

Improving and Characterizing the Performance of Stochastic Matched Subspace Detectors When Using Noisy Estimated Subspaces

Nicholas Asendorf

Department of Electrical and Computer Engineering
University of Michigan
Ann Arbor, Michigan 48105
Email: asendorf@umich.edu

Raj Rao Nadakuditi

Department of Electrical and Computer Engineering
University of Michigan
Ann Arbor, Michigan 48105
Email: rajnrao@umich.edu

Abstract—We consider a stochastic matched subspace detection problem where the signal subspace is unknown and estimated by taking the eigenvalue decomposition of the sample covariance matrix of noisy signal-bearing training data. In moderate to low signal-to-noise ratio (SNR) regimes or in the setting where the number of samples is limited, subspace estimation errors affect the performance of matched subspace detectors. We use random matrix theory to derive an optimal matched subspace detector which accounts for these estimation errors and to analytically predict the associated ROC performance curves. What emerges from the analysis is the importance of using only the $k_{\text{eff}} \leq k$ informative signal subspace components that can be reliably estimated from the noisy, limited data. Specifically, the ROC analysis shows that the performance of the optimal detector matches that of the plug-in detector that uses exactly k_{eff} components. The analytical predictions are validated using numerical simulations.

I. INTRODUCTION

Many signal processing and machine learning problems involve designing a detector to distinguish between signal and noise by using signal-bearing and noise-only training data sets. Subspace-based methods constitute a powerful and widely used class of algorithms for this purpose [1]–[5]. Matched subspace detectors solve this problem when the signal-bearing observation can be modeled as residing in a low-dimensional subspace buried in noise. The family of matched subspace detectors developed in [3]–[5] tackle the problem in the setting where the low-dimensional signal subspace is known. Much is known about the performance and optimality properties of these methods. For example, Scharf and Friedlander [3] prove that the generalized likelihood ratio test (GLRT) is optimal for solving matched subspace detection problems while [5] extend this work to the stochastic setting and conclude that the optimal detector in the known noise case is a matched subspace detector.

Much less is known, however, about these detectors in the setting where the signal subspace is unknown and estimated from noisy, limited signal-bearing data. In such settings, when signal-bearing data is available, an estimate of this low-dimensional signal subspace can be formed from an eigen-decomposition of the sample covariance matrix. Standard

plug-in detectors form a GLRT test statistic by simply substituting this subspace estimate for the true signal subspace in the expressions found in [4], [5] that were derived assuming that the signal subspace was known. However, in the noisy, finite-sample setting, the subspace estimates formed are also noisy. Recent results from random matrix theory (RMT) precisely quantify these errors in subspace estimation. The primary contribution of this paper is the derivation of a matched subspace detector which explicitly accounts for these predictable subspace estimation errors and the characterization of the receiver operating characteristic (ROC) performance of this detector relative to the standard plug-in detector.

What emerges from the ROC analysis is the importance of using only the $k_{\text{eff}} \leq k$ informative signal subspace components that can be reliably estimated from the noisy, limited data as opposed to using the (unknown) intrinsic dimension k of the signal subspace. Specifically, the ROC analysis shows that the performance of the optimal detector matches that of the plug-in detector that uses exactly k_{eff} components. When more than k_{eff} components are used, the performance of the plug-in detector degrades relative to that of the optimal detector.

The paper is organized as follows. Section II formally states the detection problem. Section III highlights pertinent results from random matrix theory that are utilized in Section IV in the derivation of the oracle, plug-in, and optimal matched subspace detectors and in Section V to derive their associated performance metrics. Section V describes a saddlepoint approximation for computing the theoretical ROC curves for the plug-in and optimal detectors. Section VI validates our analytical predictions and highlights the importance of picking k_{eff} over k . Concluding remarks are provided in Section VII.

II. PROBLEM STATEMENT

We consider the detection problem where our observation vector $y \in \mathbb{C}^{n \times 1}$ is modeled as follows:

$$y = \begin{cases} z & y \in H_0 \\ Ux + z & y \in H_1 \end{cases} \quad (1)$$

where $z \sim \mathcal{N}(0, I)$, U is an unknown $n \times k$ (real or complex) matrix with orthonormal columns, and $x \sim \mathcal{N}(0, \Sigma)$ where $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_k^2)$ with σ_k^2 unknown. The dimension of our subspace, k , is unknown and $k \ll n$. We also assume that x and z are independent.

To estimate our unknown parameters, U and Σ , we are given independent signal-bearing training data $\{y_1, \dots, y_m\}$, with $y_i \in H_1$ for $i = 1, \dots, m$ and an estimate, \hat{k} of our unknown dimension, k . After forming our subspace estimate, \hat{U} , we generate our $\hat{k} \times 1$ test vector $w = \hat{U}^H y$. Our goal is to determine a detector, $g(w) \rightarrow \{H_0, H_1\}$ which solves the following problem for an unlabeled testing point, w :

$$\begin{aligned} & \text{maximize} && P_D = P(g(w) \rightarrow H_1 | w \in H_1) \\ & \text{subject to} && P_F = P(g(w) \rightarrow H_1 | w \in H_0) \leq \alpha \end{aligned} \quad (2)$$

where $\alpha \in [0, 1]$.

III. PERTINENT RESULTS FROM RMT

The first step in any detector derivation is to form estimates \hat{U} and $\hat{\Sigma}$ to use in a GLRT. We are given signal-bearing training data $\{y_1, \dots, y_m\}$ where $y_i \in H_1$, for $i = 1, \dots, m$, which are stacked as columns in a matrix $Y = [y_1, \dots, y_m]$. To form estimates of U and Σ , we may take the eigenvalue decomposition of the sample covariance matrix, $S = \frac{1}{m} Y Y^H$. Recent results from random matrix theory allow us to classify the accuracy of eigenvectors and eigenvalues of S . From [6]–[8] we have the following result.

Theorem III.1. *As $n, m \rightarrow \infty$ with $n/m \rightarrow c$ we have that:*

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

$$\langle u_i, \hat{u}_j \rangle^2 \xrightarrow{a.s.} 0 \quad \text{for } i \neq j$$

where \hat{u}_i is the i^{th} eigenvector of S .

Note that the proof of the case where $i \neq j$ will appear in a later journal publication [7]. The key point of Theorem III.1 is that only the eigenvectors corresponding to eigenvalues above the phase transition \sqrt{c} are *informative*. When a signal eigenvalue drops below that critical threshold, the corresponding eigenvector estimate is essentially noise-like (i.e. $|\langle u_i, \hat{u}_i \rangle|^2 = o_p(1)$) and thus *uninformative*. The term $|\langle u_i, \hat{u}_i \rangle|^2$ quantifies mismatch between the estimated and underlying eigenvectors and plays an important role in the performance analysis of the resulting detectors; a similar term also comes up in the analysis of the resolving power of arrays due to model mismatch such as in [9]. Following [10], we define the effective number of identifiable subspace components k_{eff} as:

$$k_{\text{eff}} = \text{Number of } \sigma_i^2 \geq \sqrt{c} \quad (3)$$

Intuitively we expect a degradation in the performance of detectors that utilize subspace components for which $|\langle u_i, \hat{u}_i \rangle|^2 = o_p(1)$. We now turn to the signal eigenvalue estimation problem. Since we are only interested in the signal

eigenvalues above the critical value, we apply theorem 3 of [6] to our problem.

Theorem III.2. *As $n, m \rightarrow \infty$ with $n/m \rightarrow c$, when $\sigma_i^2 > \sqrt{c}$, we have that:*

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

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

We form an estimate of the signal eigenvalue, σ_i^2 , by employing maximum-likelihood (ML) estimation on $\hat{\sigma}_i$. Specifically, we form the estimate:

$$\hat{\sigma}_{i_{\text{mt}}}^2 = \underset{\sigma_i^2}{\text{argmax}} \log \left(f_{\hat{\sigma}_i^2}(\sigma_i) \right) \quad (4)$$

for only the k_{eff} signal eigenvalues for which $\sigma_i^2 > \sqrt{c}$. We estimate k_{eff} by (say) using “Algorithm 2” of [11]. Knowing $\hat{\sigma}_{i_{\text{mt}}}^2$ from data, we can now estimate $|\langle u_i, \hat{u}_i \rangle|^2$ by an application of Theorem III.1. We refer to this estimate as $|\langle u_i, \hat{u}_i \rangle|_{\text{mt}}^2$.

IV. FAMILY OF MATCHED SUBSPACE DETECTORS

The Neyman-Pearson Lemma (see [12]) states that the solution to (2) is a likelihood ratio test (LRT)

$$\Lambda(w) \underset{H_0}{\overset{H_1}{\geq}} \eta$$

where $\Lambda(w) = \frac{f(w|H_1)}{f(w|H_0)}$ and η satisfies $P(\Lambda(w) \leq \eta | H_0) = \alpha$. The LRT relies on the conditional distributions of our test statistic under each hypothesis, which by properties of Gaussian random variables are simply

$$\begin{aligned} w|H_0 &\sim \mathcal{N}(0, I_{\hat{k}}) \\ w|H_1 &\sim \mathcal{N}(0, \hat{U}^H U \Sigma U^H \hat{U} + I_{\hat{k}}) \end{aligned}$$

A. Detectors Considered

We will consider three different detectors for our test data w . The first is an oracle detector, which will assume that U and Σ are known. The purpose of this is to give an upper bound on a detector’s performance. The second is a plug-in detector which uses the oracle statistic and plugs in our estimates of \hat{U} and $\hat{\Sigma}$ for our unknown U and Σ with the additional assumption that $\hat{U} = U$ and $\hat{\Sigma} = \Sigma$. The third is the optimal, new detector which explicitly exploits the quantification of the subspace accuracy estimates in Theorem III.1 and only $\hat{k} = k_{\text{eff}}$ subspace components to form an approximation to the oracle classifier. Table I summarizes the test statistic derived for each detector and the distribution under each hypothesis.

B. Oracle Detector

The oracle detector assumes that Σ and U are both known. The LRT statistic for our processed data w is

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

Detector	Detector Statistic $\Lambda(w)$	Null Hypothesis Distribution ΛH_0	Simple Hypothesis Distribution ΛH_1
Oracle	$w^H \left[I - \left(\hat{U}^H U \Sigma U^H \hat{U} + I \right)^{-1} \right] w$		
Plug-in	$\sum_{i=1}^{\hat{k}} w_i^2 \frac{\hat{\sigma}_i^2}{1 + \hat{\sigma}_i^2}$	$\sum_{i=1}^{\hat{k}} \left(\frac{\hat{\sigma}_i^2}{1 + \hat{\sigma}_i^2} \right) \chi_{1i}^2$	$\sum_{i=1}^{\hat{k}} \left(\frac{\hat{\sigma}_i^2 (1 + \sigma_i^2 \langle u_i, \hat{u}_i \rangle ^2)}{1 + \hat{\sigma}_i^2} \right) \chi_{1i}^2$
Optimal	$\sum_{i=1}^{\min\{k_{\text{eff}}, \hat{k}\}} w_i^2 \frac{ \langle u_i, \hat{u}_i \rangle _{\text{rmt}}^2 \hat{\sigma}_{i\text{rmt}}^2}{1 + \langle u_i, \hat{u}_i \rangle _{\text{rmt}}^2 \hat{\sigma}_{i\text{rmt}}^2}$	$\sum_{i=1}^{\min\{k_{\text{eff}}, \hat{k}\}} \left(\frac{\hat{\sigma}_{i\text{rmt}}^2 \langle u_i, \hat{u}_i \rangle _{\text{rmt}}^2}{1 + \hat{\sigma}_{i\text{rmt}}^2 \langle u_i, \hat{u}_i \rangle _{\text{rmt}}^2} \right) \chi_{1i}^2$	$\sum_{i=1}^{\min\{k_{\text{eff}}, \hat{k}\}} \left(\hat{\sigma}_{i\text{rmt}}^2 \langle u_i, \hat{u}_i \rangle _{\text{rmt}}^2 \right) \chi_{1i}^2$

TABLE I

GIVEN AN OBSERVATION VECTOR y WE FORM THE VECTOR $w = \hat{U}^H y$ WHERE \hat{U} IS AN ESTIMATE OF THE SIGNAL SUBSPACE. THE TABLE SUMMARIZES THE TEST STATISTIC ASSOCIATED WITH EACH DETECTOR AND THE ASSOCIATED DISTRIBUTION UNDER H_0 AND H_1 . IN THE CFAR SETTING, THE THRESHOLD IS SET TO OBTAIN THE DESIRED FALSE ALARM PROBABILITY. NOTE THE APPEARANCE OF k_{eff} IN THE OPTIMAL DETECTOR.

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

$$\Lambda_{\text{oracle}}(w) = w^H \left[I - \left(\hat{U}^H U \Sigma U^H \hat{U} + I \right)^{-1} \right] w \quad (5)$$

and the oracle detector is

$$\Lambda_{\text{oracle}}(w) \underset{H_0}{\overset{H_1}{\geq}} \ln(\eta_{\text{oracle}}) \quad (6)$$

where η_{oracle} satisfies $P(\Lambda_{\text{oracle}}(w) \leq \ln(\eta_{\text{oracle}}) | H_0) = \alpha$.

C. Plug-in Detector

As is the case in this particular problem, U and Σ are unknown, and therefore (5) cannot be computed directly. One may then substitute ML estimates for the unknown parameters in the oracle detector to form a GLRT as in [4] and [5]. The ML estimates (in the large-sample, small matrix setting) for U and Σ are:

$$\begin{aligned} \hat{U} &= [\hat{u}_1 \dots \hat{u}_{\hat{k}}] \\ \hat{\sigma}_i^2 &= \lambda_i - 1 \text{ for } i = 1, \dots, \hat{k} \end{aligned} \quad (7)$$

where $\lambda_1, \dots, \lambda_{\hat{k}}$ are the largest \hat{k} eigenvalues of S and $\hat{u}_1, \dots, \hat{u}_{\hat{k}}$ are the corresponding eigenvectors. Define the signal covariance matrix estimate as $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_{\hat{k}}^2)$.

By replacing the unknown parameters in (5) with the estimated parameters in (7), the plug-in detector statistic becomes:

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

This simplifies to

$$\Lambda_{\text{plugin}}(w) = w^H \text{diag} \left(\frac{\hat{\sigma}_i^2}{1 + \hat{\sigma}_i^2} \right) w = \sum_{i=1}^{\hat{k}} w_i^2 \frac{\hat{\sigma}_i^2}{\hat{\sigma}_i^2 + 1} \quad (8)$$

and our detector becomes

$$\Lambda_{\text{plugin}}(w) \underset{H_0}{\overset{H_1}{\geq}} \ln(\eta_{\text{plugin}}) \quad (9)$$

where η_{plugin} satisfies $P(\Lambda_{\text{plugin}}(w) \leq \ln(\eta_{\text{plugin}}) | H_0) = \alpha$.

The plug-in detector assumes that the estimated signal subspace \hat{U} is equal to the true signal subspace U and that

the estimated signal covariance $\hat{\Sigma}$ is equal to the true signal covariance Σ . It also assumes that the provided subspace dimension estimate \hat{k} is equal to the true underlying dimension of our signal subspace k . However, as shown in Section III, choosing $\hat{k} > k_{\text{eff}}$ leads to a degradation in the performance of the plug-in detector.

D. Optimal Detector

Consider the covariance matrix of the conditional distribution $w|H_1$. By Theorem III.1 and additional analysis omitted here for brevity (see [7]), we have that in the large matrix limit:

$$\hat{U}^H U \Sigma U^H \hat{U} + I = \text{diag}(|\langle u_i, \hat{u}_i \rangle|^2 \sigma_i^2 + 1 + o_p(1)). \quad (10)$$

The plug-in detector assumes that $|\langle u_i, \hat{u}_i \rangle|^2 = 1$ and $\hat{\sigma}_i^2 = \sigma_i^2$. Theorems III.1 and III.2 show that this assumption is invalid because these eigenvalue and eigenvector estimates are biased. Although, σ_i^2 and $|\langle u_i, \hat{u}_i \rangle|^2$ are unknown, we already derived an improved signal variance estimate in (4). We plug this estimate into Theorem III.1 to obtain an estimate for $|\langle u_i, \hat{u}_i \rangle|^2$, which we denote $|\langle u_i, \hat{u}_i \rangle|_{\text{rmt}}^2$. We then derive an optimal detector by substituting these random matrix theory estimates into the diagonal covariance matrix (10) and using this covariance matrix in the GLRT. This choice is well-motivated in the high-dimensional setting where we can ignore the $o_p(1)$ terms on the right hand side of (10). After simplification of the GLRT, the optimal statistic is

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

However, as $|\langle u_i, \hat{u}_i \rangle|^2 = 0$ when $\hat{\sigma}_{i\text{rmt}}^2$ drops below \sqrt{c} , this sum ignores all indices whose signal eigenvalue estimate is below this critical value. Instead, the optimal detector only sums those components with significant signal eigenvalues. The number of significant signal eigenvalues was defined as k_{eff} . Our final optimal statistic thus becomes:

$$\Lambda_{\text{optimal}}(w) = \sum_{i=1}^{\min\{k_{\text{eff}}, \hat{k}\}} w_i^2 \frac{|\langle u_i, \hat{u}_i \rangle|_{\text{rmt}}^2 \hat{\sigma}_{i\text{rmt}}^2}{|\langle u_i, \hat{u}_i \rangle|_{\text{rmt}}^2 \hat{\sigma}_{i\text{rmt}}^2 + 1} \quad (11)$$

and the optimal detector becomes

$$\Lambda_{\text{optimal}}(w) \underset{H_0}{\overset{H_1}{\geq}} \ln(\eta_{\text{optimal}}) \quad (12)$$

where η_{optimal} satisfies $P(\Lambda_{\text{optimal}}(w) \leq \eta_{\text{optimal}} | H_0) = \alpha$.

V. THEORETICAL ROC CURVE DERIVATION

A standard way to compare the plug-in and optimal detectors derived in (9) and (12) respectively is to compute their ROC curves [13]. For a particular statistic $\Lambda(w)$, to compute theoretical ROC curves, we must compute

$$\begin{aligned} P_D &= P(\Lambda(w) \geq y | w \in H_1) \\ P_F &= P(\Lambda(w) \geq y | w \in H_0) \end{aligned} \quad (13)$$

for $-\infty < y < \infty$. To do this, we explore the conditional CDF under each hypothesis for the statistics (8) and (11).

The conditional distributions of our test samples under each hypothesis are $w|H_0 \sim \mathcal{N}(0, I)$ and $w|H_1 \sim \mathcal{N}(0, \hat{U}^H U \Sigma U^H \hat{U} + I)$. Because of the diagonal covariance matrix under H_0 , $w_i|H_0 \sim \mathcal{N}(0, 1)$ are i.i.d for $i = 1, \dots, \hat{k}$ and $w_i^2|H_0 \sim \chi_1^2$ are i.i.d for $i = 1, \dots, \hat{k}$. As argued in (10), the covariance matrix under H_1 is asymptotically diagonal. We have that $w_i|H_1 \approx \mathcal{N}(0, \sigma_i^2 |\langle u_i, \hat{u}_i \rangle|^2 + 1)$ are i.i.d for $i = 1, \dots, \hat{k}$. Therefore, for $i = 1, \dots, \hat{k}$

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

The third and fourth columns of Table I summarize the sample conditional distributions for w into each detector's statistic. An analytical expression for the asymptotic performance in the large matrix limit is obtained by substituting expressions from Theorems III.1 & III.2 for the pertinent quantities in these distributions. Note that all distributions are a weighted sum of independent chi-square random variables with one degree of freedom. That is the distributions take the form

$$\Lambda = \sum_{i=1}^r \beta_i \chi_{1i}^2$$

where β_i is the appropriate weighting, unique to each statistic under each hypothesis. The CDF of a chi-square random variable is known in closed form. However, the CDF of a weighted sum of independent chi-square random variables is not known in closed form. To evaluate (13), we use a saddlepoint approximation of the CDF of Λ by employing the generalized Lugannani-Rice formula proposed in [14]. To then compute a theoretical ROC curve, y is swept over $(0, \infty)$ and for each value of y , the saddlepoint approximation of the CDF under each hypothesis is computed using the method alluded to. This generates a set of points (P_F, P_D) which approximate the theoretical ROC curve. Figure 1 illustrates the result of one such computation.

VI. SIMULATION RESULTS AND DISCUSSION

A. ROC curves

As seen in Figure 1, the theoretical ROC curves match the empirical ROC curves thereby validating the accuracy of the saddlepoint approximation to the CDF. The performance of the optimal detector matches that of the oracle detector attesting to the accuracy of the random matrix theoretic approximations employed. Finally, we note the sub-optimality of the plug-in

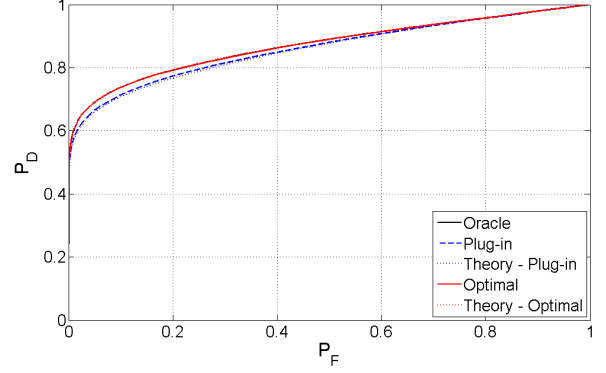


Fig. 1. Empirical and theoretical ROC curves for the oracle, plug-in, and optimal matched subspace detectors. Empirical ROC curves were simulated with $n = 100$, $m = 50$, $k = 2$, $\hat{k} = 2$, $\sigma_1 = 5$, $\sigma_2 = 0.5$. However, as σ_2 is below the critical threshold, $k_{\text{eff}} = 1$. The empirical ROC curves were computed using 2000 test samples and averaged over 25 trials using algorithms 2 and 4 of [13]. The theoretical ROC curves were computed as described in Section V. Since we chose $\hat{k} > k_{\text{eff}}$, we observe a performance gain when using the optimal detector. In fact, the optimal detector operates at almost the same (P_F, P_D) as the oracle detector. The theoretically predicted ROC curves match the empirical results reflecting the accuracy of the approximations employed.

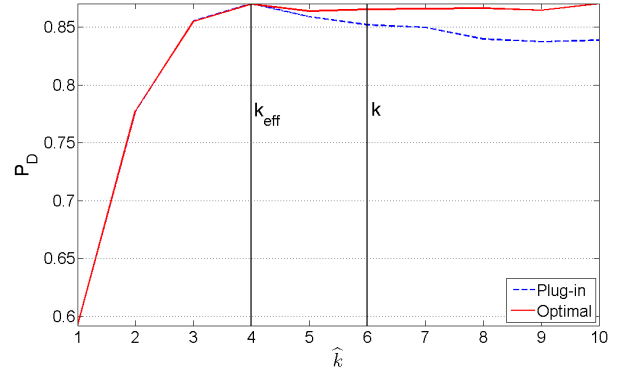


Fig. 2. Empirical exploration of the achieved probability of detection, P_D , for a fixed probability of false alarm, $P_F = 0.01$, for various \hat{k} . For the simulation, $n = 100$, $m = 50$, $k = 6$, $\Sigma^{1/2} = \text{diag}(5, 4, 3, 2, 0.6, 0.5)$ so that $k_{\text{eff}} = 4$. P_D calculation was computed using 2000 test samples and averaged over 25 trials. The optimal \hat{k} resulting in the largest P_D is not the true k , but rather k_{eff} .

detector. This is expected as we chose $\hat{k} > k_{\text{eff}}$ so that the additional subspace component chosen was uninformative.

B. Optimality of k_{eff}

Figure 2 examines the performance as a function of \hat{k} , confirming our assertion that the optimal \hat{k} equals k_{eff} . Setting $\hat{k} < k_{\text{eff}}$ drastically degrades performance for both detectors. Notably, the plug-in and optimal 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 $\hat{k} > k_{\text{eff}}$, the performance of the plug-in detector degrades while that of the optimal detector is relatively stable as if $\hat{k} = k_{\text{eff}}$. In other words, we do not

pay a price for overestimating our dimension with the optimal detector.

This makes sense (and is slightly contrived) because the optimal detector will only sum to a maximum of k_{eff} indices as evident in (11). In many applications, practitioners might employ the “play-it-safe” approach and set \hat{k} to be slightly greater than k_{eff} . The degradation in performance by adding each uninformative subspace, as seen in Figure 2, constitutes evidence to our assertion that overestimating the signal subspace dimension is a bad idea. The plug-in detector achieves optimal performance whenever $\hat{k} = k_{\text{eff}}$. When $k_{\text{eff}} < k$, setting $\hat{k} = k$ is sub-optimal.

C. Averaging over Test Samples

Figure 3 shows the performance in the setting where the test statistic is averaged over multiple independent samples. Here $k_{\text{eff}} = 2$ and the plug-in estimator sets $\hat{k} = 6$. The optimal detector requires less samples than the plug-in detector to achieve a desired (P_F, P_D) pair. This arises from the fact that the optimal detector discards the uninformative signal subspace components while the plug-in detector utilizes them. Because each test sample is pruned to ignore noisy components, we can average less samples to achieve a desired accuracy.

VII. CONCLUSION

We have considered a stochastic matched subspace detection problem where the signal subspace is unknown and estimated from noisy, limited signal-bearing training data. We used random matrix theory to analytically characterize the ROC performance of a family of matched subspace detectors including a new, optimal detector that exploits information about the “noisiness” of the signal subspace estimates.

What emerges from the analysis is the fact that quantification of whether an estimated subspace is informative or not, is more important than the relative degree of noisiness. The effective number of identifiable subspace components k_{eff} , first introduced in [10], and expressed in (3), affects detector performance the most. When $\hat{k} = k_{\text{eff}}$, the standard plug-in detector is near-optimal; when $\hat{k} \neq k_{\text{eff}}$, it is not. This highlights the importance of robust techniques [11], [15], [16] for estimating k_{eff} in subspace based detection schemes. We are working on extending the analysis to deterministic matched subspace detectors.

ACKNOWLEDGMENT

Work supported by the ONR Young Investigator Award under Grant N00014-11-1-0660.

REFERENCES

- [1] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Verlag, 2001.
- [2] J. Laaksonen and E. Oja, “Subspace dimension selection and averaged learning subspace method in handwritten digit classification,” *Artificial Neural Networks ICANN 96*, pp. 227–232, 1996.
- [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.

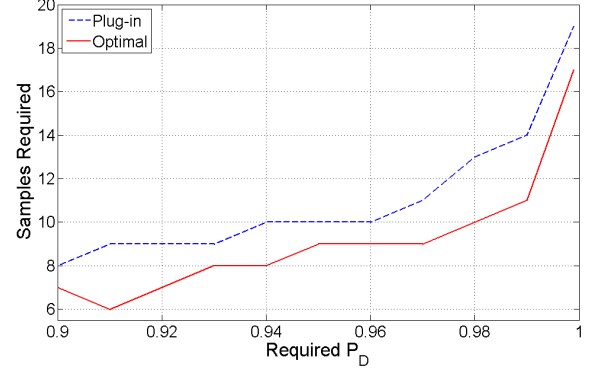


Fig. 3. Empirical exploration of the number of test samples required to achieve a desired $P_F = 0.001$ for different P_D . In this setup, the test statistic for each detector is averaged over multiple independent test samples. For the simulation, $n = 100$, $m = 50$, $k = 6$, $\hat{k} = 6$ and $\Sigma^{1/2} = \text{diag}(3, 2, 0.5, 0.4, 0.3, 0.2)$ so that $k_{\text{eff}} = 2$. P_D calculation was computed using 2000 test samples and averaged over 10 trials. Because the optimal detector only uses k_{eff} components, it requires less test samples to achieve a desired (P_F, P_D) than the plug-in detector.

- [5] T. McWhorter and L. Scharf, *Matched subspace detectors for stochastic signals*. Defense Technical Information Center, 2003.
- [6] D. Paul, “Asymptotics of sample eigenstructure for a large dimensional spiked covariance model,” *Statistica Sinica*, vol. 17, no. 4, p. 1617, 2007.
- [7] N. Asendorf and R. Nadakuditi, “Characterizing the performance of matched subspace detectors,” *In Preparation*.
- [8] F. Benaych-Georges and R. Nadakuditi, “The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices,” *Adv. in Math.*, 2011.
- [9] H. Cox, “Resolving power and sensitivity to mismatch of optimum array processors,” *J. Acoust. Soc. Amer.*, vol. 54, no. 3, pp. 771–785, 1973.
- [10] 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.
- [11] 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.
- [12] H. Van Trees, *Detection, estimation, and modulation theory: Detection, estimation, and linear modulation theory*. Wiley, 1968.
- [13] T. Fawcett, “An introduction to roc analysis,” *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [14] 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.
- [15] 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.
- [16] 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.