EUSIPCO 2013 1569744217

SPARSE REDUNDANT FORMULATIONS AND NON-NEGATIVITY IN BLIND SOURCE SEPARATION

Jérémy Rapin[⋆†] *Jérôme Bobin*[⋆] *Anthony Larue*[†] *Jean-Luc Starck*[⋆]

[⋆] CEA, IRFU, Service d'Astrophysique, 91191 Gif-sur-Yvette Cedex, France.
[†] CEA, LIST, 91191 Gif-sur-Yvette Cedex, France.

ABSTRACT

Blind Source Separation (BSS) aims at finding a factorization of multi-spectral data into a mixing matrix and a source matrix. In this field, Non-negative Matrix Factorization (NMF) assumes that both matrices are non-negative. Very few NMF algorithms are further able to encompass sparsity in a transformed domain because of the difficulty in enforcing the solution to be non-negative and sparse simultaneously in two different domains. In this article, we adapt the framework of an algorithm, non-negative GMCA, in order to overcome this issue for a redundant transform, using modern proximal calculus techniques. We therefore obtain solutions satisfying both constraints simultaneously contrarily to other algorithms which apply them alternately. We provide the first comparison of analysis and synthesis sparse formulations in BSS and show that the analysis sparse formulation dramatically improves the identification of sources from noisy mixtures of synthetic nuclear magnetic resonance (NMR) spectra.

Index Terms— BSS, NMF, sparsity, wavelets, analysis

1. INTRODUCTION

In Blind Source Separation (BSS), one has access to m measurement vectors $y_{i,\cdot}$, assumed to result from mixtures of a limited number r of unknown sources $s_{j,\cdot}$ which are n samples long. The linear mixture model further assumes that these mixtures can be written under the form:

$$y_{i,\cdot} = \sum_{j=1}^{r} a_{ij} s_{j,\cdot} + z_{i,\cdot} \ , \ \forall i \in \{1, .., m\}, \qquad (1)$$

where the a_{ij} are unknown mixture coefficients, and the $z_{i,\cdot}$ vectors are added in order to account for measurement noise and model imperfections. This can be conveniently recast under the form $\mathbf{Y} = \mathbf{AS} + \mathbf{Z}$ where the measurements $y_{i,\cdot}$, the sources $s_{j,\cdot}$ and the noise $z_{i,\cdot}$ are respectively rows of $\mathbf{Y}$, $\mathbf{S}$ and $\mathbf{Z}$. For i.i.d. Gaussian noise, the maximum-likelihood estimate is then provided by the standard problem:

$$\underset{\mathbf{A},\,\mathbf{S}}{\operatorname{argmin}} \ \frac{1}{2}||\mathbf{Y} - \mathbf{AS}||_2^2, \qquad (2)$$

emails: firstname.lastname@cea.fr

with the notation $||\mathbf{X}||_p = \sqrt[p]{\sum_{ij} |x_{ij}|^p}$ (Frobenius norm for $p = 2$). This problem however has an infinity of minima. It is therefore useful to add some more information on the sources $\mathbf{S}$ and/or the mixing matrix $\mathbf{A}$ in order to privilege solutions with a desired structure.

In this article, we focus on the case where $\mathbf{A}$ and $\mathbf{S}$ coefficients are required to be non-negative, termed Non-Negative Matrix Factorization (NMF) [1, 2]. This assumption arises naturally in many applications such as clustering [3] or audio processing [4] for instance. Indeed, sources can for example represent power spectra, which are non-negative, and the mixtures can represent concentrations, which cannot be negative either. This constraint can be dealt with in a geometrical way [5] or through the minimization of a cost function such that in problem (2). In this category, most algorithms alternately update $\mathbf{A}$ and $\mathbf{S}$. Indeed, conveniently, the subproblem of recovering a non-negative $\mathbf{S}$ from problem (2) with $\mathbf{A}$ fixed is convex; and vice-versa for the subproblem in $\mathbf{A}$.

The non-negativity of $\mathbf{A}$ and $\mathbf{S}$ is often not sufficient for their recovery. In non-negative ICA [6], one further assumes the independence of the sources. However, this approach is sensitive to noise since it does not model it. Sparsity, on the other hand, can help handle it more efficiently. We give a short introduction about this *prior* in the next section, before presenting our new sparse NMF algorithm.

2. SPARSITY AND NMF: STATE OF THE ART

A sparse signal is a signal which concentrates its information into only a few large non-zero coefficients, or can be well approximated in such a way. This knowledge about the signal can be used as a *prior*, with a sparse regularization of the type $\lambda||x||_0$ which for instance counts non-null coefficients in x and therefore limits their number. However, this regularization is not convex and leads to combinatorial problems. The use of $||\ .\ ||_1$ is consequently often preferred, as the closest convex surrogate to the ℓ_0 pseudo-norm.

Sparsity *priors* were shown to significantly improve the results in NMF and they can be applied in a variety of ways. In [7], Hoyer constrains a sparsity level $||s_{i,\cdot}||_1/||s_{i,\cdot}||_2$ for

Algorithm 1 nGMCA framework

Require: $\mathbf{Y}$, K

1: **initialize** $\mathbf{A}_0$, $\mathbf{S}_0$ and λ_1
2: **for** $k \leftarrow 1, K$ **do**
3: Normalize the columns of $\mathbf{A}_{k-1}$
4: $\mathbf{S}_k \leftarrow \underset{\mathbf{S}\geq\mathbf{0}}{\text{argmin}}\ \frac{1}{2}\|\mathbf{Y} - \mathbf{A}_{k-1}\mathbf{S}\|_2^2 + \lambda_k\|\mathbf{S}\|_1$
5: $\mathbf{A}_k \leftarrow \underset{\mathbf{A}\geq\mathbf{0}}{\text{argmin}}\ \frac{1}{2}\|\mathbf{Y} - \mathbf{A}\mathbf{S}_k\|_2^2$
6: Select $\lambda_{k+1} \leq \lambda_k$
7: **end for**
8: **return** $\mathbf{A}_K$, $\mathbf{S}_K$

each row of $\mathbf{S}$. Kim & Park [3] and Zdunek & Cichocki's [8] use a regularization of type $\sum_j ||s_{\cdot,j}||_1^2$ where the $s_{\cdot,j}$ are columns of $\mathbf{S}$. It is also possible to use the ℓ_1 regularization $||\mathbf{S}||_1$ such as in [9].

A signal can however be sparse in a domain other than the direct domain: its sparsity depends on the basis or dictionary in which it is expressed. A sinusoid is sparse in the Fourier domain for instance, since it can be encoded with one coefficient in this domain, while in the direct domain most of its coefficients are non-zero (not sparse at all). The fact that most natural signals can be expressed in a sparse way using an appropriately chosen basis such as wavelets, has been used effectively in BSS by Zibulevsky & Pearlmutter [10] and Bobin et al.'s Generalized Morphological Component Analysis (GMCA) [11]. Still, few algorithms use sparsity in a transformed domain together with non-negativity in the direct domain [12]. To our knowledge, none provides solution optimally satisfying both at the same time since they handle the constraints alternately. In this article we give an overview of Rapin et al.'s algorithm [9] and its proximal approach (section 3). We then present the major modifications allowing to find solutions optimally verifying non-negativity in the direct domain, and sparsity in a redundant transformed domain (section 4). We finally compare on realistic data the two formulations induced by this kind of *prior* —analysis and synthesis— together with other state-of-the-art algorithms (section 5).

3. NON-NEGATIVE GMCA

In [9], Rapin et al. proposed their non-negative GMCA (nGMCA)[1], described in **Algorithm 1**, which adapted GMCA to NMF by adding the non-negativity constraint. It aims at solving the following problem:

$$\underset{\mathbf{A}\geq\mathbf{0},\ \mathbf{S}\geq\mathbf{0}}{\text{argmin}}\ \frac{1}{2}||\mathbf{Y} - \mathbf{A}\mathbf{S}||_2^2 + \lambda||\mathbf{S}||_1, \tag{3}$$

The subproblem in $\mathbf{A}$ (line 4 of Algorithm 1) can be efficiently solved with proximal splitting methods such as the forward-backward (FB) algorithm [13]. This algorithm allows the minimization of the sum of a convex and differentiable function f, and a non-differentiable proper convex and lower semi-continuous function g. In the case of the update of $\mathbf{A}$, f is the data fidelity term $f(\mathbf{A}) = \frac{1}{2}||\mathbf{Y} - \mathbf{A}\mathbf{S}||_2^2$, and $g(\mathbf{A}) = i_{\mathbf{A}\geq\mathbf{0}}(\mathbf{A})$ is the characteristic function of the set of non-negative matrices, which enforces the non-negativity constraint on $\mathbf{A}$. The algorithm requires the gradient of f, which is straightforward to compute, and the proximal operator of g which is provided in table 1 (proximal #1). The subproblem in $\mathbf{S}$ is solved similarly, with proximal #3. Efficient implementation details can be sought in [14].

#	Regularization	Proximal operator
1	$i_{x\geq 0}(x)$	$[x]_+ = \max(x, 0)$
2	$\lambda\|\|x\|\|_1$	$\text{Soft}_\lambda(x) = \text{sign}(x)[\ \|x\| - \lambda\ ]_+$
3	$\lambda\|\|x\|\|_1 + i_{x\geq 0}(x)$	$[\text{Soft}_\lambda(x)]_+$
4	$i_{(x\mathbf{W})\geq 0}(x)$	$x + [-x\mathbf{W}]_+\mathbf{W}^T$
5	$\lambda\|\|x\mathbf{W}^T\|\|_1$	$x - (\underset{\|\|u\|\|_\infty\leq\lambda}{\text{argmin}}\ \|\|x - u\mathbf{W}\|\|_2^2)\mathbf{W}$
6	$i_{\|\|x\|\|_\infty\leq\lambda}(x)$	$x - \text{Soft}_\lambda(x)$

Fig. 1. Regularizations and their proximal operators (with x a row vector and $\mathbf{W}$ such that $\mathbf{W}^T\mathbf{W} = \mathbf{I}$)

The λ parameter controls the sparsity of the result. As in GMCA, it is changed during the algorithm to fulfill two main purposes. The first is to help separating the sources: a large λ at the beginning of the algorithm favors large and discriminating coefficients of $\mathbf{S}$ and therefore helps providing a rough estimation of the mixing directions in $\mathbf{A}$. Decreasing it afterwards allows refinement of the solution. The second purpose is to denoise the signal. Indeed, unlike the sources, Gaussian noise is not sparse and is spread over all the coefficients. Keeping a final λ of the order of the noise level on the gradient ($\kappa_\sigma\sigma_{MAD}$ with $\kappa_\sigma \approx 1$ and σ_{MAD} the Median Absolute Deviation estimate which can be computed online) prevents a substantial amount of noise from entering the solution. In between, λ is decreased linearly.

4. REDUNDANT SPARSE REGULARIZATION

To date, regularization in another basis has barely been used for NMF before, because of the technical difficulties in dealing with two domains at the same time. Still, signals are rarely sparse in the direct domain, while they are commonly sparse in the wavelet domain. In particular, redundant transforms were shown to significantly improve reconstructions [15]. In the next, we consider redundant wavelets $\mathbf{W}$ from the Rice wavelet toolbox[2], which satisfy $\mathbf{W}^T\mathbf{W} = \mathbf{I}$ (tight frames) after some renormalization. However, such redundant transforms can be used to apply a sparse *prior* with two formulations which have very different behaviors [16, 17].

[1] termed (S)rGMCA in [9]

[2] `http://dsp.rice.edu/software/rice-wavelet-toolbox`

4.1. Synthesis formulation

The synthesis formulation consists in reconstructing the sources in the redundant transformed domain. $\mathbf{S}$ is consequently considered as synthesized by as few atoms as possible from the redundant dictionary. The update of the wavelet coefficients $\mathbf{S}_W$ is then expressed as:

$$\underset{(\mathbf{S}_W\mathbf{W})\geq\mathbf{0}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y}-\mathbf{A}\mathbf{S}_W\mathbf{W}||_2^2+\lambda||\mathbf{S}_W||_1, \qquad (4)$$

At the end, $\mathbf{S}=\mathbf{S}_W\mathbf{W}$. The domain transform provides a new difficulty compared to sparsity in the direct domain. Indeed, there is no convenient way to compute the proximal operator of $g(\mathbf{S})=\lambda||\mathbf{S}_W||_1+i_{\mathbf{S}_W\mathbf{W}\geq\mathbf{0}}(\mathbf{S})$, so that one cannot use the FB algorithm anymore for the update of $\mathbf{S}$. Instead, one can apply the generalized forward-backward algorithm (GFB) [18], which considers g as the sum of two convex lower semi-continuous functions $g_1(\mathbf{S})=\lambda||\mathbf{S}_W||_1$ and $g_2(\mathbf{S})=i_{\mathbf{S}_W\mathbf{W}\geq\mathbf{0}}(\mathbf{S})$ and which requires their proximal operator, provided in table 1 (repectively proximal #1, and proximal #4 for which the derivation is not given because of the lack of space).

4.2. Analysis formulation

It is essential to notice that in the previous formulation, $\mathbf{S}\mathbf{W}^T$ is generally not equal to $\mathbf{S}_W$. Indeed, while $\mathbf{W}^T\mathbf{W}=\mathbf{I}$, $\mathbf{W}\mathbf{W}^T\neq\mathbf{I}$. In the analysis formulation, one therefore directly reconstructs $\mathbf{S}$ in the direct domain while penalizing correlations between the sources and the dictionary:

$$\underset{\mathbf{S}\geq\mathbf{0}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{Y}-\mathbf{A}\mathbf{S}||_2^2+\lambda||\mathbf{S}\mathbf{W}^T||_1, \qquad (5)$$

The minimization is carried out in the same way than in the synthesis formulation, but with proximal #5 instead of proximal #4. However, proximal #5 is no more analytic and needs to be computed through a subroutine (using the FB algorithm with proximal #6 for instance).

5. EXPERIMENTS

5.1. Settings

In physical applications, molecules can be identified by their specific Nuclear Magnetic Resonance (NMR) spectra. In this section, we evaluate the algorithms on simulated realistic data, using $m=32$ mixtures of $r=15$ NMR spectra of natural molecules, with $n=1024$ samples each. The mixing matrix $\mathbf{A}$ is drawn as the absolute value of an i.i.d Gaussian random matrix. The peaks localizations were found in the Spectral Database for Organic Compounds, SDBS[3]. In order to account for the acquisition imperfections, they were convoluted with a Laplacian kernel with 4 samples width at half-maximum. Some of the obtained sources are shown in figure 2. They are naturally non-negative and already significantly sparse in the direct domain. They can however benefit from wavelet-sparsity since they are continuous. This added information is especially helpful in these difficult settings, with few measurements and spectra which can be highly overlapping. Since the peaks are very narrow, we used Daubechies-4 wavelets with only 2 levels. For the three nGMCA-based algorithms, κ_σ is set to 1 and never changed. The data matrix $\mathbf{Y}$ is given as $\mathbf{Y}=\mathbf{A}\mathbf{S}+\mathbf{Z}$ where $\mathbf{Z}$ is a Gaussian noise matrix. An example of mixture is provided in figure 3.

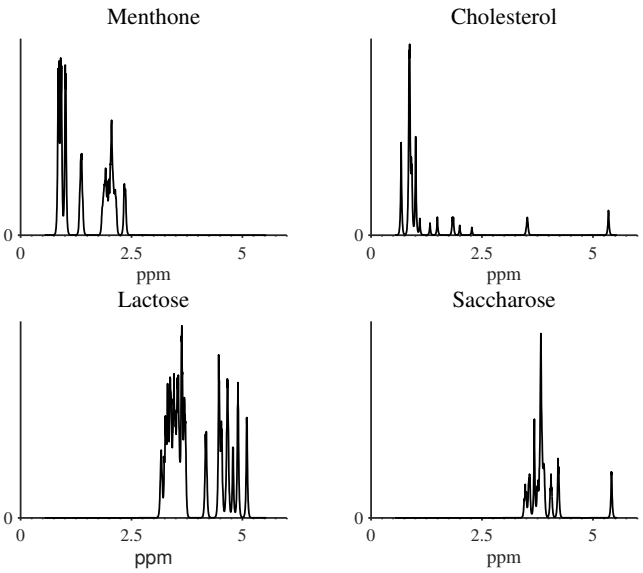

Fig. 2. NMR spectra of 4 chemical compounds.

Our algorithm is compared with ALS [1] as a standard approach for NMF; Hoyer's algorithm [7] and Kim & Park's algorithms [3] which are publicly available sparse NMF algorithms; and the regular nGMCA with sparsity in the direct domain. In Kim & Park's algorithm the standard settings are used; and in Hoyer's algorithm, the required sparsity coefficient is tuned using the ground truth data.

5.2. Influence of the Noise

In [19], Vincent et al. have proposed several scale-invariant criteria to evaluate the performances of BSS techniques. In particular, they propose the Source Distortion Ratio (SDR) for an estimated source s:

$$\mathrm{SDR}(s)=10\log_{10}\left(\frac{||s_{\text{target}}||_2^2}{||s-s_{\text{target}}||_2^2}\right), \qquad (6)$$

where s_{target} is the projection of s on the target reference source. As stated in [19], this criterion is a global performance measure taking into account all the elements of the

[3] http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi

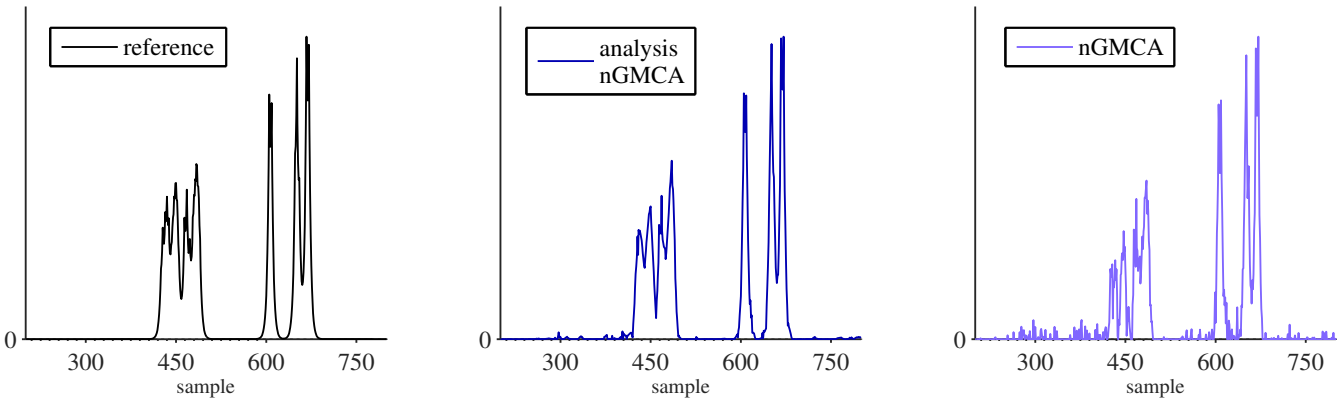

Fig. 5. Example of reconstruction (mannitol, $\mathrm{SNR}_Y = 15\mathrm{dB}$)

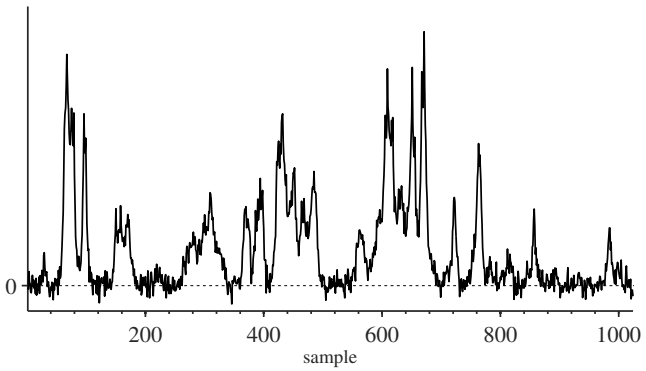

Fig. 3. Example of mixture (main component: mannitol, $\mathrm{SNR}_Y = 15\mathrm{dB}$)

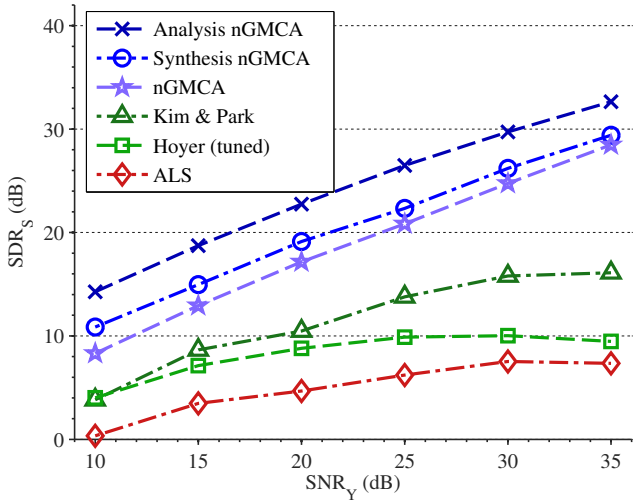

Fig. 4. SDR_S with respect to the level of noise SNR_Y (average of 24 synthetic NMR data simulations)

reconstruction, i.e. a correct separation, efficient denoising and little artifacts left by the algorithm. The larger this criterion, the better the reconstruction. Figure 4 shows the mean SDR of the sources in $\mathbf{S}$ for several levels of noise. The amount of noise is provided in term of SNR on $\mathbf{Y}$ (the smaller SNR_Y, the noisier). Logically, nGMCA performs already significantly well compared to other algorithms since it was specifically designed to handle ill-conditioned settings and noise, without any tuning. However, synthesis-sparsity provides an additional gain of 1dB, and analysis-sparsity 3dB even more. This tends to corroborate the interest of analysis formulations over synthesis formulations on natural signals which was already observed for inverse problems [16, 17].

5.3. Separation and Denoising

The influence of the wavelet transform can be seen on the example of a source reconstruction in figure 5. The added wavelet-sparsity information in the analysis formulation provides both a better denoising of the spectrum (around sample 300 for instance) and lesser interferences (around sample 450). Also important is that the wavelet-sparse *prior* preserves the sharpness of the peaks, which would have been lost with a smooth regularization of the variations.

6. CONCLUSION

nGMCA is able to exactly deal with the non-negative and sparse *priors* which provides an appreciated robustness in the complex settings we have studied (few measurements, correlated sources, noise) compared to other algorithms, while not needing any tuning of its sparsity parameter.

In this article, we have further developed this ability to handle a sparse *prior* in a redundant domain, different from the one where the non-negativity applies. This property was shown to be efficient for both separation and denoising of natural signals compared to other methods. To our knowledge, we provide here the first comparison of synthesis- and analysis-sparse BSS and corroborate the advantage of the analysis formulation for the reconstruction of natural signals.

Further work will make use of this ability in order to separate highly correlated sources, including signals with a significant baseline.

7. REFERENCES

[1] Pentti Paatero and Unto Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, pp. 111–126, 1994.

[2] Daniel D. Lee and H. Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[3] Jingu Kim and Haesun Park, "Sparse Nonnegative Matrix Factorization for Clustering," Tech. Rep., Georgia Institute of Technology, 2008.

[4] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.

[5] Massoud Babaie-Zadeh, Ali Mansour, Christian Jutten, and Farrokh Marvasti, "A Geometric Approach for Separating Several Speech Signals.," in *ICA*, 2004, pp. 798–806.

[6] Mark D Plumbley and Erkki Oja, "A Nonnegative PCA Algorithm for Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 66–76, 2004.

[7] Patrik O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, no. 5, pp. 1457–1469, 2004.

[8] Rafal Zdunek and Andrzej Cichocki, "Nonnegative matrix factorization with constrained second-order optimization," *Signal Processing*, vol. 87, no. 8, pp. 1904–1916, 2007.

[9] Jérémy Rapin, Jérôme Bobin, Anthony Larue, and Jean-Luc Starck, "Robust Non-Negative Matrix Factorization for Multispectral Data with Sparse Prior," *Proceedings of ADA7*, 2012.

[10] Michael Zibulevsky and Barak A. Pearlmutter, "Blind Source Separation by Sparse Decomposition in a Signal Dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.

[11] Jérôme Bobin, Jean-Luc Starck, Jalal Fadili, and Yassir Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2662–2674, 2007.

[12] Lingling Jiang and Haiqing Yin, "Bregman iteration algorithm for sparse nonnegative matrix factorizations via alternating L1-norm minimization," *Multidimensional Systems and Signal Processing*, vol. 23, pp. 315–328, 2012.

[13] Patrick L. Combettes and Valérie R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[14] Amir Beck and Marc Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183, 2009.

[15] Ronald R. Coifman and Dave L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*, G. Oppenheim A. Antoniadis, Ed. 1995, pp. 125–150, Springer-Verlag.

[16] Michael Elad, Peyman Milanfar, and Ron Rubinstein, "Analysis Versus Synthesis in Signal Priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, June 2007.

[17] Ivan W. Selesnick and Mario A. T. Figueiredo, "Signal Restoration with Overcomplete Wavelet Transforms: Comparison of Analysis and Synthesis Priors," *In Proceedings of SPIE*, vol. 7446 (Wavelets XIII), August 2009.

[18] Hugo Raguet, Jalal Fadili, and Gabriel Peyré, "Generalized Forward-Backward Splitting," Tech. Rep., Preprint Hal-00613637, 2011.

[19] Emmanuel Vincent, Rémi Gribonval, and Cédric Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.