

CLASSIFICATION PERFORMANCE OF RANDOM-PROJECTION-BASED DIMENSIONALITY REDUCTION OF HYPERSPECTRAL IMAGERY

James E. Fowler, Qian Du, Wei Zhu, and Nicolas H. Younan

Department of Electrical and Computer Engineering, Geosystems Research Institute,
Mississippi State University, USA

ABSTRACT

High-dimensional data such as hyperspectral imagery is traditionally acquired in full dimensionality before being reduced in dimension prior to processing. Conventional dimensionality reduction on-board remote devices is often prohibitive due to limited computational resources; on the other hand, integrating random projections directly into signal acquisition offers alternative dimensionality reduction without sender-side computational cost. Effective receiver-side reconstruction from such random projections has been demonstrated previously using compressive-projection principal component analysis (CPPCA). While this prior work has focused on squared-error quality measures, the present work reports experimental results illustrating preservation of statistical class separation and anomaly-detection performance for CPPCA reconstruction following random-projection-based dimensionality reduction.

Index Terms— random projection, dimensionality reduction, hyperspectral imagery

1. INTRODUCTION

In the traditional data pipeline used with high-dimensional data such as hyperspectral imagery, data is acquired in its full dimensionality within some typically remote signal-acquisition platform (e.g., a satellite) only to be often subsequently reduced in dimension prior to application-specific processing such as classification. However, it would be greatly beneficial if dimensionality reduction could occur before data downlink, since many signal-acquisition platforms are severely resource-constrained such that on-board dimensionality reduction could dramatically cut storage and communication burdens faced by such remote sensors. However, many approaches to dimensionality reduction are not only data dependent but also exceedingly computationally expensive so as to preclude on-board implementation. In this work, we investigate a significant departure from the traditional data flow that integrates dimensionality reduction directly into the signal-acquisition process in order to avoid not only the computational burden of explicit dimensionality reduction, but also the production of onerous quantities of data in the first place. We employ the emerging mathematical theory of random projections to this end, effectively shifting the computational burden of dimensionality reduction from the resource-constrained acquisition device to a more powerful base-station system which is presumably an earth-based central site.

Recent work in the applied-mathematics and signal-processing communities has demonstrated that projections onto randomly chosen subspaces can be a particularly powerful form of dimensionality reduction. Namely, the mathematical theory of *compressed sensing* (CS) (e.g., [1]) establishes that sparsely representable sig-

nals can be recovered exactly from data-independent random projections. Furthermore, we have recently developed *compressive-projection principal component analysis* (CPPCA) [2–4], which recovers an approximate PCA representation of the original signal from random projections. Both CS and CPPCA permit sensing platforms to enjoy the benefits of dimensionality reduction (less burdensome storage and communication requirements) without the expense of computation associated with explicit dimensionality reduction since the random projections can be accomplished simultaneously with the sensing and signal-acquisition process, while the more expensive reconstruction from the projections takes place at the receiver-side base station.

The merits of random projection plus CS or CPPCA reconstruction have previously been established mathematically with a focus on squared-error-based signal-quality measures. On the contrary, the main objective of the present work focuses on performance at classification and the corresponding issue of preservation of statistical class separability for random-projection-based dimensionality reduction, since such application-specific classification performance is typically of primary importance in many geospatial applications.

Also of interest is the detection of anomalous data points, infrequent occurrences that differ significantly from the majority of the dataset. In hyperspectral image analysis, anomalies are important pixels since they often represent man-made targets of critical application interest. During any dimensionality-reduction process, it is therefore essential to preserve these pixels in order to enable target detection and discrimination in an unsupervised setting. However, anomalous pixels are difficult to preserve well due to the fact that their spectral features are typically quite different from their surroundings. For example, anomalous pixels are often among those that experience the largest distortion after lossy compression [5].

Previous results in [2–4] indicate that CPPCA outperforms CS for a variety of bases when reconstruction quality is measured in terms of a traditional squared-error-based measure. We anticipate that this is due in no small part to the fact that CPPCA recovers an approximate PCA basis for the dataset in question, and PCA is known to be the optimal basis in several mathematical respects in terms of reduction of squared error per percentage of coefficients retained. It is known, however, that PCA is not necessarily the optimal basis for maintaining class separability [6]. Consequently, in the experimental results that follow, we empirically evaluate the performance of random-projection-based dimensionality reduction followed by CPPCA-based reconstruction at both classification and anomaly-detection tasks. In both instances, we find that performance on CPPCA-reconstructed datasets is very similar to that on the original datasets, indicating that random projections are a promising new avenue to achieve low-complexity

remote-platform dimensionality reduction.

2. AN OVERVIEW OF CPPCA

At the core of the CPPCA technique is a receiver-side process that produces an approximation to the PCA transform basis. Consider a dataset of M zero-mean vectors $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_M]$, where each $\mathbf{x}_m \in \mathbb{R}^N$. The covariance matrix of \mathbf{X} is $\mathbf{\Sigma} = \mathbf{X}\mathbf{X}^T/M$, and the PCA transform matrix is the $N \times N$ matrix \mathbf{W} of eigenvectors that emanates from the eigendecomposition of $\mathbf{\Sigma}$; i.e.,

$$\mathbf{\Sigma} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T, \quad (1)$$

where \mathbf{W} contains the N unit eigenvectors of $\mathbf{\Sigma}$ column-wise. However, central to the CPPCA paradigm is that production of the PCA transform matrix occurs at the receiver rather than at the sender as in the traditional use of PCA; that is, the CPPCA receiver cannot implement eigendecomposition (1) directly as it does not know either \mathbf{X} or $\mathbf{\Sigma}$. Instead, the receiver knows only K -dimensional projections of \mathbf{X} . Specifically, suppose we have K orthonormal vectors \mathbf{p}_k that form the basis of K -dimensional subspace \mathcal{P} such that $\mathbf{P} = [\mathbf{p}_1 \ \cdots \ \mathbf{p}_K]$ provides an orthogonal projection onto \mathcal{P} . The CPPCA sender produces $\tilde{\mathbf{Y}} = \mathbf{P}^T\mathbf{X}$, and it is from projections $\tilde{\mathbf{Y}}$ that the CPPCA receiver approximates \mathbf{W} . The projected vectors have covariance

$$\tilde{\mathbf{\Sigma}} = \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T/M = \mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}/M = \mathbf{P}^T\mathbf{\Sigma}\mathbf{P}, \quad (2)$$

which the CPPCA receiver calculates having received $\tilde{\mathbf{Y}}$ from the sender.

Rayleigh-Ritz theory [7] describes the relation between the eigenvectors of $\mathbf{\Sigma}$ and those of $\tilde{\mathbf{\Sigma}}$ as given by (2). Covariance matrix $\mathbf{\Sigma}$ has spectrum $\lambda(\mathbf{\Sigma}) = \{\lambda_1(\mathbf{\Sigma}), \dots, \lambda_N(\mathbf{\Sigma})\}$, where the eigenvalues satisfy $\lambda_1(\mathbf{\Sigma}) \geq \dots \geq \lambda_N(\mathbf{\Sigma})$, and the corresponding unit eigenvectors are \mathbf{w}_n . The eigendecomposition of $\tilde{\mathbf{\Sigma}} = \mathbf{P}^T\mathbf{\Sigma}\mathbf{P}$ is $\tilde{\mathbf{\Sigma}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{U}}^T$, where $\tilde{\mathbf{U}} = [\tilde{\mathbf{u}}_1 \ \cdots \ \tilde{\mathbf{u}}_K]$, $\tilde{\mathbf{\Lambda}} = \text{diag}(\lambda_1(\tilde{\mathbf{\Sigma}}), \dots, \lambda_K(\tilde{\mathbf{\Sigma}}))$, $\|\tilde{\mathbf{u}}_k\|_2 = 1$, and $\lambda_1(\tilde{\mathbf{\Sigma}}) \geq \dots \geq \lambda_K(\tilde{\mathbf{\Sigma}})$. The K eigenvalues $\lambda_k(\tilde{\mathbf{\Sigma}})$ are called *Ritz values*; additionally, there are K vectors, known as *Ritz vectors*, defined as

$$\mathbf{u}_k = \mathbf{P}\tilde{\mathbf{u}}_k, \quad 1 \leq k \leq K, \quad (3)$$

where $\tilde{\mathbf{u}}_k$ are the eigenvectors of $\tilde{\mathbf{\Sigma}}$. Finally, we define *normalized projection* \mathbf{v}_n as the orthogonal projection of \mathbf{w}_n onto \mathcal{P} , normalized to unit length; i.e.,

$$\mathbf{v}_n = \frac{\mathbf{P}\mathbf{P}^T\mathbf{w}_n}{\|\mathbf{P}\mathbf{P}^T\mathbf{w}_n\|_2}. \quad (4)$$

These vectors are illustrated for an example distribution in the simple case of $N = 3$ and $K = 2$ in Fig. 1.

Existing Rayleigh-Ritz theory is rather limited in that it tells us very little about the Ritz vectors for $K < N$. We know only that the Ritz vectors do not typically align with the orthogonal projections of any of the eigenvectors [7]; i.e., $\mathbf{u}_k \neq \mathbf{v}_n$ in general. However, CPPCA is built on the central idea that, if subspace \mathcal{P} is chosen randomly, and the distribution of the vectors in \mathbf{X} is highly eccentric in that eigenvalue $\lambda_k(\mathbf{\Sigma})$ is sufficiently separated in value with respect to the other eigenvalues, then it is likely that its corresponding normalized projection, \mathbf{v}_k , will be quite close to the Ritz vector, \mathbf{u}_k , corresponding to the Ritz value $\lambda_k(\tilde{\mathbf{\Sigma}})$. Under the assumption that $\mathbf{u}_k \approx \mathbf{v}_k$, a algorithm based on projections

onto convex sets (POCS) [8] was devised in [2, 4] to approximate the first L eigenvectors \mathbf{w}_n from $\tilde{\mathbf{Y}}$; the reader is referred to [2, 4] for the specific details of this process. Suffice it to say, however, that the entire feasibility of the CPPCA technique rests on the approximation $\mathbf{u}_k \approx \mathbf{v}_k$. An analysis was presented in [3, 4] that established the validity of this approximation.

3. EXPERIMENTAL RESULTS

Fig. 4 presents experimental results regarding the classification performance resulting from CPPCA reconstruction for a 100×100 spatial region of the ‘‘Moffett’’ hyperspectral image, an AVIRIS image with 224 spectral bands. The second row of images in Fig. 4 depicts the five independent components determined from the original image (shown on the first row of Fig. 4); these component images are produced using the Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm for independent component analysis (ICA). The third row of images in Fig. 4 depicts the same independent component images determined from a CPPCA reconstruction after the original image was reduced in spectral dimension by random projection onto 90-dimensional random subspaces (i.e., a 40% reduction in data size). Also included is an objective measure of similarity, the spatial correlation coefficient as denoted by ρ . Since, for each of the endmember classes, ρ is quite close to 1.0, we conclude that the random projections are successful in preserving essentially all of the class separability in this example.

To examine the feasibility of anomaly detection within random projection subspaces, we deploy the well-known RX algorithm [9] for anomaly detection on an original AVIRIS dataset (a 100×100 spatial region from the 224-band Moffett scene). We then create a random projection matrix that projects each N -band hyperspectral vector into a K -dimensional subspace. We deploy the RX algorithm within this subspace and compare the resulting anomaly-detection map to the corresponding map produced from the original N -band dataset (here, $N = 224$). The results are summarized in Fig. 2. In Fig. 2, we see that the correlation coefficient between the two anomaly-detection maps varies from about 0.6 to 0.9 as the dimensionality of the subspace increases (here, ‘‘reduction factor’’ refers to the ratio K/N). Thus, the maps are quite similar, particularly for K/N above 0.3. Figs. 3(a) and (b) present the original-data map and the projected map for $K/N = 0.4$, indicating that, at least subjectively, locations of anomalies appear to be well-detected within the projected subspace.

4. CONCLUSIONS

The results presented here suggest that, empirically, both classification as well as anomaly detection can indeed be carried out within randomly chosen projection subspaces when CPPCA is used to reconstruct from the projections. The component images resulting from ICA classification applied to the CPPCA-reconstructed dataset closely resemble those from ICA of the original AVIRIS dataset, with a correlation coefficient well above 0.9 for each component. Anomaly detection using the RX algorithm on the CPPCA reconstruction also produces similar results as compared to anomalies identified in the original dataset. Thus, we conclude that random projection at the remote sender constitutes a promising avenue for low-cost dimensionality reduction, with the brunt of the computational cost incurred at the receiver side in the form of CPPCA reconstruction.

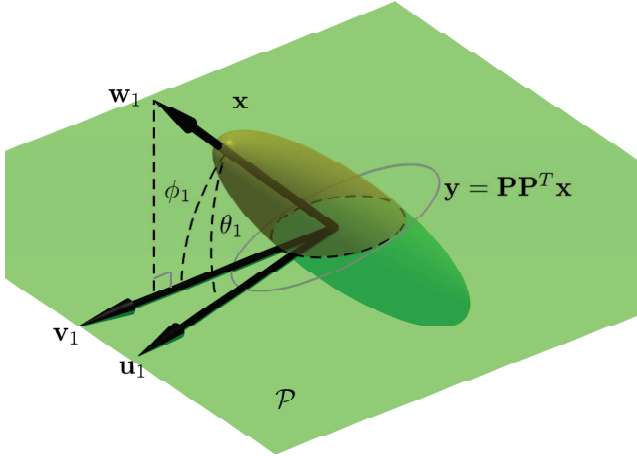


Figure 1: Data distribution of \mathbf{x} in \mathbb{R}^3 is projected onto 2D subspace \mathcal{P} as \mathbf{y} ; the first Ritz vector, \mathbf{u}_1 , lies close to the normalized projection, \mathbf{v}_1 , onto \mathcal{P} of the first eigenvector, \mathbf{w}_1 , of \mathbf{x} (from [2]).

5. REFERENCES

- [1] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.
- [2] J. E. Fowler, "Compressive-projection principal component analysis for the compression of hyperspectral signatures," in *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. W. Marcellin, Eds., Snowbird, UT, March 2008, pp. 83–92.
- [3] —, "Compressive-projection principal component analysis and the first eigenvector," in *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. W. Marcellin, Eds., Snowbird, UT, March 2009, pp. 223–232.
- [4] —, "Compressive-projection principal component analysis," *IEEE Transactions on Image Processing*, to appear.
- [5] Q. Du, W. Zhu, and J. E. Fowler, "Anomaly-based JPEG2000 compression of hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 696–700, October 2008.
- [6] S. Prasad and L. M. Bruce, "Limitations of principal component analysis for hyperspectral target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 625–629, 2008.
- [7] B. N. Parlett, *The Symmetric Eigenvalue Problem*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1998.
- [8] P. L. Combettes, "The foundations of set theoretic estimation," *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182–208, February 1993.
- [9] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1760–1770, October 1990.

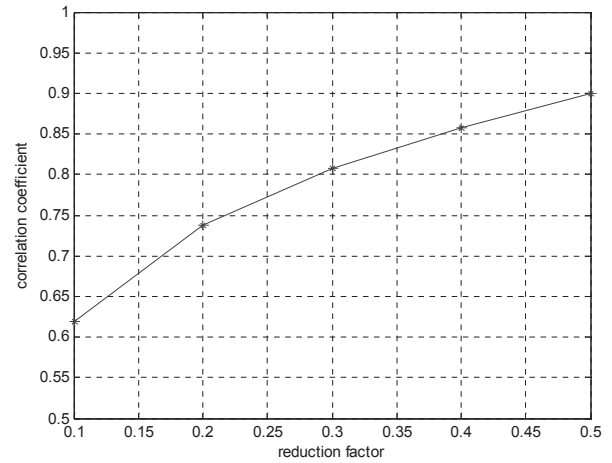
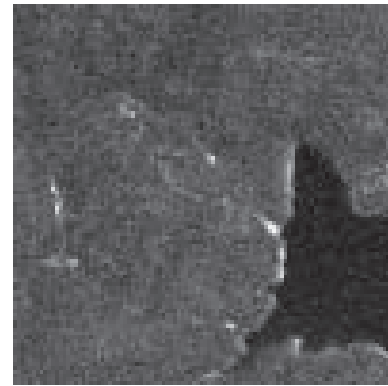


Figure 2: Correlation coefficient between anomaly-detection maps as K/N varies.



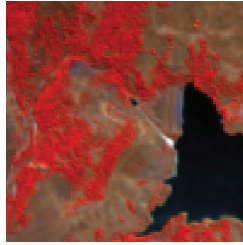
(a)



(b)

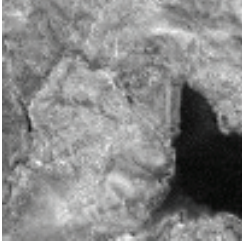
Figure 3: (a) Anomaly-detection map using the original N -band data ($N = 224$). (b) Anomaly-detection map using the data after random projection into a K -dimensional random subspace ($K/N = 0.4$); correlation coefficient $\rho = 0.8576$.

CIR color composite of original 100×100 Moffett Image

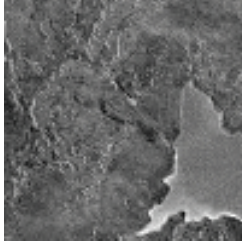


ICA classification directly from the original image

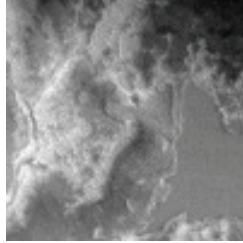
Soil



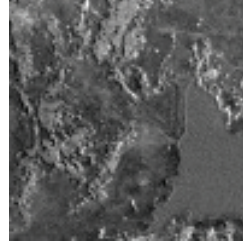
Water 1



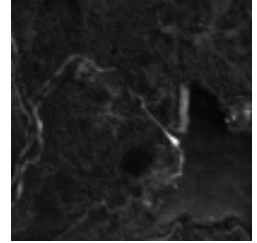
Vegetation 1



Vegetation 2

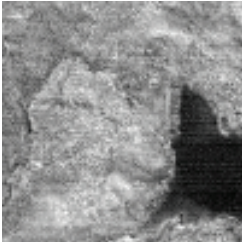


Water 2



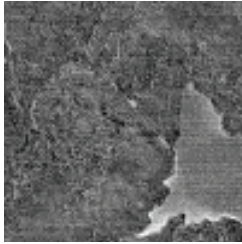
ICA classification from CPPCA reconstruction after random projection

Soil



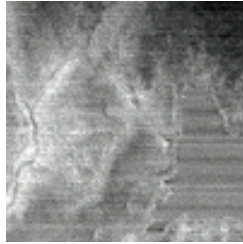
$\rho = 0.9771$

Water 1



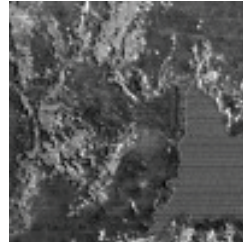
$\rho = 0.9137$

Vegetation 1



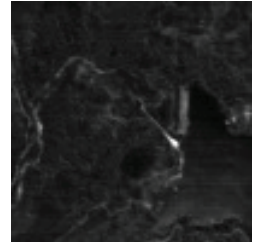
$\rho = 0.9360$

Vegetation 2



$\rho = 0.9528$

Water 2



$\rho = 0.9867$

Figure 4: Classification performance for CPPCA as compared to that on the original image