

DECODER-SIDE DIMENSIONALITY DETERMINATION FOR COMPRESSIVE-PROJECTION PRINCIPAL COMPONENT ANALYSIS OF HYPERSPPECTRAL DATA

Wei Li and James E. Fowler

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

ABSTRACT

Compressive-projection principal component analysis reconstructs vectors from random projections by recovering an approximation to the principal eigenvectors of the principal-component transform. A heuristic for the number of eigenvectors to approximate is developed to provide consistency with the Johnson-Lindenstrauss lemma and the restricted isometry property from compressed-sensing theory. The resulting heuristic is driven by only quantities known at the reconstruction side of the system. The heuristic is evaluated empirically for hyperspectral imagery and is demonstrated to provide near-optimal reconstruction quality.

Index Terms—CPPCA, dimensionality reduction, random projection, hyperspectral data

1. INTRODUCTION

Principal component analysis (PCA) is widely used in numerous data analysis, compression, and communication applications. However, PCA is a data-dependent transform which is typically determined via a computationally expensive eigendecomposition. This heavy computation burden hinders the deployment of PCA in resource-constrained environments such as satellite-borne sensor platforms. In order to make PCA more amenable to such remote devices, [1] developed a process to shift the computational burden from a resource-constrained sensor to a more powerful base-station decoder. This process, called compressive-projection PCA (CPPCA), is driven by projections at the sensor onto a randomly-chosen lower-dimensional subspace. 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. CPPCA departs from the traditional PCA paradigm in that it allows its excellent dimensionality-reduction and compression performance to be realized in a system that puts the main computational burden on the decoder.

The heart of the CPPCA framework is an approximation of the PCA transform basis via a novel eigenvector-reconstruction procedure based on convex-set optimization

This material is based upon work supported by the National Science Foundation under Grant No. CCF-0915307.

driven by Ritz vectors within the projected subspace. Specifically, in the CPPCA decoder, the first L principal eigenvectors of the PCA transform basis are approximately recovered from the random projections of the data. An open problem not discussed in [1] is how the CPPCA decoder determines how many principal eigenvectors it should attempt to recover from the random projections.

In this paper, we present a simple heuristic for determining the value of L within a CPPCA decoder for use with hyperspectral data. We motivate the choice of our heuristic by invoking several theoretical constructs, including the celebrated Johnson-Lindenstrauss (JL) lemma [2–4] as well as the restricted isometry property (RIP) [5, 6] that underlies the theory of compressed sensing (CS). We validate the proposed heuristic in a battery of empirical evaluations on hyperspectral data that demonstrate the heuristic to be quite effective in its intended task. Although the value of L provided by the heuristic occasionally differs somewhat from the best choice—i.e., the L value that yields the best reconstruction quality in terms of signal-to-noise-ratio (SNR) for the CPPCA reconstruction of a given hyperspectral dataset—the SNR for the heuristically determined L is usually extremely close to the best SNR.

2. CPPCA

We now briefly overview the CPPCA procedure; the reader is referred to [1] for a more complete description. Consider M zero-mean vectors $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M]$, where each $\mathbf{x}_m \in \mathbb{R}^N$. The covariance matrix of \mathbf{X} is $\Sigma = \mathbf{X}\mathbf{X}^T/M$, and PCA seeks a linear transformation \mathbf{W} of eigenvectors emanating from the eigendecomposition of Σ ; i.e.,

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

where \mathbf{W} contains the N unit eigenvectors of Σ column-wise. The PCA transform of \mathbf{X} is then $\tilde{\mathbf{X}} = \mathbf{W}^T\mathbf{X}$.

Suppose we have K orthonormal vectors \mathbf{p}_k that form the basis of K -dimensional subspace \mathcal{P} , such that $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K]$ provides an orthogonal projection onto \mathcal{P} . The CPPCA encoder produces $\tilde{\mathbf{Y}} = \mathbf{P}^T\mathbf{X}$, where $\tilde{\mathbf{Y}} = [\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \dots, \tilde{\mathbf{y}}_M]$ with covariance

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

The CPPCA decoder receives $\tilde{\mathbf{Y}}$ from the encoder and can thus calculate $\tilde{\Sigma}$. The goal of CPPCA is then to approximate \mathbf{W} from $\tilde{\Sigma}$ without knowledge of Σ .

Rayleigh-Ritz theory [7] describes the relation between the eigenvectors Σ and those of $\tilde{\Sigma}$ as given by (2). Covariance matrix Σ has eigenvalues satisfying $\lambda_1(\Sigma) \geq \dots \geq \lambda_N(\Sigma)$, and the corresponding unit eigenvectors are \mathbf{w}_n . On the other hand, the eigendecomposition of $\tilde{\Sigma} = \mathbf{P}^T \Sigma \mathbf{P}$ is $\tilde{\Sigma} = \tilde{\mathbf{U}} \tilde{\Lambda} \tilde{\mathbf{U}}^T$, wherein the columns of $\tilde{\mathbf{U}}$ are the unit eigenvectors $\tilde{\mathbf{u}}_k$, and the eigenvalues are $\lambda_1(\tilde{\Sigma}) \geq \dots \geq \lambda_K(\tilde{\Sigma})$. The K eigenvalues $\lambda_k(\tilde{\Sigma})$ are called *Ritz values*; additionally, there are K vectors, known as *Ritz vectors*, defined as

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

Finally, normalized projection \mathbf{v}_n is defined as the orthogonal projection of \mathbf{w}_n onto \mathcal{P} , normalized to unit length; i.e.,

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

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 $\lambda_k(\Sigma)$ 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 $\lambda_k(\tilde{\Sigma})$. Under the assumption that $\mathbf{u}_k \approx \mathbf{v}_k$, CPPCA reconstruction uses an algorithm based on projections onto convex sets to approximate the first L eigenvectors \mathbf{w}_n from $\tilde{\mathbf{Y}}$; these eigenvectors are assembled into $N \times L$ matrix Ψ , an approximation to the first L columns of the PCA transform \mathbf{W} . The CPPCA decoder then proceeds to recover the PCA coefficients by solving $\tilde{\mathbf{Y}} = \mathbf{P}^T \Psi \tilde{\mathbf{X}}$ for PCA coefficients $\tilde{\mathbf{X}}$ using a linear, least-squares solution based on a pseudoinverse.

As presented in [1], L is essentially a free parameter for the CPPCA reconstruction algorithm. If L is set too small, an insufficient number of eigenvectors will go into Ψ , and the reconstruction $\tilde{\mathbf{X}} = \Psi \tilde{\mathbf{X}}$ will be of relatively poor quality. On the other hand, if L is chosen too large, then it is likely that $\mathbf{u}_k \neq \mathbf{v}_k$ for the larger values of k since the separation between the eigenvalues $\lambda_k(\Sigma)$ tends to decrease dramatically with large k . Again, the reconstruction quality will suffer.

It is thus crucial that the CPPCA decoder select an appropriate value of L . If the decoder had access to Σ , it would be straightforward to set L by examining the relative separation between the eigenvalues $\lambda_k(\Sigma)$. However, the CPPCA decoder knows only the projection operator \mathbf{P} and the projections of the data $\tilde{\mathbf{Y}}$. Below, we develop a simple heuristic for setting L that relies only on the size of \mathbf{P} , namely, N and K .

3. THE JL LEMMA AND THE RIP

In [1], it is noted that CPPCA possesses a certain consistency with the JL lemma. Briefly, the JL lemma holds that L points in \mathbb{R}^N can be projected into a K -dimensional subspace while approximately maintaining pairwise distances as long as K is sufficiently large with respect to L . Specifically, for $\epsilon > 0$ and every set \mathcal{Q} of L points in \mathbb{R}^N , there exists mapping $f : \mathbb{R}^N \rightarrow \mathbb{R}^K$ such that, for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{Q}$,

$$(1-\epsilon) \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 \leq \|f(\mathbf{x}_1) - f(\mathbf{x}_2)\|_2^2 \leq (1+\epsilon) \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2, \quad (5)$$

as long as $K \geq O(\epsilon^{-2} \log L)$. A projection satisfying this lemma is often referred to as a *JL embedding* [4]. In CPPCA, the eigenvectors \mathbf{w}_n are necessarily mutually orthonormal; likewise, the Ritz vectors \mathbf{u}_k are also necessarily orthonormal. Consequently, for the Ritz vectors to be a reasonable approximation to the first L eigenvectors projected into the K -dimensional space (i.e., for $\mathbf{u}_k \approx \mathbf{v}_k$ for $1 \leq k \leq L$), the projections \mathbf{v}_k of the first L eigenvectors \mathbf{w}_k must also be approximately mutually orthonormal. It is observed in [1] that this will be the case if the projection is a JL embedding since pairwise distances and, via the Pythagorean theorem, mutual orthogonalities are approximately preserved by such mappings.

It has been noted that there is a close relation between the JL lemma and a fundamental construct in CS theory known as RIP [5, 6]. In short, while the JL lemma addresses the projection of a limited number of arbitrary signals, the RIP concerns the projection of any sparse signal. Specifically, suppose that \mathbf{x} is S -sparse in the domain of orthonormal $N \times N$ transform Ψ , meaning that only S coefficients in $\tilde{\mathbf{x}} = \Psi^T \mathbf{x}$ are nonzero. The projector operator for which RIP holds will approximately maintain pairwise distances between signals \mathbf{x}_1 and \mathbf{x}_2 which are both S -sparse with respect to transform Ψ . That is, [6] states the RIP condition as

$$(1-\delta) \|\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2\|_2^2 \leq \|\Phi \Psi (\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2)\|_2^2 \leq (1+\delta) \|\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2\|_2^2, \quad (6)$$

where Φ is an orthonormal $K \times N$ measurement matrix, and δ is a small positive constant [6]. Noting that distances are invariant under an orthonormal transform Ψ , we have $\|\tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_2\|_2^2 = \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2$, and (6) becomes

$$(1-\delta) \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 \leq \|\Phi (\mathbf{x}_1 - \mathbf{x}_2)\|_2^2 \leq (1+\delta) \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2. \quad (7)$$

As written in (7), the RIP bears clear resemblance to the statement of the JL lemma in (5); in fact, it is observed in [5] that every satisfactory JL embedding will also provide a projection that satisfies the RIP.

It was determined in [8] that, for the RIP to hold for the orthonormal pair Φ and Ψ , it is sufficient for

$$K \geq \gamma \cdot S (\log N)^4, \quad (8)$$

where γ is some constant. However, it has been conjectured that

$$K \geq \gamma \cdot S \log N \quad (9)$$

is actually sufficient [6].

4. A HEURISTIC FOR THE NUMBER OF CPPCA EIGENVECTORS

We now develop our heuristic for determining the value of L in CPPCA reconstruction of hyperspectral data. We adopt a simple model of the hyperspectral dataset $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M]$ such that each vector \mathbf{x}_i is some linear combination of L vectors $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_L\}$. The vectors \mathbf{e}_l are commonly called *endmembers* of the hyperspectral dataset, and they represent a number of constituent hyperspectral signatures (e.g., corn, soil, grass, water, etc.) from which each pixel vector \mathbf{x}_i is composed. Due to the limited spatial resolution of a hyperspectral image, each pixel vector is not likely to be composed of a single endmember, but rather some linear combination of multiple endmembers due to a “mixing” of several hyperspectral signatures within the spatial extent of a hyperspectral pixel. As a consequence of this L -endmember model, the PCA decomposition of dataset \mathbf{X} consists of L *principal eigenvectors*, $\{\mathbf{w}_1, \dots, \mathbf{w}_L\}$, and each \mathbf{x}_i is representable by a linear combination of these L principal eigenvectors. In this sense, each \mathbf{x}_i is L -sparse with respect to the PCA transform \mathbf{W} from (1) since \mathbf{W} contains the L principal eigenvectors \mathbf{w}_l (in addition, \mathbf{W} contains $N - L$ minor eigenvectors that span the nullspace of the principal eigenvectors).

We now assume that the CPPCA projection \mathbf{P} in (2) forms a satisfactory JL embedding such that the \mathbf{v}_k in (4) are nearly orthonormal, and the Ritz vectors in (3) satisfy $\mathbf{u}_k \approx \mathbf{v}_k$. Then, the RIP will hold in the special case in which we adopt the CPPCA projection operator as the measurement basis ($\Phi = \mathbf{P}^T$) and the CPPCA eigenvector matrix as the sparsity basis ($\Psi = \mathbf{W}^T$); in this case, the sparsity is $S = L$. Even though (8) is only sufficient and not necessary for the RIP to hold, and (9) is merely a conjecture, these relations suggest that

$$K \geq \gamma \cdot L \log N. \quad (10)$$

In other words,

$$L \leq K/\gamma \log N \quad (11)$$

is an upper bound on L that will permit CPPCA reconstruction with L eigenvalues to be consistent with both the JL and RIP constructs. Thus, we propose that

$$L = \text{round}(\alpha K/\log N), \quad (12)$$

provides a suitable heuristic for the number of CPPCA eigenvectors for some constant α ; here, $\text{round}(\cdot)$ denotes rounding to the nearest integer. The parameter α will be investigated empirically in the next section; we find that $\alpha = 1.0$

works well for the datasets we consider. Finally, we note that expressions similar to (12) have been employed for parameter setting in CS reconstructions (e.g., [9, 10]).

5. EXPERIMENTAL RESULTS

In this section, we examine the heuristic proposed above to evaluate how effective it is in setting L for CPPCA reconstruction. According to (12), the heuristic sets L to be directly proportional to $K/\log N$ which is known from the size of the projection matrix \mathbf{P} . We define the subsampling rate, or *substrate*, of \mathbf{P} to be K/N , and we will explore how L varies with substrate. In our empirical investigations, we use hyperspectral images cropped spatially to size of 100×100 . We use the popular AVIRIS images Cuprite, Jasper Ridge, and Moffett (224 spectral bands), as well as a 72-band image from a CASI sensor, which is of a scene surrounding a residential community close to Mississippi State University taken in May 2002.

We first perform an exhaustive search to find the “optimal” value of L for each hyperspectral image at four different substrates; this optimal L is the value for which the SNR is maximized for integer values of L between 1 and 30. These optimal L values and their corresponding SNRs are tabulated in the first rows of Tables 1–4.

In order to derive α in (12), zero-crossing linear regression is performed on the optimal values of L . Table 5 tabulates the resulting α values for various hyperspectral images under consideration. Clearly, α approximately equals 1.0 in all cases.

The value of L as calculated by (12) using the α values in Table 5 is tabulated in Tables 1–4 along with the resulting SNR. It is clear that the SNR as calculated by the proposed heuristic is quite close to the optimal SNR as obtained via exhaustive search. Yet, the heuristic L is not always exactly the same as the optimal L , indicating that the SNR of CPPCA reconstruction is not overly sensitive to the value L to a certain extent. Consider the Jasper Ridge dataset at 0.3 substrate, for example—the L from the heuristic is 14 which is some 50% larger than the optimal value of 9. Nevertheless, the corresponding SNR values differ by only 0.2 dB.

A further simplification of the heuristic arises from setting $\alpha = 1.0$ as Table 5 suggests to be a reasonable approximation. The resulting L and SNR values are again tabulated in Tables 1–4; this simplified heuristic results in nearly the same performance as the original heuristic which uses the exact α from Table 5.

6. CONCLUSIONS

CPPCA provides reconstruction of a dataset of vectors from random projections by recovering an approximation to the first L principal eigenvectors of the PCA transform for the data as well as the corresponding PCA coefficients. Previously, the number L of eigenvectors to recover was a free parameter to the CPPCA reconstruction algorithm. The main

contribution of the work presented here is the development of the simple heuristic, $L = \text{round}(K/\log N)$, wherein N is the dimension of the original signal space, and K is the dimension of the random projections. Experimental results on hyperspectral imagery reveal that the heuristic provides a suitable value for L that results in reconstructions with SNR very close to that of the best possible L value as determined by exhaustive search.

7. REFERENCES

- [1] J. E. Fowler, "Compressive-projection principal component analysis," *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2230–2242, October 2009.
- [2] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," *Contemporary Mathematics*, vol. 26, pp. 189–206, 1984.
- [3] S. Dasgupta and A. Gupta, "An elementary proof of a theorem of Johnson and Lindenstrauss," *Random Structures and Algorithms*, vol. 22, no. 1, pp. 60–65, January 2003.
- [4] D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *Journal of Computer and System Science*, vol. 66, no. 4, pp. 671–687, June 2003.
- [5] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, December 2008.
- [6] 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.
- [7] B. N. Parlett, *The Symmetric Eigenvalue Problem*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1998.
- [8] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," in *Proceedings of the 40th Annual Conference Information Sciences and Systems*, Princeton, NJ, March 2006, pp. 207–212.
- [9] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, December 2007.
- [10] R. Ward, "Compressed sensing with cross validation," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5773–5782, December 2009.

Table 1: SNR and L for Cuprite

subrate	0.20	0.30	0.40	0.50
optimal L	4	8	15	22
SNR (dB)	14.40	15.57	17.26	18.41
heuristic L	7	11	14	18
SNR (dB)	13.78	15.51	17.21	18.41
heuristic $L : \alpha = 1.0$	8	12	17	21
SNR (dB)	13.61	15.45	17.22	18.41

Table 2: SNR and L for Moffett

subrate	0.20	0.30	0.40	0.50
optimal L	5	10	13	21
SNR (dB)	16.67	23.17	25.07	26.92
heuristic L	8	11	15	19
SNR (dB)	16.54	23.13	25.04	26.87
heuristic $L : \alpha = 1.0$	8	12	17	21
SNR (dB)	16.54	23.13	24.95	26.92

Table 3: SNR and L for Jasper Ridge

subrate	0.20	0.30	0.40	0.50
optimal L	7	9	19	28
SNR (dB)	16.55	20.02	21.47	23.54
heuristic L	11	14	18	23
SNR (dB)	16.41	19.81	21.42	23.45
heuristic $L : \alpha = 1.0$	8	12	17	21
SNR (dB)	16.51	19.87	21.42	23.42

Table 4: SNR and L for CASI

subrate	0.20	0.30	0.40	0.50
optimal L	2	3	6	12
SNR (dB)	7.11	9.60	11.06	11.77
heuristic L	3	5	6	8
SNR (dB)	6.89	9.52	11.06	11.58
heuristic $L : \alpha = 1.0$	3	5	7	8
SNR (dB)	6.89	9.52	10.98	11.58

Table 5: Estimation of Parameter α

dataset	bands (N)	value of α
Cuprite	224	0.8545
Moffett	224	0.9219
Jasper Ridge	224	1.0961
CASI	72	0.9191