

SEMI-SUPERVISED HYPERSPECTRAL UNMIXING VIA THE WEIGHTED LASSO

Konstantinos E. Themelis^{1,2}, Athanasios A. Rontogiannis¹ and Konstantinos Koutroumbas¹

¹Inst. for Space Applications and Remote Sensing, NOA, 152 36, Athens, Greece

²Dept. of Informatics and Telecommunications, University of Athens, 157 84, Athens, Greece
{themelis, tronto, koutroumbas}@space.noa.gr

ABSTRACT

In this paper a novel approach for semi-supervised hyperspectral unmixing is presented. First, it is shown that this problem inherently accepts a sparse solution. Then, based on this observation, an efficient ℓ_1 regularized least squares algorithm is proposed, in which the constraints that are naturally imposed to the problem, are suitably incorporated. Simulations results show that the proposed method achieves the performance of quadratic programming based techniques with much lower computational requirements.

Index Terms— hyperspectral remote sensing, linear spectral unmixing, subset selection, sparse representation

1. INTRODUCTION

Hyperspectral remote sensing has gained considerable attention in recent years, due to its wide range of applications, such as environmental monitoring and terrain classification [1]. Hyperspectral sensors have the ability to collect data in hundreds of adjacent spectral bands. However, due to their low spatial resolution, more than one different materials' spectra may be mixed into each individual pixel. Identification of the materials present in mixed pixels can be done through *Spectral Unmixing* (SU). According to this procedure, the measured spectrum of a mixed pixel is decomposed into a collection of constituent spectra, called *endmembers*, and a set of corresponding fractions, called *abundances*, that indicate the proportion of each endmember present in the pixel.

Spectral unmixing can be categorized as supervised, unsupervised and semi-supervised. In supervised SU [2], it is assumed that the knowledge of the endmembers' spectral signatures contributing to a pixel is given *a priori*. On the contrary, in unsupervised SU, [3]-[4], the endmembers' spectra are estimated directly from the data. Finally, semi-supervised SU [5] assumes that a spectral library is given, containing multiple endmembers' spectral signatures. The objective in the later case, is (a) to determine how many and which endmembers are present in the mixed pixel under study and (b) to use the selected endmembers to estimate the corresponding abundance fractions. Two constraints that the abundance fractions should satisfy in order to remain physically meaningful are that they (a) should be nonnegative and (b) sum to one.

In this paper a novel approach for semi-supervised SU in hyperspectral images is presented. The approach is based on a properly modified weighted ℓ_1 -regularized least squares algorithm. The motivation of using the sparsity promoting ℓ_1 norm is that in practice, only a small number of the available endmembers are present in each pixel. Based on this observation, a weighted version [6] of the celebrated least absolute shrinkage and selection operator (lasso) criterion [7] is utilized, where weights are used for penalizing different coefficients in the ℓ_1 -regularization scheme. To efficiently solve the

ℓ_1 minimization problem, the Least Angle Regression (LARS) algorithm [8] is used. However, the ℓ_1 minimization criterion contradicts the additivity constraint of the problem, resulting in degradation of estimation performance. To tackle this effect, the additivity constraint is suitably incorporated into the minimization problem, resulting in a modified cost function. As verified by simulations for both real and simulated hyperspectral data, with this modification the proposed method achieves the performance expected by quadratic programming methods, at a much lower computational cost. It should be mentioned that a similar approach to the SU problem has been recently presented in [9]. Note, though, that not only the iterative algorithm utilized in [9], but also the way the constraints of the problem are exploited, are quite different compared to those in the current work.

2. PROBLEM FORMULATION

In a hyperspectral image, each pixel is represented by a L -dimensional vector \mathbf{y} , where L is the number of available spectral bands. The elements of \mathbf{y} correspond to the reflectance energy measured at the respective spectral bands. In this work, we concentrate on the linear mixing model, which is most commonly used for SU. This model assumes that the received pixel's spectrum is generated by a linear combination of endmembers' spectra. Suppose that we possess the spectral signatures of p materials that may exist in the image. These spectral signatures either come from a spectral library or can be obtained from pure pixels of the image. Then, \mathbf{y} can be described by the following linear regression model:

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}, \quad (1)$$

where $\Phi = [\phi_1 \ \phi_2 \ \dots \ \phi_p] \in \mathbb{R}_+^{L \times p}$, with $L \gg p$, is the mixing matrix containing the endmembers' spectra, \mathbf{x} is a $p \times 1$ vector with the corresponding abundance fractions, and \mathbf{n} is an $L \times 1$ additive noise vector. Naturally, \mathbf{x} has a sparse form since only a small number of materials coexist in each pixel.

Due to physical considerations, the fraction vector \mathbf{x} satisfies two constraints, (a) the full additivity constraint:

$$\mathbf{1}^T \mathbf{x} = \sum_{i=1}^p x_i = 1, \quad (2)$$

where $\mathbf{1}$ is a p -dimensional vector of ones and $(\cdot)^T$ denotes transposition, and (b) the nonnegativity constraint:

$$0 \leq x_i \leq 1, \quad i = 1, 2, \dots, p. \quad (3)$$

In geometrical terms, it is easily seen that the constraints define the $p - 1$ standard simplex in the p -dimensional space, denoted by \mathbb{S} .

We can write

$$\mathbf{x} \in \mathbb{S}, \quad \mathbb{S} = \left\{ x \in \mathbb{R}^p \mid \sum_{i=1}^p x_i = 1, x_i \geq 0, \forall i \right\}. \quad (4)$$

As mentioned above, \mathbf{x} is sparse, that is the cardinality $\xi = |\mathcal{S}(\mathbf{x})|$ of its support

$$\mathcal{S}(\mathbf{x}) = \{i \in \{1 \dots p\} \mid x_i \neq 0\} \quad (5)$$

satisfies $\xi < p$. In the following, the sparsity of \mathbf{x} is exploited, in order to determine the ξ endmembers which are present in the pixel under study, and then to estimate the corresponding abundance fractions, subject to the constraints.

3. PROPOSED METHOD

Given the observation model (1) and the polytope of constraints \mathbb{S} , the sparse solution follows from the optimization problem

$$\min_{\mathbf{x} \in \mathbb{S}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 \leq \epsilon, \quad (6)$$

where $\|\mathbf{x}\|_0$ is the number of nonzero x_i 's and ϵ is a small positive parameter. This problem is nonconvex and its solution requires an exhaustive combinatorial search. To alleviate this difficulty, the ℓ_0 norm in (6) can be replaced by the ℓ_1 norm [7], i.e.

$$\min_{\mathbf{x} \in \mathbb{S}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 \leq \epsilon, \quad (7)$$

where $\|\mathbf{x}\|_1 = \sum_{i=1}^p |x_i|$. The ℓ_1 penalization criterion associated with (7), also called the *lasso* [7], is

$$\mathcal{J}_{\text{lasso}} = \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \lambda_\alpha \sum_{i=1}^p |x_i|, \quad (8)$$

where λ_α is a nonnegative parameter that balances the sparsity of the solution with estimation accuracy. Minimization of $\mathcal{J}_{\text{lasso}}$ subject to $\mathbf{x} \in \mathbb{S}$ will give the lasso solution, $\mathbf{x}_{\text{lasso}}$, which is used simultaneously for endmember selection and abundance estimation.

However, in eq. (8) all x_i 's are equally penalized in the ℓ_1 penalty term. As shown in [6], this setup can lead to inconsistent endmember selection. To face this problem, different weights can be assigned to different x_i 's. This gives rise to the *weighted lasso* formulation, [6],

$$\mathbf{x}_{\text{wlasso}} = \arg \min_{\mathbf{x} \in \mathbb{S}} \left\{ \overbrace{\|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \lambda_\alpha \sum_{i=1}^p w_i |x_i|}^{\mathcal{J}_{\text{wlasso}}}, \quad (9)$$

where $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_p]^T$ is a weighting vector. A desirable property of the weights w_i 's is to be inversely proportional to their corresponding x_i 's, i.e. $w_i = 1/|x_i|$. Penalizing the nonzero elements of \mathbf{x} using small weights, and the zero entries of \mathbf{x} using very large weights, simply forces the solution $\mathbf{x}_{\text{wlasso}}$ to concentrate on the elements where \mathbf{x} is nonzero. Since \mathbf{x} is not known, a possible choice, among others, is to use the least squares estimator of \mathbf{x} , $\mathbf{x}_{\text{ls}} = [x_{\text{ls},1} \ x_{\text{ls},2} \ \dots \ x_{\text{ls},p}]^T$, as proposed in [6]. For $\gamma \geq 0$, the weights w_i can be selected as

$$w_i = \frac{1}{|x_{\text{ls},i}|^\gamma}, \quad i = 1, 2, \dots, p. \quad (10)$$

Note that (9) is a quadratic programming problem and can be solved using iterative techniques, such as interior point methods. However, these methods are computationally expensive, rendering their use prohibitive for demanding real-time applications, such as the hyperspectral unmixing problem considered here.

3.1. Enforcing the constraints

As mentioned in Section 2, the sum-to-one and nonnegativity constraints are physically imposed to the hyperspectral unmixing problem. As far as the sum-to-one constraint is concerned, we propose to incorporate it directly to the optimization problem, by substituting the ℓ_2 norm of eq. (9), with the ℓ_2 norm of an augmented problem, as shown below:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{S}} \left\{ \left\| \begin{bmatrix} \mathbf{y} \\ \lambda_\beta \end{bmatrix} - \begin{bmatrix} \Phi \\ \lambda_\beta \mathbf{1}^T \end{bmatrix} \mathbf{x} \right\|_2^2 + \lambda_\alpha \sum_{i=1}^p w_i |x_i| \right\}, \quad (11)$$

where λ_β is a regularization parameter [10]. Geometrically, parameter λ_β determines how close the solution will lie to the hyperplane of the constraint $\mathbf{1}^T \mathbf{x} = 1$. Note that in practice, the additivity constraint is enforced for large values of λ_β .

In addition, the nonnegativity constraint can be imposed using a modified version of the LARS algorithm, [8], as shown in Algorithm 1 presented below.

Algorithm 1. The positivity constrained lasso.

Let \mathcal{A} denote a subset of the indices $\{1, 2, \dots, p\}$, and let $\hat{\boldsymbol{\mu}}_{\mathcal{A}} = \Phi \hat{\mathbf{x}}_{\mathcal{A}}$ be a prediction vector of the regression coefficients $\hat{\mathbf{x}}_{\mathcal{A}}$.

1. Initialize $\hat{\mathbf{x}}_{\mathcal{A}} = 0$, $\hat{\boldsymbol{\mu}}_{\mathcal{A}} = 0$ and $\mathcal{A} = \emptyset$.
2. Compute the vector of current correlations $\hat{\mathbf{c}} = \Phi^T (\mathbf{y} - \hat{\boldsymbol{\mu}}_{\mathcal{A}})$.
3. Select j with the highest correlation, $\hat{C} = \max_j \{\hat{c}_j\}$ and set $\mathcal{A} = \mathcal{A} \cup \{j : \hat{c}_j = \hat{C}\}$.
4. Compute:

$$\Phi_{\mathcal{A}} = (\dots \phi_j \dots)_{j \in \mathcal{A}}$$

$$\mathbf{G}_{\mathcal{A}} = \Phi_{\mathcal{A}}^T \Phi_{\mathcal{A}}$$

$$\mathbf{A}_{\mathcal{A}} = (\mathbf{1}^T \mathbf{G}_{\mathcal{A}}^{-1} \mathbf{1})^{-\frac{1}{2}}$$

$$\mathbf{u}_{\mathcal{A}} = \Phi_{\mathcal{A}} \mathbf{w}_{\mathcal{A}}, \text{ where } \mathbf{w}_{\mathcal{A}} = \mathbf{A}_{\mathcal{A}} \mathbf{G}_{\mathcal{A}}^{-1} \mathbf{1} \text{ and}$$

$$\boldsymbol{\alpha} \equiv [\alpha_1, \alpha_2, \dots, \alpha_p]^T = \Phi^T \mathbf{u}_{\mathcal{A}}.$$
5. Define the p -dimensional vector \mathbf{d} , whose A_j -th entry equals w_j for $j = 1, 2, \dots, |\mathcal{A}|$, while all other entries are zero (\mathcal{A}_j denotes the j -th element while $|\mathcal{A}|$ the cardinality of \mathcal{A}).
6. (a) Set $\tilde{\gamma} = \min_{j \in \mathcal{A}} \{\max\{-\hat{x}_{\mathcal{A}_j}/d_j, 0\}\}$ and let \tilde{j} denote the position where the minimum occurs.
 (b) Set $\hat{\gamma} = \min_{j \in \mathcal{A}^C} \{\max\{\frac{\hat{C} - \hat{c}_j}{A_{\mathcal{A}} - \alpha_j}, 0\}\}$, where \mathcal{A}^C denotes the complementary set of \mathcal{A} .
7. If $\tilde{\gamma} < \hat{\gamma}$ stop the ongoing LARS step and remove \tilde{j} from \mathcal{A} , i.e. $\hat{\mathbf{x}}_{\mathcal{A}} = \hat{\mathbf{x}}_{\mathcal{A}} + \hat{\gamma} \mathbf{d}$, $\hat{\boldsymbol{\mu}}_{\mathcal{A}} = \hat{\boldsymbol{\mu}}_{\mathcal{A}} + \hat{\gamma} \mathbf{u}_{\mathcal{A}}$ and $\mathcal{A} = \mathcal{A} - \{\tilde{j}\}$.
8. Update $\hat{\mathbf{x}}_{\mathcal{A}} = \hat{\mathbf{x}}_{\mathcal{A}} + \hat{\gamma} \mathbf{d}$ and $\hat{\boldsymbol{\mu}}_{\mathcal{A}} = \hat{\boldsymbol{\mu}}_{\mathcal{A}} + \hat{\gamma} \mathbf{u}_{\mathcal{A}}$.
9. Return to step 2, until a stopping criterion is met.

Algorithm 1 builds up estimates $\hat{\boldsymbol{\mu}}_{\mathcal{A}}$ of \mathbf{y} in successive steps. As a result, the ℓ_1 norm of the current solution $\hat{\mathbf{x}}_{\mathcal{A}}$ increases in each iteration. The algorithm stops when $\|\hat{\mathbf{x}}_{\mathcal{A}}\|_1$ exceeds 1, as implied by the sum-to-one constraint.

The proposed method is summarized below:

Algorithm 2.

1. Select $\gamma \geq 0$ and calculate \mathbf{w} as in (10).
2. Define $\phi_j^* = \phi_j/w_j$, $j = 1, 2, \dots, p$.
3. Select the parameters λ_α and λ_β .
4. Using Algorithm 1, solve the augmented lasso problem

$$\hat{\mathbf{x}}^* = \arg \min_{\mathbf{x}} \left\{ \left\| \begin{bmatrix} \mathbf{y} \\ \lambda_\beta \end{bmatrix} - \sum_{i=1}^p \begin{bmatrix} \phi_i^* \\ \lambda_\beta \end{bmatrix} x_i \right\|_2^2 + \lambda_\alpha \sum_{i=1}^p w_i |x_i| \right\}.$$

5. Output $\hat{x}_j = \hat{x}_j^*/w_j$, $j = 1, 2, \dots, p$

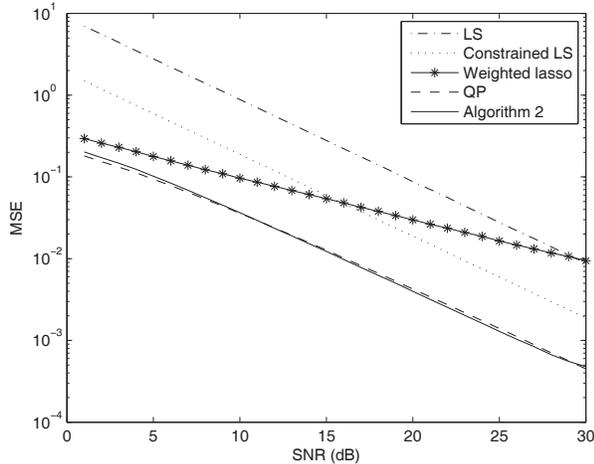


Fig. 1. Abundance estimation MSE curves versus SNR.

Tuning the parameters γ , λ_α and λ_β of Algorithm 2 is critical. As already mentioned, λ_β is given a large value (e.g. $\lambda_\beta = 1000$ in our experiments) so as to enforce the additivity constraint. Though not obvious, λ_α controls the termination of Algorithm 1. This can be seen if the conventional lasso is stated as

$$\min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \quad \text{subject to} \quad \sum_{i=1}^p |x_i| \leq T. \quad (12)$$

Stopping Algorithm 1 when $\|\hat{\mathbf{x}}_{\mathcal{A}}\|_1 > 1$ sets the parameter T . Parameter γ reflects our confidence on the least squares estimator \mathbf{x}_{ls} . For instance, in high SNRs, \mathbf{x}_{ls} detects correctly the non-zero elements of \mathbf{x} and γ can be set equal to one. On the other hand, when \mathbf{x}_{ls} fails to detect the nonzero elements of \mathbf{x} , γ can be set to zero.

4. SIMULATION RESULTS

4.1. Simulated data

To test the performance of the proposed method, we simulated a hyperspectral image consisting of 10^5 pixels. Ten endmembers were selected from the USGS spectral library [11], having 474 spectral bands. The USGS library contains spectral signatures of various materials, such as minerals, plants, man-made materials, etc. All pixels were produced using exactly three out of ten endmembers. The abundances follow a Dirichlet(1, 1, ..., 1) distribution and the additive noise is Gaussian, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, where σ^2 determines the SNR level. The proposed method is compared to the least squares algorithm (LS), [10], the equality constrained least squares algorithm (CLS), [10], the weighted lasso without enforcing the additivity constraint, (9), and the quadratic programming based technique (QP) of [12]. The simulation results are shown in Fig. 1, where the mean squared error (MSE) curves versus the SNR are displayed. In terms of performance, we observe that the proposed method is equivalent to the QP method. It should be noted, however, that such an equivalence in performance is achieved with much lower computational complexity. Moreover, it is easily seen from Fig. 1 that solving the weighted lasso problem without enforcing the sum-to-one constraint as in (11) results in significant performance degradation.



Fig. 2. Real hyperspectral data: HYDICE urban image scene, band 80.

4.2. Unmixing of a HYDICE image

In this section we apply the proposed method on a real hyperspectral image, which depicts an urban image scene, shown in Fig. 2, and has been collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor. The image is composed of 210 spectral bands, in the region from 400 to 2500nm, with a spectral resolution of 10nm. After removing the low SNR bands, 162 spectral bands remain available, i.e. $L = 162$. Six endmembers are present in the image, namely asphalt, metal, dirt, roof, grass, and tree. The spectral signatures of these endmembers have been identified using a supervised technique, which utilizes areas in the image that seem to contain pure materials, in order to extract the spectrum of each material [1]. The unmixing results are shown in Fig. 3. Due to space limitations, we show the unmixing results of only three of the available six endmembers, for each of the three algorithms, (a) LS, (b) QP and (c) the proposed method. A pure black pixel in an image indicates that the abundance of the respective endmember is zero, while a pure white pixel represents an abundance value equal to one. All other abundance values between zero and one are illustrated according to the different tones of gray. Again, we observe that the unmixing results of the proposed method are in agreement with the results produced by the quadratic programming technique, while both methods are clearly superior compared to the LS algorithm.

5. REFERENCES

- [1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, pp. 44–57, January 2002.
- [2] D. C. Heinz and C. I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 529–545, March 2001.
- [3] M. Winter, "N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Imaging Spectrometry V*. SPIE, 1999, vol. 3753, pp. 266–275.

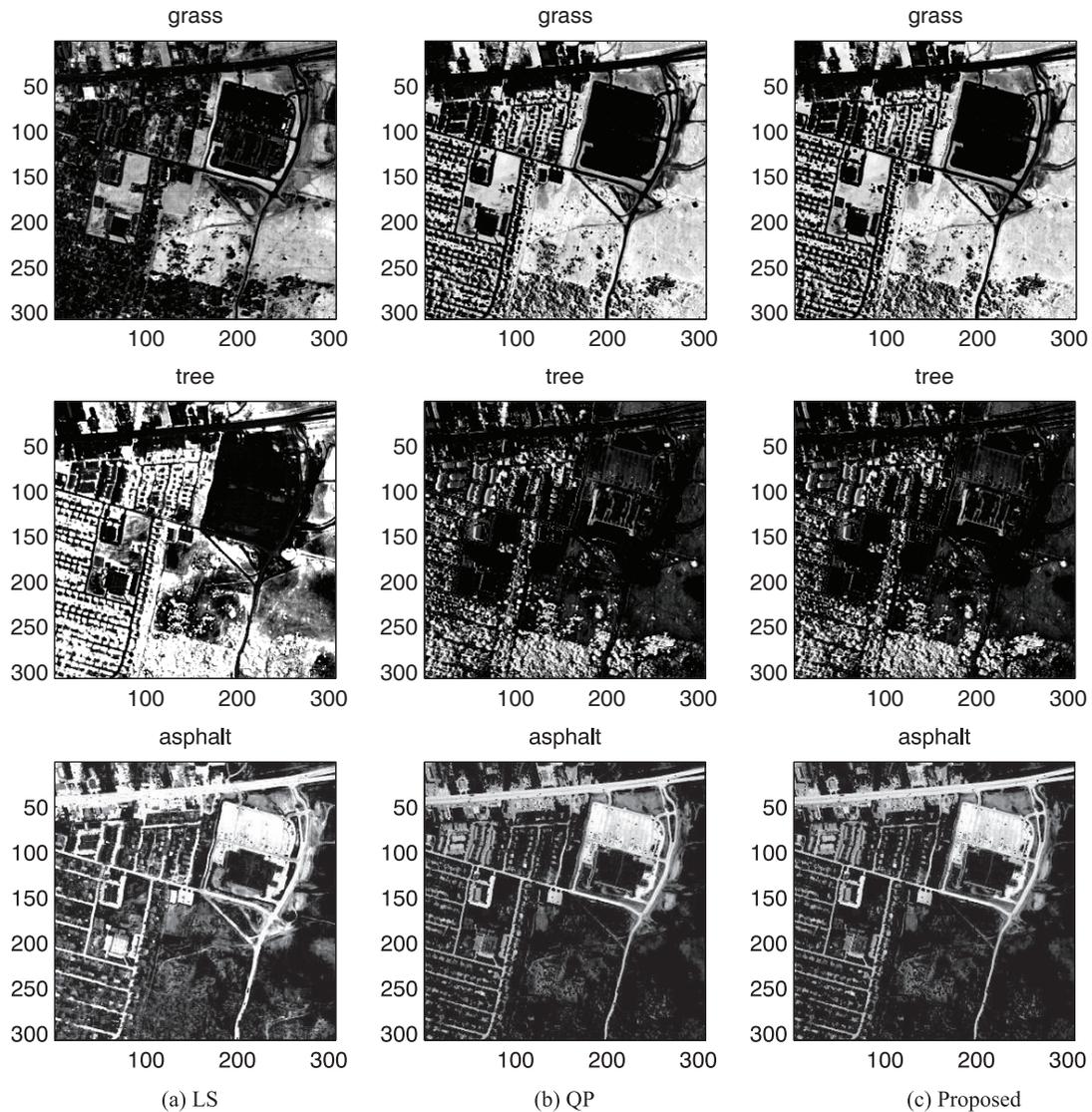


Fig. 3. Fraction maps estimated by (a) LS, (b) QP and (c) the proposed method.

- [4] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 4, pp. 898–910, April 2005.
- [5] N. Dobigeon, J.-Y. Tourneret, and C.-I. Chang, "Semi-supervised linear spectral unmixing using a hierarchical bayesian model for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 56, pp. 2684–2695, July 2008.
- [6] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, pp. 1418–1429, December 2006.
- [7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal. Statist. Soc B*, vol. 58, pp. 267–288, January 1996.
- [8] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, pp. 407–499, February 2002.
- [9] Z. Guo, T. Wittman, and S. Osher, "L1 unmixing and its application to hyperspectral image enhancement," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery*. SPIE, 2009, vol. 7334, pp. 73341M–73341M–9.
- [10] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1996.
- [11] R. N. Clark, G. A. Swayze, R. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly, and S. J. Sutley, "USGS digital spectral library," 2007, <http://speclab.cr.usgs.gov/spectral.lib06/ds231/datatable.html>.
- [12] T. F. Coleman and Y. Li, "A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM Journal on Optimization*, vol. 6, pp. 1040–1058, 1996.