U} + I_k)}{\mathcal{N}(0, I_k)}.$$

After simplification of this expression using the natural logarithm operator as a monotonic operation, the oracle statistic becomes

$$\boxed{\Lambda_{\text{oracle}}(w) = w^H \left[I_k - \left(\widehat{U}^H U \Sigma U^H \widehat{U} + I_k \right)^{-1} \right] w} \quad (9)$$

and the oracle detector is

$$\Lambda_{\text{oracle}}(w) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{oracle}} \quad (10)$$

where the threshold γ_{oracle} is chosen in the usual manner, *i.e.*, so that satisfies $P(\Lambda_{\text{oracle}}(w) > \gamma_{\text{oracle}} | H_0) = \alpha$ with α a desired false alarm rate. We note that the oracle statistic assumes that the matrix $\widehat{U}^H U \Sigma U^H \widehat{U}$ is known. Corollary 3.1 states that in the large system limit, this matrix converges almost surely to a diagonal matrix; we exploit this in Section IV-C to address the problem posed in Section II-D.

B. Plug-in Detector

When the parameters Σ and U (and the implicit variable k) in (8) are unknown, the expression in (9) cannot be computed. Suppose that we are provided a dimension estimate $\widehat{k}$ so that we may employ the generalized likelihood ratio test (GLRT) based on the distributions in (8). The GLRT statistic is

$$\Lambda(w) = \frac{\max_{U,\Sigma} \mathcal{N}(0, \widehat{U}^H U \Sigma U^H \widehat{U} + I_{\widehat{k}})}{\mathcal{N}(0, I_{\widehat{k}})}.$$

The (classical) ML estimates (in the large-sample, small matrix setting) for U and Σ are given by [16]

$$\begin{aligned} \widehat{U} &= [\widehat{u}_1 \dots \widehat{u}_{\widehat{k}}] \\ \widehat{\sigma}_i^2 &= \widehat{\lambda}_i - 1 \text{ for } i = 1, \dots, \widehat{k} \end{aligned} \quad (11)$$

where $\widehat{\lambda}_1, \dots, \widehat{\lambda}_{\widehat{k}}$ are the $\widehat{k}$ largest eigenvalues of the sample covariance matrix, S, and $\widehat{u}_1, \dots, \widehat{u}_{\widehat{k}}$ are the corresponding eigenvectors. Define the signal covariance matrix estimate as $\widehat{\Sigma} = \mathbf{diag}(\widehat{\sigma}_1^2, \dots, \widehat{\sigma}_{\widehat{k}}^2)$. We then substitute these ML estimates for the unknown parameters in (9) as in [4] and [5]. This results in the following plug-in detector statistic:

$$\Lambda_{\text{plugin}}(w) = w^H \left(I - \left[\widehat{U}^H \widehat{U} \widehat{\Sigma} \widehat{U}^H \widehat{U} + I \right]^{-1} \right) w.$$

This simplifies to

$$\boxed{\Lambda_{\text{plugin}}(w) = w^H \mathbf{diag}\left(\frac{\widehat{\sigma}_i^2}{\widehat{\sigma}_i^2 + 1} \right) w = \sum_{i=1}^{\widehat{k}} \left(\frac{\widehat{\sigma}_i^2}{\widehat{\sigma}_i^2 + 1} \right) w_i^2} \quad (12)$$

and our detector takes the form

$$\Lambda_{\text{plugin}}(w) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{plugin}} \quad (13)$$

where the threshold γ_{plugin} is chosen in the usual manner. The stochastic plug-in detector clearly takes the form of (3).

The plug-in detector assumes that the estimated signal subspace, $\widehat{U}$, is equal to the true signal subspace, U, and that the estimated signal covariance, $\widehat{\Sigma}$, is equal to the true signal covariance, Σ. In other words, the plug-in detector derivation assumes that $|\langle u_i, \widehat{u}_i \rangle|^2 = 1$ and $\widehat{\sigma}_i^2 = \sigma_i^2$ and that the provided subspace dimension estimate, $\widehat{k}$, is equal to the true underlying dimension of our signal subspace, k. Perhaps unsurprisingly, (as discussed in Section III) choosing $\widehat{k} > k_{\text{eff}}$ degrades the performance of the plug-in detector. Next we discuss an alternate viewpoint on the optimality of choosing k_{eff} components.

C. Random Matrix Theory Detector

Consider the covariance matrix of the conditional distribution $w|H_1$ in (8). By Corollary 3.1, we have that in the large matrix limit

$$\widehat{U}^H U \Sigma U^H \widehat{U} + I \xrightarrow{\text{a.s.}} \mathbf{diag}\left(|\langle u_i, \widehat{u}_i \rangle|^2 \sigma_i^2 + 1 \right). \quad (14)$$

If σ_i^2 were assumed known, this would suffice because we could plug in the results in Proposition 3.1 to get the desired statistic. We consider the setting where σ_i^2 and $|\langle u_i, \widehat{u}_i \rangle|^2$ are estimated from data and analyze the detector in the large matrix setting. In this setting, the estimate $\widehat{\sigma}_{i_{\text{rmt}}}^2$, obtained via (7), is provably consistent so that the 'plug-in' estimate of $|\langle u_i, \widehat{u}_i \rangle|^2$ based on $\widehat{\sigma}_{i_{\text{rmt}}}^2$, denoted by $|\langle u_i, \widehat{u}_i \rangle|_{\text{rmt}}^2$, is also consistent (we omit the relatively straightforward proof). Of course, there are correction terms due to finite system size effects, which we ignore, that affect the convergence properties but not the asymptotic form of the detector.

We are now in a position to address the problem posed in Section II-D. We obtain the RMT detector by substituting the aforementioned RMT estimates into the diagonal covariance matrix (14) which is subsequently used in the GLRT. After some straightforward algebra we obtain the desired RMT statistic

$$\Lambda_{\text{rmt}}(w) = \sum_{i=1}^{\widehat{k}} \left(\frac{|\langle u_i, \widehat{u}_i \rangle|_{\text{rmt}}^2 \widehat{\sigma}_{i_{\text{rmt}}}^2}{|\langle u_i, \widehat{u}_i \rangle|_{\text{rmt}}^2 \widehat{\sigma}_{i_{\text{rmt}}}^2 + 1} \right) w_i^2.$$

Note that when $i > k_{\text{eff}}$, $|\langle u_i, \widehat{u}_i \rangle|^2 \xrightarrow{\text{a.s.}} 0$ so that the sum on the right hand side (asymptotically) discards the uninformative components. Thus the RMT detector only uses the k_{eff} informative components given by (5). Consequently, we obtain the test statistic

$$\boxed{\Lambda_{\text{rmt}}(w) = \sum_{i=1}^{\min(k_{\text{eff}}, \widehat{k})} \left(\frac{|\langle u_i, \widehat{u}_i \rangle|_{\text{rmt}}^2 \widehat{\sigma}_{i_{\text{rmt}}}^2}{|\langle u_i, \widehat{u}_i \rangle|_{\text{rmt}}^2 \widehat{\sigma}_{i_{\text{rmt}}}^2 + 1} \right) w_i^2} \quad (15)$$

and the random matrix theory detector becomes

$$\Lambda_{\text{rmt}}(w) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{rmt}}, \quad (16)$$

where the threshold γ_{rmt} is chosen in the usual manner. Note that the stochastic RMT detector also takes the form of (3). The principal difference between the RMT test statistic in (15)

Detector	Detector Statistic $\Lambda(w)$	Distribution of $\Lambda\|H_0$	Distribution of $\Lambda\|H_1$
Oracle	$w^H\left[I-\left(\widehat{U}^H U\Sigma U^H\widehat{U}+I\right)^{-1}\right]w$		
Plug-in	$\sum_{i=1}^{\widehat{k}}\left(\frac{\widehat{\sigma}_i^2}{\widehat{\sigma}_i^2+1}\right)w_i^2$	$\sum_{i=1}^{\widehat{k}}\left(\frac{\widehat{\sigma}_i^2}{\widehat{\sigma}_i^2+1}\right)\chi^2_{1i}$	$\sum_{i=1}^{\widehat{k}}\left(\frac{\sigma_i^2(\sigma_i^2\|\langle u_i,u_i\rangle\|^2+1)}{\sigma_i^2+1}\right)\chi^2_{1i}$
RMT	$\sum_{i=1}^{\min(k_{\text{eff}},\widehat{k})}\left(\frac{\|\langle u_i,\widehat{u}_i\rangle\|^2_{\text{rmt}}\widehat{\sigma}^2_{i_{\text{rmt}}}}{\|\langle u_i,\widehat{u}_i\rangle\|^2_{\text{rmt}}\widehat{\sigma}^2_{i_{\text{rmt}}}+1}\right)w_i^2$	$\sum_{i=1}^{\min(k_{\text{eff}},\widehat{k})}\left(\frac{\widehat{\sigma}^2_{i_{\text{rmt}}}\|\langle u_i,\widehat{u}_i\rangle\|^2_{\text{rmt}}}{\widehat{\sigma}^2_{i_{\text{rmt}}}\|\langle u_i,\widehat{u}_i\rangle\|^2_{\text{rmt}}+1}\right)\chi^2_{1i}$	$\sum_{i=1}^{\min(k_{\text{eff}},\widehat{k})}\left(\sigma^2_{i_{\text{rmt}}}\|\langle u_i,u_i\rangle\|^2_{\text{rmt}}\right)\chi^2_{1i}$

TABLE I

GIVEN AN OBSERVATION VECTOR y FROM (1), WE FORM THE VECTOR $w=\widehat{U}^H y$ WHERE $\widehat{U}$ IS AN ESTIMATE OF THE SIGNAL SUBSPACE. THE TABLE SUMMARIZES THE TEST STATISTIC ASSOCIATED WITH EACH DETECTOR WHEN USING TESTING DATA GENERATED FROM THE STOCHASTIC MODEL. THE PLUG-IN AND RMT DETECTORS HAVE THE FORM OF (3). IN THE CFAR SETTING, THE THRESHOLD IS SET TO OBTAIN THE DESIRED FALSE ALARM PROBABILITY. NOTE THE APPEARANCE OF k_{EFF} IN THE RANDOM MATRIX THEORY DETECTOR. THE ASSOCIATED DISTRIBUTION OF EACH TEST STATISTIC UNDER H_0 AND H_1 IS PROVIDED IN THE LAST TWO COLUMNS.

and the plug-in test statistic in (12) is the role of k_{eff} in the former. The scaling factor associated with each w_i^2 for either detector is about the same; this is why the plug-in detector that uses k_{eff} components exhibits the same (asymptotic) performance as the RMT detector, which incorporates knowledge of the subspace estimate accuracy.

V. FAMILY OF DETERMINISTIC MATCHED SUBSPACE DETECTORS

We now consider the alternative deterministic test vector model (2) and derive a plug-in and RMT detector for this setting. As in the stochastic setting, we work with the processed test vector $w=\widehat{U}^H y$. When using the deterministic model, the conditional distributions of the test vector under each hypothesis are simply

$$w|H_0\sim\mathcal{N}(0,I_{\widehat{k}})$$
$$w|H_1\sim\mathcal{N}(\widehat{U}^H U\Sigma x,I_{\widehat{k}}).$$

However, as x is unknown, we employ the GLRT where $\Lambda(w)=\frac{\max_x f(w|H_1)}{f(w|H_0)}$. The GLRT statistic for our processed data w is

$$\Lambda(w)=\frac{\max_x\mathcal{N}(\widehat{U}^H U\Sigma x,I_{\widehat{k}})}{\mathcal{N}(0,I_{\widehat{k}})}. \tag{17}$$

Given a dimension estimate, $\widehat{k}$, $\widehat{U}$ is a $n\times\widehat{k}$ matrix comprised of the top $\widehat{k}$ eigenvectors of the sample covariance matrix of the training data. This is exactly the same estimate as in the stochastic setting. Employing maximum likelihood estimation on x in the GLRT in (17) yields the estimate $\widehat{x}=\left(\Sigma U^H\widehat{U}\widehat{U}^H U\Sigma\right)^{-1}\Sigma U^H\widehat{U}w$. After simplifying using $\widehat{x}$ and using the natural logarithm operator as a monotonic operation, the GLRT statistic becomes

$$\Lambda(w)=w^H\left(\widehat{U}^H U\Sigma\left(\Sigma U^H\widehat{U}\widehat{U}^H U\Sigma\right)^{-1}\Sigma U^H\widehat{U}\right)w$$

which simplifies to

$$\Lambda_(w)=w^H\left(\widehat{U}^H U\left(U^H\widehat{U}\widehat{U}^H U\right)^{-1}U^H\widehat{U}\right)w. \tag{18}$$

Notice that Σ does not appear in the test statistic.

A. Plug-in Detector

The statistic in (18) is not realizable as k, U, and Σ are unknown. One may then substitute a ML estimate for U in (18) as in [4] and [5]. The ML estimate (in the large-sample, small matrix setting) for U is the same as in the stochastic setting (see (11)) because we have not altered the assumptions on the training data.

By replacing U with the estimate $\widehat{U}$ we obtain the plug-in detector which employs the test statistic

$$\Lambda_{\text{plugin}}(w)=w^H\left(\widehat{U}^H\widehat{U}\left(\widehat{U}^H\widehat{U}\widehat{U}^H\widehat{U}\right)^{-1}\widehat{U}^H\widehat{U}\right)w.$$

This simplifies to

$$\boxed{\Lambda_{\text{plugin}}(w)=w^H w=\sum_{i=1}^{\widehat{k}}w_i^2} \tag{19}$$

and our detector becomes

$$\Lambda_{\text{plugin}}(w)\underset{H_0}{\overset{H_1}{\gtrless}}\gamma_{\text{plugin}}, \tag{20}$$

where the threshold γ_{plugin} is chosen in the usual manner. This deterministic plug-in detector is an 'energy detector' and also takes the form of (3).

The plug-in detector assumes that the estimated signal subspace $\widehat{U}$ is equal to the true signal subspace U. We saw in the stochastic setting that we should only choose k_{eff} components. We discuss the (asymptotic) optimality of choosing k_{eff} components for deterministic detectors next.

B. Random Matrix Theory Detector

Consider the term $\widehat{U}^H U$. By Corollary 3.1 and by noting that the eigenvectors are unique up to a phase, we have that $\widehat{U}^H U\xrightarrow{\text{a.s.}}SA$ where S is a $\widehat{k}\times\min(\widehat{k},k)$ matrix and A is a $\min(\widehat{k},k)\times k$ matrix defined as

$$S_{i\ell}:=\begin{cases}s_i=\exp(j\psi_i) & i=\ell\\ 0 & \text{o.w.}\end{cases},\quad A_{i\ell}:=\begin{cases}a_i=|\langle u_i,\widehat{u}_i\rangle| & i=\ell\\ 0 & \text{o.w.}\end{cases}.$$

For some ψ_i, s_i denotes the random phase ambiguity in the eigenvector computation (since eigenvectors are unique up to a phase).

The plug-in detector assumes that $A=S=I_{\widehat{k}}$, that is $s_i=1$ and $|\langle u_i,\widehat{u}_i\rangle|=1$. However, as seen in Section III, we have knowledge of $|\langle u_i,\widehat{u}_i\rangle|$ which we may exploit in deriving a new detector. Using the notation just developed, the GLRT statistic may be written as

$$\Lambda(w)=w^H SA(A^H S^H SA)^{-1}A^H S^H w.$$

We use Proposition 3.1 to estimate $a_i=\sqrt{|\langle u_i,\widehat{u}_i\rangle|^2_{\text{rmt}}}$. Recall that k_{eff} is the number of σ_i^2 above the phase transition and note that $a_i=0$ when $\sigma_i^2\leq\sqrt{c}$. Incorporating this into the detector, and noting that A and S contain only diagonal elements, the GLRT simplifies to

$$\Lambda(w)=w^H\begin{bmatrix}I_{\min(\widehat{k},k_{\text{eff}})} & 0\\ 0 & 0_{\widehat{k}-\min(\widehat{k},k_{\text{eff}})}\end{bmatrix}w.$$

After simplification, the RMT statistic becomes

$$\boxed{\Lambda_{\text{rmt}}(w) = \sum_{i=1}^{\min(\widehat{k},k_{\text{eff}})} w_i^2} \tag{21}$$

and our detector becomes

$$\Lambda_{\text{rmt}}(w) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{\text{rmt}}, \tag{22}$$

where the threshold γ_{rmt} is chosen in the usual manner. This addresses the problem posed in Section II-D for the deterministic test vector setting. We note that this deterministic RMT detector also takes on the form of (3). As in the stochastic setting, the principal difference between the RMT test statistic in (21) and the plug-in test statistic in (19) is the role of k_{eff} in the former. This is also why the plug-in detector that uses k_{eff} components exhibits the same performance as the RMT detector, which incorporates knowledge of the subspace estimates. Both of these detectors are 'energy detectors'.

VI. Theoretical ROC Curve Predictions

We saw in Sections IV and V that all the derived detectors were (exactly or asymptotically) of the form given by (3). Thus by answering the ROC curve prediction problem posed in Section II-C, we have characterized the asymptotic (or large system) performance of the detectors considered herein. For the following analysis, we are given n, m, $\widehat{k}$, D and Σ.

We first note that each previously derived detector corresponds to a specific choice of the diagonal matrix D in (3), which can be discerned by inspection of Tables I and II. In what follows, we solve the ROC prediction problem for general D; direct substitution of the relevant parameters for D will yield the performance curves for individual detectors. We plan to release a software implementation of the method described that will allow practitioners to generate the performance curves of the desired detectors.

Recall that the ROC or receiver operating characteristic curve [17] for a test statistic $\Lambda(w)$ is obtained by computing

$$\begin{aligned} P_D &= P(\Lambda(w) \geq \gamma | w \in H_1) \\ P_F &= P(\Lambda(w) \geq \gamma | w \in H_0) \end{aligned} \tag{23}$$

for $-\infty < \gamma < \infty$ and plotting P_D versus P_F. To compute these expressions in (23) for the deterministic and stochastic test vector setting, we need to characterize the conditional c.d.f under H_0 and H_1 for a detector with a test statistic of the form

$$w^H D w = \sum_{i=1}^{\widehat{k}} d_i w_i^2.$$

The results in Section III, especially an application of Corollary 3.1, simplify this analysis in the large system limit.

A. Stochastic Testing Setting

In the stochastic setting, the conditional distributions of our test samples under each hypothesis are $w|H_0 \sim \mathcal{N}(0, I_{\widehat{k}})$ and $w|H_1 \sim \mathcal{N}(0, \widehat{U}^H U \Sigma U^H \widehat{U} + I_{\widehat{k}})$. Because the covariance matrix of $w|H_0$ is diagonal, for $i = 1, \ldots, \widehat{k}$, $w_i|H_0 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ which implies that $w_i^2|H_0 \overset{\text{i.i.d.}}{\sim} \chi_1^2$. By Corollary 3.1, the covariance matrix of $w|H_1$ is asymptotically diagonal. Therefore for $i = 1, \ldots, \widehat{k}$, $w_i|H_1 \overset{\text{i.i.d.}}{\approx} \mathcal{N}(0, \sigma_i^2 |\langle u_i, \widehat{u}_i \rangle|^2 + 1)$ and

$$\frac{w_i^2|H_1}{\sigma_i^2 |\langle u_i, \widehat{u}_i \rangle|^2 + 1} \sim \chi_1^2.$$

Using this analysis, for a detector with the form of (3), the conditional distributions of its test statistic under each hypothesis are

$$\begin{aligned} \Lambda(w)|H_0 &\sim \sum_{i=1}^{\widehat{k}} d_i \chi_{1i}^2 \\ \Lambda(w)|H_1 &\sim \sum_{i=1}^{\widehat{k}} d_i (\sigma_i^2 |\langle u_i, \widehat{u}_i \rangle|^2 + 1) \chi_{1i}^2 \end{aligned} \tag{24}$$

where χ_{1i}^2 are independent chi-square random variables. The third and fourth columns of Table I use this general analysis to summarize the sample conditional distributions of $\Lambda(w)$ under each hypothesis for the stochastic plug-in and RMT detectors. An analytical expression for the asymptotic performance in the large matrix limit is obtained by substituting expressions from (7) and Propositions 3.1, 3.2, and 3.4 for the pertinent quantities in these distributions.

Note that the conditional distributions in (24) are a weighted sum of independent chi-square random variables with one degree of freedom. The c.d.f. of a chi-square random variable is known in closed form. However, the c.d.f. of a weighted sum of independent chi-square random variables is not known in closed form. To evaluate (23), we use a saddlepoint approximation of the conditional c.d.f. of $\Lambda(w)$ by employing the generalized Lugannani-Rice formula proposed in [18]. To then compute a theoretical ROC curve, we sweep γ over $(0, \infty)$ and for each value of γ, we compute the saddlepoint approximation of the conditional c.d.f. under each hypothesis using this method. This generates a set of points (P_F, P_D) which approximate the (asymptotic) theoretical ROC curve.

B. Deterministic Testing Setting

In the deterministic setting, the conditional distribution of a test sample under H_0 is $w|H_0 \sim \mathcal{N}(0, I_{\widehat{k}})$. The conditional distribution of w under H_1 is $w|H_1 \sim \mathcal{N}(\widehat{U}^H U \Sigma x, I_{\widehat{k}})$. By Corollary 3.1, $\widehat{U}^H U \xrightarrow{\text{a.s.}} SA$ is asymptotically diagonal with S and A defined in Section V-B. We have that $w_i|H_1 \overset{\text{i.i.d.}}{\approx} \mathcal{N}(a_i s_i \sigma_i x_i, 1)$ for $i = 1, \ldots, \widehat{k}$. Using this approximation, for a detector with the form of (3), the conditional distributions of its test statistic are

$$\begin{aligned} \Lambda(w)|H_0 &\sim \sum_{i=1}^{\widehat{k}} d_i \chi_{1i}^2 \\ \Lambda(w)|H_1 &\sim \sum_{i=1}^{\widehat{k}} d_i \chi_{1i}^2(\delta_i) \end{aligned} \tag{25}$$

where $\delta_i = \sigma_i^2 |\langle u_i, \widehat{u}_i \rangle|^2 x_i^2$ is the non-centrality parameter for the noncentral chi-square distribution. The deterministic plug-in and RMT detectors are a special case of these conditional

Detector	Detector Statistic $\Lambda(w)$	Distribution of $\Lambda\|H_0$	Distribution of $\Lambda\|H_1$
Plug-in	$\sum_{i=1}^{\widehat{k}} w_i^2$	$\chi^2_{\widehat{k}}$	$\chi^2_{\widehat{k}}\left(\sum_{i=1}^{\min(\widehat{k},k_{\text{eff}})} \sigma_i^2 \lvert\langle u_i, \widehat{u}_i\rangle\rvert^2 x_i^2\right)$
RMT	$\sum_{i=1}^{\min(k_{\text{eff}},\widehat{k})} w_i^2$	$\chi^2_{\min(\widehat{k},k_{\text{eff}})}$	$\chi^2_{\min(\widehat{k},k_{\text{eff}})}\left(\sum_{i=1}^{\min(\widehat{k},k_{\text{eff}})} \sigma_i^2 \lvert\langle u_i, \widehat{u}_i\rangle\rvert^2 x_i^2\right)$

TABLE II

GIVEN AN OBSERVATION VECTOR y FROM (2), WE FORM THE VECTOR $w = \widehat{U}^H y$ WHERE $\widehat{U}$ IS AN ESTIMATE OF THE SIGNAL SUBSPACE. THE TABLE SUMMARIZES THE TEST STATISTIC ASSOCIATED WITH EACH DETECTOR FOR THE DETERMINISTIC SETTING. THE PLUG-IN AND RMT DETECTORS HAVE THE FORM OF (3). IN THE CFAR SETTING, THE THRESHOLD IS SET TO OBTAIN THE DESIRED FALSE ALARM PROBABILITY. NOTE THE APPEARANCE OF k_{EFF} IN THE RMT DETECTOR. THE ASSOCIATED DISTRIBUTION OF EACH TEST STATISTIC UNDER H_0 AND H_1 IS PROVIDED IN THE LAST TWO COLUMNS. THE NOTATION $\chi^2_k(\delta)$ IS A NONCENTRAL CHI-SQUARE RANDOM VARIABLE WITH k DEGREES OF FREEDOM AND NON-CENTRALITY PARAMETER δ.

distributions. For the plug-in detector, $d_i = 1$ for $i = 1, \ldots, \widehat{k}$. For the RMT detector $d_i = 1$ for $i = 1, \ldots, \min(\widehat{k}, k_{\text{eff}})$ and $d_i = 0$ for $i = \min(\widehat{k}, k_{\text{eff}}) + 1, \ldots, \widehat{k}$.

For the plug-in and RMT detectors, $\Lambda_{\text{plugin}}(w)|H_0 \sim \chi^2_{\widehat{k}}$ and $\Lambda_{\text{rmt}}(w)|H_0 \sim \chi^2_{\min(\widehat{k},k_{\text{eff}})}$. Similarly, $\Lambda_{\text{plugin}}(w)|H_1 \sim \chi^2_{\widehat{k}}(\delta)$ and $\Lambda_{\text{rmt}}(w)|H_1 \sim \chi^2_{\min(\widehat{k},k_{\text{eff}})}(\delta)$ where

$$\delta = \sum_{i=1}^{\widehat{k}} \sigma_i^2 |\langle u_i, \widehat{u}_i\rangle|^2 x_i^2 = \sum_{i=1}^{\min(\widehat{k},k_{\text{eff}})} \sigma_i^2 |\langle u_i, \widehat{u}_i\rangle|^2 x_i^2. \quad (26)$$

Because $d_i = 1$ for $i = 1, \ldots, \min(\widehat{k}, k_{\text{eff}})$ for both the plug-in and RMT detectors, the resulting non-centrality parameter is the sum of all the individual non-centrality parameters. An analytical expression for the asymptotic performance in the large matrix limit is obtained by substituting expressions from Proposition 3.1 in (26). Unlike the stochastic setting, we can obtain a closed form expression for the deterministic plug-in and RMT ROC curves by solving for γ in terms of P_F and substituting this into the expression for P_D in (23). Doing so yields

$$\begin{aligned} P_{D_{\text{plugin}}} &= 1 - Q_{\chi^2_{\widehat{k}}(\delta)}\left(Q^{-1}_{\chi^2_{\widehat{k}}}(1 - P_F)\right) \\ P_{D_{\text{rmt}}} &= 1 - Q^2_{\chi^2_{\min(\widehat{k},k_{\text{eff}})}(\delta)}\left(Q^{-1}_{\chi^2_{\min(\widehat{k},k_{\text{eff}})}}(1 - P_F)\right) \end{aligned} \quad (27)$$

where Q is the appropriate c.d.f. function.

We have obtained an analytical characterization of the (asymptotic) ROC performance curve for the deterministic and stochastic test vector settings, thus answering the question posed in Section II-C.

VII. DISCUSSION AND INSIGHTS

We use numerical simulations to verify our theoretical ROC curve predictions from Section VI that rely on RMT approximations presented in Section III. We also demonstrate properties of the new RMT detectors that we derived in Sections IV and V using numerical simulations, as described next.

A. Simulation Protocol

To compute an empirical ROC curve, we first generate a random subspace, U, by taking the first k left singular vectors of a random matrix with i.i.d. $\mathcal{N}(0, 1)$ entries. Using this U, we generate training samples as described in Section II-A from which we form estimates $\widehat{U}$ and $\widehat{\Sigma}$ from the eigenvalue decomposition of the sample covariance matrix as described in (11).

We then generate a desired number of test samples from each hypothesis using either (1) or (2). For each test sample, we compute the test statistic for each detector. Using Fawcett's [17] 'Algorithm 2', we compute an empirical ROC curve. This is repeated for multiple realizations of U, generating multiple empirical ROC curves. We refer to a single empirical ROC curve corresponding to a realization of U as a trial. We then average the empirical ROC curves over multiple trials using Fawcett's [17] 'Algorithm 4'.

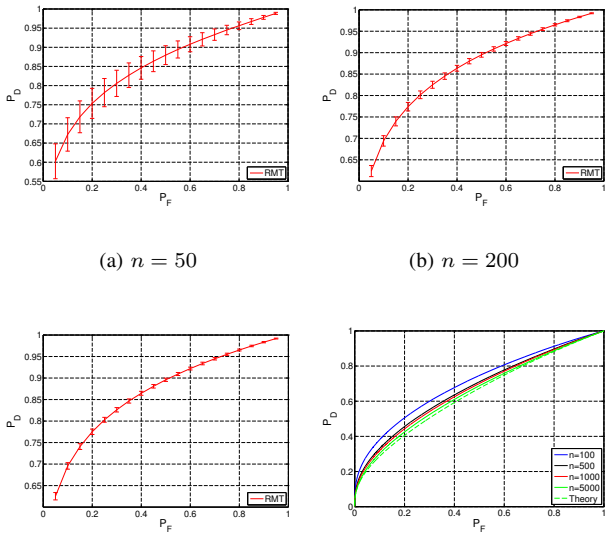
B. Convergence

The derived detectors and their subsequent theoretical ROC curve predictions rely on the asymptotic approximations which ignore finite n and m correction terms. To examine the validity of the asymptotic approximations and the rate of convergence, we first consider a setting where $\widehat{k} = k = k_{\text{eff}} = 2$ and $c = 2$. Figures 1(a)-1(c) plot the average empirical ROC curve for the stochastic RMT detector as well as one standard deviation error bars for $n = 50, 200, 1000$, respectively. As expected, as n increases the size of the error bars decreases, attesting to the asymptotic convergence of the RMT approximations. Analyzing the rate of convergence (which we conjecture to be $n^{1/2}$) is an important open problem which we shall tackle in future work.

Figure 1(d) considers the stochastic plug-in detector under a different setting where $\widehat{k} = k = 4$ but $k_{\text{eff}} = 1$ and $c = 20$. This value of c corresponds to the sample starved regime where $n > m$. We see that for small values of n, the empirical ROC curve is not well approximated by the asymptotic theoretical predictions. However, as n increases, the empirical bias decreases. Theorem 3.1 shows that the off diagonal terms in (14) asymptotically tend to zero but in the finite n and m case, these terms are $O(1/\sqrt{n})$ and thus not identically zero. For larger rank systems (increased k), there are more of these non-identically-zero terms which worsens the approximation quality for fixed, relatively small n. As n increases, this bias vanishes.

C. Accuracy of the RMT Approximations

The stochastic RMT detector approximates the oracle detector by employing the RMT approximations, for large n and m, developed in Section III. To test the accuracy of the



(a) $n = 50$ (b) $n = 200$

(c) $n = 1000$ (d) Bias

Fig. 1. (a)-(c):Empirical ROC curves with one standard deviation errorbars for the stochastic RMT detector. Empirical ROC curves were simulated with $c = 2$, $\widehat{k} = k = 2$, and $\Sigma = \mathbf{diag}\,(10, 5)$. The empirical ROC curves were computed using 10000 test samples and averaged over 100 trials using algorithms 2 and 4 of [17]. As both signal variances are above the critical threshold, $k_{\text{eff}} = 2$. For Figure 1(a) $n = 50$ and $m = 25$. For Figure 1(b) $n = 200$ and $m = 100$. For Figure 1(c) $n = 1000$ and $m = 500$. (d): Empirical and theoretical ROC curves for the stochastic plug-in detector. Empirical ROC curves were simulated with $\widehat{k} = k = 4$, $n/m = c = 20$, $\Sigma = \mathbf{diag}(10, 3, 2.5, 2)$ so that $k_{\text{eff}} = 1$. The empirical ROC curves were computed using 10000 test samples and averaged over 100 trials using algorithms 2 and 4 of [17]. As n increases, the empirical bias in the ROC curves vanishes.

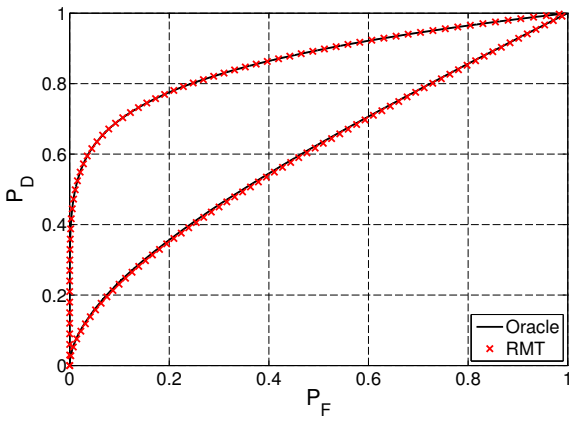


Fig. 2. (top curve): Empirical ROC curves for the oracle and RMT stochastic detectors (10) and (16) respectively. Empirical ROC curves were simulated with $n = 1000$, $m = 500$, $\widehat{k} = k = 2$, and $\Sigma = \mathbf{diag}(10, 5)$. As both $\sigma_1^2 = 10$ and $\sigma_2^2 = 5$ are above the critical threshold $\sqrt{c} = \sqrt{2}$, $k_{\text{eff}} = 2$ by equation (5). The empirical ROC curves were computed as descirbed is Section VII-A using 10000 test samples generated from (1) and averaged over 100 trials using algorithms 2 and 4 of [17]. The stochastic RMT detector realizes the same performance as the oracle detector thereby validating the approximation of Corollary 3.1. (bottom curve): This is the same setting as the top curve except $\Sigma = \mathbf{diag}(2.5, 1.25)$. In this case, as $\sigma_2^2 = 1.25 < \sqrt{2}$, $k_{\text{eff}} = 1$ by equation (5). Even when k_{eff} changes, the stochastic RMT detector is a good approximation to the oracle statistic giving further credibility to Corollary 3.1.

approximations for finite n and m, we examine the empirical performance of both detectors. Figure 2 plots empirical ROC curves for the RMT and oracle stochastic detectors. In the top curves $\widehat{k} = k = k_{\text{eff}} = 2$ and in the bottom curves $\widehat{k} = k = 2 > k_{\text{eff}} = 1$. For both values of k_{eff}, the RMT detector is a good approximation to the oracle detector thereby validating the approximation from Corollary 3.1.

D. Accuracy of ROC Curve Prediction

To test the accuracy of the ROC predictions developed in Section VI which rely on the random matrix theoretic approximations presented in Section III, we consider a setting where $\widehat{k} = k = 2$. Figure 3 plots empirical and theoretical ROC curves for the plug-in and RMT stochastic detectors for $\Sigma = \alpha\,\mathbf{diag}(10, 5)$ for three choices of α. Figure 4 plots empirical and theoretical ROC curves for the plug-in and RMT deterministic detectors for $\Sigma = \mathbf{diag}(10, 5)$ for three choices of the deterministic test vector x. The choices of Σ and x affect the performance of the resulting detectors.

In both settings, the theoretical ROC curves match the empirical ROC curves thereby validating the accuracy of the random matrix theoretic approximations employed and the accuracy of the saddlepoint approximation to the c.d.f. used in the stochastic derivation. In the stochastic setting, using $\alpha = 0.25$ results in $k_{\text{eff}} = 1$. As $\widehat{k} > k_{\text{eff}}$, the

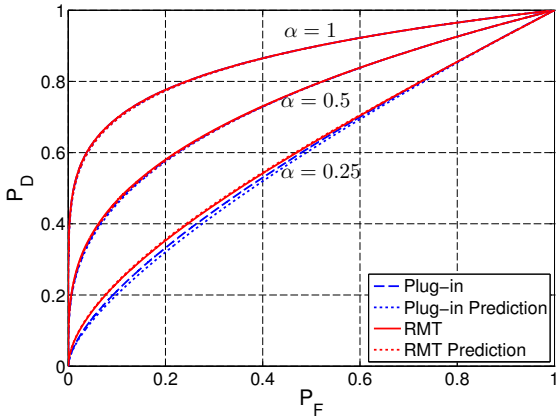


Fig. 3. Empirical and theoretical ROC curves for the plug-in and RMT stochastic detectors (13) and (16) respectively. Empirical ROC curves were simulated as in Figure 2 except with $\Sigma = \alpha\,\mathbf{diag}\,(10, 5)$. Results are plotted for $\alpha = 1, 0.5, 0.25$. For $\alpha = 1$ and $\alpha = 0.5$, $\widehat{k} = k = k_{\text{eff}} = 2$ by (5). For $\alpha = 0.25$, $k_{\text{eff}} = 1$. The theoretical ROC curves were computed as described in Section VI-A. For all values of α, all theoretically predicted ROC curves match the empirical results reflecting the accuracy of the approximations employed and the Lugannani-Rice formula. Since $\widehat{k} > k_{\text{eff}}$ when $\alpha = 0.25$, we observe a performance gain when using the RMT detector.

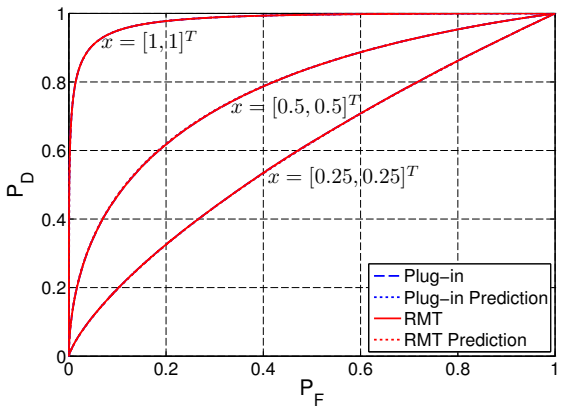


Fig. 4. Empirical and theoretical ROC curves for the plug-in and RMT deterministic detectors (20) and (22) respectively. Empirical ROC curves were simulated as in Figure 2 except with test samples generated from (2) and $\Sigma = \mathbf{diag}(10, 5)$ so that $k_{\text{eff}} = 2$. Three values of the deterministic signal vector were used: $x = [1, 1]^T$, $x = [0.5, 0.5]^T$, and $x = [0.25, 0.25]^T$. The theoretical ROC curves were computed using (27). The theoretically predicted ROC curves match the empirical results reflecting the accuracy of Corollary 3.1. The resulting ROC curves depend on the choice of x, however, since $\widehat{k} = k_{\text{eff}}$, the plug-in and RMT detector achieve the same performance for all x.

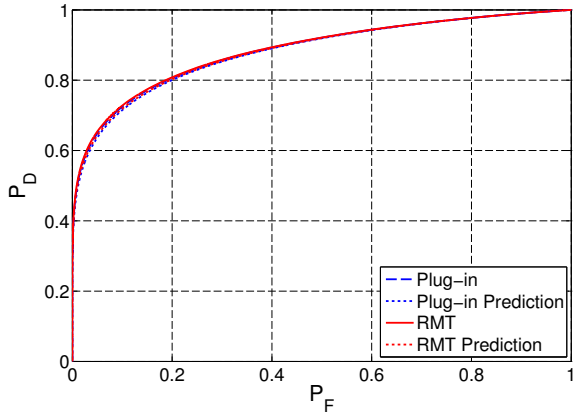


Fig. 5. Empirical and theoretical ROC curves for the plug-in and RMT stochastic detectors. This is the same setup as Figure 3 except with $n = m = 5000$, $\widehat{k} = k = 4$ and $\Sigma = \mathbf{diag}(\mathbf{10}, \mathbf{3}, \mathbf{2.5}, \mathbf{2})$ so that $k_{\text{eff}} = 4$. As the choice of m resulted in $k_{\text{eff}} = \widehat{k} = 4$, the plug-in and RMT detectors achieve relatively the same performance.

plug-in detector realizes a performance loss. As we might intuitively expect, larger values of α result in better detector performances as larger α yield larger signal variances. In the deterministic setting, larger values of $|x|$ result in better detector performance. However, as the test vector does not affect the value of $k_{\text{eff}} = \widehat{k}$, the plug-in and RMT detectors achieve the same performance because they have identical statistics.

E. Effect of the Number of Training Samples

As derived in Section III, for a fixed Σ, the number of training samples, m, directly affects k_{eff} via (5). To explore how the number of training samples affects detector performance, we consider the stochastic setting where $\widehat{k} = k = 4$. Figure 5 investigates the performance when $m = n$ so that $c = 1$.

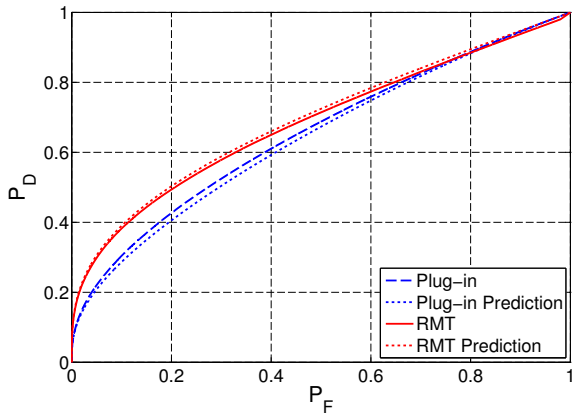


Fig. 6. Empirical and theoretical ROC curves for the plug-in and RMT stochastic detectors with the same setup as Figure 5 except with $m = 250$ so that $k_{\text{eff}} = 1$. As the choice of m resulted in $k_{\text{eff}} = 1 < \widehat{k} = 4$, the plug-in detector is suboptimal. The disagreement between the theoretical and empirical ROC curves is attributed to finite dimensionality.

With the choice of $\Sigma = \mathbf{diag}(10, 3, 2.5, 2)$, this choice of m results in $k_{\text{eff}} = \widehat{k} = 4$. In this setting, the plug-in and RMT detectors achieve relatively the same performance. Figure 6 chooses $20m = n$ so that $c = 20$ and $k_{\text{eff}} = 1$.

In the second setting, the plug-in detector becomes suboptimal because it uses $4 = \widehat{k} > k_{\text{eff}} = 1$ subspace components. Whenever $k_{\text{eff}} < \widehat{k}$ the RMT detector avoids the performance loss realized by the plug-in detector. We could have produced the same effect as this experiment by varying Σ instead of m as both of these quantities drive the value of k_{eff}.

The disagreement between the theoretical and empirical ROC curves is attributed to the finite n and m correction terms which we have neglected. As k increases, there are more off-diagonal terms in the covariance matrix in (14) that are not identically zero, as discussed earlier. Theorem 3.1 shows that asymptotically these off diagonal terms decay to zero but in the finite n and m case, they are non-zero. In the case of larger k, there are more of these non-identically-zero terms thereby worsening the approximation.

F. Effect of $\widehat{k}$

Figure 7 examines the performance of the plug-in and RMT stochastic detectors as a function of $\widehat{k}$ by plotting the empirically achieved probability of detection for a constant false alarm rate of 0.01. The result confirms that k_{eff} is the optimal choice for $\widehat{k}$. When the plug-in detector uses $\widehat{k} = k_{\text{eff}}$ it achieves an equivalent performance as the RMT detector.

Setting $\widehat{k} < k_{\text{eff}}$ drastically degrades performance for both detectors. In this regime, the plug-in and RMT detectors realize the same ROC performance, demonstrating that quantification and exploitation of the subspace estimation accuracy, while useful in ROC performance prediction, does *not* noticeably enhance detection performance. When $\widehat{k} > k_{\text{eff}}$, the performance of the plug-in detector degrades while that of the RMT detector is stable as if $\widehat{k} = k_{\text{eff}}$. In other words, we do not pay a price for overestimating the subspace dimension with the RMT detector. This makes sense (and is slightly contrived) because the RMT detector will only sum to a maximum of k_{eff}

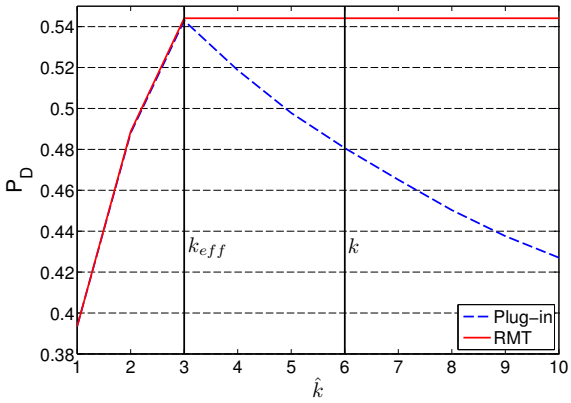


Fig. 7. Empirical exploration of the achieved probability of detection, P_D, for a fixed probability of false alarm, $P_F = 0.01$, for various $\widehat{k}$. We are in the same setting as Figure 3 except with $k = 6$ and $\Sigma = \mathbf{diag}(\mathbf{10}, \mathbf{5}, \mathbf{4}, 1, 0.75, 0.5)$ so that $k_{\text{eff}} = 3$. The optimal $\widehat{k}$ resulting in the largest P_D is not the true k, but rather k_{eff}.

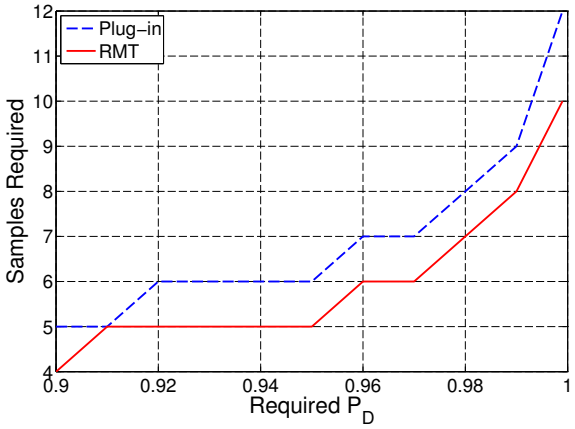


Fig. 8. Empirical exploration of the number of test samples required to achieve a desired $P_F = 0.001$ for different P_D. In this setup, the test statistic for each detector is averaged over multiple independent test samples. We are in the same setting as Figure 7 except with $\widehat{k} = k = 6$. Because the RMT detector only uses informative subspace components, it can average less test samples than the plug-in detector to achieve a desired (P_F, P_D) pair.

indices as evident in (15). In many applications, practitioners might employ the "play-it-safe" approach and set $\widehat{k}$ to be significantly greater than k_{eff}. The performance loss caused by adding each uninformative subspace, as seen in Figure 7, constitutes evidence to the assertion that overestimating the signal subspace dimension is a bad idea. When $k_{\text{eff}} < k$, even perfectly estimating the subspace dimension (i.e. setting $\widehat{k} = k$) is suboptimal.

G. Effect of Averaging Multiple Test Samples

Figure 8 considers the performance of the plug-in and RMT stochastic detectors when averaging multiple independent test samples to form the test statistic. In this setting, $k_{\text{eff}} = 3$ and the plug-in detector sets $\widehat{k} = k = 6$. The RMT detector can average less samples than the plug-in detector to achieve a desired (P_F, P_D) pair. Since the RMT detector discards the uninformative signal subspace components, each test sample is pruned to ignore noisy components. As the plug-in detector includes uninformative subspace components, its test statistic is noisier and so it must be averaged over more samples to achieve a desired level of detection accuracy. This requires waiting for additional test vectors thereby increasing the time-to-decision.

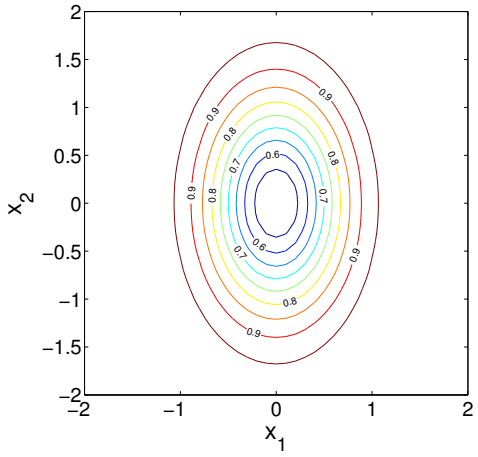


Fig. 9. Contour plot of the theoretical area under the ROC curve (AUC) for the deterministic RMT detector. We are in the same setting as Figure 4. As both signal variances are above the critical threshold, $k_{\text{eff}} = 2$. We see that the orientation of the contours follows the inverse signal variance structure. As $\sigma_1^2 = 10 > \sigma_2^2 = 5$, the vertical contour axis is longer than the horizontal one.

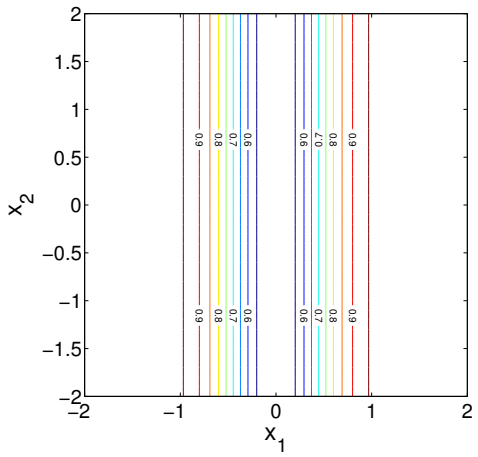


Fig. 10. Contour plot of the theoretical area under the ROC curve (AUC) for the deterministic RMT detector. We are in the same setting as Figure 9 except that $\Sigma = \mathbf{diag}(10, 1)$. As only the first signal variance is above the critical threshold, $k_{\text{eff}} = 1$. Because of this, the deterministic RMT detector will ignore the second, uninformative component of the test vector. We see this manifested in the contour plot as all contours are vertical lines indicating that x_2 has no impact on the RMT detector.

H. Effect of x in the Deterministic Setting

As seen in the derivation of the theoretical deterministic ROC curves, the choice of the signal vector, x, affects the performance of the detector through the non-centrality parameter, (26), of the noncentral chi-square distribution. To explore the effect that x has on detector performance, we consider the setting where $\widehat{k} = k = 2$. Figures 9 and 10 plot the area

under the ROC curve (AUC) for various choices of x for the theoretical RMT detector for two choices of Σ.

As seen in Figure 9, because $\widehat{k} = k_{\text{eff}}$, larger $|x|$ increase the AUC. Intuitively this makes sense because larger $|x|$ increase the non-centrality parameter, which further separates the conditional distributions of the test statistic under each hypothesis, resulting in better detection ability (higher AUC). The AUC contour lines have major axes inversely proportional to the training data signal covariance matrix, Σ. In this example, $\sigma_1^2 > \sigma_2^2$; increasing $|x_1|$ increases the AUC more than increasing $|x_2|$. In other words, as $\sigma_1^2 > \sigma_2^2$, to achieve the same increase in AUC, we would need to increase $|x_2|$ more than $|x_1|$. This is evident in the geometry of the contours.

However, in Figure 10 the choice of Σ results in $k_{\text{eff}} = 1 < \widehat{k} = 2$. We see the geometry of the AUC contour lines change to vertical lines. This makes sense because the second subspace component is uninformative. Therefore, changes in $|x_2|$ do not affect the performance of the detector, resulting in vertical AUC contour lines.

By placing a standard normal prior on x, the stochastic detector may be viewed as the expectation over all deterministic detectors. In fact, we can place any prior on x and obtain a different stochastic detector by taking the expectation over all deterministic detectors with a given prior. This facilitates ROC performance curve analysis over non-Gaussian priors on x.

VIII. Conclusion

In this paper, we considered a matched subspace detection problem where the low-rank signal subspace is unknown and must be estimated from finite, noisy, signal-bearing training data. We considered both a stochastic and deterministic model for our testing data. The subspace estimate, formed from the eigenvalue decomposition of the sample covariance matrix, is noisy due to finite training samples and noise in the training observation vectors. We used recent results from random matrix theory to precisely quantify the accuracy of the eigenvectors of a sample covariance matrix. An important insight from these results is that only the subspace components whose eigen-SNR is above a certain threshold are informative [13] with predictable bias; the remaining subspace components are uninformative.

Armed with this knowledge, we derived a new RMT detector that only uses the effective number of informative subspace components, k_{eff}. A detector that uses the uninformative components will thus incur a performance degradation, relative to the RMT detector. However, whenever $\widehat{k} = k_{\text{eff}}$, the plug-in detector that uses $\widehat{k}$ subspace components will exhibit the same performance as the RMT detector. This highlights the importance of robust techniques [14], [19], [20] for estimating k_{eff} in subspace based detection schemes.

We also used the precise quantification of subspace estimation errors, provided by random matrix theory, to predict the ROC performance of both stochastic and deterministic detectors whose test statistic takes the general form of $\Lambda(w) = w^H D w$. We saw that the conditional distributions of the test statistics can be expressed as a weighted sum of chi-square random variables. The ROC curve can then be computed using a saddlepoint approximation method in the stochastic setting and in closed form in the deterministic setting.

We conclude by noting the interplay between the stochastic and deterministic models considered in (1) and (2), respectively. Placing a standard normal prior on x in (2) yields the stochastic model in (1). Hence, we might have computed the ROC performance curve for (1) by integrating the ROC curve conditioned for the deterministic setting, *conditioned* on a value of x. We may thus use this idea to derive the asymptotic ROC performance curve for other non-normal distributions for x. In future work, we plan to investigate aspects related to the rate of convergence.

Appendix

Proof: We provide the proof of Theorem 3.1 when $i \neq j$.

Let $U_{n,k}$ be a $n \times k$ real or complex matrix with orthonormal columns, u_i for $1 \leq i \leq k$. Let $\Sigma = \mathbf{diag}\left(\sigma_1^2, \ldots, \sigma_k^2\right)$ such that $\sigma_1^2 > \sigma_2^2 > \cdots > \sigma_k^2 > 0$ for $k \geq 1$. Define $P_n = U_{n,k} \Sigma U_{n,k}^H$ so that P_n is rank-k. Let Z_n be a $n \times m$ real or complex matrix with independent $\mathcal{CN}(0,1)$ entries. Let $X_n = \frac{1}{m} Z_n Z_n^H$, which is a random Wishart matrix, have eigenvalues $\lambda_1(X_n) \geq \cdots \geq \lambda_n(X_n)$. Let $\widehat{X}_n = X_n(I_n + P_n)$. X_n and P_n are independent by assumption. Define the empirical eigenvalue distribution as $\mu_{X_n} = \frac{1}{n}\sum_{j=1}^n \delta_{\lambda_j(X_n)}$. We assume that as $n \to \infty$, $\mu_{X_n} \xrightarrow{\text{a.s.}} \mu_X$.

Let $k_{\text{eff}} = k$. For $i = 1, \ldots, k$, let $\widehat{v}_i$ be an arbitrary unit eigenvector of $\widehat{X}_n$. By the eigenvalue master equation, $\widehat{X}_n \widehat{v}_i = \widehat{\lambda}_i \widehat{v}_i$, it follows that

$$U_{n,k}^H \left(\widehat{\lambda}_i I_n - X_n\right)^{-1} X_n U_{n,k} \Sigma U_{n,k}^H \widehat{v}_i \quad = U_{n,k}^H \widehat{v}_i. \quad (28)$$

Without loss of generality, we may assume that X_n is diagonal, with entries $\lambda_1(X_n), \ldots, \lambda_n(X_n)$. We may make this assumption as we could diagonalize X_n and incorporate the eigenvectors of X_n into $U_{n,k}$, i.e. a change of basis. Using this assumption, (28) simplifies to

$$\left[T_{\mu_{r,j}^{(n)}}\left(\widehat{\lambda}_i\right)\right]_{r,j=1}^k \Sigma U_{n,k}^H \widehat{v}_i = U_{n,k}^H \widehat{v}_i \quad (29)$$

where $\mu_{r,j}^{(n)}$, $r = 1, \ldots, k$, $j = 1, \ldots, k$, is the random complex measure defined by

$$\mu_{r,j}^{(n)} = \sum_{\ell=1}^n \overline{u_{\ell,r}^{(n)}} u_{\ell,j}^{(n)} \delta_{\lambda_\ell(X_n)}$$

and $T_{\mu_{r,j}^{(n)}}$ is the T-transform defined by

$$T_\mu(z) = \int \frac{t}{z-t} d\mu(t) \quad \text{for } z \notin \text{supp } \mu.$$

We may rewrite (29) as

$$\left(I_k - \left[\sigma_j^2 T_{\mu_{r,j}^{(n)}}\left(\widehat{\lambda}_i\right)\right]_{r,j=1}^k\right) U_{n,k}^H \widehat{v}_i = 0.$$

Therefore, $U_{n,k}^H \widehat{v}_i$ must be in the kernel of $M_n\left(\widehat{\lambda}_i\right) = I_k - \left[\sigma_j^2 T_{\mu_{r,j}^{(n)}}\left(\widehat{\lambda}_i\right)\right]_{r,j=1}^k$. By Proposition 9.3 of [9]

$$\mu_{r,j}^{(n)} \xrightarrow{\text{a.s.}} \begin{cases} \mu_X & \text{for } i = j \\ \delta_0 & \text{o.w.} \end{cases}$$

where μ_X is the limiting eigenvalue distribution of X_n. Therefore,

$$M_n\left(\widehat{\lambda}_i\right) \xrightarrow{\text{a.s.}} \mathbf{diag}\left(1-\sigma_1^2 T_{\mu_X}\left(\widehat{\lambda}_i\right),\ldots,1-\sigma_k^2 T_{\mu_X}\left(\widehat{\lambda}_i\right)\right)$$

As $k_{\text{eff}} = k$, $\sigma_i^2 > 1/T_{\underbrace{\mu_X}}(b^+)$, where b is the supremum of the support of μ_X. As $\widehat{\lambda}_i$ is the eigenvalue corresponding to the eigenvector $\widehat{v}_i$, by Theorem 2.6 of [9] $\widehat{\lambda}_i \xrightarrow{\text{a.s.}} T_{\mu_X}^{-1}\left(1/\sigma_i^2\right)$. Therefore,

$$M_n\left(\widehat{\lambda}_i\right) \xrightarrow{\text{a.s.}} \mathbf{diag}\left(1-\frac{\sigma_1^2}{\sigma_i^2},\ldots,1-\frac{\sigma_{i-1}^2}{\sigma_i^2},0,1-\frac{\sigma_{i+1}^2}{\sigma_i^2},\ldots,1-\frac{\sigma_k^2}{\sigma_i^2}\right) \tag{30}$$

Recall that $U_{n,k}^H \widehat{v}_i$ must be in the kernel of $M_n\left(\widehat{\lambda}_i\right)$. Therefore, any limit point of $U_{n,k}^H \widehat{v}_i$ is in the kernel of the matrix on the right hand side of (30). Therefore, for $i \neq j$, $i = 1,\ldots,k$, $j = 1,\ldots,k$, we must have that $\left(1-\frac{\sigma_j^2}{\sigma_i^2}\right)\langle u_j, \widehat{v}_i\rangle = 0$. As $\sigma_i^2 \neq \sigma_j^2$, for this condition to be satisfied we must have that for $j \neq i$, $i = 1,\ldots,k$, $j = 1,\ldots,k$, $\langle u_j, \widehat{v}_i\rangle \xrightarrow{\text{a.s.}} 0$. When $k_{\text{eff}} = k$, this holds for all eigenvectors $\widehat{v}_i$, $1 \leq i \leq k$, of $\widehat{X}_n$ as $\widehat{v}_i$ was chosen arbitrarily.

Recall that our observed vectors $y_i \in \mathbb{C}^{n\times 1}$ have covariance matrix $U_{n,k}\Sigma U_{n,k}^H + I_n = P_n + I_n$. Therefore, our observation matrix, Y_n which is a $n \times m$ matrix, may be written $Y_n = (P_n + I_n)^{1/2} Z_n$. The sample covariance matrix, $S_n = \frac{1}{m} Y_n Y_n^H$, may be written $S_n = (I_n + P_n)^{1/2} X_n (I_n + P_n)^{1/2}$. By similarity transform, if $\widehat{v}_i$ is a unit-norm eigenvector of $\widehat{X}_n$ then $\widehat{w}_i = (I_n + P_n)^{1/2}\widehat{v}_i$ is an eigenvector of S_n. If $\widehat{u}_i = \widehat{w}_i/\|\widehat{w}_i\|$ is a unit-norm eigenvector of S_n, it follows that

$$\langle u_j, \widehat{u}_i\rangle = \frac{\sqrt{\sigma_i^2+1}\langle u_j, \widehat{v}_i\rangle}{\sqrt{\sigma_i^2|\langle u_j, \widehat{v}_i\rangle|^2 + 1}}$$

As $\langle u_j, \widehat{v}_i\rangle \xrightarrow{\text{a.s.}} 0$ for all $i \neq j$ it follows that $\langle u_j, \widehat{u}_i\rangle \xrightarrow{\text{a.s.}} 0$ for all $i \neq j$ which was desired to be shown.

An extension of this argument proves the general result when $i > k$ or when $k_{\text{eff}} < k$; however, the derivation is more intricate and we omit it here for space considerations. An important assumption needed for the proof is that for fixed i and k, $n|\lambda_{i+k} - \lambda_i|^2 \xrightarrow{\text{a.s.}} 0$, which occurs due to the 'rigidity' of the eigenvalues of X_n. Tao [21] recently proved that these conditions hold for the Wigner matrix. By the universality [22] properties of random matrices, this spacing assumption is justified for the non-zero eigenvalues of the Wishart matrix. ■

ACKNOWLEDGMENT

This work is supported by the ONR Young Investigator Program under Grant N00014-11-1-0660.

REFERENCES

[1] L. Scharf and C. Demeure, *Statistical signal processing: detection, estimation, and time series analysis*. Addison-Wesley Publishing Company, 1991, vol. 1.

[2] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer Series in Statistics, 2001, vol. 1.

[3] L. Scharf and B. Friedlander, "Matched subspace detectors," *Signal Processing, IEEE Trans. on*, vol. 42, no. 8, pp. 2146–2157, 1994.

[4] Y. Jin and B. Friedlander, "A cfar adaptive subspace detector for second-order gaussian signals," *Signal Processing, IEEE Trans. on*, vol. 53, no. 3, pp. 871–884, 2005.

[5] T. McWhorter and L. Scharf, *Matched subspace detectors for stochastic signals*. Defense Technical Information Center, 2003.

[6] F. Vincent, O. Besson, and C. Richard, "Matched subspace detection with hypothesis dependent noise power," *Signal Processing, IEEE Transactions on*, vol. 56, no. 11, pp. 5713–5718, 2008.

[7] O. Besson and L. Scharf, "Cfar matched direction detector," *Signal Processing, IEEE Transactions on*, vol. 54, no. 7, pp. 2840–2844, 2006.

[8] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statistica Sinica*, vol. 17, no. 4, p. 1617, 2007.

[9] F. Benaych-Georges and R. Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Adv. in Math.*, 2011.

[10] N. Asendorf and R. Nadakuditi, "Improving and characterizing the performance of stochastic matched subspace detectors when using noisy estimated subspaces," in *Proceedings of the Asilomar Conference of Signals and Systems*, November 2011.

[11] H. Van Trees, *Detection, estimation, and modulation theory: Detection, estimation, and linear modulation theory*. Wiley, 1968.

[12] H. Cox, "Resolving power and sensitivity to mismatch of optimum array processors," *J. Acoust. Soc. Amer*, vol. 54, no. 3, pp. 771–785, 1973.

[13] R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *Signal Processing, IEEE Trans. on*, vol. 56, no. 7, pp. 2625–2638, 2008.

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

[15] F. Benaych-Georges and R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices," *Journal of Multivariate Analysis*, 2012, to appear. Available online at `arxiv:1103.2221`.

[16] R. Muirhead, *Aspects of multivariate statistical theory*. Wiley Online Library, 1982, vol. 42.

[17] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[18] A. Wood, J. Booth, and R. Butler, "Saddlepoint approximations to the cdf of some statistics with nonnormal limit distributions," *Journal of the American Statistical Association*, pp. 680–686, 1993.

[19] I. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *The Ann. of stat.*, vol. 29, no. 2, pp. 295–327, 2001.

[20] N. El Karoui, "Tracy–widom limit for the largest eigenvalue of a large class of complex sample covariance matrices," *The Ann. of Prob.*, vol. 35, no. 2, pp. 663–714, 2007.

[21] T. Tao, "The asymptotic distribution of a single eigenvalue gap of a wigner matrix," *Arxiv preprint arXiv:1203.1605*, 2012.

[22] L. Erdős, "Universality of wigner random matrices: a survey of recent results," *Russian Mathematical Surveys*, vol. 66, p. 507, 2011.

PLACE PHOTO HERE

Nicholas Asendorf received the B.S. degree in computer engineering from the University of Maryland, College Park, MD, in 2010 and the M.S. degree in electrical engineering:systems from the University of Michigan in 2012.

He is a graduate student in the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor. His research interests include the need for data driven algorithms in statistical signal processing and machine learning particularly in low SNR and sample starved settings.

PLACE
PHOTO
HERE

Raj Rao Nadakuditi received the B.S. degree in electrical engineering from Lafayette College, Easton, PA, in 1999, and the M.S. and Ph.D. degrees in electrical engineering and oceanographic engineering from the Massachusetts Institute of Technology and Woods Hole Oceanographic Institution Joint Program in Applied Ocean Science and Engineering in 2001 and 2007, respectively.

He joined the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, in September 2009, where he is currently an Assistant Professor. His research interests are in the general area of statistical signal processing with an emphasis on the application of random matrix theory for high-dimensional problems that arise in the context of radar, sonar, wireless communications, and machine learning.