

HYPERSPECTRAL TARGET DETECTION FROM INCOHERENT PROJECTIONS: NONEQUIPROBABLE TARGETS AND INHOMOGENEOUS SNR

Kalyani Krishnamurthy, Maxim Raginsky and Rebecca Willett

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

ABSTRACT

This paper describes a computationally efficient approach for the detection of spectral targets of different strengths, contaminated by a colored Gaussian background, from relatively few incoherent projections compared to the dimension of the target. The performance of the detector is analyzed with respect to the number of observations collected, the background perturbation, the target signal strength, the similarity among different targets in a known spectral dictionary, and the known *a priori* probabilities of the targets.

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

1. HYPERSPECTRAL TARGET DETECTION

Hyperspectral imaging combines spatial and spectral information, and thus enables one to characterize the material composition of the underlying scene. It is thus very useful in applications like remote sensing and astronomical imaging. However, despite its usefulness, collecting hyperspectral images with M spatial locations, and N spectral bands typically requires MN sensors or prolonged sensing time.

Recent advances in compressive sensing have shown that it is possible to accurately reconstruct a sparse signal from a few incoherent projections relative to the signal dimension [1]. Researchers have built practical hyperspectral cameras that acquire fewer compressive measurements than hyperspectral voxels, and developed algorithms that reconstruct the underlying hyperspectral data from such measurements [2]. In our previous work, we investigated how such systems can be applied in the context of hyperspectral target detection depending on (a) the type and the number of measurements one collects, (b) the background perturbation that the system can tolerate, and (c) the similarity among different *equally likely* targets of interest, assuming that *the signal strength at every spatial location is the same* [3]. In this paper, we work in a more realistic setting where we allow the signal strengths to vary spatially, and consider targets with nonuniform *a priori* probabilities. We also demonstrate the performance of the proposed approach on simulated and real data.

1.1. Previous investigations

Several researchers have exploited compressive sensing to perform target detection and classification without reconstructing the underlying signal [4, 5]. In [4], the authors develop a classification algorithm called the *smashed filter* to classify an image in \mathbb{R}^N to one of m known classes from K incoherent projections of the image, where $K < N$. The underlying image is assumed to lie on a

low-dimensional manifold, and the algorithm finds the closest match from the m known classes by performing a nearest neighbor search over the m different manifolds. The incoherent projections are chosen to preserve the distances among the manifolds. Though [4] offers theoretical bounds on the number of measurements necessary to preserve distances among different manifolds, it is not clear how the performance scales with K , and how to incorporate background models into this setup. Moreover, this approach may be computationally intensive since it involves searching over different manifolds. Haupt, et al. use a nearest-neighbor classifier to classify an N -dimensional signal to one of m equally likely target classes from $K < N$ random projections, and provide theoretical guarantees on the detector performance [5]. This approach is computationally efficient, but it is nontrivial to extend it to the case of hyperspectral target detection with colored background noise and nonequiprobable targets. Furthermore, they only bound the probability of error, whereas false alarm performance is a more crucial metric in many applications.

2. PROBLEM SETUP

We are interested in detecting one spectral target from a spectral dictionary \mathcal{D} of known targets by measuring incoherent hyperspectral projections of the form

$$\mathbf{z}_i = \Phi(\alpha_i \mathbf{f}_i^* + \mathbf{b}_i) + \mathbf{n}_i \quad (1)$$

where

- $i \in \{1, \dots, M\}$ indexes the spatial locations of the hyperspectral data,
- $\alpha_i \in \mathbb{R}$ is a measure of the signal-to-noise ratio (SNR) at location i and is known or can be estimated,
- $\mathbf{f}_i^* \in \mathbb{R}^N$ is one of the m possible unit-norm targets from a known spectral dictionary $\mathcal{D} = \{\mathbf{f}^{(1)}, \mathbf{f}^{(2)}, \dots, \mathbf{f}^{(m)}\}$,
- $\Phi \in \mathbb{R}^{K \times N}$ is a random measurement matrix with $K < N$,
- $\mathbf{b}_i \in \mathbb{R}^N \sim \mathcal{N}(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$ is the background noise vector, whose statistics are either assumed to be known or estimated from training data, and
- $\mathbf{n}_i \in \mathbb{R}^K \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is the i.i.d. sensor noise.

We assume that Φ , \mathbf{b} and \mathbf{n} are independent of each other, and the prior probabilities of different targets in the spectral dictionary are known in advance. If these probabilities are unknown, then each target can be considered equally likely. Given this setup, our goal is to develop a suitable target detection approach, and provide theoretical guarantees on its performance.

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.

3. TARGET DETECTION APPROACH

The observation model in (1) can be written as $z_i = \alpha_i \Phi \mathbf{f}_i^* + \mathbf{n}'_i$ where $\mathbf{n}'_i = \Phi \mathbf{b}_i + \mathbf{n}_i \sim \mathcal{N}(\Phi \boldsymbol{\mu}_b, \Phi \boldsymbol{\Sigma}_b \Phi^T + \sigma^2 \mathbf{I})$ is the noise term that combines the effects of background and sensor noise. We can simplify this noise model through background mean subtraction and whitening to obtain

$$\mathbf{y}_i = C_\Phi (z_i - \Phi \boldsymbol{\mu}_b) = \alpha_i C_\Phi \Phi \mathbf{f}_i^* + \mathbf{w}_i \quad (2)$$

for $i \in \{1, \dots, M\}$, where $C_\Phi = (\Phi \boldsymbol{\Sigma}_b \Phi^T + \sigma^2 \mathbf{I})^{-1/2}$ is the whitening filter and $\mathbf{w}_i = C_\Phi \mathbf{n}'_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is white Gaussian noise. In our previous work [3] we showed that if $\Phi = \sigma \mathbf{B}^{-1/2} \mathbf{A}$, where the entries of \mathbf{A} are i.i.d. draws from $\mathcal{N}(0, 1/K)$ and $\mathbf{B} = \mathbf{I} - \mathbf{A} \boldsymbol{\Sigma}_b \mathbf{A}^T$ is a positive definite matrix, then $C_\Phi \Phi = \mathbf{A}$. (We assume \mathbf{A} is zero mean. While elements of Φ must be nonnegative in a linear optical system, in practice mean shifting can be used to preprocess the data and circumvent this difficulty whenever the SNR is sufficiently high and signal independent.) This particular choice of \mathbf{A} ensures that, with high probability, the distance between any two vectors in \mathbb{R}^N is preserved by this projection operator [6]. We have shown in [3] that, provided $\lambda_{\max} < \frac{1}{\|\mathbf{A}\|^2}$, where λ_{\max} is the spectral norm of $\boldsymbol{\Sigma}_b$ and $\|\mathbf{A}\|$ is the spectral norm of \mathbf{A} , then, with high probability, \mathbf{B} is positive definite. This condition gives an upper bound on the tolerable background perturbation in relation to the spectral norm of \mathbf{A} , which in turn varies with N and K . For a fixed N , the norm of \mathbf{A} decreases with K which implies that the system can tolerate more background if one collects more measurements.

Let $j \in \{1, \dots, m\}$ be the index of the target of interest. Let us cast the hyperspectral target detection problem as a multiple hypothesis testing problem of the form

$$\mathcal{H}_{0i}^{(j)} : \mathbf{f}_i^* = \mathbf{f}^{(j)} \quad \text{and} \quad \mathcal{H}_{1i}^{(j)} : \mathbf{f}_i^* \neq \mathbf{f}^{(j)} \quad (3)$$

for $i = 1, \dots, M$. We define our decision rule corresponding to target $\mathbf{f}^{(j)} \in \mathcal{D}$ in terms of a set of significance regions $\Gamma_i^{(j)}$ defined according to

$$\Gamma_i^{(j)} = \left\{ \mathbf{y} : \log \mathbb{P} \left(\mathbf{f}_i^* = \mathbf{f}^{(j)} \mid \mathbf{y}, \alpha_i, \mathbf{A} \right) < \log \mathbb{P} \left(\mathbf{f}_i^* = \mathbf{f}^{(\ell)} \mid \mathbf{y}, \alpha_i, \mathbf{A} \right) \text{ for some } \ell \in \{1, \dots, m\} \right\},$$

where

$$\log \mathbb{P} \left(\mathbf{f}_i^* = \mathbf{f}^{(j)} \mid \mathbf{y}, \alpha_i, \mathbf{A} \right) \propto -\frac{\|\mathbf{y}_i - \alpha_i \mathbf{A} \mathbf{f}^{(j)}\|^2}{2} + \log p^{(j)}$$

is the log of the a posteriori probability of the target $\mathbf{f}^{(j)}$ at the i^{th} spatial location given \mathbf{y} , α_i and \mathbf{A} . The decision rule can be formally expressed in terms of the significance regions as follows:

$$\text{declare } \mathcal{H}_{1i}^{(j)} \text{ if the test statistic } \mathbf{y}_i \in \Gamma_i^{(j)}. \quad (4)$$

Note that these hypothesis tests are independent of each other, but are nonidentical since the decision rule depends on $\{\alpha_i\}$ that are different at every spatial location.

In [7], Storey introduced the positive False Discovery Rate (pFDR) criterion to account for the errors encountered in testing multiple, independent and *identical* hypothesis tests simultaneously. Controlling the false discovery rate is a useful alternative to controlling the false alarm rate in multiple hypothesis tests because it reflects the fraction of incorrect declarations of $\mathcal{H}_{1i}^{(j)}$ among the total number of *declarations* instead of among the total number of *tests*.

Formally, the pFDR is defined to be the expected ratio of the number of false rejections of null hypotheses to the total number of rejected null hypotheses, subject to the condition that one rejects at least one null hypothesis. In our case, this pFDR error measure can be suitably modified to accommodate the testing of multiple, *nonidentical* hypothesis tests simultaneously as shown below. Let $\boldsymbol{\Gamma} \triangleq \{\Gamma_i^{(j)}\}$; then

$$\text{pFDR}^{(j)}(\boldsymbol{\Gamma}) = \mathbb{E} \left[\frac{V(\boldsymbol{\Gamma})}{R(\boldsymbol{\Gamma})} \mid R(\boldsymbol{\Gamma}) > 0 \right] \quad (5)$$

where $V(\boldsymbol{\Gamma}) \triangleq \sum_{i=1}^M \mathbb{I}_{\{\mathbf{y}_i \in \Gamma_i^{(j)}\}} \mathbb{I}_{\{\mathcal{H}_{0i}^{(j)}\}}$, in our case, is the number of missed targets and $R(\boldsymbol{\Gamma}) \triangleq \sum_{i=1}^M \mathbb{I}_{\{\mathbf{y}_i \in \Gamma_i^{(j)}\}}$ is the total number of spectra declared to be nontargets. The theorem below presents our main result.

Theorem 1: Suppose that one performs multiple, independent, nonidentical hypothesis tests of the form (3) and decides according to (4). As long as the background covariance matrix satisfies the condition $\lambda_{\max} < 1/\|\mathbf{A}\|^2$, the worst-case pFDR satisfies the following bound:

$$\text{pFDR}^{(\max)} \leq \frac{p_{\max}}{p_{\min}} \left(\frac{1 - p_{\max}}{1 - p_{\min}} \left(1 + \frac{\alpha_{\min}^2 d_{\min}}{4K} \right)^{\frac{K}{2}} - \frac{1}{p_{\min}} \right)^{-1}$$

where $\text{pFDR}^{(\max)} = \max_{j \in \{1, \dots, m\}} \text{pFDR}^{(j)}(\boldsymbol{\Gamma})$, $\alpha_{\min} = \min_{i \in \{1, \dots, M\}} \alpha_i$, $d_{\min} = \min_{\mathbf{f}^{(i)}, \mathbf{f}^{(j)} \in \mathcal{D}, i \neq j} \|\mathbf{f}^{(i)} - \mathbf{f}^{(j)}\|^2$, and p_{\max} and p_{\min} are the maximum and the minimum values of the a priori target probabilities respectively.

The proof of this theorem is detailed in the appendix. This bound is nonnegative and less than one as long as

$$K > 2 \log \left(\frac{(1 - p_{\min})}{p_{\min}(1 - p_{\max})} \right) / \log \left(1 + \frac{\alpha_{\min}^2 d_{\min}}{4K} \right). \quad (6)$$

For a fixed p_{\max} , p_{\min} , α_{\min} and d_{\min} , this bound enables one to choose a value of K to guarantee a desired pFDR value. The bound on K increases logarithmically with the increase in the difference between p_{\max} and p_{\min} . This is to be expected since one would need more measurements to detect a less probable target as our decision rule weights each target by its a priori probability. The dependence of the bound on K on $\alpha_{\min}^2 d_{\min}$ illustrates the interplay between the signal strength of the targets, the similarity among different targets in \mathcal{D} and the number of measurements collected. A small value of d_{\min} suggests that the targets in \mathcal{D} are very similar to each other and thus α_{\min} and K need to be high enough such that similar spectral targets can be identified.

4. EXPERIMENTAL RESULTS

4.1. Simulated Data

To test the effectiveness of our approach, we generated a dictionary of 9 spectra corresponding to trees, grass, water bodies and roads, obtained from a set of labeled HyMap data (described in Sec. 4.2), and used these to generate background-contaminated compressive measurements for different values of K . In this experiment, we constructed simulated data according to $\mathbf{g}_i = \alpha_i \mathbf{f}_i + \mathbf{b}_i$ for $i = 1, \dots, N_s$, such that the background statistics and $\{\alpha_i\}$ are known exactly. For this data, $N_s = 2025$, $p_{\min} = 1.24 \times 10^{-2}$, $p_{\max} = 3.09 \times 10^{-1}$, $d_{\min} = 1.89 \times 10^{-3}$ and $\alpha_{\min} = 1.65 \times 10^2$. The condition on the spectral norm of $\boldsymbol{\Sigma}_b$ in Theorem 1 suggests that

the proposed method is provably effective only when the background is sufficiently weak. For this simulation, however, we consider a setting where the background is stronger than the theoretical limit and are forced to use a scaled version of the background covariance matrix to form the measurement matrix Φ . (The strong background is also an issue with the real data described in the next subsection, so our treatment of the strong background here helps validate our approach to the real data.) We scale Σ_b by projecting it onto the set of covariance matrices \mathcal{C} that satisfy the condition and finding a closest match $\tilde{\Sigma}_b \in \mathcal{C}$ to Σ_b by iteratively shrinking the leading eigenvalues of $\tilde{\Sigma}_b$ until the positive definiteness condition is satisfied.

We generate the compressive measurements of this data according to (1) and transform them by a series of operations to arrive at a model of the form discussed in (2). Given the observations of the form (2), we apply our compressive MAP classifier to the simulated data and evaluate its performance for different values of K . We compare the performance of our compressive classifier to a nearest neighbor classifier obtained by finding the closest fit to each spectrum in \mathbf{g} from the dictionary of possible targets \mathcal{D} . In other words, if \mathbf{g}_i is the i^{th} spectrum, then its nearest neighbor class label L_i^{NN} is determined according to $L_i^{NN} = \arg \min_{\ell \in \{1, \dots, m\}} \|\mathbf{g}_i - \boldsymbol{\mu}_b - \alpha_i \mathbf{f}^{(\ell)}\|^2$. The compressive classifier class label L_i^{CC} for the i^{th} observed spectrum \mathbf{y}_i is determined according to $L_i^{CC} = \arg \min_{\ell \in \{1, \dots, m\}} \left(\frac{1}{2} \|\mathbf{y}_i - \alpha_i \mathbf{A} \mathbf{f}^{(\ell)}\|^2 - \log p^{(\ell)} \right)$ where $p^{(\ell)}$ is the a priori probability of target class ℓ . If the signal-to-noise ratio α_i is not known, it can be estimated from the observed data \mathbf{y}_i according to $\hat{\alpha}_i = \sqrt{\|\mathbf{y}_i\|^2 - K}$ since the sensing matrix \mathbf{A} and sensor noise \mathbf{w} are statistically independent, and are known exactly [5]. In our experiments we evaluate the performance of our classifier in both of these cases. The empirical pFDR $^{(j)}$ curve for each target spectrum j is calculated as follows: $\text{pFDR}^{(j)} = \frac{\sum_{i=1}^{N_s} \mathbb{I}_{\{L_i^{NN}=j\}} \mathbb{I}_{\{L_i^{CC} \neq j\}}}{\sum_{i=1}^{N_s} \mathbb{I}_{\{L_i^{CC} \neq j\}}}$. The empirical pFDR $^{(c)}$ is the ratio of the number of missed targets to the total number of spectra that were declared to be nontargets. The plots in Fig. 1(a) show the results obtained using our target detection approach when (a) $\{\alpha_i\}$ are known exactly and (b) $\{\alpha_i\}$ are estimated from the observations \mathbf{y} . In both these cases, the empirical pFDR curves lie below the theoretical bound and decay with the increase in the values of K .

4.2. HyMap Data

We further evaluate the performance of our target detection approach on a HyMap data set [8] that consists of 7962 labeled spectra corresponding to trees, grass, water bodies and roads, split into nonoverlapping training and validation sets, \mathbf{g}^t and \mathbf{g}^v respectively. We estimate the a priori target probabilities and background statistics from the training data, and test our target detection approach on the validation data. Our target detection approach discussed in Sec. 3 is built on the assumptions that the spectral dictionary \mathcal{D} , the signal-to-noise ratios $\{\alpha_i\}$ and the background statistics are known accurately. However, with the real HyMap data, \mathcal{D} , the background statistics, and $\{\alpha_i\}$ are unknown, and need to be estimated from the available labeled data \mathbf{g}^t and \mathbf{g}^v . We form a spectral dictionary by averaging the training spectra in each class and normalizing the resulting mean spectrum to have unit norm. The a priori probability of each target class j is estimated from the HyMap training data, and for this HyMap data, $p_{\max} = 3.42 \times 10^{-1}$, and $p_{\min} = 1.51 \times 10^{-2}$.

Estimation of the background statistics: If the signal-to-noise ratio of each training spectrum is known, then the background statistics can be estimated from the available train-

ing data and the unit-norm targets in the spectral dictionary as follows: $\hat{\boldsymbol{\mu}}_b = \frac{1}{N_t} \sum_{\ell=1}^m \sum_{i=1}^{n_\ell} \hat{\mathbf{b}}_i^{(\ell)}$ and $\hat{\Sigma}_b = \frac{1}{N_t} \sum_{\ell=1}^m \sum_{i=1}^{n_\ell} \left(\hat{\mathbf{b}}_i^{(\ell)} - \hat{\boldsymbol{\mu}}_b \right) \left(\hat{\mathbf{b}}_i^{(\ell)} - \hat{\boldsymbol{\mu}}_b \right)^T$ where N_t is the number of training spectra in \mathbf{g}^t , n_ℓ is the number of training spectra in the ℓ^{th} class, $\hat{\mathbf{b}}_i^{(\ell)} = \mathbf{g}_i^{t_\ell} - \alpha_i^{(\ell)} \mathbf{f}^{(\ell)}$, $\mathbf{g}_i^{t_\ell}$ is the i^{th} training spectrum in the ℓ^{th} class, and $\mathbf{f}^{(\ell)} \in \mathcal{D}$. Since the signal-to-noise ratios are unknown, the $\alpha_i^{(\ell)}$ in $\hat{\mathbf{b}}_i^{(\ell)}$ have to be replaced by suitable estimates. We set $\hat{\alpha}_i^{(\ell)} = \|\mathbf{g}_i^{t_\ell}\|_2$ for $\ell = 1, \dots, m$ and $i = 1, \dots, n_\ell$. (However, this estimate may be inaccurate if the background is very strong.) The estimate Σ_b did not satisfy the positive definiteness condition in Theorem 1 and we used a scaled version of it in our experiments, as discussed in Sec. 4.1.

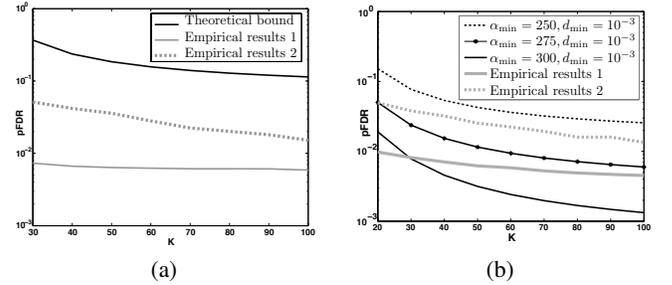


Fig. 1. Comparison of the worst-case empirical pFDR curves with the theoretical bounds. (a) Results corresponding to the simulated data. (b) Results corresponding to the real HyMap data. The plots in black are the theoretical bounds, the plots in gray are the empirical pFDR curves obtained when $\{\alpha_i\}$ were known exactly, or estimated from the validation spectra in the case of real HyMap data (solid plot, denoted by empirical results 1), and $\{\alpha_i\}$ were estimated from the observations \mathbf{y} (dashed plot, denoted by empirical results 2). In both these experiments the values of K were chosen according to (6).

We generated the observations \mathbf{y} from the validation data according to (1), transformed them according to (2), and compared the performance of our compressive classifier to that of the nearest neighbor classifier as discussed in Sec. 4.1. In the plots shown in Fig. 1(b), we compare the worst-case empirical positive false discovery ratio given by $\text{pFDR}^{(\max)} = \max_{j \in \{1, \dots, m\}} \text{pFDR}^{(j)}$ to the theoretical bound derived in the previous section. These plots are obtained by running independent experiments over 2000 different noise and sensing matrix realizations for different values of K , and averaging the pFDR values obtained in each iteration. In order to compare the empirical pFDR values to the theoretical bounds, one needs accurate knowledge of α_{\min} and d_{\min} to compute the bounds for different values of K . In this experiment, however, our knowledge of α_{\min} and d_{\min} is not accurate for two reasons: 1) We form the spectral dictionary by averaging the (quite diverse) training spectra from each target class that were obtained under different illumination, atmospheric, and spectral mixing conditions. As a result, the d_{\min} that we compute from this dictionary might not reflect the similarity among different targets in the validation data set. In practice, this issue could be resolved with larger spectral dictionaries which account for uncertainty in the target spectra. 2) Estimation errors for the background statistics propagate to errors in the estimation of α_{\min} . In practice, this issue could be circumvented by conducting separate tests to accurately estimate background statistics. Thus we plot the theoretical bound for different values of $\alpha_{\min}^2 d_{\min}$ and compare them with the $\text{pFDR}^{(\max)}$ curves obtained by estimat-

ing the values of α_i 's from (1) the validation spectra, and (2) the observations \mathbf{y} . The plots show that the theoretical bounds and the empirical pFDR curves decay at a similar rate as K increases.

5. CONCLUSION

Previous studies, notably [4], have exploited the distance preservation properties of compressive or dimension reducing projections in the context of target detection. In this paper, we derive explicit performance bounds for a special case of such a detector in the presence of background contamination, nonequiplausible targets, and spatially varying signal strengths. The theoretical contributions of this paper clearly illustrate the target detection impact of compressive measurements. While our bounds are sensitive to the values of some constants (such as d_{\min} and α_{\min}), they provide insight into how the number of compressive measurements should scale with the noise level, target probabilities, and target diversity to maintain a constant false discovery rate tolerance.

Our bounds do not reflect spatial correlations in the spectra. While this is an important avenue for future work, we note that many hyperspectral imagers have relatively low spatial resolution (and hence low spatial correlations).

APPENDIX

Proof of Theorem 1: The proof of Theorem 1 adapts the proof techniques from [7] to nonidentical independent hypothesis tests. To keep our notations simple, we drop the j superscripts on the set of significance regions and the hypothesis tests though they are a function of the target of interest j . We begin by expanding the pFDR definition in (5) as follows:

$$\text{pFDR}^{(j)}(\{\Gamma_i\}) = \sum_{k=1}^M \mathbb{E} \left[\frac{V(\{\Gamma_i\})}{R(\{\Gamma_i\})} \middle| R(\{\Gamma_i\}) = k \right] \times \mathbb{P}(R(\{\Gamma_i\}) = k | R(\{\Gamma_i\}) > 0)$$

Observe that $R(\{\Gamma_i\}) = k$ implies that there exists some subset $S_k = \{u_1, \dots, u_k\} \subseteq \{1, \dots, M\}$ of size k such that $\mathbf{y}_{u_i} \in \Gamma_{u_i}$ for $i = 1, \dots, k$. To simplify the notation, let $\Lambda_{S_k} = \prod_{u \in S_k} \Gamma_u \times \prod_{\ell \notin S_k} \Gamma_\ell$ denote the significance region that corresponds to set S_k , and $\mathbf{T} = (\mathbf{y}_1, \dots, \mathbf{y}_M)$ be a set of test statistics corresponding to each hypothesis test. Considering all such subsets we have

$$\text{pFDR}^{(j)}(\{\Gamma_i\}) = \sum_{k=1}^M \sum_{S_k} \mathbb{E} \left[\frac{V(\{\Gamma_i\})}{k} \middle| \mathbf{T} \in \Lambda_{S_k} \right] \mathbb{P}(\mathbf{T} \in \Lambda_{S_k} | R(\{\Gamma_i\}) > 0). \quad (7)$$

By plugging in the definition of $V(\{\Gamma_i\})$, we have

$$\begin{aligned} \mathbb{E}[V(\{\Gamma_i\}) | \mathbf{T} \in \Lambda_{S_k}] &= \mathbb{E} \left[\sum_{i=1}^M \mathbb{I}_{\{\mathbf{y}_i \in \Gamma_i\}} \mathbb{I}_{\{\mathcal{H}_i=0\}} \middle| \mathbf{T} \in \Lambda_{S_k} \right] \\ &= \sum_{p=1}^k \mathbb{E} \left[\mathbb{I}_{\{\mathcal{H}_{u_p}=0\}} \middle| \mathbf{y}_{u_p} \right] = \sum_{p=1}^k \mathbb{P}(\mathcal{H}_{u_p} = 0 | \mathbf{y}_{u_p} \in \Gamma_{u_p}) \end{aligned}$$

for all $u_p \in S_k$ since the tests are independent of each other. Storey has shown that if we perform only one hypothesis test, then the pFDR can be written in terms of simple posterior probability $\mathbb{P}(\mathcal{H}_0 | \mathbf{y} \in \Gamma)$ [7]. We refer to this posterior probability for the i^{th} hypothesis test by pFDR_i . In our previous work we showed that

the pFDR for i^{th} hypothesis test can be expanded using Bayes' rule and upper bounded in terms of the a priori probabilities of the targets of interest and the probability of classification error of the form $(P_e)_i = \mathbb{P}(\hat{\mathbf{f}}_i \neq \mathbf{f}_i^*)$ as shown below

$$\text{pFDR}_i^{(j)} \leq \frac{p^{(j)}(P_e)_i}{(1-p^{(j)}) \left(1 - \frac{(P_e)_i}{1-p^{(j)}}\right)}. \quad (8)$$

In [5] the authors bound the classification error $(P_e)_i$ incurred in classifying an observed signal of the form $\mathbf{y}_i = \alpha_i \mathbf{A} \mathbf{f}_i^* + \mathbf{w}_i$ discussed in Sec. 3 to one of m equally likely targets. It can be shown that their proof can be readily extended to the case of targets with non uniform a priori probabilities and that

$$(P_e)_i \leq \frac{1-p_{\min}}{p_{\min}} \left(1 + \frac{\alpha_i^2 d_{\min}}{4K}\right)^{-K/2} \quad (9)$$

for a fixed nonrandom $\mathbf{f}_i^* = \mathbf{f}^{(j)} \in \mathcal{D}$, where $\hat{\mathbf{f}}_i = \arg \min_{\mathbf{f}^{(\ell)} \in \mathcal{D}} \mathbb{P}(\mathbf{f}_i^* = \mathbf{f}^{(\ell)} | \mathbf{y}_i, \alpha_i, \mathbf{A})$. Substituting (9) in (8), and upper bounding the $p^{(j)}$ in (8) by p_{\max} we have,

$$\text{pFDR}_i^{(j)} \leq \frac{p_{\max}}{p_{\min}} \left(\frac{1-p_{\max}}{1-p_{\min}} \left(1 + \frac{\alpha_{\min}^2 d_{\min}}{4K}\right)^{\frac{K}{2}} - \frac{1}{p_{\min}} \right)^{-1}$$

where $\alpha_{\min} = \min_{i \in \{1, \dots, M\}} \alpha_i$. The result of Theorem 1 is obtained by substituting the bound above in (7).

6. REFERENCES

- [1] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [2] 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.
- [3] K. Krishnamurthy, M. Raginsky, and R. Willett, "Hyperspectral target detection from incoherent projections," in *Proceedings of the 35th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010.
- [4] 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.
- [5] 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*, pp. 1430–1434, 2006.
- [6] 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.
- [7] J. D. Storey, "The positive false discovery rate: A Bayesian interpretation and the q-value," *Annals of Statistics*, pp. 2013–2035, 2003.
- [8] T. Cocks, R. Jenssen, A. Stewart, I. Wilson, and T. Shields, "The HyMapTM airborne hyperspectral sensor: the system, calibration and performance," in *Proceedings of the 1st EARSel workshop on Imaging Spectroscopy*, pp. 37–42, 1998.