

# A Compressive Sensing and Unmixing Scheme for Hyperspectral Data Processing

Chengbo Li, Ting Sun, Kevin Kelly, and Yin Zhang

## Abstract

Hyperspectral data processing typically demands enormous computational resources in terms of storage, computation and I/O throughputs, especially when real-time processing is desired. In this paper, we investigate a low-complexity scheme for hyperspectral data compression and reconstruction. In this scheme, compressed hyperspectral data are acquired directly by a device similar to the single-pixel camera [5] based on the principle of compressive sensing. To decode the compressed data, we propose a numerical procedure to directly compute the unmixed abundance fractions of given endmembers, completely bypassing high-complexity tasks involving the hyperspectral data cube itself. The reconstruction model is to minimize the total variation of the abundance fractions subject to a pre-processed fidelity equation with a significantly reduced size, and other side constraints. An augmented Lagrangian type algorithm is developed to solve this model. We conduct extensive numerical experiments to demonstrate the feasibility and efficiency of the proposed approach, using both synthetic data and hardware-measured data. Experimental and computational evidences obtained from this study indicate that the proposed scheme has a high potential in real-world applications.

## Index Terms

Hyperspectral imaging, data unmixing, compressive sensing, total variation, augmented Lagrangian method, fast Walsh-Hadamard transform.

## I. INTRODUCTION

Hyperspectral imaging is a crucial technique and a powerful tool to identify and quantify distinct material substances from (often remotely) observed spectral data. It employs hyperspectral sensors to collect two dimensional spatial images over many contiguous spectral bands containing the visible, near-infrared, and shortwave infrared spectral bands [10]. Hyperspectral imaging has a wide range of applications such as terrain classification, mineral

Chengbo Li is with the Department of Computational and Applied Mathematics, Rice University, 6100 Main Street, Houston TX 77005 (phone: (832) 967-7907; email: Chengbo.Li@rice.edu).

Ting Sun is with the Department of Electrical and Computer Engineering, Rice University, 6100 Main Street, Houston TX 77005 (phone: (832) 755-9378; email: ting.sun@rice.edu).

Kevin Kelly is with the Department of Electrical and Computer Engineering, Rice University, 6100 Main Street, Houston TX 77005 (phone: (713) 348-3565; email: kkelly@rice.edu).

Yin Zhang is with the Department of Computational and Applied Mathematics, Rice University, 6100 Main Street, Houston TX 77005 (phone: (713) 348-5744; email: yzhang@rice.edu).

detection and exploration [13], [14], pharmaceutical counterfeiting [15], environmental monitoring, and military surveillance [16].

Typically, hyperspectral imaging is of spatially low resolution, in which each pixel, from a given spatial element of resolution and at a given spectral band, is a mixture of several different material substances, termed endmembers, each possessing a characteristic hyperspectral signature [11]. Hyperspectral unmixing is to decompose each pixel spectrum to identify and quantify the relative abundance of each endmember. In the linear mixing model, interactions among distinct endmembers are assumed to be negligible [12], which is a plausible hypothesis in most cases. Frequently, the representative endmembers for a given scene are known *a priori* and their signatures can be obtained from a spectral library (e.g., ASTER and USGS) or codebook. On the other hand, when endmembers are unknown but the hyperspectral data is fully accessible, many algorithms exist for determining endmembers in a scene, including N-FINDR [18], PPI (pixel purity index) [17], VCA (vertex component analysis) [19], SGA (simplex growing algorithm) [20]; NMF-MVT (nonnegative matrix factorization minimum volume transform) [21], SISAL (simplex identification via split augmented Lagrangian) [22], MVSA (minimum volume simplex analysis) [24], and MVES (minimum-volume enclosing simplex) [23].

Because of their enormous volume, it is particularly difficult to directly process and analyze hyperspectral data cubes in real time or near real time. On the other hand, hyperspectral data are highly compressible with two-fold compressibility: 1) each spatial image is compressible, and 2) the entire cube, when treated as a matrix, is of low rank. To fully exploit such rich compressibility, in this paper we propose a scheme that never requires to explicitly store or process a hyperspectral cube itself. In this scheme, data are acquired by means of compressive sensing (CS). The theory of CS shows that a sparse or compressible signal can be recovered from a relatively small number of linear measurements (see, for example, [1], [2], [3]). In particular, the concept of the single pixel camera [5] can be extended to the acquisition of compressed hyperspectral data, which will be used in our experiments. The main novelty of the scheme is in the decoding side where we combine data reconstruction and unmixing into a single step of much lower complexity. At this point, we assume that the involved endmember signatures are known and given, from which we then directly compute abundance fractions. In a later study we will extend this approach to blind unmixing where endmember signatures are not precisely known *a priori*. For brevity, we will call the proposed procedure *compressive sensing and unmixing* or CSU scheme.

In the following paragraphs, we introduce the main contributions, and the notation and organization of this paper.

#### A. Main Contributions

We propose and conduct a proof-of-concept study on a low-complexity, compressive sensing and unmixing (CSU) scheme, formulating a unmixing model based on total variation (TV) minimization [4], developing an efficient algorithm to solve it, and providing experimental and numerical evidence to validate the scheme. This proposed scheme directly unmixes compressively sensed data, bypassing the high-complexity step of reconstructing the hyperspectral cube itself. The validity and potential of the proposed CSU scheme are demonstrated by experiments using both synthetic and hardware-measured data.

## B. Notations

We introduce necessary notations here. Suppose that in a given scene there exist  $n_e$  significant endmembers, with spectral signatures  $w_i^T \in \mathbb{R}^{n_b}$ , for  $i = 1, \dots, n_e$ , where  $n_b \geq n_e$  denotes the number of spectral bands. Let  $x_i \in \mathbb{R}^{n_b}$  represent the hyperspectral data vector at the  $i$ -th pixel and  $h_i^T \in \mathbb{R}^{n_e}$  represent the abundance fractions of the endmembers for any  $i \in \{1, \dots, n_p\}$ , where  $n_p$  denotes the number of pixels. Furthermore, let  $X = [x_1, \dots, x_{n_p}]^T \in \mathbb{R}^{n_p \times n_b}$  denote a matrix representing the hyperspectral cube,  $W = [w_1, \dots, w_{n_e}]^T \in \mathbb{R}^{n_e \times n_b}$  the mixing matrix containing the endmember spectral signatures, and  $H = [h_1, \dots, h_{n_p}]^T \in \mathbb{R}^{n_p \times n_e}$  a matrix holding the respective abundance fractions.  $\mathbf{1}_s$  denotes the column vector of all ones with length  $s$ . We use  $A \in \mathbb{R}^{m \times n_p}$  to denote the measurement matrix in compressive sensing data acquisition, and  $F \in \mathbb{R}^{m \times n_b}$  to denote the observation matrix, where  $m < n_p$  is the number of samples for each spectral band.

## C. Organization

The paper is organized as follows. Section II focuses on formulating our unmixing model. Section III introduces a data preprocessing technique to significantly reduce the problem size and thus complexity. Section IV describes a variable splitting, augmented Lagrangian algorithm for solving the proposed unmixing model. Section V presents numerical results based on synthetic data. Section VI describes a hardware setup and its implementation to collect compressed hyperspectral data, presents and analyzes the performance of the CSU scheme on a hardware-measured dataset. Finally, Section VII gives concluding remarks.

## II. PROBLEM FORMULATION

Assuming negligible interactions among endmembers, the hyperspectral vector  $x_i$  at the  $i$ -th pixel can be regarded as a linear combination of the endmember spectral signatures, and the weights are gathered in a nonnegative abundance vector  $h_i$ . Ideally, the components of  $h_i$ , representing abundance fractions, should sum up to unity; i.e., the hyperspectral vectors lie in the convex hull of endmember spectral signatures [19]. In short, the data model has the form

$$X = HW, \quad H\mathbf{1}_{n_e} = \mathbf{1}_{n_p}, \quad \text{and} \quad H \geq 0. \quad (1)$$

However, in reality the sum-to-unity condition on  $H$  does not usually hold due to imprecisions and noise of various kinds. In our implementation, we imposed this condition on synthetic data, but skipped it for measured data.

Since each column of  $X$  represents a 2D image corresponding to a particular spectral band, we can collect the compressed hyperspectral data  $F \in \mathbb{R}^{m \times n_b}$  by randomly sampling all the columns of  $X$  using the same measurement matrix  $A \in \mathbb{R}^{m \times n_p}$ , where  $m < n_p$  is the number of samples for each column. Mathematically, the data acquisition model can be described as

$$AX = F. \quad (2)$$

Combining (1) and (2), we obtain constraints

$$AHW = F, \quad H\mathbf{1}_{n_e} = \mathbf{1}_{n_p}, \quad \text{and} \quad H \geq 0. \quad (3)$$

For now, we assume that the endmember spectral signatures in  $W$  are known, our goal is to find their abundance distributions (or fractions) in  $H$ , given the measurement matrix  $A$  and the compressed hyperspectral data  $F$ . In general, system (3) is not sufficient for determining  $H$ , necessitating the use of some prior knowledge about  $H$  in order to find it.

In compressive sensing, regularization by  $\ell_1$  minimization has been widely used. However, it has been empirically shown that the use of TV regularization is generally more advantageous on image problems since it can better preserve edges or boundaries in images that are essential characteristics of most images. TV regularization puts emphasis on sparsity in the gradient map of the image and is suitable when the gradient of the underlying image is sparse [2]. In our case, we make the reasonable assumption that the gradient of each image composed by abundance fractions for each endmember is mostly and approximately piecewise constant. Therefore, we propose to recover the abundance matrix  $H$  by solving the following unmixing model:

$$\min_{H \in \mathbb{R}^{n_p \times n_e}} \sum_{j=1}^{n_e} \text{TV}(He_j) \quad \text{s.t.} \quad AHW = F, H\mathbf{1}_{n_e} = \mathbf{1}_{n_p}, H \geq 0, \quad (4)$$

where  $e_j$  is the  $j$ -th standard unit vector in  $\mathbb{R}^{n_p}$ ,

$$\text{TV}(He_j) \triangleq \sum_{i=1}^{n_p} \|D_i(He_j)\|, \quad (5)$$

$\|\cdot\|$  is the 2-norm in  $\mathbb{R}^2$ , and  $D_i \in \mathbb{R}^{2 \times n_p}$  denotes the discrete gradient operator at the  $i$ -th pixel. Since the unmixing model directly uses compressed data  $F$ , we will call it a *compressed unmixing* model.

### III. SVD PREPROCESSING

The size of the fidelity equation  $AHW = F$  in (3) is  $m \times n_b$  where  $m$ , although less than  $n_p$  in compressive sensing, can still be quite large, and  $n_b$ , the number of spectral bands, typically ranges from hundreds to thousands. We propose a preprocessing procedure based on singular value decomposition of the observation matrix  $F$  to decrease the size of the fidelity equations from  $m \times n_b$  to  $m \times n_e$ . Since the number of endmembers  $n_e$  is typically up to two orders of magnitude smaller than  $n_b$ , the resulting reduction in complexity is significant, potentially enabling near-real-time processing speed. The proposed preprocessing procedure is based on the following result.

*Proposition 1:* Let  $A \in \mathbb{R}^{m \times n_p}$  and  $W \in \mathbb{R}^{n_e \times n_b}$  be full-rank, and  $F \in \mathbb{R}^{m \times n_b}$  be rank- $n_e$  with  $n_e < \min\{n_b, n_p, m\}$ . Let  $F = U_e \Sigma_e V_e^T$  be the economy-size singular value decomposition of  $F$  where  $\Sigma_e \in \mathbb{R}^{n_e \times n_e}$  is diagonal and positive definite,  $U_e \in \mathbb{R}^{m \times n_e}$  and  $V_e \in \mathbb{R}^{n_b \times n_e}$  both have orthonormal columns. Assume that  $\text{rank}(WV_e) = n_e$ , then the two linear systems below for  $H \in \mathbb{R}^{n_p \times n_e}$  have the same solution set; i.e., the equivalence holds

$$AHW = F \iff AHWV_e = U_e \Sigma_e. \quad (6)$$

*Proof:* Denote  $\mathcal{H}_1 = \{H : AHW = F\}$  and  $\mathcal{H}_2 = \{H : AHWV_e = U_e \Sigma_e\}$ . Given  $F = U_e \Sigma_e V_e^T$ , it is obvious that  $\mathcal{H}_1 \subseteq \mathcal{H}_2$ . To show  $\mathcal{H}_1 = \mathcal{H}_2$ , it suffices to verify that  $\dim(\mathcal{H}_1) = \dim(\mathcal{H}_2)$ .

Let “vec” denote the operator that stacks the columns of a matrix to form a vector. By well-known properties of Kronecker product “ $\otimes$ ”,  $AHW = F$  is equivalent to

$$(W^T \otimes A) \text{vec}H = \text{vec}F, \quad (7)$$

where  $W^T \otimes A \in \mathbb{R}^{(n_b m) \times (n_e n_p)}$ , and

$$\text{rank}(W^T \otimes A) = \text{rank}(W)\text{rank}(A) = n_e m. \quad (8)$$

Similarly,  $AHWW_e = U_e \Sigma_e$  is equivalent to

$$((WV_e)^T \otimes A) \text{vec}H = \text{vec}(U_e \Sigma_e), \quad (9)$$

where  $(WV_e)^T \otimes A \in \mathbb{R}^{(n_e m) \times (n_e n_p)}$  and, under our assumption  $\text{rank}(WV_e) = n_e$ ,

$$\text{rank}((WV_e)^T \otimes A) = \text{rank}(WV_e)\text{rank}(A) = n_e m. \quad (10)$$

Hence,  $\text{rank}(W^T \otimes A) = \text{rank}((WV_e)^T \otimes A)$ , which implies the solution sets of (7) and (9) have the same dimension; i.e.,  $\dim(\mathcal{H}_1) = \dim(\mathcal{H}_2)$ . Since  $\mathcal{H}_1 \subseteq \mathcal{H}_2$ , we conclude that  $\mathcal{H}_1 = \mathcal{H}_2$ . ■

This proposition ensures that under a mild condition the matrices  $W$  and  $F$  in the fidelity equation  $AHW = F$  can be replaced, without changing the solution set, by the much smaller matrices  $WV_e$  and  $U_e \Sigma_e$ , respectively, potentially leading to multi-order magnitude reductions in equation sizes.

Suppose that  $F$  is a observation matrix for a rank- $n_e$  hyperspectral data matrix  $\hat{X}$ . Then  $F = A\hat{H}\hat{W}$  for some full rank matrices  $\hat{H} \in \mathbb{R}^{n_p \times n_e}$  and  $\hat{W} \in \mathbb{R}^{n_e \times n_b}$ . Clearly, the rows of  $\hat{W}$  span the same space as the columns of  $V_e$  do. Therefore, the condition  $\text{rank}(WV_e) = n_e$  is equivalent to  $\text{rank}(W\hat{W}^T) = n_e$ , which definitely holds for  $W = \hat{W}$ . It will also hold for a random  $W$  with high probability. Indeed, the condition  $\text{rank}(WV_e) = n_e$  is rather mild.

In practice, the observation matrix  $F$  usually contains model imprecisions or random noise, and hence is unlikely to be exactly rank  $n_e$ . In this case, truncating the SVD of  $F$  to rank- $n_e$  is a sensible strategy, which will not only serve the dimension reduction purpose, but also a denoising purpose because the SVD truncation annihilates insignificant singular values of  $F$  likely caused by noise. Motivated by these considerations, we propose the following SVD preprocessing procedure.

*Algorithm 1 (SVD Preprocessing):*

**Input**  $F$ ,  $W$  and  $n_e$ .

**Do** the following:

compute the rank- $n_e$  principal SVD:  $F \approx U_e \Sigma_e V_e^T$ ;

overwrite data:  $W \leftarrow WV_e$  and  $F \leftarrow U_e \Sigma_e$ ;

**End**

**Output**  $F$  and  $W$ .

#### IV. ALGORITHM

Our computational experience indicates that, at least for the problems we tested so far, to obtain good solutions it suffices to solve a simplified compressed unmixing model that omits the nonnegativity of  $H$ ,

$$\min_H \sum_{j=1}^{n_e} \text{TV}(He_j) \quad \text{s.t.} \quad AHW = F, \quad H\mathbf{1}_{n_e} = \mathbf{1}_{n_p}. \quad (11)$$

For simplicity, we will discuss our algorithm for the above model which was actually used in our numerical experiments. In fact, in our experiments with hardware-measured data, we also omitted the second constraint above since it would not help in the presence of sizable system imprecisions and noise. It should also be emphasized that in the compressed unmixing model (11), the matrices  $W$  and  $F$  are the output from the SVD preprocessing procedure. In particular, the size of the fidelity equation has been reduced to  $m \times n_e$  from the original size  $m \times n_b$ , a factor of  $n_b/n_e$  reduction in size.

The main algorithm we proposed here is based on the augmented Lagrangian method framework and a variable splitting formulation, which is an extension to the algorithm TVAL3 [6]. Wang, Yang, Yin, and Zhang [7] first introduced the splitting formulation into TV regularization problems, and applied a penalty algorithm to the formulation. Then Goldstein and Osher [8] added Bregman regularization into the formulation, producing a faster algorithm since it is equivalent to augmented Lagrangian multiplier method. In 2009, Li, Zhang, and Yin also employed this set of ideas and developed an efficient TV regularization solver TVAL3.

To separate the discrete gradient operator from the non-differentiable TV term, we introduce splitting variables  $v_{ij} = D_i(He_j)$  for  $i = 1, \dots, n_p$  and  $j = 1, \dots, n_e$ . Then (11) is equivalent to

$$\min_{H, v_{ij}} \sum_{i,j} \|v_{ij}\| \quad \text{s.t.} \quad D_i(He_j) = v_{ij}, \quad \forall i, j, \quad AHW = F, \quad H\mathbf{1}_{n_e} = \mathbf{1}_{n_p}. \quad (12)$$

The augmented Lagrangian function for (12) can be written as

$$\begin{aligned} \mathcal{L}_A(H, v_{ij}) \quad \triangleq \quad & \sum_{i,j} \left\{ \|v_{ij}\| - \lambda_{ij}^T (D_i(He_j) - v_{ij}) + \frac{\alpha}{2} \|D_i(He_j) - v_{ij}\|_2^2 \right\} - \\ & \langle \Pi, AHW - F \rangle + \frac{\beta}{2} \|AWH - F\|_F^2, \quad -\nu^T (H\mathbf{1}_{n_e} - \mathbf{1}_{n_p}) + \frac{\gamma}{2} \|H\mathbf{1}_{n_e} - \mathbf{1}_{n_p}\|_2^2, \end{aligned} \quad (13)$$

where  $\lambda_{ij}, \Pi, \nu$  are multipliers of appropriate sizes, and  $\alpha, \beta, \gamma > 0$  are penalty parameters corresponding to the three sets of constraints in (12), respectively. For brevity, we have omitted the multipliers in the argument list of  $\mathcal{L}_A$ .

We apply the augmented Lagrangian method (see Hestenes [25] and Powell [26]) on (12) which minimizes the augmented Lagrangian function  $\mathcal{L}_A$  for fixed multipliers, then updates the multipliers. Specifically, in our case the multipliers are updated as follows. For all  $1 \leq i \leq n_p$  and  $1 \leq j \leq n_e$ ,

$$\lambda_{ij} \leftarrow \lambda_{ij} - \alpha(D_i(He_j) - v_{ij}), \quad \Pi \leftarrow \Pi - \beta(AHW - F), \quad \nu \leftarrow \nu - \gamma(H\mathbf{1}_{n_e} - \mathbf{1}_{n_p}). \quad (14)$$

### A. Alternation Minimization

To minimize  $\mathcal{L}_A(H, v_{ij})$  efficiently, we employ an alternating minimization scheme; i.e., minimizing  $\mathcal{L}_A(H, v_{ij})$  with respect to  $v$  and  $H$ , one at a time, until convergence is achieved. The minimization problem with respect to  $v_{ij}$ 's is separable and has closed-form solutions  $v_{ij}^*$  according to the well-known shrinkage formula:

$$v_{ij}^* = \max \left\{ \|\theta_{ij}\| - \frac{1}{\alpha}, 0 \right\} \frac{\theta_{ij}}{\|\theta_{ij}\|}, \quad (15)$$

where

$$\theta_{ij} \triangleq D_i(H e_j) - \frac{\lambda_{ij}}{\alpha}. \quad (16)$$

On the other hand, minimizing the augmented Lagrangian with respect to  $H$  can be excessively costly for large-scale problems. Fortunately, in the alternating minimization scheme, it is unnecessary to carry out such a minimization step to a high accuracy. All we need is to sufficiently decrease the augmented Lagrangian function. In our implementation, we take only one gradient step on  $H$  from the current iterate; i.e.,

$$H \leftarrow H - \tau \mathcal{G}(H), \quad (17)$$

where  $\mathcal{G}(H)$  denotes the gradient of  $\mathcal{L}_A(H, v_{ij})$  with respect to  $H$ , which can be derived as

$$\mathcal{G}(H) = \sum_{i,j} \left\{ -D_i^T \lambda_{ij} e_j^T + \alpha D_i^T (D_i H e_j - v_{ij}) e_j^T \right\} - A^T \Pi W^T + \beta A^T (A H W - F) W^T - \nu \mathbf{1}_{n_e}^T + \gamma (H \mathbf{1}_{n_e} - \mathbf{1}_{n_p}) \mathbf{1}_{n_e}^T. \quad (18)$$

The only remaining issue is to choose the step length  $\tau$  in (17), for which we adapt a scheme used in [6]. In this scheme, the step size  $\tau$  is determined by a non-monotone line search scheme [28] to satisfy a so-called ‘‘non-monotone Armijo condition’’. We start from an initial step proposed by Barzilai and Borwein [27] for gradient type method that we will call a BB-step, then use a backtracking technique to search for a step satisfying the non-monotone Armijo condition. To sum up, the alternating minimization algorithm for minimizing the augmented Lagrangian function (13), for fixed multipliers, is as follows.

*Algorithm 2 (Alternating Minimization):*

**Input** starting point  $H$  and all other necessary quantities.  
**While** ‘‘inner stopping criteria’’ are not satisfied,  
    compute  $v_{ij}$  by the *shrinkage formula* (15);  
    compute the *BB step*  $\tau$  (see [27]), and set  $\rho \in (0, 1)$ ;  
    **While** ‘‘non-monotone Armijo condition’’ is not satisfied,  
        Backtracking:  $\tau = \rho \tau$ ;  
    **End**  
    update  $H$  by formula (17);  
**End**

**Output**  $H$  and  $v_{ij}$  for all  $i, j$ .

### B. Overall Algorithm

Putting all components together, our algorithm for solving the compressed unmixing model (11) can be summarized as follows.

*Algorithm 3 (Compressed Unmixing):*

**Input** data  $F$  and  $W$ , and penalty parameters  $\alpha, \beta, \gamma > 0$ .

Preprocess  $F$  and  $W$  by Algorithm 1;

Initialize multipliers  $\lambda_{ij}, \Pi, \nu$ , and variable  $H$ .

**While** “outer stopping criteria” are not satisfied,

update variables  $v_{ij}$  and  $H$  by Algorithm 2;

update multipliers  $\lambda_{ij}, \Pi, \nu$  by formulas in (14);

**End**

**Output**  $H$ .

The complexity of Algorithm 3 at each iteration is dominated by two matrix multiplications involving  $W^T \otimes A$  and its transpose, respectively. In the next two sections, we will demonstrate the effectiveness of the algorithm in several sets of numerical experiments.

In Algorithm 3, the outer stopping criteria can be specified based on either relative change of variables or the optimality conditions of the compressed unmixing model (11). While the latter is more rigorous, it is also more costly. In our experiments, we used relative change of variables in both outer and inner stopping criteria. Specific parameter settings and initial values used our experiments are given in the next section.

## V. EXPERIMENTAL RESULTS: SYNTHETIC DATA

### A. Setup of Experiments

To demonstrate the feasibility, practicality and potential of the proposed CSU scheme, We will present numerical results from applying the proposed CSU scheme to two types of data. In this section, results are obtained on simulated or synthetic datasets. In the next section, we provide results from a much more realistic simulation where compressed hyperspectral data were directly measured by a hardware apparatus.

We implemented the CSU scheme in a Matlab code which is still at an early stage of development. All numerical experiments reported in this paper were performed on a SONY VGN-FZ290 laptop running Windows 7 and MATLAB R2009b (32-bit), equipped with a 1.5GHz Intel Core 2 Duo CPU T5250 and 2GB of DDR2 memory.



In both types of experiments, we use randomized Walsh-Hadamard matrices as measurement matrices,  $A$ , considering that they permit fast transformation and easy hardware implementation. A Walsh-Hadamard matrix is randomized by choosing  $m$  random row from it and applying a random permutation to its columns.

In Algorithm 3, the multipliers  $\lambda_{ij}$ ,  $\Pi$ , and  $\nu$  are always initialized to 0; the backtracking parameter is  $\rho = 0.6$ ; the penalty parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  were selected from a range of  $2^5$  to  $2^9$ , according to estimated noise levels. Despite of a lack of theoretical guidance, we have found that it is not particularly difficult to choose adequate values for these penalty parameters since the algorithm is not overly sensitive to such values as long as they fall into some appropriate but reasonably wide range. It takes a bit experience, and often a few trial-and-error attempts, to find acceptable penalty parameter values for a given class of problems.

### B. Test Results on Synthetic Data

In the first test, we generated compressed data according to data acquisition model (3). We selected 4 endmembers from the ASTER Spectral Library [30]: nontronite, ferroaxinite, trona, and molybdenite, whose spectral signatures are shown in Figure 2. A total of 211 bands were selected in the range of 0.4 to 2.5 micrometers. The distributions of abundance fractions corresponding to 4 endmembers were given in Figure 1 with a spatial resolution of  $64 \times 64$ . Figures 1 and 2 gives the “true”  $H$  and  $W$ , respectively, from which we generated an observation matrix  $F = AHW$  for some measurement matrix  $A$ . In addition, to test the robustness of the CSU scheme, in some experiments we added zero-mean Gaussian random noise with stand derivation 0.8 to the observation matrix  $F$ .

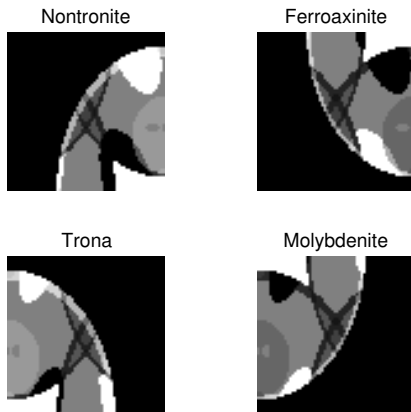


Fig. 1. Synthetic abundance distributions.

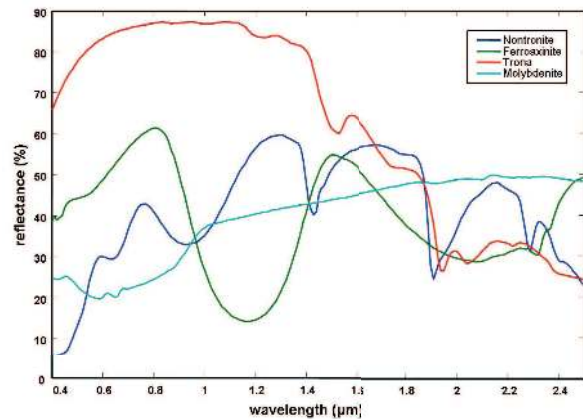


Fig. 2. Endmember spectral signatures.

In Figure 3, we plot relative errors in computed abundance fractions versus measurement rate of compressed data on 100 distinct testing points, with or without additive noise. The average elapsed time for these runs is less than 10 seconds. We observe that the CSU scheme attains relative error less than 1% when measurement rate is greater than 20% in both noisy and noise-free cases. This test empirically validates the convergence of the algorithm and the feasibility of the proposed CSU scheme, which has inspired us conducting further tests on larger and more realistic problems.

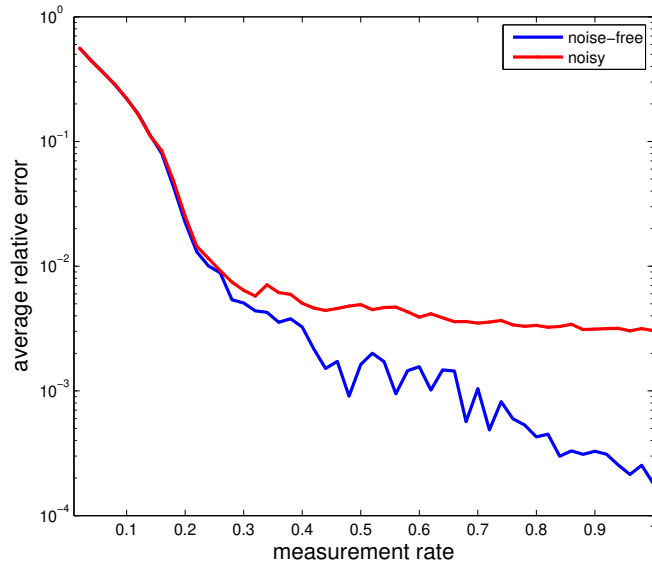


Fig. 3. Recoverability for noisy and noise-free cases.

In the second test, we generated a compressed data matrix  $F$  by applying the data acquisition model (2) to the publicly available HYDICE Urban hyperspectral data [31], which contains 163 bands in a range from 0.4 to 2.5 micrometers, after some water absorption bands, each having a  $307 \times 307$  resolution. According to the analysis of this Urban data cube in [9], there are 6 significant endmembers in the scene — road, metal, dirt, grass, tree, and roof, as is shown in Figure 4. The spectral signatures for these 6 selected endmembers are plotted in Figure 5.



Fig. 4. “Urban” image and endmember selection.

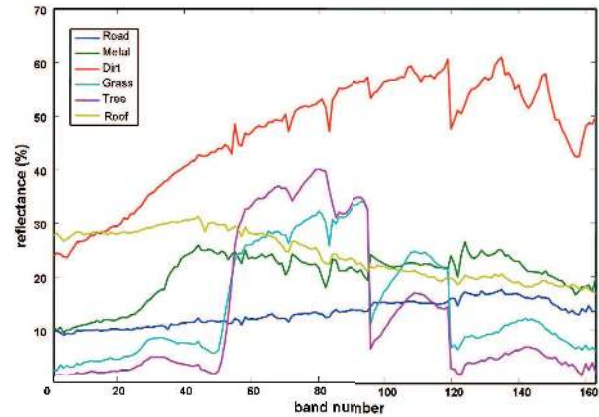


Fig. 5. Spectral signatures with water absorption bands abandoned.

Our computed unmixing result from 25% measurements are given in Figure 6, where six subfigures depict the computed distributions of abundance fractions for the six endmembers, respectively. It took about 215 seconds to run the algorithm. Qualitatively, we see that features in the original image such as roads, plants, and buildings have

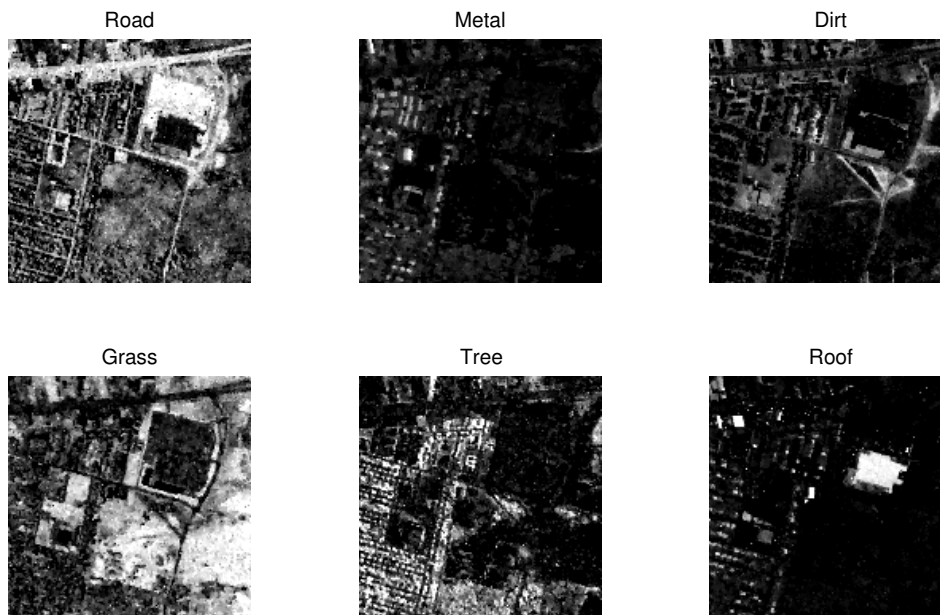


Fig. 6. Computed abundance: solution obtained from 25% of measurements.

been properly segmented by a visual comparison with Figure 4. For example, a hunk of roof marked by number 6 in Figure 4 appears prominently in the lower-right subfigure of Figure 6 for the abundance fractions of roof.

Figure 7 shows the least squares solution from directly solving  $AHW = F$  for  $H$  with 100% data, which becomes an overdetermined linear system in this case. Comparing Figure 6 with Figure 7, we observe that the proposed CSU scheme, using 25% of the data, is capable of keeping important features and most details, even though the overall quality in computed abundance fractions by the CSU scheme is slightly lower than that of the least squares solution using 100% of the data.

## VI. EXPERIMENTAL RESULTS: HARDWARE-MEASURED DATA

### A. Hardware Implementation

This section contains experimental results using hardware-measured data. Figure 8 shows the schematic of a compressing sensing hyperspectral imaging system based on a digital micro-mirror device (DMD). This system incorporates a micro-mirror array driven by pseudo-random patterns and one spectrometer. Similar to the single-pixel camera setup [5], it optically samples incoherent image measurements as dictated by the CS theory; then a reconstruction algorithm is applied to recover the acquired spatial image as well as spectral information.

The spectrometer (on the right) we employed is a USB4000 by Ocean Optics which features a 3648-element linear array detector responsive from 200-1100 nm. The spectrometer and DMD (at the top) are synchronized to take data when the pseudo-random pattern switches. For each such a pattern, the measured data from the spectrometer is represented as a linear vector with the length of 3648. The target (at the bottom) is illuminated by two 35W

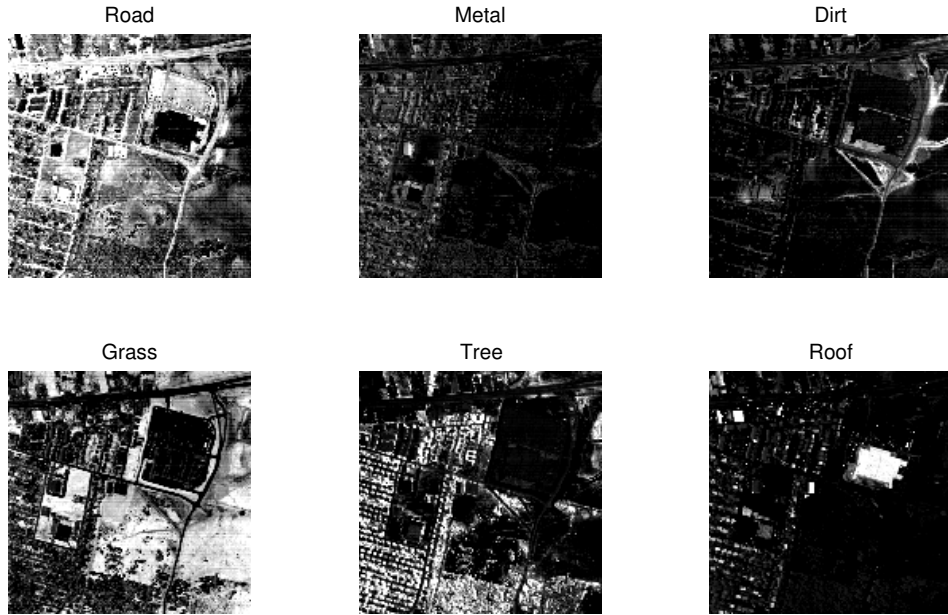


Fig. 7. Estimated abundance: least squares solution.

daylight lamps from 45 degrees on both sides in order to achieve sufficiently uniform illumination.

### B. A Test on Real Data

In this test, we use compressed hyperspectral data collected by the hardware apparatus described above with the same type of measurement matrices  $A$  as in the previous experiments. Since the light shined on the object was distributed into over 3600 spectral bands, the intensity was significantly weakened in each channel, a relatively high level of noise became inevitable in the experiments. In many aspects, this represents a realistic and revealing test on the concept of the proposed CSU scheme.

Our target image is an image of color wheel, as is shown in Figure 9, which is composed of various intensity levels of three colors: yellow, cyan, and magenta. We selected 175 uniformly distributed bands in the range of 0.4 to 0.75 micrometers, and resolution at each band was  $256 \times 256$ . For convenience, we also chose yellow, cyan, and magenta as the three endmembers, though different choices are certainly possible. In a separate experiment, we measured the spectral signatures for the three colors which are plotted in Figure 10. The parameters and initial values used in this test by Algorithm 3 are the same as those specified in Section V-A.

The abundance fractions corresponding to the three endmembers were computed from 10% measured data, and are shown in Figure 11. The elapsed time to process the compressed unmixing was about 26 seconds. As we can see, our model and algorithm detected, quite accurately, the areas corresponding to each color at various levels of brightness.

Figure 12 gives 4 slices of the computed hyperspectral cube, obtained by multiplying the estimated abundance

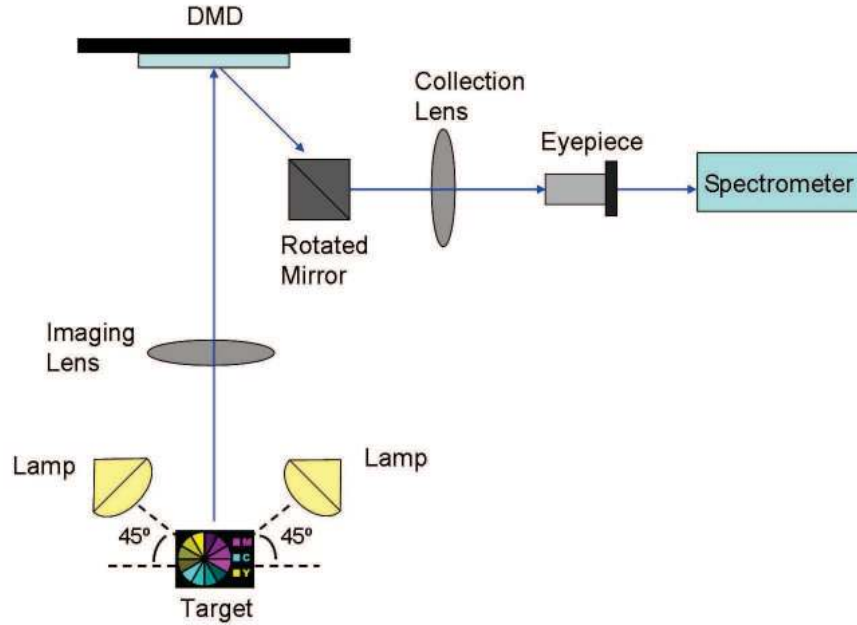


Fig. 8. Single-pixel camera schematic for hyperspectral data acquisition.

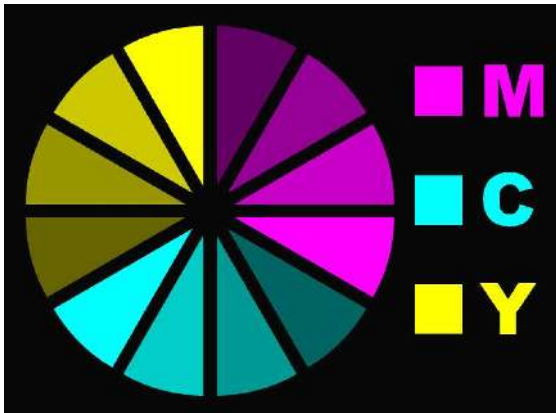


Fig. 9. Target image "Color wheel".

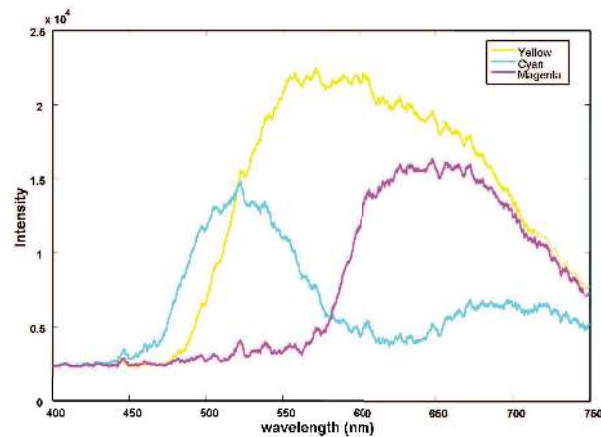


Fig. 10. Measured spectral signatures of the three endmembers.

matrix  $H$  with  $W$ , corresponding to four different spectral bands or wavelengths. For comparison, Figure 13 gives four slices, corresponding to the same four spectral bands as in Figure 12, of a computed hyperspectral cube that were computed from the same 10% of the measured dataset, one slice at a time, by the 2D TV solver TVAL3 as would be the case in the reconstruction of 2D images from compressed measurements in a standard CS setting. In this setting, neither endmember signatures nor abundance fractions was utilized. It is evident that the results in Figure 12 are much cleaner than those in Figure 13. Apparently, this remarkable superiority of the proposed CSU scheme is the consequence of two factors: 1) a thorough exploitation of both low-rankness and sparsity in 3D hyperspectral data, 2) the denoising effects of the SVD preprocessing.

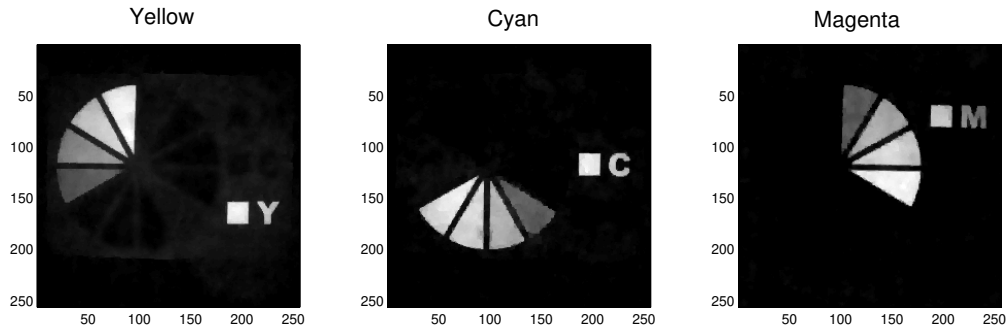


Fig. 11. Estimated abundance: CS unmixing solution from 10% measurements.

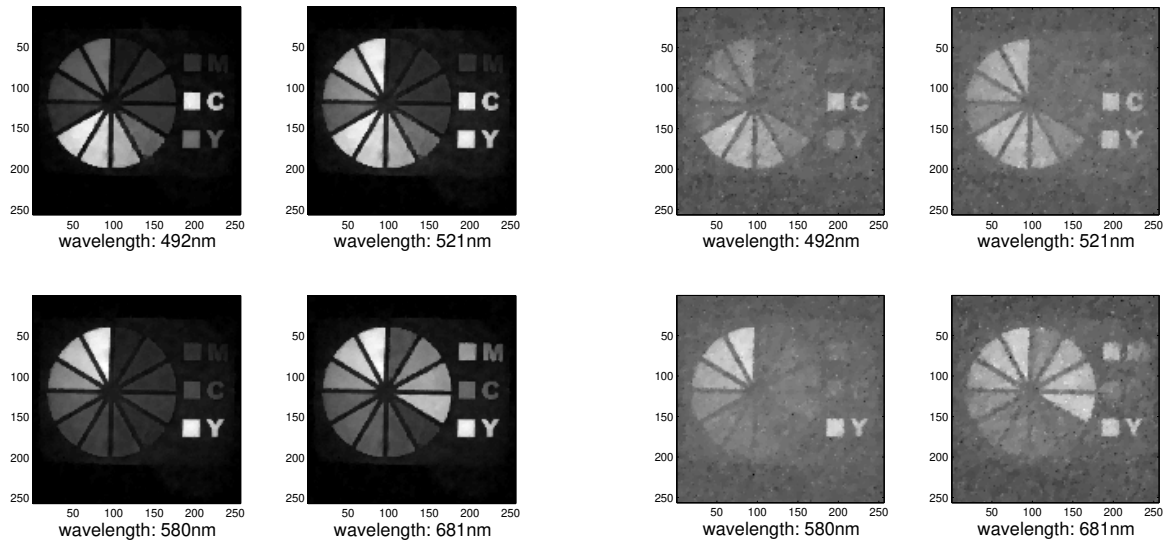


Fig. 12. Four slices computed by the proposed approach.

Fig. 13. Four slices computed slice-by-slice by TV minimization.

## VII. CONCLUSIONS

This work is a proof-of-concept study on a compressive sensing and unmixing (CSU) scheme for hyperspectral data processing that does not require forming or storing any full-size data cube. The CSU scheme consists of three major steps: 1) data acquisition by compressive sensing; 2) data preprocessing by SVD; and (3) data unmixing by solving a compressed unmixing model with total-variation regularization on abundance fraction distributions.

In this first-stage study, we only consider the situation where the spectral signatures of the endmembers are either precisely or approximately known. After performing the SVD preprocessing, data sizes to be processed become much smaller and independent of the number of spectral bands. An efficient algorithm has been constructed for solving a compressed unmixing model based on the augmented Lagrangian method and alternating minimization.

The proposed CSU scheme has been empirically, and rather convincingly, validated using both synthetic data

and measured data acquired by a hardware device similar to the single-pixel camera [5]. Our numerical results clearly demonstrate that compressively acquired data of size ranging from 10% to 25% of the full size can produce satisfactory results highly agreeable with the “ground truth”. The process speed achieved so far, which can certainly be further improved, seems to fall within a promising range.

It is certainly desirable to extend the work of this paper to more practical situations where knowledge about endmember spectral signatures are either very rough, highly incomplete, or even totally missing, leading to the much more difficult task of compressive sensing and blind unmixing. In particular, the optimization models for this task become non-convex. However, some recent successes in solving non-convex matrix factorization models, such as [29] on matrix completion, offer hopes for us to conduct further research along this direction.

## REFERENCES

- [1] E. Candès, and T. Tao, *Near optimal signal recovery from random projections: Universal encoding strategies*, IEEE Trans. on Inform. Theory, vol. 52, no. 12, pp. 5406–5425, 2006.
- [2] E. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, vol. 52, no. 2, pp. 489–509, 2006.
- [3] D. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.
- [4] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, pp. 259–268, 1992.
- [5] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, *A new compressive imaging camera architecture using optical-domain compression*, Computational Imaging IV, vol. 6065, pp. 43–52, Jan. 2006.
- [6] C. Li, *An Efficient Algorithm for Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing*, Mater Thesis, Computational and Applied Mathematics, Rice University, 2009.
- [7] Y. Wang, J. Yang, W. Yin, and Y. Zhang, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imag. Sci., vol. 1, no. 4, pp. 248–272, 2008.
- [8] T. Goldstein and S. Osher, *The split Bregman method for L1 regularized problems*, SIAM J. Imag. Sci., vol. 2, no. 2, pp. 323–343, April 2009.
- [9] Z. Guo, T. Wittman, and S. Osher, *L1 unmixing and its application to hyperspectral image enhancement*, UCLA CAM report, March 2009.
- [10] G. Vane, R. Green, T. Chrien, H. Enmark, E. Hansen, and W. Porter, *The airborne visible/infrared imaging spectrometer (AVIRIS)*, Rem. Sens. of the Environ., vol. 44, pp. 127–143, 1993.
- [11] T. Lillesand, R. Kiefer, and J. Chipman, *Remote Sensing and Image Interpretation*, John Wiley & Sons, Inc., fifth edition, 2004.
- [12] R. Clark and T. Roush, *Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications*, J. of Geophysical Research, vol. 89, pp. 6329–6340, 1984.
- [13] D. Manolakis, C. Siracusa, and G. Shaw, *Hyperspectral subpixel target detection using linear mixing model*, IEEE Trans. Geosci. Remote Sensing, vol. 39, pp. 1392–1409, 2001.
- [14] C. Chang and D. Heinz, *Subpixel spectral detection for remotely sensed images*, IEEE Trans. Geosci. Remote Sensing, vol. 38, 1144–1159, 2000.
- [15] M. B. Lopes, J. C. Wolff, J. M. Bioucas-Dias, M. A. T. Figueiredo, *Near-infrared hyperspectral unmixing based on a minimum volume criterion for fast and accurate chemometric characterization of counterfeit tablets*, Analytical Chemistry, vol. 82, pp. 1462–1469, 2010.
- [16] S. M. Chai, A. Gentile, W. E. Lugo-Beauchamp, J. L. Cruz-Rivera, and D. S. Wills, *Hyper-spectral image processing applications on the SIMD pixel processor for the digital battlefield*, IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Method and Applications, pp. 130–138, 1999.
- [17] J. Boardman, *Automating spectral unmixing of AVIRIS data using convex geometry concepts*, in JPL Pub.93-26, AVIRIS Workshop, vol. 1, pp. 11–14, 1993.
- [18] M. E. Winter, *N-FINDR: an algorithm for fast autonomous spectral endmember determination in hyperspectral data*, in Proc. of the SPIE conference on Imaging Spectrometry V, vol. 3753, pp. 266–275, 1999.

- [19] J. Nascimento and J. Bioucas-Dias, *Does independent component analysis play a role in unmixing hyperspectral data?*, IEEE Transactions on Geoscience and Remote Sensing, vol. 43, pp. 175–187, 2005.
- [20] C. Chang, C. Wu, W. Liu, and Y. Ouyang, *A new growing method for simplex-based endmember extraction algorithm*, IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 10, pp. 2804–2819, 2006.
- [21] L. Zhang, X. Tao, B. Wang and J. Zhang, *A new scheme for decomposition of mixed pixels based on nonnegative matrix factorization*, IEEE International Geoscience and Remote sensing Symposium, pp. 1759–1762, 2007.
- [22] J. Bioucas-Dias, *A variable splitting augmented Lagrangian approach to linear spectral unmixing*, In First IEEE Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing, Grenoble, France, 2009.
- [23] C. Chi T. Chan and W. Ma, *A convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing*, in IEEE International Conference in Acoustics, Speech and Signal Processing-ICASSP2009, Taiwan, 2009.
- [24] J. Li and J. Bioucas-Dias, *Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data*, in IEEE International Geoscience and Remote sensing Symposium -IGARSS2008, Boston, 2008.
- [25] M. R. Hestenes, *Multiplier and gradient methods*, Journal of Optimization Theory and Applications, vol. 4, pp. 303–320, and in Computing Methods in Optimization Problems, 2 (Eds L.A. Zadeh, L.W. Neustadt, and A.V. Balakrishnan), Academic Press, New York, 1969.
- [26] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, Optimization (Ed. R. Fletcher), Academic Press, London, New York, pp. 283–298, 1969.
- [27] J. Barzilai and J. M. Borwein, *Two-point step size gradient methods*, IMA J. Numer. Anal., vol. 8, pp. 141–148, 1988.
- [28] H. Zhang and W. W. Hager, *A nonmonotone line search technique and its application to unconstrained optimization*, SIAM J. Optim., vol. 14 , pp. 1043–1056, 2004.
- [29] Z. Wen, W. Yin, and Y. Zhang, *Solving a Low-Rank Factorization Model for Matrix Completion by a Non-linear Successive Over-Relaxation Algorithm*, CAAM Technical Report TR10-07, Department of Computational and Applied Mathematics, Rice University, March 2010.
- [30] ASTER Spectral Library, <http://speclib.jpl.nasa.gov>.
- [31] US Army Corps of Engineers, <http://www.agc.army.mil/research/products/Hypercube/>.