

# Subspace Vertex Pursuit: A Fast and Robust Near-Separable Nonnegative Matrix Factorization Method for Hyperspectral Unmixing

Qing Qu, *Student Member, IEEE*, Nasser M. Nasrabadi, *Fellow, IEEE*, and Trac D. Tran, *Fellow, IEEE*

**Abstract**—The *separability assumption* turns the nonnegative matrix factorization (NMF) problem tractable, which coincides with the pure pixel assumption and provides new insights for the hyperspectral unmixing problem. Based on this assumption, and starting from the data self-expressiveness perspective, we formulate the unmixing problem as a joint sparse recovery problem by using the data itself as a dictionary. Moreover, we present a quasi-greedy algorithm for this problem by employing a back-tracking strategy. In comparison with the previous greedy methods, the proposed method can refresh the candidate pixels by solving a small fixed-scale convex sub-problem in every iteration. Therefore, our method has two important characteristics: (*i*) enhanced robustness against noise; (*ii*) moderate computational complexity and scalability to large dataset. Finally, computer simulations on both synthetic and real hyperspectral datasets demonstrate the effectiveness of the proposed method.

**Index Terms**—Linear mixture model, spectral unmixing, end-member extraction, nonnegative matrix factorization (NMF), optimization, greedy pursuit.

## I. INTRODUCTION

HYPERSPECTRAL imaging (HSI) has received considerable attention in the past few decades [1]. With the wealth of spectral information available, HSI has been successfully applied to various domains such as agriculture, mineralogy and environment monitoring [2]. Very often, the resolution cell corresponding to a single pixel in an image contains several substances. In this situation, the scattered energy is a mixture of the endmember spectra. Thus, the spectral unmixing (SU) problem, which consists of identifying the pure materials (endmembers)

Manuscript received August 31, 2014; revised January 05, 2015 and March 23, 2015; accepted March 29, 2015. Date of publication April 02, 2015; date of current version August 12, 2015. This work was supported in part by the National Science Foundation under Grant CCF-1117545, in part by the Army Research Office under Grant 60219-MA, and in part by the Office of Naval Research under Grant N000141210765. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jose Bioucas-Dias.

Q. Qu is with the Electrical Engineering Department, Columbia University, New York, NY 10027 USA (e-mail: qq2105@columbia.edu).

N. M. Nasrabadi is with the U.S. Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: nnasraba@arl.army.mil).

T. D. Tran is with the Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: trac@jhu.edu).

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/JSTSP.2015.2419184

and estimating their associated fractions (abundances), has become a major issue in HSI and has been extensively investigated recently [1]–[4].

The most widely used model in SU is the linear mixture model (LMM) [2]. It assumes that the spectral signature of each pixel is an additive linear combination of all the pure end-members. The weights associated with the linear endmember combination, termed as abundances, should be nonnegative and sum-to-one. Let  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathbb{R}^{L \times N}$  be a hyperspectral image with  $L$  spectral bands and  $N$  pixels, where each column  $\mathbf{y}_i$  ( $1 \leq i \leq N$ ) is element-wise nonnegative and corresponds to the spectral signature of a pixel. If no noise presents, then we have the following matrix factorization

$$\mathbf{Y} = \mathbf{F}\mathbf{W}, \quad (1)$$

where  $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_r] \in \mathbb{R}^{L \times r}$  is the pure endmember matrix,  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N] \in \mathbb{R}^{r \times N}$  is the associated abundance matrix, and  $r$  is the number of pure endmembers. Both  $\mathbf{F}$  and  $\mathbf{W}$  are element-wise nonnegative. Given the data matrix  $\mathbf{Y}$ , the SU problem is to obtain a low rank matrix factorization of  $\mathbf{Y}$  that recovers both  $\mathbf{F}$  and  $\mathbf{W}$  simultaneously. This is known as the nonnegative matrix factorization (NMF) problem [5], [6], which is bilinear with respect to (w.r.t.)  $\mathbf{F}$  and  $\mathbf{W}$ , and is known to be NP-hard [7].

Recently, a sequence of elegant papers [8]–[14] have showed that the NMF problem is tractable under the separability assumption, for which all the columns of  $\mathbf{Y}$  reside in a convex hull spanned by a small number of columns in  $\mathbf{Y}$  itself, that is,  $\mathbf{F} = \mathbf{Y}_{\mathcal{I}}$ , where  $\mathbf{Y}_{\mathcal{I}}$  denotes the restriction for the columns of  $\mathbf{Y}$  to the index subset  $\mathcal{I}$ . Thus, the separable NMF reduces to finding the vertices of a convex hull. This result coincides with the study of the endmember extraction problem under the pure pixel assumption [3].

### A. Review of the Previous Work

Along with the development of the theory in the separable NMF [8], [9], many heuristic algorithms have been developed in practice [3], [10]–[20], [22]–[26], and some of them are demonstrated to work both in theory and practice [12]–[14]. These algorithms can be categorized as

a) *Random Projection*: These methods use the fact that the pixels, which have extreme projections on some random *skewers* (a normalized Gaussian vector), are located at the vertices of a polytope with probability one. The most classical algorithm in this category is the Pixel Purity Index (PPI) [15]. It

repeatedly generates a number of skewers so that all the vertices can be identified with high probability. Like the Coupon Collector's problem, it usually requires  $\Theta(r \log r)$  trials to find all the endmembers. However, the non-iterative PPI is rather sensitive to the number of endmembers and the noise due to the randomness of the skewers.

*b) Greedy Pursuit:* These algorithms are iterative and each iteration mainly consists of two steps: *(i) detection step*, which detects one extreme point of the current convex hull; *(ii) projection step*, where the whole dataset is projected to the orthogonal complement of the detected extreme points. These two steps are repeated until all the extreme points are depleted. Vertex component analysis (VCA) [16], successive projection algorithm (SPA) and its variants [13], [17]–[19], and the recently proposed extreme ray (XRAY) [20] method are the typical algorithms within this category. The differences among these algorithms are mainly in the detection step, where VCA detects a vertex by an extreme projection onto a skewer, SPA finds the one with the largest  $\ell_2$ -norm (some of its generalizations use different  $\ell_p$ -norms, see [18] and therein), and XRAY detects the one having the largest correlation with some residual vector. In the noiseless setting, VCA succeeds with probability one, while SPA and its variants are guaranteed for exact recovery [3].

*c) Simplex Volume Maximization:* This type of approach is based on the Winter's belief [21] that pure endmembers can be located by finding a collection of pixels whose simplex volume is the largest, which is, however, generally a NP-hard combinatorial problem. Alternatively, two greedy strategies, the alternating volume maximization (AVMAX) and the successive volume maximization (SVMAX) methods are introduced [22]. AVMAX employs a block coordinate ascent strategy, which maintains a set of  $r$  candidate pixels and sequentially maximize the volume by updating one column at a time while keeping the rest fixed. The procedure is repeated for several cycles until convergence. On the other hand, SVMAX employs a strategy that is similar to SPA, it successively finds endmembers that maximize the volume of the current simplex. However, these greedy methods are only guaranteed to find local optimal solutions and require data preprocessing step such as dimension reduction (i.e., PCA).

*d) Convex Optimization:* More recently, a sequence of work [10]–[12], [14] treated the problem from the data self-expressiveness perspective and formulate it as a convex programming problem. Bittorf *et al.* [12] and Gillis [14] relaxed the task as a linear programming problem, and it was shown in [14] that the linear optimization approach works better than the previous greedy methods. Elhamifar *et al.* [10] and Esser *et al.* [11] formulated the problem as a row sparse recovery problem by using the data itself as the dictionary. They relaxed the problem as a convex  $\ell_{1,p}$ -minimization ( $p = 2, \infty$ ) problem and solved it some first-order solvers such as the alternating direction method of multipliers (ADMM) [27]. In the noiseless setting, when there are no *duplicated points* (i.e., samples too similar to the pure endmember pixels), their methods are guaranteed to find the global optimal solution. Moreover, our empirical results suggest it is more robust to noise than those greedy methods. However, the major concern of this approach is that

it has  $\mathcal{O}(N^2)$  ( $N$  is the size of the dataset) variables to optimize, which is computationally prohibitive as the volume of data increases.

### B. Contributions of the Paper

In this paper, we investigate more practical and robust approaches to solve the separable NMF problem for SU. The contributions are mainly twofold.

First, starting from the data self-expressiveness perspective, our *minor* contribution is that we have introduced a modified simultaneous orthogonal matching pursuit (SOMP) method [29], [30] for the joint sparse recovery problem. Compared with the original SOMP method, our modified version, named SOMP+, solves a nonnegative least squares (NNLS) problem in the projection step instead of using the least squares (LS). The benefit of such a modification is that we do not require the endmember matrix  $\mathbf{F}$  to be full column rank [31], which is, however, a crucial assumption for most of the previous greedy methods to succeed [3], [13]. Moreover, it can better handle the case when the data are highly correlated. We provide more intuitive explanations for this behavior in a heuristic manner, though it has been rigorously analyzed in [31]. Because the data degeneracy, high spectral and spatial correlations are ubiquitous in hyperspectral images, such a property is very favorable for conquering this major challenge for hyperspectral imagery data analysis.

Second, based on the proposed SOMP+ algorithm, our *major* contribution is that we further propose a novel quasi-greedy method, named subspace vertex pursuit (SVP). It employs a back-tracking strategy similar to the subspace pursuit (SP) [32], [33]. We are the first to introduce the idea of backtracking to the hyperspectral unmixing problem, where previous greedy methods are mostly iterative and cannot refine the previously found wrong endmembers. In every iteration, the SVP method first combines  $r$  newly detected candidate pixels in the current convex hull with the  $r$  old candidates from the previous step, forming a new  $2r$  candidate set; second, SVP refreshes the candidate set by solving a constrained  $\ell_{1,2}$ -minimization sub-problem, picking only  $r$  pixels out. Unlike previous greedy methods, SVP is able to freely remove unreliable candidate pixels considered as reliable in the previous steps.

The merits of the SVP method are twofold:

- 1) In comparison with the previous greedy methods, SVP is demonstrated to be much more robust to noise, due to our clever endmember refreshment step.
- 2) Compared with the  $\ell_{1,2}$ -minimization method [10], [11], our SVP method shows comparable superior performance, but with a reduced computational complexity by a factor of  $\mathcal{O}(N)$ .

In summary, our SVP algorithm balances well for both the noise robustness performance and computational scalability, and can better handle the high correlation issue in the hyperspectral data, which makes it very practical for solving the SU problems under the pure pixel assumption.

### C. Organization of the Paper and Notations

The rest of the paper is organized as follows. In Section II, we present the basic assumptions, formulations, and key observations for the SU problem. Section III introduces the sparse

optimization and greedy pursuit techniques for solving the SU problem. Section IV presents the proposed SVP method. In Section V, we discuss some practical issues with the proposed SVP method. In Section VI, extensive experiments on both synthetic and real datasets demonstrate the effectiveness and efficiency of our proposed methods. Finally, the conclusion is drawn in Section VII.

Prior to the ensuing presentation, let us define the following notations for the ease of later use.

$\mathbb{R}_+$	the sets of nonnegative real numbers
$\mathbf{e}_i$	$i$ -th standard basis vector
$\mathbf{1}_N$	the vector with all components unitary
$\mathbf{I}_N$	$N \times N$ identity matrix
$\leq, \geq$	component-wise inequalities
$\mathbf{X}_i, \mathbf{X}^j$	the $i$ -th column and the $j$ -th row
$x_{ij}$	element at the $i$ -th row and $j$ -th column
$\mathbf{X}^{(k)}$	the variable $\mathbf{X}$ in the $k$ -th iteration
$\ \cdot\ _p(p = 1, 2, \infty)$	$\ell_p$ -norm ( $p = 1, 2, \infty$ )
$\ \cdot\ _{1,p}, \ \cdot\ _F$	matrix $(1, p)$ norm and Frobenius norm

## II. PROBLEM STATEMENT

### A. Problem Formulation and Basic Assumptions

1) *Linear Mixture Model Assumption:* Most of the methods for the SU problem are based on the simple but representative LMM, where each pixel is assumed to be a nonnegative linear combination of the pure endmembers. Given the matrix  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathbb{R}_+^{L \times N}$  constructed from the hyperspectral data cube, the endmember matrix  $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_r] \in \mathbb{R}_+^{L \times r}$  and the number of endmembers  $r$ , each pixel  $\mathbf{y}_i$  ( $1 \leq i \leq N$ ) is assumed to satisfy

$$\mathbf{y}_i = \sum_{j=1}^r \mathbf{f}_j w_{ji} + \mathbf{n}_i, \quad (2)$$

where  $\mathbf{n}_i \in \mathbb{R}^L$  is a white Gaussian noise vector with every entry independent and identically distributed (i.i.d.) and  $w_{ji}$  is the abundance value for the endmember  $j$  of the  $i$ th pixel. To be physically meaningful, the abundance  $w_{ji}$  has to satisfy the abundance sum-to-one constraint (ASC) and the abundance nonnegative constraint (ANC) as follows

$$\begin{aligned} \text{ASC : } & \sum_{j=1}^r w_{ji} = 1 \quad (1 \leq i \leq N), \\ \text{ANC : } & w_{ji} \geq 0 \quad (\top \mathbf{1} \leq j \leq r, 1 \leq i \leq N). \end{aligned} \quad (3)$$

Based on (2), notice that LMM could be written in matrix form as

$$\mathbf{Y} = \mathbf{F}\mathbf{W} + \mathbf{N} \quad (4)$$

with  $\mathbf{N} = [\mathbf{n}_1, \dots, \mathbf{n}_N]$ ,  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N]$  and  $\mathbf{w}_i = [w_{1i}, w_{2i}, \dots, w_{ri}]^\top$  ( $1 \leq i \leq r$ ).

2) *The Constrained NMF Problem for SU:* Based on the LMM assumption, the aim of the SU problem is to find a non-negative rank  $r$  factorization of the matrix  $\mathbf{Y}$  so that we can recover the matrices  $\hat{\mathbf{F}} \in \mathbb{R}_+^{L \times r}$  and  $\hat{\mathbf{W}} \in \mathbb{R}_+^{r \times N}$  simultaneously, where  $\hat{\mathbf{F}}$  and  $\hat{\mathbf{W}}$  are the minimizers of the following problem

$$\min_{\mathbf{F} \geq 0, \mathbf{W} \geq 0} \|\mathbf{Y} - \mathbf{F}\mathbf{W}\|_F^2, \text{s.t. } \mathbf{W}^\top \mathbf{1}_r = \mathbf{1}_N. \quad (5)$$

This is a NMF problem with the sum-to-one constraint, which is highly ill-posed and NP-hard [7]. In general, the problem is nonconvex and could have many spurious local minima. Moreover, due to the symmetry, it has many equally good solutions  $[\hat{\mathbf{F}}, \hat{\mathbf{W}}] = [\mathbf{FTD}, \mathbf{D}^{-1}\Gamma^\top \mathbf{W}]$ , up to some column permutation  $\Gamma \in \mathbb{R}^{r \times r}$  and scaling  $\mathbf{D} \in \mathbb{R}^{r \times r}$  (i.e.,  $\mathbf{D}$  is a positive diagonal matrix).

3) *The Separability Assumption for the NMF Problem:* Recently, Arora *et al.* [9] have provided a sufficient condition which guarantees the NMF problem to be well-posed in a sense that the solution is unique up to some column permutations, and can be solved within polynomial time. Their condition could be stated as follows:

*Definition 1 (Simplicial Vectors):* A set of vectors  $\{\mathbf{v}_1, \dots, \mathbf{v}_r\} \in \mathbb{R}^L$  are called simplicial if no vector  $\mathbf{v}_i$  lies within the convex hull of the others  $\{\mathbf{v}_j : j \neq i\}$ .

*Definition 2 (Separable NMF):* We call the factorization  $\mathbf{Y} = \mathbf{F}\mathbf{W}$  a separable NMF if the columns of  $\mathbf{F}$  are simplicial and there exists a column permutation matrix  $\Pi$ , such that

$$\mathbf{W}\Pi = [\mathbf{I}_r \quad \mathbf{U}], \quad (6)$$

where  $\mathbf{I}_r$  is a rank  $r$  identity matrix,  $\Pi$  is a column permutation matrix and  $\mathbf{U} \in \mathbb{R}_+^{r \times (N-r)}$  satisfies  $\mathbf{U}^\top \mathbf{1}_r = \mathbf{1}_{N-r}$ . Furthermore, the factorization is called near-separable, if  $\mathbf{Y} = \mathbf{F}\mathbf{W} + \mathbf{N}$  where  $\mathbf{N}$  is an additive noise matrix.

### B. Key Observation for the Separable NMF

Quite recently, Elhamifar, Esser, and Bittorf *et al.* [10]–[12] looked into the problem from data self-expressiveness perspective and formulated it as a row sparse recovery problem. Their key observations are based on the following theorem.

*Theorem 1 (Row Sparsity Property [12]):* If  $\mathbf{Y} = \mathbf{F}\mathbf{W}$  admits a separable NMF, then

$$\mathbf{Y} = \mathbf{Y}\mathbf{X}, \text{ with } \mathbf{X} = \Pi \begin{bmatrix} \mathbf{I}_r & \mathbf{U} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \Pi^\top, \quad (7)$$

where  $\Pi$  is a column permutation matrix and  $\mathbf{X} \in \mathbb{R}^{N \times N}$ . Furthermore, if  $\hat{\mathbf{Y}} = \mathbf{Y} + \mathbf{N}$  is near-separable with  $\mathbf{N}$  to be an i.i.d. white Gaussian matrix, we have

$$\hat{\mathbf{Y}} = \hat{\mathbf{Y}}\mathbf{X} + \hat{\mathbf{N}}, \quad (8)$$

where  $\hat{\mathbf{N}} = \mathbf{N}(\mathbf{I}_N - \mathbf{X})$  is still zero-mean Gaussian.

Since the column permutation matrix  $\Pi$  does not change the number of nonzero rows in  $\mathbf{X}$ , if  $r \ll N$ , Theorem 1 implies that the matrix  $\mathbf{X}$  is row sparse. Therefore, given the data matrix  $\mathbf{Y}$  under the separability assumption, we expect to exactly recover the row sparse matrix  $\mathbf{X}$  by solving

$$\min_{\mathbf{X}} \|\mathbf{X}\|_{row-0}, \text{s.t. } \mathbf{Y} = \mathbf{Y}\mathbf{X}, \mathbf{X} \geq 0, \mathbf{X}^\top \mathbf{1}_N = \mathbf{1}_N, \quad (9)$$

where  $\|\cdot\|_{row-0}$  denotes the number of non-zero rows. Then the endmembers can be identified based on the structure of  $\mathbf{X}$  according to Theorem 1.

### III. A UNIFIED VIEW OF APPROACHES FOR THE SEPARABLE NMF PROBLEM

#### A. Convex Optimization Viewpoint

Instead of solving the combinatorial problem (9) directly, Elhamifar *et al.* [10] and Esser *et al.* [11] relax the problem as a convex optimization problem of the form

$$\min_{\mathbf{X} \geq 0} \|\mathbf{X}\|_{1,p}, \text{ s.t. } \mathbf{Y} = \mathbf{Y}\mathbf{X}, \mathbf{X}^\top \mathbf{1}_N = \mathbf{1}_N, \quad (10)$$

where  $\|\mathbf{X}\|_{1,p} = \sum_{i=1}^N \|\mathbf{x}^i\|_p$  ( $p = 2, \infty$ ) is a convex surrogate of  $\|\mathbf{X}\|_{row-0}$ . For the concern of noise robustness, they further suggest to solve the following penalized problem

$$\min_{\mathbf{X} \geq 0} \frac{1}{2} \|\mathbf{Y} - \mathbf{Y}\mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_{1,p}, \text{ s.t. } \mathbf{X}^\top \mathbf{1}_N = \mathbf{1}_N, \quad (11)$$

where  $\lambda > 0$  is a penalty parameter. This problem can be solved by a first-order solver such as the ADMM algorithm [27], [28] (see Appendix). Once the matrix  $\mathbf{X}$  is recovered, the pure pixels can be detected based on the index of the non-zero rows in  $\mathbf{X}$ . Empirically, this method shows superior robustness against noise, especially for  $p = 2$ . In the rest of the paper, we only refer to the  $\ell_{1,2}$ -minimization variant of their methods (i.e.,  $p = 2$ ). However, the drawbacks of the  $\ell_{1,2}$ -minimization method are twofold:

a) *Duplicate Points*: The method requires the assumption that only one pure pixel exists for each endmember. If multiple duplicate points for one pure endmember exist, then the solution of the convex relaxed problem (10) is not unique and the convex relaxation may fail to recover the solution of the original problem (9). However, this problem could be alleviated by preprocessing such as clustering that merges similar pixels.

b) *Computational Complexity*: The more critical problem is that their method requires to solve an optimization problem with  $\mathcal{O}(N^2)$  variables, which is computationally prohibitive, especially for large-scale problems.

#### B. Greedy Viewpoint

Most of the state-of-the-art SU methods employ greedy strategies. Since operations involved in each iteration are computationally cheap, these methods are well-suited for large-scale problems. Therefore, alternatively, it is natural to consider the variants of simultaneous orthogonal matching pursuit (SOMP) method [29], [30] for solving the joint sparse recovery problem (10). Before introducing the method in detail, we provide the following useful definitions

*Definition 3 (Projection and Residual):* Let  $\mathbf{Y} \in \mathbb{R}^{L \times N}$  and  $\Phi \in \mathbb{R}^{L \times m}$ , then the projection operator  $\text{Proj}(\mathbf{Y}, \Phi)$ , which projects  $\mathbf{Y}$  onto the convex cone spanned by the columns of  $\Phi \in \mathbb{R}^{L \times m}$ , and the associated residual  $\text{Res}(\mathbf{Y}, \Phi)$  are defined to be

$$\begin{aligned} \mathbf{H} &= \text{Proj}(\mathbf{Y}, \Phi) = \arg \min_{\mathbf{B} \geq 0} \|\mathbf{Y} - \Phi \mathbf{B}\|_F^2, \\ \mathbf{R} &= \text{Res}(\mathbf{Y}, \Phi, \mathbf{H}) = \mathbf{Y} - \Phi \mathbf{H}. \end{aligned} \quad (12)$$

---

#### Algorithm 1 SOMP Algorithm for Spectral Unmixing

---

**Input:** The data  $\mathbf{Y}$ , the number of pure endmembers  $r$ ;  
**Output:** The estimated endmember and abundance  $\hat{\mathbf{F}}, \hat{\mathbf{W}}$ ;

```

1: Initialize:  $\mathcal{I}^{(0)} = \emptyset, \mathbf{R}^{(0)} = \mathbf{Y}$ 
2: for  $k = 1 : r$  do
3:   Calculate the correlation matrix
       $\mathbf{C}^{(k)} = (\mathbf{R}^{(k-1)})^\top \mathbf{Y};$ 
4:   Detection Step: Find  $i = \arg \max_j \|(\mathbf{C}^{(k)})^j\|_2$ ;
5:   Update the index set:  $\mathcal{I}^{(k)} = \mathcal{I}^{(k-1)} \cup \{i\}$ ;
6:   Projection Step: Calculate  $\mathbf{H}^{(k+1)} = \text{Proj}(\mathbf{Y}, \mathbf{Y}_{\mathcal{I}^{(k)}});$ 
7:   Update the residual  $\mathbf{R}^{(k)} = \text{Res}(\mathbf{Y}, \mathbf{Y}_{\mathcal{I}^{(k)}}, \mathbf{H}^{(k+1)});$ 
8: end for
9: The endmember dictionary  $\mathbf{F} = \mathbf{Y}_{\mathcal{I}^{(r)}}$ ;
10: Recover the abundance matrix  $\hat{\mathbf{W}}$ 

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W} \geq 0} \|\mathbf{Y} - \mathbf{F}\mathbf{W}\|_F^2, \text{ s.t. } \mathbf{1}_r^\top \mathbf{W} = \mathbf{1}_N^\top. \quad (13)$$

11: return  $\hat{\mathbf{F}} = \mathbf{F}, \hat{\mathbf{W}}$ 
```

---

Based on these definitions, the proposed approach is introduced in Algorithm 1. In every iteration, the algorithm proceeds by alternating between the following two steps:

- 1) *Detection Step*: The algorithm finds a candidate pixel  $i$  having the largest correlation with all columns of the current residual matrix  $\mathbf{R}^{(k)}$  (i.e.,  $i = \arg \max_j \|(\mathbf{C}^{(k)})^j\|_p$ , where  $\mathbf{C}^{(k)} = (\mathbf{R}^{(k-1)})^\top \mathbf{Y}$ , and adds it to the index set  $\mathcal{I}$  as a pure endmember. Notice that the XRAY (max) method [20] is a special case of our method when  $p = \infty$ .
- 2) *Projection Step*: The projection step is different from the original SOMP, where we solve a NNLS sub-problem instead of LS. The reason for this modification will be discussed later in this section.

After the pure endmembers are extracted, the abundance matrix  $\mathbf{W}$  can be recovered by solving a fully constraint least squares (FCLS) problem [34] given in (13).

Next, we provide the intuitions for solving the NNLS subproblem in the projection step instead of the classical LS subproblem. Before that, we introduce a useful definition for better explanation.

*Definition 4 (Endmember Coherence):* For a given endmember matrix  $\mathbf{F} \in \mathbb{R}^{L \times r}$  with every column as a pure endmember, the endmember coherence  $\mu(\mathbf{F})$  can be defined as

$$\mu(\mathbf{F}) = \max_{i \neq j} \frac{|\mathbf{f}_i^\top \mathbf{f}_j|}{\|\mathbf{f}_i\|_2 \|\mathbf{f}_j\|_2}. \quad (14)$$

Based on this definition, when the coherence is large, as the data samples are highly correlated, it is challenging to extract all the endmembers, and vice versa. Therefore, the endmember coherence provides a notion of the “difficulty level” of the SU problem. Next, we use this definition to explain the differences between solving the LS and NNLS sub-problems in three cases:

- **Case 1:** As illustrated in Fig. 1(a), the coherence  $\mu(\mathbf{F})$  is sufficiently small so that the given endmembers  $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$  are linearly independent and the spectral angles between each endmember are large. In this case, we have  $\mathbf{R}_{NNLS} = \mathbf{R}_{LS}$ , so that it makes no difference between solving LS and NNLS sub-problems in this scenario.

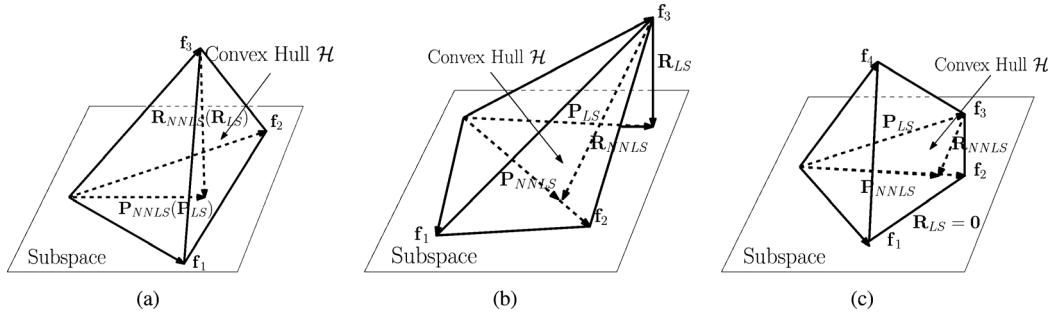


Fig. 1. Differences between NNLS and LS projections for three different cases.  $\mathbf{P}_{NNLS}$  and  $\mathbf{P}_{LS}$  denote the projection vectors for the NNLS and LS, and  $\mathbf{R}_{NNLS}$  and  $\mathbf{R}_{LS}$  denote the residual vectors for the NNLS and LS, respectively. (a) Case 1, (b) Case 2, (c) Case 3.

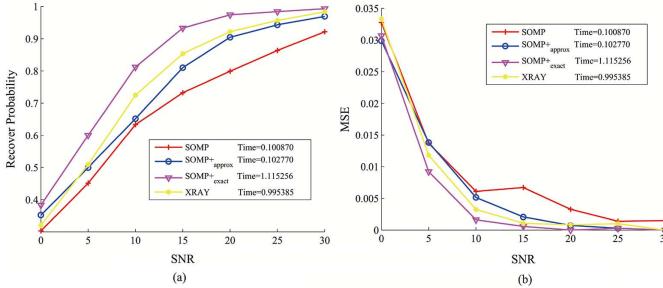


Fig. 2. Comparison of the recovery probability for the SOMP with different projection methods and XRAY. (a) Recovery Probability, (b) Mean Square Error.

- **Case 2:** As illustrated in Fig. 1(b), the coherence  $\mu(\mathbf{F})$  is large so that the angle between  $\mathbf{f}_1$  and  $\mathbf{f}_2$  is small. In this case, NNLS projects  $\mathbf{f}_3$  to the convex cone spanned by  $\{\mathbf{f}_1, \mathbf{f}_2\}$  while LS projects  $\mathbf{f}_3$  to the corresponding subspace, the two projections of  $\mathbf{f}_3$  are geometrically different. Especially, by the triangle inequality, it is obvious that  $\|\mathbf{R}_{NNLS}\|_2 > \|\mathbf{R}_{LS}\|_2$ . Intuitively, given the same noise level, the projection  $\mathbf{R}_{NNLS}$  with a larger amplitude is less likely to be immersed by the noise.
- **Case 3:** As illustrated in Fig. 1(c), the coherence  $\mu(\mathbf{F})$  is sufficiently large so that  $\mathbf{f}_3$  is linearly dependent on  $\{\mathbf{f}_1, \mathbf{f}_2\}$ . Suppose  $\mathbf{f}_3$  is the target endmember, and is projected to the previously detected  $\{\mathbf{f}_1, \mathbf{f}_2\}$ . For the LS, since  $\mathbf{f}_3$  resides in the subspace of  $\text{span}(\mathbf{f}_1, \mathbf{f}_2)$ , the residual for the subspace projection is trivial with  $\mathbf{R}_{LS} = \mathbf{0}$ ; however, for the NNLS, the cone projection  $\mathbf{R}_{NNLS}$  is generally nontrivial. Therefore, in this degenerate case, the modified SOMP succeeds even when the original SOMP fails.

The major challenge for most of the hyperspectral imagery problems is the high spectral and spatial correlation [1]. Therefore, for the SU problem, it is more interesting and meaningful to deal with the more challenging scenarios such as Case 2 and 3. This further suggests that it is preferable to solve the NNLS rather than the LS sub-problem in the projection step for greedy methods. Empirically, such a modification leads to significant performance gains for hyperspectral endmember extraction (See Fig. 2). In the literature, there are efficient ways to solve the NNLS problem [35]. One cheap approximation is projecting the LS solution  $\mathbf{H} = \arg \min_{\mathbf{B}} \|\mathbf{Y} - \mathbf{Y}_{\mathcal{I}} \mathbf{B}\|_F^2$  back onto the nonnegative orthant by

$$\mathbf{H} = \max\{(\mathbf{Y}_{\mathcal{I}}^\top \mathbf{Y}_{\mathcal{I}})^{-1} \mathbf{Y}_{\mathcal{I}}^\top \mathbf{Y}, \mathbf{0}\}, \quad (15)$$

and we let SOMP+*approx* denote the modified SOMP by approximating the solution of the NNLS sub-problem using (15).

TABLE I  
DETECTION AND PROJECTION STRATEGIES FOR DIFFERENT METHODS

Algorithm	SOMP	SOMP+ <i>approx</i>	SOMP+ <i>exact</i>	XRAY
Norm	$\ell_2$	$\ell_2$	$\ell_2$	$\ell_\infty$
Projection	LS	Inexact NNLS	NNLS	NNLS

In [20], Kumar *et al.* solve the NNLS sub-problem with higher precision by a block coordinate descent (BCD) method, and we let SOMP+*exact* to be the modified SOMP by solving the NNLS using the BCD method. In Fig. 2, we compare four algorithms (i.e., SOMP, SOMP+*approx*, SOMP+*exact* and XRAY) showing the effects of different norm selections and projection methods on a synthetic hyperspectral data, the detailed explanation of the experimental setting is postponed to Section VI.B and the choices for the norms and projection methods for the four algorithms are listed in Table I. From Fig. 2, we conclude that

- By comparing XRAY and SOMP+*exact*, we can see that the choice of  $p = 2$  is more robust to noise than  $p = \infty$ . This is because the selection of  $p = \infty$  is based on only one candidate pixel that produces the maximum inner product with a residual, while for  $p = 2$  we choose an endmember based on the inner product of the residual vector with all the data.
- Comparing the performances of SOMP, SOMP+*approx* and SOMP+*exact*, we observe that by solving the NNLS sub-problem we obtain better performance compared to solving the LS sub-problem. Moreover, the more accurate we solve the NNLS sub-problem, the better the performance.

#### IV. SUBSPACE VERTEX PURSUIT

From the previous section, our observation shows that greedy algorithms sequentially detect one pure candidate pixel at a time. However, due to the system noise and perturbations, the selected candidate pixels, found to be the optimal choice previously, may not be optimal in the current iteration, and they can never be corrected. It further indicates that, in the noisy setting, they could be stuck in some local optimal solutions. To alleviate this problem, we further introduce the idea of back-tracking, originated from the SP algorithm [32], [33] for sparse recovery, to the vertex finding problem, and we name the newly proposed algorithm *subspace vertex pursuit* (SVP).

#### A. Algorithm Description

In each iteration, the SVP method maintains a set of  $r$  columns of  $\mathbf{Y}$  and performs a simple test for the data matrix  $\mathbf{Y}$ . If the columns of  $\mathbf{Y}$  do not lie in the current estimate for the correct spanning convex cone, SVP adds  $r$  more candidates to the current candidate set, and then refines the estimate by only

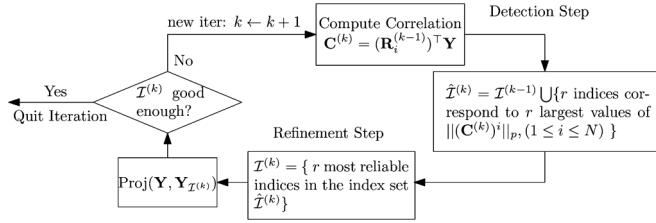


Fig. 3. One Iteration of the proposed SVP method: a list of  $r$  candidates, which is allowed to be updated during each iteration, is maintained.

retaining the  $r$  most reliable ones. The expectation is that the recursive refinements of the estimate will lead to a convex cone with a decreasing distance from the data  $\mathbf{Y}$ . To summarize, each iteration of SVP consists of 3 steps: (i) *detection step* (ii) *refinement step* (iii) *projection step*. A schematic diagram of the SVP method is depicted in Fig. 3, and the overall algorithm is summarized in Algorithm 2. It should be noticed that Step 6 ensures that the residual is always non-increasing so that the algorithm converges. In the following paragraphs, we provide detailed explanations for each step.

1) *Detection Step*: The aim of the detection step is to detect  $r$  candidate vertices simultaneously. This could be achieved by random projections, like PPI, where  $r$  different pixels are picked by extreme projections using a large amount of random skewers. However, as discussed in Section I, this method is not robust to noise due to the randomness. Alternatively, we consider the correlation matrix  $\mathbf{C}^{(k)} = (\mathbf{R}^{(k-1)})^\top \mathbf{Y}$ , and pick  $r$  indices corresponding to the  $r$  largest values of  $\|(\mathbf{C}^{(k)})^i\|_2 (1 \leq i \leq N)$ . This procedure is guaranteed to pick at least one vertex in each step and is robust to noise.

2) *Refinement Step*: In the refinement step, given the merged candidate set  $\hat{\mathcal{I}}^{(k)}$  of  $2r$  points, our aim is to find  $r$  most reliable candidates out of the  $2r$  point set  $\hat{\mathcal{I}}^{(k)}$ , as the vertex candidates, which is critical for the proposed method to work. The intuition here is that the pure endmembers are the smallest number of points to form a convex representation for all the data points. Therefore, we consider solving the following  $\ell_{1,2}$ -regularized sub-problem

$$\begin{aligned} & \min_{\mathbf{X}_1} \frac{1}{2} \|\mathbf{Y} - \mathbf{Y}_{\hat{\mathcal{I}}^{(k)}} \mathbf{X}_1\|_F^2 + \lambda \|\mathbf{X}_1\|_{1,2} \\ & \text{s.t. } \mathbf{X}_1 \geq \mathbf{0}, \mathbf{X}_1^\top \mathbf{1}_{2r} - \mathbf{1}_N = \mathbf{0}, \end{aligned} \quad (16)$$

to promote the row sparsity of  $\mathbf{X}_1 \in \mathbb{R}^{2r \times N}$ . This problem can be efficiently solved by the ADMM algorithm [27] (see Appendix B for detailed explanations). We pick  $r$  indices in  $\hat{\mathcal{I}}^{(k)}$  corresponding to the  $r$  largest  $\ell_2$ -norm of the rows in  $\mathbf{X}_1$  to form the new index set  $\mathcal{I}^{(k)}$ . To further reduce the scale of the sub-problem, one may even consider the following sub-problem

$$\begin{aligned} & \min_{\mathbf{X}_2} \frac{1}{2} \|\mathbf{Y}_{\hat{\mathcal{I}}^{(k)}} - \mathbf{Y}_{\hat{\mathcal{I}}^{(k)}} \mathbf{X}_2\|_F^2 + \lambda \|\mathbf{X}_2\|_{1,2}, \\ & \text{s.t. } \mathbf{X}_2 \geq \mathbf{0}, \mathbf{X}_2^\top \mathbf{1}_{2r} - \mathbf{1}_{2r} = \mathbf{0}, \end{aligned} \quad (17)$$

where  $\mathbf{X}_2 \in \mathbb{R}^{2r \times 2r}$ . In comparison with the  $\ell_{1,2}$ -minimization problem (11) with  $\mathcal{O}(N^2)$  variables to optimize, the number of variables for the sub-problem (16) is around the order of  $\mathcal{O}(rN)$ , and is further reduced to  $\mathcal{O}(r^2)$  for the sub-problem (17). Given that  $r \ll N$ , the scales of our sub-problems (16)

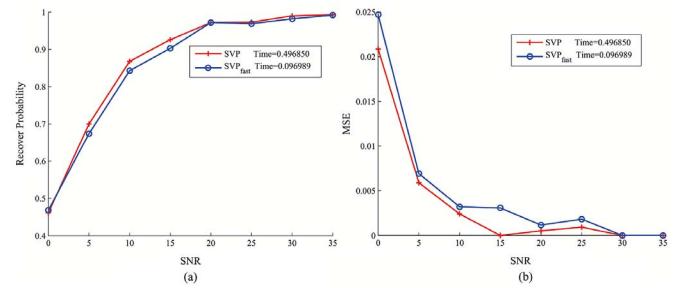


Fig. 4. A comparison of two different refinement strategies for the proposed SVP method. SVP solves the sub-problem defined in (16), and  $\text{SVP}_{\text{fast}}$  solves the sub-problem defined in (17). (a) Recovery Probability, (b) Mean Square Error.

and (17) are much reduced. In particular, the number of variables in (17) is independent of the sample number  $N$ , which is more favorable. In Fig. 4, we show a comparison of the two proposed refinement strategies for the SVP algorithm on a synthetic hyperspectral dataset (a detailed explanation for the experimental setting is postponed to Section VI.B), where SVP solves a sub-problem defined in (16) and  $\text{SVP}_{\text{fast}}$  solves a sub-problem defined in (17). From the results, we can see that

- The two methods show similar performances against noise, although SVP behaves slightly better than  $\text{SVP}_{\text{fast}}$ .
- The computation time of  $\text{SVP}_{\text{fast}}$  is much smaller compared to SVP, given the dataset  $N = 500$  and the number of endmember  $r = 20$ . If  $N$  becomes larger, the difference of computation time between  $\text{SVP}_{\text{fast}}$  and SVP will be significantly amplified.

### B. Comparison With SP

Though the proposed method is inspired from the SP method [32] for sparse recovery, our algorithm is different from theirs especially in the refinement step, where SP solves a LS sub-problem as follows

$$\mathbf{X} = \arg \min_{\mathbf{B}} \|\mathbf{Y} - \mathbf{Y}_{\hat{\mathcal{I}}^{(k)}} \mathbf{B}\|_F^2. \quad (19)$$

Based on the solution  $\mathbf{X}$  in (19), the indices in  $\hat{\mathcal{I}}^{(k)}$  which produce  $r$  largest values of  $\|\mathbf{X}^i\|_2 (i \in \hat{\mathcal{I}}^{(k)})$  are selected as reliable candidates.

However, for the separable NMF problem, such a strategy cannot guarantee that the picked candidates are the true vertices. To provide a counter-example, consider a special case in  $\mathbb{R}_+^3$  as shown in Fig. 5, where a convex hull  $\mathcal{H}$  is generated by  $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ , and  $\{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3\}$  are on the edges of  $\mathcal{H}$ . For the illustration purpose, we consider a special index set  $\hat{\mathcal{I}}^{(k)}$  containing the indices of the points  $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3\}$ . For the points within the convex hull of  $\{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3\}$  (dotted as circles in Fig. 5), they can either be represented by points  $\{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3\}$  or  $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$  or a mixed combinations of  $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3\}$ , and the situations are the same for the “disk”, “cross” and “box” points. We can neither guarantee the uniqueness of the LS solution  $\mathbf{X}$  for (19), nor that the picked pixels are the true vertices. Therefore, the candidate vectors picked by this procedure are unreliable.

Instead of solving the LS sub-problem, we propose to solve the  $\ell_{1,2}$ -regularized sub-problem (16) alternatively. The reason is quite intuitive, consider the counter-example in the previous

---

**Algorithm 2 Subspace Vertex Pursuit for Spectral Unmixing**


---

**Input:** The data  $\mathbf{Y}$ , the number of pure endmembers  $r$ ;  
**Output:** The estimated endmember and abundance  $\mathbf{F}, \mathbf{W}$ ;

- 1: Initialize:  $k = 1, \mathbf{C}^{(0)} = \mathbf{Y}^\top \mathbf{Y}$ , the residual  $\mathbf{R}^{(0)} = \text{Res}(\mathbf{Y}, \mathbf{Y}_{\mathcal{I}^{(0)}})$ , where  

$$\mathcal{I}^{(0)} = \left\{ r \text{ indices associated with } r \text{ largest } \left\| (\mathbf{C}^{(0)})^i \right\|_2 \right\},$$
- 2: **while** not converged **do**
- 3:   **Detection Step:** Compute  $\mathbf{C}^{(k)} = (\mathbf{R}^{(k-1)})^\top \mathbf{Y}$  and update  

$$\hat{\mathcal{I}}^{(k)} = \mathcal{I}^{(k-1)} \cup \{r \text{ indices associated with } r \text{ largest values of } \left\| (\mathbf{C}^{(k)})^i \right\|_2\},$$
- 4:   **Refinement Step:** Solve (16) or (17) to get  $\hat{\mathbf{X}}^{(k)}$   

$$\mathcal{I}^{(k)} = \left\{ r \text{ indices in } \hat{\mathcal{I}}^{(k)} \text{ with } r \text{ largest } \|(\hat{\mathbf{X}}^{(k)})^i\|_2 \right\},$$
- 5:   **Projection Step:** Calculate  $\mathbf{H}^{(k+1)} = \text{Proj}(\hat{\mathbf{Y}}, \mathbf{Y}_{\mathcal{I}^{(k)}})$ ;
- 6:   Update the residual  $\mathbf{R}^{(k)} = \text{Res}(\mathbf{Y}, \mathbf{Y}_{\mathcal{I}^{(k)}}, \mathbf{H}^{(k+1)})$ ;
- 7:   **If**  $\|\mathbf{R}^{(k)}\|_F > \|\mathbf{R}^{(k-1)}\|_F$ , **stop**;
- 8:    $k = k + 1$ ,
- 9: **end while**
- 10: Let  $\mathbf{F} = \mathbf{Y}_{\mathcal{I}^{(k)}}$  and calculate the abundance matrix  $\mathbf{W}$

$$\mathbf{W} = \arg \min_{\mathbf{W} \geq 0} \|\mathbf{Y} - \mathbf{FW}\|_F^2, \text{s.t. } \mathbf{1}_r^\top \mathbf{W} = \mathbf{1}_N^\top, \quad (18)$$

11: **return**  $\mathbf{F}, \mathbf{W}$ .

---

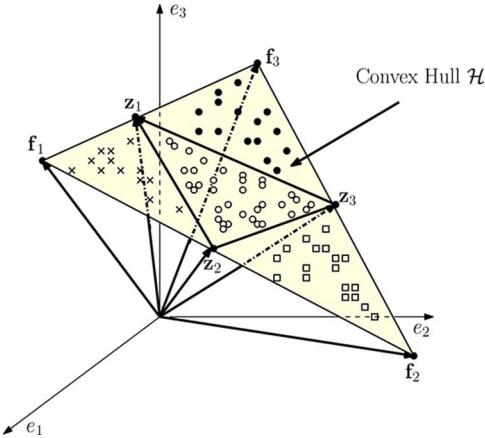


Fig. 5. A convex hull  $\mathcal{H}$  generated by  $\{f_1, f_2, f_3\}$  in  $\mathbb{R}_+^3$ , and  $\{z_1, z_2, z_3\}$  are on the edges of  $\mathcal{H}$ . The “circle” points are inside the convex hull of  $\{z_1, z_2, z_3\}$ , the “disk” points are inside the convex hull of  $\{z_1, f_3, z_3\}$ , the “cross” points are inside the convex hull of  $\{z_1, z_2, f_1\}$ , the “box” points are inside the convex hull of  $\{z_1, z_3, f_2\}$ .

paragraph: for all the points in the convex hull of  $\{f_1, f_2, f_3\}$ , the smallest number of points to represent all the data are  $\{f_1, f_2, f_3\}$ , where the solution is unique and what we desired. Hence, it is more favorable to solve the  $\ell_{1,2}$  sub-problem (16) or (17) instead.

## V. PRACTICAL ISSUES FOR SVP

In this section, we discuss some practical issues of the proposed SVP method. Some of these discussions may also apply to other existing unmixing methods as summarized in [1], [3], [10], [11].

### A. Handling Duplicate Points

The proposed SVP method solves an  $\ell_{1,2}$ -minimization subproblem, so it cannot well handle the case when there are multiple duplicate points at vertices. However, such a problem can be prevented by preprocessing the data. One simple heuristic is to merge data points with very small spectral angles by some preset threshold to prune the set of representatives, preventing from having too-close data points.

### B. Choice for the Parameter $\lambda$

Similar to the classical  $\ell_1$ -minimization problem, the parameter  $\lambda$  balances the weights of the  $\ell_1$ -penalty and data fidelity terms for the sub-problems (16) and (17) in the proposed SVP method. When the noise is small, we are then more confident in our model, it is preferred to set  $\lambda$  small in order to put more weights on the data fidelity term, and vice versa. Empirically, we find that the proposed method is robust to the choice of  $\lambda$ . In general settings, we usually set  $\lambda \sim 10^{-2}$ . However, when the noise level is high, we increase  $\lambda$  to be around the order of  $10^{-1}$  or even 1.

On the other hand, note that the parameter  $\lambda$  is also sensitive to the scaling of the data matrix. For instance, if we scale each entry of the data matrix  $\mathbf{Y}$  by a constant  $C$ , the parameter  $\lambda$  should also be scaled by  $C$  correspondingly. To address this issue, we normalize the data  $\mathbf{Y}$  by  $\hat{\mathbf{Y}} = \mathbf{Y} / \max\{|Y_{i,j}|\}$  to get rid of the data scaling dependence for the parameter  $\lambda$ .

### C. Handling Illumination Changes

In practice, due to illumination and spectral variability [1], instead of exactly lying in a convex hull  $\mathcal{H}$ , the data  $\mathbf{Y}$  resides in a convex cone  $\mathcal{C}$  defined by

$$\mathcal{C} = \{\gamma \mathbf{y} \in \mathbb{R}^L | \mathbf{y} = \mathbf{Fs}, \mathbf{s} \geq \mathbf{0}, \mathbf{1}_r^\top \mathbf{s} = 1, \gamma > 0\}. \quad (20)$$

However, we can normalize the data to lie on a hyperplane  $\mathcal{L} = \{\mathbf{y} \in \mathbb{R}^L | \mathbf{b}^\top \mathbf{y} = 1\}$  by a projective projection

$$\hat{\mathbf{y}}_i = \frac{\mathbf{y}_i}{\mathbf{b}^\top \mathbf{y}_i}, \quad \forall i, 1 \leq i \leq N. \quad (21)$$

As illustrated in Fig. 6, the projected data lies in the new convex hull  $\hat{\mathcal{H}} = \{\hat{\mathbf{y}} \in \mathbb{R}^L | \hat{\mathbf{y}} = \frac{\mathbf{y}}{\mathbf{b}^\top \mathbf{y}}, \mathbf{y} \in \mathcal{C}\}$ , which is the intersection of the hyperplane  $\mathcal{L}$  and the convex cone  $\mathcal{C}$ .

### D. Comparison of Computational Complexity

In this subsection, the computational complexities of the proposed SVP method and some existing benchmark methods are briefly discussed, and the results are summarized in Table II. We keep the notations  $L, r, N$  to be data dimension, number of end-members and number of data points, respectively.

- From [13], we know that the computational complexities for SPA, VCA and PPI are  $\mathcal{O}(rLN)$ ,  $\mathcal{O}(rLN)$  and  $\mathcal{O}(rLN \log r)$ , respectively.

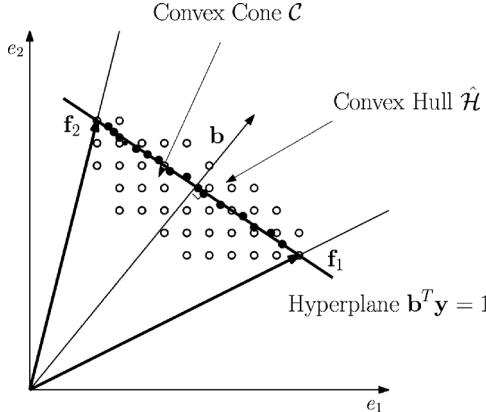


Fig. 6. An illustration of the projective projection to normalize the data in the convex cone  $\mathcal{C}$  (the circle points) to the convex hull  $\hat{\mathcal{H}}$  (the disk points).

TABLE II  
COMPARISON OF THE COMPUTATIONAL COMPLEXITY

Algorithm	Computational Complexity
PPI [15]	$\mathcal{O}(rLN \log r)$
VCA [16]	$\mathcal{O}(rLN)$
SPA [17]	$\mathcal{O}(rLN)$
AVMAX [22]	$\mathcal{O}(rLN)$
SVMAX [22]	$\mathcal{O}(rLN)$
XRAY [20]	$\mathcal{O}(rLN + C_B r^2 N)$
SOMP	$\mathcal{O}(rLN^2)$
SVP	$\mathcal{O}(LN^2 + C_\epsilon r^2 L)$
$SVP_{fast}$	$\mathcal{O}(LN^2 + C_\epsilon r^3)$
$\ell_{1,2}$ -minimization [10]	$\mathcal{O}(C_\epsilon N^3)$

- As discussed in [22], the computational complexities for SVMAX and AVMAX are both around  $\mathcal{O}(rLN)$ .
- The computational complexity for XRAY is mainly due to the NNLS problem solved by BCD method. Its total computational complexity is about  $\mathcal{O}(rLN + C_B r^2 N)$ , where  $C_B$  is the iteration number for the BCD method.
- For SOMP, the computation is around  $\mathcal{O}(rLN^2)$ .
- For SVP, based on Algorithm 2, the computational complexities for computing the correlation in the detection step, solving the  $\ell_{1,2}$  minimization problem in the refinement step, and calculating the residual in the projection step are roughly  $\mathcal{O}(LN^2)$ ,  $C_\epsilon r^2 L$  and  $\mathcal{O}(rLN)$ , respectively, where  $C_\epsilon$  is the number of iterations that depends on the error tolerance  $\epsilon$ . Therefore, the total computational complexity is around  $\mathcal{O}(LN^2 + C_\epsilon r^2 L)$ . So when  $N$  is large, it scales like  $\mathcal{O}(LN^2)$ , which implies that the detection step bottlenecks the overall computational complexity.
- For  $SVP_{fast}$ , similarly to SVP, the complexity is on the order of  $\mathcal{O}(LN^2 + C_\epsilon r^3)$ .
- For the  $\ell_{1,2}$ -minimization method, the computation for solving the ADMM algorithm is on the order of  $\mathcal{O}(C_\epsilon N^2(L + N))$ , where  $C_\epsilon$  is the number of iterations that depends on the error tolerance  $\epsilon$ .

From Table II, we observe

- The computational complexities of the proposed SOMP and  $SVP_{fast}$  methods are approximately  $\mathcal{O}(N)$  higher than the greedy methods. The increase is mainly due to the computation of the self-correlation in the detection step.

Nonetheless, as discussed in Section IV.A, such a procedure brings the merit of significant noise robustness.

- In comparison with the  $\ell_{1,2}$ -minimization method, the computational complexity of our algorithms is reduced by a factor of  $\mathcal{O}(N)$ .

In summary, in comparison with previous methods, our  $SVP_{fast}$  combines two important characteristics: (i) empirically enhanced robustness against noise; (ii) moderate computational complexity for most practical purposes.

## VI. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed method SVP on both synthetic and real datasets. First, we provide some explanations of experimental settings and standard comparison criteria. Second, we evaluate the proposed method on a synthetic dataset. Finally, the proposed method is demonstrated on real hyperspectral datasets.

### A. Experimental Settings and Comparison Criteria

1) *Simulation Environment*: All the experiments are executed on a desktop computer with a 3.4 GHz Intel quad-core i7 processor and 8 GB 1600 MHz RAM. The simulations are performed by using MATLAB 2013a, where all the codes<sup>1</sup>,<sup>2</sup> are reimplemented in efficient manners for the purpose of fair comparisons. We compare the proposed methods with VCA [16], SPA [13], AVMAX [22], SVMAX [22], XRAY (max) [20], and the  $\ell_{1,2}$ -minimization method [10]<sup>3</sup>. For all the greedy methods SOMP+, VCA, SPA, AVMAX, SVMAX and XRAY (max), there are no parameters to tune. For the  $\ell_{1,2}$ -minimization method, the penalty parameter  $\lambda$  is set to be  $1 \times 10^{-1}$  and the max iteration is set to be 500. For the proposed SVP, the penalty parameter  $\lambda$  for the sub-problems is set to be  $\lambda = 1 \times 10^{-2}$  and the max iteration is set to be 20.

#### 2) Comparison Criteria:

a) *Recovery Probability*: We define the probability of successful recovery to be

$$\mathcal{P} = \frac{|\mathcal{I}_r|}{r} \quad (22)$$

where  $r$  is the number of endmembers,  $\mathcal{I}_r$  is the index set of the recovered endmembers, and  $|\mathcal{I}_r|$  is the cardinality of the set. For  $T$  repeated simulations, the recovery probability is simply  $\mathcal{P}_s = \frac{1}{T} \sum_{i=1}^T \mathcal{P}_i$ , where  $\mathcal{P}_i$  is the recovery probability for the  $i$ th simulation.

b) *Mean Square Error (MSE)*: Let  $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_r\}$  be the ground truth endmembers, and let  $\{\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \dots, \hat{\mathbf{f}}_r\}$  be the predicted endmembers. We define the MSE as

$$\mathcal{E} = \min_{\boldsymbol{\pi} \in \Pi} \frac{1}{r} \sum_{i=1}^r \|\mathbf{f}_i - \hat{\mathbf{f}}_{\pi_i}\|_2^2, \quad (23)$$

where  $\pi_i \in \{1, 2, \dots, r\}$  ( $1 \leq i \leq r$ ),  $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_r]$ , and  $\Pi = \{\boldsymbol{\pi} \in \mathbb{R}^r | \pi_l \in \{1, 2, \dots, r\}, \pi_l \neq \pi_m, l \neq m\}$  is the set of all the permutations of  $[1, 2, \dots, r]$ . The problem

<sup>1</sup><http://bit.ly/1g03XKE>

<sup>2</sup>[http://mx.nthu.edu.tw/~tsunghan/download/AVMAX\\_SVMAX\\_WAVMAX\\_codes.zip](http://mx.nthu.edu.tw/~tsunghan/download/AVMAX_SVMAX_WAVMAX_codes.zip)

<sup>3</sup>We do not compare the proposed method with PPI [15] and Hottopixx [12], because the performances of both algorithms are not comparable to the formerly listed methods as shown in [13], [14].

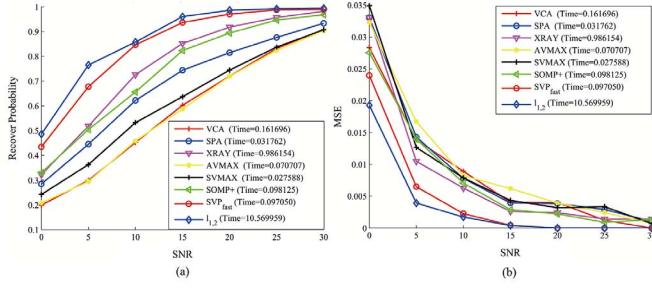


Fig. 7. Comparison of the computational time and the noise robustness for the state-of-the-art methods (i.e., VCA, SPA, AVMAX, SVMAX, XRAY and the  $\ell_{1,2}$ -minimization method) with the proposed methods on a synthetic dataset. (a) Recovery Probability, (b) Mean Square Error.

in (23) with  $N!$  permutations of  $\pi$  can be efficiently solved by the Hungarian algorithm [36]. For  $T$  repeated simulations, the average MSE =  $\frac{1}{T} \sum_{i=1}^T \mathcal{E}_i$ , where  $\mathcal{E}_i$  is the MSE for the  $i$ th simulation.

### B. Experiment Results on Synthetic Data

1) *Experiment I (Robustness to Noise):* We compare the computational time and noise robustness for the proposed SVP method with state-of-the-art algorithms on a selected USGS library<sup>4</sup> [37]. For each simulation,  $r = 20$  endmembers are randomly selected from the USGS dataset to form the feature matrix  $\mathbf{F} \in \mathbb{R}^{L \times r}$ , where the feature dimension  $L = 224$ . The weighting matrix  $\mathbf{W}$  is generated by  $[\mathbf{I}_r, \mathbf{U}] \mathbf{\Pi} \in \mathbb{R}^{20 \times 500}$ , where  $\mathbf{I}_r$  is an identity matrix so that there exists one pure endmember for each selected material. Every column of  $\mathbf{U} \in \mathbb{R}^{20 \times 480}$  is generated by the Dirichlet distribution with random parameters drawn from the uniform distribution  $\mathcal{U}_{[0,1]}$ ,  $\mathbf{\Pi}$  is a random permutation matrix so that the order of mixing weights does not affect the performance. The data matrix  $\mathbf{Y}$  is generated by  $\mathbf{Y} = \mathbf{FW} + \mathbf{N}$ , where each entry of  $\mathbf{N}$  is i.i.d. white Gaussian with  $N_{ij} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$ . We vary the signal-to-noise ratio (SNR) from 0 dB to 30 dB, where SNR is defined to be  $\text{SNR} = 10 \log_{10}(\frac{1}{NL} \|\mathbf{FW}\|_F^2 / \|\mathbf{N}\|_F^2)$ . For different levels of the Gaussian noise, the simulations are repeated for  $T = 100$  times. We perform three experiments based on the synthetic data as follows.

- 1) We compare the proposed SOMP+<sub>approx</sub>, SOMP+<sub>exact</sub> methods with the original SOMP and XRAY (max) to show the performance differences of the detection and projection steps. The corresponding results and the associated discussions are shown in Fig. 2 and Section III.B, respectively.
- 2) We compare the proposed SVP and SVP<sub>fast</sub> methods to show the computation and noise robustness performance differences between solving the sub-problem (16) and (17). The performance is shown in Fig. 4, and the discussions on the experimental results are provided in Section IV.A.
- 3) The proposed SVP<sub>fast</sub> is compared with state-of-the-art methods on a synthetic dataset, where the results are shown in Fig. 7. We can see that SVP<sub>fast</sub> shows comparable superior performance as the  $\ell_{1,2}$ -minimization method [10]. Moreover, even on a medium scale dataset  $N = 500$ , the proposed SVP<sub>fast</sub> is around  $10^2$  times faster than the  $\ell_{1,2}$ -minimization method.

<sup>4</sup>[http://www.lx.it.pt/bioucas/code/sunsal\\_demo.zip](http://www.lx.it.pt/bioucas/code/sunsal_demo.zip)

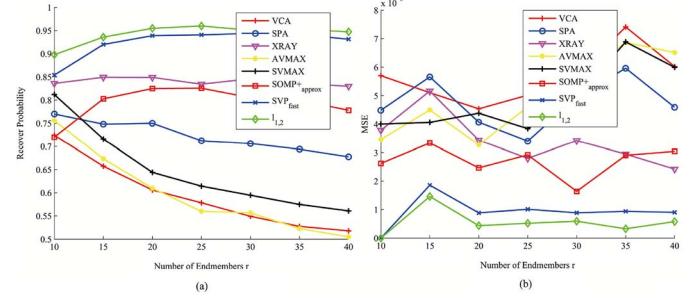


Fig. 8. Comparison of the robustness to the number of endmembers for the state-of-the-art methods (i.e., VCA, SPA, AVMAX, SVMAX, XRAY and the  $\ell_{1,2}$ -minimization method) with the proposed methods on a synthetic dataset. (a) Recovery Probability, (b) Mean Square Error.

2) *Experiment II (Robustness to the Number of Endmembers):* In this experiment, we compare the robustness of all the methods w.r.t. the number of endmembers. The experimental setting is the same as that of Experiment I, except that we fix SNR = 15 dB,  $\lambda = 1 \times 10^{-2}$  and vary the number of end-members  $r$  from 10 to 40. The results are shown in Fig. 8, from which we can see that the proposed SVP method along with the  $\ell_{1,2}$ -minimization method shows superior robustness to the change of  $r$ , while the performances of all the other greedy methods significantly decrease when  $r$  increases. In the meanwhile, we also observe that the choice of  $\lambda$  is almost independent to the number of endmembers.

### C. Experiment Results on Real Data

In this subsection, we present some numerical results for hyperspectral endmember extraction on two real hyperspectral datasets. We compare the proposed method with VCA, SPA, AVMAX, SVMAX, XRAY, and the  $\ell_{1,2}$ -minimization method.

1) *Urban Dataset:* The Urban hyperspectral data is publicly available<sup>5</sup>. Each pixel in the Urban dataset is composed of 210 spectral channels with spectral resolution 10 nm and spectrum ranging from 400 nm to 2500 nm. After removing the water absorption bands (Band 1–4, 76, 87, 101–111, 136–153, and 198–210),  $L = 162$  clean bands remain. Thus, the Urban data cube has dimension  $307 \times 307 \times 162$  with the number of data points  $N = 94249$ , which is a rather simple and well understood dataset: it is mainly composed of six types of materials (asphalt, roof, metal, grass, tree and dirt) as reported in [38]. Fig. 9(a) shows the false-color image of the Urban hyperspectral data; Fig. 9(b) shows the endmember signatures of six different materials labelled in Fig. 9(a), where the signatures are obtained by averaging ten manually selected pure endmembers in the corresponding marked area in Fig. 9(a).

First, we remove the duplicate pixels by preprocessing the data. We merge similar data points whose spectral angles  $\theta_{ij} = \arccos\left(\frac{\mathbf{y}_i^\top \mathbf{y}_j}{\|\mathbf{y}_i\|_2 \|\mathbf{y}_j\|_2}\right)$  are smaller than a certain threshold, i.e.,  $\cos(\theta_{ij}) \geq \tau = 0.995$ , reducing the size of the dataset from  $\mathbf{Y}_0 \in \mathbb{R}^{162 \times 94249}$  to  $\mathbf{Y} \in \mathbb{R}^{162 \times 1147}$ . Based on the preprocessed data, we compare our methods with state-of-the-art algorithms methods mentioned above, the results are shown in Fig. 10. From Fig. 10(h), we can see the proposed SVP method

<sup>5</sup>Available at <http://www.agc.army.mil/>

TABLE III  
COMPARISON OF THE RMSEs ON THE URBAN IMAGE

Algorithm	SPA	VCA	XRAY	AVMAX	SVMAX	SOMP+ <i>approx</i>	SVP <sub>fast</sub>	$\ell_{1,2}$
RMSE ( $\times 10^{-2}$ )	2.6506	2.9614	1.3824	1.5292	1.6715	1.2996	0.8351	0.6469

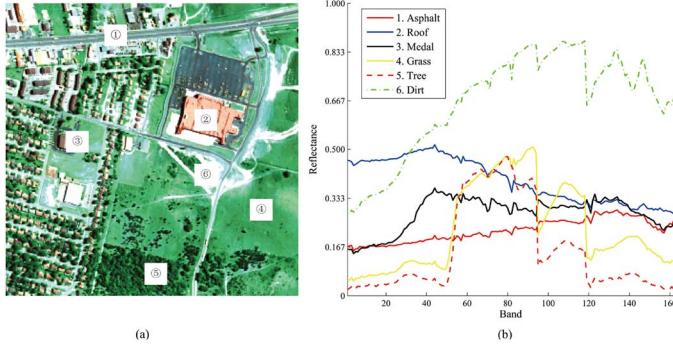


Fig. 9. (a) shows the false-color image of the Urban data of Band 50, 100 and 150; (b) displays the pure endmember signatures which are obtained by averaging 10 manually selected pure endmembers in the corresponding marked area in (a).

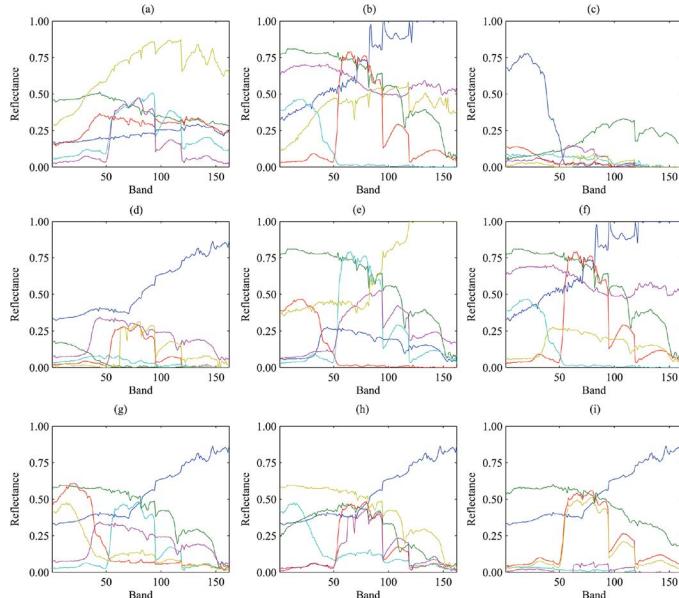


Fig. 10. For the Urban dataset, (a) shows the signatures are obtained by averaging ten manually selected pure endmembers in the corresponding marked area in Fig. 9(a); (b)-(i) are the pure endmembers extracted by SPA [13], VCA [16], XRAY [20], AVMAX [22], SVMAX [22], SOMP+*approx*, SVP<sub>fast</sub> and the  $\ell_{1,2}$ -minimization [10] method, respectively. (a) Manually Labelled, (b) SPA, Time = 0.019771, (c) VCA, Time = 0.144663, (d) XRAY, Time = 0.095890, (e) AVMAX, Time = 0.073597, (f) SVMAX, Time = 0.015299, (g) SOMP+*approx*, Time = 0.101365, (h) SVP<sub>fast</sub>, Time = 0.254073, (i)  $\ell_{1,2}$ , Time = 73.095418.

can extract distinct endmembers from the data and they are visually more similar to the manually labelled endmembers than the other methods (see Fig. 10(a)).

Because there is no ground truth available for the pure endmembers in the Urban image, next we introduce a criterion for comparing these algorithms on the real data. Once the pure endmember dictionary  $\hat{\mathbf{F}}$  is recovered, we compute the corresponding abundance maps by solving the FCLS problem as follows

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W} \geq 0} \|\mathbf{Y}_0 - \hat{\mathbf{F}}\mathbf{W}\|_F^2, \text{ s.t. } \mathbf{W}^\top \mathbf{1}_r = \mathbf{1}_N, \quad (24)$$

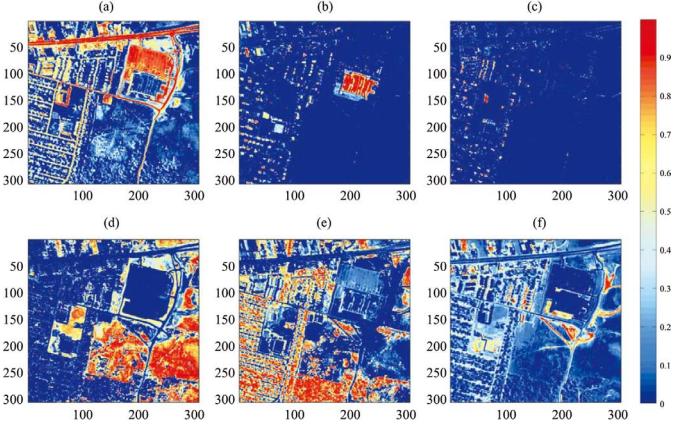


Fig. 11. The abundance map for six materials: (a) asphalt, (b) roof, (c) metal, (d) grass, (e) tree and (f) dirt, obtained by the FCLS based on the manually labelled pure endmembers in Fig. 10(a).

and then calculate the root mean square error (RMSE)

$$\text{RMSE} = \frac{1}{\sqrt{NL}} \|\mathbf{Y}_0 - \hat{\mathbf{F}}\hat{\mathbf{W}}\|_F. \quad (25)$$

Note that geometrically FCLS provides the projection of the data onto the convex hull spanned by the endmembers. If  $\hat{\mathbf{W}}$  denotes the FCLS solution of each method, then the RMSE provides some insight on how much data is contained in the convex hull and therefore, how well the endmembers describe the data. The results of the RMSE for all the algorithms are shown in Table III. From the results, we can see that our SVP method and the  $\ell_{1,2}$ -minimization method both produce significantly lower RMSEs than the other methods. While the RMSE of the  $\ell_{1,2}$ -minimization method is 22% lower than SVP<sub>fast</sub>, but our method is nearly 300 times faster than the  $\ell_{1,2}$ -minimization method on this particular dataset, as shown in Fig. 10(h) and (i).

Finally, the abundance maps for the manually labelled endmembers, and the endmembers extracted by the SOMP+*approx*, SVP<sub>fast</sub> and  $\ell_{1,2}$ -minimization methods are shown in Figs. 11, 12, 13 and 14, respectively. If we assume that the abundance maps in Fig. 11 are the “ground truth”, we conclude the following:

- For the SOMP+*approx* method, the prediction for the metal, grass and trees are not accurate, observed from Fig. 12(b), (d) and (e).
- For the SVP<sub>fast</sub> method, the estimation for the metal material is not so accurate, as observed from Fig. 13(c);
- For the  $\ell_{1,2}$ -minimization method, the materials extracted for the metal and trees are not correct, observed from Fig. 14(c) and (e), respectively.

For the methods with three smallest RMSEs, i.e., SOMP+, SVP and the  $\ell_{1,2}$ -minimization methods, we can see that: (i) none of the methods detect the metal material properly, this phenomenon needs further investigations, one reason might be that

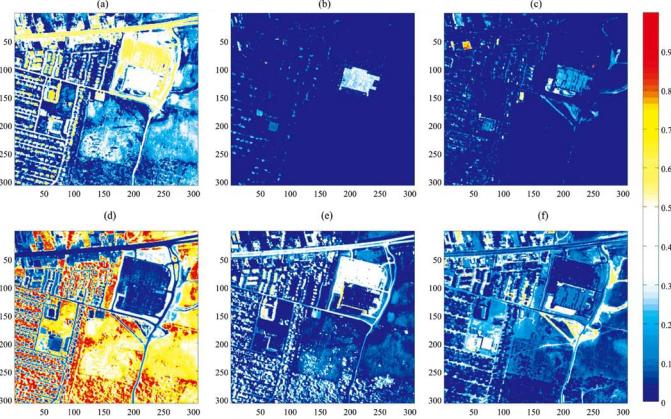


Fig. 12. The abundance map for six materials: (a) asphalt, (b) roof, (c) metal, (d) grass, (e) tree and (f) dirt, obtained by the FCLS based on the pure endmembers extracted by  $\text{SOMP} + \text{approx}$  in Fig. 10(g).

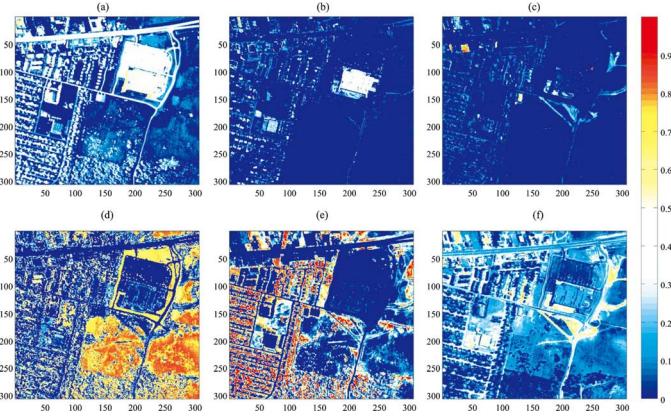


Fig. 13. The abundance map for six materials: (a) asphalt, (b) roof, (c) metal, (d) grass, (e) tree and (f) dirt, obtained by the FCLS based on the pure endmembers extracted by  $\text{SVP}_{\text{fast}}$  in Fig. 10(h).

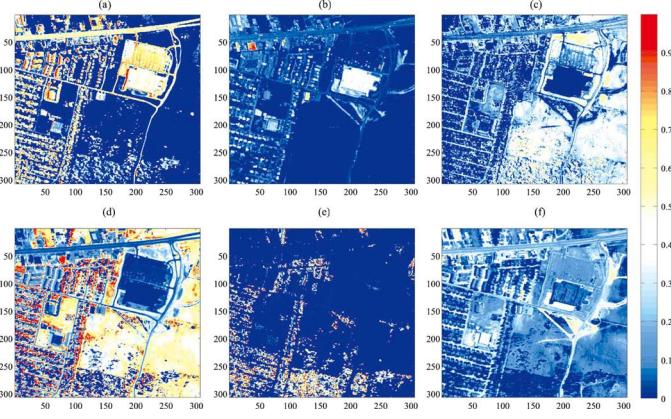


Fig. 14. The abundance map for six materials: (a) asphalt, (b) roof, (c) metal, (d) grass, (e) tree and (f) dirt, obtained by the FCLS based on the pure endmembers extracted by the  $\ell_{1,2}$ -minimization method in Fig. 10(i).

there is no pure metal pixel since metal is a kind of a rare material, another is that the pixels of metal are too noisy to be properly detected; (ii) our proposed  $\text{SVP}_{\text{fast}}$  method produces the abundance maps visually closer to the “ground truth” compared to the other methods.

2) *Cuprite Dataset:* We also test the proposed algorithm on a low-resolution AVIRIS hyperspectral data obtained over

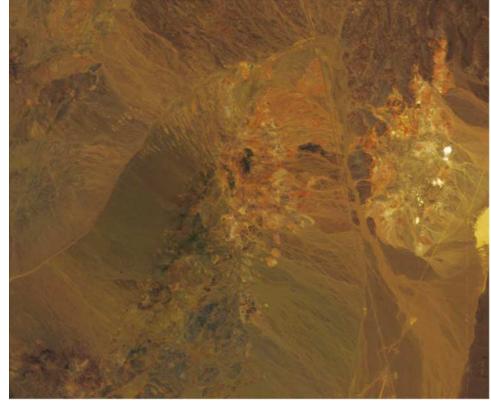


Fig. 15. False-color image of the AVIRIS Cuprite data set in the file “f970619t01p02\_r02\_sc04.a.rfl”. The bands used as RGB channels are bands (12, 22, 42) of the original 224-band image.

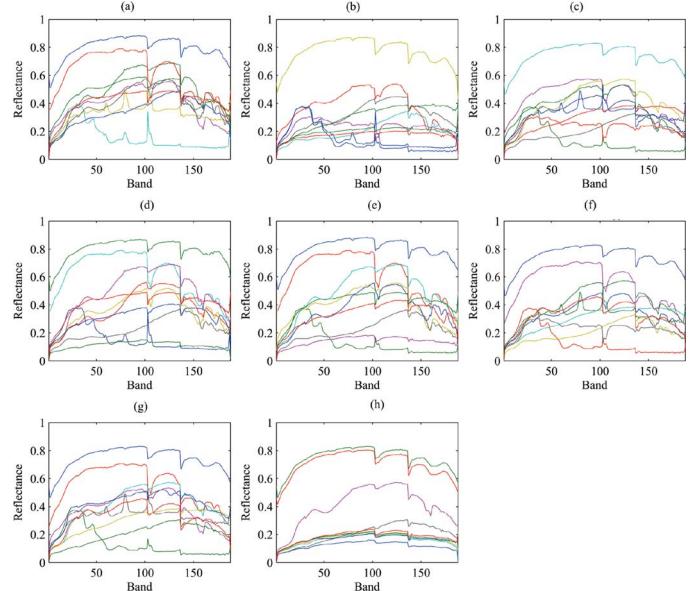


Fig. 16. For the Cuprite image, (a)-(h) are the pure endmembers extracted by SPA [13], VCA [16], XRAY [20], AVMAX [22], SVMAX [22],  $\text{SOMP} + \text{approx}$ ,  $\text{SVP}_{\text{fast}}$  and  $\ell_{1,2}$ -minimization [10] methods, respectively.

the Cuprite mining region in Nevada, USA (refer to [40] for a detailed description), which is available online in reflectance units<sup>6</sup>. The portion used in this experiment is the image of the sector labelled as “f970619t01p02\_r02\_sc04.a.rfl” of size  $512 \times 614$  in the original dataset (see Fig. 15). The scene comprises 224 spectral bands between  $0.4$  and  $2.5 \mu\text{m}$ , with a nominal spectral resolution of 10 nm. Prior to the analysis, bands 1–2, 105–115, 150–170 and 223–224 are removed due to water absorption and low SNR in those bands, leaving a total of 188 spectral bands available.

We first preprocess the data by merging similar data points whose spectral angles  $\theta_{ij} = \arccos(\frac{\mathbf{y}_i^T \mathbf{y}_j}{\|\mathbf{y}_i\|_2 \|\mathbf{y}_j\|_2})$  are smaller than a threshold, i.e.,  $\cos \theta_{ij} \leq \tau = 0.9995$ , reducing the size of the dataset from  $\mathbf{Y}_0 \in \mathbb{R}^{188 \times 314368}$  to  $\mathbf{Y} \in \mathbb{R}^{162 \times 1432}$ . Based on the preprocessed data, we compare our methods with state-of-the-art algorithms mentioned above, the extracted pure endmembers for each algorithm are shown in Fig. 16, and the

<sup>6</sup><http://aviris.jpl.nasa.gov/html/aviris.freedata.html>

TABLE IV  
COMPARISON OF THE RMSES ON THE CUPRITE IMAGE

Algorithm	SPA	VCA	XRAY	AVMAX	SVMAX	SOMP+ <i>approx</i>	SVP <sub>fast</sub>	$\ell_{1,2}$
RMSE ( $\times 10^{-3}$ )	9.6114	8.0194	6.9614	8.1117	8.6967	6.8814	5.8643	10.1949

corresponding RMSEs are shown in Table IV. For the particular Cuprite image, from Table IV, we can see that the proposed SVP<sub>fast</sub> method shows the smallest RMSE, but however the  $\ell_{1,2}$ -minimization method shows the worst performance in this case. This might because the pixels in the Cuprite image are too similar, and the  $\ell_{1,2}$ -minimization method gets trapped with those spectrally similar points near to the true endmembers, see Fig. 16(h).

## VII. CONCLUSION

In this paper, we investigated the spectral unmixing problem based on the recent development of separable NMF. In particular, starting from the data self-expressiveness perspective, we formulate the problem as a joint sparse recovery problem and provide a unified overview of the convex optimization and greedy approaches for solving the problem. More importantly, based on these methods, we introduced a fast and robust algorithm, the subspace vertex pursuit method, to solve the unmixing problem. The proposed method shows not only superior noise robustness performance but also significant reduction of the computational complexity. Finally, simulation results on both synthetic and real datasets demonstrate the effectiveness of the proposed method.

## APPENDIX

In this Appendix, we introduce a first-order ADMM solver [27] for the following  $\ell_{1,2}$ -regularized convex optimization problem

$$\begin{aligned} \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{AX}\|_F^2 + \lambda \|\mathbf{X}\|_{1,2}, \\ \text{s.t. } \mathbf{X} \geq \mathbf{0}, \mathbf{X}^\top \mathbf{1}_r - \mathbf{1}_N = \mathbf{0}, \end{aligned} \quad (26)$$

where  $\mathbf{Y} \in \mathbb{R}^{L \times N}$ ,  $\mathbf{A} \in \mathbb{R}^{L \times r}$  and  $\mathbf{X} \in \mathbb{R}^{r \times N}$ . The basic idea of the ADMM method is to introduce appropriate auxiliary variables into the optimization program, breaking the program into sub-problems with closed form solutions. The algorithm tries to find a saddle point of the augmented Lagrangian function, which iteratively minimizes the function w.r.t. the primal variables and maximizes it w.r.t. the dual.

To start, we introduce an auxiliary variable  $\mathbf{Z} \in \mathbb{R}^{r \times N}$  and consider an equivalent optimization problem of (26) as follows

$$\begin{aligned} \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{AX}\|_F^2 + \lambda \|\mathbf{Z}\|_{1,2}, \\ \text{s.t. } \mathbf{Z} \geq \mathbf{0}, \mathbf{X} - \mathbf{Z} = \mathbf{0}, \mathbf{X}^\top \mathbf{1}_r - \mathbf{1}_N = \mathbf{0}. \end{aligned} \quad (27)$$

Then, we form the augmented Lagrangian function corresponding to the two equality constraints

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{T}, \mathbf{t}) = & \frac{1}{2} \|\mathbf{Y} - \mathbf{AX}\|_F^2 + \lambda \|\mathbf{Z}\|_{1,2} + \text{tr}(\mathbf{T}^\top (\mathbf{X} - \mathbf{Z})) \\ & + (\mathbf{1}_r^\top \mathbf{X} - \mathbf{1}_N^\top) \mathbf{t} + \frac{\mu}{2} (\|\mathbf{X} - \mathbf{Z}\|_F^2 + \|\mathbf{X}^\top \mathbf{1}_r - \mathbf{1}_N\|_2^2), \end{aligned}$$

---

**Algorithm 3 The ADMM method for the joint sparse optimization problem**

---

**Input:**  $\mathbf{Y}$ ,  $\mathbf{A}$  and  $\lambda$ ;

**Output:** the optimal solution  $\mathbf{X}^*$ ;

- 1: Set  $k = 0$  and initialize  $\mathbf{X}^{(0)}$ ,  $\mathbf{Z}^{(0)}$ ,  $\mathbf{T}^{(0)}$  and  $\mathbf{t}^{(0)}$ ;
  - 2: **while** not converge **do**
  - 3:   Update  $\mathbf{X}^{(k+1)}$  by solving the linear system (28).
  - 4:   Update  $\mathbf{Z}^{(k+1)}$  by
  - $$\mathbf{Z}^{(k+1)} = \max \left\{ \mathcal{S}_{\lambda/\mu} \left( \mathbf{X}^{(k+1)} + \frac{1}{\mu} \mathbf{T}^{(k)} \right), \mathbf{0} \right\}.$$
  - 5:   Update the dual variables
  - $$\mathbf{T}^{(k+1)} = \mathbf{T}^{(k)} + \mu \left( \mathbf{X}^{(k+1)} - \mathbf{Z}^{(k+1)} \right),$$
  - $$\mathbf{t}^{(k+1)} = \mathbf{t}^{(k)} + \mu \left[ (\mathbf{X}^{(k+1)})^\top \mathbf{1}_r - \mathbf{1}_N \right],$$
  - 6:    $k \leftarrow k + 1$ , update  $\mu$ ;
  - 7: **end while**
  - 8: **return**  $\mathbf{Z}^{(k)}$
- 

where  $\mathbf{T} \in \mathbb{R}^{r \times N}$  and  $\mathbf{t} \in \mathbb{R}^N$  are Lagrangian multipliers, and  $\mu > 0$  is a penalty parameter for the augmented terms. For each iteration  $k \geq 1$ , the ADMM algorithm consists of an iterative procedure as follows:

a) *Minimize  $\mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{T}, \mathbf{t})$  w.r.t.  $\mathbf{X}$ :* The next iterate  $\mathbf{X}^{(k+1)}$  is obtained by solving the following linear system

$$\begin{aligned} & [\mathbf{A}^\top \mathbf{A} + \mu (\mathbf{1}_r \mathbf{1}_r^\top + \mathbf{I})] \mathbf{X}^{(k+1)} \\ &= \mathbf{A}^\top \mathbf{Y} + \mu \left( \mathbf{Z}^{(k)} + \mathbf{1}_r \mathbf{1}_N^\top \right) - \mathbf{T}^{(k)} - \mathbf{1}_r \left( \mathbf{t}^{(k)} \right)^\top. \end{aligned} \quad (28)$$

b) *Minimize  $\mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{T}, \mathbf{t})$  w.r.t.  $\mathbf{Z}$ :* We solve the following sub-problem

$$\mathbf{Z}^{(k+1)} = \arg \min_{\mathbf{Z} \geq \mathbf{0}} \lambda \|\mathbf{Z}\|_{1,2} + \frac{\mu}{2} \left\| \mathbf{Z} - \mathbf{X}^{(k+1)} - \frac{1}{\mu} \mathbf{T}^{(k)} \right\|_F^2.$$

One cheap estimate of the solution is

$$\mathbf{Z}^{(k+1)} = \max \left\{ \mathcal{S}_{\lambda/\mu} \left( \mathbf{X}^{(k+1)} + \frac{1}{\mu} \mathbf{T}^{(k)} \right), \mathbf{0} \right\}. \quad (29)$$

where  $\mathcal{S}_\alpha(x) = \text{sign}(x) \max\{|x| - \alpha, 0\}$  is the soft-thresholding operator.

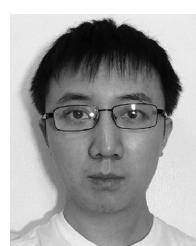
c) *Dual Ascent for  $\mathbf{T}$  and  $\mathbf{t}$ :*

$$\begin{aligned} \mathbf{T}^{(k+1)} &= \mathbf{T}^{(k)} + \mu \left( \mathbf{X}^{(k+1)} - \mathbf{Z}^{(k+1)} \right), \\ \mathbf{t}^{(k+1)} &= \mathbf{t}^{(k)} + \mu \left[ (\mathbf{X}^{(k+1)})^\top \mathbf{1}_r - \mathbf{1}_N \right]. \end{aligned} \quad (30)$$

These three steps are repeated until convergence. Finally, Algorithm 3 summarizes the whole procedure.

## REFERENCES

- [1] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
- [2] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [3] W.-K. Ma, J. M. Bioucas-Dias, P. Gader, T.-H. Chan, N. Gillis, A. Plaza, A. Ambikapathi, and C.-Y. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.
- [4] Q. Qu, N. M. Nasrabadi, and T. D. Tran, "Abundance estimation for bilinear mixture models via joint sparse and low-rank representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4404–4423, Jul. 2014.
- [5] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Aug. 1999.
- [6] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS*, Vancouver, BC, Canada, Dec. 2001, vol. 13, pp. 556–562.
- [7] S. A. Vavasis, "On the complexity of nonnegative matrix factorization," *SIAM J. Optimiz.*, vol. 20, no. 3, pp. 1364–1377, Aug. 2009.
- [8] D. Donoho and V. Stodden, "When does nonnegative matrix factorization give a correct decomposition into parts?", in *Proc. NIPS*, Whistler, BC, Canada, Dec. 2003, vol. 17, pp. 2004–2012.
- [9] S. Arora, R. Ge, R. Kannan, and A. Moitra, "Computing a nonnegative matrix factorization-provably," in *Proc. 44th STOC*, New York, NY, USA, May 2012, pp. 145–162.
- [10] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *Proc. IEEE Proc. CVPR*, Providence, RI, USA, Jun. 2012, pp. 1600–1607.
- [11] E. Esser, M. Moller, S. Osher, G. Sapiro, and J. Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3239–3252, Jul. 2012.
- [12] V. Bittorf, B. Recht, E. Re, and J. A. Tropp, "Factoring nonnegative matrices with linear programs," in *Proc. NIPS*, Lake Tahoe, CA, USA, Dec. 2012, pp. 1223–1231.
- [13] N. Gillis and S. Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 698–714, Apr. 2014.
- [14] N. Gillis and R. Luce, "Robust near-separable nonnegative matrix factorization using linear optimization," *J. Mach. Learn. Res.*, vol. 15, pp. 1249–1280, Apr. 2014.
- [15] J. W. Boardman, "Geometric mixture analysis of imaging spectrometry data," in *Proc. IEEE IGARSS*, Pasadena, CA, USA, Aug. 1994, vol. 4, pp. 2369–2371.
- [16] 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.
- [17] U. Araujo, B. Saldaña, R. Galvao, T. Yoneyama, H. Chame, and V. Visani, "The successive projections algorithm for variable selection in spectroscopic multicomponent analysis," *Chemometr. Intell. Lab. Syst.*, vol. 57, no. 2, pp. 65–73, Jul. 2001.
- [18] A. Ambikapathi, T.-H. Chan, C.-Y. Chi, and K. Keizer, "Hyperspectral data geometry-based estimation of number of endmembers using p-norm based pure pixel identification algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2753–2769, May 2013.
- [19] H. Ren and C.-I. Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, Oct. 2003.
- [20] A. Kumar, V. Sindhwan, and P. Kambadur, "Fast conical hull algorithms for near-separable nonnegative matrix factorization," in *Proc. ICML*, Atlanta, GA, USA, Jun. 2013, vol. 28, pp. 231–239.
- [21] M. E. Winter, "N-findr: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE 3753 Imaging Spectrometry V*, Denver, CO, USA, Oct. 1999, vol. 3753, pp. 266–275.
- [22] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, Nov. 2011.
- [23] Z. Shi, W. Tang, Z. Duren, and Z. Jiang, "Subspace matching pursuit for sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3256–3274, Jun. 2014.
- [24] W. Tang, Z. Shi, and Y. Wu, "Regularized simultaneous forward-backward greedy algorithm for sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5271–5288, Sep. 2014.
- [25] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.
- [26] N. Akhtar, F. Shafait, and A. Mian, "Futuristic greedy approach to sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2157–2174, May 2015.
- [27] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundat. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Nov. 2010.
- [28] M. Hong and Z. Luo, "On the linear convergence of the alternating direction method of multipliers," *arXiv:1208.3922*, Mar. 2013.
- [29] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
- [30] X. Fu, W.-K. Ma, T.-H. Chan, J. M. Bioucas-Dias, and M.-D. Iordache, "Greedy algorithms for pure pixels identification in hyperspectral unmixing: A multiple-measurement vector viewpoint," in *Proc. EUSIPCO*, Marrakech, Morocco, Sep. 2013.
- [31] N. Gillis, "Successive nonnegative projection algorithm for robust nonnegative blind source separation," *SIAM J. Imag. Sci.*, vol. 7, no. 2, pp. 1420–1450, 2014.
- [32] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
- [33] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.
- [34] D. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, May 2001.
- [35] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1974, ch. 23, p. 161.
- [36] H. W. Kuhn, "The Hungarian method for the assignment method," *Nav. Res. Logist. Quart.*, vol. 2, pp. 83–97, 1955.
- [37] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
- [38] Z. Guo, T. Wittman, and S. Osher, "L1 unmixing and its application to hyperspectral image enhancement," in *Proc. SPIE Conf. Algorithms Technol. Multispectral, Hyperspectral, Ultraspectral Imaging XV*, Orlando, FL, USA, 2009.
- [39] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low rank representation," in *Proc. NIPS*, Granada, Spain, Dec. 2011, pp. 612–620.
- [40] F. Kruse, J. Boardman, and J. Huntington, "Comparison of airborne hyperspectral data and EO-1 hyperion for mineral mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.



**Qing Qu** Ph.D. candidate in the Dep. of Electrical Engineering, Columbia University. Prior to that, he received the B.Eng. in electronic engineering from Tsinghua University, Beijing, in July 2011, and the M.Sc. in electrical engineering from the Johns Hopkins University, Baltimore, in Dec. 2012. He interned at the U.S. Army Research Laboratory from Jun. 2012 to Aug. 2013. His current research focuses on developing practical algorithms and provable guarantees for signal processing and machine learning problems with low intrinsic data structure.



**Nasser M. Nasrabadi** (S'80–M'84–SM'92–F'01) received the B.Sc. (Eng.) and Ph.D. degrees in electrical engineering from Imperial College of Science and Technology (University of London), London, England, in 1980 and 1984, respectively.

From October 1984 to December 1984, he worked for IBM (UK) as a Senior Programmer. During 1985 to 1986, he was with Philips Research Laboratory in NY as a member of technical staff. From 1986 to 1991, he was an Assistant Professor in the Department of Electrical Engineering at Worcester

Polytechnic Institute, Worcester, MA. From 1991 to 1996 he was an associate professor with the Department of Electrical and Computer Engineering at State University of New York at Buffalo, Buffalo, NY. Since September 1996, he has been a Senior Research Scientist (ST) with the US Army Research Laboratory (ARL) working on image processing and automatic target recognition.

Dr. Nasrabadi has served as an associate editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS, SYSTEMS, AND VIDEO TECHNOLOGY, and the IEEE TRANSACTIONS ON NEURAL NETWORKS. He is also a Fellow of ARL and SPIE. His current research interests are in Hyperspectral imaging, automatic target recognition, statistical machine learning theory, robotics, and neural networks applications to image processing. He is also a Fellow of ARL, SPIE, and IEEE.



**Trac D. Tran** (S'94–M'98–SM'08–F'14) received the B.S. and M.S. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1993 and 1994, respectively, and the Ph.D. degree in electrical engineering from the University of Wisconsin, Madison, WI, USA, in 1998. In July of 1998, he joined the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA, where he currently holds the rank of Professor. In the summer of 2002, he was an American Society

for Engineering Education/ONR Summer Faculty Research Fellow at the Naval Air Warfare Center Weapons Division at China Lake, CA, USA. He is currently a regular consultant for the U.S. Army Research Laboratory in Adelphi, MD, USA. His research interests are in the field of digital signal processing, particularly in sparse representation, sparse recovery, sampling, multirate systems, filter banks, transforms, wavelets, and their applications in signal analysis, compression, processing, and communications. His pioneering research on integer-coefficient transforms and pre-/postfiltering operators has been adopted as critical components of Microsoft Windows Media Video 9 and JPEG XR the latest international still image compression standard International Organization for Standardization/International Electrotechnical Commission 29199-2.

Dr. Tran was the co-director (with Prof. J. L. Prince) of the 33rd Annual Conference on Information Sciences and Systems, Baltimore, MD, USA, in March 1999. He has served as the Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING as well as IEEE TRANSACTIONS ON IMAGE PROCESSING. He was a former member of the IEEE Technical Committee on Signal Processing Theory and Methods and is a current member of the IEEE Image Video and Multidimensional Signal Processing Technical Committee. He was the recipient of the NSF CAREER Award in 2001, the William H. Huggins Excellence in Teaching Award from Johns Hopkins University in 2007, and the Capers and Marion McDonald Award for Excellence in Mentoring and Advising in 2009.