SUBSPACE DETECTION OF HIGH-DIMENSIONAL VECTORS USING COMPRESSIVE SAMPLING

Martin Azizyan and Aarti Singh
Machine Learning Department, Carnegie Mellon University

ABSTRACT

We consider the problem of detecting whether a high dimensional vector $\in \mathbb{R}^n$ lies in a r-dimensional subspace S, where $r \ll n$, given few compressive measurements of the vector. This problem arises in several applications such as detecting anomalies, targets, interference and brain activations. In these applications, the object of interest is described by a large number of features and the ability to detect them using only linear combination of the features (without the need to measure, store or compute the entire feature vector) is desirable. We present a test statistic for subspace detection using compressive samples and demonstrate that the probability of error of the proposed detector decreases exponentially in the number of compressive samples, provided that the energy off the subspace scales as n. Using information-theoretic lower bounds, we demonstrate that no other detector can achieve the same probability of error for weaker signals. Simulation results also indicate that this scaling is near-optimal.

Index Terms— subspace detection, compressed sensing

1. INTRODUCTION

We study the problem of detecting whether a high-dimensional vector $\mathbf{x} \in \mathbb{R}^n$ lies in a known low-dimensional subspace S, given few compressive measurements of the vector. The problem of testing whether a vector lies within a subspace is relevant for several tasks such as anomaly detection [1], medical imaging [2], hyperspectral target detection [3], radar signal processing [4], interference estimation [5], etc.

In high-dimensional settings, it is desirable to acquire only a small set of compressive measurements of the vector instead of measuring every coordinate. This reduces the amount of storage, communication, and computation needed. Recent advances in compressed sensing show that it is possible to *reconstruct* the original vector from few compressive measurements without significant loss in accuracy, provided that the vector is sparse or lies in a low-dimensional subspace. However, in some applications (as mentioned above) the objective is not to reconstruct the vector, but to *detect* whether the vector sensed using compressive measurements lies in a low-dimensional subspace or not. In this paper, we address this question and show that it is possible to detect when a compressively sampled vector lies in a known low-dimensional subspace using few compressive measurements. Notice that we cannot first reconstruct the vector and then test whether it lies in the subspace as the reconstruction might be arbitrarily poor if the signal did not lie in that subspace.

While a few papers have considered the problem of detection from compressive samples [6, 7, 8], these typically consider a simple hypothesis test of whether the vector $\mathbf{x}$ is $\mathbf{0}$ (i.e. observed vector is purely noise) or a known signal vector $\mathbf{s}$:

$$\mathcal{H}_0 : \mathbf{x} = 0 \quad \text{vs.} \quad \mathcal{H}_1 : \mathbf{x} = \mathbf{s} \tag{1}$$

In this paper, we consider the subspace detection setting, where the subspace is known but the exact signal vector is unknown. This set up leads to the composite hypothesis test:

$$\mathcal{H}_0 : \mathbf{x} \in S \quad \text{vs.} \quad \mathcal{H}_1 : \mathbf{x} \notin S$$

Equivalently, let $\mathbf{x}_\perp$ denote the component of $\mathbf{x}$ that does not lie in S. Now the composite hypothesis test can be stated as:

$$\mathcal{H}_0 : \|\mathbf{x}_\perp\|_2 = 0 \quad \text{vs.} \quad \mathcal{H}_1 : \|\mathbf{x}_\perp\|_2 > 0 \tag{2}$$

In [8], the authors also consider the unknown signal case, but do not consider a subspace setting. The most closely related paper is [9] where the authors consider that a few of the coordinates of the vector $\mathbf{x}$ are missing at random. Note that the missing observation case is a special case of compressive measurements where the rows of the measurement matrix have a single non-zero entry which picks out the corresponding coordinate. Thus, our results generalize that of [9], and additionally characterize the optimal information-theoretic threshold of signal to noise ratio needed to perform the testing reliably in high dimensions. While writing this paper, we also became aware of a paper that recently appeared on arXiv [10] that characterizes similar optimal threshold for detection of a sparse vector using compressive measurements. In this case, the subspace is specified by the span of a subset of the canonical basis vectors and the optimal threshold is shown to scale as n/m, however the probability of error is allowed to decrease arbitrarily slowly in the number of compressive measurements m. We propose detectors for an arbitrary subspace S, and require the probability of error to decay exponentially in the number of compressive measurements. As a result, the optimal threshold for energy off the subspace scales as n in our setting.

Our results show that it is possible to detect whether a high-dimensional vector $\in \mathbb{R}^n$ lies in a r-dimensional subspace S, where $r \ll n$ using only $m = \omega(r)$ [1] compressive

[1] $a = \omega(b)$ means that $a > b\cdot$ constant for every constant> 0.

samples, provided the energy off the subspace scales as n. In comparison, in the setting of Eq. (1), if the signal vector is known then the energy of the signal vector $\mathbf{s}$ can be a constant > 0 (with universal random measurements that do not depend on the subspace S) [6]. Even weaker signals can be detected if both the signal and subspace are known, and the measurement matrix is tailored to the subspace S [7, 11]. However, this requires knowing the subspace S at the time of making the compressive measurements. We do not consider that setting here, and instead focus on *universal* compressive measurements that can be used to test whether a high-dimensional vector lies in *any* set of known subspaces that do not have to be fixed at the time of measurement collection.

2. MEASUREMENT MODEL AND TEST STATISTIC

Let S be a known r-dimensional subspace of $\mathbb{R}^n$, spanned by the orthonormal columns of a matrix $\mathbf{U} \in \mathbb{R}^{n\times r}$. We are interested in determining whether an unknown vector $\mathbf{x} \in \mathbb{R}^n$ lies in S or not based only on a small number of compressive measurements. Specifically, for some $m \geq 1$, let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a random matrix with i.i.d. $\mathcal{N}(0,1)$ entries. Then we observe the m-dimensional vector $\mathbf{y} = \mathbf{A}(\mathbf{x}+\mathbf{w})$, where $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{n\times n})$ denotes noise with known variance σ^2 that is independent of $\mathbf{A}$. Notice that this noise model is analogous to the one studied in [12], and different from the more commonly studied case $\mathbf{y}' = \mathbf{A}\mathbf{x} + \mathbf{w}_m$ where $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{m\times m})$. In particular, for fixed $\mathbf{A}$ we have $\mathbf{y} \sim \mathcal{N}(\mathbf{A}\mathbf{x}, \sigma^2\mathbf{A}\mathbf{A}^T)$ and $\mathbf{y}' \sim \mathcal{N}(\mathbf{A}\mathbf{x}, \sigma^2\mathbf{I}_{m\times m})$. We focus on the former model as in most applications noise is inherently generated during the measurement process, while compressive linear measurements may be formed later on to optimize storage or data collection.

Define the projection operator $\mathbf{P}_{\mathbf{U}} = \mathbf{U}\mathbf{U}^T$. Then $\mathbf{x}_\perp = (\mathbf{I} - \mathbf{P}_{\mathbf{U}})\mathbf{x}$, where $\mathbf{x}_\perp$ is the component of $\mathbf{x}$ that does not lie in S, and $\mathbf{x} \in S$ iff $\|\mathbf{x}_\perp\|_2^2 = 0$. Analogously to [9], we define the test statistic $\mathbf{T} = \|(\mathbf{I} - \mathbf{P}_{\mathbf{BU}})(\mathbf{A}\mathbf{A}^T)^{-1/2}\mathbf{y}\|_2^2$ based on the observed vector $\mathbf{y}$ and study its properties, where $\mathbf{B} = (\mathbf{A}\mathbf{A}^T)^{-1/2}\mathbf{A}$. Here $\mathbf{P}_{\mathbf{BU}}$ is the projection operator onto the column space of $\mathbf{BU}$, specifically $\mathbf{P}_{\mathbf{BU}} = \mathbf{BU}((\mathbf{BU})^T\mathbf{BU})^{-1}(\mathbf{BU})^T$, if $((\mathbf{BU})^T\mathbf{BU})^{-1}$ exists.

3. MAIN RESULTS

For the sake of notational simplicity, throughout this section we directly work with the matrix $\mathbf{B}$ and its marginal distribution. Writing $\mathbf{y} = \mathbf{B}(\mathbf{x}+\mathbf{w})$, we have $\mathbf{T} = \|(\mathbf{I} - \mathbf{P}_{\mathbf{BU}})\mathbf{y}\|_2^2$. Notice that since $\mathbf{A}$ is i.i.d. normal, the distribution of the row span of $\mathbf{A}$ (and hence $\mathbf{B}$) will be uniform over m-dimensional subspaces of $\mathbb{R}^n$ [13]. Furthermore, due to the $(\mathbf{A}\mathbf{A}^T)^{-1/2}$ term, the rows of B will be orthonormal (almost surely). First we show that, in the absence of noise, the test statistic $\|(\mathbf{I} - \mathbf{P}_{\mathbf{BU}})\mathbf{B}\mathbf{x}\|_2^2$ is close to $m\|\mathbf{x}_\perp\|_2^2/n$ with high probability.

Theorem 1. *Let* $0 < r < m < n$, $0 < \alpha_0 < 1$ *and* $\beta_0, \beta_1, \beta_2 > 1$. *With probability at least* $1 - \exp[(1-\alpha_0+\log\alpha_0)m/2] - \exp[(1-\beta_0+\log\beta_0)m/2] - \exp[(1-\beta_1+\log\beta_1)m/2] - \exp[(1-\beta_2+\log\beta_2)r/2]$

$$\left(\alpha_0\frac{m}{n} - \beta_1\beta_2\frac{r}{n}\right)\|\mathbf{x}_\perp\|_2^2 \leq \|(\mathbf{I}-\mathbf{P}_{\mathbf{BU}})\mathbf{B}\mathbf{x}\|_2^2 \leq \beta_0\frac{m}{n}\|\mathbf{x}_\perp\|_2^2.$$

This implies the following corollary.

Corollary 1. *If* $m \geq c_1 r\log m$, *then with probability at least* $1 - c_2\exp[-c_3 m]$,

$$d_1\frac{m}{n}\|\mathbf{x}_\perp\|_2^2 \leq \|(\mathbf{I}-\mathbf{P}_{\mathbf{BU}})\mathbf{B}\mathbf{x}\|_2^2 \leq d_2\frac{m}{n}\|\mathbf{x}_\perp\|_2^2$$

for some universal constants $c_1 > 1$, $c_2 > 0$, $c_3 \in (0,1)$, $d_1 \in (0,1)$, $d_2 > 1$.

Corollary 1 states that given just **over** r noiseless compressive measurements, we can estimate $\|x_\perp\|_2^2$ accurately with high probability. In the presence of noise, t is natural to consider the hypothesis test:

$$\mathbf{T} = \|(\mathbf{I}-\mathbf{P}_{\mathbf{BU}})\mathbf{y}\|_2^2 \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \eta \qquad (3)$$

The following result bounds the false alarm level and missed detection rate of this test (for appropriately chosen η) assuming a lower bound on $\|\mathbf{x}_\perp\|_2^2$ under $\mathcal{H}_1$.

Theorem 2. *If the assumptions of Corollary 1 are satisfied, and if for any* $\mathbf{x} \in \mathcal{H}_1$

$$\|\mathbf{x}_\perp\|_2^2 \geq \sigma^2\frac{4e+2}{d_1}\left(1-\frac{r}{m}\right)n,$$

then

$$\mathbb{P}(\mathbf{T} \geq \eta|\mathcal{H}_0) \leq \exp[-c_4(m-r)]$$

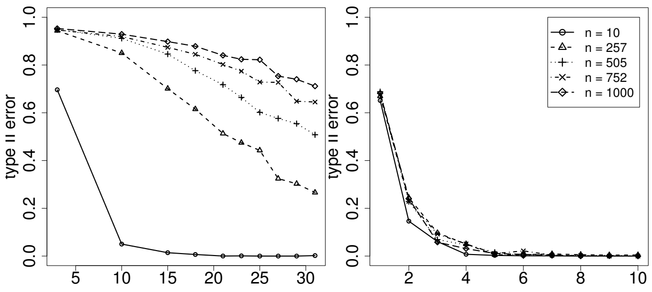
and

$$\mathbb{P}(\mathbf{T} \leq \eta|\mathcal{H}_1) \leq c_2\exp[-c_3 m] + \exp\left[-c_5(m-r)\right].$$

where $\eta = e\sigma^2(m-r)$, $c_4 = (e-2)/2$, $c_5 = (e+\log(2e+1))/2$, *and all other constants are as in Corollary 1.*

It is important to determine whether the performance of the test statistic we proposed can be improved further. The following theorem provides an information-theoretic lower bound on the probability of error of any test. A corollary of this theorem implies that the proposed test statistic is optimal, that is, every test with probability of missed detection and false alarm decreasing exponentially in the number of compressive samples m requires that the energy off the subspace scale as n.

Theorem 3. *Let* P_0 *be the joint distribution of* $\mathbf{B}$ *and* $\mathbf{y}$ *under the null hypothesis. Let* P_1 *be the joint distribution of* $\mathbf{B}$ *and* $\mathbf{y}$ *under the alternate hypothesis where* $\mathbf{y} = \mathbf{B}(\mathbf{x}+\mathbf{w})$, $\mathbf{x} \sim \pi$



(a) $\|\mathbf{x}_\perp\|_2$ v.s. error (b) $\|\mathbf{x}_\perp\|_2/\sqrt{n}$ v.s. error

Fig. 1: Simulation results. Type II error averaged over 100 runs for different values of n. Across all trials, we set $\sigma^2 = 1$, $r = 1$, $m = 5$, $\mathbf{U} = (1, 0, ..., 0)^T$, and the Type I error is $\alpha = 0.05$. The detection threshold was set to $\eta_m = F^{-1}(1 - \alpha; m-1)$ ($F^{-1}(\cdot, k)$ is the inverse CDF of χ^2_k).

independently of $\mathbf{B}$ *and* $\mathbf{w}$*, and* π *is the uniform distribution over* $\mathbf{x}$ *such that* $\mathbf{x} = \mathbf{x}_\perp$ *and* $\|\mathbf{x}\|_2 = M$ *for some* $M > 0$*. If conditions of Corollary 1 are satisfied, then*

$$\inf_{\psi} \max_{i=0,1} P_i(\psi \neq i) \geq \frac{1}{4} \exp\left[-\frac{M^2}{2\sigma^2}\frac{m}{n}\right]$$

where the infimum is over all hypothesis tests ψ*.*

Corollary 2. *If there exists a hypothesis test* ψ *based on* $\mathbf{B}$ *and* $\mathbf{y}$ *such that for all* n *and* σ^2

$$\max_{i=0,1} \mathbb{P}(\psi \neq i | \mathcal{H}_i) \leq C_0 \exp[-C_1(m-r)]$$

for some $C_0, C_1 > 0$*, then there exists some* $C > 0$ *such that* $\|\mathbf{x}_\perp\|_2^2 \geq C\sigma^2(1 - r/m)n$ *for any* $\mathbf{x} \in \mathcal{H}_1$ *and all* n *and* σ^2*.*

4. SIMULATION RESULTS

The bounds in Theorem 2 and Corollary 2 state that for fixed r, $\|\mathbf{x}_\perp\|_2$ needs to scale as approximately $\sqrt{n}$ to ensure low error under $\mathcal{H}_1$. We perform simulations to demonstrate the effect of this scaling as follows.

We measure detection error as a function of $\|\mathbf{x}_\perp\|_2$ for some values of n, for fixed false alarm level α (note that the true distribution under the null hypothesis is known, so we can construct an exact level α test). If the bounds in Theorem 2 are tight, we expect to see that for fixed $\|\mathbf{x}_\perp\|_2$, larger n leads to larger error; we observe this in Figure 1a. Moreover, we expect that if we rescale the x-axis to $\|\mathbf{x}_\perp\|_2/\sqrt{n}$, the error becomes independent of n, as is the case in Figure 1b. Thus, our simulations verify that the proposed test statistic can reliably detect if a n-dim vector lies in the given subspace provided the energy off the subspace scales as n.

5. CONCLUSION

This paper shows that it is possible to detect whether a high-dimensional vector lies in a subspace with very few compressive measurements. We precisely characterized the amount of energy outside the subspace needed for reliable detection and verified this with simulations. The test statistic we propose is optimal in the sense that no other test can detect vectors with smaller energy off the subspace while ensuring that the probability of error decays exponentially with the number of compressive samples. Since the measurement model we consider is not specialized to the problem at hand, the proposed approach is universal and can be used in to detect energy outside any given subspace. This work also has important ramifications for sequential basis learning, where the subspace of interest may not be known a priori, but needs to be learnt from a collection of high-dimensional vectors that are expected to lie in some lower-dimensional subspace. We plan to investigate this direction in future work.

6. ACKNOWLEDGEMENTS

This research is supported in part by AFOSR under grant FA9550-10-1-0382 and NSF under grant IIS-1116458.

7. PROOFS

We will use the following results for random projections.

Theorem 4. *Let* $\mathbf{z} \in \mathbb{R}^d$ *be distributed uniformly over the set of unit norm vectors in* $\mathbb{R}^d$*. For* $k < d$*, let* $\mathbf{z}' = (\mathbf{z}_1, ..., \mathbf{z}_k, 0, ..., 0) \in \mathbb{R}^d$ *be the projection of* $\mathbf{z}$ *onto the span of* $\mathbf{e}_1, ..., \mathbf{e}_k$*, the first* k *elements of the standard basis. Then for any* $0 < \alpha < 1$ *and* $\beta > 1$*,*

$$\mathbb{P}\left(\|\mathbf{z}'\|_2^2 \leq \alpha k/d\right) \leq \exp\left((1 - \alpha + \log\alpha)k/2\right) \text{ and}$$
$$\mathbb{P}\left(\|\mathbf{z}'\|_2^2 \geq \beta k/d\right) \leq \exp\left((1 - \beta + \log\beta)k/2\right).$$

As in Section 3, in the following we assume the distribution of $\mathbf{B}$ is such that its rows are orthonormal and the row span of $\mathbf{B}$ is distributed uniformly over m-dimensional subspaces of $\mathbb{R}^n$. We will also always assume $0 < r < m < n$.

Proposition 1. *Let* $\mathbf{z}_1, ..., \mathbf{z}_n \in \mathbb{R}^m$ *be the columns of* $\mathbf{B}$*, and* $\mathbf{P}_r \in \mathbb{R}^{m\times m}$ *the projection operator on the span of* $\mathbf{z}_1, ..., \mathbf{z}_r$*. For any* $\beta_1, \beta_2 > 1$*,* $\|\mathbf{P}_r\mathbf{z}_{r+1}\|_2^2 \leq \beta_1\beta_2 r/n$ *with probability at least* $1 - \exp\left((1 - \beta_1 + \log\beta_1)m/2\right) - \exp\left((1 - \beta_2 + \log\beta_2)r/2\right)$*.*

Proof. Let $\widetilde{\mathbf{z}}_{r+1} = \mathbf{z}_{r+1}/\|\mathbf{z}_{r+1}\|_2$. Then $\|\mathbf{P}_r\mathbf{z}_{r+1}\|_2^2 = \|\mathbf{P}_r\widetilde{\mathbf{z}}_{r+1}\|_2^2\|\mathbf{z}_{r+1}\|_2^2$. By Theorem 4, $\|\mathbf{z}_{r+1}\|_2^2 \leq \beta_1 m/n$ w.p. at least $1 - \exp[(1 - \beta_1 + \log\beta_1)m/2]$. Now consider $\|\mathbf{P}_r\widetilde{\mathbf{z}}_{r+1}\|_2^2$. This is the norm of the projection of a random unit vector in $\mathbb{R}^m$ onto a random r-dimensional subspace, so by Theorem 4 we have that $\|\mathbf{P}_r\widetilde{\mathbf{z}}_{r+1}\|_2^2 \leq \beta_2 r/m$ with probability at least $1 - \exp[(1 - \beta_2 + \log\beta_2)r/2]$ and the result follows by combining the two terms. ☐

Proof of Theorem 1: Define $\mathbf{x}_\parallel = \mathbf{x} - \mathbf{x}_\perp$. Clearly $\mathbf{x}_\parallel = \mathbf{P}_\mathbf{U}\mathbf{x}_\parallel$, so $(\mathbf{I} - \mathbf{P}_\mathbf{BU})\mathbf{B}\mathbf{x}_\parallel = \mathbf{0}$ and $(\mathbf{I} - \mathbf{P}_\mathbf{BU})\mathbf{B}\mathbf{x} = (\mathbf{I} - \mathbf{P}_\mathbf{BU})\mathbf{B}\mathbf{x}_\perp$. From here the case when $\|\mathbf{x}_\perp\|_2 = 0$ follows easily, so for the remainder of the proof assume $\|\mathbf{x}_\perp\|_2 > 0$. Clearly $(\mathbf{I} - \mathbf{P}_\mathbf{BU})\mathbf{B}\mathbf{x}_\perp$ and $\mathbf{P}_\mathbf{BU}\mathbf{B}\mathbf{x}_\perp$ are orthogonal. Defining $\widetilde{\mathbf{x}}_\perp = \mathbf{x}_\perp/\|\mathbf{x}_\perp\|_2$, we can write $\|(\mathbf{I}-\mathbf{P_{BU}})\mathbf{B}\widetilde{\mathbf{x}}_\perp\|_2^2 = \|\mathbf{B}\widetilde{\mathbf{x}}_\perp\|_2^2 - \|\mathbf{P_{BU}}\mathbf{B}\widetilde{\mathbf{x}}_\perp\|_2^2$. By Theorem 4, with probability at least $1-\exp[(1-\alpha_0+\log\alpha_0)m/2]-\exp[(1-\beta_0+\log\beta_0)m/2]$

$$\alpha_0 m/n \le \|\mathbf{B}\widetilde{\mathbf{x}}_\perp\|_2^2 \le \beta_0 m/n.$$

Now consider $\|\mathbf{P_{BU}}\mathbf{B}\widetilde{\mathbf{x}}_\perp\|_2^2$. Due to the rotational symmetry of the distribution of $\mathbf{B}$, we can assume $\mathbf{U} = (\mathbf{e}_1, ..., \mathbf{e}_r)$ and $\widetilde{\mathbf{x}}_\perp = \mathbf{e}_{r+1}$. So applying Proposition 1, w.p. $\ge 1-\exp[(1-\beta_1+\log\beta_1)m/2]-\exp[(1-\beta_2+\log\beta_2)r/2]$,

$$\|\mathbf{P_{BU}}\mathbf{B}\widetilde{\mathbf{x}}_\perp\|_2^2 \le \beta_1\beta_2 r/n$$

and the result follows by combining the two bounds. □

Lemma 1. *If $\mathbf{P}$ is a rank k projection operator, $\mu \in \mathbb{R}^m$, and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, then $\frac{1}{\sigma^2}\|\mathbf{P}(\mu+\mathbf{w})\|_2^2 \sim \chi_k^2\left(\frac{\|\mathbf{P}\mu\|_2^2}{\sigma^2}\right)$.*

Proof. See [14], specifically pages 64 and 146. □

Proposition 2. *For $\epsilon > 0$, if $y \sim \chi_k^2$, then*

$$\mathbb{P}(y \ge (1+\epsilon)k) \le \exp\left(-k(\epsilon - \log(1+\epsilon))/2\right).$$

If $\lambda + k > \eta > 0$ and $z \sim \chi_k^2(\lambda)$, then $\mathbb{P}(z \le \eta) \le$

$$\exp\left[\left(g(\eta)/2 - \eta + 2\eta\lambda/g(\eta) - \lambda\right)/2\right]\left(2\eta/g(\eta)\right)^{k/2}$$

where $g(\eta) = k + \sqrt{k^2 + 4\eta\lambda}$.

Proof of Theorem 2: Observe that for fixed $\mathbf{B}$, the test statistic is equal to $\mathbf{T} = \|(\mathbf{I}-\mathbf{P_{BU}})\mathbf{y}\|_2^2 = \|\mathbf{B}^T(\mathbf{I}-\mathbf{P_{BU}})\mathbf{B}\mathbf{x} + \mathbf{B}^T(\mathbf{I}-\mathbf{P_{BU}})\mathbf{B}\mathbf{w}\|_2^2$. Clearly $\mathbf{B}^T(\mathbf{I}-\mathbf{P_{BU}})\mathbf{B}$ is a rank $m-r$ projection operator. By Lemma 1, conditioned on $\mathbf{B}$, $\mathbf{T}/\sigma^2 \sim \chi_{m-r}^2\left(\lambda(\mathbf{B},\mathbf{x})/\sigma^2\right)$, where we have defined $\lambda(\mathbf{B},\mathbf{x}) = \|(\mathbf{I}-\mathbf{P_{BU}})\mathbf{B}\mathbf{x}\|_2^2$. For $\mathbf{x} \in \mathcal{H}_0$, $\lambda(\mathbf{B},\mathbf{x}) \equiv 0$ (see proof of Theorem 1), and $\mathbf{T}/\sigma^2 \sim \chi_{m-r}^2$. From the first part of Proposition 2 we see that since $\eta = e\sigma^2(m-r)$,

$$\mathbb{P}(\mathbf{T}/\sigma^2 \ge \eta/\sigma^2|\mathcal{H}_0) \le \exp[-(m-r)(e-2)/2].$$

Now consider $\mathcal{H}_1$. Let $\mathfrak{G}(\mathbf{B},\mathbf{x})$ be the indicator of the event

$$d_1 m/n\|\mathbf{x}_\perp\|_2^2 \le \lambda(\mathbf{B},\mathbf{x}) \le d_2 m/n\|\mathbf{x}_\perp\|_2^2.$$

By Corollary 1, $\mathbb{E}_\mathbf{B}(1-\mathfrak{G}(\mathbf{B},\mathbf{x})) \le c_2\exp[-c_3 m]$ for fixed $\mathbf{x}$. Together with the condition on $\|\mathbf{x}_\perp\|_2^2$ this gives

$$\mathbb{P}(\mathbf{T} \le \eta|\mathcal{H}_1) \le c_2\exp[-c_3 m] + \mathbb{P}(z \le e(m-r))$$

where $z \sim \chi_{m-r}^2\left((4e+2)(m-r)\right)$. By Proposition 2,

$$\mathbb{P}(z \le e(m-r)) \le \exp\left[-(m-r)\left(e + \log(2e+1)\right)/2\right] \square$$

Proof of Theorem 3: Let K be the Kullback-Leibler divergence. Then $\inf_\psi \max_{i=0,1} P_i(\psi \ne i) \ge e^{-K(P_0,P_1)}/4$ [15]. Let q be the density of $\mathbf{B}$ and $p(\mathbf{y};\mu,\Sigma)$ that of $\mathcal{N}(\mu,\Sigma)$. Note that, under P_0, $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{m\times m})$ since rows of $\mathbf{B}$ are orthonormal. By Jensen's inequality,

$$\begin{aligned} K(P_0,P_1) &= \mathbb{E}_\mathbf{B}\mathbb{E}_\mathbf{y}\log\frac{p(\mathbf{y};\mathbf{0},\sigma^2\mathbf{I}_{m\times m})q(\mathbf{B})}{\mathbb{E}_\mathbf{x} p(\mathbf{y};\mathbf{B}\mathbf{x},\sigma^2\mathbf{I}_{m\times m})q(\mathbf{B})} \\ &\le \mathbb{E}_\mathbf{B}\mathbb{E}_\mathbf{x}\|\mathbf{B}\mathbf{x}\|_2^2/(2\sigma^2) = M^2 m/(2\sigma^2 n). \end{aligned}$$

□

8. REFERENCES

[1] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, and A. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Mag.*, Jan 2002.

[2] B. Ardekani, J. Kershaw, K. Kashikura, and I. Kanno, "Activation detection in functional mri using subspace modeling and maximum likelihood estimation," *IEEE Trans. Medical Imaging*, vol. 18, no. 2, Feb 1999.

[3] H. Kwon and N. Nasrabadi, "Kernel matched subspace detectors for hyperspectral target detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, Feb 2006.

[4] M. Rangaswamy, F. Lin, and K. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Signal Processing*, vol. 84, pp. 1653–1665, Sept 2004.

[5] M. McCloud and L. Scharf, "Interference estimation with applications to blind multiple-access communication over fading channels," *IEEE Transactions on Information Theory*, vol. 46, no. 3, May 2000.

[6] M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Detection and estimation with compressive measurements," Tech. Rep., 2006.

[7] Jose L. Paredes, Zhongmin Wang, Gonzalo R. Arce, and Brian M. Sadler, "Compressive matched subspace detection," in *EUSIPCO*, 2009.

[8] J. Haupt and R. Nowak, "Compressive sampling for signal detection," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Proc (ICASSP)*, 2007, pp. 1509–1512.

[9] L. Balzano, B. Recht, and R. Nowak, "High dimensional matched subspace detection when data are missing," in *ISIT*, 2010.

[10] Ery Arias-Castro, "Detecting a vector based on linear measurements," `http://arxiv.org/pdf/1112.6235v1.pdf`, December, 2011.

[11] Z Zhongmin Wang, G.R. Arce, and B.M. Sadler, "Subspace compressive detection for sparse signals," in *ICASSP*, 2008, pp. 3873–3876.

[12] E. J. Candès and M. A. Davenport, "How well can we estimate a sparse vector?," *ArXiv e-prints*, Apr. 2011.

[13] A. T. James, "Normal multivariate analysis and the orthogonal group," *The Annals of Mathematical Statistics*, vol. 25, no. 1, pp. pp. 40–75, 1954.

[14] L.L. Scharf and C. Demeure, *Statistical signal processing: detection, estimation, and time series analysis*, Addison-Wesley Reading, Massachusetts, 1991.

[15] A.B. Tsybakov, *Introduction to nonparametric estimation*, Springer series in statistics. Springer, 2009.