



HAL
open science

Revisiting Sparse ICA from a Synthesis Point of View: Blind Source Separation for Over and Underdetermined Mixture

Fangchen Feng, Matthieu Kowalski

► **To cite this version:**

Fangchen Feng, Matthieu Kowalski. Revisiting Sparse ICA from a Synthesis Point of View: Blind Source Separation for Over and Underdetermined Mixture. 2017. hal-01297471v2

HAL Id: hal-01297471

<https://hal.science/hal-01297471v2>

Preprint submitted on 26 Jun 2017 (v2), last revised 22 May 2018 (v3)

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.

Revisiting Sparse ICA from a Synthesis Point of View: Blind Source Separation for Over and Underdetermined Mixture

Fangchen Feng, Matthieu Kowalski

Abstract—This paper studies the existing links between two approaches of Independent Component Analysis (ICA), projection pursuit and Infomax/maximum likelihood estimation, and the Sparse Component Analysis (SCA), mainly used in the Generalized Morphological Component Analysis (GMCA), to tackle the Blind Source Separation (BSS) of instantaneous mixtures problem. If ICA methods suit well for overdetermined and noiseless mixtures, SCA (via GMCA) has demonstrated its robustness to noise. Using the "synthesis" point of view to reformulate ICA methods as an optimization problem, we propose a new optimization framework, which encompasses both approaches. We show that the algorithms developed to minimize the proposed functional built on SCA, but imposing a numerical decorrelation constraint on the sources, aims to improve the Signal to Inference Ratio (SIR) of the estimated sources, without degrading the Signal to Distortion Ratio (SDR).

I. INTRODUCTION

The blind source separation for instantaneous mixtures appears in various applications such as speech processing [1], biomedical processing [2] and digital communications [3]. The BSS problem is also mingled with the independent component analysis (ICA), as this family of methods introduced in 1984 [4] has been developed to tackle the following linear problem [5]: Given M observations of size T $\mathbf{X} \in \mathbb{R}^{M \times T}$, estimate the mixing matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ and the N sources $\mathbf{S} \in \mathbb{R}^{N \times T}$, such that

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E} \quad (1)$$

with $\mathbf{E} \in \mathbb{R}^{M \times T}$ some additive noise.

a) Independent component analysis: Based on the independence assumption, ICA was mainly developed in the (over)determined noiseless case [6] (see references therein for extensions of ICA):

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (2)$$

with $M \geq N$.

The general idea of ICA is to estimate an unmixing matrix $\mathbf{W} = \mathbf{A}^{-1}$ by minimizing a contrast function measuring the dependencies of the sources. This unmixing matrix is then used to estimate the source signals by $\mathbf{S} = \mathbf{W}\mathbf{X}$. ICA methods have been applied with success in a wide range of applications, such as electroencephalography (EEG) [7], functional magnetic resonance imaging (fMRI) [8]–[10], and

audio source separation [11]. One can refer to [6] for a deep presentation of ICA for blind source separation.

A fundamental hypothesis of ICA, is that at most one sources can be Gaussian [6]. In order to apply ICA, it is usual to consider the mixing model (1) in a transform domain, where the "non-Gaussianity" assumption can be respected. Denoting by $\{\varphi_k, \varphi_K \in \mathbb{R}^T\}_{k=1}^K$ a dictionary of waveform (such as wavelets or time-frequency atoms)¹ and $\Phi = [\varphi_1, \dots, \varphi_K] \in \mathbb{R}^{T \times K}$, the mixing model (2) becomes

$$\begin{aligned} \mathbf{X}\Phi &= \mathbf{A}\mathbf{S}\Phi \\ \tilde{\mathbf{X}} &= \tilde{\mathbf{A}}\tilde{\mathbf{S}} \end{aligned} \quad (3)$$

where $\tilde{\mathbf{X}} = \mathbf{X}\Phi$ and $\tilde{\mathbf{S}} = \mathbf{S}\Phi$ are the analysis coefficients of \mathbf{X} and \mathbf{S} respectively in the transform domain. ICA in the transform domain has been proposed for image processing [12] with wavelets, audio source separation with the Short-Time Fourier Transform (STFT) [11] and fMRI [9] where a dictionary learning strategy is used to choose the transform domain. In [12], the authors have remarked that contrast function employed in ICA can be interpreted as a measure of sparsity.

b) Sparsity by analysis: Using the analysis form, sparse methods are based on the mixing model in the transform domain:

$$\begin{aligned} \mathbf{X}\Phi &= \mathbf{A}\mathbf{S}\Phi + \mathbf{E}\Phi \\ \tilde{\mathbf{X}} &= \tilde{\mathbf{A}}\tilde{\mathbf{S}} + \tilde{\mathbf{E}} \end{aligned} \quad (4)$$

with $\tilde{\mathbf{E}}$ the analysis coefficients of \mathbf{E} .

These methods are especially popular to deal with the underdetermined mixtures where the mixing matrix is first estimated, before estimating the sources [13]–[15]. It is also the starting point of time-frequency techniques for BSS applied to audio signals [16], [17].

Nevertheless, when the family of waveforms is overcomplete, several problems appear: first, if the noise \mathbf{E} is assumed to be Gaussian white in the time domain, $\tilde{\mathbf{E}} = \mathbf{E}\Phi$ becomes correlated in the transform domain (its density is even a degenerated Gaussian). Moreover, the estimated coefficients $\tilde{\mathbf{S}}$ used to synthesize the sources do not belong to the image of the operator Φ , thus should not be considered as analysis coefficients.

F. Feng and M. Kowalski are with Laboratoire des Signaux et Systèmes, UMR 8506 Univ Paris-Sud – CNRS – CENTRALESUPELEC, 91192 Gif-sur-Yvette Cedex, France (e-mail: fangchen.feng@u-psud.fr, matthieu.kowalski@u-psud.fr).

¹We stick to the real case for the sake of simplicity, but the dictionary can be complex.

c) *Sparsity by synthesis*: A simple way to deal with these drawbacks of the analysis operator, is to use the synthesis modeling of the sparse component analysis (SCA) [18]:

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E} = \mathbf{A}\boldsymbol{\alpha}\boldsymbol{\Phi}^* + \mathbf{E} \quad (5)$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{N \times K}$ are the synthesis coefficients of \mathbf{S} in the dictionary $\boldsymbol{\Phi}$, assumed to be sparse. The problem is then to estimate the mixing matrix \mathbf{A} and the synthesis coefficients $\boldsymbol{\alpha}$, the sources being synthesized by $\mathbf{S} = \boldsymbol{\alpha}\boldsymbol{\Phi}^*$. Using (5), [19] proposed to estimate \mathbf{A} and $\boldsymbol{\alpha}$ jointly using an alternating optimization strategy based on the maximum *a posteriori* (MAP):

$$\begin{aligned} \max_{\mathbf{A}, \boldsymbol{\alpha}} p(\mathbf{A}, \boldsymbol{\alpha} | \mathbf{X}) &= \min_{\mathbf{A}, \boldsymbol{\alpha}} -\log(p(\mathbf{A}, \boldsymbol{\alpha} | \mathbf{X})) \\ &= \min_{\mathbf{A}, \boldsymbol{\alpha}} -\log(p(\mathbf{X} | \mathbf{A}, \boldsymbol{\alpha})) - \log(p(\mathbf{A})) - \log(p(\boldsymbol{\alpha})) \end{aligned} \quad (6)$$

$$(7)$$

The chosen priors are the following:

- the noise is white Gaussian:

$$-\log(p(\mathbf{A}, \boldsymbol{\alpha} | \mathbf{X})) = \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2$$

- A Gaussian prior on \mathbf{A} :

$$-\log(p(\mathbf{A})) = \frac{\mu}{2} \|\mathbf{A}\|_F^2, \mu > 0$$

- A generalized Gaussian prior on the synthesis coefficients:

$$-\log(p(\boldsymbol{\alpha})) = \frac{\lambda}{\gamma} \sum_{n,k} |\alpha_n(k)|^\gamma, 0 < \gamma \leq 1, \lambda > 0$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Generalized Gaussian prior with $0 < \gamma \leq 1$ allows one to favor sparse coefficients [19] and this MAP approach using sparsity is also the starting point of the generalized morphological component analysis (GMCA) [20] developed for images using wavelet basis. One of the main advantages of such an approach, is its ability to deal with the additive noise compared to ICA (see the discussion in [20]).

d) *Contributions and outline*: As an extension of our previous work [21], we consider in this paper the BSS problem (1) in general: (over)determined and underdetermined, possibly with an additive white Gaussian noise. A fundamental hypothesis of this work is that the sources admit a sparse representation leading to (5).

The contributions of this article are fourfold:

- 1) we studied the existing algorithms based on the MAP approach (5) exploiting sparsity, in particular the GMCA. We proposed a convergent algorithm based on PALM [22], which is more robust in practice, in the sense that it obtains acceptable results in the underdetermined case (Sec. II).
- 2) we provide a discussion on the formal links between some ICA approaches and sparsity-based approaches (Sec. III)
- 3) we proposed a new framework exploiting sparsity and numerical time decorrelation of the sources, which generalizes ICA and SCA for BSS (Sec. IV).

- 4) we compared all these algorithms on synthetic instantaneous audio mixtures (Sec. V). Even if instantaneous mixture is not the right model for audio signals [23], this framework allows us to evaluate the separation results with objective measures, and subjectively by listening. The experiments show that the proposed framework outperforms existing algorithms in the underdetermined case, and is more robust to noise in the (over)determined case. Moreover, the proposed framework is robust to the chosen number of sources to estimate.

II. BSS WITH SYNTHESIS SPARSE MODEL

We recall the mixing model with the synthesis coefficient as defined in (5).

$$\begin{aligned} \mathbf{X} &= \mathbf{A}\mathbf{S} + \mathbf{E} \\ &= \mathbf{A}\boldsymbol{\alpha}\boldsymbol{\Phi}^* + \mathbf{E} \end{aligned} \quad (8)$$

where

- $\mathbf{X} \in \mathbb{R}^{M \times T}$ contains the M mixtures of T samples,
- $\mathbf{S} \in \mathbb{R}^{N \times T}$ are the N sources of T samples.
- $\mathbf{E} \in \mathbb{R}^{M \times T}$ is the additive noise, assumed to be white and Gaussian.
- $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the mixing matrix.
- $\boldsymbol{\Phi} \in \mathbb{R}^{K \times T}$ is the matrix dictionary of waveforms.
- $\boldsymbol{\alpha} \in \mathbb{R}^{N \times K}$ are the synthesis coefficients of \mathbf{S} .

In this article, we stick to the simple model of the white Gaussian noise. If the problem of spatially correlated noise has been widely studied and can be tackled by a whitening step (see, for example, [24]), taking into account spatially and temporally correlated noise is more complex [25]. We let the study of non white noise in inverse problems to further works, as it is out of the scope of this article, dedicated to the instantaneous BSS.

In this section, we provide a brief state of the art of the existing SCA approaches for BSS. We then present a convergent algorithm thanks to the proximal alternating linearized method (PALM) [22].

A. The optimization framework

In order to jointly estimate the mixing matrix \mathbf{A} and the synthesis coefficients $\boldsymbol{\alpha}$ in (5), the authors of [19] proposed to solve the optimization problem of the following functional, which is equivalent to the MAP of (7)

$$\min_{\mathbf{A}, \boldsymbol{\alpha}} \frac{1}{2} \|\mathbf{X} - \mathbf{A}\boldsymbol{\alpha}\boldsymbol{\Phi}^*\|_F^2 + \lambda h(\boldsymbol{\alpha}) + g(\mathbf{A}) \quad (9)$$

where

- $f(\mathbf{A}, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{X} - \mathbf{A}\boldsymbol{\alpha}\boldsymbol{\Phi}^*\|_F^2$ is the data fit term, corresponding to the white Gaussian noise prior.
- h is the regularization term employed to favor sparse solution. A popular choice is the ℓ_1 norm [26].
- $\lambda > 0$ is a hyperparameter balancing between the data term and the regularization term.

- g contains constraints on \mathbf{A} to avoid trivial solutions and limit the ambiguity scaling problem. A common choice for g is the indicator function of the unit circle:

$$g(\mathbf{A}) = \iota_{\mathcal{C}}(\mathbf{A}) = \begin{cases} 0 & \text{if } \|\mathbf{a}_n\| = 1, n = 1, 2, \dots, N \\ +\infty & \text{otherwise} \end{cases} \quad (10)$$

where \mathbf{a}_n is the n -th column of \mathbf{A} .

B. State of the art algorithms

The optimization problem (9) is non-differentiable and non-convex. In [19], the authors proposed to use a smooth relaxation of the ℓ_1 norm to solve the problem. However, [27] shows that the smooth technique has several drawbacks, mainly because of the choice of the smooth parameter which balances between the convergence rate and the approximation level. It is also pointed out in [19] that the separation results are sensitive to initialization. In [28], the GMCA method was developed based on the alternating optimization strategy to solve the problem for image separation in (over)determined setting. They first perform the optimization with respect to the signal and then with respect to the mixing matrix followed by a normalization step. One limitation of GMCA is that this block-coordinate descend-like algorithm does not have any convergence proof mainly because of the extra normalization step. It is also mentioned in [28] that GMCA does not work in underdetermined case.

C. A convergent algorithm: BSS-PALM

We propose to use the proximal alternating linearized method (PALM) [22] to solve the optimization problem (9). We show that the sequence generated by this algorithm converges to a critical point of problem (9). The experiment Section V will show some advantages of using PALM instead of GMCA, such as its relative robustness to deal with underdetermined mixtures, and its robustness to random initialization.

PALM relies on the proximal operator defined as follows [29]:

$$\text{prox}_{\Psi} := \underset{\mathbf{X}}{\text{argmin}} \Psi(\mathbf{X}) + \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2 \quad (11)$$

The proposed algorithm is based on two special cases of the proximal operator. When $\Psi(\mathbf{X}) = \lambda \|\mathbf{X}\|_1$, the proximal operator is an element-wise soft-thresholding:

$$\mathcal{S}_{\lambda}(\mathbf{X}) = \text{prox}_{\Psi}(\mathbf{X}) = \text{sign}(x) (|x| - \lambda)^+ \quad (12)$$

where $\text{sign}(x)$ is the sign function and $(x)^+ = \max(x, 0)$.

When $\Psi(\mathbf{X}) = \iota_{\mathcal{C}}(\mathbf{X})$ as defined in (10), the proximal operator reduces to a column-wise normalization projection:

$$\mathcal{P}_{\mathcal{B}}(\mathbf{X}) = \text{prox}_{\Psi}(\mathbf{X}) = \begin{cases} \frac{\mathbf{x}_n}{\|\mathbf{x}_n\|} & \text{if } \|\mathbf{x}_n\| \neq 0 \\ \mathbf{x}_n & \text{otherwise} \end{cases} \quad (13)$$

where \mathbf{x}_n is the n -th column of \mathbf{x} .

With the above proximal operators, we present the proposed algorithm in Alg. 1. In the following, we refer to this algorithm as BSS-PALM.

Using a direct application of PALM to solve (9), one can prove the next proposition

Algorithm 1: BSS-PALM

Initialization : $\alpha_1 \in \mathbb{R}^{N \times K}$, $\mathbf{A}_1 \in \mathbb{R}^{M \times N}$,
 $L_{1,1} = \|\mathbf{A}_1\|_F^2$, $L_{1,2} = \|\alpha_1 \Phi^*\|_F^2$, $j = 1$;

repeat

- 1) $\nabla_{\alpha_j} f(\mathbf{A}_j, \alpha_j) = -\mathbf{A}_j^T (\mathbf{X} - \mathbf{A}_j \alpha_j \Phi^*) \Phi$;
- 2) $\alpha_{j+1} = \mathcal{S}_{\lambda/L_{j,1}} \left(\alpha_j - \frac{1}{L_{j,1}} \nabla_{\alpha} Q(\mathbf{A}_j, \alpha_j) \right)$;
- 3) $\nabla_{\mathbf{A}_j} f(\mathbf{A}_j, \alpha_{j+1}) = -(\mathbf{X} - \mathbf{A}_j \alpha_{j+1} \Phi^*) \Phi \alpha_{j+1}^*$;
- 4) $\mathbf{A}_{j+1} = \mathcal{P}_{\mathcal{C}} \left(\mathbf{A}_j - \frac{1}{L_{j,2}} \nabla_{\mathbf{A}} f(\mathbf{A}_j, \alpha_{j+1}) \right)$;
- 5) $L_{j+1,1} = \|\mathbf{A}_{j+1}\|_F^2$;
- 6) $L_{j+1,2} = \|\alpha_{j+1} \Phi^*\|_F^2$;
- 7) $j = j + 1$;

until convergence;

Proposition 1. *The sequence (\mathbf{A}_j, α_j) generated by Alg. 1 converges to a critical point of problem (9).*

Details of the proof are postponed in Appendix A.

III. REVISITING SPARSE ICA

Recently, it was claimed in [30] that two of the most used ICA methods for fMRI (Infomax and FastICA) separate sparse sources rather than independent sources, leading to the conclusion that the mathematical design of better analysis tool for brain fMRI should emphasize on other characteristics, such as sparsity, rather than independence. One given explanation is that the sparsity-based ℓ_1 minimization can be connected with InfoMax and FastICA, because both of these ICA methods implicitly assume that the independent components have the generalized Gaussian distribution, which includes the sparse sources modeled by ℓ_1 minimization. This conclusion is balanced in [31] where the authors show that these two algorithms are indeed relevant to the recovery of independent fMRI sources. However, connections between ICA and SCA appear several times. In [32], it is shown that ICA methods work better in transform domain such as Curvelets or Ridgelets, and justify the use of kurtosis in ICA by a sparse coding point of view. This remark was already made in [12] at the end of 90's. In [13], [14], the estimation of the mixing matrix for underdetermined mixture is performed by exploiting the sparsity of the sources in the transform domain, and the independence can be viewed as a consequence of sparsity.

Back to the ICA direct model (2), we suppose in this section that $M = N$. As the sources \mathbf{S} are assumed to be independent, ICA methods take the uncorrelation constraint $E(\mathbf{S}\mathbf{S}^T) = \mathbf{D}$, \mathbf{D} being a diagonal matrix, into account. In practice, as the number of samples is large, only the numerical decorrelation is considered, that is $\mathbf{S}\mathbf{S}^T = \mathbf{D}$. Furthermore, ICA even assumes $\mathbf{D} = \mathbf{I}$, thanks to a whitening step. Another common point of ICA algorithms, is to look for an estimate \mathbf{W} of the inverse of \mathbf{A} such that $\mathbf{S} = \mathbf{W}\mathbf{X}$. After presenting maximum likelihood/Infomax and projection pursuit as optimization problem, we propose a new functional to encompass these ICA approaches and SCA, allowing us to deal with noisy mixtures as well as underdetermined mixtures.

A. ICA based on Infomax/Maximum likelihood estimation

These methods aim to estimate \mathbf{W} by maximizing the amount of mutual information or the likelihood of the model. Both lead to:

$$\max_{\mathbf{W}} E \left(\sum_{n=1}^N \log(p_n(\mathbf{w}_n^T \mathbf{X})) \right) + \log(\det \mathbf{W}) \quad (14)$$

with p_n the density of the independent components, under the orthogonality constraint on \mathbf{W} : $\mathbf{W}\mathbf{W}^T = \mathbf{W}^T\mathbf{W} = \mathbf{I}$

An important point of [19] is the equivalence of their proposed approach (9) and the problem (14) in the determined noiseless scenario (3), Φ being an orthogonal basis. Indeed, when Φ is an orthogonal basis, there is an equivalence between the synthesis coefficients α such that $\mathbf{S} = \alpha\Phi^*$ and the analysis coefficients $\tilde{\mathbf{S}}$. Then, one has

$$\tilde{\mathbf{X}} = \mathbf{A}\tilde{\mathbf{S}} = \mathbf{A}\alpha \quad (15)$$

with $\tilde{\mathbf{X}} = \mathbf{X}\Phi$ being the analysis coefficients of \mathbf{X} . Denoting by $\mathbf{W} = \mathbf{A}^{-1}$, (15) becomes

$$\mathbf{W}^{-1}\tilde{\mathbf{X}} = \tilde{\mathbf{S}} = \alpha \quad (16)$$

Then, by re-injecting (16) into (9), the optimization becomes

$$\min_{\tilde{\mathbf{S}}, \mathbf{W}} \frac{1}{2} \|\mathbf{W}\tilde{\mathbf{X}} - \tilde{\mathbf{S}}\|_F^2 + \lambda h(\tilde{\mathbf{S}}) + g(\mathbf{W}) . \quad (17)$$

Then, with the choices $g(\mathbf{W}) = -N \log(|\det \mathbf{W}|)$ and $h(\tilde{\mathbf{S}}) = h(\mathbf{W}^T \tilde{\mathbf{X}}) = \mathbb{E} \left(\sum_{n=1}^N \log(p_n(\mathbf{w}_n^T \mathbf{X})) \right)$, one recovers the objective (14) in the noiseless scenario $\mathbf{W}\tilde{\mathbf{X}} = \tilde{\mathbf{S}}$.

However, as already stressed in the introduction, there is no warranty to recover real analysis coefficients $\tilde{\mathbf{S}}$ such that there exist \mathbf{S} verifying $\tilde{\mathbf{S}} = \mathbf{S}\Phi$, unless Φ is an orthogonal basis. That is, this equivalence shows that the synthesis sparse approach (9) can be viewed as a generalization of the Infomax/maximum likelihood techniques able to deal with noisy and underdetermined mixtures, by working on the synthesis coefficients α , such that $\mathbf{S} = \alpha\Phi^*$.

B. ICA based on projection pursuit

Projection pursuit methods aim to identify the N components of the mixture, by estimating the weight vectors \mathbf{W} which maximize a measure of non gaussianity and assuring the decorrelation constraint $E(\mathbf{S}\mathbf{S}^T) = \mathbf{I}$. That is, one can re-formulate ICA by projection pursuit as:

$$\max_{\mathbf{W}} J(\mathbf{W}\mathbf{X}) \text{ s.t } \mathbf{W}\mathbf{W}^T = \mathbf{W}^T\mathbf{W} = \mathbf{I} \quad (18)$$

where J is a measure of non-gaussianity. Then, by a simple change of variable, one can reformulate (18) on the analysis coefficients as

$$\max_{\mathbf{A}, \tilde{\mathbf{S}}} J(\mathbf{S}) \text{ s.t. } \tilde{\mathbf{S}}\tilde{\mathbf{S}}^T = \mathbf{I} \text{ and } \tilde{\mathbf{X}} = \mathbf{A}\tilde{\mathbf{S}} \quad (19)$$

A possible choice for J is the kurtosis of the coefficients. That is

$$J(\tilde{\mathbf{S}}) = \sum_{n=1}^N \text{Kurt}(\tilde{s}_n) = \sum_{n=1}^N \frac{\sum_n |\tilde{s}_n|^4}{(\sum_n |\tilde{s}_n|^2)^2} .$$

Other choices are possible, such as smooth approximation of the ℓ_0 norm as proposed in [10] or proposed neg-entropy approximations used in FastICA (see [6]).

Once again, if Φ is an orthogonal basis in (9), one has an equivalence between (19) and (9) in the noiseless case, with the choice $h(\tilde{\mathbf{S}}) = J(\tilde{\mathbf{S}})$, s.t. $\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T = \mathbf{I}$.

C. Toward a general formulation of sparse ICA

We have seen that, when Φ is an orthogonal basis, some approaches of ICA can be formulated as the optimization problem (9) in the noiseless case. Moreover, using the maximum a posteriori (MAP) interpretation of the optimization problem (9), the regularization term h can reflect the statistical independence of the synthesis coefficients between the sources. Indeed, using the MAP approach in a Bayesian setting, the independence assumption reflected in the prior on α leads to

$$h(\alpha) = \sum_{n=1}^N h_n(\alpha_n) .$$

Then, in order to deal with noise, we propose the following generalization of "sparse ICA"

$$\begin{cases} \min_{\mathbf{A}, \alpha} \frac{1}{2} \|\mathbf{X} - \mathbf{A}\alpha\Phi^*\|_F^2 + h(\alpha) + g(\mathbf{A}) \\ \text{s.t. } \mathbf{S} = \alpha\Phi^* \text{ and } \mathbf{S}\mathbf{S}^T = \mathbf{D} \end{cases} \quad (20)$$

where $\mathbf{D} \in \mathbb{R}^{N \times N}$ is any diagonal matrix.

Using the ℓ_1 sparse regularization term for h implies that one considers that the synthesis coefficients of the sources are independent. However, it is interesting to note that maximizing the kurtosis is equivalent to minimizing the $\frac{\ell_p}{\ell_q}$ sparse criterion as defined in [33] with $p = 2$ and $q = 4$, used for blind deconvolution.

In this paper, for the sake of simplicity, we stick to ℓ_1 norm for the synthesis coefficients. Therefore we consider the formulation as follows:

$$\begin{cases} \min_{\mathbf{A}, \alpha} \frac{1}{2} \|\mathbf{X} - \mathbf{A}\alpha\Phi^*\|_F^2 + \lambda \|\alpha\|_1 + \iota_{\mathcal{C}}(\mathbf{A}) \\ \text{s.t. } \alpha\Phi^*\Phi\alpha^* = \mathbf{D} \end{cases} \quad (21)$$

IV. ALGORITHMS FOR SPARSE ICA

In this section, we propose three algorithms to solve the sparse ICA problem (21). First, we show that a slight modification of the BSS-PALM algorithm ensures that the estimated sources are decorrelated in the determined noiseless scenario. Then we propose an ADMM type algorithm and a simplified version to tackle the problem.

A. A simple modification of BSS-PALM

The optimization problem (21) is challenging because of the non-convex term and the quadratic constraint. However, we demonstrate that by modifying the update stage for the mixing matrix \mathbf{A} in BSS-PALM, simply by a pseudo inverse supposing that the sources are decorrelated, the decorrelation constraint is respected for determined noiseless mixtures. This algorithm was build on the block-coordinate variable metric

forward backward (BC-VMFB) method [34] to solve the following version of (9):

$$\min_{\alpha, \mathbf{A}} \frac{1}{2} \|\mathbf{X} - \mathbf{A}\alpha\Phi^*\|_2^2 + \frac{\mu}{2} \|\mathbf{A}\|_2^2 + \lambda \|\alpha\|_1 + \iota_{\mathcal{C}}(\mathbf{A}) \quad (22)$$

where μ is assumed to be very large. Adding the term $\frac{\mu}{2} \|\mathbf{A}\|_2^2$ has of course no influence on the solution, because of the constraint $\iota_{\mathcal{C}}(\mathbf{A})$. However, it has numerical consequences, letting one to replace the estimation of \mathbf{A} by a pseudo inverse supposing that the sources are decorrelated. This algorithm, referred as BSS-Regu in the following, is summarized in Alg. 2.

Algorithm 2: BSS-Regu

Initialization : $\alpha_1 \in \mathbb{R}^{N \times K}$, $\mathbf{A}_1 \in \mathbb{R}^{M \times N}$, $L_1 = \|\mathbf{A}_1\|_F^2$,
 $j = 1$;
repeat
 1) $\nabla_{\alpha} f(\mathbf{A}_j, \alpha_j) = -\mathbf{A}_j^T (\mathbf{X} - \mathbf{A}_j \alpha_j \Phi^*) \Phi$;
 2) $\alpha_{j+1} = \mathcal{S}_{\lambda/L_j}(\alpha_j - \frac{\nabla_{\alpha} f(\mathbf{A}_j, \alpha_j)}{L_j})$;
 3) $\mathbf{A}_{j+1} = \mathcal{P}_{\mathcal{C}}(\mathbf{X} \alpha_j \Phi^*)$;
 4) $L_{j+1} = \|\mathbf{A}_{j+1}\|_F^2$;
 5) $j = j + 1$;
until convergence;

Next propositions show that this algorithm still converges, and that the estimated sources are decorrelated for (over)determined mixtures in the noiseless case.

Proposition 2. *The sequence (α_j, \mathbf{A}_j) generated by Algorithm 2 converges to a critical point of (22)*

Proof. The proof is postponed in Appendix B □

Proposition 3. *Let α be a minimizer of (22) obtained by Alg. 2, and let $\mathbf{S} = \alpha\Phi^*$. Then, if $M \geq N$, $\mathbf{S}\mathbf{S}^T = \mathbf{D}$ as $\lambda \rightarrow 0$, where \mathbf{D} is a diagonal matrix.*

Proof. The proof is postponed in Appendix C. □

B. An ADMM-type approach for Sparse ICA

In this subsection, we present two algorithms to directly solve the problem (20). We use an ADMM approach before providing a simplified version, which appears to be faster and more robust in practice.

1) *ADMM approach:* We first reformulate problem (20) with a linear constraint by introducing an extra variable as follows:

$$\begin{cases} \arg \min_{\mathbf{A}, \alpha, \mathbf{S}} \frac{1}{2} \|\mathbf{X} - \mathbf{A}\alpha\Phi^*\|_F^2 + \lambda \|\alpha\|_1 + \iota_{\mathcal{B}}(\mathbf{A}) + \iota_{\text{deco}}(\mathbf{S}) \\ \text{s.t. } \mathbf{S} = \alpha\Phi^* \end{cases} \quad (23)$$

where ι_{deco} is an indicator function of the decorrelation constraint of \mathbf{S} , reading:

$$\iota_{\text{deco}}(\mathbf{S}) = \begin{cases} 0 & \text{if } \mathbf{S}\mathbf{S}^T = \mathbf{D}, \text{ with } \mathbf{D} \text{ being diagonal} \\ +\infty & \text{otherwise} \end{cases} \quad (24)$$

We then apply the linearized preconditioned alternating direction method of multipliers (LPADMM) [35] to solve the problem (23).

The general idea of ADMM is based on the alternating optimization of the corresponding augmented Lagrangian function. The linearized and preconditioned version (LPADMM) is intended to simplify the algorithm, that is, in each iteration, instead of minimizing the original function, to minimize its first order approximation. While used in practice for some non-convex problem (for example [36]), the convergence of ADMM algorithms in a non-convex setting is currently under study [37], [38]. The derivation of this algorithm is postponed in Appendix D. We refer to it as BSS-LPADMM in the following and is summarized in Alg. 3.

Algorithm 3: BSS-LPADMM

Initialization : $\alpha_1 \in \mathbb{R}^{N \times K}$, $\mathbf{S}_1 \in \mathbb{R}^{N \times T}$, $\mathbf{A}_1 \in \mathbb{R}^{M \times N}$,
 $L_{1,1} = \|\mathbf{A}_1\|_F^2 + \gamma$, $L_{1,2} = \|\alpha_1\Phi^*\|_F^2$, $j = 1$;
repeat
 1) $\nabla_{\alpha_j} f(\mathbf{A}_j, \alpha_j) = -\mathbf{A}_j^T (\mathbf{X} - \mathbf{A}_j \alpha_j \Phi^*) \Phi - \eta_j \Phi - \gamma(\mathbf{S}_j - \alpha_j \Phi^*) \Phi$
 2) $\alpha_{j+1} = \mathcal{S}_{\lambda/L_{j,1}}(\alpha_j - \frac{\nabla_{\alpha} f(\mathbf{A}_j, \alpha_j)}{L_{j,1}})$;
 3) $\mathbf{S}_{j+1/2} = \alpha_{j+1} \Phi^* - \eta_{j+1}/\gamma$;
 4) $\Sigma_{\mathbf{S}_{j+1/2}} = \mathbf{S}_{j+1/2} \mathbf{S}_{j+1/2}^T$;
 5) $\mathbf{W}_{\mathbf{S}_{j+1/2}} = \text{diag}(\Sigma_{\mathbf{S}_{j+1/2}}^{1/2}) \Sigma_{\mathbf{S}_{j+1/2}}^{-1/2}$;
 6) $\mathbf{S}_{j+1} = \mathbf{W}_{\mathbf{S}_{j+1/2}} \mathbf{S}_{j+1/2}$;
 7) $\nabla_{\mathbf{A}_j} f(\mathbf{A}_j, \alpha_{j+1}) = -(\mathbf{X} - \mathbf{A}_j \alpha_{j+1} \Phi^*) \Phi \alpha_{j+1}^*$;
 8) $\mathbf{A}_{j+1} = \mathcal{P}_{\mathcal{C}}(\mathbf{A}_j - \frac{1}{L_{2,j}} \nabla_{\mathbf{A}_j} f(\mathbf{A}_j, \alpha_{j+1}))$;
 9) $\eta_{j+1} = \eta_j + \gamma(\mathbf{S}_{j+1} - \alpha_{j+1} \Phi^*)$
 10) $L_{j+1,1} = \|\mathbf{A}_{j+1}\|_F^2 + \gamma$;
 11) $L_{j+1,2} = \|\alpha_{j+1} \Phi^*\|_F^2$;
 12) $j = j + 1$;
until convergence;

2) *A simplified version:* The BSS-LPADMM algorithm solves directly the problem (20) but is subject to big computational burden. Therefore, we design here a simplified version in Alg. 4 and refer to this algorithm as BSS-Deco.

Algorithm 4: BSS-Deco

Initialization : $\alpha_1 \in \mathbb{R}^{N \times K}$, $\mathbf{A}_1 \in \mathbb{R}^{M \times N}$, $L_1 = \|\mathbf{A}_1\|_F^2$,
 $j = 1$;
repeat
 1) $\alpha_{j+1} = \mathcal{S}_{\lambda/L_j}(\alpha_j - \frac{\nabla_{\alpha} f(\mathbf{A}_j, \alpha_j)}{L_j})$;
 2) $\mathbf{S}_{j+1/2} = \alpha_{j+1} \Phi^*$;
 3) $\mathbf{S}_{j+1} = \text{diag}(\Sigma_{\mathbf{S}_{j+1/2}}^{1/2}) \Sigma_{\mathbf{S}_{j+1/2}}^{-1/2} \mathbf{S}_{j+1/2}$;
 4) $\mathbf{A}_{j+1} = \mathcal{P}_{\mathcal{C}}(\mathbf{X} \mathbf{S}_{j+1/2}^T)$;
 5) $L_{j+1} = \|\mathbf{A}_{j+1}\|_F^2$;
 6) $j = j + 1$;
until convergence;

Compared to BSS-LPADMM, we set the dual variable η and the penalty parameter γ to zero. For the update stage of the mixing matrix, we use the same strategy of BSS-Regu

however with the decorrelated version of the estimated source \mathbf{Y} . Despite the lack of convergence proof, experiments support its good performances.

V. EXPERIMENTS

After presenting the experimental setup, we discuss the choice of the hyper-parameter λ and the robustness to the choice of the number of unknown sources. We then compare all the proposed algorithms and state-of-the-art ICA and SCA algorithms on over/underdetermined mixtures with and without additive white Gaussian noise.

A. Experimental setup

The algorithms are evaluated on mixtures created with 10 sets of signals used in [39], issued from the SiSEC2011 database [40], with a sample rate at 11 kHz and a duration of 6 s. The mixing matrix was generated randomly following a normal distribution with normalized columns. The STFT was computed with half-overlapping tight Hann window of 512 samples length (about 46.5 ms) using the LTFAT toolbox [41]. All the algorithms are initialized randomly following a normal distribution.

The separation performances were assessed using the Signal to Distortion Ratio (SDR) and Signal to Interference Ratio (SIR) [42]. The SDR indicates the overall quality of each estimated source compared to the target, while the SIR reveals the amount of residual crosstalk from the other sources. A larger value of SDR/SIR means a better quality of separation.

To show the improvement brought by the perfect knowledge of the mixing matrix, we design two "non blind" oracle settings for the proposed algorithms for comparison. These two oracles are denoted by BSS-Oracle (corresponding to the non blind version of BSS-PALM, BSS-Regu and BSS-Deco) and BSS-LPADMM-Oracle (corresponding to the non blind version of BSS-LPADMM-Oracle).

Finally, the parameter γ for BSS-ADMM was set empirically to $\gamma = 0.05$.

B. Choice of the hyper-parameter λ

Several methods have been studied for automatic choice of λ in inverse problem, such as projected GSURE (see [43] and references therein) or SUGAR [44]. However, most of the proposed methods implies to compute several solutions for several λ , and then choose the "best" solution according to some criteria. If such a blind method is needed for some applications, it can also be required to let the user decide what is the "best" solution. Particularly for signal (audio, image, video...) restoration, the best acceptable result will not always fit any "objective" criteria. Such a discussion can be found for example in [45] for audio signal.

We stick in this article to simple choices for the hyper-parameter:

- If the mixture is assumed to be free of noise, we choose $\lambda \rightarrow 0$ in order to do not perform any "denoising" on the estimation. In practice, for small value of λ , we used the continuation trick also known as warm-start or fixed-point

continuation [46]: we first run the algorithm with a large value of λ , and then iteratively decreased the parameter till the wanted value, while the algorithm is initialized with the result of the previous run.

- If some noise is added, then we choose the λ giving the best results in term of SDR for each algorithm. This choice is rarely the best choice from a subjective point of view, and cannot be automatically done in practice. However, it appears to be the most fair, by giving the best achievable result from a SDR point of view (the SDR giving the overall quality of each estimated source compared to the target).

In addition to this two "default" choices, we also provide a short discussion about the influence of λ on the results.

C. Robustness to the number of sources

Authors of [47] show that if one has only an upper bound of the number of sources, acceptable estimation of the source signals can still be obtained by the analysis sparsity minimization with ℓ_1 norm. They showed that the extra source channels will contain little energy thus do not have obvious negative effect on the source estimation. The following experiments show that the proposed algorithms have similar behaviors regarding the robustness to the number of sources.

For the sake of simplicity, we only show the results obtained in the noiseless setting for the proposed BSS-Deco in Alg. 4. Other algorithms lead to similar results, or when the mixture is corrupted by some white Gaussian. In these experiments, the number of microphones was set to $M = 3$, the real number of sources N_r and its upper bound N varied from 2 to 5.

We employed the simple strategy presented in Alg. 5 to eliminate extra source channels during the iterations.

Algorithm 5: BSS-Deco with the number of sources determination

```

repeat
    1) Update  $\alpha$ ,  $\mathbf{S}$  and  $\mathbf{A}$  according to one iteration of
       Algo. 4;
    2) if  $E_{I_M}/E_{I_m} > \epsilon$  then
        | Eliminate  $\mathbf{S}_{I_m}$ ,  $\alpha_{I_m}$  and  $\mathbf{A}_{I_m}$ 
    3)  $j = j + 1$ ;
until convergence;

```

In Algo. 5, I_M and I_m are the source index which correspond to the source of maximum and minimum energy respectively. E_{I_M} and E_{I_m} are the corresponding energy. ϵ is a threshold which is empirically fixed to 3.5 in our experiments. Fig. 1 displays the separation results and support the robustness of the proposed algorithm to the choice of the number of sources.

D. Over-determined BSS

In this setting, the number of sources is fixed to $N = 3$ and the number of microphones runs from 3 to 10 (3, 5, 7 and 10).

TABLE I

PERFORMANCES OF DIFFERENT ALGORITHMS IN NOISELESS (OVER)DETERMINED SETTING FOR $N = 3$ (SDR / SIR). ON A LINE, THE BEST PERFORMANCE IS IN BLACK BOLD. IF THE DIFFERENCE BETWEEN A PERFORMANCE AND THE BEST IS LESS THAN 1 DB, IT IS DISPLAYED IN GRAY BOLD.

	BSS-PALM	BSS-Regu	BSS-ADMM	BSS-Deco	GMCA	EFICA	SOBI
$M = 3$	17.0 / 17.0	45.0 / 45.0	41.6 / 41.6	45.8 / 45.8	17.8 / 17.8	-	-
$M = 5$	19.4 / 19.4	45.8 / 45.8	44.3 / 44.3	46.0 / 46.0	18.0 / 18.0	-	-
$M = 6$	25.6 / 25.6	46.0 / 46.0	45.7 / 45.7	46.1 / 46.1	22.8 / 22.8	-	-
$M = 7$	29.9 / 29.9	46.7 / 46.7	45.6 / 45.6	46.9 / 46.9	27.5 / 27.5	-	-
$M = 10$	39.3 / 39.2	47.5 / 47.7	48.2 / 48.2	47.6 / 47.9	38.3 / 38.3	-	-
With whitening	66.0 / 66.0	48.4 / 48.4	48.4 / 48.4	48.3 / 48.3	66.6 / 68.4	49.5 / 49.5	37.3 / 37.3

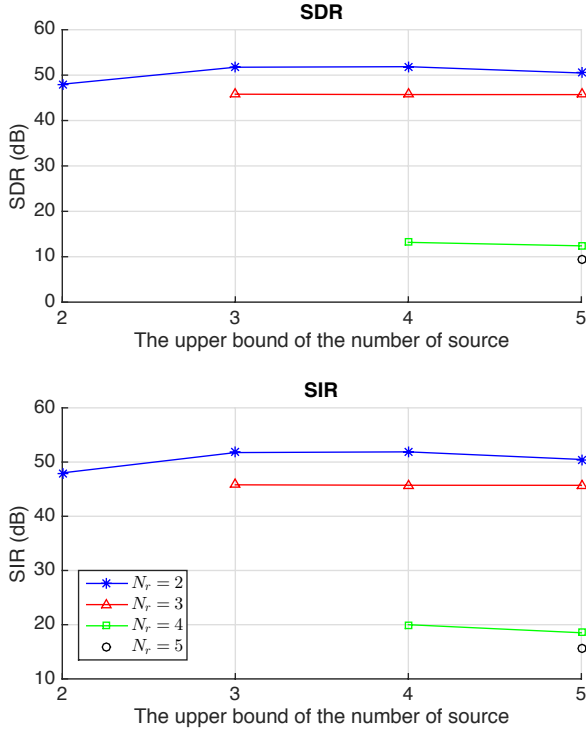


Fig. 1. Performances as a function of the upper bound of number of sources for different cases

1) *Noiseless case*: As reference for ICA approaches, we provide the results obtained by EFICA [48] and second order blind identification (SOBI) [49] using the toolbox [50].

In the noiseless case, the whitening and dimension reduction used in ICA is justified. We therefore present the results obtained by all the algorithms – except for classical ICA approach – with and without this pre-processing step. The results are summarized in Table I.

It is clear that the pre-processing step greatly improves the results for BSS-PALM and GMCA, which outperform other approaches, while their performances were the worst without this pre-processing step. However, it is also interesting to note that the performances improve with the number of observation when such a pre-processing step is not used.

2) *Noisy case*:

a) *Performance as a function of input SNR*: In this experiment, we stick to the determined case, i.e. $M = 3$. Results are summarized on Fig. 2.

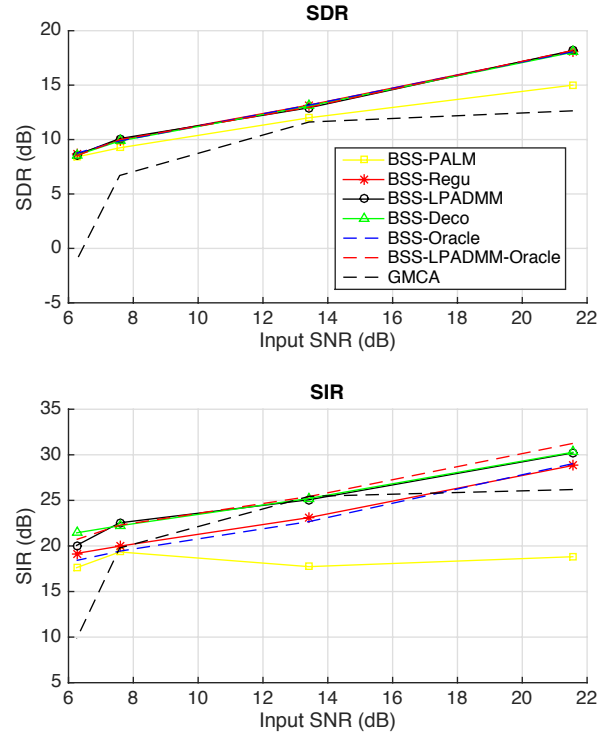


Fig. 2. Performances as a function of the input SNR in determined scenario ($M = 3, N = 3$)

Again, BSS-Regu, BSS-LPADMM and BSS-Deco perform similarly in terms of SDR and outperform BSS-PALM especially with high input SNR. In terms of SIR, as expected, BSS-LPADMM and BSS-Deco lead to the best performances, comparable with BSS-LPADMM-Oracle.

One of the most remarkable results, is that the two oracle settings perform similarly as their corresponding "blind" algorithms in terms of SDR, while BSS-PALM and BSS-oracle obtain the two worst SIR. This last point supports the intuition that the decorrelation constraint is particularly important to improve the SIR.

b) *Performance as a function of the number of observation*: In this experiment, a white Gaussian noise is added to reach an input SNR of 7.58 dB. The results are summarized on Fig. 3.

Similar remarks as previous paragraph can be made: BSS-LPADMM and BSS-Deco reach the best performances, comparable to BSS-LPADMM Oracle, for both SDR and SIR,

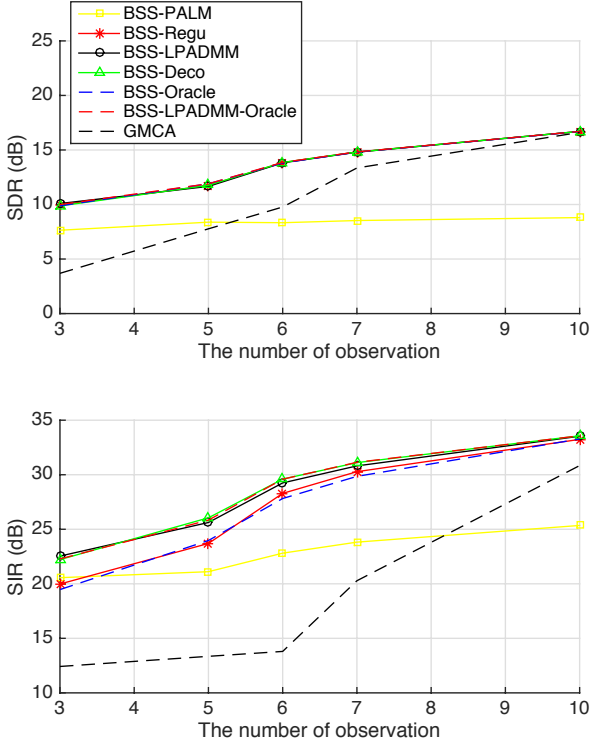


Fig. 3. Performances as a function of the number of microphones in (over)determined noisy scenario with $N = 3$ and input SNR equaling 7.6 dB

while BSS-Oracle show bad performances in term of SIR. As expected, SDR and SIR grows with M for all algorithms.

c) *Performance as a function of sparsity level:* The hyperparameter λ in the proposed algorithms is linked to the variance of the input noise and controls the sparsity level of the estimated sources. Therefore, we present the performances of the proposed algorithms as a function of the sparsity level² of the estimation on Fig. 4.

In this case, the behavior of the SDR and SIR are comparable. We can notice that, for BSS-Regu, BSS-LPADMM and BSS-Deco, the best performance is obtained when the sparsity level is around 85%. Empirically, this sparsity level corresponds to $\lambda \simeq \sigma$ where σ is the standard deviation of the input noise.

E. Underdetermined BSS

The number of microphones varies from 2 to 5 and the number of sources runs from 3 to 6. We compare the proposed algorithms to the state-of-the art approaches where the mixing matrix is first estimated using Demix [15], then the sources are estimated by the time-frequency masking (DUET) [16] or the ℓ_1 minimization of the analysis coefficients [17]. We denote these two methods by Demix-DUET and Demix- ℓ_1 respectively. We must stress that Demix cannot work when

²Sparsity level here means the percentage of zero value in the vector or matrix

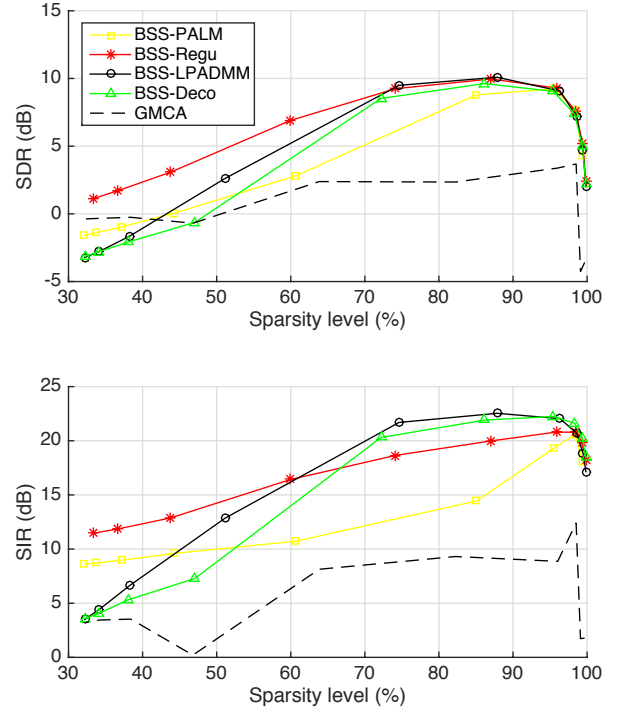


Fig. 4. Performances as a function of the sparsity level of the estimated sources for SNR=7.58 dB in determined scenario ($M = 3$, $N = 3$)

the number of sources is larger than 5 with the number of microphone $M = 2$.

1) Noiseless case:

a) *Performance as a function of the number of microphone:* We first fix the number of sources to $N = 6$. Table II shows the evolution of SDR and SIR with respect to the number of microphones. As DUET method is mainly for two-microphones setting, its performance is not shown.

Except for $M = 2$, the Demix- ℓ_1 reaches the best SDR, but the difference with BSS-Deco and BSS-Regu is less than 1 dB, while BSS-Deco outperforms other approaches in term of SIR. One can remark that BSS-LPADMM-Oracle is also outperformed by BSS-Deco. This result shows that the proposed ADMM algorithm can be sensitive to local minima.

b) *Performance as a function of the number of sources:* For this experiment, the number of microphone is fixed to $M = 2$. We provide on Fig. 5 the evolution of SDR and SIR with respect to the number of sources.

As expected, performances collapse when N grows. For $N \leq 4$, all approaches except BSS-PALM and Demix-DUET provide comparable results in term of SDR. BSS-Deco reaches the best SIR (greater than BSS-Oracle and close to BSS-LPADMM-Oracle). For $N \geq 5$, performances of BSS-Deco collapse in term of SDR, but still outperforms other non-oracle algorithms in term of SIR. BSS-Regu has the opposite behavior.

2) Noisy case:

a) *Performance as a function of the input SNR:* Fig. 6 displays the separation results of the proposed algorithms as

TABLE II

PERFORMANCES OF DIFFERENT ALGORITHMS IN NOISELESS UNDERDETERMINED SETTING WITH NUMBER OF SOURCES $N = 6$ (SDR / SIR). ON A LINE, THE BEST PERFORMANCE IS IN BLACK BOLD. IF THE DIFFERENCE BETWEEN A PERFORMANCE AND THE BEST IS LESS THAN 1 dB, IT IS DISPLAYED IN GRAY BOLD.

	BSS-PALM	BSS-Regu	BSS-ADMM	BSS-Deco	BSS-Oracle	BSS-LPADMM-Oracle	Demix- ℓ_1
$M = 2$	01.4 / 05.9	02.5 / 06.9	01.9 / 06.2	01.3 / 07.8	03.0 / 07.9	02.5 / 09.2	-
$M = 3$	07.5 / 11.5	07.7 / 12.0	06.4 / 12.6	07.5 / 14.3	07.9 / 12.5	06.7 / 13.0	08.0 / 12.1
$M = 4$	12.9 / 17.0	12.7 / 16.9	11.8 / 18.1	12.9 / 19.7	13.3 / 18.0	12.1 / 18.6	13.2 / 17.1
$M = 5$	15.3 / 17.1	19.5 / 23.6	17.7 / 24.1	20.1 / 26.3	20.2 / 24.7	18.7 / 24.6	20.1 / 24.1

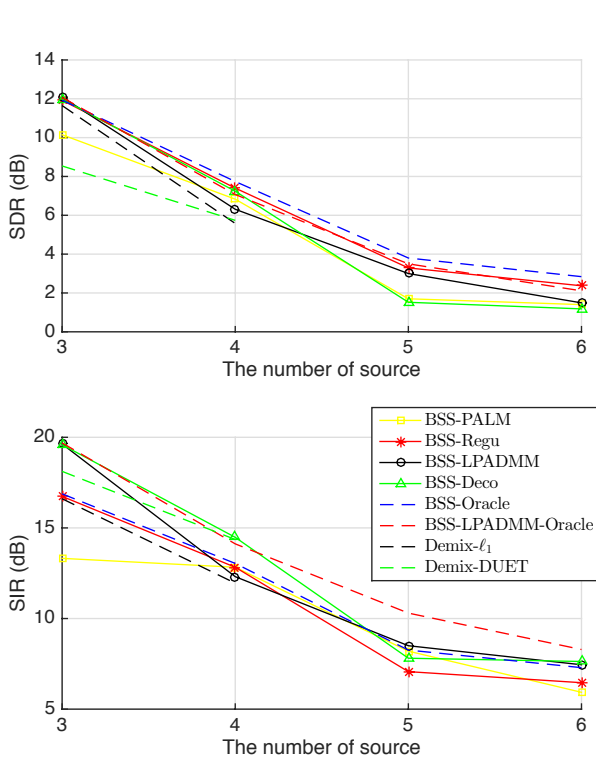


Fig. 5. Performances as a function of the number of sources in underdetermined scenario ($M = 2$)

a function of the input SNR, with a number of sources fixed to $N = 3$ and a number of microphones to $M = 2$.

Again, the two Oracle algorithms outperform the others in term of SDR, while only BSS-LPADMM-Oracle outperforms other approaches in term of SIR, still supporting the fact that taking decorrelation into account improves the SIR of the estimation.

From a SDR point view, all algorithms are comparable, except Demix-DUET. BSS-PALM and Demix- ℓ_1 perform a little worse, but the difference is less than 1 dB. The major difference between the algorithms is from a SIR point of view: BSS-LPADMM and BSS-Deco clearly outperform other approaches, including BSS-Oracle.

b) Performance as a function of the sparsity level:

As in the (over)determined case, we present on Fig. 7 the separation performances as a function of the sparsity level of the estimation for SNR=23.43 dB.

Contrary to the determined case, a compromise must be performed between the SDR and the SIR: a small improvement

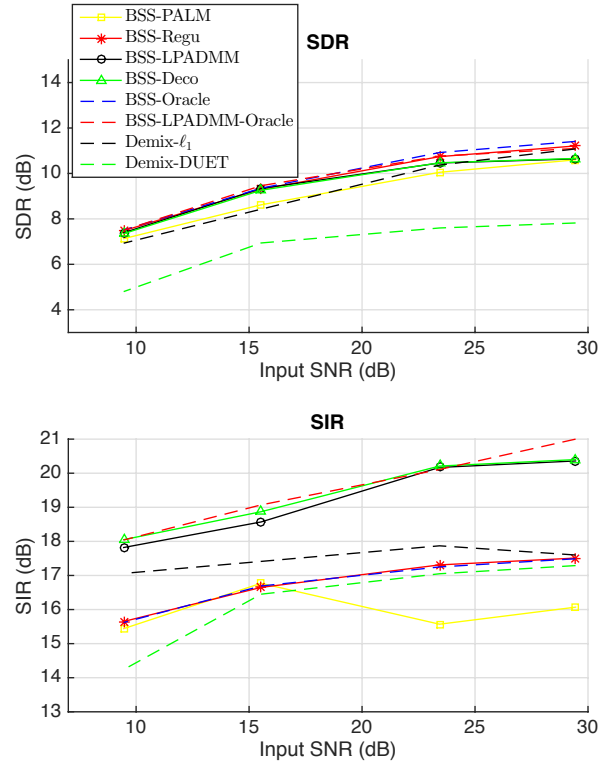


Fig. 6. Performances as a function of the input SNR in underdetermined scenario ($M = 2$, $N = 3$)

on the SDR can lead to a big decrease of the SIR.

F. Computational comparison

We end the experiment section by giving some indications about the computational time of different algorithms. Table III shows the computational time for the previously mentioned sparsity-based algorithms with 20000 iterations, which are the number of iterations used in practice for the experiments.

TABLE III
COMPUTATIONAL TIME FOR DIFFERENT ALGORITHMS

BSS-Regu	BSS-PALM	BSS-LPADMM	BSS-Deco	GMCA	Demix- ℓ_1
2790 s	1430 s	4780 s	2840 s	1600 s	1280 s

The computational times for EFICA and SOBI using ICALAB toolbox and Demix-DUET method are less than 1 second.

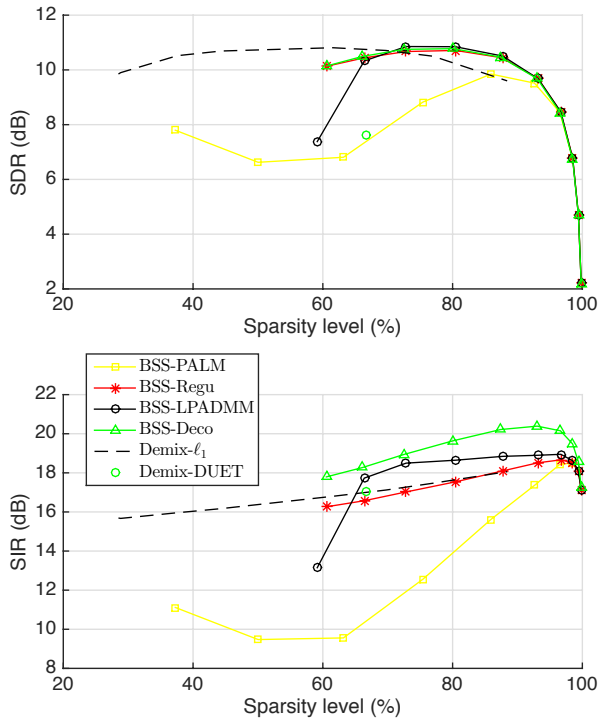


Fig. 7. Performances as a function of the sparsity level in underdetermined noiseless case ($M = 2$, $N = 3$) with $\text{SNR}=23.4$ dB

VI. DISCUSSION AND CONCLUSION

In this paper, we studied the link between some ICA methods and SCA for instantaneous BSS. By combining the decorrelation constraint in time domain and the synthesis sparsity optimization, we proposed a new framework of Sparse ICA to generalize ICA into noisy and underdetermined scenario. We designed several iterative algorithms to solve the problem.

Numerical experiment clearly support that taking the decorrelation into account greatly improve the results in term of SIR, without degrading the SDR. This claim is particularly supported by the fact that proposed algorithms BSS-LPADMM and BSS-Deco outperform the non-blind source separation Oracle algorithm without any decorrelation constraint. Moreover, except when the number of unknown sources is large, the proposed BSS-LPADMM and BSS-Deco reach comparable results of their non blind source separation Oracle with decorrelation constraint BSS-LPADMM-Oracle.

Regarding the computational time of the various approaches used in the experimental section, it appears that ICA approaches remain the most competitive for noiseless and overdetermined mixture. However, for noisy overdetermined mixtures, BSS-Deco appears to be much more robust than GMCA with respect to the input SNR and the number of unknown sources, GMCA being already known to be more robust to noise than ICA methods [20]. Finally, for underdetermined mixtures (with or without noise) BSS-Deco appears to be very competitive: while its computational cost is twice that

of Demix- ℓ_1 , the SIR improvement is around 1 dB for an input SNR of 10 dB, and 2 dB for an input SNR of 20 dB, while the SDR is slightly higher (less than 1 dB).

Future works will focus on extending the Sparse ICA framework to convolutive mixtures. Straightforward extension of this work could also be considered: studying other sparse regularization then the simple ℓ_1 norm, such as social sparsity [51], $\frac{\ell_p}{\ell_q}$ criterion [33], but also considering the sparsity constraint directly on the analysis coefficient of the sources as in [52].

APPENDIX

A. Proof of Proposition 1

Proof. PALM method is designed to deal with non-convex problems reading:

$$\min_{\mathbf{x}, \mathbf{y}} F(\mathbf{x}) + Q(\mathbf{x}, \mathbf{y}) + G(\mathbf{y}) \quad (25)$$

where $F(\mathbf{x})$ and $G(\mathbf{y})$ are proper lower semi-continuous functions, $Q(\mathbf{x}, \mathbf{y})$ is a smooth function with Lipschitz gradient on any bounded set. The proximal method proposed in [22] updates the estimate of (\mathbf{x}, \mathbf{y}) via

$$\mathbf{x}_{k+1} \in F(\mathbf{x}) + \langle \mathbf{x} - \mathbf{x}_k, \nabla_{\mathbf{x}} Q(\mathbf{x}_k, \mathbf{y}_k) \rangle + \frac{t_k^1}{2} \|\mathbf{x} - \mathbf{x}_k\|_F^2 \quad (26)$$

$$\mathbf{y}_{k+1} \in G(\mathbf{y}) + \langle \mathbf{y} - \mathbf{y}_k, \nabla_{\mathbf{y}} Q(\mathbf{x}_{k+1}, \mathbf{y}_k) \rangle + \frac{t_k^2}{2} \|\mathbf{y} - \mathbf{y}_k\|_F^2 \quad (27)$$

where t_k^1 and t_k^2 are two appropriate chosen step sizes.

With the proximal operator defined in (11), the minimization steps (26) and (27) can be written as follows:

$$\mathbf{x}_{k+1} \in \text{prox}_{F/t_k^1}(\mathbf{x}_k - \frac{1}{t_k^1} \nabla_{\mathbf{x}} Q(\mathbf{x}_k, \mathbf{y}_k)) \quad (28)$$

$$\mathbf{y}_{k+1} \in \text{prox}_{G/t_k^2}(\mathbf{y}_k - \frac{1}{t_k^2} \nabla_{\mathbf{y}} Q(\mathbf{x}_{k+1}, \mathbf{y}_k)) \quad (29)$$

It is shown in [22] that the sequence generated by (28) (29) convergences to a critical point of (25). For problem (9), we have:

$$\begin{aligned} F(\tilde{\mathbf{S}}) &= \lambda \|\tilde{\mathbf{S}}\|_1, \quad G(\mathbf{A}) = \iota_C(\mathbf{A}) \\ Q(\mathbf{A}, \tilde{\mathbf{S}}) &= \frac{1}{2} \|\mathbf{X} - \mathbf{A}\alpha\Phi^*\|_F^2 \end{aligned} \quad (30)$$

and the particular choices:

$$t_{k,1} = L_{k,1} \quad t_{k,2} = L_{k,2} \quad (31)$$

where $L_{k,1}$ and $L_{k,2}$ are the Lipschitz constants of $\nabla_{\tilde{\mathbf{S}}} Q(\mathbf{A}_k, \alpha)$ and $\nabla_{\mathbf{A}} Q(\mathbf{A}, \alpha_{k+1})$ respectively. \square

B. Proof of Proposition 2

Proof. The convergence proof is obtained by applying the block-coordinate variable metric forward backward (BC-VMFB) method [34]. BC-VMFB deals with the problem (25), using the following update rules:

$$\mathbf{x}_{k+1} \in F(\mathbf{x}) + \langle \mathbf{x} - \mathbf{x}_k, \nabla_{\mathbf{x}} Q(\mathbf{x}_k, \mathbf{y}_k) \rangle + \frac{t_k^1}{2} \|\mathbf{x} - \mathbf{x}_k\|_{\mathbf{U}_{k,1}}^2 \quad (32)$$

$$\mathbf{y}_{k+1} \in G(\mathbf{x}) + \langle \mathbf{y} - \mathbf{y}_k, \nabla_{\mathbf{y}} Q(\mathbf{x}_{k+1}, \mathbf{y}_k) \rangle + \frac{t_k^2}{2} \|\mathbf{y} - \mathbf{y}_k\|_{\mathbf{U}_{k,2}}^2 \quad (33)$$

where $\mathbf{U}_{k,1}$ and $\mathbf{U}_{k,2}$ are positive definite matrices. Thanks to the generalized proximal operator [34]:

$$\text{prox}_{\mathbf{U}, \psi} := \underset{\mathbf{y}}{\text{argmin}} \psi(\mathbf{y}) + \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_{\mathbf{U}}^2 \quad (34)$$

the update rules (32) (33) read:

$$\mathbf{x}_{k+1} \in \text{prox}_{\mathbf{U}_{k,1}, F/t_k^1}(\mathbf{x}_k - \frac{1}{t_k^1} \nabla_{\mathbf{x}} Q(\mathbf{x}_k, \mathbf{y}_k)) \quad (35)$$

$$\mathbf{y}_{k+1} \in \text{prox}_{\mathbf{U}_{k,2}, G/t_k^2}(\mathbf{y}_k - \frac{1}{t_k^2} \nabla_{\mathbf{y}} Q(\mathbf{x}_{k+1}, \mathbf{y}_k)) \quad (36)$$

which are shown in [34] to converge to a critical point of (25).

In order to apply BC-VMFB to (22), we choose

$$\begin{aligned} F(\tilde{\mathbf{S}}) &= \lambda \|\tilde{\mathbf{S}}\|_1, \quad G(\mathbf{A}) = \iota_{\mathcal{C}}(\mathbf{A}) \\ Q(\mathbf{A}, \tilde{\mathbf{S}}) &= \frac{1}{2} \|\mathbf{X} - \mathbf{A}\tilde{\mathbf{S}}\Phi^*\|_F^2 + \frac{\mu}{2} \|\mathbf{A}\|_F^2 \end{aligned} \quad (37)$$

and

$$\begin{aligned} t_{k,1} &= t_{k,2} = 1, \quad \mathbf{U}_{k,1} = L_{k,1} \mathbf{I} \\ \mathbf{U}_{k,2} &= \frac{\partial Q(\mathbf{A}, \tilde{\mathbf{S}}_{k+1})^2}{\partial^2 \mathbf{A}} = \tilde{\mathbf{S}}_{k+1} \Phi^* \Phi \tilde{\mathbf{S}}_{k+1}^* + \mu \mathbf{I} \end{aligned} \quad (38)$$

where \mathbf{I} is the identity matrix. The update step for \mathbf{A} (36) can now be written as:

$$\begin{aligned} \mathbf{A}_{k+1/2} &= \mathbf{x} \Phi \tilde{\mathbf{S}}_{k+1}^* (\tilde{\mathbf{S}}_{k+1} \Phi^* \Phi \tilde{\mathbf{S}}_{k+1}^* + \mu \mathbf{I})^{-1} \\ \mathbf{A}_{k+1} &\in \text{prox}_{\mathbf{U}_{k,2}, \iota_{\mathcal{C}}}(\mathbf{A}_{k+1/2}) \end{aligned} \quad (39)$$

As the choice of the parameter μ does not change the minimizer of (22), by choosing μ large enough, the update of \mathbf{A} simply reads:

$$\begin{aligned} \mathbf{A}_{k+1/2} &= \mathbf{X} \Phi \tilde{\mathbf{S}}_{k+1}^* \\ \mathbf{a}_{k+1,n} &= \mathbf{a}_{k+1/2,n} / \|\mathbf{a}_{k+1/2,n}\|, \quad n = 1, 2, \dots, N \end{aligned} \quad (40)$$

□

C. Proof of Proposition. 3

Proof. The minimization of (22) is equivalent to the following problem as $\lambda \rightarrow 0$:

$$\min_{\alpha, \mathbf{A}} \|\alpha\|_1 + \iota_{\mathcal{C}}(\mathbf{A}) \quad \text{s.t. } \mathbf{X} = \mathbf{A}\alpha\Phi^* \quad (41)$$

On one hand if α and \mathbf{A} is a minimiser of (41), one must have that $\mathbf{X} = \mathbf{A}\mathbf{S}$ with $\mathbf{S} = \alpha\Phi^*$. On the other hand, at convergence, the update for \mathbf{A} in BSS-Regu (step 3 in Algorithm 2) ensures that one has $\mathbf{A} = \mathbf{X}\mathbf{S}^T \mathbf{D}$ where \mathbf{D} is a diagonal matrix corresponding to the normalization step. Then one has $\mathbf{A} = \mathbf{A}\mathbf{S}\mathbf{S}^T \mathbf{D}$, leading to $\mathbf{S}\mathbf{S}^T = \mathbf{D}^{-1}$ if $M \geq N$.

□

D. Derivation of BSS-LPADMM

The augmented Lagrangian reads:

$$\begin{aligned} L(\mathbf{A}, \alpha, \mathbf{s}, \eta) &= \frac{1}{2} \|\mathbf{X} - \mathbf{A}\alpha\Phi^*\|_F^2 + \lambda \|\alpha\|_1 + \iota_{\mathcal{C}}(\mathbf{A}) \\ &\quad + \iota_{\text{deco}}(\mathbf{S}) + \langle \eta, \mathbf{S} - \alpha\Phi^* \rangle + \frac{\gamma}{2} \|\mathbf{S} - \alpha\Phi^*\|_F^2 \end{aligned} \quad (42)$$

where η is the dual variable and γ is the penalty parameter. Let

$$F_j(\alpha) = \frac{1}{2} \|\mathbf{X} - \mathbf{A}_j \alpha \Phi^*\|_F^2 + \langle \eta_j, \mathbf{S}_j - \alpha \Phi^* \rangle + \frac{\gamma}{2} \|\mathbf{S}_j - \alpha \Phi^*\|_F^2 \quad (43)$$

LPADMM minimizes the augmented Lagrangian by iteratively updating the primal and dual variables via the following update rules:

$$\alpha_{k+1} = \underset{\alpha}{\text{argmin}} \langle \nabla_{\alpha} F_k(\alpha_k), \alpha \rangle + \lambda \|\alpha\|_1 + \frac{L_{F,k}}{2} \|\alpha_k - \alpha\|_F^2 \quad (44)$$

$$\mathbf{S}_{k+1} \in \underset{\mathbf{S}}{\text{argmin}} \langle \eta_k, \mathbf{S} - \alpha_{k+1} \Phi^* \rangle + \frac{\gamma}{2} \|\mathbf{S} - \alpha_{k+1} \Phi^*\|_F^2 + \iota_{\text{deco}}(\mathbf{S}) \quad (45)$$

$$\mathbf{A}_{k+1} = \underset{\mathbf{A}}{\text{argmin}} \frac{1}{2} \|\mathbf{X} - \mathbf{A}\alpha_{k+1}\Phi^*\|_F^2 + \iota_{\mathcal{C}}(\mathbf{A}) \quad (46)$$

$$\eta_{k+1} = \eta_k + \gamma(\mathbf{s}_{k+1} - \alpha_{k+1}\Phi^*) \quad (47)$$

In sub-problem (44), $L_{F,k}$ is the Lipschitz constant of $\nabla_{\alpha} F_k(\alpha)$, with

$$\nabla_{\alpha} F_k(\alpha) = -\mathbf{A}_k^T (\mathbf{X} - \mathbf{A}_k \alpha \Phi^*) \Phi - \eta_k \Phi - \gamma (\mathbf{S}_k - \alpha \Phi^*) \Phi.$$

Using the soft-thresholding operator (12), (44) can be rewritten as:

$$\alpha_{k+1} = \mathcal{S}_{\lambda/L_{F,k}} \left(\alpha_k - \frac{\nabla_{\alpha} F_k(\alpha_k)}{L_{F,k}} \right) \quad (48)$$

Sub-problem (45) can be formulated as a the decorrelation projection

$$\begin{cases} \mathbf{S}_{k+1} = \underset{\mathbf{S}}{\text{argmin}} \frac{\gamma}{2} \left\| \mathbf{S} - \alpha_{k+1} \Phi^* + \frac{\eta_k}{\gamma} \right\|_F^2 \\ \text{s.t. } \mathbf{S}\mathbf{S}^T = \mathbf{D} \end{cases} \quad (49)$$

and can be solved thanks to the following proposition

Proposition 4. Let $\mathbf{S} \in \mathbb{R}^{N \times T}$ be a 0-mean signal matrix with a positive definite covariance matrix $\Sigma_{\mathbf{S}} = \mathbf{S}\mathbf{S}^T$. Let \mathbf{W} be the optimal decorrelation transformation that minimizes the Mean-Squared Error (MSE) between the input \mathbf{S} and the output $\mathbf{Y} = \mathbf{W}\mathbf{S}$, such that its covariance matrix $\Sigma_{\mathbf{Y}}$ is diagonal:

$$\min_{\mathbf{W}} \|\mathbf{S} - \mathbf{W}\mathbf{S}\|_F^2 \quad \text{s.t. } \Sigma_{\mathbf{Y}=\mathbf{W}\mathbf{S}} \text{ is diagonal}$$

then $\mathbf{W} = \text{diag}(\Sigma_{\mathbf{S}}^{1/2}) \Sigma_{\mathbf{S}}^{-1/2}$ where $\text{diag}(\cdot)$ is the diagonal matrix formed with its diagonal elements.

Proof. Using the change of variables $\bar{\mathbf{S}} = \Sigma_{\mathbf{S}}^{-1/2} \mathbf{S}$ and $\bar{\mathbf{W}} = \mathbf{W} \Sigma_{\mathbf{S}}^{1/2}$, one can rewrite the minimization problem as

$$\min_{\bar{\mathbf{W}}} \|\mathbf{S} - \bar{\mathbf{W}} \bar{\mathbf{S}}\|_F^2 \quad \text{s.t.} \quad \bar{\mathbf{W}} \bar{\mathbf{W}}^T = \mathbf{D}$$

where \mathbf{D} is any diagonal matrix. The MSE $e = \|\mathbf{S} - \bar{\mathbf{W}} \bar{\mathbf{S}}\|_F^2$ can then be rewritten as :

$$e = \|\mathbf{S}\|_F^2 + \|\bar{\mathbf{W}} \bar{\mathbf{S}}\|_F^2 - 2\langle \mathbf{S}, \bar{\mathbf{W}} \bar{\mathbf{S}} \rangle$$

According to the Cauchy-Schwarz inequality, one has $\langle \mathbf{S}, \bar{\mathbf{W}} \bar{\mathbf{S}} \rangle \leq \|\mathbf{S}\|_F \|\bar{\mathbf{W}} \bar{\mathbf{S}}\|_F$ with equality if and only if $\bar{\mathbf{W}}$ is a diagonal matrix. denoting by \bar{d}_i the i -th diagonal element of $\bar{\mathbf{W}}$, e reads :

$$e = \sum_{i=1}^N (\mathbf{s}_i - \bar{d}_i \bar{\mathbf{s}}_i) (\mathbf{s}_i - \bar{d}_i \bar{\mathbf{s}}_i)^T$$

where \mathbf{s}_i and $\bar{\mathbf{s}}_i$ is the i -th row of \mathbf{S} and $\bar{\mathbf{S}}$ respectively. Minimizing e with respect to \bar{d}_i is a simple problem leading to $\bar{d}_i = \mathbf{s}_i \bar{\mathbf{s}}_i^T$, i.e. $\bar{\mathbf{W}} = \text{diag}(\mathbf{S} \bar{\mathbf{S}}^T)$. \square

Finally, Sub-problem (46) is tackled by a classical projected gradient descend.

REFERENCES

- [1] S. Kim and C. D. Yoo, "Underdetermined blind source separation based on subspace representation," *IEEE Transactions on Signal processing*, vol. 57, no. 7, pp. 2604–2614, 2009.
- [2] S.-i. Amari and A. Cichocki, "Adaptive blind signal processing-neural network approaches," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2026–2048, 1998.
- [3] A. Mansour, A. K. Barros, and N. Ohnishi, "Blind separation of sources: Methods, assumptions and applications," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 83, no. 8, pp. 1498–1512, 2000.
- [4] J. Hérault and B. Ans, "Réseau de neurones à synapses modifiables: Décodage de messages sensoriels composites par apprentissage non supervisé et permanent," *Comptes rendus des séances de l'Académie des sciences. Série 3, Sciences de la vie*, vol. 299, no. 13, pp. 525–528, 1984.
- [5] P. Comon, "Independent component analysis, a new concept?" *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [6] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.
- [7] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 02, pp. 163–178, 2000.
- [8] M. J. McKeown, T. J. Sejnowski *et al.*, "Independent component analysis of fMRI data: examining the assumptions," *Human brain mapping*, vol. 6, no. 5-6, pp. 368–372, 1998.
- [9] M. U. Khalid and A.-K. Seghouane, "Multi-subject fMRI connectivity analysis using sparse dictionary learning and multiset canonical correlation analysis," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015, pp. 683–686.
- [10] R. Ge, Y. Wang, J. Zhang, L. Yao, H. Zhang, and Z. Long, "Improved FastICA algorithm in fMRI data analysis using the sparsity property of the sources," *Journal of neuroscience methods*, vol. 263, pp. 103–114, 2016.
- [11] M. Plumbley, S. Abdallah, J. Bello, M. Davies, J. Klingseisen, G. Monti, and M. Sandler, "ICA and related models applied to audio analysis and separation," in *Proceedings of the 4th International ICSC Symposium on Soft Computing and Intelligent Systems for Industry*. Citeseer, 2001.
- [12] J. Hurri, A. Hyvärinen, and E. Oja, "Wavelets and natural image statistics," in *In Proc. Scandinavian Conf. on Image Analysis' 97*, 1997.
- [13] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal processing*, vol. 81, no. 11, pp. 2353–2362, 2001.
- [14] M. Babaie-Zadeh, C. Jutten, and A. Mansour, "Sparse ICA via cluster-wise PCA," *Neurocomputing*, vol. 69, no. 13, pp. 1458–1466, 2006.
- [15] S. Arberet, R. Gribonval, and F. Bimbot, "A robust method to count and locate audio sources in a multichannel underdetermined mixture," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 121–133, 2010.
- [16] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, 2004.
- [17] 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.
- [18] R. Gribonval and S. Lesage, "A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges," in *Proceedings of the 14th European Symposium on Artificial Neural Networks (ESANN)*. d-side publi., 2006, pp. 323–330.
- [19] 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.
- [20] 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, pp. 221–302, 2008.
- [21] F. Feng and M. Kowalski, "A unified approach for blind source separation using sparsity and decorrelation," in *Proceedings of the 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 1736–1740.
- [22] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, pp. 1–36, 2013.
- [23] M. Davies, "Audio source separation," in *Institute of mathematics and its applications conference series*, vol. 71. Oxford; Clarendon; 1999, 2002, pp. 57–68.
- [24] A. Belouchrani and A. Cichocki, "Robust whitening procedure in blind source separation context," *Electronics letters*, vol. 36, no. 24, pp. 2050–2051, 2000.
- [25] M. Kowalski and A. Gramfort, "Inverse problems with time-frequency dictionaries and non-white gaussian noise," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 1741–1745.
- [26] E. Vincent, "Complex nonconvex ℓ_p norm minimization for underdetermined source separation," in *Proceedings of International Conference on Independent Component Analysis and Signal Separation (LVA/ICA)*. Springer, 2007, pp. 430–437.
- [27] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical programming*, vol. 103, no. 1, pp. 127–152, 2005.
- [28] J. Mc, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2662–2674, 2007.
- [29] R. T. Rockafellar and R. J. Wets, "Variational analysis: Grundlehren der mathematischen wissenschaften," 1998.
- [30] I. Daubechies, E. Roussos, S. Takerkart, M. Benharrosh, C. Golden, K. D'ardenne, W. Richter, J. Cohen, and J. Haxby, "Independent component analysis for brain fMRI does not select for independence," *Proceedings of the National Academy of Sciences*, vol. 106, no. 26, pp. 10415–10422, 2009.
- [31] V. D. Calhoun, V. K. Potluru, R. Phlypo, R. F. Silva, B. A. Pearlmutter, A. Caprihan, S. M. Plis, and T. Adali, "Independent component analysis for brain fMRI does indeed select for maximal independence," *PLoS one*, vol. 8, no. 8, p. e73309, 2013.
- [32] D. L. Donoho and A. G. Flesia, "Can recent innovations in harmonic analysis "explain" key findings in natural image statistics?" *Network: computation in neural systems*, vol. 12, no. 3, pp. 371–393, 2001.
- [33] L. Li, "Sparsity-promoted blind deconvolution of ground-penetrating radar (GPR) data," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 8, pp. 1330–1334, 2014.
- [34] 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*, pp. 1–26, 2013.
- [35] Y. Ouyang, Y. Chen, G. Lan, and E. Pasilio Jr, "An accelerated alternating direction method of multipliers," *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, pp. 644–681, 2015.
- [36] R. Chartrand and B. Wohlberg, "A nonconvex ADMM algorithm for group sparsity with sparse groups," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2013, pp. 6009–6013.
- [37] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *arXiv preprint arXiv:1511.06324*, 2015.

- [38] L. Yang, T. K. Pong, and X. Chen, "Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction," *arXiv preprint arXiv:1506.07029*, 2015.
- [39] F. Feng and M. Kowalski, "Hybrid model and structured sparsity for under-determined convolutive audio source separation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014.
- [40] S. Araki, F. Nesta, E. Vincent, Z. Koldovský, G. Nolte, A. Ziehe, and A. Benichoux, "The 2011 signal separation evaluation campaign (SiSEC2011):-audio source separation," in *Proceedings of Latent Variable Analysis and Signal Separation (LVA/ICA)*. Springer, 2012, pp. 414–422.
- [41] P. L. Sondergaard, B. Torr sani, and P. Balazs, "The linear time frequency analysis toolbox," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 10, no. 04, 2012.
- [42] E. Vincent, R. Gribonval, and C. F votte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.
- [43] R. Giryes, M. Elad, and Y. C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Applied and Computational Harmonic Analysis*, vol. 30, no. 3, pp. 407–422, 2011.
- [44] C.-A. Deledalle, S. Vaiter, J. Fadili, and G. Peyr , "Stein Unbiased Gradient estimator of the Risk (SUGAR) for multiple parameter selection," *SIAM Journal on Imaging Sciences*, vol. 7, no. 4, pp. 2448–2487, 2014.
- [45] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann, "Subjective and objective quality assessment of audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2046–2057, 2011.
- [46] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1107–1130, 2008.
- [47] Y. Li, A. Cichocki, and S.-i. Amari, "Analysis of sparse representation and blind source separation," *Neural computation*, vol. 16, no. 6, pp. 1193–1234, 2004.
- [48] Z. Koldovsk , P. Tichavsk , and E. Oja, "Efficient variant of algorithm fastica for independent component analysis attaining the Cram r-Rao lower bound," *IEEE Transactions on neural networks*, vol. 17, no. 5, pp. 1265–1277, 2006.
- [49] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on signal processing*, vol. 45, no. 2, pp. 434–444, 1997.
- [50] A. Cichocki, S.-I. Amari, K. Siwek, T. Tanaka, A. H. Phan, R. Zdunek, S. Cruces, P. Georgiev, Y. Washizawa, and Z. Leonowicz, "ICALAB toolboxes," URL: <http://www.bsp.brain.riken.jp/ICALAB>, 2007.
- [51] M. Kowalski, K. Siedenburg, and M. D rfler, "Social sparsity! neighborhood systems enrich structured shrinkage operators," *IEEE transactions on signal processing*, vol. 61, no. 10, pp. 2498–2511, 2013.
- [52] S. Arberet, P. Vanderghelynst, R. Carrillo, J. Thiran, and Y. Wiaux, "Sparse reverberant audio source separation via reweighted analysis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 7, pp. 1391–1402, 2013.