SPARSE MATRIX TRANSFORM FOR IMPROVED DETECTION OF SIGNALS, ANOMALIES, AND CHANGES

Charles A. Bouman, Guangzhi Cao

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 87545

1. INTRODUCTION

For a variety of remote sensing detection problems, the covariance matrix is a key statistical quantity for characterizing the "normal" variability of the data. Particularly for high-dimensional data (*e.g.,* hyperspectral imagery), it provides a concise characterization of the data distribution that includes *all* pairwise correlations. For this reason, it remains a cornerstone in the statisitical analysis of remote sensing data. In particular, it is key to standard approaches for detecting signals [1], anomalies [2], and changes [3, 4, 5, 6, 7].

To apply these algorithms requires as a first step that the covariance matrix be estimated from the data. Particularly when there are limitations on the number of samples available for making that estimate, the sample covariance is often a sub-optimal choice.[1] For this reason, various kinds of regularization have been introduced [8, 9, 10, 11, 12, 13], including, very recently, regularization based on the sparse matrix transform (SMT) [14, 15].

In this paper, we will describe the SMT, and investigate its utility in improving the performance of algorithms for detecting signals, anomalies, and changes.

2. HYPERSPECTRAL DETECTION ALGORITHMS

2.1. Signals

If $\mathbf{t}$ corresponds to a target signal that we seek to detect over the background clutter $\mathbf{x}$ with mean $\langle \mathbf{x} \rangle = 0$ and covariance $R = \langle \mathbf{x}\mathbf{x}^T \rangle$, then the matched filter is given by $\mathbf{q} = R^{-1}\mathbf{t}$. Applied to an observation $\mathbf{s} = \alpha t + \mathbf{x}$, where α is a measure of target strength, and $\mathbf{x}$ is the background, the scalar quantity $\mathbf{q}^T\mathbf{s}$ provides an indicator of whether the signal is present (*i.e.,* whether $\alpha \neq 0$). When R is not known precisely, then $\hat{\mathbf{q}} = \hat{R}^{-1}\mathbf{t}$ is an approximate matched filter, and the performance of $\hat{\mathbf{q}}$ depends on the quality of the estimator $\hat{R}$ [1, 15].

[1] Actually, the sample covariance *is* the optimal estimator, in the sense of maximum likelihood, but most detection algorithms actually require an *inverse* covariance matrix. Unfortunately, the optimal estimate of the inverse covariance matrix is not the same as the inverse of the (maximum likelihood) optimal estimate of the covaricance matrix.

2.2. Anomalies

For anomaly detection, there is not a known target, and the aim is to decide whether the observed signal $\mathbf{s}$ is consistent with the background distribution, or if it is an ourlier. The RX detector [2] uses squared Mahalanobis distance from the data centroid as the discriminating statistic: $\mathbf{s}^T R^{-1}\mathbf{s}$. A good RX anomaly detector will enclose a fraction $f < 1$ (with $1 - f \ll 1$) of the data in as small a volume as possible. For Gaussian data, the nearer the estimate of $\hat{R}^{-1}$ to R^{-1}, the better the anomaly detector.

2.3. Changes

Given two images of the same scene, taken at different times and under different conditions, the aim of anomalous change detection is to determine whether any unusual changes have occurred in the scene. A variety of quadratic covariance-based algorithms have been developed for this purpose (see review in Ref. [7]), and their performance depends on the accuracy with with the covariance is estimated. For instance, if R_{aa} and R_{bb} are the covariance matrices for the two images, and R_{ab} is the cross-covariance, then the hyperbolic anomalous change detector [6] is given by

$$\mathcal{A}(\mathbf{s}_a, \mathbf{s}_b) = \begin{bmatrix}\mathbf{s}_a^T & \mathbf{s}_b^T\end{bmatrix} \left(\begin{bmatrix} R_{aa} & R_{ab}^T \\ R_{ab} & R_{bb} \end{bmatrix}^{-1} - \begin{bmatrix} R_{aa} & 0 \\ 0 & R_{bb} \end{bmatrix}^{-1} \right) \begin{bmatrix} \mathbf{s}_a \\ \mathbf{s}_b \end{bmatrix}, \tag{1}$$

where $\mathbf{s}_a$ and $\mathbf{s}_b$ are the observed signals in the two images.

3. 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

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

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 [14]

$$\hat{E} = \arg\min_{E\in\Omega} \left\{\left|\text{diag}(E^T S E)\right|\right\} \tag{3}$$

$$\hat{\Lambda} = \text{diag}\left(\hat{E}^T S \hat{E}\right), \tag{4}$$

where $S = \left\langle \mathbf{x}\mathbf{x}^T \right\rangle = \frac{1}{m} X X^T$ 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: $\hat{R} = S$.

3.1. Sparse Matrix Transform (SMT)

The Sparse Matrix Transform (SMT) provides a 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

$$\Theta(i,j,\theta)_{rs} = \begin{cases} \cos(\theta) - 1 & \text{if } r = s = i \text{ or } r = s = j \\ \sin(\theta) & \text{if } r = i \text{ and } s = j \\ -\sin(\theta) & \text{if } r = j \text{ and } s = i \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

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. (3) with $\Omega = \Omega_K$ and Eq. (4). (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.)

4. 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] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1760–1770, 1990.

[3] A. Schaum and A. Stocker, "Long-interval chronochrome target detection," *Proc. 1997 International Symposium on Spectral Sensing Research*, 1998.

[4] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF post-processing in multispectral bi-temporal image data: new approaches to change detection studies," *Remote Sensing of the Environment*, vol. 64, pp. 1–19, 1998.

[5] A. Schaum and A. Stocker, "Hyperspectral change detection and supervised matched filtering based on covariance equalization," *Proc. SPIE*, vol. 5425, pp. 77–90, 2004.

[6] J. Theiler and S. Perkins, "Proposed framework for anomalous change detection," *ICML Workshop on Machine Learning Algorithms for Surveillance and Event Detection*, pp. 7–14, 2006.

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

[8] 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.

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

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

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

[12] 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.

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

[14] 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.

[15] G. Cao, C. A. Bouman, and J. Theiler, "Weak signal detection in hyperspectral imagery using sparse matrix transform (SMT) covariance estimation," *Proc. Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2009, doi: 10.1109/WHISPERS.2009.5289043.