

THE PERFORMANCE OF DETERMINISTIC MATCHED SUBSPACE DETECTORS WHEN USING SUBSPACES ESTIMATED FROM NOISY, MISSING DATA

Nicholas Asendorf, Raj Tejas Suryaprakash, Raj Rao Nadakuditi

Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, Michigan 48104

ABSTRACT

We consider a matched subspace detection problem where a signal vector residing in an unknown low-rank k subspace is to be detected using a subspace estimate obtained from noisy signal-bearing training data with missing entries. The resulting subspace estimate is inaccurate due to limited training data, missing entries, and additive noise. Recent results from random matrix theory (RMT) precisely quantify these subspace estimation errors for the setting where the signal has low coherence. We analytically quantify the ROC performance of the resulting plug-in detector and derive a new detector which explicitly accounts for these subspace estimation errors. The realized increase in performance can be attributed to the new detector only using the $k_{\text{eff}} \leq k$ “informative” signal subspace components. The fraction of observed entries determines k_{eff} via a simple relationship that we describe. Detection performance better than random guessing is only achievable when the percent of observed data is above a critical threshold which we explicitly characterize.

Index Terms— Matched subspace detector, random matrix theory, missing data, ROC analysis

1. INTRODUCTION

The matched subspace detector (MSD) is a widely used tool in signal processing and machine learning to detect a signal embedded in a low-rank subspace in the presence of additive noise [1, 2, 3]. The performance of the MSD has been explored when the low-rank signal subspace is known. Scharf and Friedlander [1] consider the MSD when the signal is placed deterministically at an unknown location in a known subspace while McWhorter and Scharf [3] extend this work to allow the signal to be placed randomly with a known (or assumed) distribution in the known subspace.

There is little work characterizing the performance of the matched subspace detector when the signal subspace is unknown and estimated from data. Recently, we used RMT to quantify the performance of stochastic MSDs in such a setting [4]. That work brought into sharp focus the importance

of using $k_{\text{eff}} \leq k$ informative signal subspace components in detector statistics; here k_{eff} is the effective number of informative components identifiable from limited, noisy data as described in [5]. In this paper, we extend the analysis to the deterministic MSD setting where the training data is noisy *and* has missing entries. The missing entry context is motivated in [6] by distributed detection scenarios where it might be prohibitive to collect and transmit only a (randomly chosen) fraction p of the training data entries. Alternately one might think of $1 - p \in (0, 1)$ as a compression factor as in compressed sensing.

The main contribution of this paper is a precise quantification of the resulting performance of the MSD. We uncover a phase transition phenomenon by showing that there is a critical fraction, p_{crit} , which is a simple function of the eigen-SNR, the number of training samples, and the number of sensors, below which detection performance deteriorates to random guessing. Compressing the training dataset below this critical fraction is undesirable.

The paper is organized as follows. Section 2 formally states the detection problem. Section 3 presents pertinent results from RMT. These results are used in Section 4 to derive a plug-in and random matrix theory MSD and in Section 5 to derive theoretical ROC curves for each detector. Section 6 validates our analytical predictions, highlights the importance of selecting k_{eff} over k subspace components, and demonstrates the effect of missing data. Section 7 presents concluding remarks.

2. PROBLEM STATEMENT

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

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

where $z \sim \mathcal{CN}(0, I_n)$, U is an unknown $n \times k$ (real or complex) matrix with orthonormal columns, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_k)$ is assumed known with $\sigma_1 > \sigma_2 > \dots > \sigma_k > 0$, and $x \in \mathbb{C}^{k \times 1}$ is an unknown deterministic vector. In the oracle setting, the subspace dimension, k , is known with $k \ll n$.

To estimate our unknown subspace, U , we are given independent signal-bearing training data $\{y_1, \dots, y_m\}$, such that

$$y_i = Ux_i + z_i \quad (2)$$

where $x_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{CN}(0, \Sigma^2)$, $z_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{CN}(0, I_n)$, and Σ and U are as defined above; z_i and x_i are independent. We observe only a fraction $p \in (0, 1)$ of the entries of the $n \times m$ matrix $Y = [y_1 \dots y_m]$; p is independent of n and m and we choose which entries are observed at random with i.i.d. probability p ; the unobserved/missing entries are set to zero.

The resulting matrix is denoted \tilde{Y} . We form our estimate of the signal subspace, \hat{U} , by taking the first k left singular vectors of the SVD of \tilde{Y} . After forming our subspace estimate, \hat{U} , and given a new (independent) observation y from (1), we generate a $k \times 1$ test vector $w = \hat{U}^H y$. Our objective is to formulate a detector, $g(w) \rightarrow \{H_0, H_1\}$, which maximizes the probability of detection subject to a specified false alarm rate α .

3. PERTINENT RESULTS FROM RMT

By modifying an argument in [7], we obtain the following result.

Theorem 3.1. *Assume that $x_i \sim \mathcal{CN}(0, \Sigma^2)$ as in (2) and that U in (2) obeys a low coherence condition. Then as $n, m \rightarrow \infty$ with $n/m \rightarrow c$ we have that for $i, j = 1, \dots, k$:*

$$|\langle u_i, \hat{u}_i \rangle|^2 \xrightarrow{a.s.} \begin{cases} 1 - \frac{c(1 + p\sigma_i^2)}{p\sigma_i^2(p\sigma_i^2 + c)} & \text{if } \sigma_i > \frac{c^{1/4}}{\sqrt{p}} \\ 0 & \text{otherwise} \end{cases}.$$

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

where \hat{u}_i are the left singular vectors of \tilde{Y} .

The low coherence condition appears in, for example, [6] with the idea being that the matrix $U [x_1 \dots x_m]$ has entries of about the same magnitude. With the Gaussianity assumption for x , all we need is U to have low coherence. Recall that the coherence of a matrix with orthonormal columns is $\max_{i,j} |U_{i,j}|$. When a matrix is spiky, random sampling of its entries may result in a loss of information; matrices with low coherence behave better under random sampling and it is this setting that we focus on in this paper.

The key insight from Theorem 3.1 is that only the singular vectors corresponding to signal singular values above the phase transition $\frac{c^{1/4}}{\sqrt{p}}$ are *informative*. The fraction of missing entries p regulates this phase transition point as $O(1/\sqrt{p})$. When a signal singular value drops below this critical threshold, the corresponding singular vector 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 singular vectors; when

$p < p_{\text{crit.}} := \sqrt{c}/\max_i(\sigma_i^2)$ then *all* singular vectors are uninformative. 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 refer to the estimate in Theorem 3.1 as $|\langle u_i, \hat{u}_i \rangle|_{\text{rmt}}^2$.

4. FAMILY OF MATCHED SUBSPACE DETECTORS

The Neyman-Pearson solution to the detection problem posed is a likelihood ratio test (LRT)

$$\Lambda(w) \stackrel{H_1}{\underset{H_0}{\gtrless}} \eta$$

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

$$w|H_0 \sim \mathcal{N}(0, I_k) \quad \text{and} \quad w|H_1 \sim \mathcal{N}(\hat{U}^H U \Sigma x, I_k).$$

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

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

Employing maximum likelihood estimation on x yields the estimate $\hat{x} = (\Sigma U^H \hat{U} \hat{U}^H U \Sigma)^{-1} \Sigma U^H \hat{U} w$. After simplification of the GLRT using \hat{x} and using the natural logarithm operator as a monotonic operation, the GLRT statistic becomes

$$\Lambda(w) = w^H \left(\hat{U}^H U \Sigma \left(\Sigma U^H \hat{U} \hat{U}^H U \Sigma \right)^{-1} \Sigma U^H \hat{U} \right) w. \quad (3)$$

4.1. Plug-in Detector

In our problem statement, U is unknown, and therefore (3) cannot be computed directly. One may then substitute a ML estimate for U in (3) as in [2] and [3]. The ML estimate (in the large-sample, small matrix setting) for U is the top k left singular vectors $\hat{U} = [\hat{u}_1 \dots \hat{u}_k]$ of the training data matrix, \tilde{Y} .

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

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

This simplifies to

$$\Lambda_{\text{plugin}}(w) = w^H w = \sum_{i=1}^k w_i^2 \quad (4)$$

and our detector becomes

$$\Lambda_{\text{plugin}}(w) \underset{H_0}{\overset{H_1}{\geq}} 2 \ln(\eta_{\text{plugin}}) \quad (5)$$

where η_{plugin} satisfies $P(\Lambda_{\text{plugin}}(w) > 2 \ln(\eta_{\text{plugin}}) | H_0) = \alpha$.

The plug-in detector assumes that the estimated signal subspace \hat{U} is equal to the true signal subspace U . However, as shown in Section 3, using uninformative components of w is not well motivated.

4.2. Random Matrix Theory Detector

Consider the term $\hat{U}^H U$. By Theorem 3.1 and by noting that the singular vectors are unique up to a phase, we have that in the large matrix limit $\hat{U}^H U \approx SA$ where $S = \text{diag}(s_1, \dots, s_k)$ with $s_i = \exp(j\psi_i)$ where for some ψ_i , s_i denotes the random phase ambiguity in the computation of the SVD and $A = \text{diag}(a_1, \dots, a_k)$ with $a_i = |\langle u_i, \hat{u}_i \rangle|$.

The plug-in detector assumes that $A = S = I_k$, that is $s_i = 1$ and $|\langle u_i, \hat{u}_i \rangle| = 1$. However, as seen in Section 3, this assumption is invalid. Let us now re-derive the GLRT using Theorem 3.1. Our conditional distributions are

$$w|H_0 \sim \mathcal{N}(0, I_k) \text{ and } w|H_1 \sim \mathcal{N}(SA\Sigma x, I_k)$$

and the GLRT is

$$\Lambda(w) = \frac{\max_x \mathcal{N}(SA\Sigma x, I_k)}{\mathcal{N}(0, I_k)}.$$

Since the covariance matrices are diagonal and S , A , and Σ are diagonal, we have that

$$\max_x \mathcal{N}(SA\Sigma x, I_k) = \prod_{i=1}^k \max_{x_i} \mathcal{N}(\sigma_i s_i a_i x_i, 1).$$

We use Theorem 3.1 to estimate $a_i = \sqrt{|\langle u_i, \hat{u}_i \rangle|_{\text{rmt}}^2}$. Now, $a_i = 0$ when $\sigma_i \leq \frac{c^{1/4}}{\sqrt{p}}$. We define k_{eff} as the number of signal singular values above the phase transition $\frac{c^{1/4}}{\sqrt{p}}$. Our GLRT becomes

$$\Lambda(w) = \frac{\mathcal{N}(0, 1)^{k-k_{\text{eff}}} \prod_{i=1}^{k_{\text{eff}}} \max_{x_i} \mathcal{N}(\sigma_i s_i a_i x_i, 1)}{\mathcal{N}(0, I_k)}.$$

Employing maximum likelihood estimation on the remaining x_i yields the estimates $\hat{x}_i = \frac{w_i}{s_i a_i}$. Notice that for these values of x_i , dividing by a_i is justified as the corresponding signal singular value is above the phase transition and so $a_i > 0$. After simplification of the GLRT using this \hat{x}_i and the natural logarithm operator, the statistic becomes

$$\Lambda_{\text{rmt}}(w) = \sum_{i=1}^{k_{\text{eff}}} w_i^2 \quad (6)$$

and our detector becomes

$$\Lambda_{\text{rmt}}(w) \underset{H_0}{\overset{H_1}{\geq}} 2 \ln(\eta_{\text{rmt}}) \quad (7)$$

where η_{rmt} satisfies $P(\Lambda_{\text{rmt}}(w) > 2 \ln(\eta_{\text{rmt}}) | H_0) = \alpha$.

5. THEORETICAL ROC CURVE DERIVATION

A standard way to compare the plug-in and RMT detectors derived in (5) and (7) respectively is to compute their ROC curves. For a particular statistic $\Lambda(w)$, to compute theoretical ROC curves, we must compute

$$\begin{aligned} P_D &= P(\Lambda(w) > \gamma | w \in H_1) \\ P_F &= P(\Lambda(w) > \gamma | w \in H_0) \end{aligned} \quad (8)$$

for $-\infty < \gamma < \infty$. To do this, we explore the conditional CDF under each hypothesis for the statistics (4) and (6).

The conditional distribution of a test sample under H_0 is $w|H_0 \sim \mathcal{N}(0, I_k)$. Clearly $\Lambda_{\text{plugin}}(w)|H_0 \sim \chi_k^2$ and $\Lambda_{\text{rmt}}(w)|H_0 \sim \chi_{k_{\text{eff}}}^2$. The conditional distribution under H_1 is $w|H_1 \sim \mathcal{N}(\hat{U}^H U \Sigma x, I_k)$. As argued in Section 4.2, $\hat{U}^H U = SA$ is asymptotically diagonal. We have that $w_i|H_1 \approx \mathcal{N}(a_i s_i \sigma_i x_i, 1)$ are i.i.d. for $i = 1, \dots, k$. Therefore $\Lambda_{\text{plugin}}(w)|H_1 \sim \chi_k^2(\delta)$ and $\Lambda_{\text{rmt}}(w)|H_1 \sim \chi_{k_{\text{eff}}}^2(\delta)$ where

$$\delta = \sum_{i=1}^k \sigma_i^2 |\langle u_i, \hat{u}_i \rangle|^2 x_i^2 = \sum_{i=1}^{k_{\text{eff}}} \sigma_i^2 |\langle u_i, \hat{u}_i \rangle|^2 x_i^2 \quad (9)$$

is the non-centrality parameter for the noncentral chi-square distribution with k and k_{eff} degrees of freedom. An analytical expression for the asymptotic performance in the large matrix limit is obtained by substituting expressions from Theorem 3.1 in (9). To compute a theoretical ROC curve, we may solve for γ in terms of P_F and substitute this into the expression for P_D in (8). Doing so yields

$$\begin{aligned} P_{D_{\text{plugin}}} &= 1 - Q_{\chi_k^2(\delta)} \left(Q_{\chi_k^2}^{-1}(1 - P_F) \right) \\ P_{D_{\text{rmt}}} &= 1 - Q_{\chi_{k_{\text{eff}}}^2(\delta)} \left(Q_{\chi_{k_{\text{eff}}}^2}^{-1}(1 - P_F) \right) \end{aligned} \quad (10)$$

where Q is the cdf.

6. SIMULATION RESULTS AND DISCUSSION

6.1. ROC Curves

We consider a setting where $k_{\text{eff}} = 1 < k = 2$. For this setting, as seen in Figure 1, for any false alarm rate (P_F), the RMT detector achieves a higher probability of detection (P_D), demonstrating the sub-optimality of the plug-in detector. This is expected because $k_{\text{eff}} < k$ so that the plug-in detector is employing uninformative subspace components. The

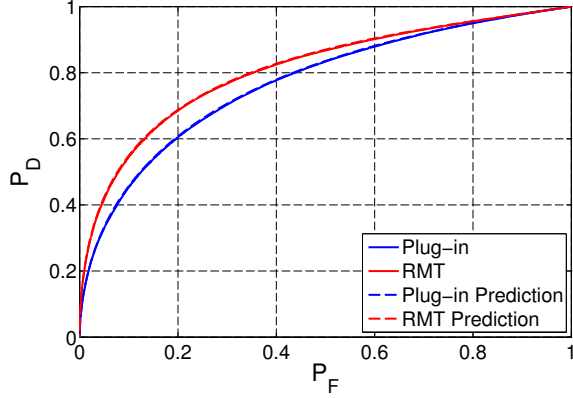


Fig. 1. Empirical and theoretical ROC curves for the plug-in and RMT matched subspace detectors. Empirical ROC curves were simulated with $n = 500$, $m = 500$, $k = 2$, $\Sigma = \text{diag}(3, 0.1)$, and $p = 0.8$. However, as σ_2 is below the critical threshold, $k_{\text{eff}} = 1$. The empirical ROC curves were computed using 5000 test samples and averaged over 25 trials. x was generated randomly for training samples but fixed for test samples. The theoretical ROC curves were obtained using (10). Note the excellent agreement and the performance gain realized by the RMT detector.

theoretical ROC curves in (10) match the empirically generated ROC curves validating the performance predictions of (10) which rely on Theorem 3.1.

6.2. Effect of Missing Data

Figure 2 examines the performance of each detector as a function of p . Again we observe the sub-optimality of the plug-in detector. The theoretical P_D prediction in (10) matches empirically achieved P_D for both detectors. As expected, as p decreases, the achieved probability of detection decreases. We note the presence of a critical $p_{\text{crit.}} := \sqrt{c}/\max_i(\sigma_i^2)$ obtained from Theorem 3.1, below which (in the large system limit) we may only achieve $P_D = P_F$; the rounding in Figure 2 is attributed to finite system approximation error.

7. CONCLUSION

We considered a deterministic MSD problem where the unknown low-rank signal subspace is estimated from noisy, limited, signal-bearing training data with missing entries. We used RMT to characterize the resulting performance and showed that using $k_{\text{eff}} \leq k$ subspace components is optimal. The relationship between k_{eff} and p was made explicit in Theorem 3.1 and we showed that detection better than random guessing (in the large system limit) is only achievable for $p > p_{\text{crit.}} := \sqrt{c}/\max_i(\sigma_i^2)$.

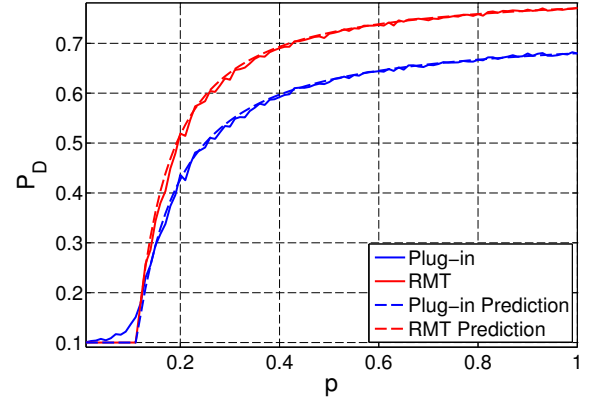


Fig. 2. Empirically computed probability of detection, P_D , for a fixed probability of false alarm, $P_F = 0.1$, for various p . Here, $n = 1000$, $m = 1000$, $k = 2$, $\Sigma = \text{diag}(3, 0.1)$. P_D was computed using (10) and x was generated as described in Figure 1. For values of $p \leq 1/9$, $k_{\text{eff}} = 0$ and performance degrades to $P_D = P_F + o(1)$ for both detectors. As p increases, $k_{\text{eff}} = 1$ allowing the detectors to achieve better than random guessing performance. When $k_{\text{eff}} > 0$ the plug-in detector is sub-optimal for all values of p .

8. REFERENCES

- [1] L.L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Sig. Proc.*, vol. 42, no. 8, pp. 2146–2157, 1994.
- [2] Y. Jin and B. Friedlander, "A CFAR adaptive subspace detector for second-order gaussian signals," *IEEE Trans. Sig. Proc.*, vol. 53, no. 3, pp. 871–884, 2005.
- [3] T. McWhorter and L. Scharf, *Matched subspace detectors for stochastic signals*, Defense Technical Information Center, 2003.
- [4] N. Asendorf and R.R. Nadakuditi, "Improving and characterizing the performance of stochastic matched subspace detectors when using noisy estimated subspaces," in *Proc. Asilomar SS&C*, November 2011.
- [5] R.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.
- [6] L. Balzano, B. Recht, and R. Nowak, "High-dimensional matched subspace detection when data are missing," *Proc. ISIT*, 2010, pp. 1638–1642.
- [7] F. Benaych-Georges and R.R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices," *Arxiv preprint arXiv:1103.2221*, 2011.