WEAK SIGNAL DETECTION IN HYPERSPECTRAL IMAGERY USING SPARSE MATRIX TRANSFORM (SMT) COVARIANCE ESTIMATION

*Guangzhi Cao, Charles A. Bouman**

Purdue University
School of Electrical and Computer Engineering
West Lafayette, IN 47907

James Theiler[†]

Los Alamos National Laboratory
Space and Remote Sensing Sciences
Los Alamos, NM 87544

ABSTRACT

Many detection algorithms in hyperspectral image analysis, from well-characterized gaseous and solid targets to deliberately uncharacterized anomalies and anomalous changes, depend on accurately estimating the covariance matrix of the background. In practice, the background covariance is estimated from samples in the image, and imprecision in this estimate can lead to a loss of detection power.

In this paper, we describe the sparse matrix transform (SMT) and investigate its utility for estimating the covariance matrix from a limited number of samples. The SMT is formed by a product of pairwise coordinate (Givens) rotations. Experiments on hyperspectral data show that the estimate accurately reproduces even small eigenvalues and eigenvectors. In particular, we find that using the SMT to estimate the covariance matrix used in the adaptive matched filter leads to consistently higher signal-to-clutter ratios.

Index Terms— covariance matrix, hyperspectral imagery, matched filter, signal detection, sparse matrix transform

1. INTRODUCTION

The covariance matrix is the cornerstone of multivariate statistical analysis. From radar [1] and remote sensing [2] to high finance [3], algorithms for the detection and analysis of signals require the estimation of a covariance matrix, often as a way to characterize the background clutter. For applications – such as hyperspectral imagery – where the covariance matrix is large, the estimation of that matrix from a limited number of samples is especially challenging.

The sample covariance is the most natural estimator, but particularly when the number of samples m (*e.g.,* number of pixels) is not much larger than the dimension p (*e.g.,* number of spectral channels), this is not necessarily the best estimate. Sliding window [4] and segmentation methods [5] are just two

examples where only a small number of samples is available for each covariance matrix that needs to be estimated. To mitigate the effect of undersampling, various kinds of regularization have been proposed [6, 7, 8]. We will consider two recently developed regularizing schemes for covariance matrix estimation based on the sparse matrix transform (SMT) [9].

In Ref. [9], the effectiveness of the SMT estimator was demonstrated in terms of eigenvalues and Kullback-Leibler distances between Gaussian distributions based on true and approximate covariance matrices. In this paper, we investigate the performance of the adaptive matched filter, which depends on a covariance matrix estimate, when different regularizers are used. This work extends previous work by others investigating different approaches for regularizing the adaptive matched filter [4, 10, 11].

2. MAXIMUM LIKELIHOOD ESTIMATION

Given a p-dimensional Gaussian distribution with zero mean and covariance matrix $R \in \mathbb{R}^{p \times p}$, the likelihood of observing m samples, organized into a data matrix $X = [\mathbf{x}_1 \mathbf{x}_2 \ldots \mathbf{x}_m] \in \mathbb{R}^{p \times m}$, is given by

$$\mathcal{L}(R; X) = \frac{|R|^{-m/2}}{(2\pi)^{mp/2}} \exp\left[-\frac{1}{2}\text{trace}\left(X^T R^{-1} X\right)\right]. \quad (1)$$

If the covariance is decomposed as the product $R = E\Lambda E^T$ where E is the orthogonal eigenvector matrix and Λ is the diagonal matrix of eigenvalues, then one can jointly maximize the likelihood with respect to E and Λ, which results in the maximum likelihood (ML) estimates [9]

$$\hat{E} = \arg\min_{E \in \Omega} \left\{ \left| \text{diag}(E^T SE) \right| \right\} \quad (2)$$

$$\hat{\Lambda} = \text{diag}\left(\hat{E}^T SE\right), \quad (3)$$

where $S = \langle \mathbf{x}\mathbf{x}^T \rangle = \frac{1}{m} X X^T$ is the sample covariance, and Ω is the set of allowed orthogonal transforms. Then $\hat{R} = \hat{E}\hat{\Lambda}\hat{E}^T$ is the ML estimate of the covariance.

Note that if S has full rank and Ω is the set of all orthogonal matrices, then the ML estimate of the covariance is given by the sample covariance: $\hat{R} = S$.

2.1. Sparse Matrix Transform (SMT)

The Sparse Matrix Transform (SMT) provides as way to regularize the estimate of the covariance matrix by restricting the set Ω to a class of sparse eigenvector matrices E.

The most sparse nontrivial orthogonal transform is the *Givens rotation*, which corresponds to a rotation by an angle θ in the plane of the i and j axes; specifically, it is given by $E = I + \Theta(i, j, \theta)$ where

$$
\Theta(i,j,\theta)_{mn} = \begin{cases} \cos(\theta) - 1 & \text{if } m = n = i \text{ or } m = n = j \\ \sin(\theta) & \text{if } m = i \text{ and } n = j \\ -\sin(\theta) & \text{if } m = j \text{ and } n = i \\ 0 & \text{otherwise.} \end{cases}
\tag{4}
$$

Let E_k denote a Givens rotation, and note that a product of orthogonal rotations $E_k E_{k-1} \cdots E_1$ is still orthogonal. Let Ω_K be the set of orthogonal matrices that can be expressed as a product of K Givens rotations. The SMT covariance estimate is then given by Eq. (2) with $\Omega = \Omega_K$ and Eq. (3). (Actually, the effective Ω is more restrictive than this, since we do not optimize over *all* possible products of K rotations, but instead greedily choose Givens rotations one at a time.)

3. MATCHED FILTER CRITERION

One practical use for a covariance estimate is the detection of signals using a matched filter. This detector depends on the covariance matrix of the background data, and the better the covariance is estimated, the more effective the detector.

A filter $\mathbf{q} \in \mathbb{R}^p$ is a vector of coefficients which is applied to an observation $\mathbf{x}$ to give a scalar value $\mathbf{q}^T \mathbf{x}$ which is large when $\mathbf{x}$ contains signal $\mathbf{t}$ and is small otherwise. The signal-to-clutter ratio for a filter $\mathbf{q}$ is given by

$$
\text{SCR} = \frac{(\mathbf{q}^T \mathbf{t})^2}{\langle (\mathbf{q}^T \mathbf{x})^2 \rangle} = \frac{(\mathbf{q}^T \mathbf{t})^2}{\mathbf{q}^T \langle \mathbf{x} \mathbf{x}^T \rangle \mathbf{q}} = \frac{(\mathbf{q}^T \mathbf{t})^2}{\mathbf{q}^T R \mathbf{q}}
\tag{5}
$$

The *matched filter* is the vector that optimizes the SCR, and it is given, up to a constant multiplier, by $\mathbf{q} = R^{-1} \mathbf{t}$. Using this $\mathbf{q}$ in Eq. (5), we get the optimal SCR:

$$
\text{SCR}_o = \frac{(\mathbf{t}^T R^{-1} \mathbf{t})^2}{\mathbf{t}^T R^{-1} R R^{-1} \mathbf{t}} = \mathbf{t}^T R^{-1} \mathbf{t}.
\tag{6}
$$

If we approximate the matched filter using an approximate covariance matrix $\hat{R}$, then $\hat{\mathbf{q}} = \hat{R}^{-1} \mathbf{t}$ gives

$$
\text{SCR} = \frac{(\mathbf{t}^T \hat{R}^{-1} \mathbf{t})^2}{\mathbf{t}^T \hat{R}^{-1} R \hat{R}^{-1} \mathbf{t}}
\tag{7}
$$

and the SCRR is the ratio

$$
\frac{\text{SCR}}{\text{SCR}_o} = \frac{(\mathbf{t}^T \hat{R}^{-1} \mathbf{t})^2}{(\mathbf{t}^T \hat{R}^{-1} R \hat{R}^{-1} \mathbf{t})(\mathbf{t}^T R^{-1} \mathbf{t})}.
\tag{8}
$$

If $\hat{R} = R$, then SCRR $= 1$, but in general SCRR ≤ 1.

4. NUMERICAL EXPERIMENTS

The analysis in this paper is based on two hyperspectral datasets: one is a 224-channel AVIRIS image of the Florida coastline that was used in Ref. [12] and one is a 191-channel image of the Washington DC mall that was used in Ref. [2]. For the Florida image, we used all 75,000 of the pixels in the image to estimate the "true" covariance; for the Washington data, we limited ourselves to the 1224 pixels that were labeled as "water." We have performed comparisons with a number of other data sets, and observed similar results.

We compare SMT and SMT-S to other regularization schemes. Shrinkage estimators are a widely used class of estimators which regularize the covariance matrix by shrinking it toward some target structures. Shrinkage estimators generally have the form $\hat{R} = \alpha D + (1 - \alpha) S$ where D is some positive definite matrix. Two popular choices of D are the scaled identity matrix, $\text{trace}(S)/p \cdot I$ (called "Shrinkage-trI" in the plots), and the diagonal entries of S, $\text{diag}(S)$ (called "Shrinkage-D" in the plots), and the corresponding shrinkage estimators are used for comparison in this paper. SMT-S is the shrinkage estimator with D being the SMT estimate.

The parameter K required by SMT can be efficiently determined by a simple cross-validation procedure. Specifically, we partition the samples into three subsets, and choose K to maximize the average likelihood of the left-out subset given the estimated covariance using the other two subsets. This cross-validation requires little additional computation since every E_k is a sparse operator. As the number of samples grows (*i.e.,* $m \approx 10p$), the optimal K for SMT can be very large. In experiments, we considered K up to $p(p-1)/2$.

For SMT-S, we found that we could more agressively limit the number of Givens rotations (we used $K \leq 10p$) and still obtain effective performance by choosing the shrinkage coefficient α to maximize the cross-validated likelihood.

In Fig. 1, the performance of different covariance estimators is compared using the SCRR statistic. In Fig. 1(a,b,d,e), we see that SMT and the SMT-S algorithms outperformed the other regularization schemes over a range of values of m, with the most dramatic improvement at the smallest sample sizes. This behavior was *not* seen in Fig. 1(c,f), however, and we discuss this in the following section.

4.1. Structure in the covariance matrix

It is widely appreciated that real hyperspectral data is not fully modeled by a multivariate Gaussian distribution [13, 14]. But in addition to this structure *beyond* the Gaussian, there also appears to be structure *within* the covariance matrix.

We hypothesize that this structure is exploited by SMT and that that explains how SMT can outperform the other regularizers. We tested this hypothesis by randomly rotating the covariance matrices. A random orthogonal matrix Q (obtained from a QR decomposition of a matrix whose elements
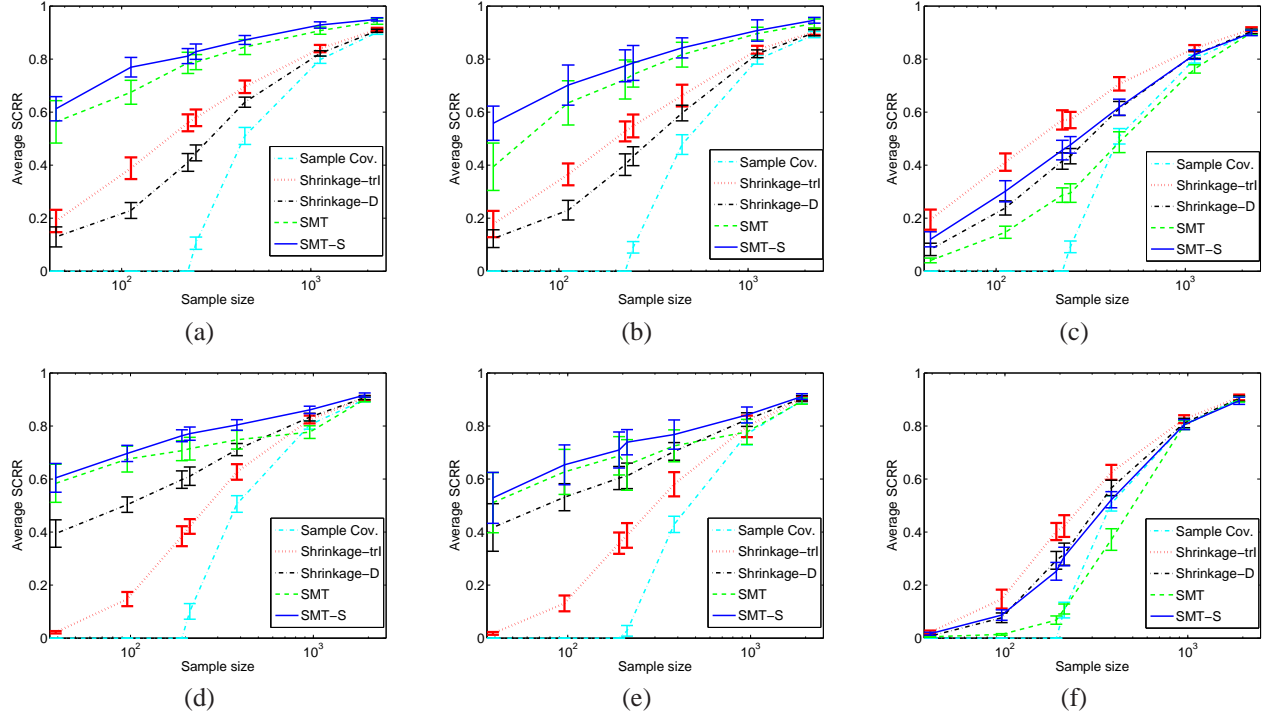
Fig. 1. Average of SCRR as a function of sample size for (a,b,c) Florida image and (d,e,f) Washington image. In all cases, the target signals are randomly generated from a Gaussian distribution, and the error bars are based on runs with 30 trials. (a,d) Gaussian samples are generated from the "true" covariance matrices for these two images. (b,e) Non-Gaussian samples are drawn at random with replacement from the image data itself. (c,f) Gaussian samples generated from randomly rotated covariance matrices. All plots are based on 30 trials, and each trial used a different rotation (for the randomly rotated covariances) and a different target **t**.

were independently chosen from a Gaussian distribution) was used to rotate the matrices: that is, $R' = QRQ^T$. Fig. 1(c,f) shows the performance of different covariance estimators applied to randomly rotated covariance matrices. Indeed, these panels show that the rotationally invariant Shrinkage-trI is the estimator with the best performance.

The performance of the sample covariance was similar for all of these cases. This is consistent with the theoretical result, due to Reed *et al.* [1], that for $m \gg p$, the expected value of SCRR is given by $1 - p/m$. (For $m < p$, the sample covariance is singular and the matched filter is undefined.)

One appeal of algorithms based on the covariance matrix is that they are often rotationally invariant. If we rotate our data **x** via some linear transform $\hat{\mathbf{x}} = L\mathbf{x}$, then the analysis on $\hat{\mathbf{x}}$ uses a different covariance matrix $\hat{R} = LRL^T$, but rotationally invariant algorithms will detect signal at the same pixels and achieve the same performance.

Rotationally invariant algorithms are attractive, but there are properties of the data which are manifestly not rotationally invariant. For instance, radiance or reflectance is always non-negative but arbitrary linear combinations can produce negative values. This is not a "problem" in the sense that the

algorithms which exploit this data do not require non-negative values; but it does point to unexploited structure in the data.

This structure can be illustrated by considering the covariance matrix R from a hyperspectral image. We can express this matrix in terms of a product of eigenvectors and eigenvalues, $R = E\Lambda E^T$, and then observe an image of the eigenvector matrix E. Fig. 2 illustrates such images, based on the Florida data and on the Washington data. We see that the eigenvector images, in Fig. 2(b,f) are sparse, particularly compared to their randomly rotated counterparts in Fig. 2(c,g). The histograms in Fig. 2(d,h), with their sharp peaks at zero, further emphasize this sparsity.

5. REFERENCES

[1] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems*, vol. 10, pp. 853–863, 1974.

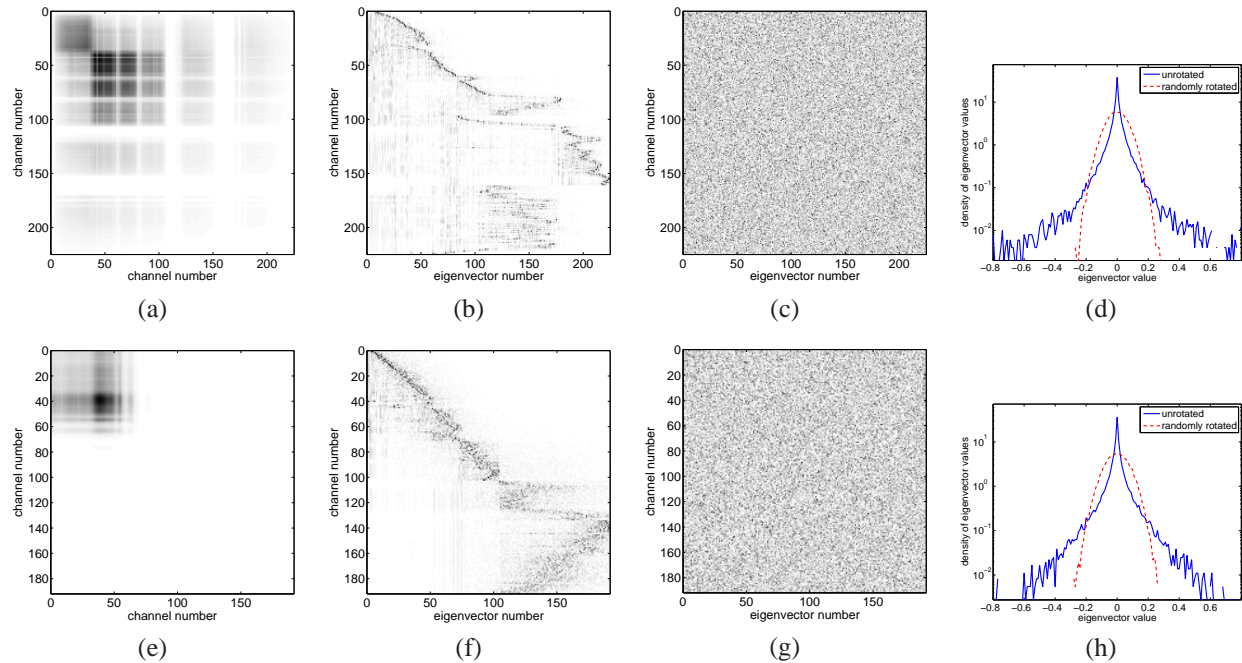[2] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley & Sons, 2003.

Fig. 2. Non-rotationally-invariant structure in the covariance matrix of real hyperspectral data is evident in the image of eigenvectors for (a,b,c,d) Florida data and for (e,f,g,h) Washington data. In (a,e) the covariance matrix is shown with larger values of R_{ij} plotted darker; in (b,f) the matrix E of eigenvectors of R is shown with larger values of the absolute value $|E_{ij}|$ shown darker; in (c,g) the eigenvectors are shown for a randomly rotated covariance matrix; and in (d,h) a histogram of eigenvector values is shown for both the original and the randomly rotated covariance matrix.

[3] O. Ledoit and M. Wolf, "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *J. Empirical Finance*, vol. 10, pp. 603–621, 2003.

[4] N. M. Nasrabadi, "Regularization for spectral matched filter and RX anomaly detector," *Proc. SPIE*, vol. 6966, pp. 696604, 2008.

[5] C. C. Funk, J. Theiler, D. A. Roberts, and C. C. Borel, "Clustering to improve matched filter detection of weak gas plumes in hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing*, vol. 39, pp. 1410–1420, 2001.

[6] J. P. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, pp. 763–767, 1996.

[7] M. J. Daniels and R. E. Kass, "Shrinkage estimators for covariance matrices," *Biometrics*, vol. 57, no. 4, pp. 1173–1184, 2001.

[8] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Annals of Statistics*, vol. 36, no. 1, pp. 199–227, 2008.

[9] G. Cao and C. A. Bouman, "Covariance estimation for high dimensional data vectors using the sparse matrix transform," in *Advances in Neural Information Processing Systems 21*. 2009, pp. 225–232, MIT Press.

[10] P. V. Villeneuve, H. A. Fry, J. Theiler, B. W. Smith, and A. D. Stocker, "Improved matched-filter detection techniques," *Proc. SPIE*, vol. 3753, pp. 278–285, 1999.

[11] A. De Maio, "Fast converging adaptive matched filter and adaptice cosine/coherence estimator," *Signal Processing*, vol. 82, pp. 1417–1423, 2002.

[12] J. Theiler, "Quantitative comparison of quadratic covariance-based anomalous change detectors," *Applied Optics*, vol. 47, pp. F12–F26, 2008.

[13] D. Manolakis, D. Marden, J. Kerekes, and G. Shaw, "On the statistics of hyperspectral imaging data," *Proc. SPIE*, vol. 4381, pp. 308–316, 2001.

[14] J. Theiler, B. R. Foy, and A. M. Fraser, "Characterizing non-Gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery," *Proc. SPIE*, vol. 5806, pp. 182–193, 2005.