HYCA: A New Technique for Hyperspectral Compressive Sensing

Gabriel Martín, José M. Bioucas-Dias, *Member, IEEE*, and Antonio Plaza, *Senior Member, IEEE*

***Abstract*—Hyperspectral imaging relies on sophisticated acquisition and data processing systems able to acquire, process, store, and transmit hundreds or thousands of image bands from a given area of interest. In this paper, we exploit the high correlation existing among the components of the hyperspectral data sets to introduce a new compressive sensing methodology, termed hyperspectral coded aperture (HYCA), which largely reduces the number of measurements necessary to correctly reconstruct the original data. HYCA relies on two central properties of most hyperspectral images, usually termed data cubes: 1) the spectral vectors live on a low-dimensional subspace; and 2) the spectral bands present high correlation in both the spatial and the spectral domain. The former property allows to represent the data vectors using a small number of coordinates. In this paper, we particularly exploit the high spatial correlation mentioned in the latter property, which implies that each coordinate is piecewise smooth and thus compressible using local differences. The measurement matrix computes a small number of random projections for every spectral vector, which is connected with coded aperture schemes. The reconstruction of the data cube is obtained by solving a convex optimization problem containing a data term linked to the measurement matrix and a total variation regularizer. The solution of this optimization problem is obtained by an instance of the alternating direction method of multipliers that decomposes very hard problems into a cyclic sequence of simpler problems. In order to address the need to set up the parameters involved in the HYCA algorithm, we also develop a constrained version of HYCA (C-HYCA), in which all the parameters can be automatically estimated, which is an important aspect for practical application of the algorithm. A series of experiments with simulated and real data shows the effectiveness of HYCA and C-HYCA, indicating their potential in real-world applications.**

***Index Terms*—Coded aperture, compressive sensing (CS), hyperspectral imaging, optimization, signal subspace, total variation (TV).**

Manuscript received January 26, 2014; revised August 9, 2014; accepted October 18, 2014. This work work was supported by the European Community's Marie Curie Research Training Networks Program under Contract MRTN-CT-2006-035927, Hyperspectral Imaging Network (HYPER-I-NET); by the Portuguese Science and Technology Foundation under Project PEst-OE/EEI/LA0008/2013, Project PTDC/EEI-PRO/1470/2012, and Project SFRH/BPD/94160/2013; and by the CEOS-SPAIN (reference AYA2011-29334-C02-02).

G. Martín and J. M. Bioucas-Dias are with the Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, 1649-004 Lisboa, Portugal.

A. Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres E-10071, Spain.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2014.2365534

I. 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]. For instance, the NASA Jet Propulsion Laboratory's Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) covers the wavelength region from 0.4 to 2.5 μm using 224 spectral channels at nominal spectral resolution of 10 nm [2]. The resulting multidimensional data cube[1] typically comprises several gigabytes per flight.

Due to the extremely large volumes of data collected by imaging spectrometers, hyperspectral data compression has received considerable interest in recent years [3], [4]. These data are usually acquired by a satellite or an airborne instrument and sent to a ground station on Earth for subsequent processing. Usually the bandwidth connection between the satellite/airborne platform and the ground station is reduced, which limits the amount of data that can be transmitted. As a result, there is a clear need for (either lossless or lossy) hyperspectral data compression techniques that can be applied onboard the imaging instrument [5]–[7].

A. Proposed Approach and Contributions

In this paper, we develop a new compressive sensing (CS) framework [8], [9] for hyperspectral images, termed hyperspectral coded aperture (HYCA), which exploits two characteristics of hyperspectral images:

(I) Hyperspectral vectors generally belong to a low-dimensional subspace.

(II) The data cube components display very high correlation in both the spatial and the spectral domain.

These two characteristics imply that the hyperspectral data cubes are compressible, i.e., they admit a representation in a given base or frame, in which most of the coefficients are small and, thus, the data can be well approximated with just a small number of large coefficients. In turn, conventional compression techniques based on transform coding are subject to potential drawbacks (e.g., all transform coefficients must be computed, and the locations of the large coefficients must be encoded, thus introducing an overhead).

Compressibility, or sparsity,[2] is a necessary condition for the success of CS. In our approach, and having in mind

[1]The concept of a cube comes from envisaging the hyperspectral image as a stack of square images, one per channel.

[2]A vector is k-sparse if only k of its components are different from zero.

the properties (I) and (II), we represent the spectral vectors using a basis of the signal subspace and particularly model the high spatial correlation present in hyperspectral scenes by promoting small local differences on the images of coefficients by minimizing their 2-D total variation (TV) [10]. Under the linear mixing model assumption that has been widely used in hyperspectral imaging [11], if the we use the spectral signatures of the pure spectral components (called *endmembers*) to represent the spectral vectors, then, the representation coefficients are the abundance fractions of the pure materials. This way, the proposed approach is strongly connected with spectral unmixing. To be more precise, and assuming that we use the spectral signatures of the endmembers to represent the spectral vectors, our methodology implements hyperspectral unmixing in addition to hyperspectral CS.

1) Measurement Strategy: Contrarily to conventional compression schemes, which first acquire the full data set and then implement some compressing technique, CS acquires directly the compressed signal. This is achieved by computing inner products, also termed measurements, between known vectors and the original data. This process is sometimes called "coded aperture" because we can conceive the inner products as the total light that is transmitted trough masks acting on the aperture of the instrument. The single-pixel camera [12] is a paradigmatic illustration of this concept. The number of measurements is usually much smaller than the number of components of the original data. In our approach, we compute for each spectral vector a few random projections using Gaussian independent identically distributed (i.i.d.) vectors. Among the many possible strategies to obtain the measurements from hyperspectral data cubes, ours has two desirable characteristics: 1) it decouples the spectral and spatial domains, which greatly simplifies the reconstruction algorithms; and 2) the use of Gaussian i.i.d. vectors to compute the inner products yields incoherent CS matrices, a desirable property for the success of data recovery [9].

2) Reconstruction Algorithm: In order to reconstruct the original data from the compressed measurements, we minimize a convex objective function containing a quadratic data misfit term and the 2-D TV regularizer. The minimization of this objective function is a hard optimization problem owing to its large-dimensionality and the presence of nonsmooth terms. To solve it, we adopt the alternating direction method of multipliers (ADMMs) introduced in [13], which allows to convert a hard optimization problem into a cyclic sequence of simpler problems.

B. Related Work

The application of CS to hyperspectral images is an active area of research, both in terms of the hardware and the signal processing algorithms [14]–[19]. Works [14], [15] introduce two variants of the coded aperture snapshot spectral imager (CASSI). In [14], the measurement of the input scene is equivalent to projective measurement in the spectral domain. The reconstruction algorithm computes an optimal solution to this CS underdetermined problem using the expectation maximization algorithm, combined with a wavelet-based denoising technique. In [15] the measurement is a sum over the wavelength dimension of a mask-modulated and later sheared data cube. The reconstruction algorithm solves an $\ell_2 - \ell_1$ optimization problem to determine the wavelet coefficients of the original data cube.

The CASSI systems take only a single snapshot from which the original hyperspectral image is inferred. In most cases, even assuming that the reconstructed image is sparse in some basis or frame, this inference is a highly underdetermined and ill-posed inverse problem. Aiming at improving the conditioning of that inverse problem, this paper [17] adopts the linear mixing model and proposes a joint segmentation and reconstruction of the original dual disperser CASSI measurements described in [14]. In addition, the mixing matrix containing the signatures of the spectral endmembers is also estimated.

In [18], the authors reconstruct the hyperspectral data cube by minimizing a convex functional that penalizes both the trace norm and the sum TV norms of the all image bands. These two regularizers promote, respectively, low-rank and piecewise smoothness on the reconstructed data cube. The measurement matrix acts independently over the channels.

Last but not least, the works [16], [19] introduce a methodology with similarities to the one described in this work, but with two major differences: our subspace representation is more flexible, in the sense that we just need to infer a basis for it and not necessarily the mixing matrix. On the other hand, our measurement matrix acts on the spectral domain, whereas [16] acts on the spatial domain. This has strong implications on the reconstruction algorithm and in the quality of the reconstructions. The main ideas of our presented technique were discussed in [20] and [21]. Here, we significantly expand these ideas to present in full detail our new CS-based approach for hyperspectral data compression.

C. Paper Organization

The remainder of this paper is organized as follows. Section II formulates the problem. Section III describes a new CS methodology, termed HYCA and its constrained version (C-HYCA), in which all the parameters can be automatically estimated. Section IV presents a series of experiments with simulated and real data intended to show the effectiveness of HYCA and C-HYCA in real-world applications. Finally, Section V concludes the paper with some remarks and hints at plausible future research lines.

II. Problem Formulation

Let $\mathbf{X} \in \mathbb{R}^{n_b \times n_p}$ represent, in matrix format, a hyperspectral image with n_b spectral bands and $n_p := n_r \times n_c$ 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 $\mathbf{x}^T$ represent the transposed of $\mathbf{x}$, $\|\mathbf{x}\|$ represent the standard Euclidean norm of $\mathbf{x}$, $\|\mathbf{X}\|_F := \sqrt{\text{trace}\{\mathbf{X}\mathbf{X}^T\}}$ represent the Frobenius norm of the matrix $\mathbf{X}$, and $\mathbf{I}$ represent the identity matrix of suitable size. With all

these definitions in place, the CS measurements vector $\mathbf{y} \in \mathbb{R}^m$ is 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 that computes m inner products between m known vectors and the elements of $\mathbf{X}$ and $\mathbf{w}$ model additive perturbation, hereinafter 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 $\mathbf{x} := \mathrm{vec}(\mathbf{X})$ is the vectorization of matrix $\mathbf{X}$ obtained by stacking its columns and $\mathbf{A} \in \mathbb{R}^{m \times n}$, with $n := n_b \times n_p$, is the matrix representation of the linear operator A.

The objective of CS is to recover $\mathbf{x}$ from $\mathbf{y}$ with $m \ll n$, in order to have compression in the acquisition. Without any further information, this recovering is impossible even in the absence of noise because the matrix $\mathbf{A}$ is undetermined. If, however, vector $\mathbf{x}$ admits a sparse representation with respect to a given frame $\mathbf{\Phi}$, i.e., $\mathbf{x} = \mathbf{\Phi}\boldsymbol{\theta}$ with $\boldsymbol{\theta}$ being sparse, then the solution of the optimization problem

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_0 \quad \text{subject to: } \|\mathbf{y} - \mathbf{A}\mathbf{\Phi}\boldsymbol{\theta}\| \leq \delta \tag{2}$$

yields, in given conditions, a good approximation of $\mathbf{x}$. In (2), the notation $\|\mathbf{x}\|_0$, abusively termed the ℓ_0 norm of $\mathbf{x}$, denotes the number on non-null components of $\mathbf{x}$ and $\delta \geq 0$ is a parameter depending on the "size" of the noise.

For $\delta = 0$, if the system of linear equations $\boldsymbol{y} = \mathbf{A}\mathbf{\Phi}\boldsymbol{\theta}$ has a solution satisfying $2\|\boldsymbol{\theta}\|_0 < \mathrm{spark}(\mathbf{A}\mathbf{\Phi})$, where $\mathrm{spark}(\mathbf{A}) \leq \mathrm{rank}(\mathbf{A}) + 1$ is the smallest number of linearly dependent columns of $\mathbf{A}\mathbf{\Phi}$, it is necessarily the unique solution of (2) [22], [23]. For $\delta > 0$, the concept of uniqueness of the sparsest solution is now replaced with the concept of stability [24]–[26]. For example, in [25], it is shown that, given a solution vector $\boldsymbol{\theta}$ of the noiseless ($\delta = 0$) problem (2), satisfying the sparsity constraint $\|\boldsymbol{\theta}\|_0 < (1/\mu(\mathbf{A}\mathbf{\Phi}) + 1)/2$, where $\mu(\mathbf{A}\mathbf{\Phi})$ is the mutual coherence of matrix $\mu(\mathbf{A}\mathbf{\Phi})$, then $\boldsymbol{\theta}$ is unique and any solution $\boldsymbol{\theta}^\delta$ of (2) for a given $\delta > 0$ such that $\|\mathbf{y} - \mathbf{A}\mathbf{\Phi}\theta^\delta\| \leq \delta$ satisfies

$$\|\boldsymbol{\theta}^\delta - \boldsymbol{\theta}\|^2 \leq \frac{4\delta^2}{1 - \mu(\mathbf{A}\mathbf{\Phi})(2\|\boldsymbol{\theta}\|_0 - 1)}.$$

A. Optimization Strategies

The problem (2) is NP-hard [27], and therefore, there is little hope in solving it in a straightforward way. Currently, there are two possible approaches to solve (2). The first one is given by convex relaxation methods, and the second one is given by approximation algorithms. Examples of convex relaxation methods are basis pursuit (BP) and BP denoising [28], which replace the ℓ_0 norm with the ℓ_1 norm in order to solve the problem. Another example of a convex relaxation method is the well-known least absolute shrinkage and selection operator (LASSO) method [29]. On the other hand, approximation algorithms try to solve the $\ell_2 - \ell_0$ problem directly. In this group, we can mention the Bayesian CS [30], iterative signal recovery from incomplete and inaccurate samples [31], iterative hard thresholding [32], gradient descent sparsification [33], and hard thresholding pursuit (HTP) [34]. The solution of our problem is guaranteed if the restricted isometric property is satisfied, which happens if there is a restricted isometric constant $\sigma_s(\mathbf{A})$, which satisfies the following inequality (see [34] and [35] for details):

TABLE I
RATIOS BETWEEN t AND σ_* FOR DIFFERENT ALGORITHMS IN THE LITERATURE (REPRODUCED FROM [34])

Algorithm	IHT	GDS	CoSaMP	HTP	BP
$\sigma_t < \sigma$	$\sigma_{3s} < 0.5$	$\sigma_{2s} < 0.3333$	$\sigma_{4s} < 0.384$	$\sigma_{3s} < 0.5777$	$\sigma_{2s} < 0.465$
$(t/s)/\sigma_*^2$	12	18	27.08	9	9.243

$$(1 - \sigma_s)\|\mathbf{x}\|^2 \leq \|\mathbf{A}\mathbf{x}\|^2 \leq (1 + \sigma_s)\|\mathbf{x}\|^2 \tag{3}$$

where s denotes the level of sparsity of $\mathbf{x}$, so that $\|\mathbf{x}\|_\mathbf{0} \leq \mathbf{s}$. Many of the aforementioned algorithms guarantee the exact recovery for some values of $\sigma_t(\mathbf{A}) \leq \sigma_*$, for a given value of t related to s and for some specific value of σ_*. In the specific case of i.i.d. random matrices, the condition $\sigma_t(\mathbf{A}) \leq \sigma_*$ holds true provided that the number of measurements scale as

$$m \approx c\frac{t}{\sigma_*^2}\log(n/t) \tag{4}$$

for a given constant c [34]. Since our aim is to make as few measurements as possible, we may heuristically assess a sufficient condition by the smallness of the ratio $(t/s)/\sigma_t^2$. Table I (reproduced from [34]) shows this ratio for the different algorithms used to solve the problem. Here, we can see that although the HTP algorithm provides the best results, the BP algorithm is still competitive. In addition, there are some works, which have demonstrated that relaxation methods solving the $\ell_2 - \ell_1$ problem perform much better than the approximation algorithms in the specific case of hyperspectral data sets.

In this paper, we use convex relaxation via analysis-based regularization. In other words, we use the ℓ_1 norm instead of the ℓ_0 due to the fact that it is computationally simpler to solve the problem with ℓ_1 than with ℓ_0. Specifically, the problem that we aim to solve is the following one:

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{subject to: } \|\mathbf{y} - \mathbf{A}\mathbf{\Phi}\boldsymbol{\theta}\| \leq \delta. \tag{5}$$

Note that this problem is in reality a synthesis one due to the fact that, in this case, we are synthesizing $\mathbf{x} = \mathbf{\Phi}\boldsymbol{\theta}$ from the coefficients of θ. However, there is empirical evidence of the superiority of analysis-based approaches [36]–[38]. A synthesis prior represents the sought signal as a weighted sum of *atoms*. On the other hand, an analysis prior models the coefficients obtained by applying the forward transform to the signal. In [37], analysis and synthesis ℓ_1-norm regularization with overcomplete transforms are compared, and evidence is given for the superiority of algorithms based on analysis instead of synthesis. As a result, we adopt an analysis-based approach, i.e., we use an analysis operator $\mathbf{\Psi}$ to transform our synthesis problem into an analysis problem, so that the transformed coefficients $\mathbf{\Psi}\mathbf{x}$ are sparse, thus obtaining

$$\min_{\mathbf{x}} \|\mathbf{\Psi}\mathbf{x}\|_1 \quad \text{subject to: } \|\mathbf{y} - \mathbf{A}\mathbf{x}\| \leq \delta. \tag{6}$$

Specifically, we use 2-D TV regularization [10] so that the $\boldsymbol{\Psi}$ operator computes the horizontal and vertical differences, as indicated in Section II-D.

B. Measurement Strategy

Measuring instruments generally acquire data in an uncompressed form and then perform compression for improved storage or communication. However, CS involves the acquisition of the data in an already compressed way, reducing the amount of data that needs to be measured in the first place. In order to compress the data at the acquisition time, CS computes inner products (also termed measurements) between known vectors and the original data, so that the number of measurements obtained is less than the number of measurements for the original data. There are many possible strategies for performing CS in hyperspectral imaging problems. For instance, we can apply CS in the spectral domain by performing inner products to the image pixels. Another possibility is to perform CS in the spatial domain by performing inner products to the image bands [19], [16]. Finally, we can also develop a hybrid strategy by performing CS in both the spectral and the spatial domain. In this paper, we perform CS in the spectral domain, which means that the inner products are related to the image pixels. In our approach we use matrices with Gaussian i.i.d. entries (Bernoulli i.i.d. entries would work as well [39]) for the compression. Specifically, we use different compression matrices $\mathbf{A}_i \in \mathbb{R}^{q\times n_b}$ for each pixel, where q is the number of measurements to be performed at each pixel. This way, we can express our compression matrix as a block diagonal one $\mathbf{A}$, which contains the matrices $\mathbf{A}_i$ used to perform CS measurements at each pixel

$$\mathbf{A} := \text{bdiag}(\mathbf{A}_1, \ldots, \mathbf{A}_{n_p}) \tag{7}$$

where $\text{bdiag}(\cdot)$ stands for block diagonal. The different matrices $\mathbf{A}_i \in \mathbb{R}^{q\times n_b}$ act on each spectral vector $\mathbf{x}_i$ by computing q projections. This way, we obtain $q \times n_p$ measurements and a compression ratio of n_b/q.

It is worth noting that the collection of matrices $\mathbf{A}_1, \ldots, \mathbf{A}_{n_p}$ have $q \times n_b \times n_p$ elements, which corresponds to q data cubes. Owing to the stringent hardware requisites of the typical onboard systems, the use and the manipulation of such amount of data would raise serious problems. In order to address this issue, instead of using a different matrix $\mathbf{A}_i$ for each pixel, we use a small subset of matrices $\mathbf{H}_i$, so that $\mathbf{A}_i \in \{\mathbf{H}_1, \ldots, \mathbf{H}_{n_h}\}$, i.e., we will have only n_h different matrices $\mathbf{A}_i$.

As aforementioned, we perform CS in the spectral domain. In order to exploit the usual high correlation of hyperspectral data cubes in the spatial domain, we select the compression matrices $\mathbf{A}_i \in \{\mathbf{H}_1, \ldots, \mathbf{H}_{n_h}\}$, as illustrated in Fig. 1: the data cube is split into nonoverlapping square windows of size $n_h := ws \times ws$, where ws denotes the window size, and the CS measurements on each pixel in the window are performed via matrices $\mathbf{H}_1, \ldots, \mathbf{H}_{n_h}$. Given that inside each window, the spectral vectors tend to be piecewise smooth, and therefore most of them are similar, the proposed measurement strategy yields a number of independent measures per spectral vector larger than q. More details on this issue will be provided in Section II-E.

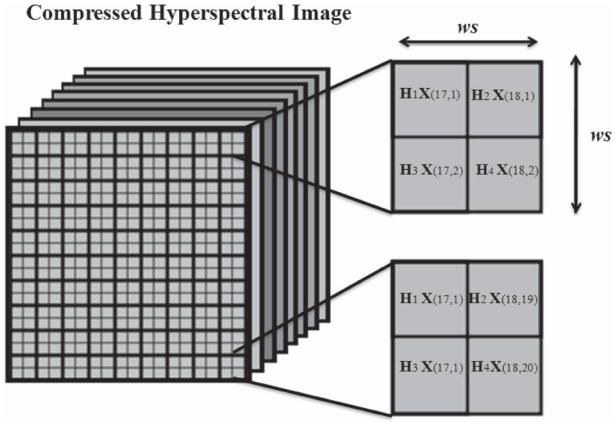

Fig. 1. Measurement strategy illustrated using a toy hyperspectral data cube.

C. Exploiting the Low Dimensionality of the Hyperspectral Data

Hyperspectral imagers collect the reflectance of the surface of the Earth at different "narrow" channels; thus, the resulting bands of the hyperspectral data cube are highly correlated. The high correlation of the hyperspectral bands make the data sets live systematically in low-dimensional subspaces [40]. We can model this situation using the following expression:

$$\mathbf{X} = \mathbf{EZ} \tag{8}$$

where $\mathbf{E} \in \mathbb{R}^{n_b\times p}$ is any full column rank matrix, possibly orthogonal, which spans the signal subspace. It is important to note that, owing to the smoothness of the spectral vectors, model (8) is a good approximation, with $p \ll n_b$, even in the presence of nonlinear mixing effects. Of course, if the linear mixing model is a good approximation for the cube under consideration, then $\mathbf{E}$ may be taken as the mixing matrix containing a number columns equal to the number of endmembers [11]. More on this in the following.

In this paper, we work under the assumption that $\mathbf{E}$ is known or may be estimated from the original data. In the later case, we assume that the sensor acquires the complete image cube $\mathbf{X}$, and then, we can run a fast algorithm for estimating $\mathbf{E}$ and compute the CS measurements. We are aware that we are not implementing a canonical CS strategy. We believe, however, that our line of attack is feasible and effective in most real applications. In fact, if we assume that the bottleneck is mainly in the data transmission stage, the application of CS techniques makes perfect sense in this scenario, since we can send $m = q \times n_p$ samples with $q < p \ll n_b$.

The main advantage of working in a low-dimensional subspace instead of with the original data cube is the fact that, in the former case, we have $p \times n_p$ optimization variables instead of $n = n_b \times n_p$. As aforementioned, in hyperspectral data the spectral bands are often highly correlated; thus, the size of the subspace p is much lower than the original number of bands $p \ll n_b$.

If we assume that the linear mixture model is valid in our case, we have that $\mathbf{E}$ is the mixing matrix containing the endmembers of the image by columns, and the coefficients $\mathbf{Z}$ are the fractional abundances maps associated to the endmembers

[11]. During the last few years, there have been many advances in the development of endmember extraction techniques onboard the imaging instrument (in other words, the trend is to have the capacity onboard to compute the mixing matrix); thus, it is reasonable to assume that we can estimate $\mathbf{E}$ before compressing the data. As a result, when we reconstruct the coefficients $\mathbf{Z}$ on the Earth station, we are performing unmixing and CS at the same time. In the unmixing process, usually the nonnegativity constraint is enforced to impose the abundances to be greater or equal than zero, i.e., $\mathbf{Z} \geq 0$.

D. Proposed Recovery Criteria

Given that $\mathbf{x} = \text{vec}(\mathbf{X}) = \text{vec}(\text{EZ}) = (\mathbf{I} \otimes \mathbf{E})\mathbf{z}$, with $\mathbf{z} := \text{vec}(\mathbf{Z})$, then we have

$$\mathbf{Ax} = \mathbf{A}(\mathbf{I} \otimes \mathbf{E})\mathbf{z} \tag{9}$$
$$= \text{bdiag}(\mathbf{A}_1\mathbf{E}, \ldots, \mathbf{A}_{n_p}\mathbf{E})\mathbf{z} \tag{10}$$
$$= \mathbf{Kz} \tag{11}$$

where

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

Similarly to criteria (6), we propose the following convex optimization problem to recover $\mathbf{z}$:

$$\min_{\mathbf{x}} \|\mathbf{\Psi z}\|_1 \quad \text{subject to: } \|\mathbf{y} - \mathbf{Kz}\| \leq \delta. \tag{12}$$

Our regularizer $\|\mathbf{\Psi z}\|_1$ is the sum of TVs of the p images of coefficients $\mathbf{Z}$

$$\|\mathbf{\Psi z}\|_1 = \sum_{i=1}^{p} \text{TV}(\mathbf{Z}^i) \tag{13}$$
$$:= \text{TV}(\mathbf{z}) \tag{14}$$

where $\mathbf{Z}^i$ is the ith image of representation coefficients with respect to matrix $\mathbf{E}$. Therefore, $\mathbf{\Psi}$ is the discrete gradient operating over the images $\mathbf{Z}^i$. By minimizing $\text{TV}(\mathbf{z})$, we are promoting piecewise-smooth images of coefficients. When $\mathbf{E}$ is the mixing matrix, we can interpret the aforementioned results in terms of abundance fractions.

A convenient way to write $\text{TV}(\mathbf{z})$, in terms of the algorithms introduced in the next section, is as follows:

$$\text{TV}(\mathbf{z}) = \phi(\mathbf{Dz})$$

where $\mathbf{D} := [\mathbf{D}_h^T \ \mathbf{D}_v^T]^T$ and $\mathbf{D}_h$, $\mathbf{D}_v$ denote $(pn_p) \times (pn_p)$ matrices computing, respectively, the horizontal and vertical backward differences, assuming a cyclic boundary; i.e., $\mathbf{D}_h\mathbf{z} := [\mathbf{D}_{h(i,j)}(\mathbf{z}),\ i = 1, \ldots, n_p,\ j = 1, \ldots, p]^T$, where $\mathbf{D}_{h(i,j)}(\mathbf{z})$ is the horizontal backward difference computed at pixel $i \in \{i = 1, \ldots, n_p\}$ and image of coefficients $j \in \{i = 1, \ldots, p\}$. The action of $\mathbf{D}_v$ parallels that of $\mathbf{D}_h$ replacing the operator $\mathbf{D}_{h(i,j)}(\mathbf{z})$ by $\mathbf{D}_{v(i,j)}(\mathbf{z})$, which computes vertical backward differences at pixels $i \in \{i = 1, \ldots, n_p\}$ and image of coefficients $j \in \{i = 1, \ldots, p\}$. In addition, we define

$$\phi(\mathbf{Dz}) := \sum_{j=1}^{p} \sum_{i=1}^{n_p} \sqrt{\left(\mathbf{D}_{h(i,j)}(\mathbf{z})\right)^2 + \left(\mathbf{D}_{h(i,j)}(\mathbf{z})\right)^2}. \tag{15}$$

The regularizer (15) is the sum of the so-called isotropic TV [41] of the images of coefficients $\mathbf{Z}^i$, for $i = 1, \ldots, p$. Variations of the isotropic TV are the nonisotropic TV [41]

$$\phi(\mathbf{Dz}) := \sum_{j=1}^{p} \sum_{i=1}^{n_p} \left|\mathbf{D}_{h(i,j)}(\mathbf{z})\right| + \left|\mathbf{D}_{v(i,j)}(\mathbf{z})\right| \tag{16}$$

which promotes horizontal and vertical discontinuities, and the vector TV (see, e.g., [42] and [43])

$$\phi(\mathbf{Dz}) := \sum_{i=1}^{n_p} \sqrt{\sum_{i=1}^{p} \left(\mathbf{D}_{h(i,j)}(\mathbf{z})\right)^2 + \left(\mathbf{D}_{h(i,j)}(\mathbf{z})\right)^2} \tag{17}$$

which promotes aligned discontinuities over all bands.

The three variations of TV just presented are convex and lead to very similar algorithms. In this paper, we use the isotropic TV regularizer (15), as it is the more flexible regarding the type of discontinuities present in the hyperspectral data cubes.

Based on the aforementioned considerations, we address the following two criteria:

1) C-HYCA

$$\min_{\mathbf{Z}} \text{TV}(\mathbf{z}) \quad \text{subject to: } \|\mathbf{y} - \mathbf{Kz}\| \leq \delta, \quad \mathbf{z} \geq 0. \tag{18}$$

2) HYCA

$$\min_{\mathbf{z}} (1/2)\|\mathbf{y} - \mathbf{Kz}\|^2 + \lambda \text{TV}(\mathbf{z}) \quad \text{subject to: } \mathbf{z} \geq 0. \tag{19}$$

E. Brief Note About the Measurement Rate

TV minimization, similar to (18) or (19), has been widely used in compressed sensing applications with remarkable empirical results (see e.g., [9], [16], [19]–[21], and [44]). However, because the operator $\mathbf{\Psi}$ in (13) is not unitary, the standard theory of compressed sensing does not apply to the analysis formulations (18) and (19).

The research on provable guarantees that the TV-based CS is robust has just started. We refer to the work [45] that provides near-optimal guarantees for stable and robust image recovery from undersampled noisy measurements using TV minimization. More specifically, it is proved that an image can be reconstructed from $O(s \log n_p)$ CS measurements within the best s-term approximation of its gradient up to a logarithmic factor. We remark that our imaging problem is multidimensional and therefore not covered by the results introduced in [45].

A detailed treatment of the recovering guaranties for our CS problem scenario is beyond the reach of this paper. We develop, however, a few guidelines. Let us consider the cases: 1) $q \geq p$, i.e., the number of CS measurements per pixel is greater or equal than the rank of $\mathbf{X}$; and 2) $q < p$, the other way around. Here, we address the two cases.

1) $q \geq p$: In this regime, the matrices $\mathbf{A}_i\mathbf{E} \in \mathbb{R}^{q \times p}$, for $i = 1, \ldots, n_p$, are full column rank and therefore, in the absence of noise, we have

$$\mathbf{z} = \mathbf{K}^\dagger \mathbf{y}$$

where

$$\mathbf{K}^\dagger := \text{bdiag}\left((\mathbf{A}_1\mathbf{E})^\dagger, \ldots, (\mathbf{A}_{n_p}\mathbf{E})^\dagger\right)$$

where the symbol $\mathbf{B}^{\dagger}$ stands for the pseudoinverse of $\mathbf{B}$. Clearly, in this case, we do not need TV regularization. However, this is not the case in the presence of noise, where the TV regularization helps to stabilize the estimates of $\mathbf{z}$, as empirically illustrated in Section IV.

2) $q < p$: This case is much more complex, and its detailed treatment is object of current research efforts. Herein, we just provide motivation for the adopted CS measurement strategy. Let us focus our attention on the optimization variables associated with a given acquisition window of size $n_h = ws \times ws$, as schematized in Fig. 1, and define the optimization variables linked with the pixels inside the window as $\mathbf{z}_i \in \mathbb{R}^p$, for $i = 1, \ldots, n_h$. Since the TV regularization promotes piecewise smooth images $\mathbf{Z}^i$, for $i = 1, \ldots, p$, it is expectable, therefore, that a number of vectors $\mathbf{z}_i$, for $i = 1, \ldots, n_h$, are similar. Let $S_1, \ldots, S_{n_c}$ be a partition of the index set $\{1, \ldots, n_h\}$ and $\mathbf{a}_k \in \mathbb{R}^p$, for $k = 1, \ldots, n_c$, such that $\mathbf{z}_i = \mathbf{a}_k$ if $i \in S_k$. With this construction, recovering the set $\mathbf{z}_i$, for $i = 1, \ldots, n_h$, is equivalent to recover the set $\mathbf{a}_k$, for $k = 1, \ldots, n_c$. It is not hard to prove that the linear system linking the $q \times n_h$ CS measurements in the window with the variables $\mathbf{a}_k$, for $k = 1, \ldots, n_c$, has full column rank if

$$(\min_k |S_k|) \times q \geq p \tag{20}$$

where $|S_k|$ stands for the number of elements of S_k. Condition (20) provides a guideline concerning the size of the windows: it should be large enough to have $\min_k |S_k|$ as large as possible, and thus allowing for smaller values of q and consequently higher compression rates. However, there should be no clear advantage in increasing the size of the window beyond a given limit because, for most real-world images, is not expectable that $\min_k |S_k|$ keep increasing with the size of the window.

III. HYCA and C-HYCA Algorithms

Here, we introduce HYCA and C-HYCA algorithms. In both cases, the objective is to compute $\widehat{\mathbf{z}}$, or $\widehat{\mathbf{Z}}$ in the matrix format, and then, we infer the original hyperspectral data set by computing $\widehat{\mathbf{X}} = \mathbf{E}\widehat{\mathbf{Z}}$.

A. HYCA Optimization

To solve the problem in (19), we follow closely the methodology introduced in [13]. The core idea is to introduce a set of new variables per regularizer and then use the ADMM [46] to solve the resulting constrained optimization problem. By a careful choice of the new variables, the initial problem is converted into a sequence of much simpler problems. With this in mind, an equivalent way of writing the optimization problem (19) is

$$\min_{\mathbf{z}} \frac{1}{2}\|\mathbf{y} - \mathbf{K}\mathbf{z}\|^2 + \lambda_{TV}\, \phi(\mathbf{D}\mathbf{z}) + \iota_{R+}(\mathbf{z}) \tag{21}$$

where $\iota_{R+}(\mathbf{z}) = \sum_{i=1}^{pn_p} \iota_{R+}(z_i)$ is the indicator function (z_i represents the ith element of $\mathbf{z}$ and $\iota_{R+}(z_i)$ is zero if z_i belongs to the nonnegative orthant and $+\infty$ otherwise). Given the objective function in (21), we can write the following equivalent formulation:

$$\begin{aligned} \min_{\mathbf{z},\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3} \quad & \frac{1}{2}||\mathbf{y} - \mathbf{K}\mathbf{v}_1||^2 + \iota_{R+}(\mathbf{v}_2) + \lambda_{TV}\, \phi(\mathbf{D}\mathbf{z}) \\ \text{subject to} \quad & \mathbf{v}_1 = \mathbf{z} \\ & \mathbf{v}_2 = \mathbf{z} \\ & \mathbf{v}_3 = \mathbf{D}\mathbf{z} \end{aligned} \tag{22}$$

which in compact form becomes

$$\min_{\mathbf{z},\mathbf{v}} g(\mathbf{v}) \quad \text{subject to} \quad \mathbf{v} = \mathbf{G}\mathbf{z} \tag{23}$$

with

$$\mathbf{v} \equiv \left[\mathbf{v}_1^T, \mathbf{v}_2^T, \mathbf{v}_3^T\right]^T$$

$$g(\mathbf{v}) \equiv \frac{1}{2}||\mathbf{y} - \mathbf{K}\mathbf{v}_1||^2 + \iota_{R+}(\mathbf{v}_2) + \lambda_{TV}\, \phi(\mathbf{v}_3) \tag{24}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \\ \mathbf{D} \end{bmatrix}. \tag{25}$$

Algorithm 1 ADMM pseudocode for solving problem (23).

1. Initialization: set $k = 0$, choose $\mu > 0$, $\mathbf{z}^{(0)}$, $\mathbf{v}^{(0)}$, $\mathbf{d}^{(0)}$
2. repeat:
  3. $\mathbf{z}^{(k+1)} \leftarrow \arg\min_{\mathbf{z}} \mathcal{L}(\mathbf{z}, \mathbf{v}^{(k)}, \mathbf{d}^{(k)})$
  4. $\mathbf{v}^{(k+1)} \leftarrow \arg\min_{\mathbf{v}} \mathcal{L}(\mathbf{z}^{(k+1)}, \mathbf{v}, \mathbf{d}^{(k)})$
  5. $\mathbf{d}^{(k+1)} \leftarrow \mathbf{d}^{(k)} - [\mathbf{G}\mathbf{z}^{(k+1)} - \mathbf{v}^{(k+1)}]$
6. until some stopping criterion is satisfied.

The ADMM for solving (23) is shown in Algorithm 1, where (see [13] and [47])

$$\mathcal{L}(\mathbf{z}, \mathbf{v}, \mathbf{d}) \equiv g(\mathbf{z}, \mathbf{v}) + \frac{\mu}{2}\|\mathbf{G}\mathbf{z} - \mathbf{v} - \mathbf{d}\|^2 \tag{26}$$

is the augmented Lagrangian for problem (23), $\mu > 0$ is a positive constant, and $\mathbf{d}/\mu$ denotes the Lagrange multipliers associated to the constraint $\mathbf{G}\mathbf{z} = v$. In each iteration, Algorithm 1 sequentially optimizes $\mathcal{L}$ with respect to $\mathbf{z}$ (step 3) and $\mathbf{v}$ (step 4), and then updates the Lagrange multipliers (step 5).

Expanding the augmented Lagrangian introduced in (26), we obtain

$$\begin{aligned} &\mathcal{L}(\mathbf{z}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3) \\ &\quad = \frac{1}{2}||\mathbf{y} - \mathbf{K}\mathbf{v}_1||^2 + \iota_{R+}(\mathbf{v}_2) + \lambda_{TV}\, \phi(\mathbf{v}_3) \\ &\qquad + \frac{\mu}{2}\|\mathbf{z} - \mathbf{v}_1 - \mathbf{d}_1\|^2 + \frac{\mu}{2}\|\mathbf{z} - \mathbf{v}_2 - \mathbf{d}_2\|^2 \\ &\qquad + \frac{\mu}{2}\|\mathbf{D}\mathbf{z} - \mathbf{v}_3 - \mathbf{d}_3\|^2. \end{aligned} \tag{27}$$

Algorithm 2 Pseudocode of HYCA algorithm.

1. Initialization: set $k=0$, choose $\mu \geq 0$, $\mathbf{z}^{(0)}$, $\mathbf{v}_1^{(0)}$, $\mathbf{v}_2^{(0)}$, $\mathbf{v}_3^{(0)}$, $\mathbf{d}_1^{(0)}$, $\mathbf{d}_2^{(0)}$, $\mathbf{d}_3^{(0)}$
2. repeat:
  3. $\mathbf{z}^{(k+1)} \leftarrow \arg\min_{\mathbf{z}} \mathcal{L}(\mathbf{z}, \mathbf{v}_1^{(k)}, \mathbf{v}_2^{(k)}, \mathbf{v}_3^{(k)}, \mathbf{d}_1^{(k)}, \mathbf{d}_2^{(k)}, \mathbf{d}_3^{(k)})$
  4. $\mathbf{v}_1^{(k+1)} \leftarrow \arg\min_{\mathbf{v}_1} \mathcal{L}(\mathbf{z}^{(k+1)}, \mathbf{v}_1, \mathbf{v}_2^{(k)}, \mathbf{v}_3^{(k)})$
  5. $\mathbf{v}_2^{(k+1)} \leftarrow \arg\min_{\mathbf{v}_2} \mathcal{L}(\mathbf{z}^{(k+1)}, \mathbf{v}_1^{(k+1)}, \mathbf{v}_2, \mathbf{v}_3^{(k)})$
  6. $\mathbf{v}_3^{(k+1)} \leftarrow \arg\min_{\mathbf{v}_3^{(k)}} \mathcal{L}(\mathbf{z}^{(k+1)}, \mathbf{v}_1^{(k+1)}, \mathbf{v}_2^{(k+1)}, \mathbf{v}_3)$
  7. Update Lagrange multipliers:
  $\mathbf{d}_1^{(k+1)} \leftarrow \mathbf{d}_1^{(k)} - \mathbf{z}^{(k+1)} + \mathbf{v}_1^{(k+1)}$
  $\mathbf{d}_2^{(k+1)} \leftarrow \mathbf{d}_2^{(k)} - \mathbf{z}^{(k+1)} + \mathbf{v}_2^{(k+1)}$
  $\mathbf{d}_3^{(k+1)} \leftarrow \mathbf{d}_3^{(k)} - \boldsymbol{\Psi}_h \mathbf{z}^{(k+1)} + \mathbf{v}_3^{(k+1)}$
  8. Update iteration: $k \leftarrow k+1$
9. until some stopping criterion is satisfied.

The pseudocode of HYCA is shown in Algorithm 2. It is the expansion of the ADMM algorithm presented in Algorithm 1. The goal of step 3 in Algorithm 2 is to determine the value of the variable $\mathbf{z}$ at each iteration. This is a quadratic problem with a block circulant system matrix, thus effectively solved in the Fourier domain

$$\mathbf{z}^{(k+1)} \leftarrow (\mathbf{D}^T\mathbf{D} + 2\mathbf{I})^{-1}(\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2 + \mathbf{D}^T\boldsymbol{\xi}_3) \tag{28}$$

where $\boldsymbol{\xi}_1 = \mathbf{v}_1^{(k)} + \mathbf{d}_1^{(k)}$, $\boldsymbol{\xi}_2 = \mathbf{v}_2^{(k)} + \mathbf{d}_2^{(k)}$, $\boldsymbol{\xi}_3 = \mathbf{v}_3^{(k)} + \mathbf{d}_3^{(k)}$.

Steps 4–6 of HYCA compute the minimization of the augmented Lagrangian with respect to $\mathbf{v}$, which is decoupled with respect to $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. The minimization with respect to $\mathbf{v}_1$ is

$$\mathbf{v}_1^{(k+1)} \leftarrow \arg\min_{\mathbf{v}_1} \frac{1}{2}\|\mathbf{y} - \mathbf{K}\mathbf{v}_1\|^2 + \frac{\mu}{2}\left\|\mathbf{z}^{(k+1)} - \mathbf{v}_1 - \mathbf{d}_1^{(k)}\right\|^2 \tag{29}$$

whose solution is

$$\mathbf{v}_1^{(k+1)} \leftarrow (\mathbf{K}^T\mathbf{K} + \mu\mathbf{I})^{-1}\left(\mathbf{K}^T\mathbf{y} + \mu\left(\mathbf{z}^{(k+1)} - \mathbf{d}_1^{(k)}\right)\right). \tag{30}$$

The inverse of the matrix $(\mathbf{K}^T\mathbf{K} + \mu\mathbf{I})$, given the block diagonal structure of $\mathbf{K}$, is the block diagonal and can be easily precomputed. We remark that we just need to save the diagonal blocks.

To compute $\mathbf{v}_2$, the optimization problem to be solved is

$$\mathbf{v}_2^{(k+1)} \leftarrow \arg\min_{\mathbf{v}_2} \iota_{R+}(\mathbf{v}_2) + \frac{\mu}{2}\left\|\mathbf{z}^{(k+1)} - \mathbf{v}_2 - \mathbf{d}_2^{(k)}\right\|^2 \tag{31}$$

which is the projection the $\mathbf{z}^{(k+1)} - \mathbf{d}_2^{(k)}$ onto the nonnegative orthant, i.e.,

$$\mathbf{v}_2^{(k+1)} \leftarrow \max\left(0, \mathbf{z}^{(k+1)} - \mathbf{d}_2^{(k)}\right) \tag{32}$$

where the max operation is to be understood componentwise.

Finally, the minimization with respect to $\mathbf{v}_3$, amounts to compute

$$\mathbf{v}_3^{(k+1)} \leftarrow \arg\min_{\mathbf{v}_3} \lambda_{TV}\,\phi(\mathbf{v}_3) + \frac{\mu}{2}\left\|\mathbf{D}\mathbf{z}^{(k+1)} - \mathbf{v}_3 - \mathbf{d}_3^{(k)}\right\|^2 \tag{33}$$

whose solution is the well-known *vect-soft threshold* (see, e.g., [48])

$$\mathbf{v}_{3(i,j)}^{(k+1)} \leftarrow \text{vect-soft}\left(\left(\mathbf{D}\mathbf{z}^{(k+1)} - \mathbf{d}_3^{(k)}\right)_{(i,j)}, \lambda_{TV}/\mu\right) \tag{34}$$

where $(\mathbf{a})_{i,j} := [\mathbf{a}_h^T\mathbf{a}_v^T]_{(i,j)} = (\mathbf{a}_{h(i,j)}, \mathbf{a}_{v(i,j)})$, for $i = 1, \ldots, n_p$ and $j = 1, \ldots, p$, and vect-soft$(\cdot, \tau)$ denotes the application of the vect-soft-threshold function $\mathbf{b} \mapsto \mathbf{b}(\max\{\|\mathbf{b}\|_2 - \tau, 0\}/\max\{\|\mathbf{b}\|_2 - \tau, 0\} + \tau)$.

The convergence conditions from [46, Theorem 1] are met: matrix $\mathbf{G}$ is full column rank and function g introduced in (23) is closed, proper, and convex. Under these conditions, the same theorem states that, for any $\mu > 0$, if (23) has a solution, e.g., $\mathbf{z}^*$, then the sequence $\{\mathbf{z}^{(k)}\}$ converges to $\mathbf{z}^*$. If (23) does not have a solution, then, at least one of the sequences $\{\mathbf{z}^{(k)}\}$ or $\{\mathbf{d}^{(k)}\}$ diverges. The stopping criterion adopted in the algorithm is $k = 200$. The practical justification will be provided in Section IV.

In the ADMM scheme, the setting of the parameter μ has a strong impact over the convergence speed. In our case we set the μ parameter by hand tuning for optimal performance.

B. C-HYCA Optimization

As in HYCA, we also used the ADMM methodology to solve the C-HYCA minimization problem

$$\min_{\mathbf{z}\geq 0} \text{TV}(\mathbf{z}) \quad \text{subject to:} \quad \|\mathbf{y} - \mathbf{K}\mathbf{z}\|^2 \leq \delta \tag{35}$$

where δ is a scalar value linked to the noise statistics.

As mentioned before in the case of HYCA, here we also introduce a set of new variables per term of the objective function and then use the ADMM [46] previously described to solve the resulting convex constrained optimization problem. By a careful choice of the new variables, the initial problem is converted into a sequence of much simpler problems. With this in mind, let us define $\iota_{B(\epsilon)}$ as the indicator on a ball of radius ϵ, i.e., $\iota_{B(\epsilon)}(\mathbf{z}) = 0$ if $\|\mathbf{z}\| \leq \varepsilon$ and $+\infty$ otherwise. With these definition in place, an equivalent way of writing the optimization problem in (35) is

$$\min_{\mathbf{z}} \phi(\mathbf{D}\mathbf{z}) + \iota_{B(\delta)}(\mathbf{y} - \mathbf{K}\mathbf{z}) + \iota_{R+}(\mathbf{z}). \tag{36}$$

Given the objective function (36), we now write the following convex constrained equivalent formulation:

$$\begin{aligned} \min_{\mathbf{z},\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4} \quad & \phi(\mathbf{v}_1) + \iota_{B(\delta)}(\mathbf{v}_2) + \iota_{R+}(\mathbf{v}_3) \\ \text{subject to:} \quad & \mathbf{v}_1 = \mathbf{D}\mathbf{z} \\ & \mathbf{v}_2 = \mathbf{y} - \mathbf{K}\mathbf{v}_4 \\ & \mathbf{v}_3 = \mathbf{z} \\ & \mathbf{v}_4 = \mathbf{z} \end{aligned} \tag{37}$$

which we solve via ADMM in a way similar to Algorithm 1.

Here, notice that the constraint $\mathbf{v}_2 = \mathbf{y} - \mathbf{K}\mathbf{v}_4$ is asymmetric with regard to the other constraints. This asymmetry underlies a great improvement from the computational point of view, due to the fact that we decouple the optimization in the spatial domain from the optimization in the spectral domain. If we had followed [13] exactly, we would have $\mathbf{v}_2 = \mathbf{K}\mathbf{z}$ instead of the couple $(\mathbf{v}_2 = \mathbf{y} - \mathbf{K}\mathbf{v}_4, \mathbf{v}_4 = \mathbf{z})$. However, in that case, the optimization with regard to $\mathbf{z}$ would be much more complex, as we would have matrix $\mathbf{K}$ acting over the spectral dimension and matrix $\mathbf{D}$ acting over the spatial domain in the same system.

Expanding the augmented Lagrangian in this case, we obtain

$$\begin{aligned}\mathcal{L}(\mathbf{z}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4) \\ = \phi(\mathbf{v}_1) + \frac{\mu}{2}\|\mathbf{D}\mathbf{z} - \mathbf{v}_1 - \mathbf{d}_1\|^2 + \iota_{B(\delta)}(\mathbf{v}_2) \\ + \frac{\mu}{2}\|\mathbf{v}_2 - (\mathbf{y} - \mathbf{K}\mathbf{v}_4) - \mathbf{d}_2\|^2 + \iota_{R+}(\mathbf{v}_3) \\ + \frac{\mu}{2}\|\mathbf{z} - \mathbf{v}_3 - \mathbf{d}_3\|^2 + \frac{\mu}{2}\|\mathbf{z} - \mathbf{v}_4 - \mathbf{d}_4\|^2. \quad (38)\end{aligned}$$

The minimization of the augmented Lagrangian with respect to $\mathbf{z}$ is a quadratic problem with a block circulant system matrix, thus effectively solved in the Fourier domain

$$\mathbf{z}^{(k+1)} \leftarrow (\mathbf{D}^T\mathbf{D} + 2\mathbf{I})^{-1}(\mathbf{D}^T\boldsymbol{\xi}_1 + \boldsymbol{\xi}_3 + \boldsymbol{\xi}_4) \quad (39)$$

where $\boldsymbol{\xi}_1 = \mathbf{v}_1^{(k)} + \mathbf{d}_1^{(k)}$, $\boldsymbol{\xi}_3 = \mathbf{v}_3^{(k)} + \mathbf{d}_3^{(k)}$, and $\boldsymbol{\xi}_4 = \mathbf{v}_4^{(k)} + \mathbf{d}_4^{(k)}$.

The minimization with respect to $\mathbf{v}_1$ is similar to the minimization with respect to $\mathbf{v}_3$ in (34) yielding

$$\mathbf{v}_{1(i,j)}^{(k+1)} \leftarrow \text{vect-soft}\left(\left(\mathbf{z}^{(k+1)} - \mathbf{d}^{(k)}\right)_{1(i,j)}, 1/\mu\right). \quad (40)$$

The minimization with respect to $\mathbf{v}_2$ is a projection on a ball of radius δ

$$\mathbf{v}_2^{(k+1)} \leftarrow \begin{cases} \boldsymbol{\xi}_2 & \|\boldsymbol{\xi}_2\| \le \delta \\ \frac{\delta\boldsymbol{\xi}_2}{\|\boldsymbol{\xi}_2\|} & \text{otherwise} \end{cases} \quad (41)$$

where $\boldsymbol{\xi}_2 = \mathbf{y} - \mathbf{K}\mathbf{v}_4^{(k)} + \mathbf{d}_2^{(k)}$.

The minimization with respect to $\mathbf{v}_3$ is a component-wise projection on the nonnegative orthant yielding

$$\mathbf{v}_3^{(k+1)} \leftarrow \max\left(0, \mathbf{z}^{(k+1)} - \mathbf{d}_3^{(k)}\right). \quad (42)$$

The minimization with respect to $\mathbf{v}_4$ is a quadratic problem involving the inverse of the operator $(\mathbf{K}^T\mathbf{K} + \mathbf{I})$ which, given the structure of the matrix $\mathbf{K}$, can be precomputed

$$\begin{aligned}\mathbf{v}_4^{(k+1)} \leftarrow (\mathbf{K}^T\mathbf{K} + \mathbf{I})^{-1}\left[\left(\mathbf{z}^{(k+1)} - \mathbf{d}_4^{(k)}\right)\right. \\ \left. + \mathbf{K}^T\left(-\mathbf{v}_2^{(k+1)} + \mathbf{z} + \mathbf{d}_2^{(k)}\right)\right]. \quad (43)\end{aligned}$$

The optimizations with respect to $\mathbf{v}_2$ alone and to $\mathbf{v}_4$ alone are very light. Therefore, a very simple way to achieve joint optimization with respect to $(\mathbf{v}_2, \mathbf{v}_4)$ is to cycle over these two optimizations until convergence. However, in this case, we cycle only once due to having observed systematically faster convergence with just one step than with more steps. A proof of the convergence of C-HYCA with just one step is, however, beyond the scope of this paper.

We remark that C-HYCA does not use any regularization parameter. It uses instead the parameter δ, which is linked to the noise statistics of the image and thus much easier to infer than the 2-D TV regularization parameter λ_{TV} present in HYCA.

IV. Experimental Results

Here, we conduct a series of experiments using different versions of HYCA and C-HYCA algorithms on real and simulated data. Specifically, we have addressed three cases. In the first case, we have not enforced the nonnegativity constraint and $\mathbf{E}$ is an orthonormal matrix. In the second case, we have used the sampling matrix $\mathbf{H}_i = \mathbf{A}_i\mathbf{E}^\dagger$, where $\mathbf{A}_i\mathbb{R}^{q\times p}$ is a random matrix following a Gaussian i.i.d. and $\mathbf{E}^\dagger$ is the pseudoinverse of the matrix $\mathbf{E}$. In the third case, we have enforced the nonnegativity constraint and assumed that $\mathbf{E}$ is the mixing matrix, which contains the endmembers.

A. Synthetic Data

The synthetic data set used in this experiments were generated from spectral signatures randomly selected from the U.S. Geological Survey (USGS).[3] The simulated images 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. Zero-mean Gaussian noise was added to the synthetic scenes with signal-to-noise ratios (SNRs) defined as $\text{SNR} := 10 \cdot \log_{10}(\mathbb{E}\|\mathbf{E}\mathbf{Z}\|_F^2/\mathbb{E}\|\mathbf{N}\|_F^2)$, where $\mathbb{E}$ denotes mean value and $\mathbf{N}$ is a matrix version of the noise vector $\mathbf{w}$ shown in (1), to simulate contributions from ambient and instrumental noise sources. Fig. 2 displays the ground-truth abundances maps used for generating the simulated imagery.

In order to evaluate the performance of the HYCA and C-HYCA, we use as performance indicator the normalized mean-squared error (NMSE) of the reconstruction given by

$$\text{NMSE} = \|(\widehat{\mathbf{Z}} - \mathbf{Z})\|_F^2/\|\mathbf{Z}\|_F^2 \quad (44)$$

where $\mathbf{Z}$ and $\widehat{\mathbf{Z}}$ denote the original and reconstructed abundance fraction images, respectively. We set the window size $ws = 2$ so that $m = 4$. In the first case, we disabled the nonnegativity constraint and we assume that the mixing matrix $\mathbf{E}^*$ an orthogonal matrix computed as $\mathbf{E}^* := \text{orth}(\mathbf{E})$ where $\mathbf{E}$ denotes the original mixing matrix used to generate the data set and $\mathbf{E}^*$ denotes the matrix used in the reconstruction algorithm. In the second case, we used for sampling $\mathbf{H}_i = \mathbf{A}_i\mathbf{E}^\dagger$ without the nonnegativity constraint; and finally, in the third case, we used the nonnegativity constraint considering the original mixing matrix used for the data set generation.

In this experiment, we set $q = 3$. Since the original data set has $l = 224$ bands, the compression ratio is $l/q = 74.67$.

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

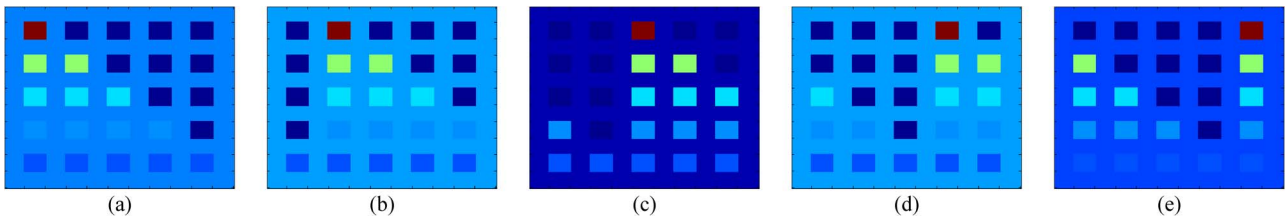

Fig. 2. True abundance maps of endmembers in the synthetic hyperspectral data. (a) Endmember #1. (b) Endmember #2. (c) Endmember #3. (d) Endmember #4. (e) Endmember #5.

TABLE II
SIMULATED DATA. AVERAGE NMSE BETWEEN THE ORIGINAL AND THE RECONSTRUCTED DATASET FOR $q = 3$ AND DIFFERENT SNR VALUES AFTER 10 MONTE CARLO RUNS

Version	SNR=30db	SNR=50db	SNR=70db	SNR=∞
HYCA *	$8.27 \cdot 10^{-4}$	$0.52 \cdot 10^{-4}$	$0.52 \cdot 10^{-4}$	$0.30 \cdot 10^{-4}$
HYCA †	$0.44 \cdot 10^{-4}$	$0.45 \cdot 10^{-4}$	$0.40 \cdot 10^{-4}$	$0.37 \cdot 10^{-4}$
HYCA §	$21.08 \cdot 10^{-4}$	$0.68 \cdot 10^{-4}$	$0.33 \cdot 10^{-4}$	$0.20 \cdot 10^{-4}$
C-HYCA *	$22.7 \cdot 10^{-4}$	$3.35 \cdot 10^{-4}$	$3.24 \cdot 10^{-4}$	$3.36 \cdot 10^{-4}$
C-HYCA †	$2.73 \cdot 10^{-4}$	$2.00 \cdot 10^{-4}$	$1.31 \cdot 10^{-4}$	$1.59 \cdot 10^{-4}$
C-HYCA §	$7.26 \cdot 10^{-4}$	$0.51 \cdot 10^{-4}$	$0.29 \cdot 10^{-4}$	$0.28 \cdot 10^{-4}$

* First case without non-negativity and $\mathbf{E}^* := \text{orth}(\mathbf{E})^\dagger$. Second case with $\mathbf{H}_i = \mathbf{A}_i\mathbf{E}^\dagger$. § Third case using the non-negativity constraint.

TABLE III
CUPRITE DATA SET. AVERAGED NMSE BETWEEN THE ORIGINAL AND THE RECONSTRUCTED DATA SET OVER 10 MONTE CARLO RUNS, FOR DIFFERENT VERSIONS OF THE ALGORITHM WITH DIFFERENT COMPRESSIONS RATIOS

Version	$q = 5$	$q = 9$	$q = 13$	$q = 17$
HYCA *	$98.13 \cdot 10^{-4}$	$5.21 \cdot 10^{-4}$	$2.59 \cdot 10^{-4}$	$4.54 \cdot 10^{-4}$
HYCA †	$376.83 \cdot 10^{-4}$	$200.39 \cdot 10^{-4}$	$108.80 \cdot 10^{-4}$	$34.56 \cdot 10^{-4}$
HYCA §	$4.66 \cdot 10^{-4}$	$2.50 \cdot 10^{-4}$	$1.28 \cdot 10^{-4}$	$1.05 \cdot 10^{-4}$
C-HYCA *	$24.05 \cdot 10^{-4}$	$5.77 \cdot 10^{-4}$	$3.54 \cdot 10^{-4}$	$2.63 \cdot 10^{-4}$
C-HYCA †	$590.82 \cdot 10^{-4}$	$351.57 \cdot 10^{-4}$	$268.54 \cdot 10^{-4}$	$225.88 \cdot 10^{-4}$
C-HYCA §	$3.84 \cdot 10^{-4}$	$2.24 \cdot 10^{-4}$	$1.74 \cdot 10^{-4}$	$1.31 \cdot 10^{-4}$

* First case without non-negativity and $\mathbf{E}^* := \text{orth}(\mathbf{E})$. † Second case with $\mathbf{H}_i = \mathbf{A}_i\mathbf{E}^\dagger$. § Third case using the non-negativity constraint.

Table II shows the NMSE for both versions in different cases. We performed 10 Monte Carlo runs, sampling not only the noise but also the elements of the matrix $\mathbf{H}_i$. The regularization parameter λ_{TV} in (19) was hand tuned for optimal performance in the case of the HYCA algorithm. Having in mind the linear model (8), the parameter δ in (18) is set to $\delta := \|\mathbf{A}(\mathbf{N})\|_F$.

As shown in Table II by disabling the nonnegativity constraint, the algorithm is more robust to noise; however, with the nonnegativity constraint, it provides the better results when there is no noise. This is expected because the noise introduces outliers in the model, which may give rise to errors in the reconstruction when we use the nonnegativity constraint. In addition, we can see that in the second case, the results are very good even with SNR $= 30$ db, which indicates that, in this case, the algorithm is more robust to noise. If we compare the HYCA and C-HYCA criteria, we can see that both versions provide very good results with very low errors. We can also see that HYCA outperforms C-HYCA, which is due to all the parameters were hand tuned for optimal performance in the case of HYCA.

B. Cuprite

In this experiment, we use the well-known AVIRIS Cuprite data set, available online in reflectance units after atmospheric correction. This scene has been widely used to validate the performance of endmember extraction algorithms. The portion used in experiments corresponds to a 250×190 pixels subset of the sector labeled as f970619t01p02_r02_sc03a.rfl in the online data. The scene comprises 224 spectral bands between 0.4 and 2.5 μm, with full-width at half-maximum of 10 nm and spatial resolution of 20 m per pixel. Prior to the analysis, several bands were removed due to water absorption and low SNR in those bands, leaving a total of 188 reflectance channels to be used in the experiments. We used a window size of $ws = 2$, so that $m = 4$ and estimated the number of endmembers with the Hysime algorithm [40].

Here, we have also considered the same three cases as before. However, in this case, the mixing matrix was estimated from the original real data using the VCA algorithm [49], so that in the first case, we use $\mathbf{E}^* := \text{orth}(\widehat{\mathbf{E}})$ where $\widehat{\mathbf{E}}$ is the mixing matrix estimated by VCA algorithm. In the second case $\mathbf{H}_i = \mathbf{A}_i\widehat{\mathbf{E}}^\dagger$ and $\mathbf{E}^* := \widehat{\mathbf{E}}$. Finally, in the third case, due to the nonlinear mixtures and outliers present in the real images, the nonnegativity constraint may be violated. In order to ensure that the mixing matrix encloses the whole data-set and then the nonnegativity constraint is satisfied, we open the cone defined the mixing matrix $\mathbf{E}^*$ as follows:

$$\mathbf{E}^* := \widehat{\mathbf{E}} + \Delta \cdot (\widehat{\mathbf{E}} - \overline{\mathbf{E}}) \tag{45}$$

where Δ is a scalar that defines how much the cone is opened, and $\overline{\mathbf{E}}$ is a matrix containing the mean spectrum of the endmembers. By choosing a value of Δ large enough, then all observed spectral vectors are inside the cone implying that $\mathbf{Z} \geq 0$. In the current data set, $\Delta = 6$ ensures this constraint.

In order to evaluate the performance of HYCA and C-HYCA with the real data set, we perform experiments with the compression ratios $188/q$ with $q = 5, 9, 13, 17$. In all cases, we used a window size of $ws = 2$, so that $m = 4$.

In Table III, we show the value of NMSE over 10 Monte Carlo runs for several versions of the algorithm with different

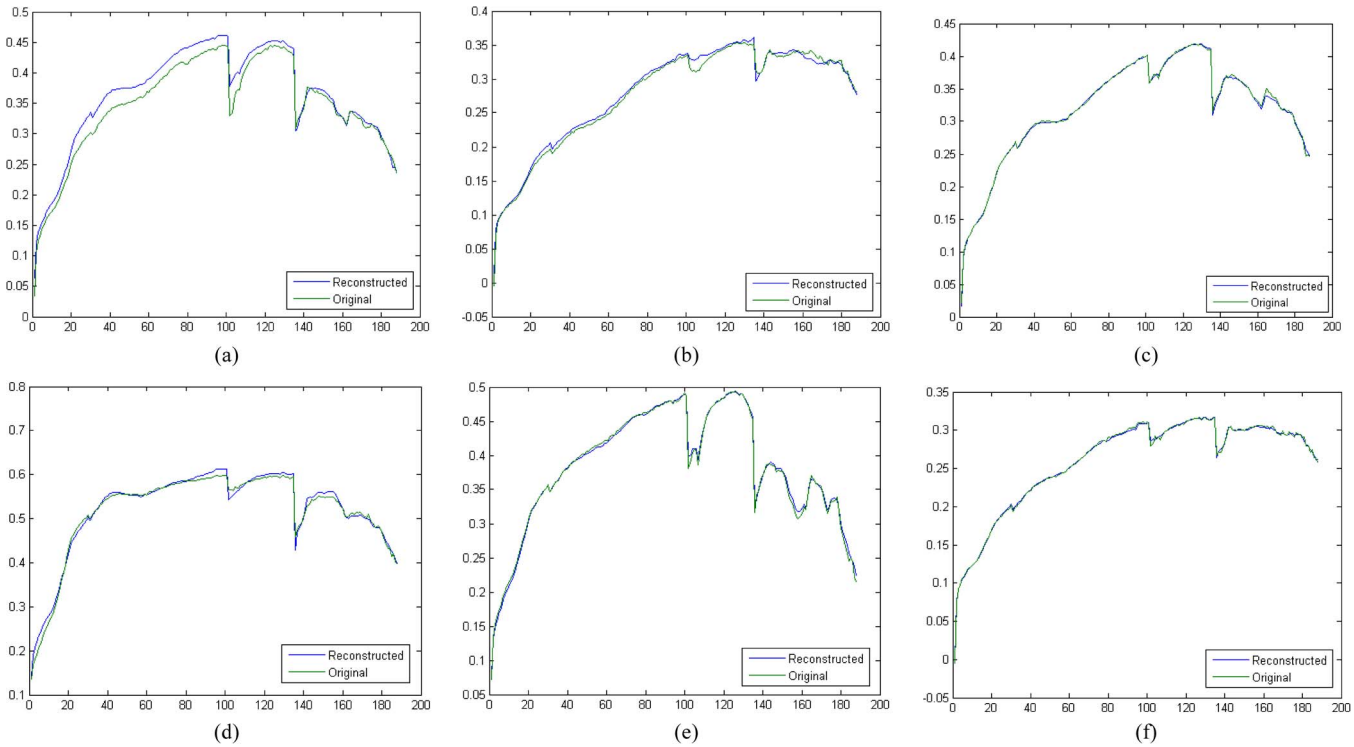

Fig. 3. (a) Worst, (b) mean, and (c) best reconstructed pixel in the Cuprite scene for values of $q = 5$ and the same in the case of $q = 15$ (d, e, f). (a) $q = 5$, worst case. (b) $q = 5$, mean case and (c) $q = 5$, best case (d) $q = 15$, worst case and (e) $q = 15$, mean case and (f) $q = 15$, best case.

compression ratios over the Cuprite data set. In this experiment, we can see that the version with the nonnegativity constraint provide better results than the other versions. Here, the second case is not working, as well as with the synthetic data set, because here, we are using an estimation of the endmembers instead of the original ones in order to compress the data, which introduces a penalty in this case.

Fig. 3 shows the reconstructed and the original signatures with highest, mean and lowest error for the C-HYCA algorithm with the nonnegativity constraint for different compression ratios. In this figure, we can see that even in the worst case although there is a scale error between the original and the reconstructed signature, the reconstructed pixel preserves the shape of the original pixel, which means that the features of the signature are preserved. In the average case and the better case the reconstructed and the original pixel are extremely similar.

Fig. 4 shows the NMSE between the original and the reconstructed images by the C-HYCA algorithm with the nonnegativity constraint for different compression ratios in the case of the Cuprite data set. Note that the scale of these figures is between 0 and $2 \cdot 10^{-3}$. Most of errors are in the transitions areas between different land-cover classes. This was expected as the 2-D TV regularizer promotes smoothness in the smooth regions. Furthermore, we can see that the most of pixels have a very low error, which means that the reconstructed and the original spectrum will be very similar in most of the cases. Note that the spectra showed in Fig. 3(a) and (d) correspond to the worst case situation, in which the error is much higher than the errors of the large majority of the pixels in the image.

Table IV shows the NMSE values after the Cuprite data set reconstruction by the C-HYCA algorithm with the nonnegativity constraint with different windows sizes $ws = [4, 6, 8]$ and the compression ratio of l/q with $q = 5$. As we can see the results are very similar for the different windows sizes. They can, however, improve a little bit by choosing an adequate window size; in this case the optimum window size of $ws = 4$. This is in line with the analysis provided in Section II-E.

Finally, in order to conclude this section, Table V shows the execution time of the reconstruction algorithms for the real Cuprite data set running the algorithms during 200 iterations. This experiment was performed in a desktop computer Intel Core i7 920 CPU AT 2.67 GHz with 4 GB of RAM. The table shows that both algorithms spend a similar time, although C-HYCA outperforms in a few seconds the HYCA algorithm. However, although these times have been obtained with a serial implementation, these algorithms can be implemented in high-performance computing architectures such as clusters or graphics processing units (GPUs), which could outperform drastically the execution time.

V. Conclusion and Future Research Lines

In this paper, we have developed a new framework for hyperspectral CS called HYCA and its C-HYCA. This approach takes advantage of two main properties of hyperspectral data, namely the high spatial correlation of abundance fractions and the low number of endmembers that are generally required to explain the observed data. While HYCA depends on the tuning of a regularization parameter controlling the relative weight

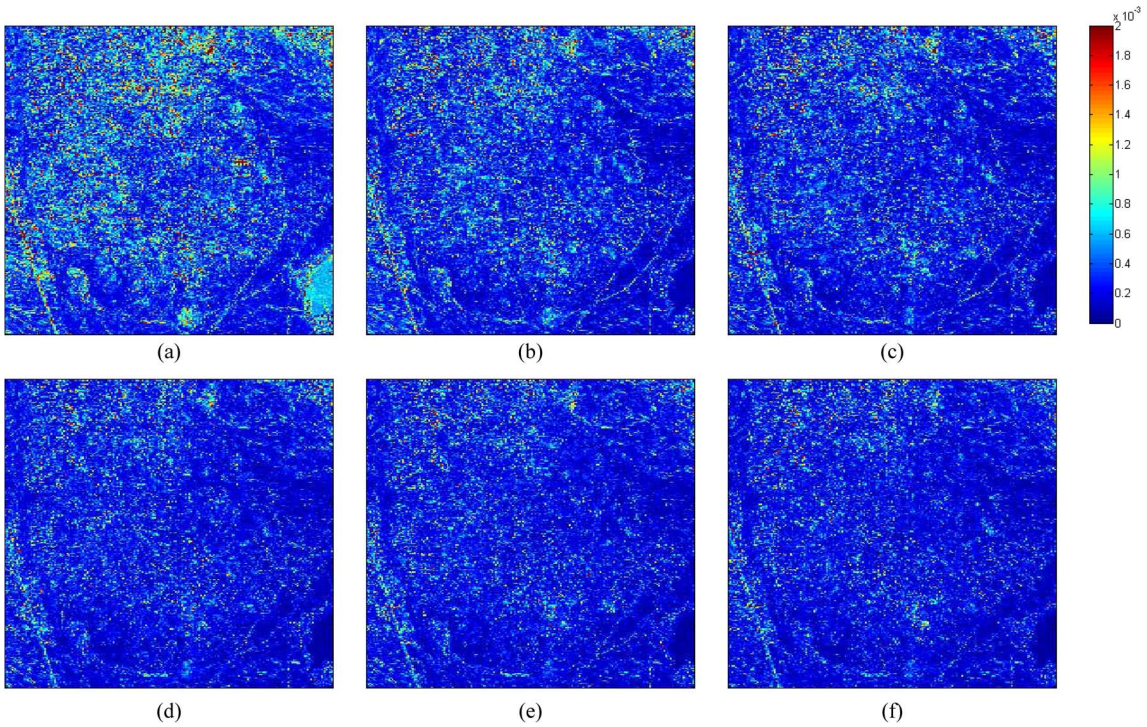

Fig. 4. NMSE between the original and the reconstructed Cuprite data set for different compression ratios (a)–(f). (a) $q = 5$. (b) $q = 7$. (c) $q = 9$. (d) $q = 11$. (e) $q = 13$. (f) $q = 15$.

TABLE IV
NMSE BETWEEN THE ORIGINAL AND THE RECONSTRUCTED CUPRITE DATASET OVER 10 MONTE CARLO RUNS AND VALUES OF WS $= [4, 6, 8]$ WITH A COMPRESSION RATIO OF $188/q$ WITH $q = 5$

Version	$ws = 4$	$ws = 6$	$ws = 8$
C-HYCA	$2.95 \cdot 10^{-4}$	$3.01 \cdot 10^{-4}$	$3.01 \cdot 10^{-4}$

TABLE V
HYCA AND C-HYCA EXECUTION TIMES FOR THE RECONSTRUCTION OF CUPRITE SCENE WITH 48 000 PIXELS RUNNING 200 ITERATIONS IN AN INTEL CORE I7 920 CPU AT 2.67 GHZ WITH 4 GB OF RAM

Version	Time (secs)	Time/#pixels
HYCA	319.78	$6.662 \cdot 10^{-3}$
C-HYCA	303.78	$6.633 \cdot 10^{-3}$

between the TV regularizer and the data term, C-HYCA does not depend on any regularization parameters. This represents an important contribution since reference information is generally difficult to obtain *a priori* in real applications. An exhaustive quantitative and comparative assessment has been performed in order to evaluate the accuracy and computational performance of these methods. The experiments have been conducted using both synthetic and real data. The obtained results show that both approaches provide good results in the task of compressing remotely sensed hyperspectral data sets, which is mainly due to the fact that they exploit the aforementioned properties, which are intrinsic of remotely sensed hyperspectral data sets. In other words, the proposed approaches can achieve high compression ratios with very small errors. In fact, most of the errors are located in the transition areas between different land-cover classes, but even in these cases, the reconstructed pixels preserve the shape of the original pixel quite accurately. This is very important due to the fact that most hyperspectral imaging applications require high fidelity in the reconstruction of the shape of the spectral signatures in order to discriminate between different materials.

An important requirement of the proposed algorithms is the need to know the endmember mixing matrix in advance in order to accurately reconstruct the original hyperspectral data set. In the future, we want to relax this requirement by resorting to spectral libraries and sparse techniques. Spectral library signatures could be used instead in order to reconstruct the image. A more detailed analysis of the sensitivity of the method to the size of the window used in the proposed measurement strategy is also a topic deserving future research. From a computational point of view, and although the presented algorithms are not expensive, the ideal scenario would to be able to perform the reconstruction process in real time. Currently, we need to measure the original image, then estimate matrix $\mathbf{E}$ and after that, we compress and transmit the data to the Earth station jointly with the estimated matrix. In applications requiring strict real-time processing, the ideal scenario would to be able to perform the full reconstruction process in real time. For this purpose, we are currently developing high-performance implementations of the proposed approaches in clusters of computers and commodity GPUs. Due to the inherently parallel nature of the discussed algorithms, we believe that the development of parallel versions can considerably reduce the processing time needed to perform the reconstruction process.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor, who handled the manuscript, and the anonymous reviewers for their insightful comments and suggestions, which greatly helped us improve the technical quality and presentation of our manuscript.

REFERENCES

[1] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for Earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, Jun. 1985.

[2] R. O. Green *et al.*, "Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.

[3] G. Motta, F. Rizzo, and J. A. Storer, *Hyperspectral Data Compression*. Berlin, Germany: Springer-Verlag, 2006.

[4] B. Huang, *Satellite Data Compression*. Berlin, Germany: Springer-Verlag, 2011.

[5] Q. Du and J. E. Fowler, "Low-complexity principal component analysis for hyperspectral image compression," *Int. J. High Perform. Comput. Appl.*, vol. 22, no. 4, pp. 438–448, Nov. 2008.

[6] C. Song, Y. Li, and B. Huang, "A GPU-accelerated wavelet decompression system with SPIHT and Reed-Solomon decoding for satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 683–690, Sep. 2011.

[7] S.-C. Wei and B. Huang, "GPU acceleration of predictive partitioned vector quantization for ultraspectral sounder data compression," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 677–682, Sep. 2011.

[8] D. Donoho, "Compressed Sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[9] J. E. Candès, T. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[10] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenom.*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

[11] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing: geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[12] D. Takhar *et al.*, "A new compressive imaging camera architecture using optical-domain compression," *Comput. Imag. IV*, vol. 6065, p. 4317, 2006.

[13] 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, Mar. 2011.

[14] M. E. Gehm, R. John, D. D. Brady, R. Willett, and T. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Exp.*, vol. 15, no. 21, pp. 14 013–14 027, Oct. 2007.

[15] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, Apr. 2008.

[16] C. Li, T. Sun, K. Kelly, and Y. Zhang, "A compressive sensing and unmixing scheme for hyperspectral data processing," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1200–1210, Mar. 2011.

[17] Q. Zhang, R. Plemmons, D. Kittle, D. Brady, and S. Prasad, "Joint segmentation and reconstruction of hyperspectral data with compressed measurements," *Appl. Opt.*, vol. 50, no. 22, pp. 4417–4435, Aug. 2011.

[18] M. Golbabaee and P. Vandergheynst, "Joint trace/TV norm minimization: A new efficient approach for spectral compressive imaging," in *Proc. 19th IEEE ICIP*, 2012, pp. 933–936.

[19] M. Golbabaee, S. Arberet, and P. Vandergheynst, "Compressive source separation: Theory and methods for hyperspectral imaging," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5096–5110, Dec. 2013.

[20] G. Martin, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral coded aperture: A new algorithm for hyperspectral compressive sensing," in *Proc. IEEE IGARSS*, 2012, vol. 1, pp. 1–4.

[21] G. Martin, J. M. Bioucas-Dias, and A. Plaza, "A new technique for hyperspectral compressive sensing using spectral unmixing," in *Proc. SPIE Opt. Photon., Satell. Data Compression, Commun., Process. Conf.*, 2012, vol. 1, pp. 1–12.

[22] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," in *Proc. Nat. Acad. Sci.*, 2003, vol. 100, pp. 2197–2202.

[23] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[24] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via lq-minimization for 0," *Appl. Comput. Harmonic Anal.*, vol. 26, no. 3, pp. 395–407, May 2009.

[25] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[26] J. Tropp, "Just relax: Convex programming methods for subset selection and sparse approximation," in *Ptoc. ICES Rep.*, 2004, pp. 1–39.

[27] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[28] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[29] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Statist. Soc. Ser. B Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[30] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[31] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmonic Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[32] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmonic Anal.*, vol. 27, no. 3, pp. 265–274, Nov. 2009.

[33] R. Garg and R. Khandekar, "Gradient descent with sparsification: An iterative algorithm for sparse recovery with restricted isometry property," in *Proc. 26th Annu. ACM Int. Conf. Mach. Learning*, 2009, pp. 337–344.

[34] S. Foucart, "Hard thresholding pursuit: An algorithm for compressive sensing," *SIAM J. Numer. Anal.*, vol. 49, no. 6, pp. 2543–2563, Nov. 2011.

[35] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[36] J. Portilla, "Image restoration through l0 analysis-based sparse optimization in tight frames," in *Proc. 16th IEEE ICIP*, 2009, pp. 3909–3912.

[37] I. W. Selesnick and M. A. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," in *SPIE Opt. Eng. Appl. Int. Soc. Opt. Photon.*, 2009, pp. 1–15.

[38] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Appl. Comput. Harmonic Anal.*, vol. 34, no. 1, pp. 30–56, Jan. 2013.

[39] R. D. R. Baraniuk, M. Davenport, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

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

[41] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1/2, pp. 89–97, Jan. 2004.

[42] X. Bresson and T. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inv. Prob. Imag.*, vol. 2, no. 4, pp. 455–484, Nov. 2008.

[43] L. Z. Q. Yuan and H. She, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, p. 36603677, 2012.

[44] M. Lustig, D. L. Donoho, J. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.

[45] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," *SIAM J. Imag. Sci.*, vol. 6, no. 2, pp. 1035–1058, 2013.

[46] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programm.*, vol. 5, no. 1–3, pp. 293–318, Apr. 1992.

[47] E. Esser, "Applications of Lagrangian-based alternating direction methods and connections to split-bregman," Univ. California, Oakland, CA, USA, Tech. Rep. 09-31, 2009.

[48] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.

[49] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

Gabriel Martín received the Computer Engineer degree in 2008, the M.Sc. degree in 2010, and the Ph.D. in 2013 from the University of Extremadura, Plasencia, Spain.

He was a Predoctoral Research Associate (funded by the Spanish Ministry of Science and Innovation) with the Hyperspectral Computing Laboratory and is now a Postdoctoral Researcher (funded by FCT) in the Instituto Superior Técnico, Technical University of Lisbon, Lisbon, Portugal, where he is developing research on compressive sensing and efficient hardware implementations for remotely sensed hyperspectral images. His research interests comprise the development of new techniques for unmixing remotely sensed hyperspectral data sets, as well as the efficient processing and interpretation of these data in different types of high-performance computing architectures.

Dr. Martín has served as a Reviewer for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

José M. Bioucas-Dias (S'87–M'95) received the EE, M.Sc., Ph.D., and "Agregado" degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), the Engineering School of the Technical University of Lisbon (TULisbon), Lisbon, Portugal, in 1985, 1991, 1995, and 2007, respectively.

Since 1995, he has been with the Department of Electrical and Computer Engineering, IST, where he was an Assistant Professor from 1995 to 2007 and an Associate Professor since 2007. Since 1993, he has been also a Senior Researcher with the Pattern and Image Analysis Group of the Instituto de Telecomunicaçoes, which is a private nonprofit research institution. His research interests include inverse problems, signal and image processing, pattern recognition, optimization, and remote sensing.

Dr. Bioucas-Dias was an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS(1997–2000) and is an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Guest Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING for the Special Issue on Spectral Unmixing of Remotely Sensed Data and of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING for the Issue on Hyperspectral Image and Signal Processing, and he is a Guest Editor of IEEE SIGNAL PROCESSING MAGAZINE for the Special Issue on Signal and Image Processing in Hyperspectral Remote Sensing, He was the General Cochair of the third IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS'2011) and has been a member of program/technical committees of several international conferences.

Antonio Plaza (SM'05) was born in Caceres, Spain, in 1975.

He is an Associate Professor (with accreditation for Full Professor) with the Department of Technology of Computers and Communications, University of Extremadura, Plasencia, Spain, where he is the Head of the Hyperspectral Computing Laboratory (HyperComp). His main research interests comprise remotely sensed hyperspectral image analysis and efficient implementations of large-scale scientific problems on high-performance computing architectures. He has been the Advisor of 12 Ph.D. dissertations and more than 30 M.Sc. dissertations. He was the Coordinator of the Hyperspectral Imaging Network, a European project with total funding of 2.8 million Euros. He authored more than 400 publications, including 126 JCR journal papers (78 in IEEE journals), 20 book chapters, and over 240 peer-reviewed conference proceeding papers (94 in IEEE conferences). He has edited the book *High-Performance Computing in Remote Sensing* (CRC Press/Taylor and Francis) (the first book on this topic in the published literature) and guest edited eight special issues on hyperspectral remote sensing for different journals.

Dr. Plaza is a recipient of the recognition of Best Reviewers of the IEEE Geoscience and Remote Sensing Letters (in 2009) and a recipient of the recognition of Best Reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (in 2010), a journal for which he served as Associate Editor in 2007–2012. He is also an Associate Editor for IEEE ACCESS, and was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012) and the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE (2013). He was also a member of the steering committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He is currently an Associate Editor for the *Journal of Real-Time Image Processing*. He is a recipient of the 2013 Best Paper Award of the JSTARS journal, and a recipient of the most highly cited paper (2005–2010) in the *Journal of Parallel and Distributed Computing*. He is a coauthor of the 2011 Best Student Paper at the IEEE International Conference on Space Technology, and a recipient of the 2008 Best Paper award at the IEEE Symposium on Signal Processing and Information Technology. He is a recipient of the Best Ph.D. Dissertation award at University of Extremadura in 2002. Four of his students have received the Best Ph.D. Dissertation award at University of Extremadura, and one of his students has received the Best Ph.D. Dissertation award at Complutense University of Madrid, Spain. He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012, and is currently serving as President of the Spanish Chapter of IEEE GRSS (since November 2012). He has served as a proposal evaluator for the European Commission (Marie Curie Actions, Engineering Panel), the European Space Agency, the Belgium Science Policy, the Israel Science Foundation, and the Spanish Ministry of Science and Innovation. He has participated in the Tenure Track Selection Committee of different universities in Italy, Spain, and Australia. He has reviewed more than 500 manuscripts for over 50 different journals. He is also currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal.