HYPERSPECTRAL TARGET DETECTION FROM INCOHERENT PROJECTIONS

Kalyani Krishnamurthy, Maxim Raginsky and Rebecca Willett

Department of Electrical and Computer Engineering
Duke University, Durham, NC 27708

ABSTRACT

This paper studies the detection of spectral targets corrupted by a colored Gaussian background from noisy, incoherent projection measurements. Unlike many detection methods designed for incoherent projections, the proposed approach a) is computationally efficient, b) allows for spectral backgrounds behind potential targets, and c) yields theoretical guarantees on detector performance. In particular, the theoretical performance bounds highlight fundamental tradeoffs among the number of measurements collected, the spectral resolution of targets, the amount of background signal present, signal-to-noise ratio, and the similarity between potential targets in a dictionary.

Index Terms— Hyperspectral imaging, Target detection, False Discovery Rate, Incoherent projections, Compressive sensing

1. HYPERSPECTRAL TARGET DETECTION

Hyperspectral imaging is a useful tool in such applications as remote sensing and astronomical imaging because it allows inference of the material composition of a scene and localization of materials of interest. Target recognition performance is typically affected by the spatial and spectral resolution of this data, and there are several challenges associated with collecting and using high-resolution hyperspectral data. To collect a hyperspectral image of size $M \times M \times N$ (i.e. M^2 spatial locations, each with N spectral bands), conventional cameras require M^2N detectors, which may be prohibitively large or expensive.

Recent developments in the field of compressed sensing offer an attractive potential solution to this problem. The core idea of compressed sensing is that if a signal or image of interest is sparse in some domain, then it can be reconstructed accurately from very few (relative to the dimension of the signal or image) non-adaptive measurements [1, 2]. Several researchers exploit compressed sensing to build practical systems that acquire such measurements and use sophisticated reconstruction algorithms to recover the true underlying intensity [3, 4]. One example is a single-shot compressive spectral imaging system which collects coded projections of each spectrum in the spectral image [3]. The number of detectors required in this setup scales with the number of spatial locations M^2 and the number of projections $K \leq N$. This paper explores how such hyperspectral cameras can be used in target detection applications. In particular, to investigate the applicability of such systems to target detection, we consider the following questions: 1) What kinds of measurements will enable efficient target detection? 2) How can one perform target detection from incoherent projections? 3) How can one incorporate background models into this setup? 4) How does performance scale with the number of measurements collected, the spectral resolution of targets, the amount of background signal present, the signal-to-noise ratio, and the similarity between potential targets in a dictionary? We address these questions in this paper.

This work was supported by NSF Award No. DMS-08-11062, DARPA Grant No. HR0011-09-1-0036, and AFRL Grant No. FA8650-07-D-1221.

1.1. Previous investigations

Over the past few years, researchers have investigated the benefits of exploiting compressed sensing in classification and target detection [6, 7, 8]. In [6], the authors propose a matching pursuit based algorithm, called the *Incoherent Detection and Estimation Algorithm* (IDEA), to detect the presence of a signal of interest against a strong interfering signal from noisy and incoherent projections. The algorithm is shown to perform well on experimental data sets under some strong assumptions on the sparsity of the signal of interest and the interfering signal. However, theoretical results are not provided on the performance of this detector relative to the number of measurements that one collects. Davenport, et al. [7] propose a matched filter based algorithm, called the *smashed filter*, to classify an image in $\mathbb{R}^N$ to one of m predetermined classes based on K incoherent projections, where $K \ll N$. In this approach, the authors model the images in each class using a manifold and discuss the kinds of incoherent projections that would preserve the manifold structure under each class. An image is classified based on a nearest-neighbor search on each of the m different manifolds. This approach offers theoretical results on the number of measurements needed relative to the size of each image and the intrinsic dimensionality of the manifolds. However, it is computationally expensive since it involves searching over different manifolds. Also, it is not immediately clear how to incorporate background information in such a manifold model. In [8], the authors present a hypothesis testing framework to classify a single signal in $\mathbb{R}^N$ as one of m different signals using $K < N$ random projections and give bounds on the probability of classification error. Though this approach is simple to implement, its extension to hyperspectral target detection problems is nontrivial because of background and the need for multiple testing performance bounds when considering M^2 spatial locations.

2. PROBLEM SETUP

Let us consider the problem of detecting spectral targets from incoherent hyperspectral projections of the form

$$\boldsymbol{y}(x_1, x_2, \cdot) = \boldsymbol{\Phi}(\alpha \boldsymbol{f}^*(x_1, x_2, \cdot) + \boldsymbol{b}) + \boldsymbol{n} \tag{1}$$

where

- $x_1, x_2 \in \{1, \ldots, M\}$ correspond to the spatial coordinates of the hyperspectral data,
- α is a measure of the signal-to-noise ratio which is assumed to be the same at every spatial location,

- $\boldsymbol{f}^*(x_1, x_2, \cdot)$ is one of the m possible unit-norm targets from a known spectral dictionary $\mathcal{D} = \{\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}, \ldots, \boldsymbol{f}^{(m)}\}$,
- $\boldsymbol{\Phi} \in \mathbb{R}^{K \times N}$ is a random measurement matrix with $K < N$,
- $\boldsymbol{b} \in \mathbb{R}^N \sim \mathcal{N}(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$ is the background noise vector and
- $\boldsymbol{n} \in \mathbb{R}^K \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ is the i.i.d. sensor noise that corrupts the observations.

For simplicity, let us index each observed spectrum by $\boldsymbol{y}_i = \boldsymbol{\Phi}(\alpha \boldsymbol{f}_i^* + \boldsymbol{b}) + \boldsymbol{n}$ for $i = 1, \ldots, M^2$. Assume that $\boldsymbol{\Phi}$, $\boldsymbol{b}$ and $\boldsymbol{n}$ are independent of each other, and that α is known or can be estimated. Typically, in hyperspectral target detection, the background corresponding to the scene of interest and the sensor noise are modeled together by a colored multivariate Gaussian distribution [9]. However, in our case, it is important to distinguish the two because of the presence of the projection operator $\boldsymbol{\Phi}$. The projection operator acts upon the background spectrum in the same way as on the target spectrum, but it does not affect the sensor noise. Suppose that the end user wants to test for the presence of one target versus the rest, but it is not known a priori which target from $\mathcal{D}$ the user wants to detect. With this problem setup, our goal is to find a way to perform hyperspectral target detection from $\boldsymbol{y} \in \mathbb{R}^{M \times M \times K}$, characterize the detector performance, and relate it to the number of observations we collect.

3. OUR APPROACH

We begin by transforming the observation model in (1) through a series of operations to arrive at a model of the form

$$\boldsymbol{y}_i = \alpha \boldsymbol{A} \boldsymbol{f}_i^* + \boldsymbol{z}$$

for $i \in \{1, \ldots, M^2\}$, where the entries of $\boldsymbol{A}$ are i.i.d. draws from $\mathcal{N}(0, 1/K)$ and $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ is white Gaussian noise. We will later justify this particular choice of $\boldsymbol{A}$.

Suppose that there is enough background training data available to estimate the background mean $\boldsymbol{\mu}_b$ and covariance matrix $\boldsymbol{\Sigma}_b$. We can assume without loss of generality that $\boldsymbol{\mu_b} = \boldsymbol{0}$ since $\boldsymbol{\Phi \mu_b}$ can be subtracted from $\boldsymbol{y}$ in practice. Given $\boldsymbol{\Phi}$, it is straightforward to show that $\boldsymbol{n}' \triangleq \boldsymbol{\Phi b} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Phi \Sigma_b \Phi}^T + \sigma^2 \boldsymbol{I})$. Next, we apply the whitening filter given by $C_{\boldsymbol{\Phi}} = (\boldsymbol{\Phi \Sigma_b \Phi}^T + \sigma^2 \boldsymbol{I})^{-1/2}$, and it can be easily shown that $C_{\boldsymbol{\Phi}} \boldsymbol{n}' \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$; that is, the transformed noise is white and Gaussian. The subscript on C indicates the C is a function of $\boldsymbol{\Phi}$. After whitening we observe,

$$\boldsymbol{y}_i' = C_{\boldsymbol{\Phi}} \boldsymbol{y} = \alpha \boldsymbol{A} \boldsymbol{f}_i^* + C_{\boldsymbol{\Phi}} \boldsymbol{n}'$$

where $\boldsymbol{A} = C_{\boldsymbol{\Phi}} \boldsymbol{\Phi}$ and $i \in \{1, \ldots, M^2\}$.

It is important to consider the impact of $\boldsymbol{A}$ on the spectral dictionary of interest to develop a suitable target detection approach. Let $\ell \in \{1, \ldots, m\}$ be the index of the target spectrum of interest. The presence of the target spectrum $\boldsymbol{f}^{(\ell)} \in \mathcal{D}$ at every spatial location $i = 1, \ldots, M^2$ can be detected using the following criterion:

$$\text{Declare } \boldsymbol{f}_i^* = \boldsymbol{f}^{(\ell)} \text{ if } \|\boldsymbol{y}_i - \alpha \boldsymbol{A} \boldsymbol{f}^{(\ell)}\|^2 \leq \|\boldsymbol{y}_i - \alpha \boldsymbol{A} \widetilde{\boldsymbol{f}}\|^2 \quad (2)$$

for all $\widetilde{\boldsymbol{f}} \in \mathcal{D} \backslash \boldsymbol{f}^{(\ell)}$, where $\mathcal{D} \backslash \boldsymbol{f}^{(\ell)}$ refers to the dictionary $\mathcal{D}$ excluding the target spectrum $\boldsymbol{f}^{(\ell)}$. In words, we declare that target $\boldsymbol{f}^{(\ell)}$ is present if no other target in $\mathcal{D}$ is a better fit to the data. It is well known that, with high probability, a projection matrix whose entries are i.i.d. draws from $\mathcal{N}(0, 1/K)$ preserves distances between vectors in $\mathbb{R}^N$ [10]. Thus, if we choose the entries of $\boldsymbol{A}$ as $a_{i,j} \sim \mathcal{N}(0, 1/K)$ for $i = 1, \ldots, K$ and $j = 1, \ldots, N$, then the corresponding $\boldsymbol{\Phi}$ can be found by solving the following equation:

$$\boldsymbol{\Phi} = C_{\boldsymbol{\Phi}}^{-1} \boldsymbol{A} = \left(\boldsymbol{\Phi \Sigma_b \Phi}^T + \sigma^2 I\right)^{1/2} \boldsymbol{A}. \quad (3)$$

Using linear algebra and matrix theory, it is possible to show that if $\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{A \Sigma_b A}^T$ is positive definite, then

$$\boldsymbol{\Phi} = \sigma \boldsymbol{B}^{-1/2} \boldsymbol{A}$$

satisfies (3) [1]. The following lemma gives the conditions under which $\boldsymbol{B}$ is positive definite.

Lemma 1: If the largest eigenvalue of $\boldsymbol{\Sigma}_b$ satisfies

$$\lambda_{\max} < \frac{1}{\|\boldsymbol{A}\|^2},$$

where $\|\boldsymbol{A}\|$ is the spectral norm of $\boldsymbol{A}$, then $\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{A \Sigma_b A}^T$ is positive definite.

The proof of this lemma is provided in the appendix. This lemma draws an interesting relationship between the maximum background perturbation that the system can tolerate and the spectral norm of the measurement matrix, which in turn varies with K and N. The lemma below gives an upper bound on $\lambda_{\max}$ as a function of the dimensions of $\boldsymbol{A}$.

Lemma 2: There exists an absolute constant $c > 0$, such that if $\lambda_{\max} < \frac{1}{c^2\left(\sqrt{\frac{N}{K}}+1\right)^2}$, then, with high probability, B is positive definite.

The proof of this lemma is given in the appendix. For a given N, the upper bound on $\lambda_{\max}$ increases as K increases, which implies that the system can tolerate more background perturbation if we collect more measurements.

The hyperspectral target detection problem can then be cast as the following multiple hypothesis testing problem:

$$\mathcal{H}_{0i}^{(\ell)} : \boldsymbol{f}_i^* = \boldsymbol{f}^{(\ell)} \quad (4)$$

$$\mathcal{H}_{1i}^{(\ell)} : \boldsymbol{f}_i^* \neq \boldsymbol{f}^{(\ell)} \quad (5)$$

for $i = 1, \ldots, M^2$. Note that the hypotheses are defined so that the i^{th} null hypothesis corresponds to the presence of the target spectrum in the i^{th} spatial location. Let us define

$$\widehat{\boldsymbol{f}}_i \triangleq \underset{\widetilde{\boldsymbol{f}} \in \mathcal{D}}{\arg\min} \|\boldsymbol{y}_i - \alpha \boldsymbol{A} \widetilde{\boldsymbol{f}}\|^2. \quad (6)$$

The decision rule in (2) can be equivalently written as

$$\text{Declare } \mathcal{H}_{0i}^{(\ell)} \text{ if } \widehat{\boldsymbol{f}}_i = \boldsymbol{f}^{(\ell)} \quad (7)$$

for $i = 1, \ldots, M^2$. Note that it is not always necessary to compute $\widehat{\boldsymbol{f}}_i$ at every spatial location to declare $\mathcal{H}_{1i}^{(\ell)}$. To see this, observe that it takes $\mathcal{O}(NK)$ operations to compute $\|\boldsymbol{y}_i - \alpha \boldsymbol{A} \widetilde{\boldsymbol{f}}\|^2$ for any $\widetilde{\boldsymbol{f}} \in D$ and *at most* m such computations to find some $\widetilde{\boldsymbol{f}} \in \mathcal{D} \backslash \boldsymbol{f}^{(\ell)}$ such that $\|\boldsymbol{y}_i - \alpha \boldsymbol{A} \widetilde{\boldsymbol{f}}\|^2 < \|\boldsymbol{y}_i - \alpha \boldsymbol{A} \boldsymbol{f}^{(\ell)}\|^2$. However, it always takes $\mathcal{O}(mNK)$ operations to compute $\widehat{\boldsymbol{f}}_i$.

We characterize the performance of this detector using the positive False Discovery Rate (pFDR) criterion introduced by Storey

[1] The authors would like to thank Prof. Roummel Marcia for fruitful discussions related to this work.

[11] since it lets us analyze the errors that one encounters in testing multiple hypotheses simultaneously. The pFDR in this context is the expected ratio of the number of missed targets to the total number of pixels declared not to have a target present, subject to at least one pixel being target-free. (Note that this ratio is traditionally referred to as the positive false *nondiscovery* rate (pFNDR), but is technically the pFDR in this context because of our definitions of the null and alternate hypotheses.) With our problem setup, for each hypothesis test, the distribution of the test statistic $\boldsymbol{y}_i$ under the null hypothesis is $\boldsymbol{y}_i|\mathcal{H}_{0i}^{(\ell)} \sim \mathcal{N}(\alpha \boldsymbol{A}\boldsymbol{f}^{(\ell)}, 1)$, and under the alternate hypothesis is $\boldsymbol{y}_i|\mathcal{H}_{1i}^{(\ell)} \sim \sum_{j=1,j\neq\ell}^{m} \frac{1}{m-1}\mathcal{N}(\alpha \boldsymbol{A}\boldsymbol{f}^{(j)}, 1)$; under both hypotheses, $\boldsymbol{y}_i$'s are i.i.d. across all spatial locations. Storey [11] showed that the pFDR for such identical hypothesis tests is given by

$$\mathrm{pFDR}(\ell) = \mathbb{P}\left(\mathcal{H}_0^{(\ell)}|\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(\ell)}\right) \equiv \mathbb{P}\left(\boldsymbol{f}^* = \boldsymbol{f}^{(\ell)}|\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(\ell)}\right) \tag{8}$$

where the subscript i is dropped to signify that the tests are identical. The following theorem presents our main result:

Theorem 1: Suppose that one performs M^2 identical hypotheses tests of the form (4) and (5) such that $\boldsymbol{y}_i|\mathcal{H}_{zi}^{(\ell)}$ are i.i.d for $z \in \{0, 1\}$, and makes decisions according to (7). Then the worst-case pFDR *satisfies the following bound as long as the background covariance matrix satisfies the bound of Lemma 1:*

$$\max_{\ell\in\{1,\ldots,m\}} \mathrm{pFDR}(\ell) \leq \left(\left(1 + \frac{\alpha^2 d_{\min}}{4K}\right)^{K/2} - m\right)^{-1} \tag{9}$$

where

$$d_{\min} = \min_{\boldsymbol{f}^{(i)},\boldsymbol{f}^{(j)}\in\mathcal{D}, i\neq j} \|\boldsymbol{f}^{(i)} - \boldsymbol{f}^{(j)}\|^2.$$

The proof of this theorem is provided in the appendix. Observe that this bound is independent of N, and is only a function of K, m, the signal-to-noise ratio α, and $d_{\min}$. The lack of dependence on N is not unexpected. Indeed, when we are interested in preserving pairwise distances among the members of a fixed dictionary of size m, the celebrated Johnson–Lindenstrauss lemma [10, **?**] says that, with high probability, $K \asymp \log m$ random Gaussian projections suffice, regardless of the ambient dimension N. This is precisely the regime we are working with here. For the bound in (9) to be nonnegative and ≤ 1, one needs to ensure that $\left(1 + \frac{\alpha^2 d_{\min}}{4K}\right)^{K/2} \geq m + 1$, which leads to the following condition on K:

$$K \geq \frac{2\log(m+1)}{\log\left(1 + \frac{\alpha^2 d_{\min}}{4K}\right)}. \tag{10}$$

Thus, if α and $d_{\min}$ are such that $\alpha^2 d_{\min} \asymp K$, then $K \asymp \log m$ measurements at each spatial location (i.e. of each spectrum) suffice to guarantee a given value of the pFDR *regardless of the spectral dimension,* N, where the proporotionality constants depend on the desired pFDR level. This observation suggests that if $d_{\min}$ is very small, which implies that the spectra in $\mathcal{D}$ are very similar to each other, then the signal-to-noise ratio (α) has to be high enough that any two spectra can be reliably distinguished. The dependence of K on $\log m$ implies that, as the number of possible target signals increases, the difficulty of detecting the presence of one target versus the others increases, and thus more measurements are needed to identify different targets. Fig. 1 illustrates how the pFDR bound changes as a function of K, α^2 and $d_{\min}$ for given values of N and m. Our simulation studies showed that the number of targets do not seem to affect the bounds as much as α^2 and $d_{\min}$ does. Thus the pFDR bound in (9) brings out interesting relationships between various quantities such as K, m and $d_{\min}$.

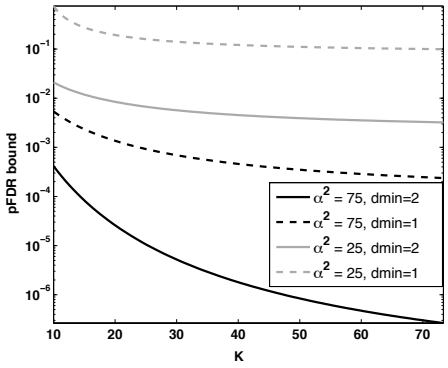

Fig. 1. Plot of the pFDR bound as a function of K, α^2 and $d_{\min}$ for a fixed $N = 220$ and $m = 10$. The plots in black correspond to $\alpha^2 = 75$ and the ones in gray correspond to $\alpha^2 = 25$ respectively. The dotted curves correspond to $d_{\min} = 1$ and the solid lines correspond to $d_{\min} = 2$ respectively. From the plots it is clear that the bound worsens as the signal-to-noise ratio decreases, and as the ambiguity among different spectra in $\mathcal{D}$ increases.

4. CONCLUSION

In this work we proposed a way to perform hyperspectral target detection from incoherent projections and derived worst-case theoretical bounds on the detector performance. Our future work on this topic will involve extending the analysis to the case when the SNR may depend on the spatial location, and incorporating other background models in our framework. Also, we will investigate other ways to construct the measurement matrix such that its entries are nonnegative, which will make it practical to implement in a real system.

APPENDIX

Proof of Lemma 1: To ensure positive definiteness of $\boldsymbol{B}$, we must have

$$\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{x} - \boldsymbol{x}^T\left(\boldsymbol{A}\boldsymbol{\Sigma_b}\boldsymbol{A}^T\right)\boldsymbol{x} > 0 \tag{11}$$

for any nonzero $\boldsymbol{x} \in \mathbb{R}^K$. Note that since $\boldsymbol{\Sigma_b}$ is positive semidefinite, $\boldsymbol{x}^T\left(\boldsymbol{A}\boldsymbol{\Sigma_b}\boldsymbol{A}^T\right)\boldsymbol{x} \geq 0$. However, the right hand side of (11) is > 0 only if the spectral norm of $\boldsymbol{A}\boldsymbol{\Sigma_b}\boldsymbol{A}^T$ is < 1, since $\boldsymbol{x}^T\left(\boldsymbol{A}\boldsymbol{\Sigma_b}\boldsymbol{A}^T\right)\boldsymbol{x} \leq \|\boldsymbol{x}\|^2 \cdot \|\boldsymbol{A}\boldsymbol{\Sigma_b}\boldsymbol{A}^T\|$. The norm of $\boldsymbol{A}\boldsymbol{\Sigma_b}\boldsymbol{A}^T$ is in turn bounded above by

$$\begin{aligned}\|\boldsymbol{A}\boldsymbol{\Sigma_b}\boldsymbol{A}^T\| &\leq \|\boldsymbol{A}\|\|\boldsymbol{\Sigma_b}\|\|\boldsymbol{A}^T\| \\ &= \|\boldsymbol{A}\|^2\|\boldsymbol{\Sigma_b}\| = \|\boldsymbol{A}\|^2\lambda_{\max}\end{aligned} \tag{12}$$

since $\|\boldsymbol{A}\| = \|\boldsymbol{A}^T\|$ and $\|\boldsymbol{\Sigma_b}\| = \lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{\Sigma_b}$. To ensure $\|\boldsymbol{A}\boldsymbol{\Sigma_b}\boldsymbol{A}^T\| < 1$, the right hand side of (12) has to be < 1, which leads to the result of Lemma 1.

Proof of Lemma 2: By construction, the entries of $\boldsymbol{A}$ are drawn from $\mathcal{N}(0, 1/K)$. Vershynin [12] has shown that if $\boldsymbol{G}$ is a $K \times N$ random matrix with i.i.d. zero mean subgaussian entries with unit variance, then, with probability at least $1 - \exp(-c(K+N))$, the following holds:

$$\|\boldsymbol{G}\| \leq c\left(\sqrt{K} + \sqrt{N}\right)$$

where c is an absolute constant. Applying this result to $\boldsymbol{A}$ yields $\|\boldsymbol{A}\| \leq c\left(\sqrt{N/K} + 1\right)$ with high probability. This result, together with the result of Lemma 1 yields Lemma 2.

Proof of Theorem 1: The proof of Theorem 1 follows by writing out the Bayesian interpretation of pFDR in (8) and bounding the numerator and the denominator terms. Using Bayes' rule, (8) can be written as

$$\mathrm{pFDR}(\ell) = \frac{\mathbb{P}\left(\boldsymbol{f}^* = \boldsymbol{f}^{(\ell)}\right)\mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(\ell)} | \boldsymbol{f}^* = \boldsymbol{f}^{(\ell)}\right)}{\mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(\ell)}\right)}. \quad (13)$$

If each spectrum in $\mathcal{D}$ is equally likely, then $\mathbb{P}\left(\boldsymbol{f}^* = \boldsymbol{f}^{(\ell)}\right) = 1/m$ and $\mathbb{P}\left(\boldsymbol{f}^* \neq \boldsymbol{f}^{(\ell)}\right) = 1 - 1/m$. Haupt, et al. [8] have shown that the probability of error incurred in classifying a signal $\boldsymbol{f}^* \in \mathbb{R}^N$ to one of m different classes from a known dictionary $\mathcal{D}$, given $K < N$ measurements of the form $\boldsymbol{y} = \alpha \boldsymbol{A}\boldsymbol{f}^* + \boldsymbol{n}$ and the knowledge of the measurement matrix $\boldsymbol{A}$, can be upper bounded by

$$\begin{aligned}\mathrm{P_e} \triangleq \mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^*\right) &= \mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(j)} | \boldsymbol{f}^* = \boldsymbol{f}^{(j)}\right) \\ &\leq (m-1)\left(1 + \frac{\alpha^2 d_{\min}\sigma_{\boldsymbol{A}}^2}{4}\right)^{-K/2} \quad (14)\end{aligned}$$

for a fixed nonrandom $\boldsymbol{f}^* = \boldsymbol{f}^{(j)} \in \mathcal{D}$, where $\widehat{\boldsymbol{f}}$ is estimated according to (6), the entries of $\boldsymbol{A}$ are i.i.d. draws from $\mathcal{N}(0, \sigma_{\boldsymbol{A}}^2)$, and $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, I)$. In their setup, $\sigma_{\boldsymbol{A}}^2 = 1/N$, whereas in our case $\sigma_{\boldsymbol{A}}^2 = 1/K$. The numerator in (13) can be upper bounded by relating it to the probability of error in a classification setting as follows:

$$\mathbb{P}\left(\boldsymbol{f}^* = \boldsymbol{f}^{(\ell)}\right)\mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(\ell)} | \boldsymbol{f}^* = \boldsymbol{f}^{(\ell)}\right) = \frac{1}{m}\mathrm{P_e}. \quad (15)$$

The denominator in (13) can be written as

$$\begin{aligned}\mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(\ell)}\right) &= \mathbb{P}\left(\boldsymbol{f}^* = \boldsymbol{f}^{(\ell)}\right)\mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(\ell)} | \boldsymbol{f}^* = \boldsymbol{f}^{(\ell)}\right) \\ &\quad + \mathbb{P}\left(\boldsymbol{f}^* \neq \boldsymbol{f}^{(\ell)}\right)\mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(\ell)} | \boldsymbol{f}^* \neq \boldsymbol{f}^{(\ell)}\right), \quad (16)\end{aligned}$$

and can be bounded by considering each of its terms separately. Observe that $\mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(\ell)} | \boldsymbol{f}^* = \boldsymbol{f}^{(\ell)}\right)$ is nonnegative, and

$$\begin{aligned}&\mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(\ell)} | \boldsymbol{f}^* \neq \boldsymbol{f}^{(\ell)}\right) = \mathbb{P}\left(\widehat{\boldsymbol{f}} \in \mathcal{D} \backslash \boldsymbol{f}^{(\ell)} | \boldsymbol{f}^* \neq \boldsymbol{f}^{(\ell)}\right) \\ &\geq \mathbb{P}\left(\widehat{\boldsymbol{f}} = \boldsymbol{f}^* | \boldsymbol{f}^* \neq \boldsymbol{f}^{(\ell)}\right) = 1 - \mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^* | \boldsymbol{f}^* \neq \boldsymbol{f}^{(\ell)}\right) \\ &= 1 - \frac{\mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^*, \boldsymbol{f}^* \neq \boldsymbol{f}^{(\ell)}\right)}{\mathbb{P}\left(\boldsymbol{f}^* \neq \boldsymbol{f}^{(\ell)}\right)} \geq 1 - \frac{\mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^*\right)}{1 - 1/m} \\ &= 1 - \frac{\mathrm{P_e}}{1 - 1/m} \quad (17)\end{aligned}$$

Substituting (17) in (16) one can see that

$$\mathbb{P}\left(\widehat{\boldsymbol{f}} \neq \boldsymbol{f}^{(\ell)}\right) \geq (1 - 1/m)\left(1 - \frac{\mathrm{P_e}}{1 - 1/m}\right) \quad (18)$$

The result of Theorem 1 follows directly from (14), (18) and (15).

5. REFERENCES

[1] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Info. Th.*, vol. 52, no. 12, 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Th.*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express*, vol. 15, no. 21, pp. 14013–14027, 2007.

[4] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," in *SPIE*, 2006, vol. 6065.

[5] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Sparse signal detection from incoherent projections," in *ICASSP 2006*, May 2006, vol. 3, pp. III–III.

[6] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proceedings of SPIE*, 2007, vol. 6498, p. 64980H.

[7] J. Haupt, R. Castro, R. Nowak, G. Fudge, and A. Yeh, "Compressive sampling for signal classification," in *Fortieth Asilomar Conference on Signals, Systems and Computers*, Nov 2006, pp. 1430–1434.

[8] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29–43, 2002.

[9] R. G. Baraniuk, M. Davenport, R. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[10] J. D. Storey, "The positive false discovery rate: A bayesian interpretation and the q-value," *Annals of Statistics*, pp. 2013–2035, 2003.

[11] S. Dasgupta and A. Gupta, "An elementary proof of a theorem of Johnson and Lindenstrauss," *Random Struct. Alg.*, vol. 22, no. 1, pp. 60–65, 2003.

[12] R. Vershynin, "Lecture notes on non-asymptotic random matrix theory," 2006-2007.