

# BSS with corrupted data in transformed domains

Cécile Chenot and Jérôme Bobin\*

CEA, IRFU, Service d'Astrophysique - SEDI, 91191 Gif-sur-Yvette Cedex, France.  
{cecile.chenot, jerome.bobin}@cea.fr

**Abstract.** Most techniques of Blind Source Separation (BSS) are highly sensitive to the presence of gross errors while these last are ubiquitous in many real-world applications. This mandates the development of *robust* BSS methods, especially to handle the determined case for which there is currently no strategy able to separate the outliers from the sources contributions. We propose a new method which exploits the difference of structural contents that is naturally exhibited by the sources and the outliers in many applications to accurately separate the two contributions. More precisely, we exploit the sparse representations of the signals in two adapted and different dictionaries to estimate jointly the mixing matrix, the sources and the outliers. Preliminary results show the good accuracy of the proposed algorithm in various settings.

**Keywords:** Blind source separation, robust recovery, outliers, sparse signal modeling, morphological diversity.

## 1 Introduction

Multichannel data are nowadays encountered in various domains such as astrophysics [4] or remote sensing [8]. Recovering the underlying signals in these data is generally necessary to analyze them. This extraction of the meaningful information can be done using Blind Source Separation (BSS). The standard instantaneous linear mixture model assumes that BSS aims at recovering the  $n$  sources  $\{\mathbf{S}_i\}_{i=1..n}$  linearly mixed into  $m \geq n$  observations  $\{\mathbf{X}_j\}_{j=1..m}$  with  $t > n$  samples. This model can be conveniently recast in the following matrix form:

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

where  $\mathbf{X} \in \mathbf{R}^{m \times t}$  designates the linear observations,  $\mathbf{A} \in \mathbf{R}^{m \times n}$  the unknown mixing matrix,  $\mathbf{S} \in \mathbf{R}^{n \times t}$  the sources and  $\mathbf{N} \in \mathbf{R}^{m \times t}$  a Gaussian noise term accounting for model imperfections.

This model is too simplistic to represent satisfactorily complex real-world applications. Indeed, the data can be corrupted by localized and large errors, designated in the following as *outliers*  $\mathbf{O}$ . These deviations from the linear model

---

\* This work is supported by the European Community through the grants PHySIS (contract no. 640174), DEDALE (contract no. 665044) and LENA (ERC StG no. 678282) within the H2020 Framework Program.

(1) encompass unexpected physical events such as the presence of spectral variability in hyperspectral unmixing [8], the presence of point-source emissions in astrophysics [15], and also malfunctions of captors [12], to name only a few. In the following, we will assume that the data can be better expressed by:

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{O} + \mathbf{N}, \quad (2)$$

where  $\mathbf{O} \in \mathbf{R}^{m \times t}$  stands for the outliers.

### Robust BSS methods in the literature

Despite the unavoidable presence of outliers in some applications, most of the BSS methods in the literature are highly sensitive to their presence [7] and only few strategies dedicated to this problem have been developed. They can mainly be divided into three classes:

- Within the ICA-framework, the authors of [13] promote the mutual independence of the sources by using the robust  $\beta$ -divergence instead of the standard and sensitive Kullback-Leibler divergence. However, since this method only estimates  $\mathbf{A}$ , no separation between  $\mathbf{O}$  and  $\mathbf{S}$  is performed.
- The “two-step methods” reside in: i) eliminating  $\mathbf{O}$  from the data and ii) performing the separation on the “outliers-free” observations. This strategy has been particularly popularized in hyperspectral imaging [12], [19] for which a precise separation between  $\mathbf{O}$  and the low-rank matrix  $\mathbf{A}\mathbf{S}$  has been shown to be possible with the algorithm PCP [6].
- The component separation techniques aim at recovering simultaneously  $\mathbf{A}$ ,  $\mathbf{S}$  and  $\mathbf{O}$ . It has essentially been used in the NMF framework [1, 8, 10, 11]. The efficiency of these methods strongly depends on the non-negativity assumption, which is not valid in a large number of applications.

In [7], we proposed a component separation method exploiting the sparse representations of  $\mathbf{S}$  and  $\mathbf{O}$  in a same dictionary to jointly estimate  $\mathbf{A}$ ,  $\mathbf{S}$  and  $\mathbf{O}$ . Even though  $\mathbf{A}$  is well estimated, this method is unable to accurately separate  $\mathbf{O}$  from the sources contributions, especially when the number of sources is close to the number of observations [7]. Indeed, even if  $\mathbf{A}$  is perfectly known, separating  $\mathbf{O}$  from  $\mathbf{S}$  is an ill-posed problem since it amounts recovering the sought-after signals  $[\mathbf{S}^T \mathbf{O}^T]^T$  from the observations  $\mathbf{X}$  obtained with the sensing matrix  $[\mathbf{A} \mathbf{I}]$ :

$$\text{Recover } \mathbf{S} \text{ and } \mathbf{O} \text{ given } \mathbf{A} \text{ and } \mathbf{X} \text{ such that } \mathbf{X} = (\mathbf{A} \mathbf{I}) \begin{pmatrix} \mathbf{S} \\ \mathbf{O} \end{pmatrix}, \quad (3)$$

where  $\mathbf{I}$  denotes the identity matrix of size  $m \times m$ .

Solving (3) requires additional assumptions on the signals such as:

- The outliers do not lie in the span of  $\mathbf{A}$  or  $\mathbf{A}\mathbf{S}$  is low rank while  $\mathbf{O}$  is sparse and broadly distributed [6]. Consequently, if  $m \gg n$ , the outliers can be separated precisely from the sources contribution but this is not valid if  $m$  is close to  $n$ .
- This ill-posed problem can also be handled using sparsity-based regularization [7], [2]. Nonetheless, the compressibility of  $\mathbf{S}$  and  $\mathbf{O}$  in a same dictionary

is not sufficient to solve (3): it is also necessary that every sample of  $[\mathbf{S}^T \mathbf{O}^T]^T$  be sparse. This condition is rarely verified in practice (*e.g.*  $\mathbf{O}$  is column-sparse such as in [7], [8]). However, if the structural contents of the sources and the outliers are different, it is possible to separate the two signals by representing sparsely each signal in one specific dictionary (morphological diversity principle [14]). The two dictionaries then help discriminating between the two contributions.

In the following, we will assume that the morphologies of the outliers and the sources are different in order to separate the two contributions [14]. This additional assumption is usually valid in imaging problems. For instance, in hyperspectral imaging, stripping lines created by malfunctions of captors (the outliers) have a different geometry than the spatial distributions of the observed components (the sources) [12]. Similarly in astrophysics, point source emissions (outliers) have a different morphology than the components of interest which are more broadly distributed [15], [4].

### Contributions

We introduce a new robust BSS algorithm, coined rGMCA, enforcing the sparsity of the sources and the one of the outliers in different transformed domains. It exploits the difference of morphology between outliers and sources to separate the two contributions and estimates precisely the mixing matrix, the sources and outliers, without restrictive hypothesis on low-rankness or non-negativity. A review of the morphological diversity principle is provided in section 2. The algorithm rGMCA is detailed in section 3. Last, numerical experiments are presented in section 4, showing the good performances of the rGMCA algorithm.

### Notations

The Moore-Penrose pseudo-inverse of the matrix  $\mathbf{M}$  is designated by  $\mathbf{M}^\dagger$  and its transpose by  $\mathbf{M}^T$ . The  $j$ th column of  $\mathbf{M}$  is denoted  $\mathbf{M}^j$ , the  $i$ th row  $\mathbf{M}_i$ , and the  $i, j$ th entry  $\mathbf{M}_{i,j}$ . The norm  $\|\mathbf{M}\|_2$  denotes the Frobenius norm of  $\mathbf{M}$ , and more generally  $\|\mathbf{M}\|_p$  designates the  $p$ -norm of the matrix  $\mathbf{M}$  seen as a long vector. The soft-thresholding operator is denoted  $\mathcal{S}_\lambda(\mathbf{M})$ , where

$$[\mathcal{S}_\lambda(\mathbf{M})]_{i,j} = \begin{cases} \mathbf{M}_{i,j} - \text{sign}(\mathbf{M}_{i,j}) * \lambda_i & \text{if } |\mathbf{M}_{i,j}| > \lambda_i \\ 0 & \text{otherwise} \end{cases}$$

## 2 Sparsity and morphological diversity

We aim at separating the outliers from the sources by assuming that their morphological/structural contents are different. For this purpose, we introduce two appropriate dictionaries:  $\Phi_{\mathbf{O}}$  and  $\Phi_{\mathbf{S}}$ . These dictionaries are key to separating the two contributions. They are chosen so that the corresponding expansion coefficients of  $\mathbf{O}$  and  $\mathbf{S}$  are sparse:

$$\mathbf{O}_j = \alpha_{\mathbf{O}_j} \Phi_{\mathbf{O}}, \forall j \in \{1..m\} \quad \text{and} \quad \mathbf{S}_i = \alpha_{\mathbf{S}_i} \Phi_{\mathbf{S}}, \forall i \in \{1..n\},$$

where  $\{\alpha_{\mathbf{O}_j}\}_{j=1..m}$  and  $\{\alpha_{\mathbf{S}_i}\}_{i=1..n}$  are composed of few significant samples. For instance, wavelets can be used to represent sparsely natural images and curvelets for smooth curves to cite only two [14].

The morphological diversity between the sources and the outliers implies that each component  $\{\mathbf{O}_j\}_{j=1..m}$  or  $\{\mathbf{S}_i\}_{i=1..n}$  has its sparsest expansion coefficients in  $\Phi_{\mathbf{O}}$  or  $\Phi_{\mathbf{S}}$  respectively:

$$\forall i \in \{1..n\}, \forall j \in \{1..m\}, \|\mathbf{O}_j \Phi_{\mathbf{O}}^T\|_0 < \|\mathbf{O}_j \Phi_{\mathbf{S}}^T\|_0 \text{ and } \|\mathbf{S}_i \Phi_{\mathbf{S}}^T\|_0 < \|\mathbf{S}_i \Phi_{\mathbf{O}}^T\|_0.$$

Therefore, it is possible to solve (3) by seeking for the sparsest representations, in the spirit of the MCA (*Morphological Component Analysis*) algorithm. The latest aims at separating  $k$  different morphological components of a monochannel signal, given  $k$  appropriate dictionaries, by maximizing the sparsity of the expansion coefficients of each morphological component in its corresponding dictionary. The good performances of MCA support the utilization of the sparsity to separate different morphological components [14].

In the next section, we will present how, we exploit the morphological diversity between the sources and the outliers to separate the two contributions. Besides, the sparse representations of  $\mathbf{S}$  will be also used to discriminate between the sources. Indeed, sparsity has been shown to be a powerful criterion to unmix the sources [5].

### 3 The algorithm rGMCA

The use of sparsity in our strategy is twofold: it allows for an accurate separation between the outliers and the sources by exploiting their morphological diversity and it is also used to discriminate between the sources. In order to exploit simultaneously these two aspects, we propose to estimate jointly  $\mathbf{A}$ ,  $\mathbf{S}$  and  $\mathbf{O}$  by minimizing the following cost function:

$$\underset{\mathbf{A}, \mathbf{S}, \mathbf{O}}{\text{minimize}} \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S} - \mathbf{O}\|_2^2 + \lambda \|\mathbf{S}\Phi_{\mathbf{S}}^T\|_1 + \beta \|\mathbf{O}\Phi_{\mathbf{O}}^T\|_{p,q}, \quad (4)$$

where the first term designates the data fidelity term, well suited to deal with the remaining Gaussian noise, and the second and third terms enforce respectively the sparsity of  $\mathbf{S}$  and  $\mathbf{O}$  in their corresponding dictionary. In the following, we will assume that the outliers corrupt entire columns of the data such as in [8]. Consequently, we will promote this structure by using the  $\ell_{2,1}$  norm ( $p = 2, q = 1$ ). Despite the non-convexity of the proposed problem, it can be tackled using Block Coordinate Relaxation [16]. Alternatively estimating  $\mathbf{A}$ ,  $\mathbf{O}$  and  $\mathbf{S}$  propagates the errors from one variable to the others, and thus performs poorly if not initialized with a good accuracy. We propose instead to fully exploit the structure of the problem by using the scheme presented in Alg.1 to minimize (4):

- Estimating  $\mathbf{A}$  and  $\mathbf{S}$  jointly for fixed  $\mathbf{O}$ : it exploits the joint sparsity of the sources to retrieve more precisely  $\mathbf{A}$  from the denoised observations  $\mathbf{X} - \mathbf{O}$ .
- Estimating  $\mathbf{O}$  and  $\mathbf{S}$  for fixed  $\mathbf{A}$  such as in (3): it provides a precise separation of the two contributions by using their morphological diversity.

We found that this scheme was the less prone to be trapped into local minima.

**Algorithm 1** rGMCA Algorithm

---

```

1: procedure rGMCA( $\mathbf{X}, n$ )
2:   Initialize  $\tilde{\mathbf{A}}^{(0)}$  (randomly or with a PCA),  $\tilde{\mathbf{S}}^{(0)} = \mathbf{0}$  and  $\tilde{\mathbf{O}}^{(0)} = \mathbf{0}$ .
3:   while  $k < K$  do
4:     Set  $\tilde{\mathbf{S}}^{(0,k)} \leftarrow \tilde{\mathbf{S}}^{(k-1)}$  and  $\tilde{\mathbf{A}}^{(0,k)} \leftarrow \tilde{\mathbf{A}}^{(k-1)}$ 
5:     while  $i < I$  do ▷ Joint estimation of  $\mathbf{A}$  and  $\mathbf{S}$ 
6:       Update  $\tilde{\mathbf{S}}^{(i,k)}$  with (6)
7:       Update  $\tilde{\mathbf{A}}^{(i,k)}$  with (7)
8:     Set  $\tilde{\mathbf{S}}^{(k)} \leftarrow \tilde{\mathbf{S}}^{(i-1,k)}$  and  $\tilde{\mathbf{A}}^{(k)} \leftarrow \tilde{\mathbf{A}}^{(i-1,k)}$ 
9:     Set  $\tilde{\mathbf{S}}^{(0,k)} \leftarrow \tilde{\mathbf{S}}^{(k)}$  and  $\tilde{\mathbf{O}}^{(0,k)} \leftarrow \tilde{\mathbf{O}}^{(k-1)}$ 
10:    while  $j < J$  do ▷ Joint estimation of  $\mathbf{S}$  and  $\mathbf{O}$ 
11:      Update  $\tilde{\mathbf{S}}^{(j,k)}$  with (5)
12:      Update  $\tilde{\mathbf{O}}^{(j,k)}$  with (8)
13:    Set  $\tilde{\mathbf{S}}^{(k)} \leftarrow \tilde{\mathbf{S}}^{(i-1,k)}$  and  $\tilde{\mathbf{O}}^{(k)} \leftarrow \tilde{\mathbf{O}}^{(i-1,k)}$ 
  return  $\tilde{\mathbf{S}}^{(k-1)}, \tilde{\mathbf{A}}^{(k-1)}, \tilde{\mathbf{O}}^{(k-1)}$ .

```

---

**3.1 Estimation of  $\mathbf{A}$  and  $\mathbf{S}$** 

Estimating  $\mathbf{A}$  and  $\mathbf{S}$  for fixed  $\mathbf{O}$  amounts to minimize the following cost function:

$$\operatorname{argmin}_{\mathbf{S}, \mathbf{A}} \frac{1}{2} \|\mathbf{X} - \mathbf{O} - \mathbf{AS}\|_2^2 + \lambda \|\mathbf{S}\Phi_{\mathbf{S}}^T\|_1.$$

This problem is similar to the GMCA algorithm [5], performed on the residual  $\mathbf{X} - \mathbf{O}$ . This algorithm was first proposed in [5], as well as its fast version that we use to speed-up rGMCA. This fast version seeks directly for the sparse coefficients  $\alpha_{\mathbf{S}}$  and  $\mathbf{A}$ :

- The estimate of  $\alpha_{\mathbf{S}}$ , for fixed  $\mathbf{A}$ , is obtained by minimizing:

$$\operatorname{argmin}_{\alpha_{\mathbf{S}}} \frac{1}{2} \|(\mathbf{X} - \mathbf{O})\Phi_{\mathbf{S}}^T - \mathbf{A}\alpha_{\mathbf{S}}\|_2^2 + \lambda \|\alpha_{\mathbf{S}}\|_1. \quad (5)$$

This can be solved using ISTA or FISTA [3] or with a projected least square as it is proposed in [5] (generally faster than using a proximal method):

$$\alpha_{\mathbf{S}} = \mathcal{S}_{\lambda}(\mathbf{A}^{\dagger}((\mathbf{X} - \mathbf{O})\Phi_{\mathbf{S}}^T)). \quad (6)$$

- The estimate of  $\mathbf{A}$  is given by:

$$\mathbf{A} = ((\mathbf{X} - \mathbf{O})\Phi_{\mathbf{S}}^T) \alpha_{\mathbf{S}}^{\dagger}. \quad (7)$$

Details on GMCA can be found in [5].

**3.2 Estimation of  $\mathbf{S}$  and  $\mathbf{O}$** 

Estimating  $\mathbf{S}$  and  $\mathbf{O}$  for fixed  $\mathbf{A}$  corresponds to the ill-posed problem presented in (3). In the spirit of the MCA algorithm [14], we estimate alternatively the sparse coefficients  $\alpha_{\mathbf{S}}$  and  $\alpha_{\mathbf{O}}$  by working directly in their associated transformed domains with the following updates:

- The estimation of the  $\alpha_{\mathbf{S}}$  is given by (5), which is solved using FISTA.
- The estimation of the  $\alpha_{\mathbf{O}}$  is given by :

$$\operatorname{argmin}_{\alpha_{\mathbf{O}}} \frac{1}{2} \left\| (\mathbf{X} - \mathbf{A}\mathbf{S}) \Phi_{\mathbf{O}}^T - \alpha_{\mathbf{O}} \right\|_2^2 + \beta \|\alpha_{\mathbf{O}}\|_{2,1}.$$

Every entry  $k \in \{1..t\}$  is obtained with the closed form:

$$\tilde{\alpha}_{\mathbf{O}}^k = ((\mathbf{X} - \mathbf{A}\mathbf{S}) \Phi_{\mathbf{O}}^T)^k \times \max \left( 0, 1 - \frac{\beta}{\left\| ((\mathbf{X} - \mathbf{A}\mathbf{S}) \Phi_{\mathbf{O}}^T)^k \right\|_2} \right). \quad (8)$$

### 3.3 Choice of the parameters

The parameters  $\lambda$  and  $\beta$  are automatically set.

**Strategy for  $\lambda$ :** It has been shown in [5] that using a decreasing strategy for  $\lambda$  in the GMCA algorithm increases its robustness against local minima. Practically, an increasing number of entries are selected. The final threshold  $\lambda_i$  for each source  $\mathbf{S}_i$  is  $k\sigma_i$ , where  $k \in (1, 3)$  and  $\sigma_i$  is the standard deviation of the noise contaminating the  $i$ th source. If  $\sigma_i$  is not known, it can be estimated with the MAD (median absolute deviation) operator. A large  $k$  prevents the incorporation of Gaussian noise in the source estimate.

When estimating jointly  $\mathbf{O}$  and  $\mathbf{S}$ , the values of  $\lambda_i$  are directly set to the final thresholds  $k\sigma_i$ .

**Strategy for  $\beta$ :** The value of  $\beta$  is fixed to the value  $\sigma \times \sqrt{2} \times \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})}$  which corresponds to an estimation of  $\mathbb{E} \left\{ \left\| (\mathbf{N}\Phi_{\mathbf{O}}^T)^k \right\|_2 \right\}$ , ( $\operatorname{mad}((\mathbf{X} - \mathbf{A}\mathbf{S} - \mathbf{O})\Phi_{\mathbf{O}}^T)$  corresponds to a good estimate of the standard deviation of  $\mathbf{N}\Phi_{\mathbf{O}}^T$  if it is not known), and thus limits the impact of the Gaussian noise.

## 4 Numerical experiments

In this section, we compare rGMCA with the standard robust BSS methods presented in the introduction: minimization of the  $\beta$ -divergence in  $\Phi_{\mathbf{O}}$  [13] (tuned implementation from [9]), the combination PCP+GMCA (the outliers are first discarded from the observations with a tuned implementation of PCP in  $\Phi_{\mathbf{O}}$  [6] and then  $\mathbf{A}$  and  $\mathbf{S}$  are estimated with GMCA in  $\Phi_{\mathbf{S}}$  [5]) and also with GMCA to illustrate the benefits of using robust strategies. We investigate the performances of the algorithms with respect to the following criteria:

- The unmixing precision is measured with the global criterion  $\Delta_A = \frac{\|\tilde{\mathbf{A}}^\dagger \mathbf{A} - \mathbf{I}\|_1}{n^2}$  [5] and the maximal angle made between the estimated and true columns of  $\mathbf{A}$  defined as  $\max_{i,i \in \{1, \dots, n\}} \arccos \langle \mathbf{A}^i, \tilde{\mathbf{A}}^i \rangle$  in degree.
- The accuracy of the separation between  $\mathbf{S}$  and  $\mathbf{O}$  is assessed with the minimal SDR (signal distortion ratio [17]) obtained for each estimation of  $\mathbf{S}$ .

Since the minimization of the  $\beta$ -divergence does not estimate  $\mathbf{S}$ , we will compute the SDR from the estimated sources  $\mathcal{S}_\lambda(\tilde{\mathbf{A}}^\dagger \mathbf{X}_{\Phi_{\mathbf{S}}}) \Phi_{\mathbf{S}}$ , where  $\tilde{\mathbf{A}}$  is the mixing matrix estimated with the algorithm.

#### 4.1 Monte Carlo simulations

In this first part, we investigate the robustness of the algorithms with Monte-Carlo simulation (80 runs) on 1D signals with varying parameters (the amplitude and the percentage of the corrupted data) and the following setting:

- A total of 8 sources, sparse in DCT, are mixed into 20 observations which are corrupted with the Gaussian noise  $\mathbf{N}$  (standard deviation of 0.1, SNR around 40dB), and the outliers which are sparse in the direct domain.
- The columns of  $\mathbf{O}$  follow a Bernoulli-Gaussian law with an activation parameter  $\rho$  (default value 10%) and a standard deviation  $\sigma_O$  (default value 100).
- The sparse coefficients of  $\mathbf{S}$  are also drawn from a Bernoulli-Gaussian law (activation parameter of 5%, standard deviation of 100).

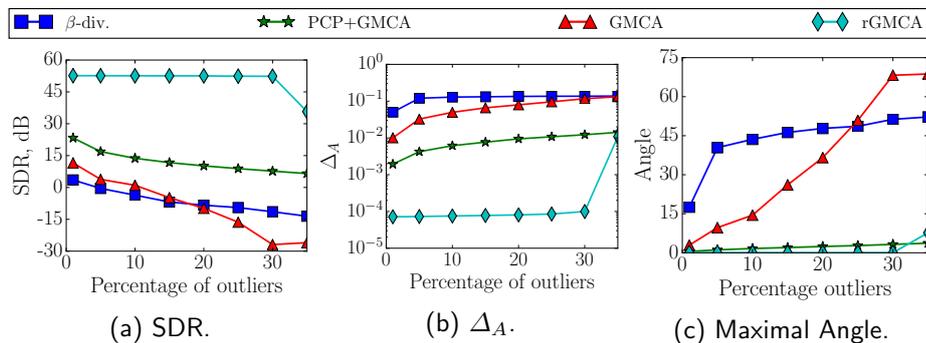


Fig. 1. Influence of the percentage of corrupted entries.

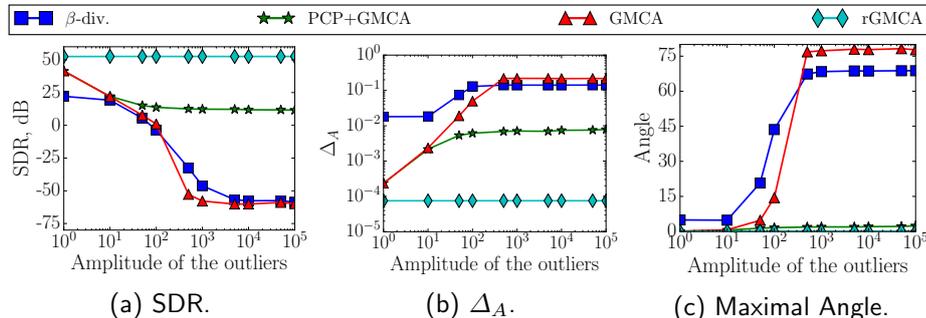


Fig. 2. Influence of the amplitude of the outliers.

**Percentage of corrupted data** As shown in fig.1, the minimization of the  $\beta$ -divergence and GMCA are highly sensitive to the increasing percentage of outliers. Not only  $\mathbf{S}$  is poorly estimated fig.1a, but also the sources are not correctly unmixed fig.1c: the unmixing process is challenging without the explicit estimation of  $\mathbf{O}$ . The combination PCP+GMCA provides the most robust unmixing process fig.1c, but cannot separate precisely the outliers from the source contribution without further hypothesis [18], fig.1a. The algorithm rGMCA returns the most accurate estimations of  $\mathbf{S}$  and  $\mathbf{A}$  while  $\rho$  is lower than 30%, what

should be achieved in practice if the dictionary  $\Phi_{\mathbf{O}}$  is wisely chosen.

**Amplitude of the outliers** The minimization of the  $\beta$ -divergence and GMCA have similar performances: they fail whenever the amplitude of  $\mathbf{O}$  and the one of  $\mathbf{S}$  are of the same order of magnitude. The outliers which are detrimental for the unmixing are discarded from  $\mathbf{X}$  with PCP (fig.2c: the maximal angle is small), but on the overall, PCP struggles to separate accurately  $\mathbf{O}$  from  $\mathbf{AS}$  (fig.2a: the SDR is lower than the one obtained with rGMCA). Besides, whenever the amplitude of  $\mathbf{O}$  is larger than the one of the sources, the performances of PCP+GMCA are constant. Last, rGMCA is not influence by the amplitude of the outliers with this setting since the precision reached for the estimation of  $\mathbf{A}$  and  $\mathbf{S}$  stays constant.

## 4.2 2D simulations

In this section, we compare PCP+GMCA and rGMCA which were significantly the most successful in section 4.1 on 2D applications. The first row of fig.3 shows the sources (four  $128 \times 128$  images, approximately sparse in wavelets [14]) and the outliers corresponding to a high-frequencies texture (approximately sparse in DCT). We observe the influence of the amplitude of  $\mathbf{O}$  and the number of observations  $m$ . When varying  $m$ , the maximal amplitude of  $\mathbf{O}$  is set to the maximal amplitude of  $\mathbf{AS}$ , and respectively, when varying the amplitude of  $\mathbf{O}$ , we set  $m = 20$ . The metrics are averaged for four experiments.

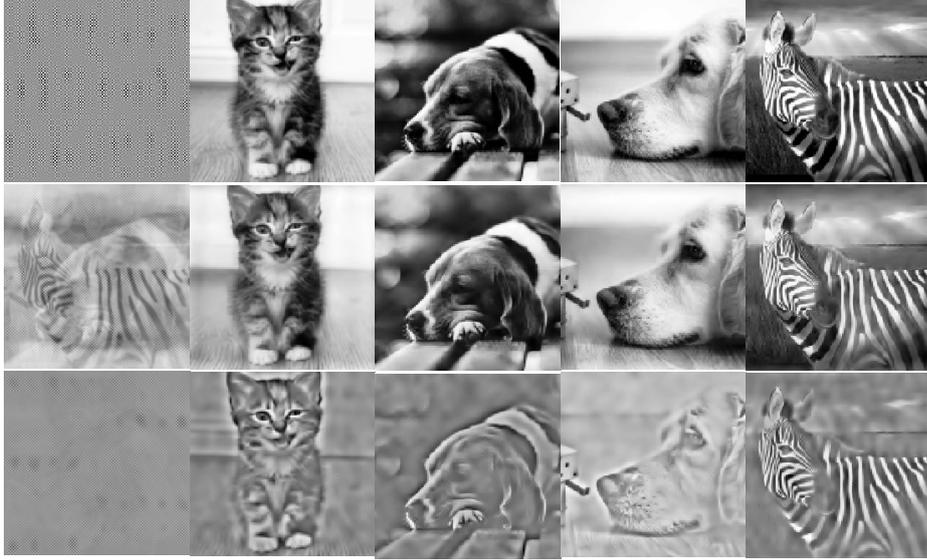
Method		$m$					$\frac{\ \mathbf{O}\ _{\infty}}{\ \mathbf{AS}\ _{\infty}}$					
		4	6	10	18	34	0.01	0.1	1	10	100	
PCP+GMCA	SDR, dB	-4.3	-3.5	-2.6	-2.4	-1.9	-2.1	-2.1	-1.6	-3.1	-4.7	
	Max. Angle	1.1	1.5	1.3	0.8	0.7	0.8	0.8	0.8	1.2	19.6	
rGMCA	SDR, dB	9.3	12.6	13.8	14.1	14.0	13.6	13.7	14.0	14.9	15.7	
	Max.Angle	3.6	2.4	1.3	0.8	0.7	0.9	0.9	0.9	0.9	0.9	

**Table 1.** Results obtained for the simulations with the four images with different numbers of observations  $m$  or amplitudes of the outliers.

**Amplitude of the outliers** Contrary to the previous 1D-case, PCP+GMCA is shown to be sensitive to the amplitude of  $\mathbf{O}$  since it becomes unable to estimate  $\mathbf{A}$  for the largest amplitude tab.1. Moreover, even if  $\mathbf{A}$  is correctly retrieved, the SDR of the sources estimated with PCP+GMCA is very low: the separation between outliers and sources is not correct (see the second row of fig.3). The method rGMCA is more reliable as it returns fair estimates of the mixing matrix and the sources for almost all the experiments. More surprisingly, the SDR obtained with rGMCA increases with the amplitude of  $\mathbf{O}$ : it becomes easier to distinguish the contribution of  $\mathbf{O}$  from the one of  $\mathbf{N}$ . It is also proportionally less influenced by the bias introduced by the different thresholding processes. This improved estimation of  $\mathbf{O}$  leads to accurate estimates of  $\mathbf{S}$ .

**Number of observations** It has been emphasized in [7] and in the introduction that the ratio  $\frac{m}{n}$  is crucial for BSS in the presence of outliers. The unmixing

process and the estimation of  $\mathbf{S}$  should be easier if  $m \gg n$  for both algorithms. The results obtained by the two strategies are indeed improved for a larger  $m$  tab.1. Besides, even if the estimated  $\mathbf{A}$  is slightly more precise for PCP+GMCA, the sources returned by rGMCA are much more accurate (see tab.1 and fig.3).



**Fig. 3.** First row: illustration of  $\mathbf{O}_1$ , then the four initial sources. Second row: illustration of  $\mathbf{X}_1$  for  $m = 34$  and  $\frac{\|\mathbf{O}\|_\infty}{\|\mathbf{AS}\|_\infty} = 1$  and then the sources estimated with rGMCA. Third row: illustration of  $\mathbf{X}_1$  for  $m = 20$  and  $\frac{\|\mathbf{O}\|_\infty}{\|\mathbf{AS}\|_\infty} = 10$  and then sources estimated with PCP+GMCA. The sources were estimated for  $m = 34$  and  $\frac{\|\mathbf{O}\|_\infty}{\|\mathbf{AS}\|_\infty} = 1$ .

## 5 Conclusion

The BSS problem in the presence of outliers is challenging since it requires a robust sources unmixing but also a precise separation of the outliers from the sources contributions. This task is not properly handled by the standard robust BSS methods without restrictive hypothesis. We propose a new method coined rGMCA that estimates jointly the sources, the outliers and the mixing matrix. It exploits the difference of morphology between the sources and the outliers to separate precisely the two contributions, including in the challenging determined case. Preliminary experiments show that rGMCA yields a precise estimation of the mixing matrix and also of the sources in various settings. The discrepancy between rGMCA and the standard robust methods is particularly important for the sources estimations in the proposed experiments. This supports the use of the morphological diversity to discriminate efficiently between the outliers and the sources.

## References

1. Altmann, Y., McLaughlin, S., Hero, A.: Robust Linear Spectral Unmixing Using Anomaly Detection. *IEEE Transactions on Computational Imaging* 1(2), 74–85 (June 2015)
2. Amini, S., Sadeghi, M., Joneidi, M., Babaie-Zadeh, M., Jutten, C.: Outlier-aware dictionary learning for sparse representation. In: *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*. pp. 1–6. IEEE (2014)
3. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* 2(1), 183–202 (2009)
4. Bobin, J., Sureau, F., Starck, J.L., Rassat, A., Paykari, P.: Joint Planck and WMAP CMB map reconstruction. *A&A* 563(A105) (2014)
5. Bobin, J., Starck, J.L., Fadili, J., Moudden, Y.: Sparsity and morphological diversity in blind source separation. *Image Processing, IEEE Transactions on* 16(11), 2662–2674 (2007)
6. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? *Journal of the ACM (JACM)* 58(3), 11 (2011)
7. Chenot, C., Bobin, J., Rapin, J.: Robust Sparse Blind Source Separation. *Signal Processing Letters, IEEE* 22(11), 2172–2176 (2015)
8. Fevotte, C., Dobigeon, N.: Nonlinear Hyperspectral Unmixing With Robust Non-negative Matrix Factorization. *Image Processing, IEEE Transactions on* 24(12), 4810–4819 (Dec 2015)
9. Gadhok, N., Kinsner, W.: An Implementation of  $\beta$ -Divergence for Blind Source Separation. In: *Electrical and Computer Engineering, 2006. CCECE'06. Canadian Conference on*. pp. 1446–1449. IEEE (2006)
10. Halimi, A., Bioucas-Dias, J., Dobigeon, N., Buller, G.S., McLaughlin, S.: Fast hyperspectral unmixing in presence of nonlinearity or mismodelling effects. *arXiv preprint arXiv:1607.05336* (2016)
11. Li, C., Ma, Y., Mei, X., Liu, C., Ma, J.: Hyperspectral unmixing with robust collaborative sparse regression. *Remote Sensing* 8(7), 588 (2016)
12. Li, Q., Li, H., Lu, Z., Lu, Q., Li, W.: Denoising of Hyperspectral Images Employing Two-Phase Matrix Decomposition. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 7(9), 3742–3754 (Sept 2014)
13. Mihoko, M., Eguchi, S.: Robust blind source separation by beta divergence. *Neural computation* 14(8), 1859–1886 (2002)
14. Starck, J.L., Murtagh, F., Fadili, J.M.: *Sparse image and signal processing: wavelets, curvelets, morphological diversity*. Cambridge University Press (2010)
15. Sureau, F., Starck, J.L., Bobin, J., Paykari, P., Rassat, A.: Sparse point-source removal for full-sky cmb experiments: application to wmap 9-year data. *Astronomy & Astrophysics* 566, A100 (2014)
16. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of optimization theory and applications* 109(3), 475–494 (2001)
17. Vincent, E., Gribonval, R., Févotte, C.: Performance measurement in blind audio source separation. *Audio, Speech, and Language Processing, IEEE Transactions on* 14(4), 1462–1469 (2006)
18. Xu, H., Caramanis, C., Sanghavi, S.: Robust pca via outlier pursuit. In: *Advances in Neural Information Processing Systems*. pp. 2496–2504 (2010)
19. Zhang, H., He, W., Zhang, L., Shen, H., Yuan, Q.: Hyperspectral Image Restoration Using Low-Rank Matrix Recovery. *Geoscience and Remote Sensing, IEEE Transactions on* 52(8), 4729–4743 (Aug 2014)