SUBSPACE VERTEX PURSUIT FOR SEPARABLE NON-NEGATIVE MATRIX FACTORIZATION IN HYPERSPECTRAL UNMIXING

Qing Qu[‡], *Xiaoxia Sun*[*], *Nasser M. Nasrabadi*[†], *Trac D. Tran*[*]

[‡] EE Department, Columbia University, W. 120th Street, New York, NY, 10027
[*] Department of ECE, the Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD, 21218
[†] U.S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, MD, 20783

ABSTRACT

Recently, the separability assumption turns the nonnegative matrix factorization (NMF) into a tractable problem. The assumption coincides with the pixel purity assumption and provides new insights for the hyperspectral unmixing problem. In this paper, we present a quasi-greedy algorithm for solving the problem by employing a back-tracking strategy. Unlike the current greedy methods, the proposed method can refresh the endmember index set in every iteration. Therefore, our method has two important characteristics: (i) low computational complexity comparable to state-of-the-art greedy methods but (ii) empirically enhanced robustness against noise. Finally, computer simulations on synthetic hyperspectral data demonstrate the effectiveness of the proposed method.

***Index Terms*—** nonnegative matrix factorization, subspace pursuit, hyperspectral unmixing, endmember detection

1. INTRODUCTION

Due to the low spatial resolution of imaging sensors, spectral unmixing [1, 2], which consists of pure endmember extraction and the abundance estimation, is a major issue in hyperspectral imagery. To deal with this problem, the most widely used model is the linear mixture model (LMM) that every mixed pixel is an additive linear combination of the pure endmembers. Based on the LMM, it is natural to consider formulating the problem as a NMF problem that simultaneously factorizes the data into two parts: the endmember and the abundance matrices [3]. However, the NMF problem is notoriously difficult to solve and the solution is not unique [4].Recently, promising alternative theories have been developed based on the separability assumption on the data which guarantees the problem to be solved within polynomial time [5, 6]. Geometrically, the separability assumption states the following: all columns of the data lies in the convex hull of some representative columns within the data itself. It is essentially the pure pixel assumption in the hyperspectral unmixing, that each material has at least one pure pixel in the data [1].

In the literature, various methods have been developed for the separable NMF problem [3]. Boardman [7] proposed the pixel purity index via random projections. Winter [8] and Chan et al. [9] looked into the problem from the perspective of convex geometry and proposed simplex volume maximization algorithms such as N-FINDR [8]. Various greedy methods, e.g., vertex component analysis (VCA) [10], successive projection algorithm (SPA) [11, 12] and extreme ray (XRAY) [13], have been proposed and studied. Quite recently, Esser et al. [14] and Elhamifar et al. [15] formulated the problem as an joint sparse recovery (i.e., $l_{1,p}$-norm minimization) problem and solved it by convex optimization techniques; Iordache et al. [16] applied a similar method for the spectral unmixing.

In this paper, we propose a novel quasi-greedy method, named subspace vertex pursuit (SVP), to solve the separable NMF problem. It employs a back-tracking strategy similar to the subspace pursuit (SP) [17, 18]. In every iteration, if we assume that the number of materials r is known, SVP first identifies r candidate points in the current convex hull and then merges them with the r candidate points from the previous step, forming a new set of $2r$ candidate points; second, SVP selects r points out of the $2r$ candidate set by solving a joint sparse optimization sub-problem. Unlike state-of-the-art greedy methods, SVP could freely remove a candidate point, which was considered to be reliable in previous iterations but shown to be wrong in the current iteration, from the candidate dataset. Therefore, SVP shows more robust performances in comparison with current greedy methods for the separable NMF problem. On the other hand, compared with the joint sparse minimization method [14, 15] with $O(N^2)$ variables to be optimized (N is the size of the dataset), as SVP solves an optimization sub-problem of much smaller scale and converges in only a few iterations, the overall computational complexity of SVP is much smaller.

2. PROBLEM STATEMENT AND ASSUMPTIONS

2.1. Linear Mixture Model and the NMF problem

Suppose that $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N] \in \mathbb{R}_+^{L\times N}$ is a hyperspectral data cube, where L is the data dimension and N

This work has been partially supported by NSF under Grant CCF-1117545 and ARO under Grant 60219-MA.

is the number of data samples. If we assume that $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_r] \in \mathbb{R}^{L\times r}$ is the pure endmember matrix and $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_r] \in \mathbb{R}^{r\times N}$ is the associated abundance matrix, where r is the number of materials. Then the data $\mathbf{Y}$ can be represented by the following model

$$\mathbf{Y} = \mathbf{F}\mathbf{W} + \mathbf{N} \tag{1}$$

where $\mathbf{N} \in \mathbb{R}^{L\times N}$ is a Gaussian white noise matrix. To be physically meaningful, the matrix $\mathbf{W}$ should also satisfy

$$\mathbf{W}\mathbf{1}_r^T = \mathbf{1}_N, \mathbf{W} \succeq \mathbf{0} \tag{2}$$

where $\mathbf{1}_r \in \mathbb{R}^r$ and $\mathbf{1}_N \in \mathbb{R}^N$ with all entries equaling to unitary. Based on the LMM, we can formulate the endmember extraction problem as

$$\begin{aligned} \{\hat{\mathbf{F}}, \hat{\mathbf{W}}\} &= \arg\min_{\mathbf{P},\mathbf{Q}} ||\mathbf{Y} - \mathbf{P}\mathbf{Q}||_F^2, \\ s.t.\ \mathbf{P} &\succeq \mathbf{0}, \mathbf{Q} \succeq \mathbf{0},\ \mathbf{Q}^T\mathbf{1}_r = \mathbf{1}_N \end{aligned} \tag{3}$$

This is an NMF problem with the sum-to-one constraint, which is a highly ill-posed, NP-hard problem and does not have a unique solution [4].

2.2. Separable Assumption for NMF

Recently, Donoho et al. [5] and Arora et al. [6] have provided a sufficient condition which guarantees the NMF problem to have unique solutions and could be solved within polynomial time. Their condition could be stated as follows.

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

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

$$\mathbf{W}\mathbf{\Pi} = \left[\begin{array}{cc} \mathbf{I}_r & \mathbf{U} \end{array}\right] \tag{4}$$

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

2.3. Key Observations for the problem

Based on the LMM and separable assumption for the data, we have the following observations for the problem.

Proposition 1 (Convex Geometry) *If the data* $\mathbf{Y}$ *are separable and satisfy the LMM assumption, then all the columns of* $\mathbf{Y}$ *lie in the convex hull*

$$\mathcal{H} = \{\mathbf{y} \in \mathbb{R}^L |\ \mathbf{y} = \mathbf{F}\mathbf{w},\ \mathbf{w} \succeq \mathbf{0}, \mathbf{1}_r^T\mathbf{w} = 1\} \tag{5}$$

which is generated by the columns of $\mathbf{F}$*, where* $\mathbf{F}$ *is a submatrix of* $\mathbf{Y}$ *and each column of* $\mathbf{F}$ *is a vertex of* $\mathcal{H}$.

Proposition 2 (Row Sparsity Property [15, 19]) *If the data* $\mathbf{Y}$ *are separable and satisfy the LMM assumption, then we have*

$$\mathbf{Y} = \mathbf{Y}\mathbf{X},\ where\ \mathbf{X} = \mathbf{\Pi}\left[\begin{array}{cc} \mathbf{I}_r & \mathbf{U} \\ \mathbf{0} & \mathbf{0} \end{array}\right]\mathbf{\Pi}^T \tag{6}$$

where $\mathbf{\Pi}$ *is a column permutation matrix,* $\mathbf{X} \in \mathbb{R}^{N\times N}$ *and* $\mathbf{U} \in \mathbb{R}_+^{r\times(N-r)}$ *satisfying* $\mathbf{U}^T\mathbf{1}_r = \mathbf{1}_{N-r}$*. Furthermore, if* $\hat{\mathbf{Y}} = \mathbf{Y} + \mathbf{N}$ *is near-separable, we have*

$$\hat{\mathbf{Y}} = \hat{\mathbf{Y}}\mathbf{X} + \hat{\mathbf{N}} \tag{7}$$

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

From Proposition 1, we can see that the separable NMF problem reduces to finding the vertices of a convex hull; from Proposition 2, as generally $r << N$, the problem is equivalent to a multiple measurement vector (MMV) problem [20] that tries to recover a row sparse matrix $\mathbf{X}$. Therefore, it is natural to consider the following problem

$$\min_{\mathbf{X}} ||\mathbf{X}||_{row-0},\ s.t.\ \mathbf{X} \in \Phi(\mathbf{C}) \tag{8}$$

where $||\cdot||_{row-0}$ denotes the number of nonzero rows and $\Phi(\mathbf{C}) = \{\mathbf{Y} = \mathbf{Y}\mathbf{C},\ \mathbf{C} \geq \mathbf{0},\ \mathbf{C}^T\mathbf{1}_N = \mathbf{1}_N\}$ is the feasible set. The problem in (8) is NP-hard in general, Esser et al. [14] and Elhamifar et al. [15] proposed to solve the following convex relaxed problem

$$\min_{\mathbf{X}} ||\mathbf{X}||_{1,p},\ s.t.\ \mathbf{X} \in \Phi(\mathbf{C}) \tag{9}$$

where $||\mathbf{X}||_{1,p} = \sum_{i=1}^N ||\mathbf{x}^i||_p$ for some $p \geq 1$.

However, the problem (9) has $O(N^2)$ variables to optimize which is impractical for large-scale dataset. Therefore, greedy methods such as the simultaneous orthogonal matching pursuit (SOMP) [20] are interesting alternatives for solving the problem [21]. Nonetheless, because SOMP sequentially detects the nonzero rows and cannot remove the rows that were considered to be reliable in the previous step but found to be wrong in the current iteration, it is not robust to noise. Therefore, it encourages us to consider the simultaneous SP method [17, 18, 22] for the separable NMF problem by incorporating a back-tracking strategy.

3. SUBSPACE VERTEX PURSUIT

In each iteration, the SVP method maintains a set of r columns of Y, performs a simple test in the current convex hull, and then refines the subset. If the data Y does not lie in the current estimation for the correct convex hull, one refines the estimate by only retaining the reliable candidates, discarding the unreliable ones while adding the same number of more reliable new candidates. The expectation is that the recursive refinements of the estimate will consecutively lead to subspaces with strictly decreasing distance from the data Y. By this scheme, we can refresh the candidate pixels found unreliable in the current iteration. Specifically, every iteration of SVP consists of 3 steps: (i) *detection step*, that finds r new candidate vertices simultaneously and (ii) *refinement step*, that select r most reliable candidate out of $2r$ index set and (iii) *projection step*, that projects the data onto the current convex hull. Next, we describe each step in detail.

3.1. Detection Step

In the kth iteration, based on the r indices in the $(k-1)$th step, we have to find r new indices simultaneously that are supposed to contain as many vertex candidates as possible. Similar to SOMP, we consider the correlation matrix $\mathbf{C}^{(k)} = (\mathbf{R}^{(k-1)})^T\mathbf{Y}$ and pick r indices corresponding to the r largest values of $||(\mathbf{C}^{(k)})^i||_2$ $(1 \leq i \leq N)$, where the matrix $\mathbf{R}^{(k-1)}$ is the residual matrix calculated in the previous projection step that will be described below. In the noiseless case, such a strategy is guaranteed to find at least one vertex in each iteration. In the noisy case, as the candidate is selected based on the correlations of all data vectors, it is robust to noise.

3.2. Refinement Step

For the kth step, suppose we have a merged $2r$ index set $\hat{I}^{(k)}$, we need to find r most reliable candidates as the set $I^{(k)}$ from the $2r$ set $\hat{I}^{(k)}$. In the SP method for sparse recovery, they solve a least squares problem as follows

$$\mathbf{X}_0 = \arg\min_{\mathbf{B}} ||\mathbf{Y} - \mathbf{Y}_{\hat{I}^{(k)}}\mathbf{B}||_F^2 \tag{10}$$

The indices in $\hat{I}^{(k)}$ which produces r largest values of $||\mathbf{X}_0^i||$ $(i \in \hat{I}^{(k)})$ are selected as new candidates. However, in our case, such a strategy does not guarantee that the selected candidates are more reliable. As the vertices are the most representative points in the data, based on Proposition 1 and 2, we consider solving the following sub-problem instead

$$\begin{aligned} &\min_{\mathbf{X}_1} \frac{1}{2}||\mathbf{Y}_{\hat{I}^{(k)}} - \mathbf{Y}_{\hat{I}^{(k)}}\mathbf{X}_1||_F^2 + \lambda||\mathbf{X}_1||_{1,2}, \\ &s.t.\ \mathbf{X}_1 \succeq \mathbf{0},\ \mathbf{X}_1^T\mathbf{1}_{2r} - \mathbf{1}_N = \mathbf{0}, \end{aligned} \tag{11}$$

where $\lambda > 0$ is a regularization parameter. This sub-problem can be efficiently solved by the alternating direction method of multipliers [23]. Because we impose $l_{1,2}$ penalty on the representation matrix $\mathbf{X}_1 \in \mathbb{R}^{2r\times 2r}$, $\mathbf{X}_1$ is supposed to be row sparse. The set of points corresponding to the nonzero rows of $\mathbf{X}_1$ are the smallest number of points to form a convex hull for the rest of the data, they are more reliable candidates to be vertices. Therefore, we pick r indices in $\hat{I}(k)$ corresponding to the largest r rows of $\mathbf{X}_1$ to form the new index set $I^{(k)}$. Compared with the joint sparse optimization problem (9) with $O(N^2)$ variables, the sub-problem (11) has only $O(r^2)$ variables to be optimized. Given $r << N$, the scale of the problem (11) is much smaller than (9) and the computational complexity will not change as the sample number N varies.

3.3. Projection Step

In the projection step, given the new candidate set $I^{(k)}$ from the refinement step, all the data are projected to the convex hull generated by the candidates $\mathbf{Y}_{I^{(k)}}$ to get the residual. The projection and residual can be computed by solving the nonnegative least squares problem

$$\begin{aligned} \mathbf{H} &= Proj(\mathbf{Y}, \mathbf{Y}_{I^{(k)}}) = \arg\min_{\mathbf{B}\succeq\mathbf{0}} ||\mathbf{Y} - \mathbf{Y}_{I^{(k)}}\mathbf{B}||_F^2, \\ \mathbf{R}^{(k)} &= Resid(\mathbf{Y}, \mathbf{Y}_{I^{(k)}}) = \mathbf{Y} - \mathbf{Y}_{I^{(k)}}\mathbf{H}, \end{aligned} \tag{12}$$

where the matrix $\mathbf{H}$ in the problem (12) can be solved exactly using block coordinate descent method as in [13]. Here we cheaply approximate the solution by projecting the least squares solution back to the nonnegative orthant $\mathbf{H} = \max\{(\mathbf{Y}_{I^{(k)}}^T\mathbf{Y}_{I^{(k)}})^{-1}\mathbf{Y}_{I^{(k)}}\mathbf{Y}, \mathbf{0}\}$.

Algorithm 1 Subspace Vertex Pursuit for Spectral Unmixing

Input: Data matrix $\mathbf{Y}$, number of endmembers r;

Output: The matrices $\mathbf{F}$, $\mathbf{W}$;

1: Initialize: $k = 1$, $\mathbf{C}^{(0)} = \mathbf{Y}^T\mathbf{Y}$, the residual $\mathbf{R}^{(0)} = Resid(\mathbf{Y}, \mathbf{Y}_{I^{(0)}})$, where $I^{(0)}$ is the index set

$$I^{(0)} = \{r \text{ indices correspond to the } r \text{ largest values of } ||(\mathbf{C}^{(0)})^i||_2\ (1 \leq i \leq N)\}$$

2: **while** not converged **do**

3: **Detection Step:** Update $\mathbf{C}^{(k)} = (\mathbf{R}^{(k-1)})^T\mathbf{Y}$ and

$$\hat{I}^{(k)} = I^{(k-1)} \bigcup \{r \text{ indices correspond to the } r \text{ largest values of } ||(\mathbf{C}^{(k)})^i||_p\ (1 \leq i \leq N)\},$$

4: **Refinement Step:** Update the index set

$$I^{(k)} = \{r \text{ most reliable indices in the index set } \hat{I}^{(k)}\}$$

5: **Projection Step:** Update $\mathbf{R}^{(k)} = Resid(\mathbf{Y}, \mathbf{Y}_{I^{(k)}})$;

6: If $||\mathbf{R}^{(k)}||_F > ||\mathbf{R}^{(k-1)}||_F$, $I^{(k)} = I^{(k-1)}$;

7: $k = k + 1$,

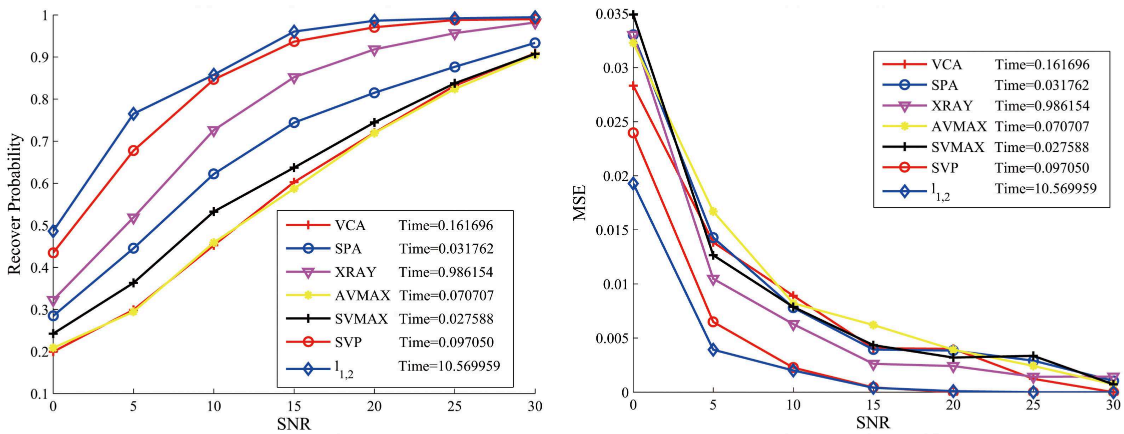
8: **end while**

9: Let $\mathbf{F} = \mathbf{Y}_{I^{(k)}}$ and calculate the abundance matrix $\mathbf{W}$

$$\mathbf{W} = \arg\min_{\mathbf{W}} ||\mathbf{Y} - \mathbf{F}\mathbf{W}||_F^2,\ s.t.\ \mathbf{W} \succeq \mathbf{0},\ \mathbf{1}_r^T\mathbf{W} = \mathbf{1}_N^T,$$

3.4. Further Explanation

The overall algorithm is summarized in Algorithm 1. In Step 4, we solve (11) to find the index set $I^{(k)}$; Step 6 is to guaran-



(a) Recovery Probability (b) Mean Square Error

Fig. 1. Comparison of state-of-the-art methods with the proposed methods on a synthetic dataset.

tee the monotonic decreasing of the residual so that the algorithm converges, and empirically the algorithm converges in $5 \sim 10$ iterations; in Step 9, the endmember matrix is recovered by setting $\mathbf{F} = \mathbf{Y}_{I(k)}$ and the abundance $\mathbf{W}$ is recovered by solving a fully constraint least squares problem [24].

4. EXPERIMENT RESULT

We compare the noise robustness and the computational complexity for the proposed SVP method with state-of-the-art algorithms on a synthetic hyperspectral dataset for endmember detection. Our comparisons are based on two criteria:

(1) (Recovery probability) Suppose the number of endmember is r and we repeat the simulation for N_s times, the recovery probability is defined by

$$Pr = \frac{the\ number\ of\ recovered\ indices}{N_s \times r} \tag{13}$$

(2) (Mean square error (MSE)) For each simulation, suppose the estimated endmembers are $\{\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \cdots, \hat{\mathbf{f}}_N\}$, the MSE is defined as

$$MSE = \min_{\boldsymbol{\pi} \in \Pi} \frac{1}{r} \sum_{j=1}^{r} ||\hat{\mathbf{f}}_j - \mathbf{f}_{\pi_j}||^2 \tag{14}$$

where $\boldsymbol{\pi} = [\pi_1, \pi_2, \cdots, \pi_r]$, and $\Pi = \{\boldsymbol{\pi} \in \mathbb{R}^r | \pi_l \in \{1, 2, \cdots, r\}, \pi_l \neq \pi_m, l \neq m\}$ is the set of all the permutations of $\{1, 2, \cdots, r\}$. The problem in (14) can be efficiently solved by the Hungarian algorithm [25].

To generate the synthetic data, first, for each simulation we randomly extract $r = 20$ pure endmembers from a pruned USGS library[1] (the angle between each endmember in the library is larger than 10^o) to form the feature matrix $\mathbf{F} \in \mathbb{R}^{224\times 20}$, where the dimension for each endmember is $L = 224$. The weighting matrix $\mathbf{W}$ is generated by $[\mathbf{I}_r, \mathbf{U}]\boldsymbol{\Pi} \in \mathbb{R}^{20\times 500}$, where $\mathbf{I}_r$ is an identity matrix so that there exists one pure endmember for each selected material. Each column of $\mathbf{U} \in \mathbb{R}^{20\times 480}$ is generated by the Dirichlet distribution. $\boldsymbol{\Pi}$ is a random permutation matrix so that the order of the pure endmembers will not affect the performance of an algorithm. The mixed matrix $\mathbf{Y}$ is generated as $\mathbf{Y} = \mathbf{FW} + \mathbf{N}$, where $\mathbf{N}$ is an i.i.d. white Gaussian matrix. We compare the proposed SVP method with the following algorithms for endmember detection: VCA [10], SPA [11], successive volume maximization (SVMAX) [9], alternating volume maximization (AVMAX) [9], XRAY [13] and the joint sparse recovery method ($l_{1,2}$) [15]. The recovery probability and the MSE for different signal-to-noise ratios (SNR) are shown in Fig. 1. From the results, we can conclude that (i) the proposed SVP method shows much enhanced noise robustness compared with state-of-the-art greedy methods with approximately the same computational time; (ii) our method shows comparable result with $l_{1,2}$-optimization method but much reduced computational complexity.

[1]http://www.lx.it.pt/ bioucas/code/sunsal demo.zip

5. CONCLUSION

In this paper, we present a novel quasi-greedy approach, named SVP, to solve the separable NMF problem. The proposed approach adopts a back-tracking strategy and solves a small-scale sub-optimization problem in each iteration. Our method has superior noise robustness and low computational cost, which has been demonstrated in the numerical experiments. In the future, we would like to make further analysis of the proposed method and apply it to other separable NMF problem besides hyperspectral unmixing.

6. 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 Observ. 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, “An signal processing perspective on hyperspectral unmixing,” *IEEE Signal Process. Mag.*, to be appear, 2014.

[4] S. A. Vavasis, “On the complexity of nonnegative matrix factorization,” *SIAM Journal on Optimization*, vol. 20, no. 3, pp. 1364-1377, 2009.

[5] D. Donoho and V. Stodden, “When does non-negative matrix factorization give a correct decomposition into parts?” *NIPS*, 2003.

[6] S. Arora, R. Ge, R. Kannan, and A. Moitra, “Computing a nonnegative matrix factorization-provably,” *STOC*, pp.145-162, 2012.

[7] J.W. Boardman, “Geometric mixture analysis of imaging spectrometery data,” *Proc. Int. Geosci. and Remote Sens. Symp.*, vol. 4, pp. 2369-2371, 1994.

[8] M. E. Winter, “N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data,” *in Proc. SPIE Conf. Imaging Spectrometry*, Pasadena, CA, pp. 266-275, Oct. 1999.

[9] 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, 2011.

[10] J. M. P. Nascimento and J. M. B. 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.

[11] N. Gillis and S. Vavasis, “Fast and robust recursive algorithms for separable nonnegative matrix factorization,” arXiv:1208.1237, 2012.

[12] 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, 2013.

[13] A. Kumar, V. Sindhwani, and P. Kambadur, “Fast conical hull algorithms for near-separable non-negative matrix factorization,” *ICML*, 2013.

[14] 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, 2012.

[15] E. Elhamifar, G. Sapiro, and R. Vidal, “See all by looking at a few: Sparse modeling for finding representative objects,” *CVPR*, 2012.

[16] 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.

[17] W. Dai, O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction”, *IEEE Trans. on Information Theory*, vol. 55, no. 5, pp. 2230-2249, May 2009.

[18] D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 301-321, 2009.

[19] V Bittorf, B. Recht, E. Re, and J.A. Tropp, “Factoring nonnegative matrices with linear programs,” *NIPS*, 2012.

[20] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Algorithms for simultaneous sparse approximation. part I: greedy pursuit,” *Signal Process.*, 2006.

[21] 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,” *EUSIPCO*, 2013.

[22] Y. Chen, N. M. Nasrabadi, T. D. Tran, “Hyperspectral image classification using dictionary-based sparse representation”, *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973-3985, May 2011.

[23] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1-122, Nov. 2010.

[24] 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, 2001.

[25] H. W. Kuhn, “The Hungarian method for the assignment method,” *Nav. Res. Logist. Quart.*, vol. 2, pp. 83-97, 1955.