



HAL
open science

Simultaneous regularized sparse approximation for wood wastes NIR spectra features selection

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

► **To cite this version:**

Leila Belmerhnia, El-Hadi Djermoune, Cédric Carteret, David Brie. Simultaneous regularized sparse approximation for wood wastes NIR spectra features selection. International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMPSAP 2015, Dec 2015, Cancun, Mexico. hal-01241851

HAL Id: hal-01241851

<https://hal.science/hal-01241851>

Submitted on 11 Dec 2015

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.

Simultaneous Regularized Sparse Approximation for Wood Wastes NIR Spectra Features Selection

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

*Centre de Recherche en Automatique de Nancy, Université de Lorraine, CNRS
Boulevard des Aiguillettes, BP 70239, 54500, Vandœuvre France

†Laboratoire Chimie Physique et Microbiologie pour l'Environnement, Université de Lorraine, CNRS
405 rue de Vandœuvre, 54500, Villers-lès-Nancy France
Email: firstname.lastname@univ-lorraine.fr

Abstract—This paper presents a new technique of simultaneous sparse approximation incorporating a regularity constraint along the coefficients matrix rows. This approach is decomposed in two steps: first a sparse representation of the coefficients matrix is obtained using a simultaneous greedy method. Then, a ℓ_1 penalty regularization on the derivative of nonzero coefficients enforces a piecewise constant variation along the rows of the solution. The regularization problem is solved efficiently using the ADMM (Alternate Direction Method of Multipliers) optimization method. The approach is applied on near-infrared spectrometry dataset of wood wastes. This allows to select among the 1647 wavelengths of the spectra those suitable for classification. The experimental tests validate the advantages of regularization in terms of classification rates.

I. INTRODUCTION

Sparse approximation is one of the most studied linear inversion problems in different application areas: compression [1], [2], spectral analysis [3], regression [4] and classification [5]. It consists in representing a signal using a minimum number of vectors from an overdetermined dictionary. Unlike the standard sparse approximation, the simultaneous version seeks to reconstruct an observations matrix \mathbf{Y} , with elementary signals \mathbf{y}_i sharing the same sparsity profil. This approach was originally proposed by Cotter et al. [6] developing the simultaneous versions of Matching Pursuit (MP) and FOCUSS algorithms, namely the M-FOCUSS. They also demonstrated that reconstruction rates using simultaneous sparse approximation are significantly higher than those obtained with the standard version. As has been pointed out by Tropp et al. [7], this approach corresponds to a convex relaxation of the exact simultaneous sparse approximation problem which seeks to find the solution with the minimum number of nonzero rows. This is a NP-hard problem for which the greedy methods provide a good compromise between efficiency and computational cost [8], [9]. Conditions of exact recovery in this case are given in [10] and [11]. Other approaches formulate the problem in the Bayesian framework [12], [13]. More recently, works have focused on the development of simultaneous sparse approximation methods that consider a correlation between the coefficients of the solution [14], [15], [16]. In particular, [17] proposes an approach to impose a smoothness constraint along the solution rows. In this paper we propose to develop a simultaneous sparse method which enforces a regularity constraint on the rows of the coefficients matrix. This work is

motivated by IR spectroscopy data classification application. Indeed, in the near-IR spectral range (2.5-1 μm) the molecular overtone and combination bands are typically very broad, leading to complex spectra. As it can be difficult to assign specific features to specific chemical components, multivariate statistical methods are often employed to extract the desired chemical information. Selecting from the full spectrum the wavelengths that result in the maximum accuracy is still a challenging task, mainly when spectra strongly overlap and have imperceptible distinctive features, as is the case with NIR spectra of wood samples. To overcome this, several approaches have been proposed to select optimal sets of variables for multivariate calibration or classification [18], [19]. The advantage of the proposed approach is twofold: it allows to exploit the sparse representation to reduce the problem dimension and the form of variation (piecewise constant) is suitable to classification problems.

The paper is organized as follows: The problem formulation of sparse approximation including a regularity constraint on the coefficient matrix is presented in II section. In section III we develop the proposed approach to solve the problem. The classification results of wood wastes NIR spectra are presented in IV. Section V concludes the paper.

II. PROBLEM FORMULATION

The simultaneous sparse approximation problem consists in finding the solution of the following cost function:

$$J_0(\mathbf{X}) = \frac{1}{2} \|\mathbf{Y} - \Phi\mathbf{X}\|_F^2 \text{ s.t. } \|\mathbf{X}\|_0 \leq s \quad (1)$$

for the observation matrix $\mathbf{Y} \in \mathbb{C}^{M \times K}$ and the dictionary $\Phi \in \mathbb{C}^{M \times N}$ ($M < N$). The coefficient matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K]$ is the unknown sparse matrix where \mathbf{x}_i is the i -th column of \mathbf{X} . $\|\mathbf{X}\|_0 = |\text{supp}(\mathbf{X})|$, where $\text{supp}(\mathbf{X}) = \{1 \leq i \leq N \mid \mathbf{x}^i \neq 0\} = \Omega$ is called set of active atoms, and \mathbf{x}^i stands for the i -th row of \mathbf{X} . s is the sparsity parameter. In this paper, we propose to include a term promoting the reconstruction of piecewise constant rows in the cost function. So the new criterion is written as:

$$J_1(\mathbf{X}) = \frac{1}{2} \|\mathbf{Y} - \Phi\mathbf{X}\|_F^2 + \lambda \|\mathcal{D}\mathbf{X}\|_1 \text{ s.t. } \|\mathbf{X}\|_0 \leq s \quad (2)$$

where \mathcal{D} is a finite difference operator acting in the horizontal direction (applied to the rows of \mathbf{X}):

$$\|\mathcal{D}\mathbf{X}\|_1 = \sum_{i=1}^N \sum_{j=1}^{K-1} |\mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}|. \quad (3)$$

Thus we seek to find \mathbf{X} such as:

$$\mathbf{X} = \arg \min_{\mathbf{X}, \|\mathbf{X}\|_0 \leq s} J_1(\mathbf{X}). \quad (4)$$

The main contribution of this work lies in the inclusion of the two regularization terms (sparsity constraint and consistency constraint on the rows). Indeed, these two constraints are of different nature (the first one is convex while the second is not), and a direct solution does not exist. Therefore, we propose to solve problem (2) in two steps by considering one constraint at a time.

III. SPARSE REGULARIZED SIMULTANEOUS APPROXIMATION (SR-SA)

A. General structure of the algorithm

The resolution of problem (4) is not easy because it is not convex. Moreover, the sparsity constraint $\|\mathbf{X}\|_0$ leads to an NP-hard problem. Hence, we propose a suboptimal approach which amounts to decompose the problem into two sub-problems:

$$(SP_1) : \Omega = \arg \min_{\Omega} \frac{1}{2} \|\mathbf{Y} - \Phi_{\Omega} \mathbf{X}_{\Omega}\|_F^2 \text{ s.t. } |\Omega| \leq s \quad (5)$$

$$(SP_2) : \mathbf{X}_{\Omega} = \arg \min_{\mathbf{X}_{\Omega}} \frac{1}{2} \|\mathbf{Y} - \Phi_{\Omega} \mathbf{X}_{\Omega}\|_F^2 + \lambda \|\mathcal{D}\mathbf{X}_{\Omega}\|_1. \quad (6)$$

Note that this scheme has the advantage of being computationally efficient. However, other minimization approaches of (4) may be considered and are currently under development. To solve problem in (5), we use the SOLS algorithm (i.e the simultaneous version of the OLS algorithm: *Orthogonal Least Squares*) proposed in [9]. This provides the support Ω . Then, we propose to rewrite the cost function (6) in standard form in order to use an efficient solver (ADMM).

B. Rewriting the problem

Suppose that Ω is known and let $\mathbf{y} = \text{vec}(\mathbf{Y}) \in \mathbb{C}^{MK \times 1}$, and $\mathbf{x} = \text{vec}(\mathbf{X}_{\Omega}) \in \mathbb{C}^{|\Omega|K \times 1}$. It is easy to see that:

$$\|\mathbf{Y} - \Phi_{\Omega} \mathbf{X}_{\Omega}\|_F^2 = \|\mathbf{y} - \mathbf{A}_{\Omega} \mathbf{x}\|_2^2 \quad (7)$$

where $\mathbf{A}_{\Omega} \in \mathbb{C}^{MK \times |\Omega|K}$ is a block-diagonal matrix formed with the columns of Φ corresponding to the sparse solution:

$$\mathbf{A}_{\Omega} = \text{blkdiag}(\underbrace{\Phi_{\Omega}, \dots, \Phi_{\Omega}}_{K \text{ times}}). \quad (8)$$

In addition, for all active atoms Ω we have:

$$\|\mathcal{D}\mathbf{X}_{\Omega}\|_1 = \sum_{i \in \Omega} \sum_{j=1}^{K-1} |\mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}|. \quad (9)$$

Let

$$\Delta = \text{blkdiag}(\underbrace{\mathbf{D}, \dots, \mathbf{D}}_{|\Omega| \text{ times}}) \quad (10)$$

where \mathbf{D} is a matrix of finite differences of order 1 and dimension $K \times K$ such that:

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & & & \mathbf{0} \\ & & \ddots & \ddots & \\ & \mathbf{0} & & -1 & 1 \\ & & & & 0 & 0 \end{bmatrix}. \quad (11)$$

Note that the last row is zero to take into account the line break in the matrix \mathbf{X}_{Ω} . Then, we can write:

$$\|\mathcal{D}\mathbf{X}_{\Omega}\|_1 = \|\Delta \text{vec}(\mathbf{X}_{\Omega}^T)\|_1. \quad (12)$$

Now there exists a permutation matrix \mathbf{P} such that: $\text{vec}(\mathbf{X}_{\Omega}^T) = \mathbf{P} \cdot \text{vec}(\mathbf{X}_{\Omega}) \triangleq \mathbf{P} \cdot \mathbf{x}$. Finally we obtain:

$$\|\mathcal{D}\mathbf{X}_{\Omega}\|_1 = \|\Delta \mathbf{P} \mathbf{x}\|_1 \triangleq \|\mathbf{F} \mathbf{x}\|_1. \quad (13)$$

We can then equivalently reformulate the subproblem (6) as:

$$(SP_2) : \mathbf{x} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}_{\Omega} \mathbf{x}\|_2^2 + \lambda \|\mathbf{F} \mathbf{x}\|_1. \quad (14)$$

C. Regularization problem solution by ADMM

In view of equation (14), the ℓ_1 regularization problem is the same as generalized LASSO. Therefore it can be solved by convex optimization methods. Here, we use the augmented Lagrangian method [20]:

$$\min_{\mathbf{x}, \mathbf{z}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}_{\Omega} \mathbf{x}\|_2^2 + \lambda \|\mathbf{z}\|_1 \text{ s.t. } \mathbf{F} \mathbf{x} - \mathbf{z} = \mathbf{0} \quad (15)$$

where \mathbf{x} and \mathbf{z} are the variables to estimate. We use the ADMM algorithm to solve this problem. ADMM is an iterative algorithm which performs three steps at each iteration k where \mathbf{x} and \mathbf{z} are the variables to estimate as follows:

$$\mathbf{x}^{k+1} = (\mathbf{A}_{\Omega}^T \mathbf{A}_{\Omega} + \rho \mathbf{F}^T \mathbf{F})^{-1} (\mathbf{A}_{\Omega}^T \mathbf{y} + \rho \mathbf{F}^T (\mathbf{z}^k - \mathbf{u}^k)) \quad (16)$$

$$\mathbf{z}^{k+1} = S_{\lambda/\rho}(\mathbf{F} \mathbf{x}^{k+1} + \mathbf{u}^k) \quad (17)$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{F} \mathbf{x}^{k+1} - \mathbf{z}^{k+1} \quad (18)$$

where \mathbf{u} is a scaled dual variable associated to the constraint $\mathbf{x} = \mathbf{z}$ and $\rho > 0$ is the nonnegative penalty parameter. $S_{\lambda/\rho}$ is the soft thresholding operator given by:

$$S_{\lambda/\rho}(x) = \max\{0, x - \lambda/\rho\} - \max\{0, -x - \lambda/\rho\}. \quad (19)$$

This procedure performs alternating minimization of the augmented Lagrangian over \mathbf{x} and \mathbf{z} . Indeed, at the first iteration, \mathbf{z} and \mathbf{u} are fixed and the augmented Lagrangian is minimized over \mathbf{x} ; next \mathbf{x} and \mathbf{u} are fixed and the we minimize over \mathbf{z} ; finally the dual variable \mathbf{u} is updated. It should be noted here that the approach we propose in two stages (finding Ω and \mathbf{x}) yields a suboptimal solution, but that guarantees a low computational cost. Moreover, estimating \mathbf{x} in (16) may become very expensive because of the size of the matrices involved in this equation. However, the block-diagonal structure of the matrix \mathbf{A}_{Ω} avoids its storage. Furthermore, the inversion of the block tri-diagonal matrix $\mathbf{A}_{\Omega}^T \mathbf{A}_{\Omega} + \rho \mathbf{F}^T \mathbf{F}$ can be done using dedicated methods that have a linear cost with respect to the support cardinality ($|\Omega|$).

TABLE I. DIFFERENT CATEGORIES OF THE THREE GROUPS OF WOOD WASTES

Group 1			Group 2			Group 3		
Cat.	Name	Samples	Cat.	Name	Samples	Cat.	Name	Samples
1.1	raw wood	123	2.5	painted solid wood	143	3.10	MDF-HDF	28
1.2	raw plywood	40	2.6	vanished solid wood	139	3.11	painted MDF-HDF	50
1.3	surfaced plywood	23	2.7	painted particle board	24	3.12	surfaced MDF-HDF	57
1.4	vanished plywood	53	2.8	raw particle board	52	3.13	raw fiber board	7
-	-	-	2.9	surfaced particle board	84	3.14	surfaced fiber board	39
-	-	-	-	-	-	3.15	solid wood metal salts	35

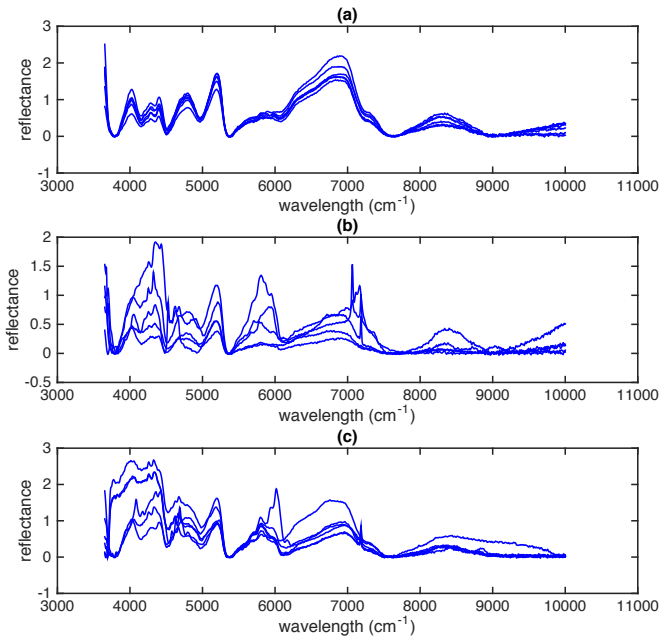


Fig. 1. Spectra from the three groups of wood wastes. (a) Group 1 to value, (b) Group 2 to recycle, (c) Group 3 polluted one to reject.

IV. CLASSIFICATION OF WOOD WASTES NIR SPECTRA

A. Data processing

Experiments were conducted on a wide range of wood wastes collected from a wastes recycling park. Different wood samples were classified into 15 labeled categories by experts. The purpose of this application is to optimize the recycling process by increasing quantities of wood to value (Group 1), to recycle (Group 2) and to reject polluted ones (Group 3). The wood categories and groups are given in Table I.

The acquisition of the infrared spectra was performed on a Nicolet 8700 FTIR spectrometer continuously purged with ultrapure N₂. A MCT detector, and a CaF₂ beam splitter equipped the apparatus. NIR reflectance spectra, recorded in the range 3000-10000 cm⁻¹ (3.33-1 μm), were carried out with near-normal specular reflectance accessory (fixed 10 degree angle of incidence) provided by Pike Technologies. The spectral resolution was 16 cm⁻¹ and 100 scans were co-added for each spectrum which includes 1647 wavelengths. Once the database of wood wastes spectra is completed, we proceed to data processing. This consists of three major steps:

- 1) data pre-processing including baseline removal using the algorithm in [21] and normalization,
- 2) variable selection using the proposed method,

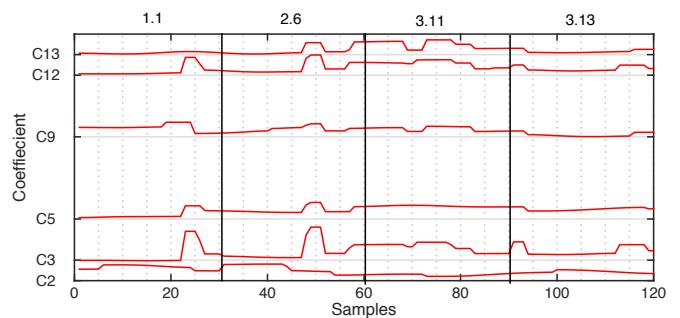


Fig. 2. Example of piecewise constant coefficients of wood samples. Here four categories (1.1, 2.6, 3.11, 3.13) with 30 samples each

- 3) data classification using SVM algorithm with a quadratic kernel function.

Some spectra from the three aforementioned groups are presented in Figure 1. Here, we consider only the binary classification which determines if a sample is to keep (wood to value or to recycle) or to reject (polluted wood).

B. Variable selection

Each column is spectrum and the spectra are grouped by categories in the observation matrix. It is this grouping which induces the piecewise evolution of coefficient matrix, thus justifying the use of the SR-SA method. The dictionary is composed of gaussian functions whose widths σ are varying from 30 to 600 with step 30 cm⁻¹. Their locations (centers) start at 3400 cm⁻¹ and are separated by σ . Sparsity parameter is fixed to 40 and regularization parameter λ to 0.6. Penalty parameter ρ is set to 0.4. The sparse approximation has a major interest in the dimensionality reduction of spectroscopic data. Indeed, it reduces the number of wavelengths from 1647 to 40. Note that samples share a lot of common features corresponding to the wood spectral signature. As shown in Figure 2, SR-SA produces piecewise constant coefficients which favor the grouping of data sharing the same spectral characteristics. This is confirmed, in the next paragraph, by the classification error rate in each class obtained in our tests.

C. SVM classification results

SVM algorithm with quadratic kernel function is used for data classification. As depicted in Table II, four tests are performed. In each test, the NIR spectra are decomposed into 2 datasets of equal size: a training set and a validation set. SVM classification results on the validation set are reported in terms of error rates for each class in Table II for the SR-SA algorithm, compared to the standard SOLS [9] and Sparse

TABLE II. CLASSIFICATION ERROR RATES FOR THE FOUR TESTS PERFORMED

Test #	Categories	SR-SA+SVM		SOLS+SVM		SGLasso+SVM	
		Type I error	Type II error	Type I error	Type II error	Type I error	Type II error
Test 1	1.1 vs 3.10	6.25%	0%	6.25%	0%	6.25%	7.14%
Test 2	2.5 vs 3.11	7.25%	0%	8.7%	4%	27.78%	8%
Test 3	1.1+1.2 vs 3.10+3.13	8.33%	5.56%	12.5%	8.33%	16.67%	5.56%
Test 4	Group 1 + Group 2 vs Group 3	20.54%	5.26%	25.85%	10.53%	26.79%	14.04%

Group Lasso (SGLasso) implemented in SLEP package¹. As compared to standard Lasso, SGLasso fits better our application because it favors the grouping of common coefficients (the rows of X). We present for each test the true negative rate (Type I error) and the false positive rate (Type II error). The results confirm that the proposed approach for variable selection is more appropriate to promote the grouping of samples with common characteristics. Indeed, while the three methods produce similar error rate for the first test, the regularization-based method seems more effective to group different categories of wood in the same class. This is justified by the fact that the proposed approach allows to select coefficients with piecewise constant shape (not necessarily fitted to real spectra shapes) but still potentially discriminative for SVM classifier. Note also that the error rates obtained for the second test are justified by the presence of painted surfaces in the two categories which makes them more difficult to discriminate.

The choice of λ has a major importance to balance between signal reconstruction and the piecewise variation of the coefficients. To evaluate the impact of this parameter we consider Test 3 for which different values of λ are tested to find the optimal one. We present in Table III results of the classification test according to this parameter. We observe that minimum error rates is reached for $\lambda = 0.6$, and in general we have found that the best results are obtained for $\lambda \in [0.5, 0.9]$ for tests 1–4.

TABLE III. IMPACT OF REGULARIZATION PARAMETER VALUE ON CLASSIFICATION RATES

	Type I error	Type II error
$\lambda = 0$ (SOLS solution)	12.5%	8.33%
$\lambda = 0.5$	8.33%	5.56%
$\lambda = 1$	10.42%	5.56%
$\lambda = 1.5$	25.11%	11.11%

V. CONCLUSION

This paper presents a new approach for simultaneous sparse approximation for piecewise signals. This consists of two steps: first, solving the simultaneous sparse approximation problem provides the set of active atoms. Then, the regularized solution is obtained by rewriting the problem in standard form and using ADMM. We showed the advantage of the method in the context features selection and classification of wood wastes NIR spectra. In particular, we found that the regularity constraint on matrix coefficients increases the classification rates compared to the tested sparsity-based variable selection methods.

REFERENCES

- [1] D. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

- [2] E.J. Candès, J. Romberg, and T. Tao, “Stable signal recovery for incomplete and annucarate measurements,” *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, May 2006.
- [3] M.D. Iordache, J. Bioucas-Dias, and A. Plaza, “Sparse unmixing of hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, pp. 2014–2039, June 2011.
- [4] H. Hoefling, “A path algorithm for the fused lasso signal approximator,” *J. Comp. Graph. Stat.*, vol. 19, no. 4, pp. 984–1006, 2010.
- [5] J. Kim and H. Park, “Sparse nonnegative matrix factorization for clustering,” *Technical Report*, 2008.
- [6] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488.
- [7] J.A. Tropp, “Algorithms for simultaneous sparse approximation. Part II: Convex relaxation,” *Signal Process.*, vol. 86, pp. 589–602, March 2006.
- [8] J.A. Tropp, A.C. Gilbert, and M.J. Strauss, “Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit,” *Signal Process.*, vol. 86, pp. 572–588, 2006.
- [9] L. Belmerhnia, E.-H. Djermoune, and D. Brie, “Greedy methods for simultaneous sparse approximation,” in *Proc. EUSIPCO*, 2014, pp. 1851–1855.
- [10] J.D. Blanchard, M. Cermak, D. Hanle, and Y. Jing, “Greedy algorithms for joint sparse recovery,” *IEEE Trans. Signal Process.*, vol. 62, pp. 1694–1704, 2014.
- [11] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, 2006.
- [12] D.P. Wipf and B.D. Rao, “An empirical bayesian strategy for solving the simultaneous sparse approximation problem,” *IEEE Trans. Signal Process.*, vol. 55, pp. 3704–3716, July 2007.
- [13] M.E. Tipping, “Sparse bayesian learning and the relevance vector machine,” *J. Mach. Lear. Res.*, vol. 1, pp. 211–244, 2001.
- [14] Z. Zhang and B.D. Rao, “Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning,” *IEEE J. Select. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.
- [15] Z. Zhang and B.D. Rao, “Exploiting correlation in sparse signal recovery problems: Multiple measurement vectors, block sparsity, and time-varying sparsity,” *arXiv:1105.0725*, 2011.
- [16] Z. Zhang and B.D. Rao, “Sparse signal recovery in the presence of correlated multiple measurement vectors,” in *Proc. IEEE ICASSP*, 2010, pp. 3986–3989.
- [17] M. Luessi, S. Derin Babacan, R. Molina, and A.K. Katsaggelos, “Bayesian simultaneous sparse approximation with smooth signals,” *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5716–5729, 2013.
- [18] Berwin A Turlach, William N Venables, and Stephen J Wright, “Simultaneous variable selection,” *Technometrics*, vol. 47, no. 3, pp. 349–363, 2005.
- [19] Bo Xin, Yoshinobu Kawahara, Yizhou Wang, and Wen Gao, “Efficient generalized fused lasso and its application to the diagnosis of alzheimers disease,” in *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014, pp. 2163–2169.
- [20] 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, 2011.
- [21] Vincent Mazet, Cédric Carteret, David Brie, Jérôme Idier, and Bernard Humbert, “Background removal from spectra by designing and minimising a non-quadratic cost function,” *Chemometrics and intelligent laboratory systems*, vol. 76, no. 2, pp. 121–133, 2005.

¹<http://yelab.net/software/SLEP/>