Heuristics for Efficient Sparse Blind Source Separation

Christophe Kervazo
CEA Saclay
France, Gif-sur-Yvette, 91191 cedex
Email: christophe.kervazo@cea.fr

Jérôme Bobin
CEA Saclay
France, Gif-sur-Yvette, 91191 cedex

Cécile Chenot
Institute for Digital Communications
University of Edinburgh

***Abstract*—Sparse Blind Source Separation (sparse BSS) is a key method to analyze multichannel data in fields ranging from medical imaging to astrophysics. However, since it relies on seeking the solution of a *non-convex* penalized matrix factorization problem, its performances largely depend on the optimization strategy. In this context, Proximal Alternating Linearized Minimization (PALM) has become a standard algorithm which, despite its theoretical grounding, generally provides poor practical separation results. In this work, we first investigate the origins of these limitations, which are shown to take their roots in the sensitivity to both the initialization and the regularization parameter choice. As an alternative, we propose a novel strategy that combines a heuristic approach with PALM. We show its relevance on realistic astrophysical data.**

I. Introduction

A. Blind Source Separation problem

In the BSS [1] framework, the data are composed of m observations, each of which has t samples. These observations are supposed to be some linear combinations of n sources. The objective of BSS is to retrieve the sources as well as the mixing coefficients. In matrix form, the goal is therefore to find two matrices $\mathbf{S}$ (of size $n \times t$) and $\mathbf{A}$ (of size $m \times n$), called respectively the source and the mixing matrices, such that: $\mathbf{X} = \mathbf{AS} + \mathbf{N}$, where $\mathbf{X}$ (of size $m \times t$) is the observation matrix that is corrupted with some unkwown noise $\mathbf{N}$. Since it requires tackling an ill-posed unsupervised matrix factorization problem, further assumptions are needed, including the statistical independance of the sources (ICA - [1]), the non-negativity of $\mathbf{A}$ and $\mathbf{S}$ [2]. In this work, we will focus on the sparsity of the sources [3]–[6]. In this framework, sparse BSS will aim at finding a (local) minimum of:

$$\min_{\mathbf{A},\mathbf{S}} \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \iota_{\{\forall i; \|\mathbf{A}^i\|_2^2 \leq 1\}}(\mathbf{A}) + \left\| \mathbf{R_S} \odot (\mathbf{S}\mathbf{\Phi_S^T}) \right\|_{\mathbf{1}} \quad (1)$$

The first data fidelity term promotes a faithful reconstruction of the data. The use of the Frobenius norm $\|.\|_F$ stems from the assumption of a Gaussian noise $\mathbf{N}$. The second term involving the characteristic function ι corresponds to the oblique constraint ensuring that the columns $\mathbf{A}^i$of $\mathbf{A}$ are all in the ℓ_2 ball. This avoids degenerated $\mathbf{A}$ and $\mathbf{S}$ matrices where $\|\mathbf{A}\| \to \infty$ and $\|\mathbf{S}\| \to 0$. The last term involving the Hadamard product $\odot$ enforces a ℓ_1 sparsity constraint in a transformed domain $\mathbf{\Phi_S}$. In the following, $\mathbf{\Phi_S}$ will be taken equal to either the identity (in which case the sparsity is enforced in the direct domain) or the starlet transform [7]. $\mathbf{R_S}$ controls the trade-off between the data fidelity and the sparsity terms. It can be decomposed into $\mathbf{R_S} = \mathbf{\Lambda_S W}$ where $\mathbf{\Lambda_S}$ is a diagonal matrix of the regularization parameters $\lambda_1, \lambda_2, ..., \lambda_n$ and $\mathbf{W}$ is a matrix used to introduce individual penalization coefficients in the context of reweighted ℓ_1 [8] (when no reweighting is used, $\mathbf{W}$ is equal to the identity matrix).

B. Sparse BSS in practice

Since sparse BSS requires solving a penalized matrix factorization problem, it is important to highlight that the separation quality strongly depends on the optimization strategy. For that purpose, different algorithmic frameworks have been used so far: projected Alternate Least-Square (ALS - [2]), Block-Coordinate Descent (BCD - [9]) and PALM [10]. However, any practitioner can draw the same conclusion: the solution of sparse BSS methods is highly sensitive to the initial point and the values of the regularization parameters, which are generally tricky to tune without any first guess of the solution.

Initialization: As problem (1) is not convex but multi-convex, the algorithm performing its minimization can be trapped in spurious critical points, depending on the initial matrices $\mathbf{A}$ and $\mathbf{S}$. Different optimization strategies can be affected differently.

Regularization parameters: The practical choice of the regularization parameters $\mathbf{\Lambda_S}$ is of paramount significance:

- The parameters are directly impacting the shape of the solution through the trade-off between the sparsity level and fidelity to the data.
- The problem being non-convex, a change in the parameters can also bring the optimization algorithm to stabilize in the neighborhood of a different critical point, potentially leading to a very different solution.

C. Contributions

While PALM is theoretically well rooted and yields rather fast minimization schemes (in contrast to BCD), it generally provides poor separation results. We first investigate the origins of this poor practical efficiency, and especially the sensitivity to both initialization and regularization parameter settings. We further show how PALM-based implementations can benefit from the information provided by heuristic approaches which are in pratice more robust. The robustness of

ISBN 978-90-827970-1-5

the proposed combined strategy is demonstrated on realistic astrophysical data.

II. Optimization strategy: PALM and GMCA

As emphasized before, the optimization strategy is decisive to avoid stationary points. In the following, we will discuss PALM [10] and projected ALS such as Generalized Morphological Component Analysis (GMCA - [3]). BCD will not be studied since it requires an exact minimization at each iteration, leading in general to a high computational cost.

A. PALM algorithm

PALM has been introduced in [10] and has benefited from several extensions [11], [12]. Its attractiveness partially comes from the fact that it is proved to converge to a local minimum of (1) under mild conditions [10].

PALM is an iterative algorithm. At each iteration, it alternates between a proximal gradient step on $\mathbf{A}$ and $\mathbf{S}$. In our case, it can be shown that the proximal operator corresponding to the ℓ_1 sparsity term in (1) is the soft-thresholding operator (with for each source $\mathcal{S}_\lambda(.) = \text{sign}(.) \odot [|.| - \lambda]_+$) applied in the transform domain. Furthermore, the operator corresponding to the oblique constraint is the projection of each column of $\mathbf{A}$ on the ℓ_2 unit ball, which we shall denote as $\Pi_{\|.\|_2}(.)$. Each iteration (k) can then be read as:

1 - Update of $\mathbf{S}$ using a fixed $\mathbf{A}$:

$$\hat{\mathbf{S}} = \mathbf{S}^{(k-1)} - \frac{\gamma}{\|\mathbf{A}^{(k-1)T}\mathbf{A}^{(k-1)}\|_2}\mathbf{A}^{(k-1)T}(\mathbf{A}^{(k-1)}\mathbf{S}^{(k-1)} - \mathbf{X}) \quad (2)$$

$$\mathbf{S}^{(k)} = \mathcal{S}_{\frac{\gamma \mathbf{R_S}}{\|\mathbf{A}^{(k-1)T}\mathbf{A}^{(k-1)}\|_2}}\left(\hat{\mathbf{S}}\mathbf{\Phi_S^T}\right)\mathbf{\Phi_S} \quad (3)$$

2 - Update of $\mathbf{A}$ using a fixed $\mathbf{S}$:

$$\hat{\mathbf{A}} = \mathbf{A}^{(k-1)} - \frac{\delta}{\|\mathbf{S}^{(k)}\mathbf{S}^{(k)T}\|_2}(\mathbf{A}^{(k-1)}\mathbf{S}^{(k)} - \mathbf{X})\mathbf{S}^{(k)T} \quad (4)$$

$$\mathbf{A}^{(k)} = \Pi_{\|.\|_2}\left(\hat{\mathbf{A}}\right) \quad (5)$$

B. GMCA algorithm

GMCA algorithm is based on projected ALS. At each iteration (k), the gradient step of (2) and (4) is replaced by a multiplication by a pseudo-inverse:

1 - $\mathbf{S}$ is updated assuming a fixed $\mathbf{A}$.

$$\hat{\mathbf{S}} = \mathbf{A}^{(k-1)\dagger}\mathbf{X} \quad (6)$$

Where $\mathbf{A}^{(k-1)\dagger}$ is the pseudo-inverse of $\mathbf{A}^{(k-1)}$.

$$\mathbf{S}^{(k)} = \mathcal{S}_{\mathbf{R_S}}(\hat{\mathbf{S}}\mathbf{\Phi_S^T})\mathbf{\Phi_S} \quad (7)$$

2 - $\mathbf{A}$ is updated assuming a fixed $\mathbf{S}$:

$$\hat{\mathbf{A}} = \mathbf{X}\mathbf{S}^{(k)\dagger} \quad (8)$$

$$\mathbf{A}^{(k)} = \Pi_{\|.\|_2}(\hat{\mathbf{A}}) \quad (9)$$

Compared to PALM, GMCA is only a proxy for the minimization of (1) and cannot be proved to converge (while stabilizing in most of practical cases after some iterations). However, thanks to heuristics, it benefits from an automatic thresholding strategy (see Sec.IV-A1) which has been empirically shown to improve its robustness with respect to the initialization.

III. PALM-based sparse BSS in practice

The objective of this section is to empirically shed light on the lack of robustness of PALM to motivate the need of finding a relevant strategy to enable its use on practical problems. The first subsection presents our experimental protocol, while the second and third ones focus on the sensibility of PALM with respect to the regularization parameters and the initialization.

A. Experimental protocol

To bring out PALM sensibility and the mechanisms at stake, our experiments focus in this subsection on simple data coming from $n = 2$ equilibrated (with equal ℓ_2 norm) sources. The sources are assumed to be exactly sparse in the direct domain, a proportion $p = 0.1$ of the $t = 500$ samples being non-zeros and drawn according to a standard normal distribution. The mixing is performed through a $\mathbf{A}$ matrix drawn randomly following a standard normal distribution and modified to have unit columns. Its condition number is $C_d = 10$. There are $m = n$ observations. To complete the creation of the $\mathbf{X}$ data, a Gaussian additive noise is added to the mixing, such that the Signal to Noise Ratio is SNR = 60 dB.

Once the algorithm launched, it stops based on a convergence criterion computed through the angular distance of the columns of $\mathbf{A}$ between two successive iterations. The separation quality is then mesured using a global criterion on $\mathbf{A}$ [13]: $C_A = -10\log_{10}(\texttt{median}(|\mathbf{P}\mathbf{A}^\dagger\mathbf{A}^*| - \mathbf{I_d}))$, with $\mathbf{A}^*$ the true mixing matrix and $\mathbf{P}\mathbf{A}^\dagger$ the pseudo-inverse of the solution found by the algorithm corrected through $\mathbf{P}$ for the scale and permutation indeterminacies. The higher C_A, the better the separation.

To study PALM robustness, we perform an exhaustive search for the parameters, testing many different values. The feasability of this approach is due to both the fact that we work on simulated data, for which we know the true $\mathbf{A}$ and $\mathbf{S}$, and the low number of sources, asking for only two parameters. Such an approach would be intractable on practical cases.

B. Sensibility to the thresholding strategy

Following our experimental protocol, we obtain Figure 1, on which C_A is displayed as a function of the threshold of each of the 2 sources ((λ_1, λ_2) are part of $\mathbf{R_S}$ and therefore impact PALM through Eq. 3). On this figure, the lack of robustness and sensibility to the parameter choice is underlined by the following remarks:

- High quality separation happens for values close to the diagonal. A slight deviation from it can cause a 30 dB drop for a 3×10^{-4} modification of one parameter and lead to a very bad separation. One could argue that this problem could be circumvented by imposing $\lambda_1 = \lambda_2$. First, it has to be emphasized that the maximum is *not* on the diagonal but slightly outside, for $\lambda_1 = 3.81 \times 10^{-3}$

ISBN 978-90-827970-1-5 © EURASIP 2018

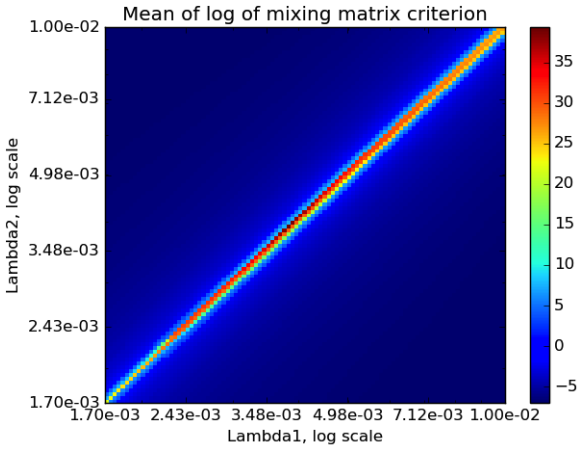

Fig. 1. Mean of C_A as a 2D function of the thresholds for 5 different initializations of PALM algorithm.

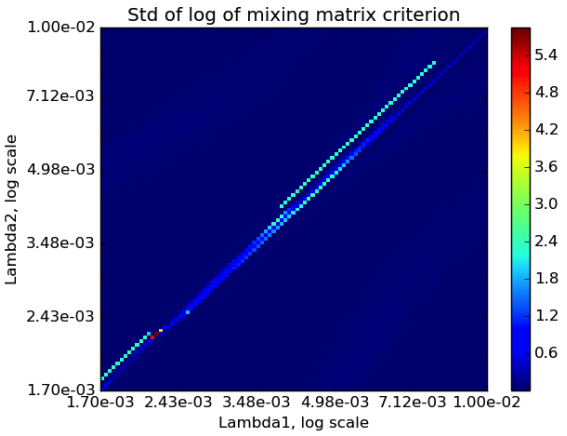

Fig. 2. Standard deviation of C_A as a 2D function of the thresholds for 5 different initializations of PALM algorithm.

and $\lambda_2 = 3.88 \times 10^{-3}$. Merely setting $\lambda_1 = \lambda_2 = 3.81 \times 10^{-3}$ yields a loss of more than 10 dB. Second, in this simple toy example, the sources are equilibrated. It would not necessarily be the case with real-life sources, which could lead to a shift of the diagonal;

- Inside of the diagonal itself, a 2×10^{-3} change of $\lambda_1 = \lambda_2$ can lead to a loss of about 7 dB.

C. Sensibility to the initialization

As described in section I-B, the initialization impacts the results due to the non-convex nature of problem (1) because the algorithm can stay trapped in spurious critical points. The standard deviation of C_A over different initializations of $\mathbf{A}$ is plotted in Fig. 2 to quantify this impact. It can reach up to 5 dB for some regularization parameters choices. The best parameter setting corresponds to a quite high standard deviation and can shift depending on the initialization. It makes the practical choice of an initialization a critical point which directly impact the quality of the local minimum found by the algorithm.

It further highlights the sensitivity to the values of the regularization parameters. A key question is thus *how to tune PALM in practice and if it can be done in an automatic way without prior knowledge about the solution.*

IV. Complexity of introducing heuristics in PALM

A. Heuristic motivation and description

Building on the automatic thresholding strategy of GMCA, the goal of this part is to try to derive one for PALM.

1) Automatic parameter choice in GMCA: In GMCA, the thresholding is directly performed on a minimizer of an approximation of the data fidelity term. Said differently, it is performed in an approximation of the source space. This has lead to a simple interpretation of the parameters, allowing for an heuristic thresholding strategy that yields good practical results and robustness. The true sources being sparse, the thresholding should remove a dense signal, partially coming from the back-projection in the source space of the data noise $\mathbf{N}$. In practice, this dense signal removal is performed computing the Median Absolute Deviation (MAD) for each source and setting the corresponding threshold to a multiple k of this value:

$$(\lambda_1, \lambda_2, ..., \lambda_n)^T = k \times \texttt{MAD}(\hat{\mathbf{S}}) \tag{10}$$

In this equation and in the remaining of this section, $\mathbf{\Phi_S^T}$ was supposed to be the identity matrix without loss of generality. k is a positive number and the MAD operator is computed row-by-row with $\texttt{MAD}(z) = \texttt{median}(|z - \texttt{median}(z)|)$ for $z \in \mathbb{R}^t$. If we suppose that GMCA has found the true $\mathbf{A}$ matrix denoted by $\mathbf{A}^*$, it can be shown that:

$$k \times \texttt{MAD}(\hat{\mathbf{S}}) \simeq k \times \texttt{MAD}\left(\mathbf{A}^{*\dagger}\mathbf{N}\right) \tag{11}$$

It has however to be emphasized that this data noise removal interpretation of the thresholding is limited to the case when the estimation of the mixing matrix is close to $\mathbf{A}^*$. In the opposite case, the imperfect demixing results in an increase of the thresholds due to an extra noise coming from interferences between the sources.

2) Heuristic in PALM: Let us assume that the thresholds in PALM are computed the same way as in GMCA through eq. (10), and that the algorithm has converged to *both* the true matrices $\mathbf{A}^*$ and $\mathbf{S}^*$. Then:

$$\begin{aligned}(\lambda_1, \lambda_2, ..., \lambda_n)^T &= k \times \texttt{MAD}\left(\mathbf{S}^* - \frac{\gamma}{\|\mathbf{A}^{*T}\mathbf{A}^*\|_2}\mathbf{A}^{*T}(\mathbf{A}^*\mathbf{S}^* - \mathbf{X})\right) \\ &\simeq k \times \frac{\gamma}{\|\mathbf{A}^{*T}\mathbf{A}^*\|_2}\texttt{MAD}(\mathbf{A}^{*T}\mathbf{N})\end{aligned} \tag{12}$$

where the last line is obtained thanks to the assumption of $\mathbf{S}^*$ sparsity. Therefore, using the MAD enables a thresholding of a projection of the noise $\mathbf{N}$, which yields a similar interpretation as in GMCA. This observation must however be tempered:

- It only holds when and if PALM has converged towards good $\mathbf{A}$ *and* $\mathbf{S}$;
- The projection is not performed in the same space as GMCA, for which the noise is back-projected in the *source* space. Both spaces however merge when $\mathbf{A}$ is orthogonal.

ISBN 978-90-827970-1-5 © EURASIP 2018

B. Experimental protocol

The mixing and the metric on the separation quality are the same as in Sec. III-A. The automatic parameter setting described in the previous subsection is used inside of a PALM algorithm for different k values. k is the same for both sources, since the MAD is supposed to be adaptive enough to the noise.

C. Empirical performances of the heuristic

The results are displayed in Fig 3. The low values of C_A, reaching at most 6dB, are to be compared with the 39 dB obtained during the exhaustive parameter search in III-A. While efficient in GMCA, the MAD heuristic leads to a bad separation quality due to the different space in which the thresholding is performed (cf. IV-A2), yielding a different role to the thresholds. To better understand this role, let us assume that the algorithm is initialized with the true mixing matrix $\mathbf{A}^*$ *only* (and not with $\mathbf{S}^*$, contrary to what is described in IV-A2). In PALM, the use of the transpose $\mathbf{A}^{*T}$ in eq. (2) leads towards a re-mixing of the sources at each iteration whose goal is to gradually decrease the data fidelity term. This re-mixing however increases the sparsity term and should be impeded by the thresholds, giving to them a demixing role.
On the opposite, when using GMCA, such a re-mixing between the sources does not exist due to the use of the pseudo-inverse $\mathbf{A}^{*\dagger}$ instead of $\mathbf{A}^{*T}$, which alleviate somehow the issue of finding relevant thresholds. Therefore, using the MAD heuristic in PALM does not results in good separations in practice, because the interpretation in terms of noise removal is less relevant than in GMCA.

V. Combining GMCA and PALM: a hybrid strategy

In this part, we propose to combine the best of GMCA and PALM in a two step approach. The algorithm consists in a warm-up stage, in which GMCA is performed, followed by a refinement stage during which PALM is performed retaining as much information as possible coming from the warm-up stage.

A. Motivation and full description of the algorithm

Our approach is motivated by several remarks:

- *PALM theoretical background:* while GMCA is only a proxy, the 2-step algorithm will attempt to solve exactly eq. (1) as PALM does. It further benefit from the high potential accuracy of PALM. Moreover, PALM is proved to converge under mild assumptions [10], so is the 2-step algorithm.
- *GMCA robustness with regards to initialization:* we propose to re-use GMCA outputs $\mathbf{A_{GMCA}}$ and $\mathbf{S_{GMCA}}$ as an initialization of PALM. Due to the almost constant results of GMCA when confronted to different initializations, the quality of the output of PALM in the refinement stage will have a much decreased variance.
- *Benefit from GMCA solution:* the results of GMCA can be re-used by PALM in the refinement stage as a first guess about the shape of the solution. It is done through two mechanisms:
  - Since *both* $\mathbf{A_{GMCA}}$ and $\mathbf{S_{GMCA}}$ are close to $\mathbf{A}^*$ and $\mathbf{S}^*$, they can be used to derive the thresholds using the MAD. It has however to be emphasized that after a given number of iterations the thresholds become fixed, which is a necessary condition to ensure convergence.
  - Furthermore, $\mathbf{S_{GMCA}}$ should already give a good approximation of the most prominent peaks. This information can be exploited in the refinement stage through the introduction of reweighted L1 [8] using the reweighting matrix $\mathbf{W}$ in problem (1). In brief, the thresholds are lowered for the biggest samples of the estimated sources, reducing the bias introduced by soft-thresholding. This is particularly appealing in PALM because the biases in these coefficients behave as errors that are transmitted to the other sources through the remixing introduced by $\mathbf{A}^T$.

These remarks lead to the following algorithm using a smart initialization and thresholding strategy:

Input : $\mathbf{X}$ (data matrix)

- Random initialization $\mathbf{A_0}$ and $\mathbf{S_0}$
- *Warm-up stage*:
  $\mathbf{A_{GMCA}}$, $\mathbf{S_{GMCA}}$ = GMCA($\mathbf{X}$,$\mathbf{A_0}$,$\mathbf{S_0}$)
- *Refinement stage*:
  $\mathbf{A_{PALM}}$, $\mathbf{S_{PALM}}$ = PALM($\mathbf{X}$,$\mathbf{A_{GMCA}}$,$\mathbf{S_{GMCA}}$)
  The initialization, thresholding strategy and reweighting information come from the warm-up stage.

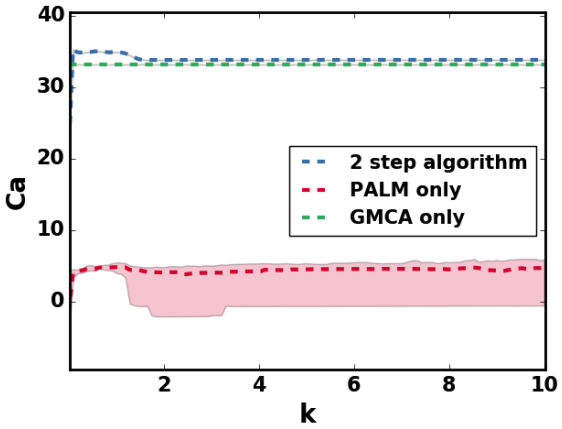

Fig. 3. Mean and quartiles of C_A for 10 different initializations of 3 different algorithms : PALM, GMCA and 2 steps.

B. Empirical results

The experimental protocol is the same as described in Sec. IV-B except that the 2-step algorithm is used instead of PALM. The results are plotted in Fig. 3. With values of C_A higher than 33 dB, the demixing is close to the best ones obtained with the exhaustive search in Sec. III. Compared to PALM only, the variance of results over different initializations is also much decreased, becoming almost zero, which shows the robustness of the algorithm with regards to the initialization.

ISBN 978-90-827970-1-5 © EURASIP 2018

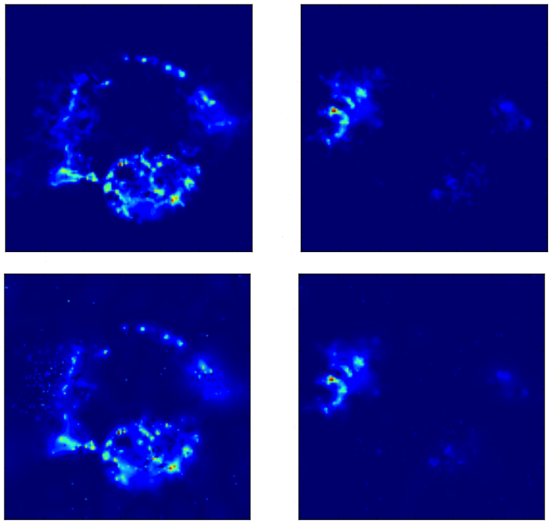

Fig. 4. *Up:* true sources. *Down:* sources estimated by the two step algorithm.

	10 dB	15dB	20 dB
2 step	**11.57**	**16.92**	**22.09**
PALM	9.38	10.94	11.01
GMCA	8.87	12.15	15.87
EFICA	-6.92	5.11	9.41
RNA	-6.68	-5.49	6.27

TABLE I
AVERAGE C_A (10 DIFFERENT INITIALIZATIONS) FOR 3 SNR VALUES AND 5 ALGORITHMS. THE BEST VALUES ARE IN BOLD.

VI. EXPERIMENT ON REALISTIC DATA

The goal of this part is to apply our algorithm to realistic data to show its efficiency.

A. Data description and experimental protocol

The $n = 2$ sources come from simulations obtained from real data of Cassiopeia A supernova remnant. These data originate from the Chandra X-ray observatory. The sources in these wavelength values correspond to the thermal emission and the iron. As displayed in Fig. 4, they each consists in a 2D image of resolution $t = 128 \times 128$ pixels, supposed to be approximately sparse in the starlet domain. The mixing matrix is simulated as described in III-A. To increase the realism of the data and further test the algorithm, we tried three relatively low SNR values: 10, 15 and 20 dB.

The sparsity is enforced *in the starlet domain*. k is set to 3, which corresponds to a standard hypothesis in terms of Gaussian noise removal.

B. Empirical results

The mean of C_A values is displayed in Fig. I. To visually assess the separation quality, the estimation for a SNR of 15 dB is shown in Fig. 4. The 2-step approach always achieve better results than the two classical BSS algorithms with which we performed the comparison, namely Relative Newton Algorithm (RNA) and Efficient FastICA (EFICA). It also outerpeforms both GMCA and PALM, being always better by at least 2 dB than the best of them. The bigger differences with GMCA than in Fig. 3 are probably explained by the fact that the approximation made by GMCA is less well verified due to both a higher noise level and less sparse sources.
In addition to the results displayed in Fig. (I), the standard deviation of C_A over different initializations is almost 0, which shows the robustness of the algorithm.

CONCLUSION

In this work, we show that the ability of recent and theoretically grounded optimization strategies like PALM to provide solutions to sparse BSS problems is highly sensitive to both the initialization and the values of the regularization parameters. To further design an effective as well as robust algorithm, we introduce a 2-step strategy combining PALM with robust heuristic methods such as GMCA. Beyond improving the robustness of PALM-based implementations with respect to initialization, the regularization parameters can be automatically set in the proposed approach. Numerical experiments on both simulated and realistic data demonstrate a high separation quality and good robustness on mixings with low SNR.

ACKNOWLEDGMENT

This work is supported by the European Community through the grant LENA (ERC StG - contract no. 678282), and by the MOD University Defence Research Collaboration (UDRC) for signal processing in the networked battlespace, number P/K014277/1.

REFERENCES

[1] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic Press, 2010.
[2] N. Gillis and F. Glineur, "Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization," *Neural Computation*, vol. 24, no. 4, pp. 1085–1105, 2012.
[3] J. Bobin, J.-L. Starck, Y. Moudden, and M. J. Fadili, "Blind source separation: The sparsity revolution," *Advances in Imaging and Electron Physics*, vol. 152, no. 1, pp. 221–302, 2008.
[4] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.
[5] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Sparse ICA for blind separation of transmitted and reflected images," *International Journal of Imaging Systems and Technology*, vol. 15, no. 1, pp. 84–91, 2005.
[6] Y. Li, S.-I. Amari, A. Cichocki, D. W. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 423–437, 2006.
[7] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press, 2010.
[8] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted 1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.
[9] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of Optimization Theory and Applications*, vol. 109, no. 3, pp. 475–494, 2001.
[10] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.
[11] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function," *Journal of Optimization Theory and Applications*, vol. 162, no. 1, pp. 107–132, 2014.
[12] ——, "A block coordinate variable metric forward–backward algorithm," *Journal of Global Optimization*, vol. 66, no. 3, pp. 457–485, 2016.
[13] J. Bobin, J. Rapin, A. Larue, and J.-L. Starck, "Sparsity and adaptivity for the blind separation of partially correlated sources." *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1199–1213, 2015.

ISBN 978-90-827970-1-5 © EURASIP 2018