



HAL
open science

A regularized sparse approximation method for hyperspectral image classification

Leila Belmerhnia, El-Hadi Djermoune, David Brie, Cédric Carteret

► To cite this version:

Leila Belmerhnia, El-Hadi Djermoune, David Brie, Cédric Carteret. A regularized sparse approximation method for hyperspectral image classification. 19th IEEE Workshop on Statistical Signal Processing, SSP 2016, Jun 2016, Palma de Majorque, Spain. hal-01310059v2

HAL Id: hal-01310059

<https://hal.science/hal-01310059v2>

Submitted on 20 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A REGULARIZED SPARSE APPROXIMATION METHOD FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Leïla Belmerhnia, El-Hadi Djermoune, David Brie

Cédric Carteret

CRAN, Université de Lorraine, CNRS
Boulevard des Aiguillettes, BP 70239,
54506, Vandœuvre, France

LCPME, Université de Lorraine, CNRS
405 rue de Vandœuvre,
54500, Villers-lès-Nancy, France

ABSTRACT

This paper presents a new technique for hyperspectral images classification based on simultaneous sparse approximation. The proposed approach consists in formulating the problem as a convex multi-objective optimization problem which incorporates a term favoring the simultaneous sparsity of the estimated coefficients and a term enforcing a regularity constraint along the rows of the coefficient matrix. We show that the optimization problem can be solved efficiently using FISTA (Fast Iterative Shrinkage-Thresholding Algorithm). This approach is applied to a wood wastes classification problem using NIR hyperspectral images.

Index Terms— Hyperspectral image classification, simultaneous sparse approximation, regularization constraint, FISTA

1. INTRODUCTION

The sparse representation has been widely used in different areas such as compression [1, 2], sparse modeling of signals and images [3] and pattern recognition [4]. Recently, this framework has been also introduced in hyperspectral unmixing and spectral analysis [5], classification problems [6] and hyperspectral image classification [7]. It consists in representing a signal using a minimum number of vectors from an overcomplete dictionary. Structured, group and simultaneous sparse models are extensions allowing to enforce certain structural constraints. For instance, in the simultaneous case, one seeks to reconstruct an observations matrix with elementary signals sharing the same sparsity profile. In the context of classification, Turlach *et al.* [8] investigated a sparse representation for simultaneous variables selection using the Lasso algorithm. In [9], the generalized fused Lasso algorithm is used for diagnosis of Alzheimer’s disease established as a classification task.

The present work is motivated by NIR hyperspectral image classification of wood wastes. The goal is to determine

two categories, namely: recyclable and not recyclable (polluted). Wood wastes may contain various wood species with different compositions and sources. Moreover, their spectra share several common characteristics. In fact, only a few imperceptible peaks in the spectra can be used as distinctive features, thus making good discrimination difficult to achieve. Hence, a variables selection stage is necessary to select, from the whole spectrum, the wavelengths that result in a maximum accuracy of spectra reconstruction while discriminating the wood samples at best.

In our recent work [10], we proposed a features selection algorithm based on the ℓ_0 norm, for classification issue. This algorithm performs in two stages. First, using a greedy algorithm, it computes a simultaneous sparse decomposition of the observation matrix. Then, assuming that the columns of this matrix have a meaningful ordering (i.e the spectra corresponding to the same group of wood appear in adjacent columns), it applies a regularization constraint using ℓ_1 norm along the coefficients matrix rows. In contrast, the approach proposed in the present work deals with the sparsity constraint and regularization jointly by substituting the ℓ_0 norm by the ℓ_1 norm. This relaxation enforces the selection of the most regular variables for the training phase. Moreover, we exploit spatial informations of hyperspectral images of wood wastes acquired according to the scheme depicted in Figure 1. In such equipment, wood wastes samples are put on a conveyor which moves along the y dimension. The spectra are captured line by line, thus defining the x direction. This type of application motivates the regularization applied along the rows (corresponding to the x direction) of the hyperspectral image.

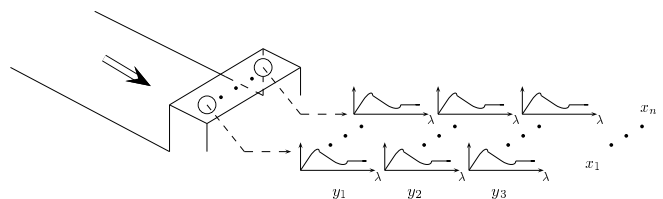


Fig. 1. Hyperspectral images acquired by an industrial NIR spectro-imager

In this context, it has been shown that combining spectral and spatial information improves the classification performances of hyperspectral data (detailed overview of these techniques is provided in [11]).

The paper is organized as follows: Section 2 introduces the general framework including the formulation of the simultaneous sparse fused Lasso problem and resolution scheme using FISTA algorithm. The experiments conducted on real hyperspectral images of wood wastes and classification results obtained using support vector machines (SVM) are presented in 3. Section 4 concludes the paper.

2. FUSED SPARSE LASSO ALGORITHM (FSLA)

2.1. Problem statement

Given an observation matrix $\mathbf{Y} \in \mathbb{C}^{M \times K}$ and a dictionary $\Phi \in \mathbb{C}^{M \times N}$ ($M < N$), the simultaneous sparse approximation problem consists in finding the solution \mathbf{X} which minimizes the following cost function:

$$J_0(\mathbf{X}) = \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{X}\|_F^2 \text{ s.t. } \|\mathbf{X}\|_{0,2} \leq s \quad (1)$$

where s is the sparsity parameter. The coefficient matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K]$ is a sparse matrix where \mathbf{x}_i is the i -th column of \mathbf{X} . The support of \mathbf{X} corresponds to the set of active atoms Γ such that: $\text{supp}(\mathbf{X}) = \{1 \leq i \leq N \mid \mathbf{x}^i \neq \mathbf{0}\} := \|\mathbf{X}\|_{0,2}$, where \mathbf{x}^i stands for the transpose of the i -th row of \mathbf{X} . Here, we propose to relax the mixed ℓ_0/ℓ_2 constraint into the tractable ℓ_1/ℓ_1 norm¹. Moreover, we include a term promoting the reconstruction of piecewise constant rows in the coefficient matrix. So the new criterion is written as:

$$J_1(\mathbf{X}; \lambda_1, \lambda_2) = \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{X}\|_F^2 + \lambda_1 \|\mathbf{X}\|_{1,1} + \lambda_2 \|\mathcal{D}\mathbf{X}\|_{1,1} \quad (2)$$

where $\lambda_1, \lambda_2 \geq 0$, $\|\mathbf{X}\|_{1,1}$ is the mixed ℓ_1/ℓ_1 -norm of \mathbf{X} and \mathcal{D} is a finite difference operator applied along the rows of \mathbf{X} :

$$\|\mathcal{D}\mathbf{X}\|_{1,1} := \sum_{i=1}^N \sum_{j=1}^{K-1} |\mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}|. \quad (3)$$

Thus we seek to find \mathbf{X} that minimizes (2). This problem can be seen as an extension of the 1D sparse fused Lasso problem to the simultaneous case. The single measurement problem was first studied and solved by Tibshirani in [12]. Here, we propose to develop the multiple measurement version which we solve using appropriate transformations combined to the proximal gradient method FISTA [13].

¹In fact, the ℓ_1/ℓ_2 mixed norm is more appropriate to preserve the simultaneous character of the problem. We already investigated this alternative. The resulting algorithm will be presented in a future work.

2.2. Resolution scheme

The criterion (2) is rewritten as:

$$J_1(\mathbf{X}; \lambda_1, \lambda_2) = f(\mathbf{X}) + g(\mathbf{X}) + h(\mathbf{X}) \quad (4)$$

where

$$f(\mathbf{X}) = \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{X}\|_F^2 \quad (5)$$

$$g(\mathbf{X}) = \lambda_1 \|\mathbf{X}\|_{1,1} \quad (6)$$

$$h(\mathbf{X}) = \lambda_2 \|\mathcal{D}\mathbf{X}\|_{1,1} \quad (7)$$

Equation (5) is the least-squares error in the cost function. The non-smooth terms in (6) and (7) correspond to the sparsity and the fusion penalties, respectively. Due to these penalties, solutions of (4) tend to be sparse and piecewise smooth, i.e. two consecutive columns of \mathbf{X} are similar.

Let $\mathbf{x} = \text{vec}(\mathbf{X}^T)$ and $\mathbf{y} = \text{vec}(\mathbf{Y}^T)$. Then, we have:

$$f(\mathbf{X}) = \frac{1}{2} \|\text{vec}(\mathbf{Y}^T) - \text{vec}(\mathbf{X}^T \Phi^T)\|_2^2 \quad (8)$$

$$= \frac{1}{2} \|\mathbf{y} - (\Phi \otimes \mathbf{I}_K) \mathbf{x}\|_2^2. \quad (9)$$

where \otimes stands for the Kronecker product. Using similar tricks, we can write:

$$g(\mathbf{X}) = \lambda_1 \|\mathbf{x}\|_1 \quad (10)$$

$$h(\mathbf{X}) = \lambda_2 \|(\mathbf{I}_N \otimes \mathbf{D}) \mathbf{x}\|_1. \quad (11)$$

where \mathbf{D} is a matrix of finite differences of order 1 and dimension $(K-1) \times K$. Finally, the cost function may be rewritten in the following vector form:

$$J'_1(\mathbf{x}; \lambda_1, \lambda_2) = \frac{1}{2} \|\mathbf{y} - \mathbf{A} \mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{F} \mathbf{x}\|_1 \quad (12)$$

with $\mathbf{A} = \Phi \otimes \mathbf{I}_K$ and $\mathbf{F} = \mathbf{I}_N \otimes \mathbf{D}$.

2.3. FISTA algorithm solution

FISTA is an extension of the Nesterov's gradient-based method ISTA used to solve problems including smooth and non-smooth terms. The advantage of FISTA is its global convergence rate which is proven to be significantly better [13]. Denoting by k the FISTA iteration, the update of \mathbf{x} is based on the proximal regularization of the objective function at the previous iteration $\mathbf{x}_{(k)}$. So, considering the non-smooth regularization term:

$$\Omega(\mathbf{x}) = \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{F} \mathbf{x}\|_1 \quad (13)$$

the update of vector $\mathbf{x}_{(k)}$ is:

$$\mathbf{x}_{(k+1)} = \arg \min_{\mathbf{x}} \left(\Omega(\mathbf{x}) + \frac{L}{2} \|\mathbf{x} - \mathbf{v}_{(k)}\|_2^2 \right) \quad (14)$$

where $\mathbf{v}_{(k)} = \mathbf{x}_{(k)} - \frac{1}{L} \nabla f(\mathbf{x}_{(k)})$, $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ and $\nabla f(\mathbf{x})$ is the gradient of $f(\mathbf{x})$ with Lipschitz constant $L > 0$. The optimization problem in (14) can be rewritten as follow:

$$\mathbf{x}_{(k+1)} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{v}_{(k)}\|_2^2 + \frac{\lambda_1}{L} \|\mathbf{x}\|_1 + \frac{\lambda_2}{L} \|\mathbf{F}\mathbf{x}\|_1 \quad (15)$$

This problem is similar to the fused Lasso signal approximator (FLSA) problem [14]. The difference lies in the structure of the fused term. Indeed, in (15), the matrix \mathbf{F} imposes no regularity between the last and the first elements of two successive rows in \mathbf{X} . In fact, the expression in (15) can be seen as a particular case of the generalized fused Lasso problem [9]. Besides, an in-depth examination of (15) reveals that the problem is decomposable and, thus, may be solved for each row i of the coefficient matrix \mathbf{X} :

$$\begin{cases} \mathbf{x}_{(k+1)}^i = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{v}_{(k)}^i\|_2^2 + \frac{\lambda_1}{L} \|\mathbf{x}\|_1 + \frac{\lambda_2}{L} \|\mathbf{D}\mathbf{x}\|_1 \\ \mathbf{V}_{(k)} = \mathbf{X}_{(k)} + \frac{1}{L} \Phi^T (\mathbf{Y} - \Phi \mathbf{X}_{(k)}). \end{cases} \quad (16)$$

for $i = 1, \dots, N$. Consequently, the optimization cost is significantly reduced allowing its use for high dimensional problems. Moreover, the matrices $\Phi^T \mathbf{Y}$ and $\Phi^T \Phi$ may be computed once and stored, which also helps to reduce memory requirements. For each i , the problem is solved using the FLSA routine implemented in SLEP package².

3. CLASSIFICATION OF HYPERSPECTRAL IMAGES OF WOOD WASTES

3.1. Data processing

Different wood samples are labeled by experts according to their groups. We distinguish two categories of wood wastes: recyclable and not recyclable as presented in Table 1. First, we extract the spectra corresponding to these samples using a Nicolet 8700 FTIR spectrometer continuously purged with ultrapure N₂. The NIR reflectance spectra are recorded in the range 3000-10000 cm⁻¹. A gold mirror was used as reference and the spectral resolution was 16 cm⁻¹ with 100 scans co-added for each spectrum. Then, the spectra are subsampled by a factor 4 and we proceed to data pre-processing including baseline removal and normalization. As a result, the 419 spectra from category 1 and 2 are grouped in the columns of the data matrix $\mathbf{Y} \in \mathbb{R}^{412 \times 419}$.

3.2. Variables selection

Variables selection is performed using equation (16). The dictionary Φ is composed of normalized Gaussian atoms. Each gaussian has a width σ varying from 31 to 600 with step 31 cm⁻¹. Their locations (centers) start at 3405 cm⁻¹ and

Table 1. Composition of the two categories of wood wastes

| Category 1: To recyclable | | Category 2: To reject | |
|---------------------------|-------------------------|-----------------------|------------------------|
| Groups | Name | Groups | Name |
| 1.1 | raw wood | 2.1 | MDF-HDF |
| 1.2 | painted solid wood | 2.2 | painted MDF-HDF |
| 1.3 | vanished solid wood | 2.3 | surfaced MDF-HDF |
| 1.4 | vanished plywood | 2.4 | raw fiber board |
| 1.5 | raw particle board | 2.5 | surfaced fiber board |
| 1.6 | surfaced particle board | 2.6 | solid wood metal salts |
| 1.7 | painted particle board | - | - |
| 1.8 | raw plywood | - | - |
| 1.9 | surfaced plywood | - | - |

are separated by σ . We tested several values for both sparsity and regularization parameters. A good reconstruction of the spectra with the desired sparsity level is obtained for $\lambda_1 = \lambda_2 = 100$. This configuration produces 48 variables (i.e set of active atoms Γ) that are presented in Figure 2. Some of these variables exhibit low dispersion of intensity. They often correspond to the most discriminative variables. The proposed approach leads to a good approximation of the spectra. For instance, the reconstruction of two spectra from the two categories are presented in Figure 3. Finally, the training is performed on the re-estimated coefficients (to reduce the bias introduced by the ℓ_1 penalties) of the selected subset of variables using the dictionary of active atoms Φ_Γ and $\lambda_1 = \lambda_2 = 0$. Hence, training model is obtained using quadratic kernel SVM by learning the separating hyperplane.

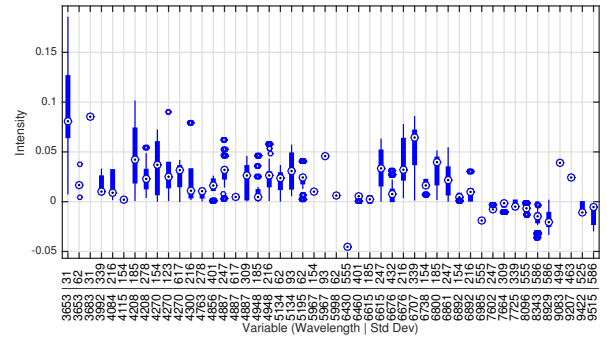


Fig. 2. Intensity dispersion of the 48 variables obtained using the proposed FLSA

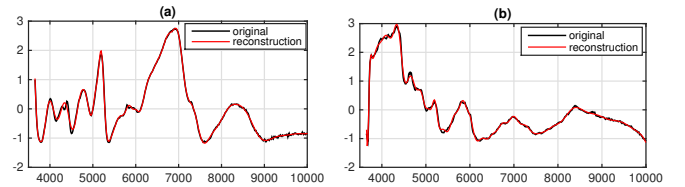


Fig. 3. Reconstruction of spectra from the two categories using the selected coefficients. (a) Category 1; (b) Category 2

²<http://yelab.net/software/SLEP/>

Table 2. Classification error rates for four different noise levels

| Reg. parameter | SNR = -5 dB | | SNR = 0 dB | | SNR = 5 dB | | SNR = 10 dB | |
|-----------------|---------------|---------------|--------------|---------------|--------------|---------------|--------------|---------------|
| | Type I error | Type II error | Type I error | Type II error | Type I error | Type II error | Type I error | Type II error |
| $\lambda_2 = 0$ | 34.73% | 35.40% | 20.37% | 25.58% | 6.27% | 19.37% | 0.61% | 8.71% |
| $\lambda_2 = 1$ | 24.37% | 25.39% | 8.88% | 13.62% | 0.52% | 3.99% | 0.35% | 0.09% |
| $\lambda_2 = 2$ | 18.54% | 24.37% | 5.48% | 9.36% | 3.05% | 1.85% | 0.35% | 0.00% |
| $\lambda_2 = 3$ | 17.67% | 18.87% | 4.26% | 3.89% | 1.22% | 0.37% | 0.61% | 0.00% |

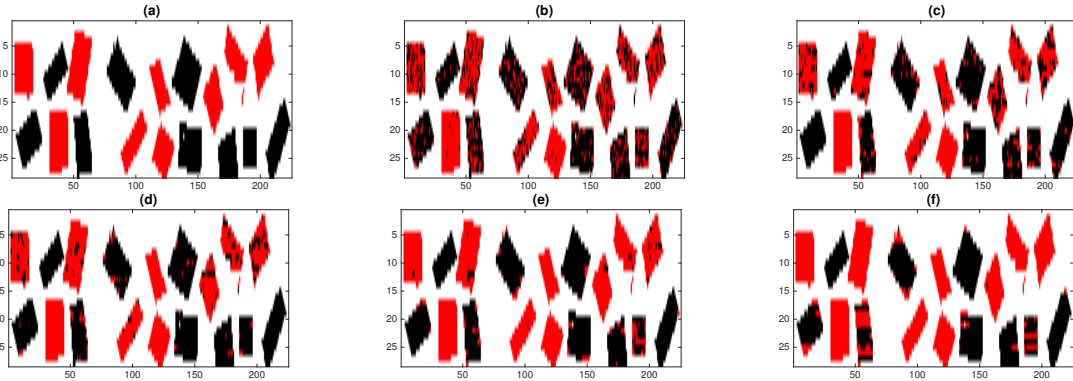


Fig. 4. Image reconstruction results for SNR = 0 dB. Black and red colors for category 1 and 2, respectively. (a) Ground truth; (b) $\lambda_2 = 0$; (c) $\lambda_2 = 1$; (d) $\lambda_2 = 2$; (e) $\lambda_2 = 3$; (f) $\lambda_2 = 10$

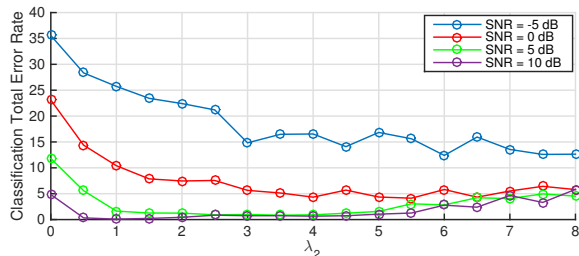


Fig. 5. Total classification error rate (in %) versus λ_2 for different values of SNR

3.3. Hyperspectral image classification

The acquisition process of hyperspectral images is composed of a conveyor on which the wood wastes samples progress at a speed of 3 m/s, a focused illumination system and an acquisition cabin which extracts the NIR spectra of samples row by row at a resolution of 12 mm. The experiments presented here consist in simulating this type of applications using *real-world* wood spectra. We start with an image of resolution 28×225 pixels representing wood samples, as depicted in Figure 4(a). The three-dimensional hyperspectral data cube \mathcal{S} is then generated by affecting to each active pixel (black or red) a real spectrum from our database. As a result, the size of the data cube is $28 \times 225 \times 412$. Finally, \mathcal{S} is perturbed by an additive white Gaussian noise \mathcal{N} such that the SNR defined by $\text{SNR} := \|\mathcal{S}\|_F^2 / \|\mathcal{N}\|_F^2$ is varying between -5 and 10 dB.

The proposed algorithm is applied to the noisy hyperspec-

tral cube slice by slice³. The aim is to study how the spatial regularization influences the classifier performances. Thus, the coefficients are estimated using the dictionary Φ_T with parameters $\lambda_1 = 0$ (the variables are already selected) and λ_2 varying from 0 to 8 for each slice successively. Thereafter, the classification is carried out on the resulting decomposition using the training model. Classification error rates are reported in Table 2 for three values of SNR. For each test we compute the true negative rate (Type I error) and the false positive rate (Type II error).

The classification error rates obtained confirm the importance of using joint spectral and spatial regularization. Indeed, for $\text{SNR} \geq 0$ dB, the classification error is less than 5% when $\lambda_2 = 3$. At low SNR, the value of the regularization parameter λ_2 has to be increased in order to achieve a minimum classification error as shown in Figure 5. We note in particular that the total classification error decreases significantly for the four tests for λ_2 between 0.5 and 5.5. The images obtained from the labeled coefficients are presented in Figure 4 for some values of λ_2 . When no regularization is applied ($\lambda_2 = 0$), we observe the salt-and-pepper appearance resulting from classification errors. Here, the best reconstruction is obtained for $\lambda_2 = 3$.

4. CONCLUSION

In this paper, we presented a sparse simultaneous fused Lasso algorithm for hyperspectral image classification. It is based on the minimization of a multi-objective convex cost function including, in particular, a sparse and a piecewise constant

³The size of each slice is 412×225 .

regularization constraints. Using a FISTA iteration, we have shown that the criterion may be solved efficiently thanks to the fused Lasso signal approximator (FLSA), applied on each row of the coefficient matrix. The resulting algorithm has a low computational cost suitable for large-scale problems. The proposed approach is used for wood wastes NIR spectra variables selection. Then, spatial information of hyperspectral images is exploited to increase classification performances. Results in terms of classification error rates show the effectiveness of the proposed algorithm.

5. REFERENCES

- [1] D.L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E.J. Candès, J. Romberg, and T. Tao, "Stable signal recovery for incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, May 2006.
- [3] A.M Bruckstein, D.L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM review*, vol. 51, no. 1, pp. 34–81, 2009.
- [4] J. Wright, A.Y Yang, A. Ganesh, S.S Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [5] M.D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 2014–2039, June 2011.
- [6] J. Kim and H. Park, "Sparse nonnegative matrix factorization for clustering," *Technical Report*, 2008.
- [7] Y. Chen, N.M. Nasrabadi, and T.D Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3973–3985, 2011.
- [8] B. A Turlach, W.N Venables, and S.J Wright, "Simultaneous variable selection," *Technometrics*, vol. 47, no. 3, pp. 349–363, 2005.
- [9] B. Xin, Y. Kawahara, Y. Wang, and W. Gao, "Efficient generalized fused lasso and its application to the diagnosis of Alzheimer's disease," in *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014, pp. 2163–2169.
- [10] L. Belmerhnia, E-H. Djermoune, C. Carteret, and D. Brie, "Simultaneous regularized sparse approximation for wood wastes nir spectra features selection," in *International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMPSAP 2015*, 2015.
- [11] M. Fauvel, Y. Tarabalka, J.A. Benediktsson, J. Chanussot, and J.C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.
- [12] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 1, pp. 91–108, 2005.
- [13] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [14] H. Hoefling, "A path algorithm for the fused lasso signal approximator," *Journal of Computational and Graphical Statistics*, vol. 19, no. 4, pp. 984–1006, 2010.