B-HYCA: BLIND HYPERSPECTRAL COMPRESSIVE SENSING

Gabriel Martín[*]*, José M. Bioucas-Dias*[*] *and Antonio Plaza*[**]

[*]Instituto de Telecomunicações, Instituto Superior Técnico, Lisbon, 1049-1, Portugal.
[**]Hyperspectral Computing Laboratory, University of Extremadura, Caceres, Spain.

ABSTRACT

Compressive Sensing has raised as a very useful way to save costs in the acquisition equipment due to the fact that with this technique we can measure the signal in an already compressed form. This is very interesting in hyperspectral applications due to the large amount of data that the hyperspectral sensors collect, store and transmit to the ground stations. Over the last years many compressive sensing methods have been applied to hyperspectral images, and others have been proposed for exploiting the unique features of this kind of images. Over the last years, many techniques have been proposed to perform compressive sensing in hyperspectral imaging. One of them is the Hyperspectral Coded Aperture (HYCA), which exploits two characteristics of hyperspectral imagery: 1) the hyperspectral vectors belong to a low dimensional subspace, and 2) the data cube components exhibit very high correlation in the spatial and in the spectral domains. However, HYCA requires the knowledge of the subspace in advance, which, very often, may compromise its applicability. In this paper it is presented a new technique similar to HYCA which does not require the knowledge of the subspace in advance; the proposed technique is termed blind HYCA (B-HYCA) and it performs a form of blind hyperspectral compressed sensing.

Index Terms— Hyperspectral imaging, Compressive sensing, Linear spectral unmixing, Blind compressive sensing

1. INTRODUCTION

Hyperspectral imaging spectrometers collect hundreds or thousands of bands (at different wavelength channels) for the same area on the surface of the Earth [1]. Due to the extremely large volumes of data collected by imaging spectrometers, hyperspectral data compression has received considerable interest in recent years [2]. Contrarily to the conventional compression schemes, which first acquire the full data set and then implement some compressing technique, Compressive Sensing (CS) acquires directly the compressed signal [3–5]. Over the last years, many techniques have been proposed to perform compressive sensing in hyperspectral imaging. One of them is the Hyperspectral Coded Aperture (HYCA) [6], which exploits two characteristics of hyperspectral imagery: 1) the hyperspectral vectors belong to a low dimensional subspace [7], and 2) the data cube components exhibit very high correlation in the spatial and in the spectral domains. However, HYCA requires the knowledge of the subspace in advance, which, very often, may compromise its applicability. The aim of this work is the development of a new technique similar to HYCA which does not require the knowledge of the subspace in advance; that is, the proposed technique, termed blind HYCA (B-HYCA), performs a form of blind hyperspectral compressed sensing.

This work was supported by the Portuguese Science and Technology Foundation under Project UID/EEA/50008/2013, Project PTDC/EEI-PRO/1470/2012, and Project SFRH/BPD/94160/2013.

2. HYCA: HYPERSPECTRAL CODED APERTURE

Let $\mathbf{X} \in \mathbb{R}^{n_b \times n_p}$ represent, in matrix format, a hyperspectral image (HSI) with n_b spectral bands and $n_p := n_r n_c$ pixels, where n_r and n_c denote, respectively, the number of rows and columns of the hyperspectral image in the spatial domain. The columns of $\mathbf{X}$ correspond to a column-wise ordering of the spectral vectors, one per image pixel.

Let $\mathbf{y} \in \mathbb{R}^m$ denote the CS measurements modeled as

$$\mathbf{y} = A(\mathbf{X}) + \mathbf{w}, \tag{1}$$

where $A : \mathbb{R}^{n_b \times n_p} \to \mathbb{R}^m$ is a linear operator which computes m inner products between known m vectors and the elements of $\mathbf{X}$ and $\mathbf{w}$ models an additive perturbation, hereafter termed noise, accounting for, *e.g.*, modeling errors and system noise. Since A is a linear operator, then we have $A(\mathbf{X}) = \mathbf{A}\mathbf{x}$, where $\mathbf{x} := \text{vec}(\mathbf{X})$ is the vectorization of matrix $\mathbf{X}$ by stacking its columns and $\mathbf{A} \in \mathbb{R}^{m \times n}$, with $n := n_b n_p$, is the matrix representation of the linear operator A.

The measurement matrix in HYCA has block diagonal structure; that is, $\mathbf{A} = \text{bdiag}(\mathbf{A}_1, \ldots, \mathbf{A}_{n_p})$, where $\mathbf{A}_i \in \mathbb{R}^{q \times n_b}$ acts on the spectral vector $\mathbf{x}_i$ computing q projections. In this way, we obtain qn_p measurements and thus a compression rate of q/n_b. A qualitative justification for this sampling strategy is that it is easy to implement and it allows for the recovery of the piecewise smooth HSIs even if q (the number of measurements per pixel) is lower than the dimension

of the subspace to which the spectral vectors belong. See [6] for further details.

The use of a different sampling matrix for each pixel leads to a considerable complexity of the recovery algorithm. To alleviate this complexity and simultaneously preserve the recovery ability above described, we set $\mathbf{A}_i \in \{\mathbf{H}_1, \ldots, \mathbf{H}_{n_h}\}$, *i.e.*, there are only n_h different matrices $\mathbf{A}_i$. These n_h matrices populate windows of size $ws \times ws$ containing precisely n_h pixels and all with the same distribution pattern. The windows cover the image grid without overlapping.

A fundamental assumption in HYCA is that the spectral vectors $\mathbf{x}_i$, for $i = \ldots, n_p$, live in a known low dimensional subspace. We may then write

$$\mathbf{X} = \mathbf{E}\mathbf{Z}, \tag{2}$$

where $\mathbf{E} \in \mathbb{R}^{n_b \times p}$ is a full column rank matrix, possibly orthogonal, which spans the signal subspace, and $\mathbf{Z} \in \mathbb{R}^{p \times n_p}$ contains the representation coefficients with respect to $\mathbf{E}$. We remark that the above assumption is a very good approximation for most real HSIs, as a result of their high spectral and spatial correlations [7].

Using the vec$(\cdot)$ properties, we have $\mathbf{x} := \text{vec}(\mathbf{X}) = \text{vec}(\mathbf{E}\mathbf{Z}) = (\mathbf{I} \otimes \mathbf{E})\mathbf{z}$, with $\mathbf{z} := \text{vec}(\mathbf{Z})$; it follows then that

$$\mathbf{A}\mathbf{x} = \mathbf{A}(\mathbf{I} \otimes \mathbf{E})\mathbf{z} = \text{bdiag}(\mathbf{A}_1\mathbf{E}, \ldots, \mathbf{A}_{n_p}\mathbf{E})\mathbf{z} = \mathbf{K}\mathbf{z} \tag{3}$$

where

$$\mathbf{K} := \text{bdiag}(\mathbf{A}_1\mathbf{E}, \ldots, \mathbf{A}_{n_p}\mathbf{E}).$$

To recover $\mathbf{z}$, HYCA solves the convex optimization problem

$$\min_{\mathbf{z}} (1/2)\|\mathbf{y} - \mathbf{K}\mathbf{z}\|^2 + \lambda_{TV}\|\mathbf{\Psi}\mathbf{z}\|_1, \tag{4}$$

where $(1/2)\|\mathbf{y} - \mathbf{K}\mathbf{z}\|^2$ is a data fidelity term, $\|\mathbf{\Psi}\mathbf{z}\|_1$ is a form of isotropic total variations [8] applied to the images of coefficients corresponding to the rows of $\mathbf{Z}$, and λ_{TV} is a regularization parameter setting the relative weight between the two terms. The optimization (4) aims at finding piecewise-smooth images of coefficients, promoted by the total variation term, compatible with the measurement vector $\mathbf{y}$.

If the linear mixing model (LMM) is a good approximation for the HSI $\mathbf{X}$ [7], then $\mathbf{E}$ may be the mixing matrix and $\mathbf{z}$ is interpretable as abundance fractions. Because the abundance fractions are nonnegative, the inclusion of the constraint $\mathbf{z} \geq \mathbf{0}$ in (4) improves the conditioning of the underlying inverse problem.

3. BLIND HYCA

As mentioned before, the assumption that the signal subspace (equivalently, the matrix $\mathbf{E}$) is known beforehand imposes limitations to the use of HYCA. To remove these limitations, we address the following optimization problem in which matrix $\mathbf{E}$ is inferred simultaneously with the $\mathbf{z}$:

$$\begin{aligned} &\min_{\mathbf{z},\mathbf{E}} (1/2)\|\mathbf{y} - \mathbf{K}\mathbf{z}\|^2 + \lambda_{TV}\|\mathbf{\Psi}\mathbf{z}\|_1 + \lambda_E\|\mathbf{D}\mathbf{E}\|_F^2 \\ &\text{subject to:} \quad \mathbf{K} = \text{bdiag}(\mathbf{A}_1\mathbf{E}, \ldots, \mathbf{A}_{n_p}\mathbf{E}), \end{aligned} \tag{5}$$

where the term $||\mathbf{X}||_F := \sqrt{\text{trace}(\mathbf{X}\mathbf{X}^T)}$ stands for the Frobenius norm of $\mathbf{X}$, $\mathbf{D}$ is a matrix such that $\mathbf{D}\mathbf{x}$ computes the first order differences of $\mathbf{x}$, and $\lambda_E > 0$ is a regularization parameter setting the relative weight of the term $\|\mathbf{D}\mathbf{E}\|_F^2$. This term promotes smooth basis vectors, which make sense, at least when $\mathbf{E}$ is the mixing matrix and then its columns are the spectral signatures of the endmembers, which are often smooth.

We remark that, owing to the presence of $\mathbf{K}\mathbf{z}$ in the data fidelity term, which involves multiplications between $\mathbf{E}$ and parts of $\mathbf{z}$, the optimization (5) is not convex anymore. In order to find critical points of (5), we use an Alternating Direction Method of Multipliers (ADMM) strategy similar to that of HYCA [6]. Following closely [9], problem (5) is converted into the equivalent form:

$$\begin{aligned} \min_{\mathbf{z},\mathbf{E},\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\mathbf{V}_4,\mathbf{V}_5} \quad & \frac{1}{2}||\mathbf{y} - \mathbf{V}_4\mathbf{v}_1||^2 + \lambda_{TV}\,\|\mathbf{v}_3\|_1 + \lambda_E\|\mathbf{V}_5\|_F^2 \\ \text{subject to} \quad & \mathbf{v}_1 = \mathbf{z} \\ & \mathbf{v}_3 = \mathbf{\Psi}\mathbf{z}, \\ & \mathbf{V}_4 = \text{bdiag}(\mathbf{A}_1\mathbf{E}, \ldots, \mathbf{A}_{n_p}\mathbf{E}).\text{,} \\ & \mathbf{V}_5 = \mathbf{D}\mathbf{E}. \end{aligned} \tag{6}$$

Then, the ADMM iterative scheme is applied to the optimization variables $\mathbf{z}$, $\mathbf{E}$, $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{V}_4$,$\mathbf{V}_5$, as detailed in [9]. We remark that, the convergence of B-HYCA is not guaranteed as the optimization is nonconvex. We have observed, however, systematic convergence as far ar as the ADMM parameter weighting the quadratic augmented Lagrangian terms is large enough.

4. RESULTS

This section presents B-HYCA reconstruction results obtained with synthetic and and real data. The quality of the reconstructions is assessed with the Normalized Mean Squared Error given by

$$\text{NMSE} := ||(\widehat{\mathbf{X}} - \mathbf{X})||_F^2 / ||\mathbf{X}||_F^2, \tag{7}$$

where $\mathbf{X}$ and $\widehat{\mathbf{X}}$ denote the true and the estimated HSIs, respectively.

The size of the window linked with the direct operator $\mathbf{K}$ was set to $ws = 55$ so that $n_h = 3025$. The parameters λ_{TV} and λ_E were empirically optimized to minimize NMSE between the original and the reconstructed data. Matrix $\mathbf{E}$ was initialized randomly, and the variable $\mathbf{V}_4^0$ was initialized as $\mathbf{V}_4^0 \leftarrow \mathbf{A}\mathbf{E}$.

Table 1. Simulated data. Average NMSE computed from 10 Monte-Carlo runs.

	SNR=30 dB	SNR=40 dB	SNR=50 dB	SNR=∞
$q = 5$	$2.87 \cdot 10^{-4}$	$0.44 \cdot 10^{-4}$	$0.14 \cdot 10^{-4}$	$0.14 \cdot 10^{-4}$
$q = 3$	$6.43 \cdot 10^{-4}$	$4.87 \cdot 10^{-4}$	$4.66 \cdot 10^{-4}$	$2.55 \cdot 10^{-4}$

Table 2. Real data. Average NMSE comuted from 10 Monte-Carlo runs.

	$q = 5$	$q = 7$	$q = 9$	$q = 11$	$q = 13$	$q = 15$	$q = 17$	$q = 19$
NMSE	$2.78 \cdot 10^{-4}$	$2.03 \cdot 10^{-4}$	$1.71 \cdot 10^{-4}$	$1.50 \cdot 10^{-4}$	$1.36 \cdot 10^{-4}$	$1.28 \cdot 10^{-4}$	$1.19 \cdot 10^{-4}$	$1.14 \cdot 10^{-4}$

4.1. Synthetic data

A synthetic dataset was generated from spectral signatures randomly selected from the United States Geological Survey (USGS)[1]. The simulated image consist of a set of 5×5 squares of 10×10 pixels each one, for a total size of 110×110 pixels. The first row of squares contains the endmembers, the second row contains mixtures of two endmembers, the third row contains mixtures of three endmembers, and so on. Fig. 1 displays the ground-truth abundances maps used for generating the simulated imagery. The synthetic scenes were degraded with additive Gaussian iid noise with different signal-to-noise ratios (SNRs) defined as SNR $= 10\log_{\mathbf{10}} \frac{\mathbb{E}\|\mathbf{X}\|_F^2}{\mathbb{E}\|\mathbf{W}\|_F^2}$, where $\mathbb{E}$ denotes mean value.

The number of measurements per pixel, q , and the SNR take values in $q \in \{3, 5\}$ and SNR $\in \{30, 40, 50, \infty\}$ dB. Table 1 shows the NMSE results, from which we conclude that B-HYCA yields reconstructions of good quality even in noisy cases and with $q < p$.

4.2. Real data

In this experiment, we use the well-known AVIRIS Cuprite data set, available online in reflectance units after atmospheric correction. The portion used in experiments corresponds to a 240 by 180 pixels subset. The bands affected by water absorption and low SNR were removed, leaving a total of $n_b = 188$ bands. In this scene, the number of endmembers estimated with Hysime algorithm [10] is $p = 14$. This information led us to test B-HYCA with a number of measurements per pixel in the set $q \in \{5, 7, 9, 11, 13, 15, 17, 19\}$. Table 2 shows the results in terms of NMSE computed from 10 monte carlo runs. We conclude that B-HYCA yields reconstruction of good quality even when q is considerably smaller than the true subspace dimension. In addition, the reconstruction errors are of same order as those provided by HYCA (see [6]).

5. CONCLUDING REMARKS

In this work we have introduced B-HYCA, a new hyperspectral compressed sensing methodology which learns the signal subspace and simultaneously estimates the original HSI. B-HYCA is inspired by the HYCA method [6] with the additional capability of estimating the signal subspace. The experimental results indicate that the reconstructed data is as accurate as the one provided by the original (nonblind) HYCA method. Further work will be developed in order to set the algorithm parameters in a blind or semi-supervised way.

[1] http://speclab.cr.usgs.gov/spectral-lib.html

6. REFERENCES

[1] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for Earth remote sensing," *Science*, vol. 228, pp. 1147–1153, 1985.

[2] B. Huang, *Satellite data compression*, Berlin: Springer, 2011.

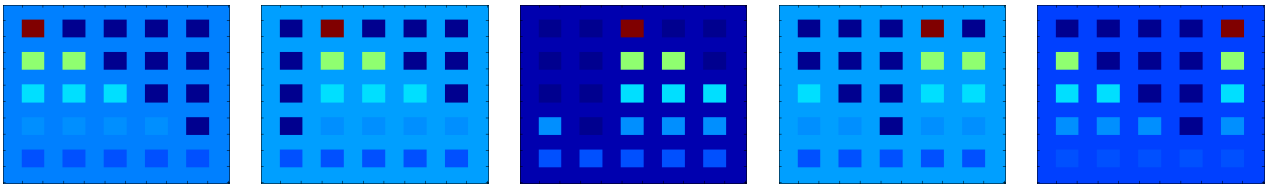
[3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] J. E. Candès, T. Romberg, and Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207, 2006.

[5] M. B. Wakin M. F. Duarte D. Baron S. Sarvotham K. F. Kelly D. Takhar, J. N. Laska and R. G. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," *Computational Imaging IV*, vol. 6065, pp. 43–52, 2006.

[6] G. Martin, J. M Bioucas-Dias, and A. Plaza, "Hyca: A new technique for hyperspectral compressive sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2819–2831, 2015.

[7] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing: geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.



(a) Endmember #1 (b) Endmember #2 (c) Endmember #3 (d) Endmember #4 (e) Endmember #5

Fig. 1. True abundance maps of endmembers in the synthetic hyperspectral data.

.

[8] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical imaging and vision*, vol. 20, no. 1-2, pp. 89–97, 2004.

[9] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.

[10] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, 2008.