

Approximation parcimonieuse simultanée constante par morceaux

Leila BELMERHIA¹, El-Hadi DJERMOUNE¹, Cédric CARTERET², David BRIE¹

¹Centre de Recherche en Automatique de Nancy, Université de Lorraine, CNRS
Boulevard des Aiguillettes, BP 70239, 54506, Vandœuvre France

²Laboratoire Chimie Physique et Microbiologie pour l'Environnement, Université de Lorraine, CNRS
405 rue de Vandœuvre, 54500, Villers-lès-Nancy France

leila.belmerhnia@univ-lorraine.fr, el-hadi.djermoune@univ-lorraine.fr
cedric.carteret@univ-lorraine.fr, david.brie@univ-lorraine.fr

Résumé – Ce papier présente une nouvelle technique d'approximation parcimonieuse simultanée intégrant une contrainte de régularité le long des lignes de la matrice de coefficients. Cette approche se décompose en deux étapes: en premier lieu une représentation parcimonieuse de la matrice des coefficients est obtenue en utilisant une méthode d'approximation gloutonne simultanée. Une régularisation avec une pénalité ℓ_1 sur la dérivée des coefficients non nuls est par la suite réalisée de façon à imposer une forme de variation constante par morceaux le long des lignes de la solution. Le problème de régularisation est résolu efficacement en utilisant la méthode d'optimisation ADMM (*Alternate Direction Method of Multipliers*). Des simulations numériques montrent l'intérêt de cette approche pour l'estimation d'une solution dont les coefficients sont constants par morceaux.

Abstract – This paper presents a new technique of simultaneous sparse approximation incorporating a regularity constraint along the lines of the coefficients matrix. This approach comprises two steps: first a sparse representation of the coefficients matrix is obtained using a simultaneous greedy method. Then, a regularization with ℓ_1 penalty on the derivative of nonzero coefficients is performed to enforce a piecewise constant variation form along the rows of the solution. The regularization problem is solved efficiently using the ADMM (Alternate Direction Method of Multipliers) optimization method. Numerical simulations show the advantage of this approach in the estimation of a solution with piecewise constant coefficients.

1 Introduction

L'approximation parcimonieuse est un problème d'inversion linéaire étudié dans différents domaines tels que la compression [1, 2], l'analyse spectrale [3] ainsi que la la regression [4] et la classification [5]. Elle consiste à représenter un signal en utilisant le minimum de vecteurs à partir d'un dictionnaire surdéterminé. A la différence de l'approximation parcimonieuse standard, la version simultanée consiste à reconstruire une matrice d'observations \mathbf{Y} , dont les signaux élémentaires \mathbf{y}_i partagent la même structure de parcimonie. Cette approche a été proposée par Cotter et al. [6] qui ont développé les versions simultanées des algorithmes Matching Pursuit (MP) et FOCUSS. Ils ont également démontré que cette approche permet d'obtenir des taux de bonne reconstruction nettement supérieurs à ceux obtenus avec la version standard. Comme cela a été souligné par Tropp et al. [7], ce type d'approche correspond à une relaxation convexe du problème d'approximation parcimonieuse simultanée exact consistant à trouver la solution ayant le minimum de lignes non nulles. C'est un problème NP-complet pour lequel les méthodes gloutonnes fournissent un bon

compromis entre complexité et coût de calcul [8, 9]. Des conditions de reconstruction exacte sont données dans [10]. D'autres approches forment le problème dans un cadre bayésien [11, 12]. Plus récemment, des travaux se sont intéressés au développement de méthodes d'approximation parcimonieuse simultanée qui imposent une corrélation entre les coefficients de la solution [13, 14]. Mentionnons en particulier [15] qui propose une approche pour imposer une contrainte de douceur selon les lignes de la solution. Notre travail est motivé par une application de classification de données de spectroscopie IR. L'intérêt de l'approche proposée est double. D'une part, elle permet d'exploiter la représentation parcimonieuse afin de réduire la dimension du problème. D'autre part la forme de variation (variation en crêteaux) est adaptée aux problèmes de classification.

Le papier est organisé comme suit: La formulation du problème intégrant les contraintes de parcimonie et de régularité sur la matrice des coefficients est présentée dans la section 2. Dans la section 3 nous développons l'approche proposée pour résoudre le problème. Les résultats expérimentaux sont présentés dans la section 4 et la section 5 conclut le papier.

2 Formulation du problème

Le problème d'approximation parcimonieuse simultanée consiste à trouver la solution de la fonction de coût suivante:

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

pour la matrice d'observations $\mathbf{Y} \in \mathbb{C}^{M \times K}$ et le dictionnaire $\Phi \in \mathbb{C}^{M \times N}$ ($M < N$). La matrice des coefficients $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K]$ est la matrice parcimonieuse à estimer telle que \mathbf{x}_i est la i -ème colonne de \mathbf{X} . $\|\mathbf{X}\|_0 = |\text{supp}(\mathbf{X})|$, où $\text{supp}(\mathbf{X}) = \{1 \leq i \leq N \mid \mathbf{x}^i \neq 0\} = \Omega$ est appelé ensemble des atomes actifs et \mathbf{x}^i représente la i -ème ligne de \mathbf{X} . s est le paramètre de parcimonie. Nous proposons ici d'inclure dans la fonction de coût un terme favorisant la reconstruction des lignes \mathbf{x}^i constantes par morceaux. On pose donc le nouveau critère:

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

où \mathcal{D} est l'opérateur de différence finie appliqué aux lignes de \mathbf{X} et $\mathbf{X}_{i,j}$ l'élément (i, j) de \mathbf{X} :

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

On cherche donc \mathbf{X} tel que:

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

La principale originalité de ce travail réside dans la prise en compte des deux termes de régularisation (la contrainte de parcimonie et la contrainte de régularité sur les lignes).

3 Algorithme proposé

3.1 Structure générale de l'algorithme

La résolution du problème d'optimisation (4) est délicate car il est non convexe. De plus, la contrainte de parcimonie $\|\mathbf{X}\|_0$ conduit à un problème NP-complet. Nous proposons ici une approche sous-optimale qui consiste à décomposer le problème en deux:

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

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

Mentionons que cette structure offre l'avantage d'être peu coûteuse en temps de calcul. Cependant, d'autres schémas de minimisation de (4) sont envisageables et sont actuellement en cours de développement. Le sous problème (5) est résolu en utilisant l'algorithme S-OLS (i.e la version simultanée de l'algorithme OLS: *Orthogonal Least Squares*) proposé dans [9]. Cela permet d'obtenir le support Ω . Il reste à résoudre le sous-problème (6). Pour cela, nous proposons de réécrire la fonction de coût correspondante sous forme standard de façon à pouvoir utiliser un solveur efficace (ADMM).

3.2 Réécriture du problème

Supposons donc que Ω est connu et posons $\mathbf{y} = \text{vec}(\mathbf{Y}) \in \mathbb{C}^{MK \times 1}$, $\mathbf{x} = \text{vec}(\mathbf{X}_{\Omega}) \in \mathbb{C}^{|\Omega|K \times 1}$. On montre alors que:

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

où $\mathbf{A}_{\Omega} \in \mathbb{C}^{MK \times |\Omega|K}$ est une matrice diagonale par blocs obtenue en considérant uniquement les colonnes de Φ correspondant à la solution parcimonieuse:

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

De plus, pour l'ensemble des atomes actifs Ω on a:

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

Soit maintenant

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

où \mathbf{D} est une matrice de différences finies d'ordre 1 et de dimension $K \times K$ telle que:

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

Notons que la dernière ligne est nulle pour prendre en compte le saut de ligne dans la matrice \mathbf{X}_{Ω} . On peut alors écrire:

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

Or il existe une matrice de permutation \mathbf{P} telle que: $\text{vec}(\mathbf{X}_{\Omega}^T) = \mathbf{P} \cdot \text{vec}(\mathbf{X}_{\Omega}) \triangleq \mathbf{P} \cdot \mathbf{x}$. Finalement, on obtient:

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

On peut alors reformuler le sous-problème (6) de façon équivalente selon:

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

3.3 Résolution du problème de régularisation par ADMM

Au regard de l'équation (14), le problème de régularisation ℓ_1 correspond au problème LASSO généralisé. Il peut donc être résolu par une méthode d'optimisation convexe. Pour cela, on propose d'utiliser la méthode du Lagrangien augmenté [16]. D'autres méthodes de calcul peuvent être utilisées comme celle basée sur le gradient proposée dans [17] mais notre choix de l'ADMM est justifié par la facilité d'implémentation de cet algorithme:

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

où \mathbf{x} et \mathbf{z} sont les variables à estimer. La résolution de ce problème se fait via l'algorithme ADMM. Cet algorithme itératif réalise trois étapes à chaque itération k dans lesquelles \mathbf{x} et \mathbf{z} sont estimés respectivement comme suit:

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

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

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

où \mathbf{u} est le multiplicateur du Lagrangien augmenté et $\rho > 0$ est le paramètre de pénalité. $S_{\lambda/\rho}$ est l'opérateur de seuillage doux donné par:

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

Il est à noter ici que l'approche que nous proposons en deux étapes (trouver Ω puis \mathbf{x}) permet d'obtenir une solution sous-optimale mais qui garantit un faible coût de calcul. De plus, l'estimation de \mathbf{x} dans (16) peut vite devenir très coûteuse à cause de la taille des matrices impliquées dans cette équation. Néanmoins, la structure diagonale par blocs de la matrice \mathbf{A}_Ω permet d'éviter son stockage. Par ailleurs, l'inversion de la matrice $\mathbf{A}_\Omega^T \mathbf{A}_\Omega + \rho \mathbf{F}^T \mathbf{F}$ peut se faire efficacement avec des méthodes d'inversion des matrices tri-diagonales par blocs qui ont un coût linéaire.

4 Résultats expérimentaux

Afin de montrer l'intérêt de l'approche proposée, nous présentons dans ce paragraphe quelques résultats de reconstruction obtenus sur des données simulées. Le dictionnaire $\Phi \in \mathbb{R}^{30 \times 200}$ est composé de 200 gaussiennes normalisées avec une largeur de 0.25 et dont les centres couvrent uniformément l'intervalle $[0, 30]$. Chaque colonne de la matrice des observations $\mathbf{Y} \in \mathbb{R}^{30 \times 20}$ est la somme de 3 gaussiennes. On impose à chacune des gaussiennes une évolution en créneau dont les instants de début et de fin sont tirés aléatoirement. Ce signal est par la suite perturbé avec un bruit additif gaussien de puissance 0.04. L'algorithme utilisé pour déterminer l'ensemble actif dans la première étape est S-OLS. Les coefficients estimés avec l'approche proposée sont comparés à ceux obtenus avec les moindres carrés ($\lambda = 0$) et la régularisation de Tikhonov (pénalité ℓ_2). Les paramètres de réglage sont: $s = 3$, et $\rho = 0.5$. Les résultats sont donnés pour deux valeurs différentes du paramètre de régularisation $\lambda = 0.9$ et $\lambda = 0.4$ sur la figure 1. En comparant les trois méthodes nous constatons que les solutions obtenues par les méthodes des moindres carrés et de la régularisation de Tikhonov ne reproduisent pas bien la variation constante par morceaux contrairement à la méthode basée sur une régularisation ℓ_1 . L'intérêt d'inclure une pénalité ℓ_1 à la fonction de coût initiale afin de favoriser cette variation constante par morceaux est observable. Les résultats

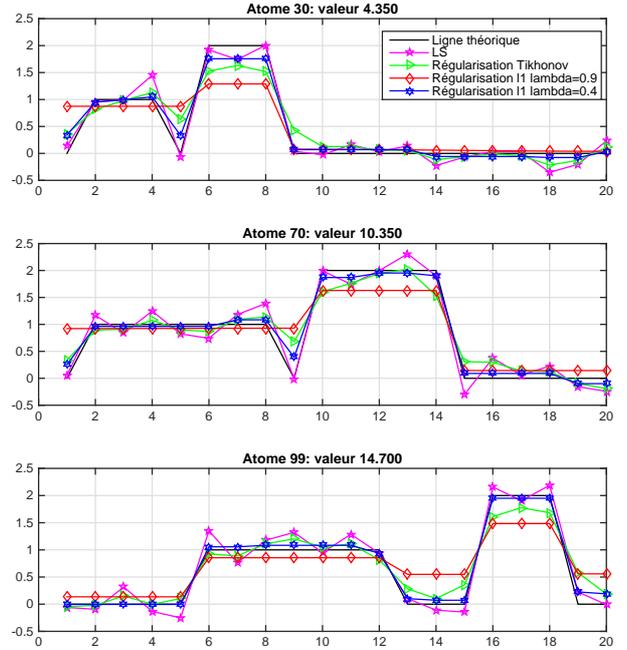


FIG. 1: Solution estimée des atomes actifs. Résultats de la méthode proposée avec $\lambda = 0.9$ et $\lambda = 0.4$. Comparaison avec l'estimation par les moindres carrés et la régularisation de Tikhonov

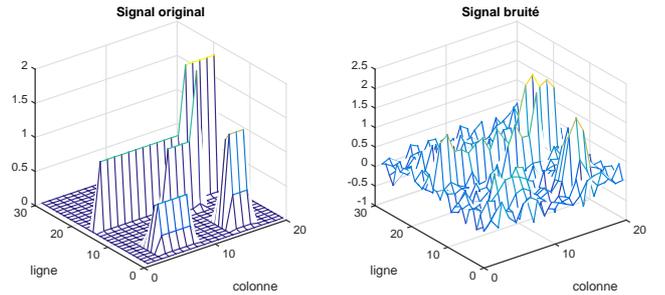


FIG. 2: Allure du signal d'entrée constant par morceaux et la version perturbée par un bruit additif.

obtenus en variant la valeur du paramètre λ mettent en évidence son influence sur la régularisation. En effet, la régularisation est plus forte quand ce paramètre est élevé et le résultat est donc plus lisse. Alors que le choix d'une valeur moyenne pour ce paramètre (0.4 par exemple) permet d'avoir un bon compromis entre la régularité et la précision dans l'estimation des coefficients. L'estimation du signal d'origine de la figure 2 perturbé par un bruit additif est présentée sur la figure 3. On constate que le signal est correctement reconstruit tout en maintenant une évolution constante par morceaux. Ceci est confirmé par l'évolution de l'erreur de reconstruction de la figure 4 qui montre clairement l'intérêt de l'approche proposée, en particulier dans le cas d'un fort rapport de signal à bruit.

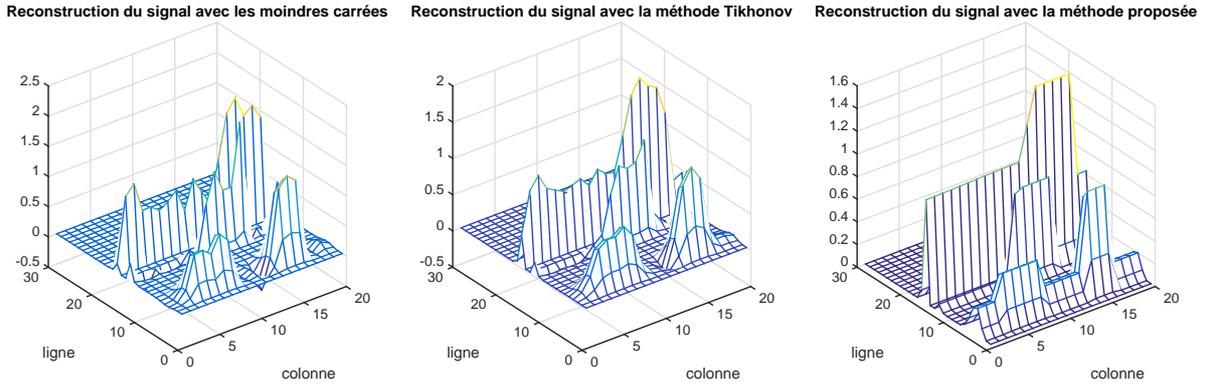


FIG. 3: Comparaison entre les résultats de reconstruction des trois méthodes du signal bruité.

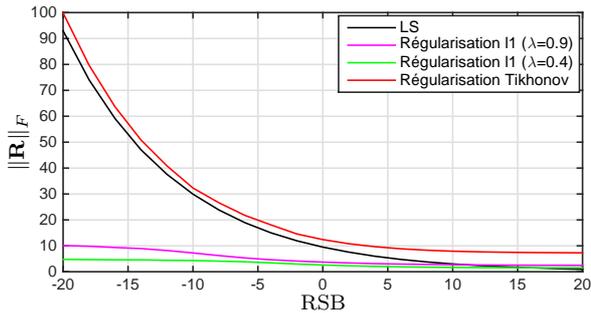


FIG. 4: Evolution de l'erreur de reconstruction $\mathbf{R} = \mathbf{Y} - \Phi_{\Omega}\mathbf{X}_{\Omega}$ des trois méthodes selon le niveau du RSB.

5 Conclusion

Ce papier présente une nouvelle approche pour l'approximation parcimonieuse simultanée adaptée aux signaux constants par morceaux. Cette technique se décompose en deux étapes: d'une part, la résolution du problème d'approximation parcimonieuse permet d'obtenir l'ensemble des atomes actifs. D'autre part, la solution régularisée est obtenue en réécrivant le problème sous forme standard et en utilisant l'algorithme ADMM. Notre motivation est d'exploiter le potentiel de ce type d'approche dans la classification de données car la représentation constante par morceau traduit parfaitement l'appartenance ou non à une classe donnée. L'intérêt de cette approche dans le cadre de la classification de données spectroscopiques est démontré dans [18].

Références

- [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 annuncarate 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. EU-SIPCO*, 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] 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.
- [12] M.E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Lear. Res.*, vol. 1, pp. 211–244, 2001.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] Amir Beck and Marc Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *Image Processing, IEEE Transactions on*, vol. 18, no. 11, pp. 2419–2434, 2009.
- [18] L. Belmerhnia, E.-H. Djermoune, C. Carteret, and D. Brie, "Simultaneous regularized sparse approximation for wood wastes nir spectra features selection," *Tech. Rep.*, 2015.