# Compressive-Projection Principal Component Analysis for the Compression of Hyperspectral Signatures

James E. Fowler

Department of Electrical and Computer Engineering
GeoResources Institute (GRI)
Mississippi State University, Starkville, Mississippi

## Abstract

A method is proposed for the compression of hyperspectral signature vectors on severely resource-constrained encoding platforms. The proposed technique, compressive-projection principal component analysis, recovers from random projections not only transform coefficients but also an approximation to the principal-component basis, effectively shifting the computational burden of principal component analysis from the encoder to the decoder. In its use of random projections, the proposed method resembles compressed sensing but differs in that simple linear reconstruction suffices for coefficient recovery. Existing results from perturbation theory are invoked to argue for the robustness under quantization of the eigenvector-recovery process central to the proposed technique, and experimental results demonstrate a significant rate-distortion performance advantage over compressed sensing using a variety of popular bases.

## Introduction

Hyperspectral sensors typically produce a large quantity of data, often necessitating compression prior to storage or transmission. For such hyperspectral compression, there is substantial interest in principal component analysis (PCA) for the decorrelation of the spectral bands. Indeed, spectral PCA has shown significant performance advantage in the compression of hyperspectral imagery (e.g., [1, 2]). However, PCA is a data-dependent transform arising from the eigendecomposition of the covariance matrix of the signal in question. In traditional compression applications, the encoder must calculate the PCA transform before it can be applied to the data. Unfortunately, the computational burden that this process entails may well exceed the limited capabilities of many hyperspectral sensing platforms which are often severely resource-constrained, e.g., satellite-borne devices. As a consequence, there is increasing interest in avoiding the burden of compression algorithms like PCA by integrating dimensionality reduction directly into the signal sensing and acquisition process.

In this paper, we consider the compression of hyperspectral signatures with each signature vector describing a single spatial location across typically hundreds of contiguous spectral bands. We propose a process that effectively shifts the computational burden of PCA of these spectral signatures from the resource-constrained encoder to the decoder which presumably resides on a significantly more powerful "base-station" system. Our approach, compressive-projection PCA (CPPCA), is driven by projections at the sensor onto lower-dimensional subspaces chosen at random. The CPPCA decoder, given only these random projections, recovers not only the coefficients associated with the PCA transform, but also an approximation to the PCA transform basis itself.

In *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. W. Marcellin, Eds., Snowbird, UT, March 2008, pp. 83–92.

True data compression, and not just dimensionality reduction, must necessarily involve some form of quantization. In CPPCA, quantization of the projections will produce distortions in both the coefficient-recovery process as well as in the eigenvector recovery used to approximate the PCA transform. However, as we show below, known results from perturbation theory argue that the eigenvector-recovery procedure central to CPPCA is robust under quantization.

We note that, in its reliance on random projections to achieve integrated acquisition and compression, CPPCA can be considered to be, in essence, an alternative to the emerging mathematical paradigm of compressed sensing (CS) (e.g., [3]). In experimental results below, we compare the rate-distortion performance of CPPCA against that of CS wherein both approaches incur quantization loss in the form of uniform scalar quantization (SQ). We find that CPPCA achieves rate-distortion performance substantially superior to that of CS using several bases commonly applied to hyperspectral data. This gain comes in spite of the fact that CPPCA can be viewed as being doubly affected—coefficients as well as transform—by the quantization, whereas CS is affected in the coefficients alone, its transform being fixed.

## Compressive-Projection Principal Component Analysis (CPPCA)

Consider a dataset of $M$ vectors $\mathbf{X} = \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_M\end{bmatrix}$, where each $\mathbf{x}_m$ is an $N$-dimensional hyperspectral signature vector; we assume that the vectors have zero mean. In the CPPCA encoder, the vectors of $\mathbf{X}$ are subjected to random projection, the projections are quantized, and a compressed bitstream is delivered to the CPPCA decoder. The CPPCA decoder then must recover not only the PCA transform coefficients, but also the basis vectors of the transform itself, all from the quantized projections. We assume that the decoder knows only the projection operator and its resulting projections, but not $\mathbf{X}$ or its statistics (e.g., covariance). Below, we present the CPPCA approach, starting with a consideration of PCA with respect to projections. We delay consideration of the effects of quantization to the next section.

### *PCA and Projections*

The covariance matrix of $\mathbf{X}$ is $\mathbf{\Sigma} = \mathbf{X}\mathbf{X}^T/M$. For a given spectral signature, $\mathbf{x}_m$, in $\mathbf{X}$, the PCA of $\mathbf{x}_m$ results from the application of a linear transform, $\check{\mathbf{x}}_m = \mathbf{W}^T\mathbf{x}_m$, where $N \times N$ transform matrix $\mathbf{W}$ emanates from the eigendecomposition of $\mathbf{\Sigma}$; i.e.,

$$\mathbf{\Sigma} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T, \tag{1}$$

where $\mathbf{W}$ contains the $N$ unit eigenvectors of $\mathbf{\Sigma}$ column-wise. Suppose we have $K$ orthonormal vectors $\mathbf{p}_k$ that form the basis of $K$-dimensional subspace $\mathcal{P}$ such that $\mathbf{P} = \begin{bmatrix}\mathbf{p}_1 & \cdots & \mathbf{p}_K\end{bmatrix}$ provides an orthogonal projection onto $\mathcal{P}$. That is, the orthogonal projection of $\mathbf{x}_m$ onto $\mathcal{P}$ is $\mathbf{y}_m = \mathbf{P}\mathbf{P}^T\mathbf{x}_m$; expressed with respect to the basis $\{\mathbf{p}_k\}$, we have $\widetilde{\mathbf{y}}_m = \mathbf{P}^T\mathbf{x}_m$, such that $\mathbf{y}_m = \mathbf{P}\widetilde{\mathbf{y}}_m$. The projected vectors $\widetilde{\mathbf{Y}} = \begin{bmatrix}\mathbf{y}_1 & \cdots & \mathbf{y}_M\end{bmatrix}$ have covariance

$$\widetilde{\mathbf{\Sigma}} = \widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^T/M = \mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}/M = \mathbf{P}^T\mathbf{\Sigma}\mathbf{P}. \tag{2}$$

Rayleigh-Ritz theory [4] describes the relation between the eigenvectors of $\mathbf{\Sigma}$ and those of $\widetilde{\mathbf{\Sigma}}$. Assume covariance matrix $\mathbf{\Sigma}$ has eigenvalues $\lambda_n$ and corresponding eigenvectors $\mathbf{w}_n$,

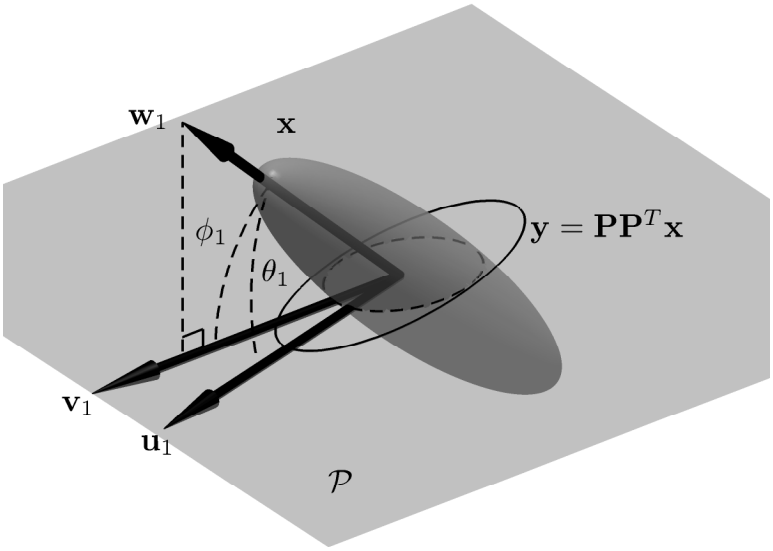



Figure 1: Data distribution of $\mathbf{x}$ in $\mathcal{R}^3$ is projected onto two-dimensional subspace $\mathcal{P}$ as $\mathbf{y}$. The first Ritz vector, $\mathbf{u}_1$, of the projected distribution lies close to the normalized projection, $\mathbf{v}_1$, onto $\mathcal{P}$ of the first eigenvector, $\mathbf{w}_1$, of $\mathbf{x}$.

$1 \leq n \leq N$; i.e., $\mathbf{\Sigma} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T$, where $\mathbf{W} = \begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_N\end{bmatrix}$, $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, $\|\mathbf{w}_n\|_2 = 1$, and $\lambda_1 \geq \cdots \geq \lambda_N$. The eigendecomposition of $\widetilde{\mathbf{\Sigma}} = \mathbf{P}^T\mathbf{\Sigma}\mathbf{P}$ is $\widetilde{\mathbf{\Sigma}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}^T$, where $\widetilde{\mathbf{U}} = \begin{bmatrix}\widetilde{\mathbf{u}}_1 & \cdots & \widetilde{\mathbf{u}}_K\end{bmatrix}$, $\widetilde{\mathbf{\Lambda}} = \mathrm{diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_K)$, $\|\widetilde{\mathbf{u}}_k\|_2 = 1$, and $\widetilde{\lambda}_1 \geq \cdots \geq \widetilde{\lambda}_K$. The $K$ eigenvalues $\widetilde{\lambda}_k$ of $\widetilde{\mathbf{\Sigma}}$ are called *Ritz values*; additionally, there are $K$ vectors, known as *Ritz vectors*, defined as

$$\mathbf{u}_k = \mathbf{P}\widetilde{\mathbf{u}}_k, \quad 1 \leq k \leq K, \tag{3}$$

where $\widetilde{\mathbf{u}}_k$ are the eigenvectors of $\widetilde{\mathbf{\Sigma}}$. Note that $\|\mathbf{u}_k\|_2 = 1$. 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.,

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

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

Traditional design methods for PCA produce the transform $\mathbf{W}$ via the eigendecomposition of (1); however, in the CPPCA decoder, one has access to merely $\widetilde{\mathbf{\Sigma}}$ and not $\mathbf{\Sigma}$ as required in (1). The goal of CPPCA is thus to approximate $\mathbf{W}$ from $\widetilde{\mathbf{\Sigma}}$ without knowledge of $\mathbf{\Sigma}$, given that $\widetilde{\mathbf{\Sigma}}$ results from random projection. CPPCA is built on the heuristic that, if subspace $\mathcal{P}$ is chosen randomly, and the distribution of the vectors in $\mathbf{X}$ is highly eccentric in that the largest eigenvalue $\lambda_1$ is very large with respect to the other eigenvalues, then it is likely that the normalized projection, $\mathbf{v}_1$, will be quite close to the Ritz vector, $\mathbf{u}_1$, corresponding to the largest Ritz value $\widetilde{\lambda}_1$. Of course, it is possible for $\mathcal{P}$ to be oriented such that this is not the case (i.e., if $\mathbf{w}_1$ happens to be close to being orthogonal to $\mathcal{P}$); however, for a randomly chosen $\mathcal{P}$ and highly eccentric $\mathbf{X}$ distribution, we see from Fig. 1 that it appears likely that $\mathbf{u}_1$ and $\mathbf{v}_1$ will be close. We also anticipate that, if the remaining eigenvalues are also mutually well-separated, we can form similar approximations of the

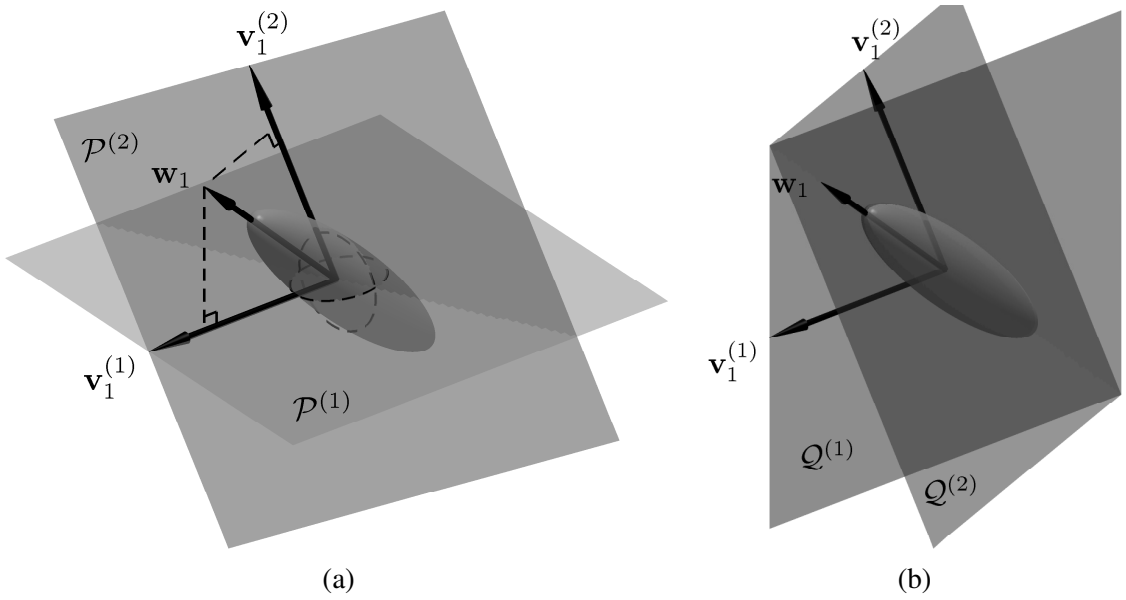



Figure 2: (a) Two two-dimensional subspaces $\mathcal{P}^{(1)}$ and $\mathcal{P}^{(2)}$ with corresponding normalized projections $\mathbf{v}_1^{(1)}$ and $\mathbf{v}_1^{(2)}$. (b) Subspaces $\mathcal{Q}^{(1)}$ and $\mathcal{Q}^{(2)}$ whose intersection uniquely determines eigenvector $\mathbf{w}_1$ up to a sign.

normalized projections using the corresponding Ritz vectors; i.e., $\mathbf{u}_2$ to approximate $\mathbf{v}_2$, $\mathbf{u}_3$ to approximate $\mathbf{v}_3$, etc. Our empirical experience suggests that this heuristic works well in practice, as experimental results presented later illustrate.

*Eigenvector Recovery*

The CPPCA decoder first recovers an approximation to the PCA transform basis by recovering approximations to the first $L$ eigenvectors of $\mathbf{\Sigma}$ from random projections. We observe that, if we knew normalized projection $\mathbf{v}$ of eigenvector $\mathbf{w}$ in subspace $\mathcal{P}$, we could form subspace $\mathcal{Q}$ as

$$\mathcal{Q} = \mathcal{P}^{\perp} \oplus \operatorname{span}\{\mathbf{v}\}, \tag{5}$$

the direct sum of the orthogonal complement of $\mathcal{P}$ with a 1D space containing $\mathbf{v}$. Clearly, $\mathbf{w}$ would lie in $\mathcal{Q}$. Suppose then that we produce $J$ distinct random $K$-dimensional subspaces, $\mathcal{P}^{(1)}$ through $\mathcal{P}^{(J)}$, each containing a normalized projection, $\mathbf{v}^{(1)}$ through $\mathbf{v}^{(J)}$, respectively, produced via (4) using the corresponding projection matrices, $\mathbf{P}^{(1)}$ through $\mathbf{P}^{(J)}$. We could then form subspaces $\mathcal{Q}^{(1)}$ through $\mathcal{Q}^{(J)}$ via (5) using $\mathcal{P}^{(1)}$, ..., $\mathcal{P}^{(J)}$ and $\mathbf{v}^{(1)}$, ..., $\mathbf{v}^{(J)}$. The eigenvector $\mathbf{w}$ would thus be in the intersection $\mathcal{Q}^{(1)} \cap \cdots \cap \mathcal{Q}^{(J)}$. This situation is illustrated in Fig. 2 for the case of $N = 3$, $K = 2$, $J = 2$, and the eigenvector in question being $\mathbf{w}_1$.

In the CPPCA decoder, though, we do not have access to normalized projections; instead, we can form Ritz vectors in each subspace $\mathcal{P}^{(j)}$ via an eigendecomposition of the corresponding projected covariance matrix $\widetilde{\mathbf{\Sigma}}^{(j)}$. We use the Ritz vectors then to form the spaces $\mathcal{Q}^{(j)}$. Since the Ritz vectors will differ slightly from the true normalized projections, the intersection $\mathcal{Q}^{(1)} \cap \cdots \cap \mathcal{Q}^{(J)}$ is almost certain to be empty. However, since the $\mathcal{Q}^{(j)}$

are closed and convex, a parallel implementation of projections onto convex sets (POCS) will converge to a least-squares solution minimizing the average distance to the subspaces $\mathcal{Q}^{(j)}$ [5]; this POCS solution can then be used to approximate $\mathbf{w}$. Specifically, for iteration $i = 1, 2, \ldots$, we form an estimate of the eigenvector as

$$\widehat{\mathbf{w}}^{(i)} = \frac{1}{J} \sum_{j=1}^{J} \mathbf{Q}^{(j)} \mathbf{Q}^{(j)^T} \widehat{\mathbf{w}}^{(i-1)}, \tag{6}$$

where projection onto $\mathcal{Q}^{(j)}$ is performed by the matrix $\mathbf{Q}^{(j)}$. (6) will converge to $\widehat{\mathbf{w}}$; normalizing this $\widehat{\mathbf{w}}$ will approximate the desired normalized eigenvector $\mathbf{w}$ (up to sign).

In order to avoid producing multiple random projections for each spectral signature in our dataset, the CPPCA encoder splits the dataset of $M$ vectors $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_M \end{bmatrix}$ into $J$ partitions $\mathbf{X}^{(j)}$, each associated with its own randomly chosen projection $\mathbf{P}^{(j)}$, $1 \leq j \leq J$. It is assumed that the dataset splitting is conducted such that each $\mathbf{X}^{(j)}$ closely resembles the whole dataset $\mathbf{X}$ statistically and so has approximately the same eigendecomposition. The encoder transmits the projected data $\widetilde{\mathbf{Y}}^{(j)} = \mathbf{P}^{(j)} \mathbf{X}^{(j)}$ to the decoder which is assumed to know the projections $\mathbf{P}^{(j)}$ *a priori*. In the CPPCA decoder, a set of Ritz vectors $\mathbf{u}_k^{(j)}$ is produced, and the Ritz vectors are used in place of the normalized projections to drive the POCS recovery. The CPPCA decoder repeats this POCS procedure using the first $L$ Ritz vectors to approximate the first $L$ principal eigenvectors which are assembled into $N \times L$ matrix $\mathbf{\Psi}$, an approximation to the $L$-component PCA transform, $L \leq K$.

*Coefficient Recovery*

Once obtaining $\mathbf{\Psi}$, the CPPCA decoder then proceeds to recover the PCA coefficients by solving $\widetilde{\mathbf{Y}}^{(j)} = \mathbf{P}^{(j)^T} \mathbf{\Psi} \check{\mathbf{X}}^{(j)}$ for PCA coefficients $\check{\mathbf{X}}^{(j)}$ in the least-squares sense for each $j$. This linear reconstruction is accomplished simply by using the pseudoinverse,

$$\check{\mathbf{X}}^{(j)} = \left( \mathbf{P}^{(j)^T} \mathbf{\Psi} \right)^{+} \widetilde{\mathbf{Y}}^{(j)}. \tag{7}$$

## Quantization Issues

In CPPCA, original spectral signature vector $\mathbf{x} \in \mathcal{R}^N$ is projected into a $K$-dimensional subspace $\mathcal{P}$ as $\widetilde{\mathbf{y}} = \mathbf{P}^T \mathbf{x}$. For compression purposes, we apply SQ to the components $\widetilde{\mathbf{y}}$, followed by some form of entropy coding. In order to analyze the effect of the quantization process on the performance of CPPCA, we adopt a simplified, high-resolution model [6] of uniform SQ as additive noise of variance $q^2/12$, where $q$ is the quantizer stepsize. That is, $\widetilde{\mathbf{y}}$ is quantized as $\widehat{\mathbf{y}} = \widetilde{\mathbf{y}} + \mathbf{n}$ where noise $\mathbf{n}$ has covariance $\mathbf{N} = E\left[\mathbf{n}\mathbf{n}^T\right] = (q^2/12)\mathbf{I}$ and zero mean. The covariance of $\widehat{\mathbf{y}}$ is then

$$\widehat{\mathbf{\Sigma}} = E\left[\widehat{\mathbf{y}}\widehat{\mathbf{y}}^T\right] = E\left[ (\widetilde{\mathbf{y}} + \mathbf{n}) (\widetilde{\mathbf{y}} + \mathbf{n})^T \right] = \widetilde{\mathbf{\Sigma}} + \mathbf{N} = \widetilde{\mathbf{\Sigma}} + \frac{q^2}{12}\mathbf{I}. \tag{8}$$

CPPCA will recover both the PCA coefficients as well as the basis vectors of the PCA transform itself from the quantized projections $\widehat{\mathbf{y}}$.

To gauge the effects of quantization on the eigenvector-recovery process of CPPCA, we invoke known perturbation bounds on eigenvectors:

**Theorem 1** *Suppose $\widetilde{\mathbf{u}}_k$ is a unit eigenvector of $\widetilde{\mathbf{\Sigma}}$ with corresponding eigenvalue $\widetilde{\lambda}_k$ such that $\widetilde{\lambda}_k$ is separated from the other eigenvalues by at least $\delta_k$; i.e., $\delta_k = \min_{\widetilde{\lambda}_{k'} \neq \widetilde{\lambda}_k} \left|\widetilde{\lambda}_{k'} - \widetilde{\lambda}_k\right|$. If $q^2/12 \leq \delta_k/5$, then there exists $\widehat{\mathbf{u}}_k$, a unit eigenvector of $\widehat{\mathbf{\Sigma}} = \widetilde{\mathbf{\Sigma}} + (q^2/12)\mathbf{I}$, such that*

$$\sin \zeta_k \leq \frac{q^2}{3\delta_k}, \tag{9}$$

*where $\zeta_k$ is the angle between $\widetilde{\mathbf{u}}_k$ and $\widehat{\mathbf{u}}_k$.*

Pf: See App. A.1 of [7] which in turn derives from Th. 8.1.10 of [8].

We note that the heuristic driving CPPCA in the first place—that Ritz vectors form reasonable approximations to normalized projections of eigenvectors—relies on the distribution of $\mathbf{X}$ being eccentric, i.e., the eigenvalues being sufficiently distinct from one another. We see thus that the same phenomenon that permits eigenvector recovery—mutually distinct eigenvalues—also encourages stability of the result under scalar quantization. The bound of (9) ensures a graceful degradation in the accuracy of the CPPCA eigenvector-recovery procedure as the quantizer stepsize $q$ increases. Furthermore, eigenvector recovery is more robust to quantization distortion as the gap between the eigenvalues and, consequently, the eccentricity of the distribution, increases. In terms of the coefficient-recovery procedure of CPPCA, robustness of (7) is not an issue since least-squares estimation forms the foundation of signal recovery in noise.

## Relation to Compressed Sensing (CS)

In brief, CS (see [3] for an overview) produces a sparse signal representation directly from a small number of projections onto another basis, recovering the sparse transform coefficients via nonlinear reconstruction. CS is similar to CPPCA proposed above in its ability to permit compressive sensors to couple dimensionality reduction with signal acquisition.[1] However, CPPCA differs from CS in that it benefits from a simple linear reconstruction for the transform coefficients.

The main tenet of CS theory holds that, if signal $\mathbf{x} \in \mathcal{R}^N$ can be sparsely represented (i.e., using only $L$ nonzero coefficients) with some basis $\mathbf{\Psi} = \begin{bmatrix}\boldsymbol{\psi}_1 & \cdots & \boldsymbol{\psi}_N\end{bmatrix}$, then we can recover $\mathbf{x}$ from $K$-dimensional projections $\widetilde{\mathbf{y}} = \mathbf{P}^T\mathbf{x}$ under certain conditions; here $\mathbf{P} = \begin{bmatrix}\mathbf{p}_1 & \cdots & \mathbf{p}_K\end{bmatrix}$ and $K < N$. Specifically, it is required that $K$ must be sufficiently large with respect to the sparsity $L$ (but still much less than $N$) and that $\mathbf{\Psi}$ and $\mathbf{P}$ be mutually *incoherent*, meaning that $\mathbf{P}$ cannot sparsely represent the $\boldsymbol{\psi}_n$ vectors. It has been shown that, if $\mathbf{P}$ is chosen randomly, then $\mathbf{P}$ and $\mathbf{\Psi}$ are incoherent for any arbitrary fixed $\mathbf{\Psi}$ with high probability. The ideal recovery procedure searches for the $\check{\mathbf{x}}$ with the smallest $\ell_0$ norm consistent with the observed $\widetilde{\mathbf{y}}$. However, this $\ell_0$ optimization being, in general, NP complete, several alternative solution procedures have been proposed. Perhaps the most prominent of these is basis pursuit (BP) which applies a convex relaxation to the $\ell_0$ problem resulting in an $\ell_1$ optimization,

$$\check{\mathbf{x}}^* = \arg\min_{\check{\mathbf{x}}} \left\|\check{\mathbf{x}}\right\|_1, \quad \text{such that } \widetilde{\mathbf{y}} = \mathbf{P}^T\mathbf{\Psi}\check{\mathbf{x}}. \tag{10}$$

[1] Indeed, there is increasing interest in integrating CS methodology directly into imaging spectrometers so as to implement dimensionality reduction simultaneously with hyperspectral data acquisition (e.g., [9]).

On the surface, CPPCA and CS appear quite similar—encoders in both approaches invoke merely dimensionality-reducing projections while the decoders, constituting the bulk of the computational complexity, reconstruct transform coefficients from the projections. There is, however, a crucial difference between the two. CS reconstruction operates under an assumption of sparsity in a *known* basis $\mathbf{\Psi}$, but the pattern of sparsity (i.e., which $L$ components are nonzero) is *unknown*. On the other hand, CPPCA reconstruction operates under a *known* sparsity pattern (i.e., the first $L$ principal components), but the transform $\mathbf{\Psi}$ itself is *unknown*. This difference requires CPPCA to perform the additional step of basis recovery, but, in return, CPPCA benefits from the simple linear reconstruction of (7) in place of the more complicated nonlinear reconstruction of (10) as typically used by CS.

## Results

Eigenvector recovery in CPPCA is based on the heuristic that Ritz vectors closely approximate the corresponding normalized projections. We now present some empirical evidence that this heuristic works rather well in practice. Let us consider $N \times N$ matrix $\mathbf{\Sigma} = \mathrm{diag}(1000, 100, 10, 1, 1, \ldots, 1)$. Note that the eigenvectors $\mathbf{w}_i$ for this $\mathbf{\Sigma}$ are simply the columns of the $N \times N$ identity matrix. Given these eigenvalues, the heuristic suggests that we should be able to closely approximate normalized projection $\mathbf{v}_1$ with Ritz vector $\mathbf{u}_1$ since $\lambda_1 = 1000$ is large and distant from the other eigenvalues. We should also have some luck with $\mathbf{v}_2$ ($\lambda_2 = 100$) and $\mathbf{v}_3$ ($\lambda_3 = 10$) but likely no luck at all with $\mathbf{v}_4$, whose corresponding eigenvalue is identical to those of a number of the others. We generate 1000 random projections by filling $N \times K$ matrices with values from independent zero-mean unit-variance Gaussian random variables and then orthogonalizing the columns. Fig. 3 illustrates the histograms of the angles $\omega_i$ for $N = 100$ and $K = 40$. Here, $\omega_i$ is the angle between the Ritz vector $\mathbf{u}_i$ and the corresponding normalized projection $\mathbf{v}_i$. We see that, as expected, $\omega_1$ is typically close to $0°$, indicating that, indeed, $\mathbf{u}_1$ is usually very close to $\mathbf{v}_1$. We see that $\omega_2$ and $\omega_3$ also cluster at $0°$, although not as tightly as $\omega_1$. On the other hand, $\omega_4$ actually clusters about $90°$. As expected, we see decreasing success in the approximation as the eigenvalues $\lambda_i$ decrease. For $\lambda_4$, which is the same as $\lambda_5$ through $\lambda_N$ in this example, we see that we are likely to end up with a Ritz vector that is orthogonal to the desired normalized projection $\mathbf{v}_4$.

We now examine the rate-distortion performance of CPPCA using a dataset extracted from a real hyperspectral image consisting of $M = 1000$ spectral pixel vectors from the hyperspectral image "cuprite," an AVIRIS dataset of $N = 224$ spectral bands; the mean vector has been removed from the vectors to impose a zero-mean condition. For CPPCA, we use $J = 20$ projection partitions as our empirical observations indicate a reasonable tradeoff between reconstruction accuracy and computational complexity for this value. We set $L = 5$ since recovery of a greater number of eigenvectors is unreliable. For rate-distortion performance, we apply uniform SQ with a quantizer stepsize of $q$, measuring the entropy of the resulting quantizer indices to estimate the rate that would be obtained via suitable entropy coding; we choose the stepsize from $q \in [0.001, 0.04]$. For CS, we consider several popular transforms: an $N$-point DCT (CS-DCT) as well as orthonormal DWTs using both the Haar basis (CS-Haar) and the length-4 Daubechies basis (CS-D4). We apply the same random projections as used for CPPCA and quantize using the same

In *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. W. Marcellin, Eds., Snowbird, UT, March 2008, pp. 83–92.

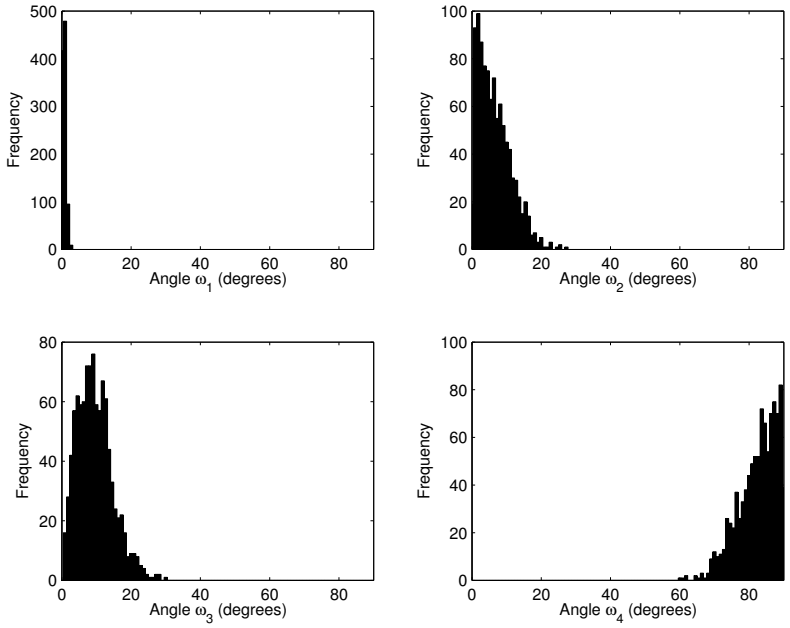



Figure 3: Histogram of angle $\omega_i$ between Ritz vector $\mathbf{u}_i$ and normalized projection $\mathbf{v}_i$ for $i = 1, \ldots, 4$. Average $\omega_1 = 0.7°$; average $\omega_2 = 6.6°$; average $\omega_3 = 9.6°$; average $\omega_3 = 82.4°$.

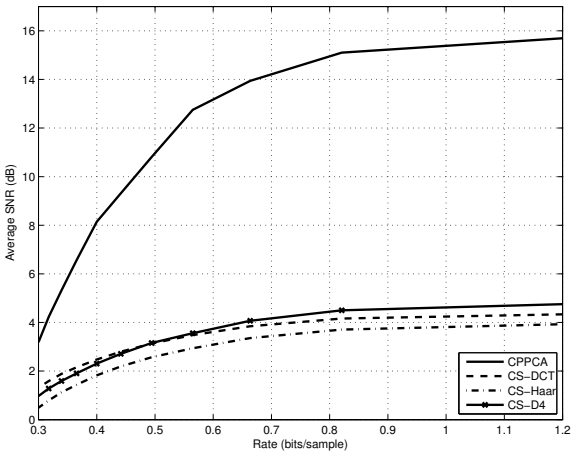
uniform-SQ stepsizes. Once again, we measure rate as entropy. For CS, we employ the BP coefficient recovery of (10) with a quadratically relaxed constraint such as is suited to robust CS with noisy data [3, 10]:

$$\check{\mathbf{x}}^* = \arg\min_{\check{\mathbf{x}}} \|\check{\mathbf{x}}\|_1, \quad \text{such that } \left\|\mathbf{P}^T\mathbf{\Psi}\check{\mathbf{x}} - \widehat{\mathbf{y}}\right\|_2 \leq \epsilon. \tag{11}$$

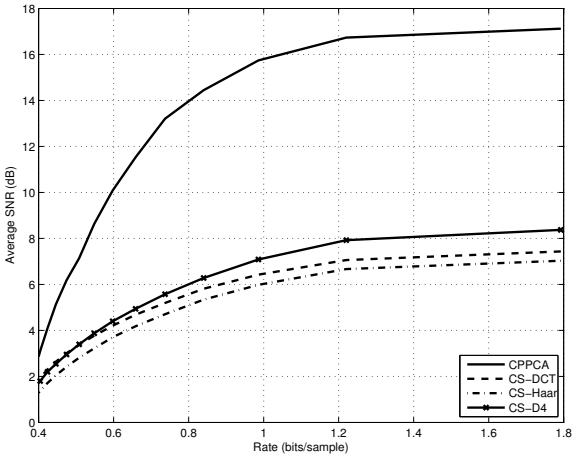
Since the error variance in each component of $\widehat{\mathbf{y}}$ is $q^2/12$, we set the error bound as $\epsilon = q\sqrt{K/12}$, where $q$ is the quantization stepsize. We use $\ell_1$-MAGIC[2] for the implementation of (11).

Clearly, the rate-distortion performance of both CPPCA and CS will depend on not only the quantizer applied, but also the degree of dataset reduction inherent in the projections; this latter quantity is characterized as a relative projection dimensionality in the form of $K/N$, expressed as a percentage. Rate-distortion results for CPPCA as well as the fixed-basis CS-DCT, CS-Haar, and CS-D4 techniques are illustrated in Fig. 4 for several projection dimensionalities. In all cases, CPPCA achieves rate-distortion performance substantially superior to those of all of the CS techniques. In terms of computational complexity, none of the implementations we employ are optimized for execution speed. However, we have observed that both the POCS-based eigenvector recovery of (6) as well as the linear coefficient recovery of (7) are substantially faster than the nonlinear reconstruction of (11), resulting in our CPPCA decoder implementation being about 12 times faster than the corresponding CS decoder.

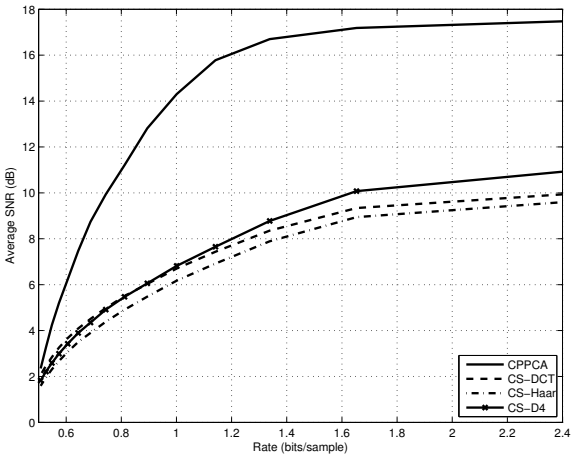
[2] `http://www.acm.caltech.edu/l1magic/`




(a) "Cuprite," $K/N = 20\%$




(b) "Cuprite," $K/N = 30\%$




(c) "Cuprite," $K/N = 40\%$

Figure 4: Rate-distortion performance for the "cuprite" hyperspectral dataset.

## Conclusions

In this paper, we presented an approach that exploits the compressive projections in sensors that integrate dimensionality reduction and signal acquisition to effectively shift the computational burden of PCA from the encoder to the decoder. This CPPCA technique couples random projections at the encoder with a Rayleigh-Ritz process for approximating eigenvectors at the decoder. In its use of random projections, CPPCA can be considered to possess a certain duality with CS—whereas CS recovery can be characterized as finding coefficients of an unknown sparsity pattern in a known basis, CPPCA recovers coefficients of a known sparsity pattern in an unknown basis. Accordingly, CPPCA requires the additional step of a POCS-based eigenvector recovery but benefits from a greatly simplified linear reconstruction as opposed to the nonlinear reconstruction fundamental to CS. Invoking known perturbation bounds, we established that the CPPCA eigenvector-recovery process is robust to quantization noise, and experimental results revealed that CPPCA achieves rate-distortion performance substantially superior to that of CS with a variety of popular bases for a dataset of hyperspectral signature vectors.

## References


[1] Q. Du and J. E. Fowler, "Hyperspectral image compression using JPEG2000 and principal component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 201–205, April 2007.

[2] B. Penna, T. Tillo, E. Magli, and G. Olmo, "A new low complexity KLT for lossy hyperspectral data compression," in *Proceedings of the International Geoscience and Remote Sensing Symposium*, vol. 7, Denver, CO, August 2006, pp. 3525–3528.

[3] 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.

[4] B. N. Parlett, *The Symmetric Eigenvalue Problem*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1998.

[5] P. L. Combettes, "The foundations of set theoretic estimation," *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182–208, February 1993.

[6] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer Academic Publishers, 1992.

[7] I. Johnstone and A. Y. Lu, "Sparse principal component analysis," *Journal of the American Statistical Association*, to appear.

[8] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins University Press, 1996.

[9] R. M. Willett, M. E. Gehm, and D. J. Brady, "Multiscale reconstruction for computational spectral imaging," in *Computational Imaging V*, C. A. Bouman, E. L. Miller, and I. Pollak, Eds. San Jose, CA: Proc. SPIE 6498, January 2007, p. 64980L.

[10] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, August 2006.



In *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. W. Marcellin, Eds., Snowbird, UT, March 2008, pp. 83–92.