HAL
open science

Joint deconvolution and unsupervised source separation for data on the sphere

R. Carloni Gertosio, Jerome Bobin

To cite this version:

R. Carloni Gertosio, Jerome Bobin. Joint deconvolution and unsupervised source separation for data on the sphere. Digital Signal Processing, 2021, 110, pp.102946. 10.1016/j.dsp.2020.102946 . hal-03177873

HAL Id: hal-03177873

https://hal.science/hal-03177873

Submitted on 23 Mar 2021

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.

arXiv:2012.12740v1 [eess.SP] 23 Dec 2020

Joint deconvolution and unsupervised source separation for data on the sphere

R. Carloni Gertosio*, J. Bobin

IRFU, CEA, Université Paris-Saclay, F-91191, Gif-sur-Yvette, France

Abstract

Tackling unsupervised source separation jointly with an additional inverse problem such as deconvolution is central for the analysis of multi-wavelength data. This becomes highly challenging when applied to large data sampled on the sphere such as those provided by wide-field observations in astrophysics, whose analysis requires the design of dedicated robust and yet effective algorithms. We therefore investigate a new joint deconvolution/sparse blind source separation method dedicated for data sampled on the sphere, coined SDecGMCA. It is based on a projected alternate least-squares minimization scheme, whose accuracy is proved to strongly rely on some regularization scheme in the present joint deconvolution/blind source separation setting. To this end, a regularization strategy is introduced that allows designing a new robust and effective algorithm, which is key to analyze large spherical data. Numerical experiments are carried out on toy examples and realistic astronomical data.

Keywords: blind source separation, sparsity, deconvolution, spherical data

Notations. Vectors are written in bold lowercase letters, such as $\mathbf{x}$. $\mathbf{x}_i$ denotes the ith entry of $\mathbf{x}$. Matrices are written in bold uppercase letters, such as $\mathbf{X}$. $\mathbf{X}_i$ and $\mathbf{X}^j$ denote respectively the ith row and j^{th} column of $\mathbf{X}$, and $\mathbf{X}_i^j$ denotes the $(i,\, j)^{\text{th}}$ entry of $\mathbf{X}$. $\|\cdot\|_p$ denotes the p-norm for vectors and the induced p-norm for matrices. $\|\cdot\|_{\ell_p}$ denotes the "entrywise" p-norm for matrices. The transpose operator is written with a superscript $\top$, such as $\mathbf{X}^\top$. The transpose-conjugate operator is noted with a superscript $\dagger$, such as $\mathbf{X}^\dagger$. $\mathbf{X} \geq 0$ means that the coefficients of $\mathbf{X}$ are non-negative. $\text{diag}(\mathbf{x})$ returns the diagonal matrix constituted of the coefficients of $\mathbf{x}$. $\odot$ denotes the element-wise product, also known as the Hadamard product. $*$ denotes the convolution operator on the sphere. The spherical harmonic projection of a vector $\mathbf{x}$ is written $\hat{\mathbf{x}}$. $\hat{\mathbf{X}}$ is defined as the stack of the spherical harmonic projections of the rows of $\mathbf{X}$. The estimate of a signal at iteration i is noted with a superscript in brackets, such as $\mathbf{X}^{(i)}$. The ground truth signal is written with an asterisk, such as $\mathbf{X}^*$.

1. Introduction

In a large number of applications, ranging from biomedical imaging to astrophysics, retrieving the relevant information from multichannel observations requires tackling an unsupervised matrix factorization problem dubbed Blind Source Separation (BSS). More precisely, we assume that such data are composed of N_c channels. For channel ν, a single observation, whether it represents an image or a 1D signal, can be described by the following linear mixture model:

$$\mathbf{X}_\nu = \sum_{n=1}^{N_s} \mathbf{A}_\nu^n \mathbf{S}_n + \mathbf{N}_\nu, \tag{1}$$

*Corresponding author.

Email addresses: remi.carlonigertosio@cea.fr (R. Carloni Gertosio), jerome.bobin@cea.fr (J. Bobin)

where the source $\mathbf{S}_n$ is composed of N_p entries or samples, $\mathbf{A}_\nu^n$ is a scalar that quantifies the mixture weight of source n in the observation ν, and $\mathbf{N}_\nu$ stands for some additive noise. The number of channels N_c is assumed greater or equal to the number of sources N_s; the resulting problem is complete or over-complete. Unsupervised source separation seeks estimates of $\mathbf{A}$ and $\mathbf{S}$ from the knowledge of $\mathbf{X}$ only. As this problem boils down to an unsupervised matrix factorisation problem, BSS is an ill-posed inverse problem, which requires additional prior information about the sources $\mathbf{S}$ and/or the so-called mixing matrix $\mathbf{A}$. So far, common assumptions are the statistical independence of the sources (Independent Component Analysis [1]), the non-negativity of the sources and the mixing matrix (Non-negative Matrix Factorisation [2, 3]). In the present paper, we will more specifically focus on imposing the sparsity of the sources in some representation domain, which has been widely showed to provide efficient separation procedures in a large number of applications [4, 5, 6, 7].

While this problem has now become standard and widely studied during the last three decades, it is much less commonplace when the observations are further distorted by channel-dependent measurement operators $\mathbf{H}_\nu$. For instance, large-band multi-wavelength observations have the particularity to have resolutions which can be significantly different between the observation channels. In this case, coping with now heterogeneous data requires tackling an extra deconvolution step, thus leading to a joint deconvolution and blind source separation (DBSS) problem. A mathematically similar problem arises when the observations are composed of incomplete measurements such as in interferometric measurements [8, 9] or compressive hyperspectral imaging [10, 11]. The above mixture model is then substituted with the following:

$$\mathbf{X}_\nu = \mathbf{H}_\nu(\mathbf{A}_\nu \mathbf{S}) + \mathbf{N}_\nu. \tag{2}$$

Being an ill-posed matrix factorization problem, BSS alone is already a challenging inverse problem. This is all-the-more complex when a channel-dependent operator further needs to be inverted as it can be ill-conditioned or even not invertible. Furthermore, standard BSS methods cannot be applied directly to the problem in Eq. (2) unless the data $\mathbf{X}$ are pre-processed so as to obtain new data with a common resolution. However, jointly solving both deconvolution and separation is expected to yield much better results, allowing to more precisely account for the forward observation model and noise in a single pass.
To the best of our knowledge, joint DBSS has been seldom investigated. The closest work known so far has been introduced by Kleinsteuber *et al.*, who proposed a BSS algorithm to analyze incomplete data in the framework of compressed sensing [12]. This can be regarded as a special case of DBSS where the measurement operator is defined as a projection onto a low-dimensional measurement subspace. However, the proposed method is not compatible in our case since it only applies to compressively sensed measurements.
More recently, we introduced the first joint DBSS method [9]. The proposed DecGMCA algorithm enforces the sparsity of the sources in some domain, that shall be represented by its transfer matrix $\mathbf{\Phi}$, by seeking a stationary point of the following cost function:

$$\min_{\mathbf{A},\mathbf{S}} \sum_{n=1}^{N_s} \lambda_n \left\| \mathbf{S}_n \mathbf{\Phi}^\top \right\|_1 + \frac{1}{2} \sum_{\nu=1}^{N_c} \left\| \mathbf{X}_\nu - \mathbf{H}_\nu \left(\mathbf{A}_\nu \mathbf{S} \right) \right\|_2^2 . \tag{3}$$

For that purpose, DecGMCA builds upon a projected Alternate Least-Squares (projected ALS) minimization procedure, where the sources and the mixing matrix are estimated sequentially. Projected ALS has been showed to provide computationally efficient and fast BSS algorithms. Furthermore, it allows the use of heuristics to automatically tune the regularization parameters $\{\lambda_n,\ n \in [1, N_s]\}$ [7]. However, in the setting of DBSS, resorting to a projected ALS optimization scheme also raises a major difficulty: the least-squares problem with respect to $\mathbf{S}$ is generally ill-conditioned – if not ill-posed – and needs to be regularized, which has a significant impact on the quality of the separation results.

Contributions. In this article, we investigate a new joint DBSS algorithm to analyze data that are sampled on the sphere. This is essential to cope with the kind of wide-field spherical data, which are now common

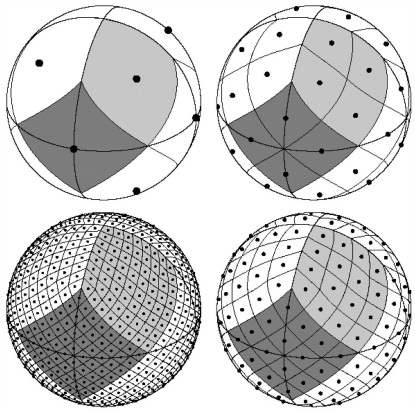


Figure 1: HEALPix discretization of the sphere (source: [13]). Moving clockwise from the top left, $N_{side} = 1, 2, 4, 8$

in scientific fields such as astronomy. This includes the analysis of the forthcoming Square Kilometer Array radiotelescope (SKA[1]) or X-ray observatories such as the future European mission Athena[2] to cite only two examples. In contrast to the standard case, analyzing spherical data raises extra difficulties due to the high computational cost of their manipulation, which makes essential the design of a computationally efficient and reliable algorithm. We therefore first aim at extending the algorithm DecGMCA [9] to tackle joint deconvolution and separation problems from spherical data. As described in Section 2, the method is based on a projected alternate least-squares minimization in order to combine rapidity and precision. Compared to the BSS problem, the procedure calls for an extra regularization to deal with a naturally ill-conditioned if not ill-posed problem. Yet, the regularization strategy and its impact on the solution has not been examined. Fot that purpose, we introduce in Section 2.3.1 several regularization strategies, which significantly improve the separation quality. Based on these results, we introduce in Section 3 a new algorithm, coined SDecGMCA, to tackle efficiently joint deconvolution and blind source separation problems. Finally, in Section 4, numerical experiments, which involve both toy examples and realistic astrophysical simulations, are presented.

2. The spherical DecGMCA algorithm

In this section, we first adapt the DBSS method proposed in [9] to tackle spherical data. Furthermore, we investigate in depth the *ad hoc*, yet necessary, regularization procedure in the update of the sources used in the DecGMCA algorithm. We show that it has a significant impact on the reconstruction quality of the sources. We then introduce two new regularization strategies, which noticeably outperform the one of DecGMCA.

2.1. Toward a DBSS method for spherical data

In the following, the multichannel data $\mathbf{X}$ are assumed to be sampled on the sphere with HEALPix [13]. As showed in Fig. 1, the sphere is divided in 12 quadrilateral sections of equal area, which are hierarchically subdivided in N_{side}^2 pixels. The resolution parameter N_{side} is the number of pixels along the sides of the 12 initial sections. HEALPix is the most commonly used pixelization of the sphere in astrophysics and geophysics, since it is well adapted for hierarchical analysis and spherical harmonic projections.

The measurement operators $\mathbf{H}_\nu$ are assumed linear and isotropic. Therefore, the mixture model (2)

[1] https://www.skatelescope.org/
[2] https://sci.esa.int/web/athena

becomes:

$$\mathbf{X}_\nu = (\mathbf{A}_\nu \mathbf{S}) * \mathbf{H}_\nu + \mathbf{N}_\nu, \tag{4}$$

where $*$ denotes the convolution operator on the sphere between a spherical function and an isotropic kernel [14]. The number of channels is recalled to be greater or equal to the number of sources.
Quite similarly to the Euclidean case, the spherical convolution operator is diagonal in the spherical harmonics basis. Spherical harmonics are ordered by degree $l \geq 0$ and mode $m \in [-l, l]$. The harmonic coefficients $\mathcal{D} = \{(l, m) \in \mathbb{N} \times \mathbb{Z},\ |m| \leq l\}$, also called multipoles, are comparable to spatial frequencies. Equivalently to a Fourier decomposition, any spherical function is uniquely decomposed in a linear combination of spherical harmonics. Let $\mathbf{y_{l,m}}$ denote the $(l, m)^{\text{th}}$ normalized spherical harmonic sampled with HEALPix. The projection in the spherical harmonic space of a vector $\mathbf{x}$ is given by:

$$\hat{\mathbf{x}}_{l,m} = \frac{4\pi}{N_p} \mathbf{y_{l,m}}^\dagger \mathbf{x}, \tag{5}$$

where $N_p = 12N_{side}^2$ is the number of pixels. The spherical harmonics sampled with HEALPix form a linearly independent set up to $l_{max} = 3N_{side} - 1$. Therefore, the projections will be limited in frequency to l_{max}. For the sake of clarity, this limitation will not be indicated in the following equations.
Equation (4) can be recast in the spherical harmonic domain:

$$\hat{\mathbf{X}}_\nu^{l,m} = \hat{\mathbf{H}}_\nu^l \mathbf{A}_\nu \hat{\mathbf{S}}^{l,m} + \hat{\mathbf{N}}_\nu^{l,m}, \tag{6}$$

where the superscript (l, m) refers to the column of the matrix with degree l and mode m, in accordance with Eq. (5). The convolution kernels being isotropic, $\hat{\mathbf{H}}$ does not depend on m. The mixing model is finally rewritten by grouping the channels, yielding:

$$\hat{\mathbf{X}}^{l,m} = \operatorname{diag}\left(\hat{\mathbf{H}}^l\right) \mathbf{A}\hat{\mathbf{S}}^{l,m} + \hat{\mathbf{N}}^{l,m}. \tag{7}$$

To build an estimator for the sources $\mathbf{S}$ and the mixing matrix $\mathbf{A}$, the following assumptions are made:

- *Data-fidelity term:* Thereafter, the noise is assumed to be Gaussian; the data-fidelity term is then defined as the Euclidean distance between the observations and the forward model as follows:

$$h(\mathbf{A}, \mathbf{S}) = \sum_{(l,m)\in\mathcal{D}} \left\| \hat{\mathbf{X}}^{l,m} - \operatorname{diag}\left(\hat{\mathbf{H}}^l\right) \mathbf{A}\hat{\mathbf{S}}^{l,m} \right\|_2^2. \tag{8}$$

- *Source regularization:* The sources are assumed to be sparse in a given dictionary $\mathbf{\Phi}$. This sparsity can be enforced by a ℓ_1-penalization in the transformed domain, hence the constraint term on $\mathbf{S}$:

$$\mathcal{G}(\mathbf{S}) = \left\| \mathbf{\Lambda} \odot \left(\mathbf{S}\mathbf{\Phi}^\top\right) \right\|_{\ell_1}, \tag{9}$$

  where $\mathbf{\Lambda}$ is a matrix that stores the thresholding parameters, which will be described in detail in Section 3.1. For the sake of simplicity, we will assume that the sparse domain is orthogonal, *ie.* that $\mathbf{\Phi}$ is a square matrix verifying $\mathbf{\Phi}^\top\mathbf{\Phi} = \mathbf{I}$, with $\mathbf{I}$ the identity matrix. However, as mentioned in [6], the forthcoming mathematical derivations also apply to tight frames at first order approximation [15].

  In some applications, the sources are expected to be non-negative. In this case, one can also impose a non-negativity constraint on $\mathbf{S}$, yielding:

$$\mathcal{G}(\mathbf{S}) = \left\| \mathbf{\Lambda} \odot \left(\mathbf{S}\mathbf{\Phi}^\top\right) \right\|_{\ell_1} + \chi_{\mathcal{K}_\mathcal{S}}(\mathbf{S}), \tag{10}$$

  where $\chi_{\mathcal{K}_\mathcal{S}}$ is the characteristic function of the positive orthant for the sources $\mathcal{K}_\mathcal{S} = \{\mathbf{S} \in \mathbb{R}^{N_s \times N_p},\ \mathbf{S} \geq 0\}$.

- *Mixing matrix regularization:* To mitigate the scale indeterminacy of the product $\mathbf{A}\hat{\mathbf{S}}$, the columns of $\mathbf{A}$ are enforced to belong to the ℓ_2-hypersphere, which is defined as the set of vectors with unit ℓ_2 norm:

$$\mathcal{J}(\mathbf{A}) = \chi_{\mathcal{O}}(\mathbf{A}), \tag{11}$$

where $\chi_{\mathcal{O}}$ is the characteristic function of the oblique ensemble $\mathcal{O} = \{\mathbf{A} \in \mathbb{R}^{N_c \times N_s}, \forall n \in [1, N_s], \|\mathbf{A}^n\|_2 = 1\}$.

In addition, it is possible to impose the non-negativity of the mixing matrix; in this case:

$$\mathcal{J}(\mathbf{A}) = \chi_{\mathcal{O}}(\mathbf{A}) + \chi_{\mathcal{K}_{\mathcal{A}}}(\mathbf{A}), \tag{12}$$

where $\chi_{\mathcal{K}_{\mathcal{A}}}$ is the characteristic function of the positive orthant for the mixing matrices.

To sum up, the mixing matrix and the sources will be estimated by looking for a stationary point of the following cost function:

$$\min_{\mathbf{A},\hat{\mathbf{S}}} \left\| \mathbf{\Lambda} \odot \left(\hat{\mathbf{S}} \mathcal{F}^{\dagger} \mathbf{\Phi}^{\top} \right) \right\|_{\ell_1} + \chi_{\mathcal{O}}(\mathbf{A}) + \sum_{(l,m)\in\mathcal{D}} \left\| \hat{\mathbf{X}}^{l,m} - \operatorname{diag}\left(\hat{\mathbf{H}}^{l}\right) \mathbf{A} \hat{\mathbf{S}}^{l,m} \right\|_2^2, \tag{13}$$

plus the characteristic functions of the positive orthants, if the sources and/or the mixing matrix are non-negative. $\mathcal{F}^{\dagger}$ is the inverse spherical harmonic transform.

Being an unsupervised matrix factorization task, the problem in Equation (13) is a challenging non-convex problem. More precisely, it is multi-convex [16] since it is convex with respect to $\mathbf{S}$ (*resp.* $\mathbf{A}$) when $\mathbf{A}$ (*resp.* $\mathbf{S}$) is fixed. In this context, most standard algorithms perform with alternating minimization steps, with respect to the mixing matrix $\mathbf{A}$ and the sources $\mathbf{S}$ in a sequential manner.
When it comes to DBSS, a major challenge is the need for a robust and yet effective algorithm, but with a reasonable computational burden. As the above optimization combines non-differentiable terms, such as the sparsity-enforcing ℓ_1 regularization term or the characteristic function of the Oblique ensemble $\chi_{\mathcal{O}}$, it would make perfect sense to design a minimizer based on plain proximal algorithms [17]. However, this would be lead to algorithms with very high computational cost, especially if one considers that DBSS problems are generally ill-conditioned. Moreover, when BSS only is considered, tuning the regularization parameters, which has a paramount impact on the separation performances, is particularly complex with standard proximal algorithms [7].
In contrast, projected Alternating Least-Squares (pALS), which were first introduced in the field of Non-Negative Matrix Factorization [18], have been proved to be particularly effective at providing fast sparse BSS algorithms. Furthermore, they come with almost automatic strategies to fix the regularization parameters in a robust way [6, 7]. In a nutshell, when either $\mathbf{A}$ or $\mathbf{S}$ is updated, a first least-squares estimate is computed by minimizing the data-fidelity term. This step is then followed by the application of the proximal operator of the corresponding regularization term[3]. These two steps are detailed in the following sections.

2.2. Update of $\mathbf{A}$

Updating the mixing matrix $\mathbf{A}$ when the sources $\mathbf{S}$ are fixed requires solving the least-squares problem $\underset{\mathbf{A}}{\operatorname{argmin}}\, h(\mathbf{A}, \mathbf{S})$, which yields for all channel $\nu \in [1, N_c]$:

$$\mathbf{A}_{\nu} \leftarrow \left(\sum_{(l,m)\in\mathcal{D}} \hat{\mathbf{X}}_{\nu}^{l,m} \hat{\mathbf{H}}_{\nu}^{l} \hat{\mathbf{S}}^{l,m\,\dagger} \right) \left(\sum_{(l,m)\in\mathcal{D}} \hat{\mathbf{H}}_{\nu}^{l\;2} \hat{\mathbf{S}}^{l,m} \hat{\mathbf{S}}^{l,m\,\dagger} \right)^{-1}. \tag{14}$$

[3] We refer to [17] for insightful details about proximal calculus and algorithms.

In this equation, the estimate is defined with sums over all multipoles, which are significantly larger than the number of sources N_s. Consequently, the matrix $\left(\sum_{(l,m)\in\mathcal{D}} \hat{\mathbf{H}}_\nu^{l\,2} \hat{\mathbf{S}}^{l,m} \hat{\mathbf{S}}^{l,m\,\dagger}\right)$ is generally invertible if not well conditioned.
In a second phase, the proximal operator of $\mathcal{J}$ is applied on the solution. If there is no constraint on the sign of $\mathbf{A}$, it is the projection on the multidimensional ℓ_2-hypersphere $\mathcal{O}$:

$$\mathbf{A}^n \leftarrow \Pi_{\mathcal{O}}(\mathbf{A}) = \frac{\mathbf{A}^n}{\|\mathbf{A}^n\|_2}. \tag{15}$$

If $\mathbf{A}$ is non-negative, the proximal operator of $\mathcal{J}$ is the composition of the proximal operators of the oblique and non-negativity constraints:

$$\mathbf{A} \leftarrow \Pi_{\mathcal{O}}\left(\Pi_{\mathcal{K}_{\mathcal{A}}}(\mathbf{A})\right). \tag{16}$$

2.3. Update of $\mathbf{S}$

Updating $\mathbf{S}$ assuming $\mathbf{A}$ is fixed leads to the following optimization problem:

$$\min_{\hat{\mathbf{S}}} \left\|\mathbf{\Lambda} \odot \left(\hat{\mathbf{S}}\mathcal{F}^H \mathbf{\Phi}^\top\right)\right\|_{\ell_1} + \sum_{(l,m)\in\mathcal{D}} \left\|\hat{\mathbf{X}}^{l,m} - \operatorname{diag}\left(\hat{\mathbf{H}}^l\right) \mathbf{A}\hat{\mathbf{S}}^{l,m}\right\|_2^2, \tag{17}$$

to which the term $\chi_{\mathcal{K}_{\mathcal{S}}}(\mathbf{S})$ is added if the sources are non-negative.
According to the pALS minimization scheme, the solution of the above minimization problem is first approximated with a least-squares estimate by finding a solution to:

$$\min_{\hat{\mathbf{S}}} \sum_{(l,m)\in\mathcal{D}} \left\|\hat{\mathbf{X}}^{l,m} - \operatorname{diag}\left(\hat{\mathbf{H}}^l\right) \mathbf{A}\hat{\mathbf{S}}^{l,m}\right\|_2^2. \tag{18}$$

Solving the least-squares problem $\underset{\mathbf{S}}{\operatorname{argmin}}\, h(\mathbf{A}, \mathbf{S})$ yields for all harmonic coefficient $(l,m) \in \mathcal{D}$:

$$\hat{\mathbf{S}}^{l,m} \leftarrow \mathbf{M}[l]^{-1}\mathbf{A}^\top \operatorname{diag}\left(\hat{\mathbf{H}}^l\right) \hat{\mathbf{X}}^{l,m}, \tag{19}$$

with $\mathbf{M}[l] = \mathbf{A}^\top \operatorname{diag}\left(\hat{\mathbf{H}}^l\right)^2 \mathbf{A}$.
It is important to notice that, in contrast to standard BSS problems, the above solution is not necessarily stable with respect to noise as the matrices $\mathbf{M}[l]$ may be ill-conditioned. Additionally, it is not always unique as $\mathbf{M}[l]$ may not be invertible; this might occur when the convolution kernels vanish for some spherical harmonics multipoles.
Borrowing ideas from sparsity enforcing deconvolution methods such as ForWard [19], we proposed in [9] to add an extra Tikhonov regularization [20] to the least-squares problem:

$$\min_{\hat{\mathbf{S}}} \frac{1}{2}\sum_{n=1}^{N_s} \sum_{(l,m)\in\mathcal{D}} \varepsilon_{n,l} \left|\hat{\mathbf{S}}_n^{l,m}\right|^2 + \left\|\hat{\mathbf{X}}^{l,m} - \operatorname{diag}\left(\hat{\mathbf{H}}^l\right) \mathbf{A}\hat{\mathbf{S}}^{l,m}\right\|_2^2. \tag{20}$$

The set $\{\varepsilon_{n,l} \geq 0,\, n \in [1, N_s],\, l \in \mathbb{N}\}$ are the Tikhonov regularization coefficients. They depend on the frequency l and on the source n. For all $(l,m) \in \mathcal{D}$, updating the sources can now be recast as:

$$\hat{\mathbf{S}}^{l,m} \leftarrow \left(\mathbf{M}[l] + \operatorname*{diag}_{n\in[1,N_s]}(\varepsilon_{n,l})\right)^{-1} \mathbf{A}^\top \operatorname{diag}\left(\hat{\mathbf{H}}^l\right) \hat{\mathbf{X}}^{l,m}. \tag{21}$$

Similarly to standard deconvolution problems [19], the way the Tikhonov regularization coefficients are fixed has a dramatic impact on the quality of the regularized least-squares solution, and eventually on the whole

separation process. This is discussed in depth in the following section.

In a second step, the proximal operator of $\mathcal{G}$ is applied to the above least-squares solution. If there is no constraint on the sign of $\mathbf{S}$, this amounts to applying the soft-thresholding operator $\mathcal{S}_{\mathbf{\Lambda}}$, with thresholds $\mathbf{\Lambda}$:

$$\mathbf{S} \leftarrow \mathcal{S}_{\mathbf{\Lambda}} \left(\mathbf{S}\mathbf{\Phi}^\top \right) \mathbf{\Phi}. \tag{22}$$

Under the non-negativity constraint, the proximal operator of $\mathcal{G}$ has no analytical form. It is approximated as the composition of the proximal operator of the sparsity constraint and the proximal operator of the non-negativity constraint, which is the projection on the non-negative orthant $\Pi_{\mathcal{K}_\mathcal{S}}$:

$$\mathbf{S} \leftarrow \Pi_{\mathcal{K}_\mathcal{S}} \left(\mathcal{S}_{\mathbf{\Lambda}} \left(\mathbf{S}\mathbf{\Phi}^\top \right) \mathbf{\Phi} \right). \tag{23}$$

2.3.1. Regularization strategies

In [9], the regularization strategy is defined quite arbitrarily with hyperparameters that are fixed to an *ad hoc* small value (*e.g.* 10^{-3}). In this section, we will particularly highlight that this regularization significantly impacts the estimation precision. We further investigate different strategies allowing more efficient and adaptive way of tuning these key parameters.
Let c be a positive number, that will be called the regularization hyperparameter. We will further investigate four different regularization strategies:

- **Strategy #1** Let us first consider the naive strategy where the regularization parameters are chosen independently of the source n and the frequency l:

$$\varepsilon_{n,l} = c. \tag{24}$$

- **Strategy #2** The strategy presented in [9] is also considered:

$$\varepsilon_{n,l} = c\, \lambda_{\max}(\mathbf{M}[l]), \tag{25}$$

  where $\lambda_{\max}(\cdot)$ returns the greatest eigenvalue. In [9], the main motivation of this choice was to set a regularization parameters that scales with the sources.

- **Strategy #3** In Eq. (19), the errors that contaminate the observations are amplified by the inverse of the smallest eigenvalue of $\mathbf{M}[l]$, denoted $\lambda_{\min}(\mathbf{M}[l])^{-1}$. Limiting the noise amplification to c amounts to choosing $\varepsilon_{n,l}$ such that $\varepsilon_{n,l} + \lambda_{\min}(\mathbf{M}[l]) \leq c \iff \varepsilon_{n,l} \leq c - \lambda_{\min}(\mathbf{M}[l])$. Bearing in mind that $\varepsilon_{n,l} \geq 0$, it is possible to set $\varepsilon_{n,l} = \max\left(0, c - \lambda_{\min}\left(\mathbf{M}[l]\right)\right)$. A change of variable is finally operated to facilitate the interpretation of the hyperparameter c, yielding for all source $n \in [1, N_s]$ and frequency $l \in \mathbb{N}$:

$$\varepsilon_{n,l} = \max\left(0, c - \frac{\lambda_{\min}(\mathbf{M}[l])}{\lambda_{\min}(\mathbf{A}^\top\mathbf{A}) + \epsilon}\right), \tag{26}$$

  with $\epsilon = 1\text{e}{-2}$, to prevent numerical issues. Since the sequence $(\lambda_{\min}(\mathbf{M}[l]))_{l\in\mathbb{N}}$ decreases, the sequence $(\varepsilon_{n,l})_{l\in\mathbb{N}}$ increases. Consequently, the higher frequencies are more penalized, while the lower ones are preserved. This is advantageous because most of the information from a physical source generally lies in the lower frequencies. Moreover, the maximum operator in (26) allows to have $\varepsilon_{n,l} = 0$ for the smaller frequencies, provided that c is small enough. This allows to keep the smaller frequencies unbiased.

- **Strategy #4** As in the Euclidean case, it is possible to define an angular power spectrum, which describes the power distribution along the degrees l. It is likewise deduced from the harmonic decomposition:

$$\mathbf{c}_{\mathbf{x}l} = \frac{1}{2l+1} \sum_{m=-l}^{l} \left|\hat{\mathbf{x}}_{l,m}\right|^2. \tag{27}$$

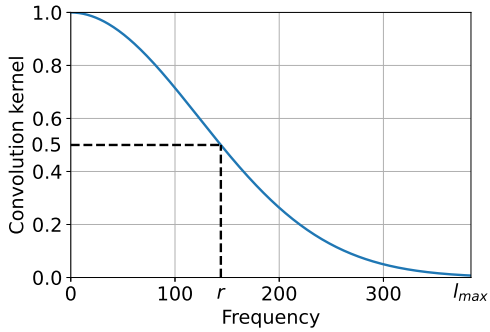


Figure 2: Example of a Gaussian filter $\hat{\mathbf{H}}_\nu$, with a resolution $r = 144$

If the angular power spectra of the sources $\{\mathbf{c}_{\mathbf{S}_n}, n \in [1, N_s]\}$ and the noise $\mathbf{c}_\mathbf{N}$ (which is assumed observation independent) are known, the optimal strategy for the unpenalized least-squares problem is $\varepsilon_{n,l} = \mathbf{c}_{\mathbf{N}l}/\mathbf{c}_{\mathbf{S}_n l}$. This is reminiscent of a Wiener deconvolution filter. Yet, it has to be reminded that the update of $\mathbf{S}$ is further followed by a thresholding step, which likely alters the properties of the regularization strategy. Therefore, the regularization strategy is adapted by adding a hyperparameter c, yielding for all sources $n \in [1, N_s]$ and frequencies $l \in \mathbb{N}$:

$$\varepsilon_{n,l} = c\,\frac{\mathbf{c}_{\mathbf{N}l}}{\mathbf{c}_{\mathbf{S}_n l}}. \tag{28}$$

c tips the balance between the Tikhonov regularization and the sparsity regularization.

2.3.2. Numerical comparisons of the regularization strategies

In this paragraph, we propose illustrating the impact of the above regularization strategies with numerical experiments. For that purpose, we consider the non-blind separation case, *ie.* with the ground truth mixing matrix $\mathbf{A}^*$ known.

In these experiments, the data are sampled on the sphere using the HEALPix pixelization, with parameters $N_{side} = 128$ and $l_{max} = 384$. The observations are unmixed and deconvolved at the resolution of the best-resolved channel. This amounts to replacing $\hat{\mathbf{H}}^l_\nu$ by $\hat{\mathbf{H}}^l_\nu/\hat{\mathbf{H}}^l_{\nu_b}$, where ν_b is the index of the best-resolved channel.

In these comparisons, the data are generated as follows:

- the sources $\mathbf{S}$ are random non-negative signals that are sparse in the spherical starlet (isotropic undecimated wavelets) domain [21]; they are also band-limited, with a frequency cut-off of $l_{max}/6 = 64$. See example in Fig. 4a.
- the mixing matrices $\mathbf{A}$ are random non-negative matrices with a given condition number.
- the convolution kernels $\hat{\mathbf{H}}^l_\nu = \exp\left(-\frac{l(l+1)}{r(\nu)(r(\nu)+1)}\log 2\right)$ are Gaussian-shaped, with resolutions r evenly spread between the minimum resolution r_{min}, which is a parameter to be set, and l_{max}. The resolution is defined as the full width at half maximum (FWHM) in the spherical harmonic domain of the convolution kernel (see example in Fig. 2).

Throughout this section, we consider $N_s = 4$ sources. Moreover, we define four parameters that characterize the observations:

- number of observations N_c ($N_c = 8$ by default),
- mixing matrix condition number $\mathrm{cond}(\mathbf{A})$ ($\mathrm{cond}(\mathbf{A}) = 2$ by default),

- minimum resolution r_{min} ($l_{max}/8 = 48$ by default),
- overall signal-to-noise ratio (10 dB by default).

The joint deconvolution and non-blind separation problem is solved by performing the update of $\mathbf{S}$ proposed in 2.3, using the ground truth mixing matrix $\mathbf{A}^*$. The update comprises a Tikhonov-penalized least square followed by a soft-thresholding. For strategy #4, the regularization parameters are calculated with the angular power spectra of the ground truth sources. The sparsity of the sources is enforced in the spherical starlet domain. The thresholds are set in the same way than during the finale iteration of SDecGMCA; the strategy for the choice of the thresholds in SDecGMCA is discussed in detail in Section 3.1.

The results are assessed with the normalized mean square error (NMSE) on the sources, defined by:

$$\text{NMSE} = -10 \log_{10} \left(\frac{\|\mathbf{S}^* - \mathbf{S}\|_{\ell_2}^2}{\|\mathbf{S}^*\|_{\ell_2}^2} \right), \tag{29}$$

with $\mathbf{S}^*$ the ground truth sources and $\mathbf{S}$ the estimated sources.

We test how the four regularization strategies we defined in the previous section behave with respect to the number of observations, the mixing matrix condition number, the minimum resolution of the convolution kernel and the SNR. Each experiment is composed of 100 trials, with varying random sources, mixing matrices and noise realizations. For each point, the optimal regularization hyperparameter is fine-tuned using a grid-search. The performances of each strategy is evaluated based on the NMSE of the estimated sources.
The results are reported in detail in Appendix C. Strategy #4 clearly provides the best reconstruction qualities, and does so in various observation scenarios (varying number of observations, mixing matrix condition number, minimum resolution and SNR). Moreover, the optimal regularization hyperparameter of strategy #4 is remarkably insensitive to the observation parameters; typically $c_{opt} \sim 0.5$. Yet, strategy #4 requires to know the spectra of the sought sources, or at least an estimation of these. Among the other regularization strategies, which do not necessitate prior information on the sources, strategy #3 achieves better results both in terms of reconstruction quality and hyperparameter sensitivity.

3. Implementation of the SDecGMCA algorithm

We shedded light on how critical the choice of the Tikhonov regularization strategy is. In this section, the details of implementation of SDecGMCA are described. In particular, a novel two-stage algorithm which exploits the proposed regularization strategies is introduced.

SDecGMCA is summarized in Algorithm 1. The mixing matrix $\mathbf{A}$ is initialized with a Principal Component Analysis (PCA) performed on the observations. For this step only, the data are first re-convolved so as they share a common resolution. To avoid noise amplification, such resolution is the one of the worse-resolved channel. SDecGMCA is comprised of two stages: the first stage estimates a first guess of the mixing matrix and the sources (**warm-up**); it is required to provide robustness with respect to the initial point. The second stage refines the separation by employing a more precise regularization strategy (**refinement**). The values of the parameters of SDecGMCA according to the stage are summarized in Table 1.

3.1. Choice of the thresholding parameters

As emphasized in [7], the thresholding parameters play a central role for the robustness of the minimization scheme with respect to spurious local minima and the precision of the estimates. To that respect, this role is twofold:

Algorithm 1 SDecGMCA

Inputs: $\mathbf{X}$, $\hat{\mathbf{H}}$, N_s, c_{wu}, c_{ref}, σ^2, $\mathbf{\Phi}$, $k = 3$, $K_{max} = 0.5$

$i \leftarrow 0$
PCA initialization of $\mathbf{A}^{(0)}$
stage $\leftarrow$ warm-up `\\Stage 1: warm-up`
while convergence not reached **do**
  $i \leftarrow i + 1$
  Update c and K according to Table 1
  Estimate $\hat{\mathbf{S}}$ with $\mathbf{A}$ fixed: $\hat{\mathbf{S}}^{(i)} \leftarrow \text{UpdateS}\left(\hat{\mathbf{X}},\ \hat{\mathbf{H}},\ \mathbf{A}^{(i-1)},\ \text{stage},\ c,\ \sigma^2,\ \mathbf{\Phi},\ k,\ K\right)$
  Estimate $\mathbf{A}$ with $\hat{\mathbf{S}}$ fixed: $\mathbf{A}^{(i)} \leftarrow \text{UpdateA}\left(\hat{\mathbf{X}},\ \hat{\mathbf{H}},\ \hat{\mathbf{S}}^{(i)}\right)$
stage $\leftarrow$ refinement `\\Stage 2: refinement`
while convergence not reached **do**
  $i \leftarrow i + 1$
  Estimate $\hat{\mathbf{S}}$ with $\mathbf{A}$ fixed: $\hat{\mathbf{S}}^{(i)} \leftarrow \text{UpdateS}\left(\hat{\mathbf{X}},\ \hat{\mathbf{H}},\ \mathbf{A}^{(i-1)},\ \text{stage},\ c_{ref},\ \sigma^2,\ \mathbf{\Phi},\ k,\ K_{max}\right)$
  Estimate $\mathbf{A}$ with $\hat{\mathbf{S}}$ fixed: $\mathbf{A}^{(i)} \leftarrow \text{UpdateA}\left(\hat{\mathbf{X}},\ \hat{\mathbf{H}},\ \hat{\mathbf{S}}^{(i)}\right)$
return $\mathbf{A}^{(i)},\ \mathbf{S}^{(i)}$

Algorithm 2 UpdateS

Inputs: $\hat{\mathbf{X}}$, $\hat{\mathbf{H}}$, $\mathbf{A}$, stage, c, σ^2, $\mathbf{\Phi}$, k, K

Calculate the regularization parameters $\{\varepsilon_{n,l}\}$ according to the current stage
for $l = 0, 1, ..., l_{max}$ and $m = -l, -l+1, ..., l$ **do**
  Tikhonov-penalized least-squares update of $\hat{\mathbf{S}}$:

$$\hat{\mathbf{S}}^{l,m} \leftarrow \left(\mathbf{A}^\top \operatorname{diag}\left(\hat{\mathbf{H}}^l\right)^2 \mathbf{A} + \underset{n\in[1,N_s]}{\operatorname{diag}}\left(\varepsilon_{n,l}\right)\right)^{-1} \mathbf{A}^\top \operatorname{diag}\left(\hat{\mathbf{H}}^l\right) \hat{\mathbf{X}}^{l,m}$$

Calculate the thresholds $\mathbf{\Lambda}$, with a support-based strategy and a ℓ_1-reweighting if need be, according to the current stage
Soft-threshold $\mathbf{S}$: $\mathbf{S} \leftarrow \mathcal{S}_{\mathbf{\Lambda}}\left(\mathbf{S}\mathbf{\Phi}^\top\right)\mathbf{\Phi}$
Project $\mathbf{S}$ on the positive orthant (if non-negativity constraint): $\mathbf{S} \leftarrow \max\left(\mathbf{S},\ \mathbf{0}\right)$
return $\hat{\mathbf{S}}$

Algorithm 3 UpdateA

Inputs: $\hat{\mathbf{X}}$, $\hat{\mathbf{H}}$, $\hat{\mathbf{S}}$

for $\nu = 1, 2, ..., N_p$ **do**
  Least-squares update of $\mathbf{A}$:

$$\mathbf{A}_\nu \leftarrow \left(\sum_{(l,m)\in\mathcal{D}} \hat{\mathbf{X}}_\nu^{l,m} \hat{\mathbf{H}}_\nu^l \hat{\mathbf{S}}^{l,m\,\dagger}\right)\left(\sum_{(l,m)\in\mathcal{D}} \hat{\mathbf{H}}_\nu^{l\,2} \hat{\mathbf{S}}^{l,m} \hat{\mathbf{S}}^{l,m\,\dagger}\right)^{-1}$$

for $n = 1, 2, ..., N_s$ **do**
  Normalization of $\mathbf{A}^n$: $\mathbf{A}^n \leftarrow \dfrac{\mathbf{A}^n}{\|\mathbf{A}^n\|_2}$
Project $\mathbf{A}$ on the positive orthant (if non-negativity constraint): $\mathbf{A} \leftarrow \max\left(\mathbf{A},\ \mathbf{0}\right)$
return $\mathbf{A}$

Stage	1: warm-up	2: refinement
Regularization strategy	#3	#4
Regularization hyperparameter c	$10\,c_{wu} \rightarrow c_{wu}$	c_{ref}
Active source support K	$0 \rightarrow K_{max}$	K_{max}
ℓ_1-reweighting	No	Yes
Non-negativity constraint on $\mathbf{S}$	No	Yes

Table 1: Parameters of SDecGMCA for each stage

- Robustness with respect to noise: The strength of GMCA lies in proposing an adapted thresholding strategy. The thresholds $\mathbf{\Lambda}$ are calculated at each iteration, for each source $n \in [1, N_s]$ by:

$$\mathbf{\Lambda}_n = k\,\sigma_{\mathbf{S}_n}\,\mathbf{1}, \tag{30}$$

  where $\sigma_{\mathbf{S}_n}$ is the standard deviation of the noise which contaminates $(\mathbf{S}\mathbf{\Phi}^\top)_n$ and k is a parameter adjusting the sparsity intensity. In practice, $k = 3$ in order to reject samples that are likely to be noise-related. When the noise level that contaminates the data is known, $\sigma_{\mathbf{S}_n}$ can be evaluated analytically; this is described in details in Appendix B. This is however not always the case in most applications in BSS; when the noise level remains unknown, it can be computed empirically with the Median Absolute Deviation applied to the sources in the sparse domain. We refer the interested reader to [7] for more details.

- Decreasing thresholding strategy: It has long been showed empirically that applying a decreasing threshold strategy dramatically improves the robustness of the separation process with respect to local spurious minima [22, 6, 7]. In the present article, we adapt this strategy to the context of DBSS. To that end, the thresholds are set so as to keep only a fixed amount of samples, say the K % samples with the greatest amplitudes, above the noise-related threshold $k\,\sigma_{\mathbf{S}_n}$. The percentage K linearly increases along the iterations from 0 to K_{max}. This procedure (i) enforces the decrease of the thresholds, thus improving the robustness of the separation, and (ii) provides an additional implicit regularization, by selecting only the most significant samples. K_{max} sets the intensity of the two aforementioned phenomena. More specifically, the smaller the K_{max}, the slower the decrease of the thresholds and the greater the regularization. In practice, K_{max} can be set to 0.5.
  Assuming that the coefficients of $(\mathbf{S}\mathbf{\Phi}^\top)_n$ are sorted in order of descending modulus, this procedure is implemented for all source $n \in [1, N_s]$ by:

$$\mathbf{\Lambda}_n = \left|(\mathbf{S}\mathbf{\Phi}^\top)_n^{p_0}\right| \mathbf{1}, \tag{31}$$

  with $p_0 = \left\lfloor K \operatorname{card}\left(\left|(\mathbf{S}\mathbf{\Phi}^\top)_n\right| \geq k\,\sigma_{\mathbf{S}_n}\right)\right\rfloor$, $K \in]0, 1]$ the active source support and $\mathbf{1}$ a vector of ones.

The soft-thresholding induced by the ℓ_1-penalization introduces a bias in the estimation of the sources. Resorting to a ℓ_1-reweighting scheme [23, 7] allows to reduce this phenomenon. It consists in deriving sample-wise thresholds based on the samples' value in the sparse domain; the greater the sample's amplitude, the smaller the threshold and thus the smaller the bias. Assuming that $\lambda_n^{(i)}$ is the initial non-sample-dependent threshold of source $n \in [1, N_s]$ at iteration i, the ℓ_1-reweighted threshold for sample $p \in [1, N_p]$ is given by:

$$\mathbf{\Lambda}_n^{p\,(i)} = \frac{\lambda_n^{(i)}}{1 + \frac{\left|(\mathbf{S}\mathbf{\Phi}^\top)_n^{p\,(i-1)}\right|}{\lambda_n^{(i)}}}. \tag{32}$$

3.2. A two-stage minimization approach

We highlighted in Section 2.3.2 that the regularization strategy #4 provides significantly better results in the non-blind source separation case. Since, this strategy is defined based on some estimate of the sources, we propose proceeding with a two-step approach: i) a warm-up step, whose goal is to provide a quick rough estimate of $\mathbf{A}$ and $\mathbf{S}$, with increased robustness with respect to the initialization and spurious local minima, and ii) a refinement step that makes use of regularization strategy #4.

Warm-up stage. During the warm-up, regularization strategy #3 is employed. Indeed, as seen in Section 2.3.2, it provides the best results among the three first regularization strategies, which require no additional information about the sources. During this stage, the estimates of the sources are likely to be dominated by the artifacts due to the noise amplification, especially if the regularization hyperparameter

at warm-up c_{wu} is too small. The estimation of $\mathbf{A}$ is then erroneous and the algorithm might get stuck in a local minimum. This phenomenon can be alleviated by over-regularizing the sources during the first iterations, *i.e.* by overestimating c_{wu} [9]. Doing so, the noise contamination is reduced and only the major features of the sources are kept. In the spirit of [9], the regularization hyperparameter is then progressively decreased along the iterations to the input value, in order to refine the estimations of $\mathbf{S}$ and consequently $\mathbf{A}$. The decrease of the warm-up regularization hyperparameter markedly improves the robustness of the separation in terms of convergence. Moreover, the earlier mentioned support-based strategy is achieved during the warm-up. Since the starting point of this step is likely to be quite far from the sought-after sources, no reweighted ℓ_1 is applied as it would tend to favor spurious solutions. For the same reason, the non-negativity constraint is likewise not applied on the sources during this stage (but it is for $\mathbf{A}$). The warm-up ends when the decrease of c_{wu} and the increase of K are completed, and when the estimations of the sources have converged, that is when $||\hat{\mathbf{S}}^{(i)} - \hat{\mathbf{S}}^{(i-1)}||_{\ell_2}/||\hat{\mathbf{S}}^{(i)}||_{\ell_2} \leq \epsilon_{wu}$. In practice, ϵ_{wu} does not require to be very small (*e.g.* $\epsilon_{wu} = 1\text{e}{-2}$).

Refinement stage. During the second stage, the estimations are refined by using the more precise regularization strategy #4 introduced in 2.3.1. The regularization parameters are calculated with the angular power spectra of the sources estimated at the previous iteration, which are assumed to be close enough to the ground truth ones:

$$\varepsilon_{n,l}^{(i)} = c_{ref} \frac{\mathbf{c}_{\mathbf{N}l}}{\mathbf{c}_{\mathbf{S}_n^{(i-1)}l}}, \tag{33}$$

where c_{ref} is the regularization hyperparameter at refinement. $\mathbf{c}_{\mathbf{N}l}$ is deduced from the SNR of the observations. Concerning the choice of the thresholds, the active support is kept constant at the same K_{max} as the previous stage. The ℓ_1-reweighting of the sources we described above is applied during this stage, as well as the non-negativity constraint on the sources. The refinement ends when the estimations of the sources have converged, that is when $||\hat{\mathbf{S}}^{(i)} - \hat{\mathbf{S}}^{(i-1)}||_{\ell_2}/||\hat{\mathbf{S}}^{(i)}||_{\ell_2} \leq \epsilon_{ref}$ (for instance $\epsilon_{ref} = 1\text{e}{-5}$).
Setting K_{max} to a different value than 1 allows to keep the major features of the sources to improve the separation. However, it biases the estimation of the sources. Consequently, after the refinement stage, a final estimation of the sources with $\mathbf{A}$ fixed and with $K = 1$ is performed. This final step does not appear in Algorithm 1 for the sake of clarity; it is however described in Appendix B.

3.3. Convergence proprieties

DBSS requires solving a multi-convex optimization problem. No method can guarantee finding in general the global minimum. At best, it is possible to guarantee a convergence toward a local minimum (*e.g.* BCD [24], PALM [25]). SDecGMCA is built upon a projected ALS scheme. To the best of our knowledge, is has not been theoretically proved that projected ALS algorithms converge. However, we empirically show that the regularization parameters tend to stabilize along the iteration, as well as the estimates of $\mathbf{A}$ and $\mathbf{S}$ (see Fig. 3).

4. Numerical experiments

In this section, we investigate the performances of the SDecGMCA algorithm on synthetic toy examples that allow performing Monte-Carlo trials, as well as realistic simulations with partial sky coverage. The code that is used is open source (see Appendix A).
For all these experiments, we make use of the HEALPix pixelisation on the sphere.
Comparison criteria are first based on the NMSE (see Eq. (29)), which measures the reconstruction quality of the sources. As well, we make use of the mixing matrix criterion [6] to assess the quality of the estimated mixing matrices:

$$\mathrm{C_A} = -10 \log_{10} \left(\text{mean} \left(\left| \mathbf{A}^+ \mathbf{A}^* - \mathbf{I} \right| \right) \right), \tag{34}$$

where the mean operator is an average over all entries, $\mathbf{A}^*$ is the ground truth mixing matrix and $\mathbf{A}^+$ the Moore–Penrose inverse of the estimated mixing matrix. The mixing matrix criterion is more appropriate to compare DBSS and BSS methods with different source regularization, since it only depends on $\mathbf{A}$.

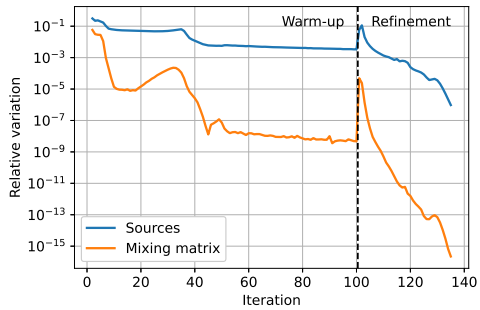


Figure 3: Empirical stabilization. The peak at iteration 101 corresponds to the switch from the warm-up stage to the refinement stage.

Parameter	Notation	Value
Minimum number of iterations at warm-up	N_{wu}	100
Number of detail scales	J	3
Thresholding parameter	k	3
Maximum source support	K_{max}	0.5

Table 2: Input parameters of SDecGMCA for the toy example

Let us define the oracle as the solution of the non-blind problem (*ie.* knowing the ground truth mixing matrix) with regularization strategy #4 and the optimal regularization hyperparameter. It provides an upper bound of the NMSE that SDecGMCA can reach.

4.1. Toy examples

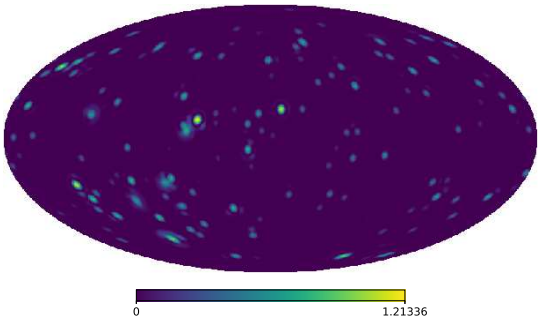
The same toy examples as in Section 2.3.2 are employed, with a similar parameterization. The input parameters of SDecGMCA are summarized in Table 2.

4.1.1. Impact of the Tikhonov regularization

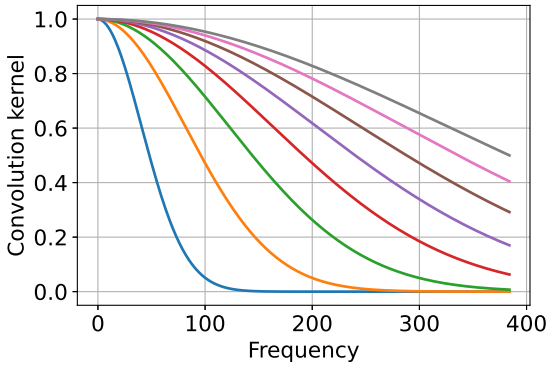
In standard BSS, the expected sources of error are the interferences (*i.e.* the leakage between the sources due to the mis-estimation of the mixing matrix), the noise contamination (*i.e.* the backprojected noise of the observations which is not filtered out) and the artifacts (*i.e.* the remaining errors) [26]. Within the scope of the joint deconvolution and BSS, this linear decomposition is not suitable anymore. For instance, when the mixing matrix is known, the estimation of the sources on its own can generate interferences in addition to the deconvolution artifacts due to the biasing Tikhonov regularization. Generally speaking, we can consider that a separation is successful when the reconstructions errors are dominated by the deconvolution artifacts. In this subsection, the data have the default observation parameters (see Section 2.3.2). Figure 4 shows an example of a solution given by SDecGMCA. The reconstructions errors are dominated by the deconvolution artifacts (see Fig. 4f).

We first assess the impact of the regularization hyperparameters on the performances of SDecGMCA. To this end, we execute SDecGMCA with different warm-up and refinement regularization hyperparameters. Since the sources are updated similarly in the non-blind problem, we consider the mean optimal regularization hyperparameters found in Section 2.3.2 as reference regularization hyperparameters.

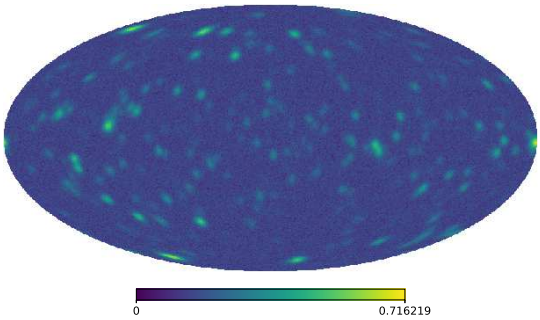
Each experiment is performed 100 times, with varying sources, mixing matrices and noise realizations. The mean NMSE and the mean $\mathrm{C_A}$ are reported in Table 3. The choice of c_{wu} for the warm-up stage has little impact on the performance metrics; indeed, the loss of NMSE and $\mathrm{C_A}$ due to a poor choice of c_{wu} is respectively at most 0.42 dB and 0.77 dB in the tested range. On the contrary, the selection of c_{ref} for the refinement stage may be more critical. It is however very interesting to highlight that, in a range of one order of magnitude around the optimal hyperparameter, the loss of NMSE and $\mathrm{C_A}$ remains quite



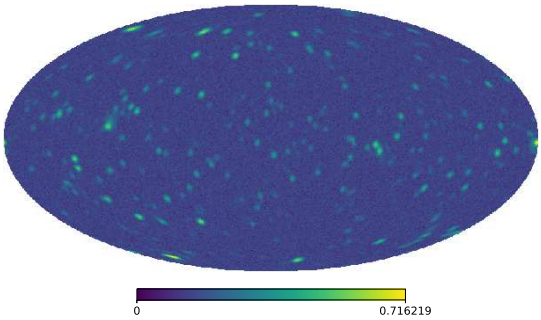
(a) Ground truth source $\mathbf{S}_1^*$



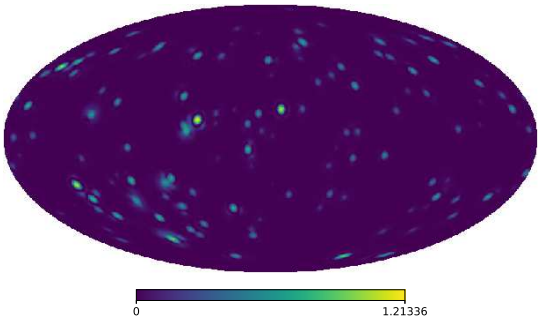
(b) Convolution kernels $\{\hat{\mathbf{H}}_\nu, 1 \leq \nu \leq 8\}$



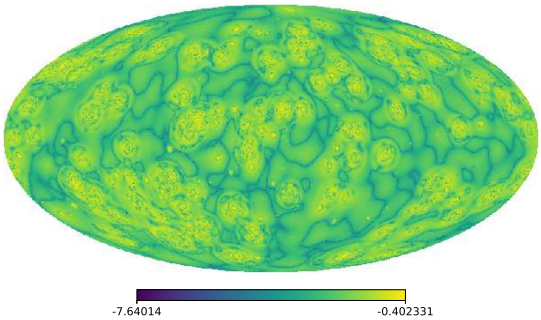
(c) Worse-resolved observation $\mathbf{X}_1$



(d) Best-resolved observation $\mathbf{X}_8$



(e) Estimated source $\mathbf{S}_1$



(f) Absolute error $|\mathbf{S}_1^* - \mathbf{S}_1|$

Figure 4: DBSS example with SDecGMCA (sources: arbitrary unit, logarithmic scale)

		c_{ref} ($\times c_{ref_{opt}}$)				
		10^{-1}	$10^{-0.5}$	$\mathbf{10^0}$	$10^{0.5}$	10^1
c_{wu} ($\times c_{wu_{opt}}$)	$10^0 \to 10^{-1}$	22.86	24.43	24.61	22.6	18.11
	$10^{0.5} \to 10^{-0.5}$	22.99	24.58	**24.79**	22.83	18.35
	$\mathbf{10^1 \to 10^0}$	23.00	24.58	24.79	22.82	18.34
	$10^{1.5} \to 10^{0.5}$	23.06	24.59	24.65	22.44	18.10
	$10^2 \to 10^1$	23.06	24.23	24.4	22.43	*17.93*
	$10^0 \to 10^{-1}$	**25.62**	25.20	24.86	24.41	22.66
	$10^{0.5} \to 10^{-0.5}$	25.60	25.19	24.86	24.46	22.81
	$\mathbf{10^1 \to 10^0}$	25.57	25.15	24.81	24.39	22.80
	$10^{1.5} \to 10^{0.5}$	25.28	24.80	24.37	23.75	22.07
	$10^2 \to 10^1$	25.28	24.56	24.25	23.84	*22.04*

Table 3: Mean NMSE (top) and $\mathrm{C_A}$ (bottom) in dB, over 100 realizations, performed by SDecGMCA as a function of c_{wu} and c_{ref}. These are given as multiples of $c_{wu\,opt}$ and $c_{ref_{opt}}$, which are the mean optimal hyperparameters for the non-blind problem. It is noted that the oracle mean NMSE is 25.74 dB.

limited (*i.e.* respectively about -2.21 dB and -1.05 dB at most). It is noted that the oracle mean NMSE is 25.74 dB; thus, the extra estimation of the mixing matrix is only the origin of a 0.95 dB loss in the NMSE.

4.1.2. Comparison with other blind source separation methods

In this paragraph, compararisons with other blind source separation methods are carried out. Since few DBSS methods have been investigated so far, a natural comparison would be with DecGMCA. However, since it has not directly been designed for data sampled on the sphere, a direct comparison cannot be performed. We rather propose to substitute DecGCMA's regularization strategy #2 within SDecGMCA to quantify the impact of the regularization strategy. In contrast to [9], where the regularization parameter is chosen in an *ad hoc* manner, we employe the optimal regularization hyperparameters; this method is therefore referred to as oDecGMCA (for optimized DecGMCA). Moreover, in order to highlight the benefit of combining the deconvolution to the BSS, we propose to compare SDecGMCA to three non-deconvolving BSS methods:

- GMCA (including the decreasing thresholding strategy and the ℓ_1-reweighting)
- Hierarchical Alternate Least-Squares (HALS) [2, 3]: non-negative matrix factorization algorithm solving $\underset{\mathbf{A}\geq 0,\, \mathbf{S}\geq 0}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{AS}\|_{\ell_2}^2$. It is based on block coordinate descent (one column of $\mathbf{A}$ and one row of $\mathbf{S}$ is updated at each iteration), with multiplicative updates.
- Beta Sparse Non-negative Matrix Factorization (β-SNMF) [27]: non-negative matrix factorization algorithm promoting the sparsity of $\mathbf{S}$ in the direct domain, which solves $\underset{\mathbf{A}\geq 0,\, \mathbf{S}\geq 0}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{AS}\|_{\ell_2}^2 + \lambda \|\mathbf{S}\|_{\ell_1}$. The minimization is also based on block coordinate descent with multiplicative updates.

As these methods can only process observations with the same resolution, the observations are convolved beforehand to the resolution of the worse-resolved observation so as to avoid noise amplification. The NMSE is adapted to take account of the resolution loss:

$$\mathrm{NMSE_w} = -10 \log_{10} \left(\frac{\|\mathbf{H}_{\nu_w} * \mathbf{S}^* - \mathbf{S}\|_{\ell_2}^2}{\|\mathbf{H}_{\nu_w} * \mathbf{S}^*\|_{\ell_2}^2} \right), \tag{35}$$

where ν_w is the worse-resolved observation channel. The $\mathrm{NMSE_w}$ of the DBSS methods can be calculated by deteriorating the estimated sources.

	C_A	$NMSE_w$	NMSE
SDecGMCA	**24.81**	**27.08**	**24.79**
oDecGMCA	23.01	20.94	15.03
GMCA	21.98	19.35	N/A
HALS	8.17	5.83	N/A
β-SNMF	8.43	7.01	N/A

Table 4: Mean performance metrics in dB, over 100 realizations, achieved by different algorithms

The performance metrics achieved by the different DBSS and BSS algorithms are reported in Table 4. Compared to oDecGMCA, SDecGMCA performs a significant gain in NMSE and a moderate increase in C_A. This result confirms that the choice of the regularization strategy is crucial for the estimation of the sources. Moreover, the BSS algorithms achieve poor results; indeed, high-frequency information, which is essential for the separation process, is lost when the data are deteriorated.

4.1.3. Varying observation parameters

Let us evaluate the sensitivity of the different source separation algorithms to the observation parameters, that is the mixing matrix condition number, the minimum resolution of the convolution kernels, the number of observations and the SNR. For SDecGMCA and oDecGMCA, the mean optimal regularization hyperparameter found above with the non-blind problem are used. At each point, the algorithms are executed 30 times with varying sources, mixing matrices and noise realization. The mean performance metrics are reported in Figure 5.

The NMSE achieved by SDecGMCA is close to the oracle; the loss is typically of 1 dB. In every scenario, SDecGMCA clearly outperforms oDecGMCA in terms of NMSE and C_A. Overall, the tendencies are consistent; the performance metrics increase with increasing minimum resolution, number of observations and SNR, while decrease with increasing mixing matrix condition number. The first notable exception is the $NMSE_w$ against the minimum resolution. It is due to the fact that the reference $\mathbf{H}_{\nu_w} * \mathbf{S}$ in the definition of the $NMSE_w$ varies from one point to the other. The second exception concerns the SNR; the performance metrics stabilize or decrease when there is little noise. This is an effect of the implicit regularization provided by the noise. According to the proposed threshold tuning strategy, when the noise level is low, the thresholds are low and the sparsity constraint is loosened. On the contrary, a higher noise level yields higher thresholds that tend to select high amplitude coefficients, which better discriminate between the sources.

Finally, let us assess the sensitivity of SDecGMCA to the regularization hyperparameter at refinement as a function of the observation parameters (see Figure 6). Both NMSE and C_A losses are limited in a range of one order of magnitude around the optimal regularization hyperparameter (typically -2 dB). The noticeable exception is when $\mathbf{A}$ is ill-conditioned. The higher sensibility to the regularization hyperparameter may come from the induced ill-condition of the $(\mathbf{M}[l])_{l\in\mathbb{N}}$ (matrices which are inverted in Eq. (21)).

4.2. Application to realistic astrophysical data

Joint deconvolution and blind source separation is now performed on realistic X-ray astrophysical simulations, whose spectra are representation of the emissions provided by the Chandra space telescope[4] in range of energy $2-6$ keV [28]. These data are composed of 25 observations, which are built as mixtures of three sources. The latter are associated with a synchrotron, a thermal and an iron line emissions as displayed in Fig. 7, whose spectra are displayed in the panel (d). It is commonplace in such an application that only a partial sky coverage is observed, which is simulated by projecting the sources on a limited portion of the sphere (see Fig. 7). Similarly to the synthetic data, we set the pixelization parameters to $N_{side} = 128$ and $l_{max} = 384$. The 25 observations have resolutions evenly spread between $l_{max}/8$ and l_{max}.

[4] `http://chandra.harvard.edu/`

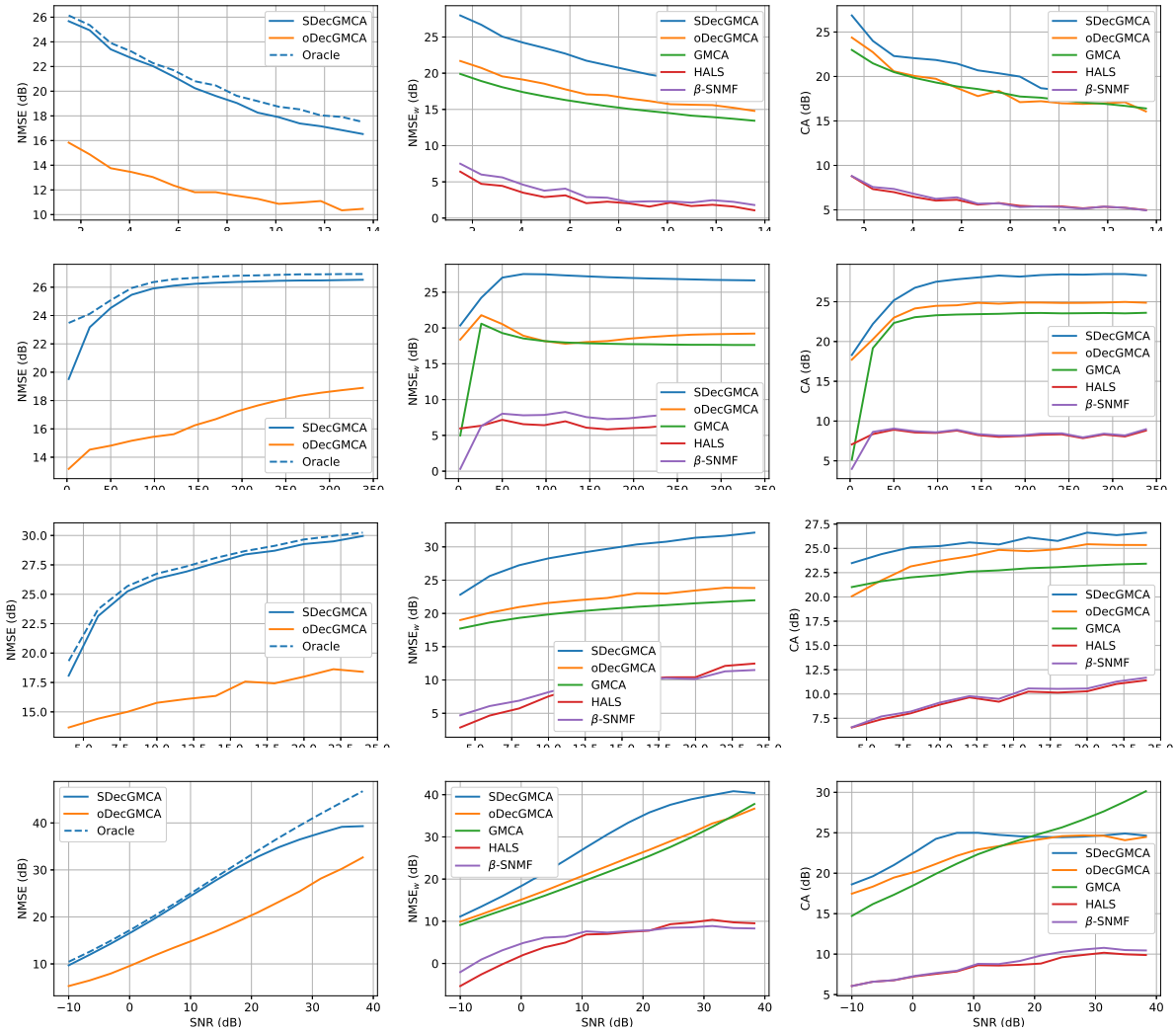


Figure 5: Mean performance metrics over more than 30 realizations as a function of observation parameters. The default values of the parameters are summarized in Section 2.3.2.

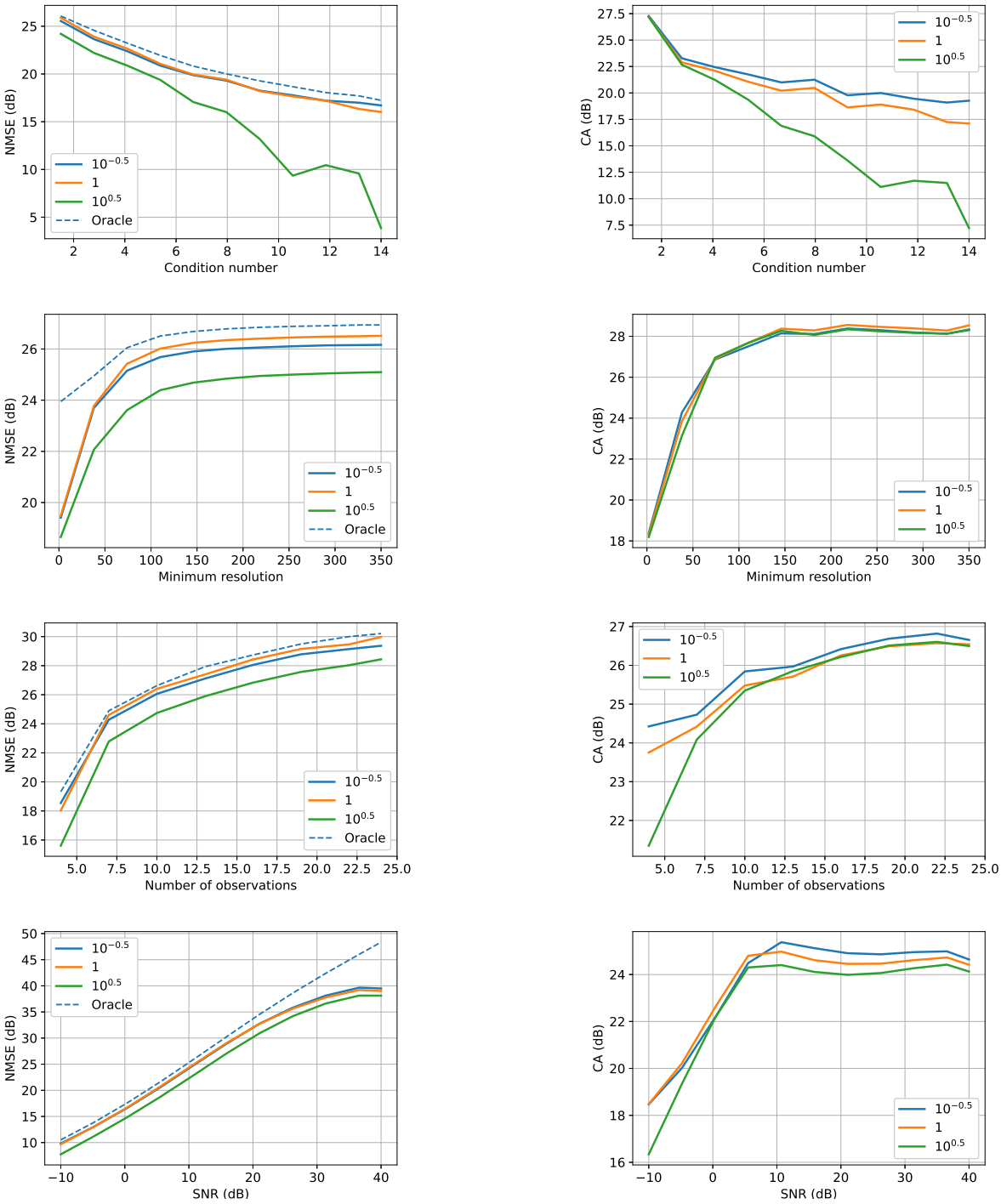
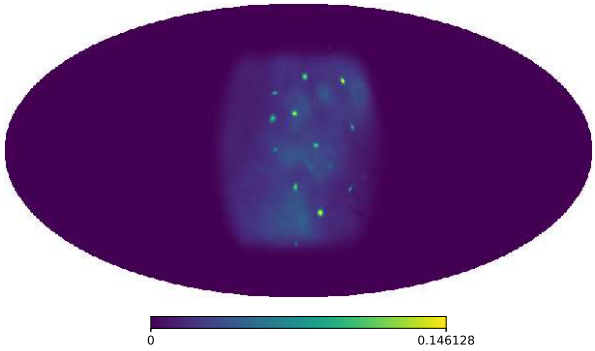
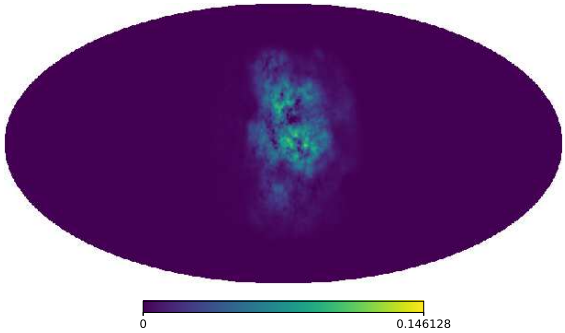


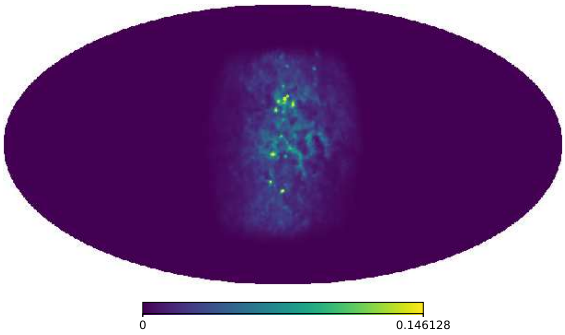
Figure 6: Mean performance metrics over more 30 realizations performed by SDecGMCA with the optimal regularization hyperparameter at warm-up $c_{wu_{opt}}$ and with different regularization hyperparameters at refinement c_{ref}, which are indicated as multiples of $c_{ref_{opt}}$ in the legends.



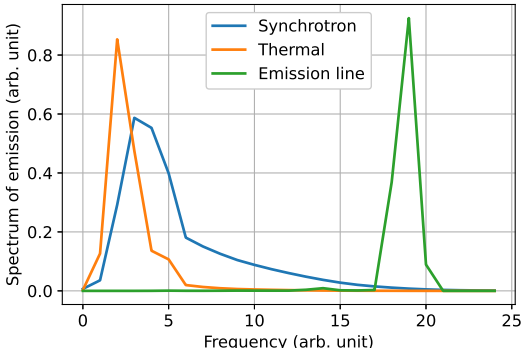
(a) Synchrotron source $\mathbf{S}_1^*$



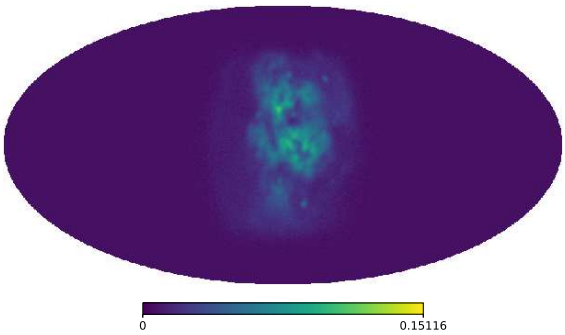
(b) Thermal source $\mathbf{S}_2^*$



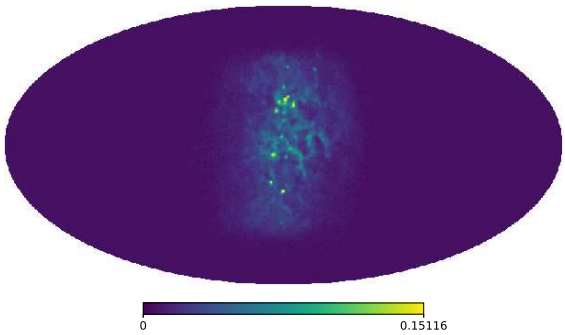
(c) Emission line source $\mathbf{S}_3^*$



(d) Mixing matrix $\mathbf{A}^*$



(e) Low resolution observation $\mathbf{X}_3$, with synchrotron and thermal emissions



(f) High resolution observation $\mathbf{X}_{20}$, with mostly emission line source

Figure 7: Realistic data and example of observations with a SNR of 10 dB.

	C_A	$NMSE_w$	NMSE
SDecGMCA	**18.02**	**26.92**	**21.17**
oDecGMCA	15.78	25.89	18.16
GMCA	16.60	25.84	N/A
HALS	7.78	10.29	N/A
β-SNMF	7.93	10.38	N/A

Table 5: Mean performance metrics in dB, over 50 realizations, achieved by different algorithms. The oracle mean NMSE is 22.58 dB.

We firstly consider that the observations are corrupted by a noise of 10 dB (see two observations on Fig. 7e and 7f). Figure 8 shows a solution given by SDecGMCA (with $K_{max} = 0.2$ to overcome the correlations between the sources). The errors are dominated by the deconvolution artifacts. Actually, the estimated sources are very close to the oracle estimation with the ground truth mixing matrix $\mathbf{A}^*$ (NMSE = 21.17 dB *vs.* oracle NMSE = 22.58 dB). Moreover, putting aside a small leakage in the lower frequencies of the emission line source in the synchrotron emission, the spectra of the three sources are well reconstructed ($C_A = 18.02$ dB).

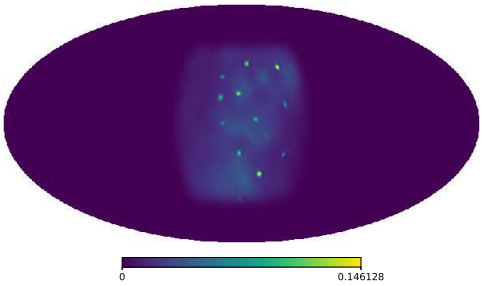
Let us compare the results returned by the previously considered DBSS and BSS algorithms. The mean performance metrics over 50 noise realizations are reported in Table 5. Equivalently to the synthetic data, SDecGMCA markedly overcomes oDecGMCA in terms of estimation error on both $\mathbf{S}$ and $\mathbf{A}$.

The estimates of the thermal source by the five algorithms are reported in Fig. 9, along with the oracle estimation. As said earlier, the SDecGMCA and oracle estimates are particularly similar. oDecGMCA reconstructs the source with a slightly lower resolution. More importantly, it is contaminated by outlier pixels, which are likely due to the regularization favoring the higher frequencies combined with a too low threshold. All the finer details are lost in the GMCA estimate, because the observations are degraded beforehand. Both HALS and β-SNMF are unable to correctly denoise the sources; this highlights the advantage of the sparsity constraint in a transformed domain.

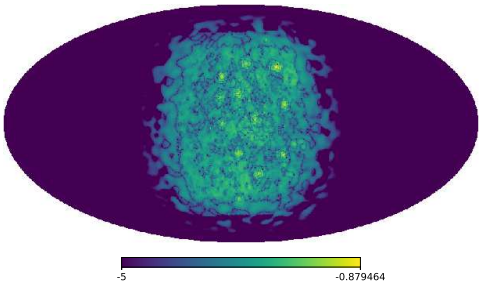
We then focus on the behavior of SDecGMCA with respect to the SNR; the SNR is in fact the observation parameter that has the greatest impact on the results and the robustness of the algorithm. In the first place, the optimal regularization parameters are estimated (by resolving the non-blind problem); the results are plotted in Fig. 10. As observed with the synthetic data, the optimal regularization hyperparamater of strategy #4 (refinement strategy) is relatively insensitive to the SNR and is worth approximately 0.5. The optimal regularization hyperparameter at warm-up (strategy #3) is more sensitive to the SNR. However, as with the synthetic data, it has limited influence on the result of the separation (thereafter, we set $c_{wu} = 1\text{e}{-3}$, independently of the SNR).

The performance metrics of SDecGMCA as a function of the SNR and for different regularization hyperparameters at refinement are plotted in Fig. 11. Similarly to the synthetic data, the NMSE and C_A stabilize when there is little noise. In overall, the choice of c_{ref} has little impact on the finale result (approximately -2 dB at most for both metrics). Overestimating c_{ref} improves the estimation of $\mathbf{A}$ to the detriment of the estimation of $\mathbf{S}$, and *vice versa*.

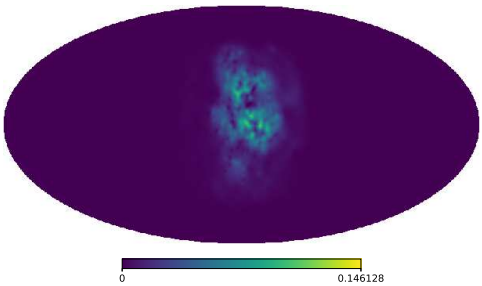
The different DBSS and BSS algorithms are compared for different noise levels; the results are reported in Fig. 12. Compared to the synthetic sources, the realistic sources are more correlated and less sparse in the starlet domain. This can result in a decrease in robustness. In order to compare the different methods, we calculated the mean performance metrics only over the successful realizations (that is with a C_A close enough to the maximum C_A along the realizations). SDecGMCA demonstrates a satisfactory robustness in terms of convergence to noise. As with the synthetic data, the performance metrics tend to stabilize when there is little noise. These experiments highlight that oDecGMCA is not robust at a low noise level; the few cases which converge return good metrics, hence the apparent better mean C_A and NMSE at high SNR than SDecGMCA. The performances of GMCA notably decrease at a low noise level. Indeed, the degree of sparsity of the sources is markedly decreased at low resolution; when combined with a low noise level (and



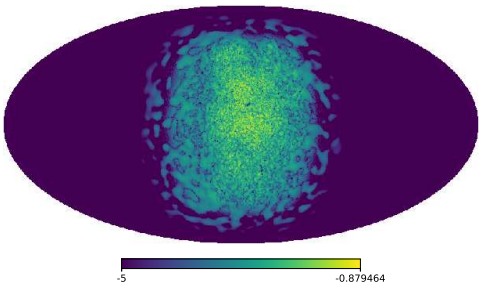
(a) Estimated source $\mathbf{S}_1$



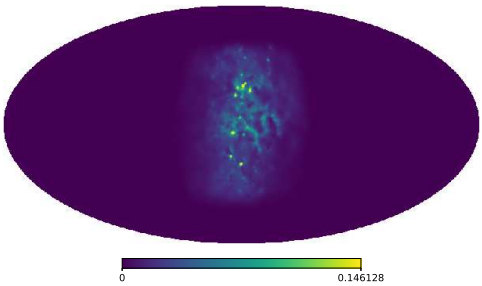
(b) Absolute error $|\mathbf{S}_1^* - \mathbf{S}_1|$

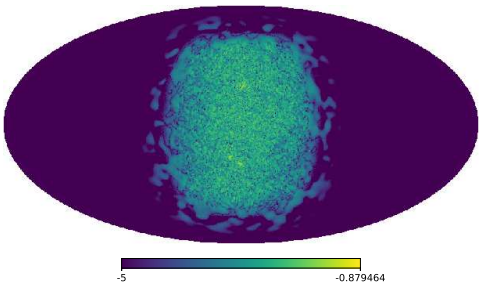


(c) Estimated source $\mathbf{S}_2$

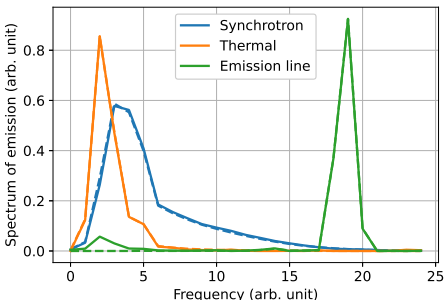


(d) Absolute error $|\mathbf{S}_2^* - \mathbf{S}_2|$



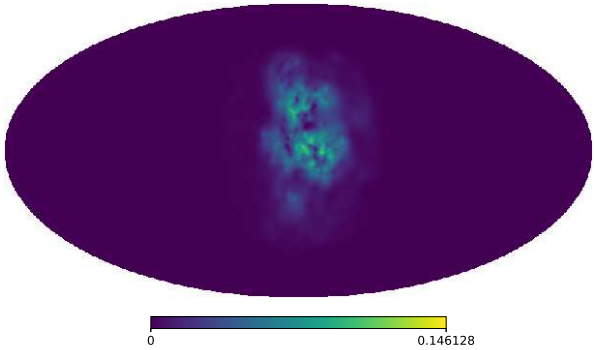


(f) Absolute error $|\mathbf{S}_3^* - \mathbf{S}_3|$

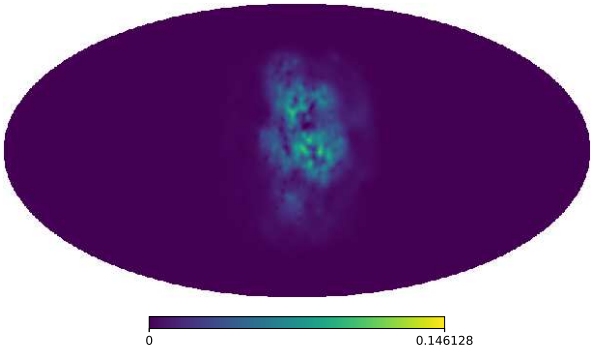


(g) Estimated mixing matrix $\mathbf{A}$ (dashed: ground truth $\mathbf{A}^*$)

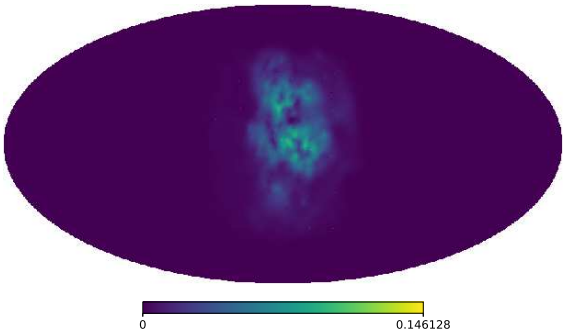
Figure 8: DBSS example with SDecGMCA (sources: arbitrary unit, logarithmic scale)



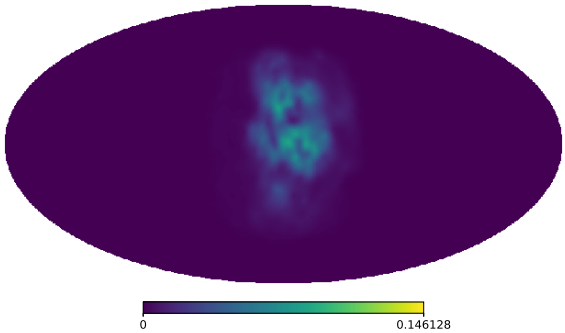
(a) Oracle



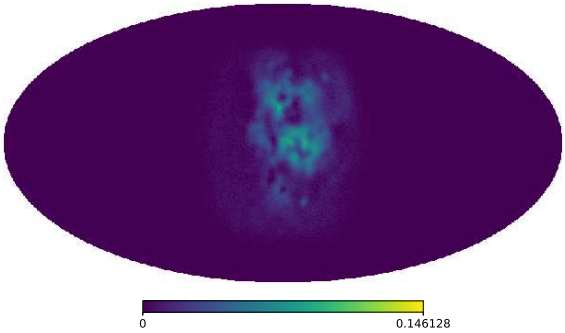
(b) SDecGMCA



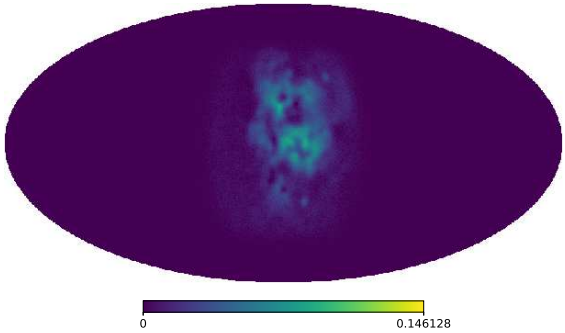
(c) oDecGMCA



(d) GMCA



(e) HALS



(f) β-SNMF

Figure 9: Estimated source $\mathbf{S}_2$ by different algorithms (arbitrary unit, logarithmic scale)

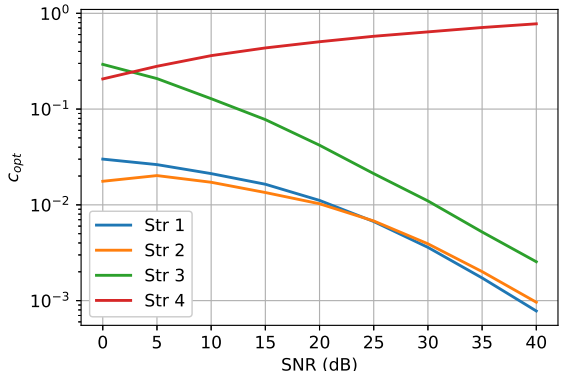


Figure 10: Mean optimal regularization hyperparameter of the non-blind problem, over 20 noise realizations, as a function of the SNR and the regularization strategy

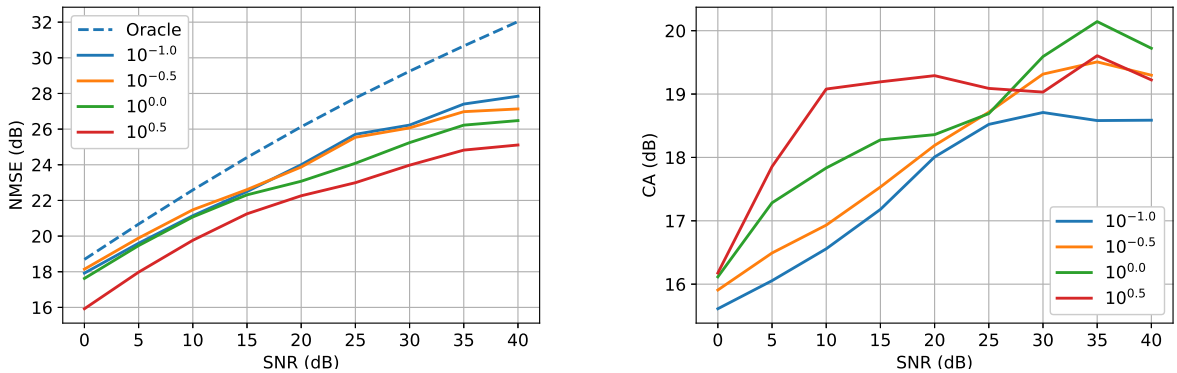


Figure 11: Mean performance metrics over 20 noise realizations performed by SDecGMCA as a function of the SNR, with $c_{wu} = 1\text{e}{-3}$ at warm-up and with different regularization hyperparameters c_{ref} at refinement, which are indicated as multiples of $c_{ref_{opt}}$ in the legends.

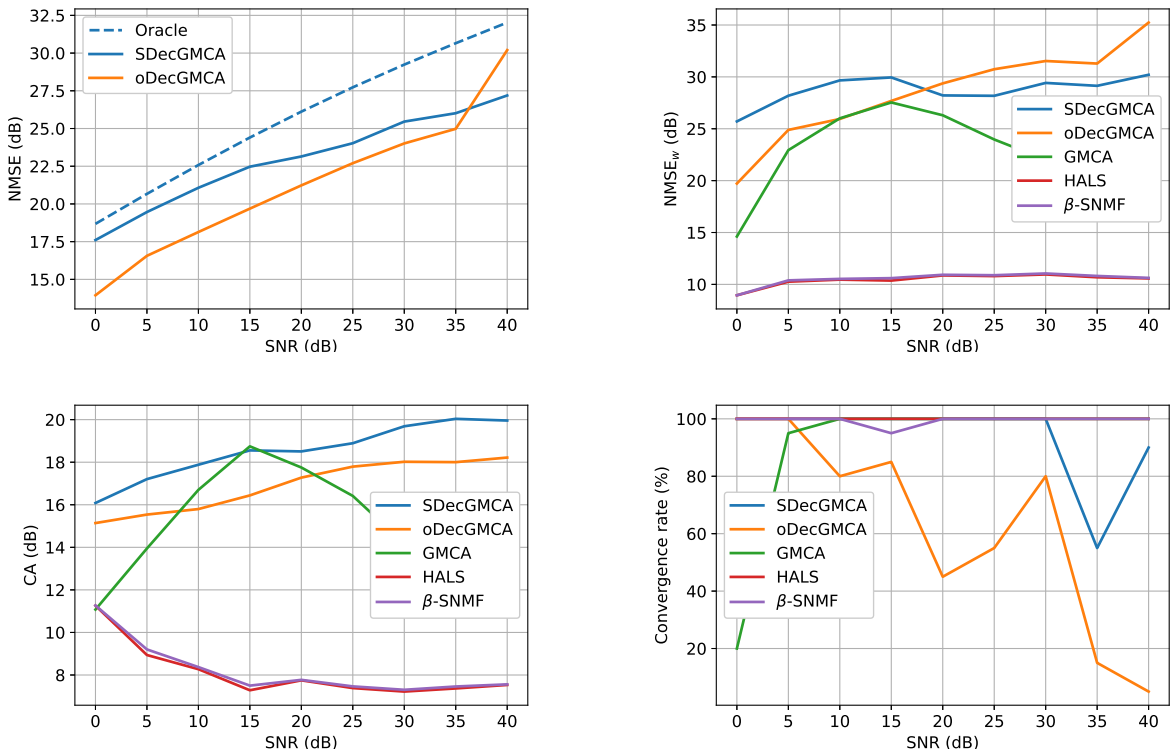


Figure 12: Mean performance metrics and convergence rate over 20 noise realizations performed by different DBSS and BSS algorithms as a function of the SNR

thus a smaller threshold regularization), the separation process is deteriorated.

5. Conclusion

In this article, we investigate a new joint deconvolution and sparse blind source separation algorithm to analyze multichannel spherical data. We first thoroughly investigate the impact of the regularization scheme in the least-squares minimization, which is proved to largely impact the quality of the separation process. We further introduce dedicated regularization schemes that much better adapt to the statistics of the sources to be estimated. Based on these regularization techniques, we introduce a two-step minimization algorithm coined SDecGMCA, which is showed to provide a robust and effective minimization procedure. Numerical experiments on both toy examples and realistic astronomical simulations are achieved to evaluate how the proposed algorithm performs in a wide range of mixing scenarios. Comparisons with standard BSS methods are further carried out, which show that the proposed SDecGMCA algorithm is competitive.

Acknowledgement

This work is supported by the European Community through the grant LENA (ERC StG - contract no. 678282).

References

[1] P. Comon, C. Jutten, Handbook of blind source separation, Academic Press, 2010.

[2] A. Cichocki, R. Zdunek, S.-i. Amari, Hierarchical ALS Algorithms for Nonnegative Matrix and 3D Tensor Factorization, in: Proceedings of ICA, Springer, 2007, pp. 169–176.
URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.106.5641

[3] N. Gillis, F. Glineur, Accelerated Multiplicative Updates and Hierarchical ALS Algorithms for Nonnegative Matrix Factorization, Neural Computation 24 (4) (2012) 1085–1105.

[4] M. Zibulevski, Blind source separation with relative newton method, Proceedings of ICA, Independent Component Analysis, 2003897-902 (2003).

[5] Y. Li, A. Cichocki, S. ichi Amari, S. L. Shishkin, J. Cao, F. Gu, Sparse Representation and Its Applications in Blind Source Separation, in: Proceedings of NIPS, Vancouver, BC, Canada, 2003, pp. 241–248.

[6] J. Bobin, J. Rapin, J.-L. Starck, A. Larue, Sparsity and adaptivity for the blind separation of partially correlated sources, IEEE Transactions on Signal Processing 63 (5) (2015) 1199–1213.

[7] C. Kervazo, J. Bobin, C. Chenot, F. Sureau, Use of palm for ℓ_1 sparse matrix factorization: Difficulty and rationalization of an heuristic approach., Digital Signal Processing 97 (February 2020).

[8] E. Chapman, S. Zaroubi, F. Abdalla, Foreground removal vs. foreground avoidance: Contamination of the eor window, ArXiv:1501.04429 (2015).

[9] M. Jiang, J. Bobin, Starck, J.-L., Joint multichannel deconvolution and blind source separation, SIAM Journal on Imaging Sciences 10 (4) (2017) 1997–2021. doi:10.1137/16m1103713.
URL http://dx.doi.org/10.1137/16M1103713

[10] M. Golbabaee, S. Arberet, P. Vandergheynst, Compressive source separation: Theory and methods for hyperspectral imaging, IEEE Transactions on Image Processing 22 (12) (December 2013).

[11] J. Kobarg, J. Maass, P. qnd Oetje, O. Tropp, E. Hirsch, C. Sagiv, M. Golbabaee, P. Vandergheynst, Numerical experiments with maldi imaging data, Computational Mathematics 40 (3) (2014).

[12] M. Kleinsteuber, H. Shen, Blind source separation with compressively sensed linear mixtures, IEEE Signal Processing Letters 19 (2) (2012).

[13] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, M. Bartelmann, HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere, Astrophysical Journal 622, 759–771 (Apr. 2005). arXiv:arXiv:astro-ph/0409513, doi:10.1086/427976.

[14] J. R. Driscoll, D. M. Healy, Computing fourier transforms and convolutions on the 2-sphere, Adv. Appl. Math. 15 (2) (1994) 202–250. doi:10.1006/aama.1994.1008.
URL https://doi.org/10.1006/aama.1994.1008

[15] M. Elad, Why simple shrinkage is still relevant for redundant representations?, IEEE Transactions on Information Theory 52 (12), 5559–5569 (2006).

[16] Y. Xu, W. Yin, A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Applications to Nonnegative Tensor Factorization and Completion, SIAM Journal on Imaging Sciences 6 (3) (2013) 1758–1789.

[17] N. Parikh, S. Boyd, Proximal algorithms, Foundations and Trends in Optimization 1 (3) (2014).

[18] P. Paatero, U. Tapper, Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values, Environmetrics 5 (2) (1994) 111–126.

[19] R. Neelamani, H. Choi, R. G. Baraniuk, ForWard: Fourier-wavelet regularized deconvolution for ill-conditioned systems, IEEE Transactions on Signal Processing 52 (2), 418–433 (2004).

[20] M. Bertero, P. Boccacci, Introduction to Inverse Problems in Imaging, Institute of Physics, 1998.

[21] J.-L. Starck, Y. Moudden, A. P., M. Nguyen, Wavelets, ridgelets and curvelets on the sphere, Astronomy and Astrophysics 446 (3) (2006) 1191–1204. doi:10.1051/0004-6361:20053246.
URL http://dx.doi.org/10.1051/0004-6361:20053246

[22] J. Bobin, J.-L. Starck, J. Fadili, Y. Moudden, Sparsity and morphological diversity in blind source separation, IEEE Transactions on Image Processing 16 (11) (2007) 2662–2674.
URL http://www.ncbi.nlm.nih.gov/pubmed/17990743

[23] E. J. Candes, M. B. Wakin, S. P. Boyd, Enhancing sparsity by reweighted L1 minimization, Journal of Fourier Analysis and Applications 14 (5), 877–905 (2008).

[24] P. Tseng, Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization, Optimization 109 (3) (2001) 475–494.
URL http://www.springerlink.com/index/q327675221126243.pdf

[25] J. Bolte, S. Sabach, M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, Mathematical Programming (2013) 1–36doi:10.1007/s10107-013-0701-9.

[26] E. Vincent, R. Gribonval, C. Fevotte, Performance measurement in blind audio source separation, IEEE Transactions on Audio, Speech & Language Processing 14 (4) (2006) 1462–1469.

[27] A. Cherni, S. Anthoine, C. Chaux, β-NMF and sparsity promoting regularizations for complex mixture unmixing, application to 2D HSQC in: 45th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2020), Barcelone, Spain, 2020.
URL https://hal.archives-ouvertes.fr/hal-02457468

[28] A. Picquenot, F. Acero, J. Bobin, P. Maggi, J. Ballet, G. Pratt, Novel method for component separation of extended sources in x-ray astronomy, A&A A139 (627) (2019).

[29] J.-L. Starck, F. Murtagh, J. Fadili, Sparse Image and Signal Processing - Wavelets, Curvelets, Morphological Diversity, Cambridge University Press, 2010.

Appendix A. Open source code

The code is open source and can be found online at `https://github.com/RCarloniGertosio/SDecGMCA` on version 3 of the LGPL.

Appendix B. Implementation of SDecGMCA with a multiresolution representation

Multiresolution representations (*e.g.* wavelets, curvelets, to only name two) are versatile and yet effective analytic signal representations that are well adapted to provide sparse representations for a wide range of natural data. Consequently, we choose to use the spherical starlet representation [21] for the numerical experiments of the present paper. In this appendix, we describe in detail the implementation of SDecGMCA with a multiresolution representation.

Dealing with the coarse resolution. As detailed in [29], such signal representations decompose the data into detail scales and a coarse scale. The former bears invaluable information to disentangle the sources while the latter is a mere low frequency approximation of the signals that is useless for the separation process. Therefore, the coarse scales of the observations are removed during the separation procedure (that is the warm-up and refinement stages described in Algorithm 1). The low frequency information included in the coarse scales of the raw data $\mathbf{X}$ is then reincorporated in the sources with the obtained mixing matrix. An iterative procedure is still necessary due to the reweighting and the regularization based on the spectra (strategy #4). Moreover, as described earlier, the maximum source support is set to $K = 1$ during this finale step in order to improve the estimation of the sources.

Non-negativity constraint on $\mathbf{S}$. The non-negativity of the sources is applied by projecting the sources, expressed in the direct space, on the positive orthant. To do so, it is necessary to estimate the sources in the pixel or sample domain. Therefore, at each step of the SDecGMCA algorithm, each source is fully reconstructed, which requires an extra projection in the harmonic space per iteration. In total, two spherical harmonics transforms are performed at each iteration: one for the estimation of the sources and one for the update of the mixing matrix.

Managing the regularization thresholds in a multiresolution representation. Selecting adequate thresholds plays a key role in the achievement of a separation. As pointed out earlier, the thresholds depend on the level of the noise corrupting the estimated sources. In the case where the level of the noise which contaminates the data is known, it is possible to determine analytically the noise level affecting the sources in the chosen multiresolution representation, thus ensuring a suitable thresholding.
We recall that the observations are contaminated by a white Gaussian noise $\mathbf{N}$ of variance σ^2. Quite similarly to the plane case, the harmonic transform of such a noise is Gaussian; its covariance matrix multipole by multipole is:

$$\mathrm{E}\left[\hat{\mathbf{N}}^{l,m}\hat{\mathbf{N}}^{l,m\,\dagger}\right] = \frac{4\pi}{N_p}\sigma^2\mathbf{I}. \tag{B.1}$$

The least-squares update of the sources colors and correlates the noise:

$$\mathrm{E}\left[\hat{\mathbf{P}}^{l,m}\hat{\mathbf{P}}^{l,m\,\top}\right] = \frac{4\pi}{N_p}\sigma^2\left(\mathbf{M}[l] + \underset{n\in[1,N_s]}{\mathrm{diag}}\left(\varepsilon_{n,l}\right)\right)^{-1}\mathbf{M}[l]\left(\mathbf{M}[l] + \underset{n\in[1,N_s]}{\mathrm{diag}}\left(\varepsilon_{n,l}\right)\right)^{-1} \tag{B.2}$$

where $\mathbf{P}$ is the noise projected in the source domain by Eq. (21) and $\mathbf{M}[l] = \mathbf{A}^\top \mathrm{diag}\left(\hat{\mathbf{H}}^l\right)^2\mathbf{A}$. The multiresolution analysis preceding the thresholding further colors the noise; more specifically, the term $\left|\mathbf{h}_{\mathbf{j}l}\right|^2$ is added to each detail scale j, where $\mathbf{h}_\mathbf{j}$ is the filter giving the jth detail scale of the multiresolution representation $\mathbf{\Phi}$.
The constraint term on $\mathbf{S}$ (Eq. 9) does not take into account the noise correlation in the source domain (principally for computational reasons, in order for the associated proximal operator to have an analytical

Parameter	Range	Regularization strategy			
		#1	#2	#3	#4
cond($\mathbf{A}$)	1.5 to 14	1.04	1.84	0.31	**0.10**
r_{min}	2 to 350	> 3.30	> 2.60	0.75	**0.11**
N_c	4 to 24	0.49	2.02	1.11	**0.25**
SNR (dB)	-10 to 40	1.18	1.08	**0.41**	0.76

Table C.6: Range of variation of the mean optimal regularization hyperparameter c_{opt}, in terms of order of magnitude, when the observation parameters vary. In some cases, c_{opt} is smaller than the lower bound of the research interval (1e−5), hence the lower bound for the range of variation.

form and be applied fast). Thus, the non-diagonal terms of the covariance matrix are neglected[5]. This leads to the following variance in the direct space for source n and scale j:

$$\mathrm{E}\left[\left[\mathbf{P}\mathbf{\Phi}^\top\right]_{n,j}^{p\quad 2}\right] = \frac{\sigma^2}{N_p}\sum_{l=0}^{l_{max}}(2l+1)\left[\left(\mathbf{M}[l] + \underset{n\in[1,N_s]}{\mathrm{diag}}\left(\varepsilon_{n,l}\right)\right)^{-1}\mathbf{M}[l]\left(\mathbf{M}[l] + \underset{n\in[1,N_s]}{\mathrm{diag}}\left(\varepsilon_{n,l}\right)\right)^{-1}\right]_n^n \left|\mathbf{h}_{\mathbf{j}_l}\right|^2 \tag{B.3}$$

Appendix C. Numerical comparisons of the regularization strategies: detailed results

In this appendix, we present the results of the numerical experiments performed to compare the regularization strategies. To do so, we recall that we resort to the non-blind separation case, *ie.* with the ground truth mixing matrix $\mathbf{A}^*$.

Figure C.13 shows the mean NMSE of 100 random trials as a function of the four aforementioned observation parameters: number of observations, mixing matrix condition number, minimum resolution and SNR. Unsurprisingly, strategy #4 clearly provides the best reconstruction qualities. Among the other strategies, that do not assume the sources to be known, strategy #3 achieves better results. It is mostly thanks to the non-linear maximum operator, which allows to keep the lower frequencies unbiased, where most of the sources energy is located (see example Fig. C.14, where $\varepsilon_{n,l} = 0$ for $l \leq 44$). Strategy #2 gives poor results; indeed, it biases more significantly the lower frequencies than the higher ones.

Table C.6 shows the range of variation of the mean optimal regularization hyperparameters when the observation parameters vary. Contrary to the 3 first strategies, the optimal regularization hyperparameter for strategy #4 is rather insensitive to the observation parameters (typically $c_{opt} \sim 0.5$). The noticeable exception is regarding the SNR. When the SNR is low, the denoising by the sparsity regularization is particularly promoted, hence a marked decrease of c_{opt}.

[5]In the numerical experiments, the non-diagonal terms are in fact small compared to the diagonal terms.

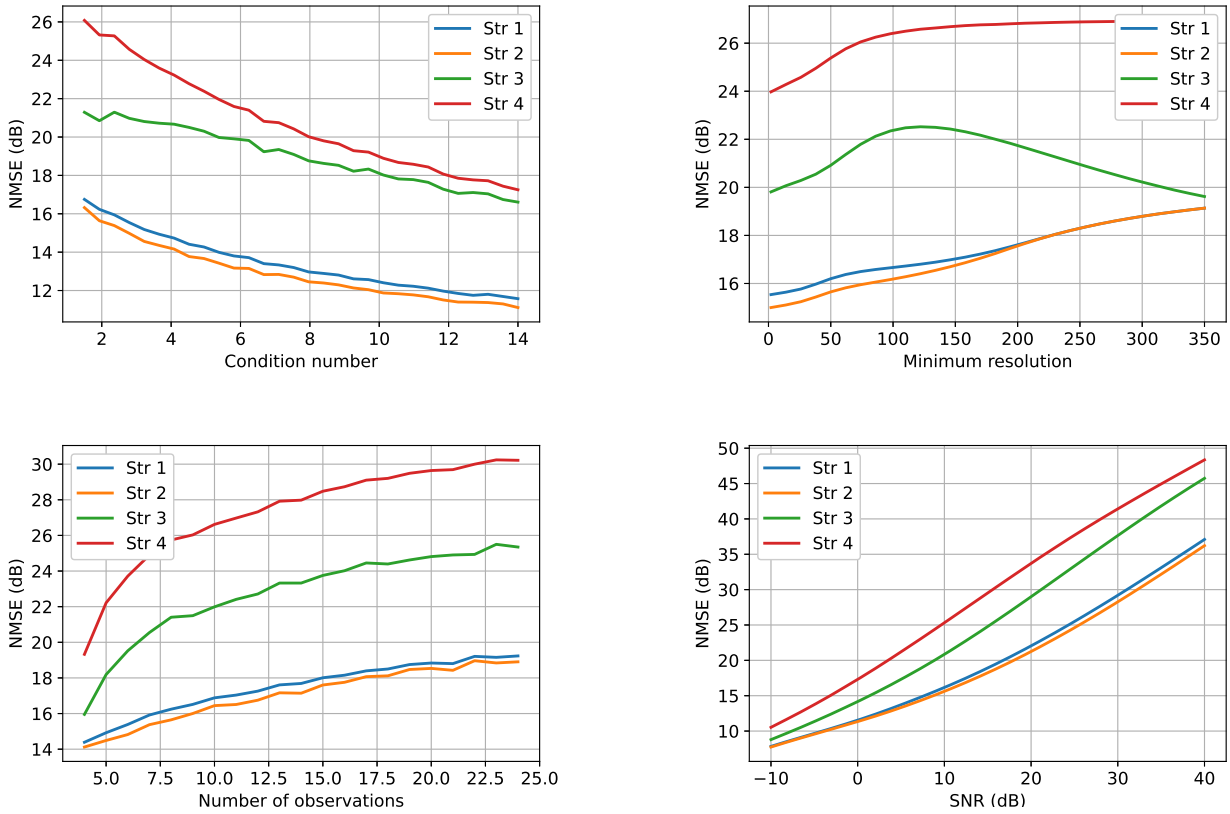


Figure C.13: Mean NMSE over 100 realizations as a function of observation parameters. The default values of the parameters are summarized in Section 2.3.2. The decrease in NMSE for str. 3 is due to the decrease of $\varepsilon_{n,0} - \varepsilon_{n,l_{max}}$, which reduces the ability of the Tikhonov regularization to discriminate the higher frequencies compared to the lower frequencies.

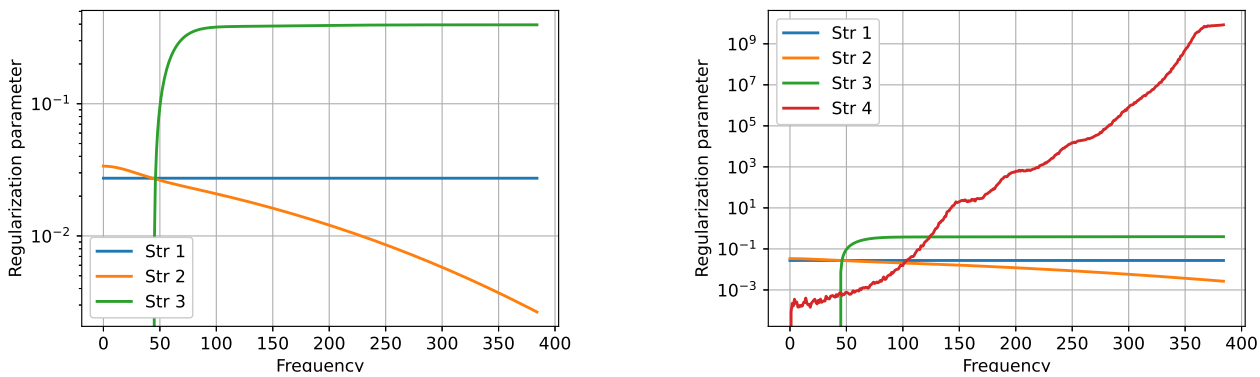
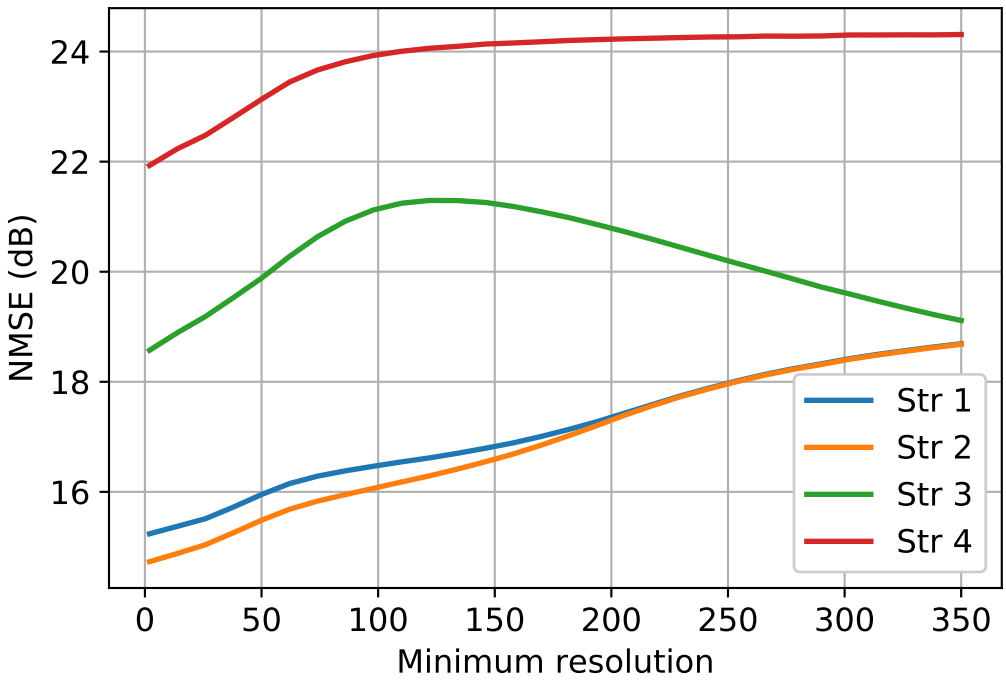
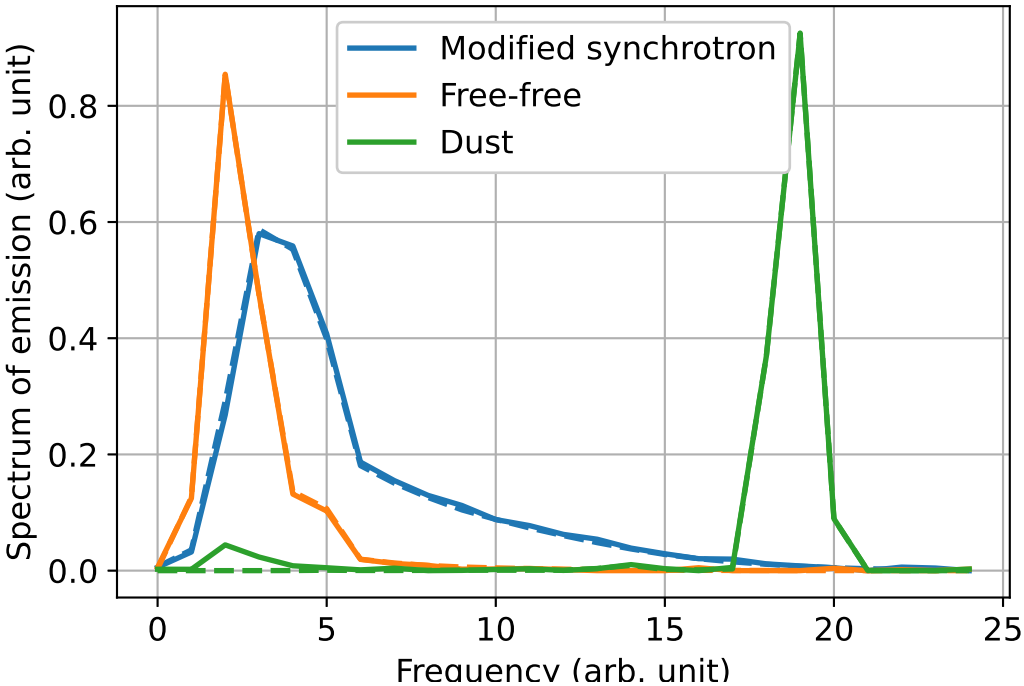
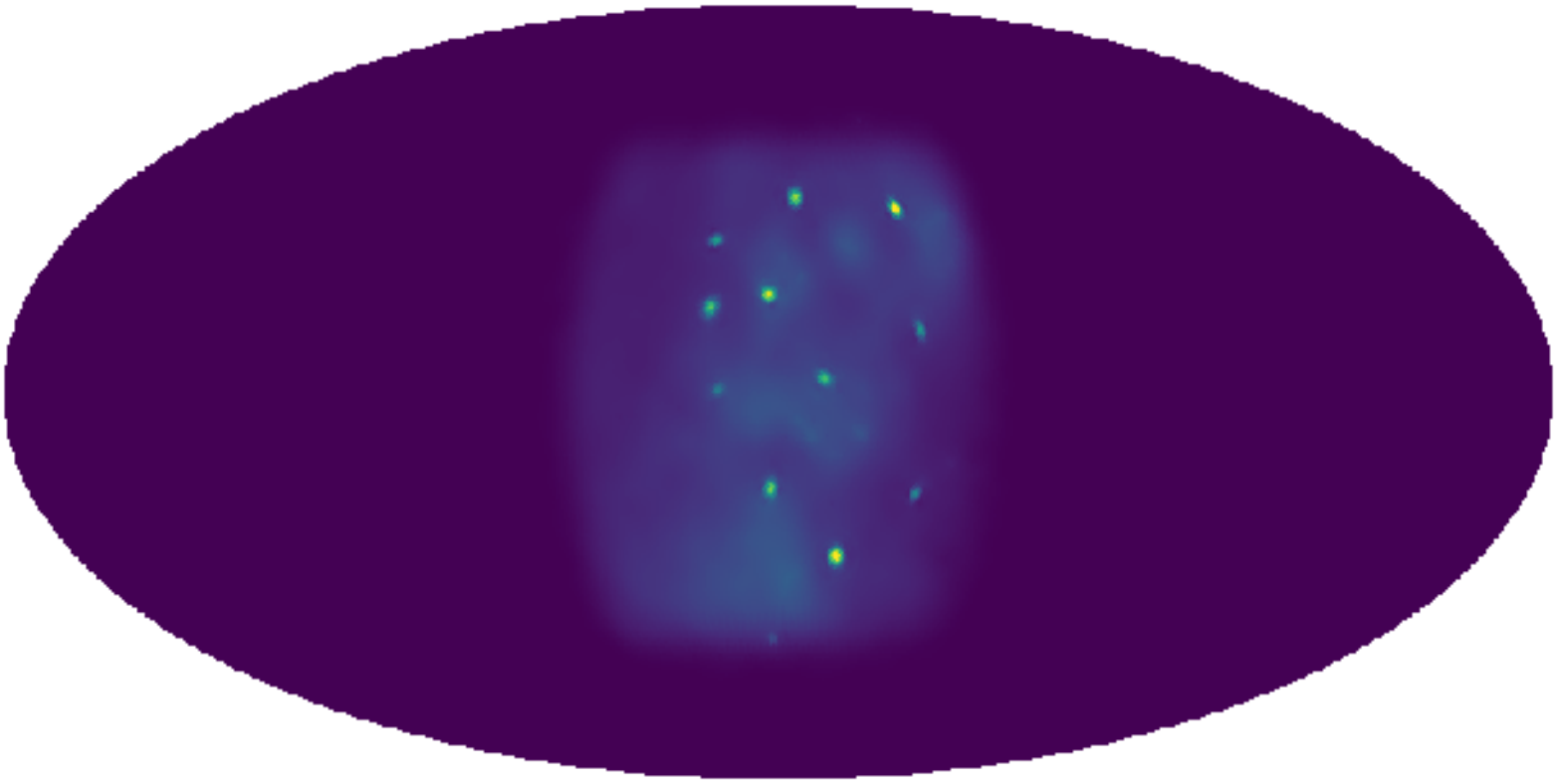
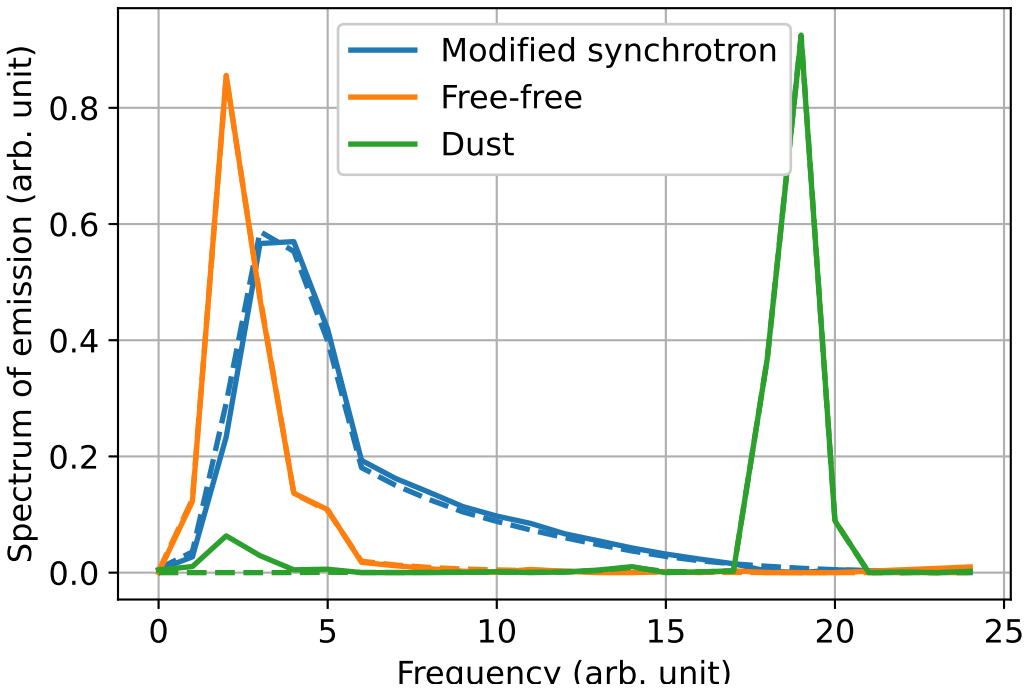


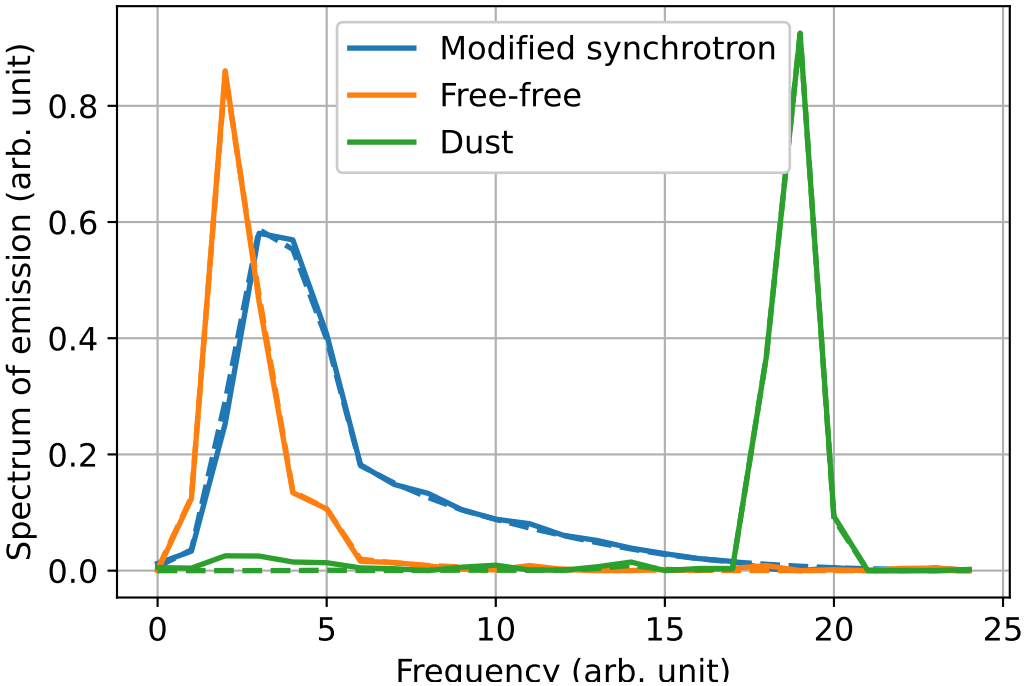
Figure C.14: Example of regularization parameters $\varepsilon_{n,l}$ as a function of the frequency l for the default observation parameters and with the mean optimal hyperparameters.

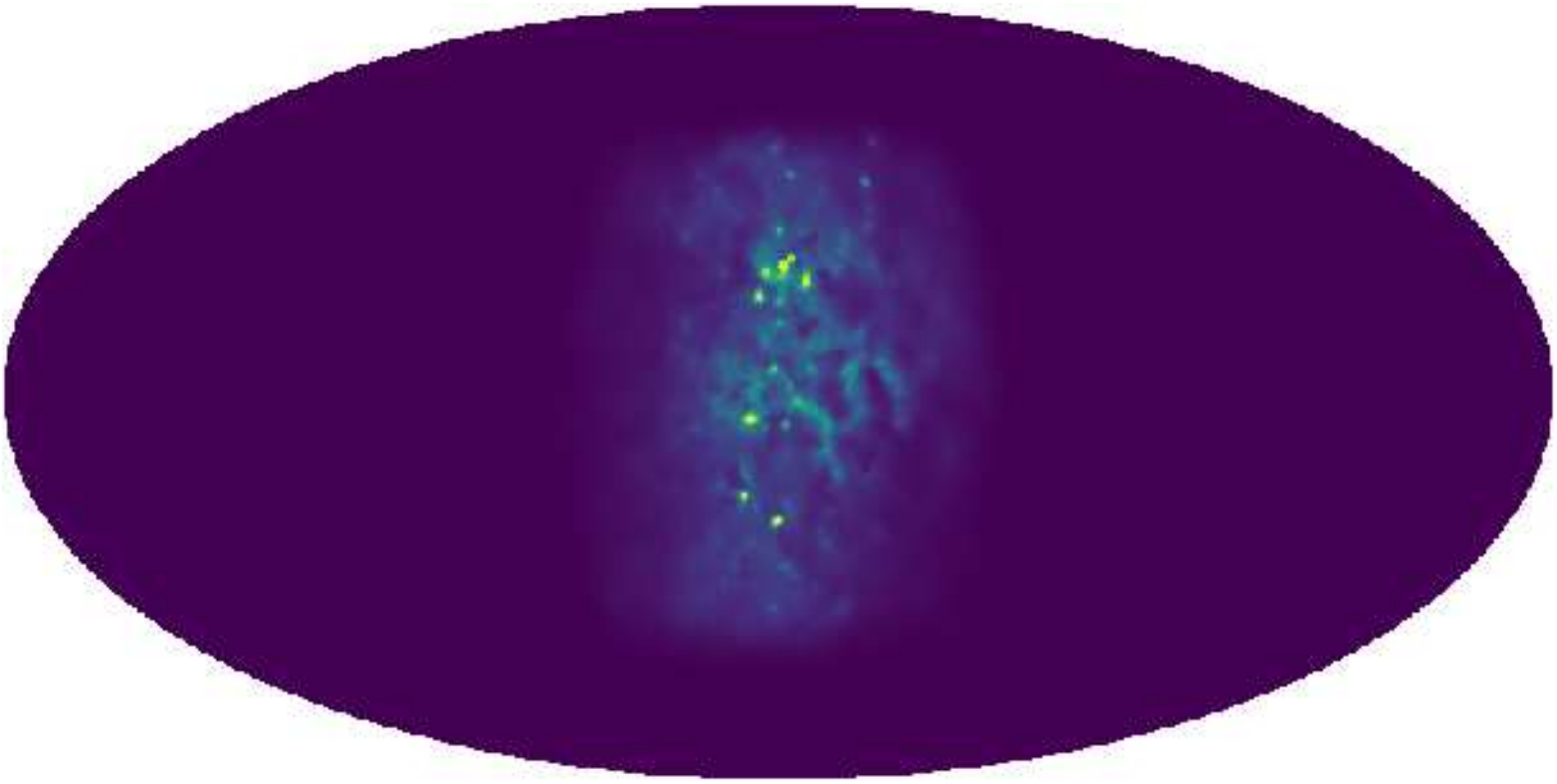

24
22
20
18
16
NMSE (dB)
0
50
100
150
200
250
300
350
Minimum resolution
Str 1
Str 2
Str 3
Str 4

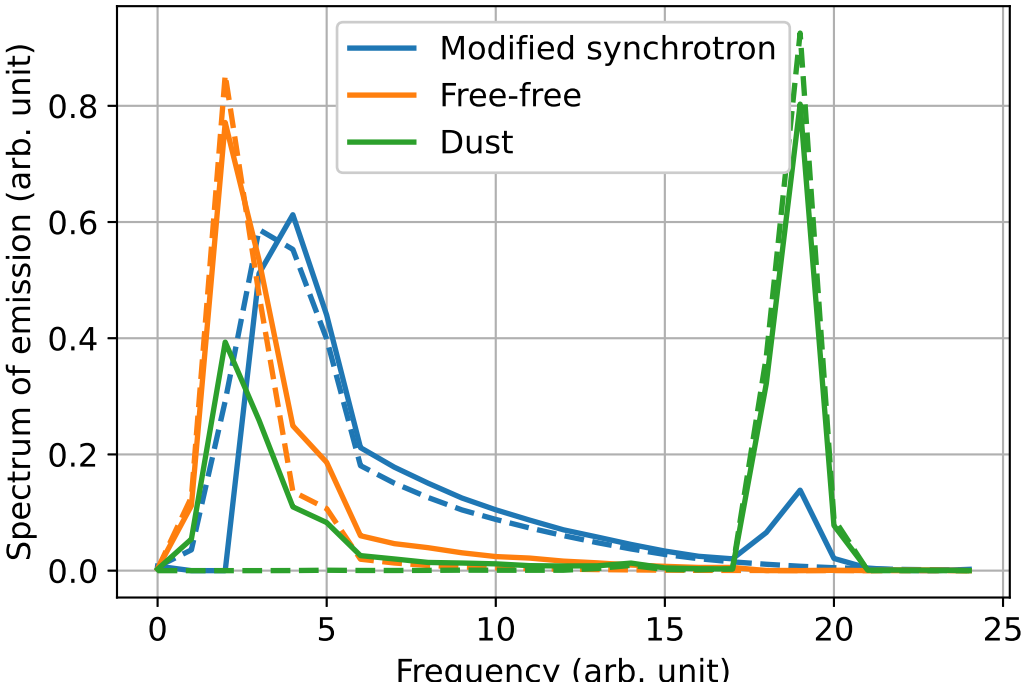

Modified synchrotron
Free-free
Dust
Spectrum of emission (arb. unit)
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
Frequency (arb. unit)

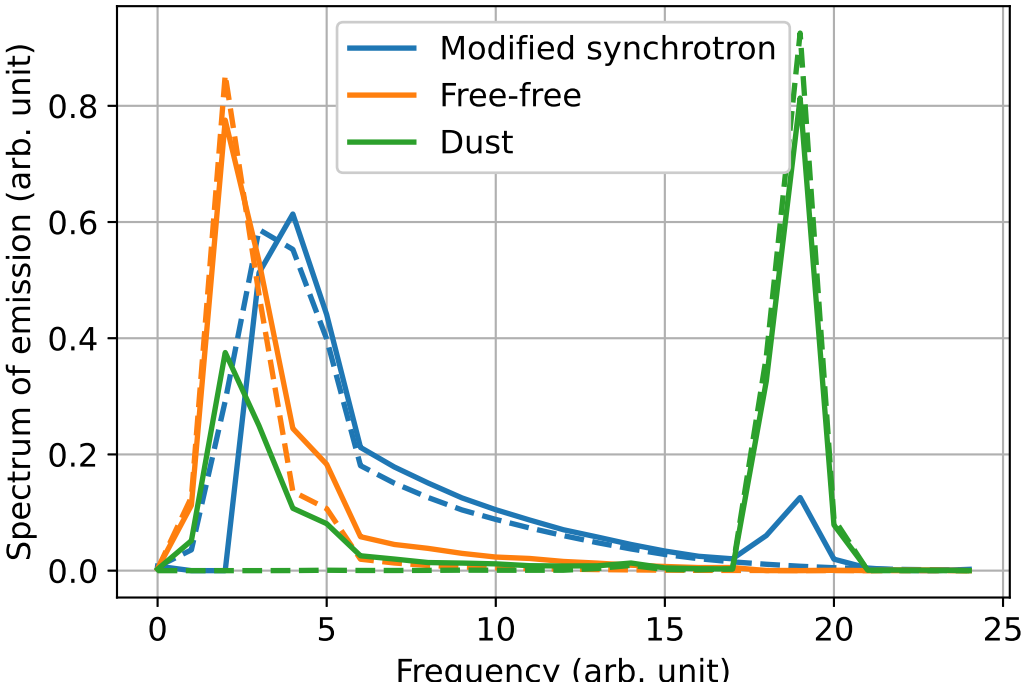
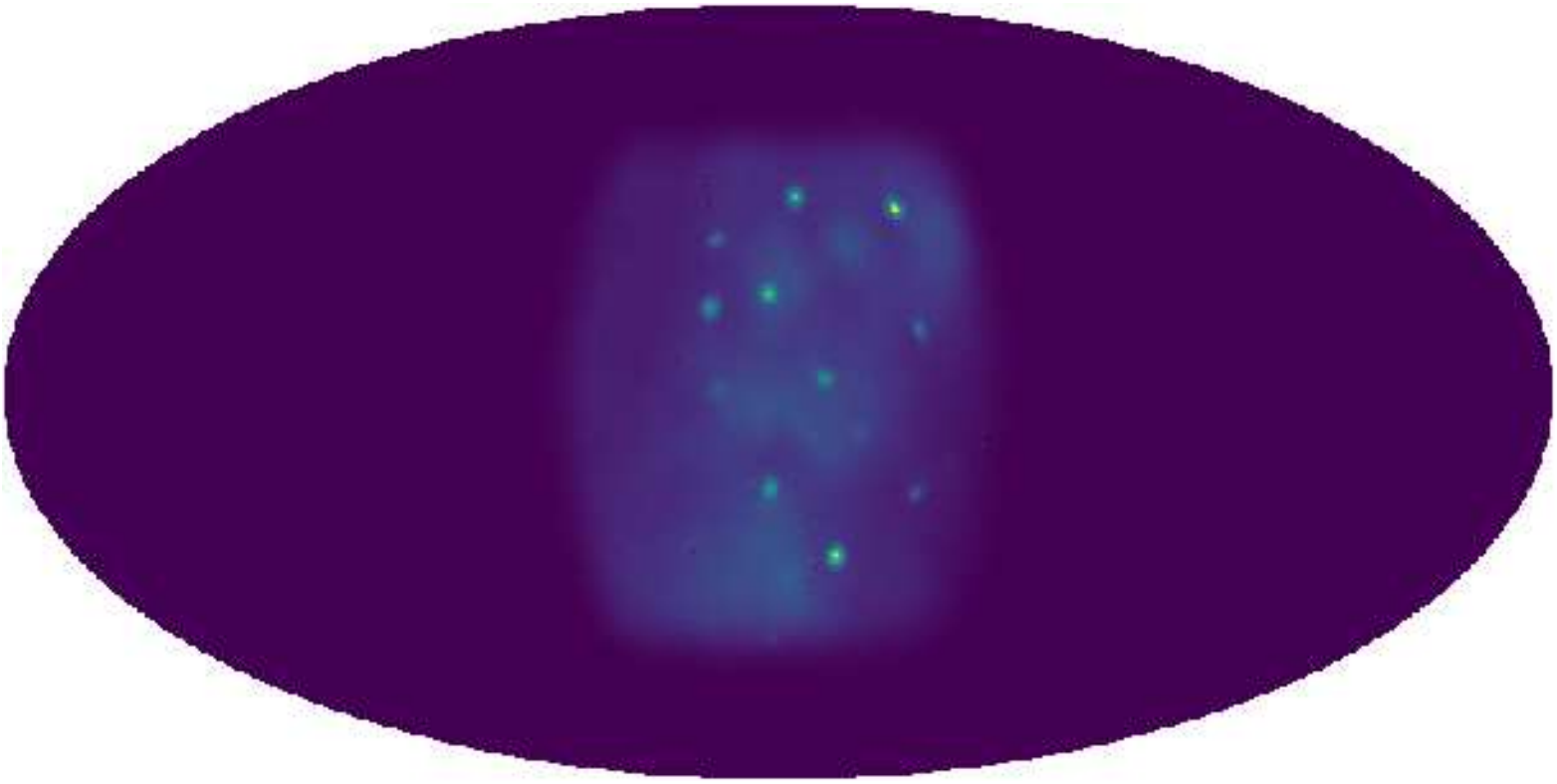
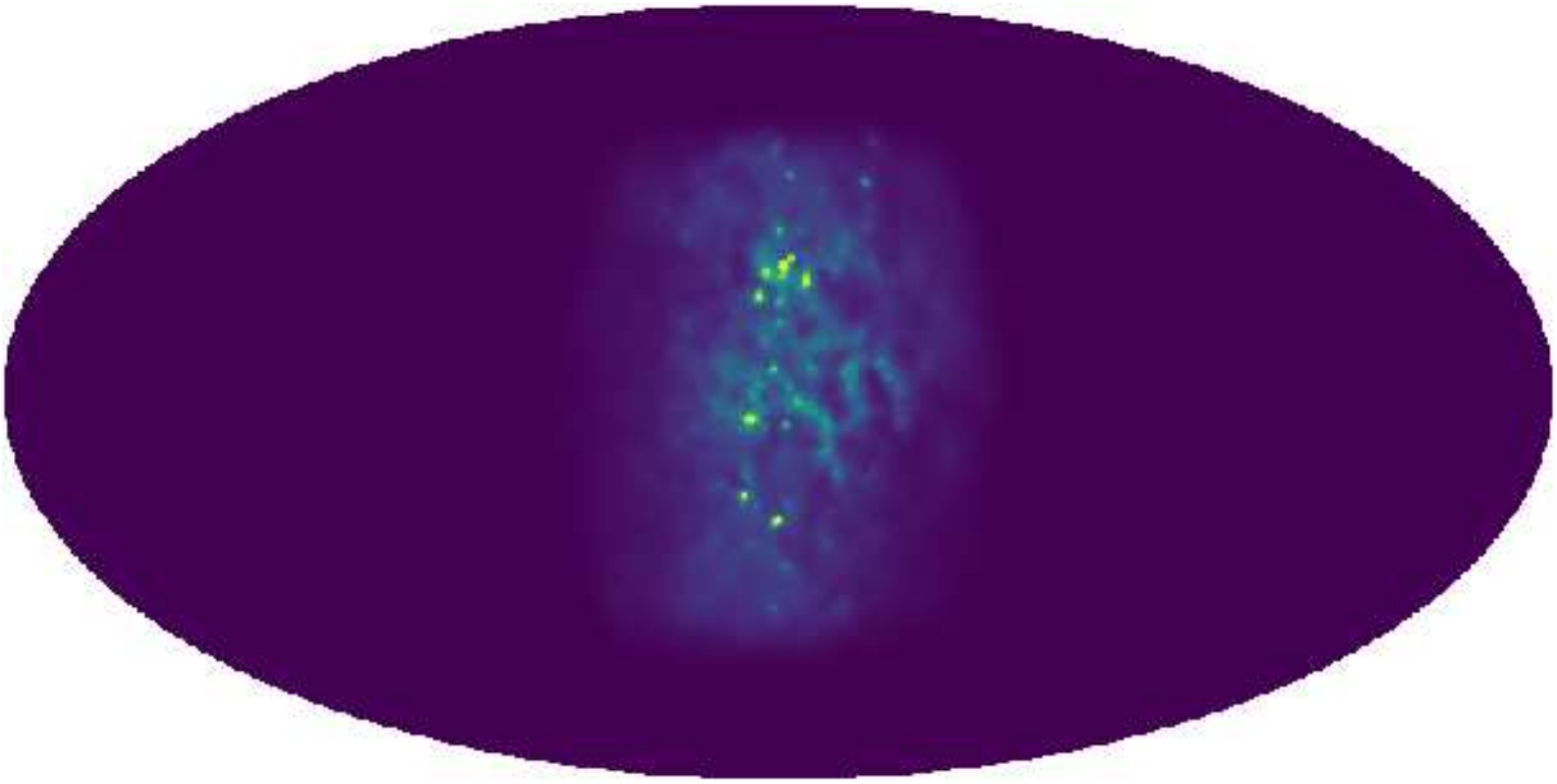
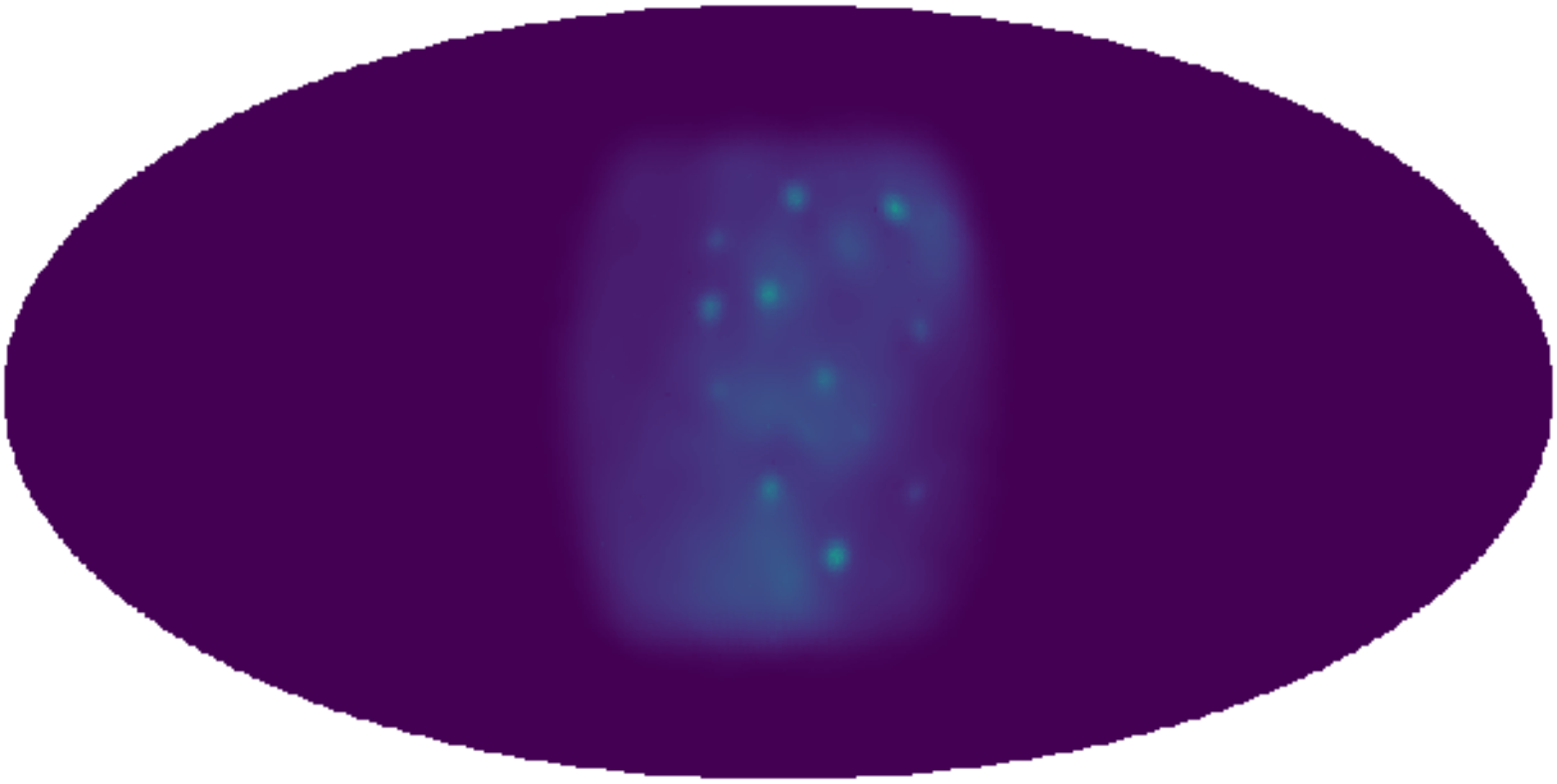
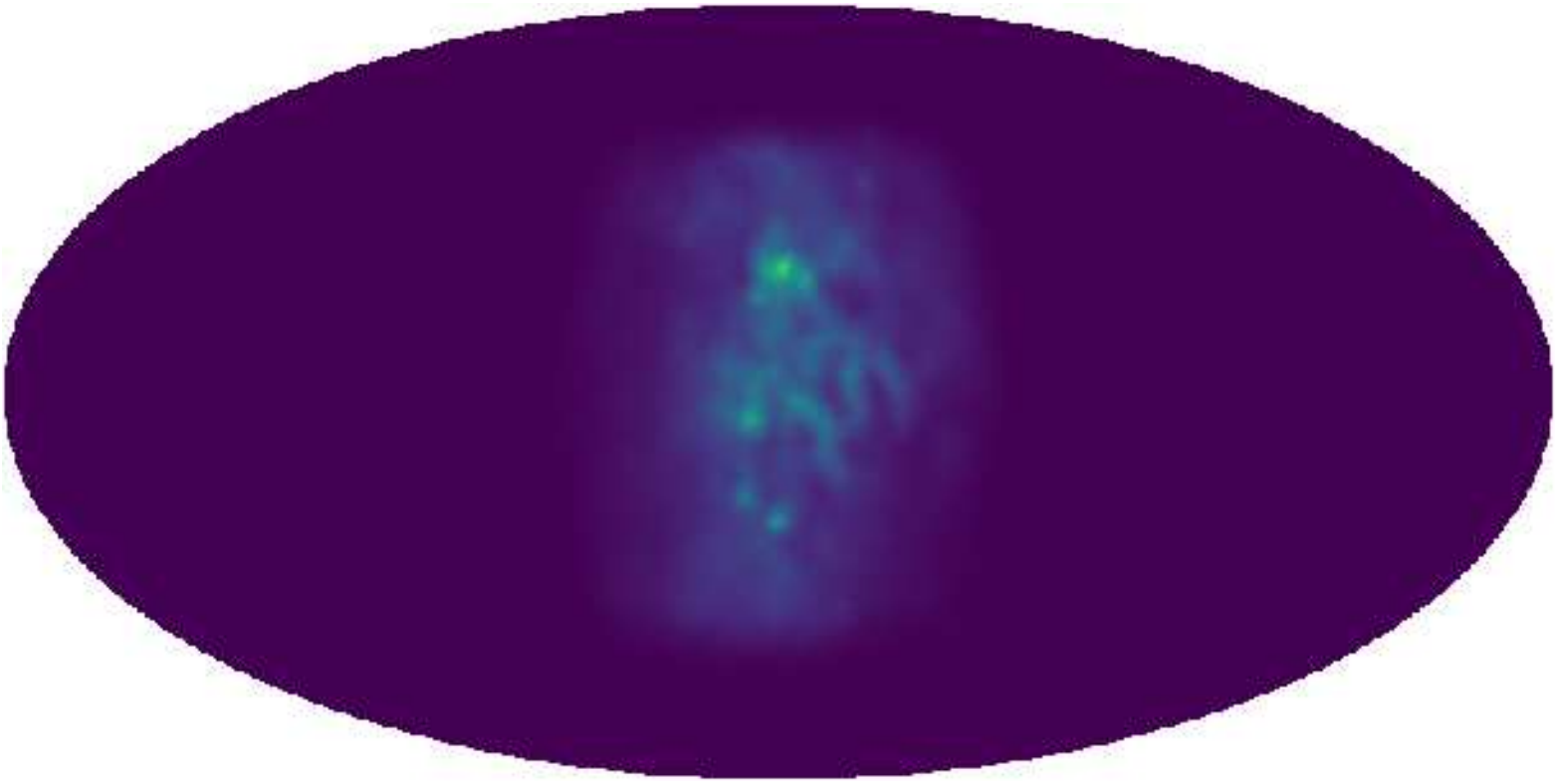
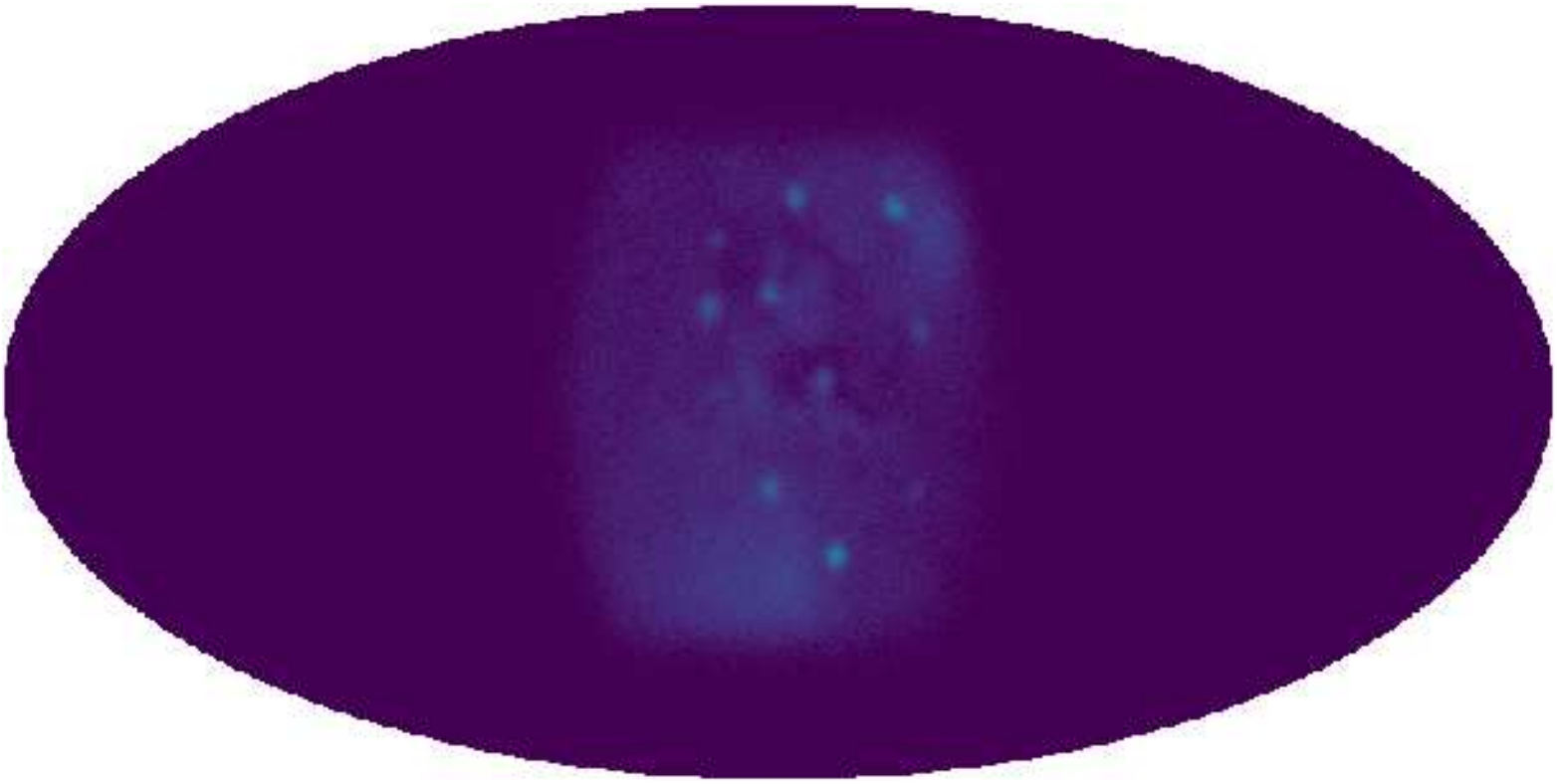
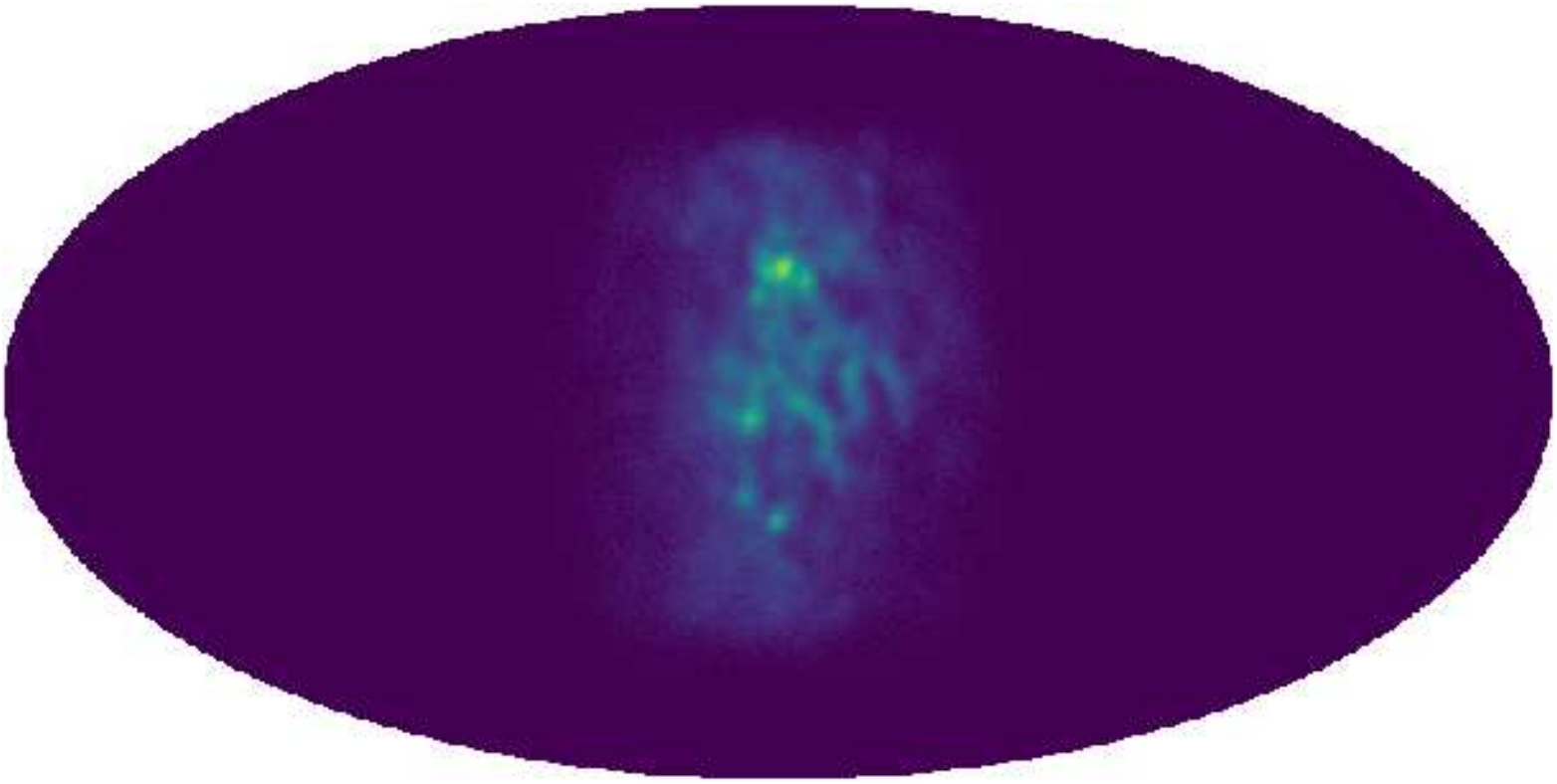
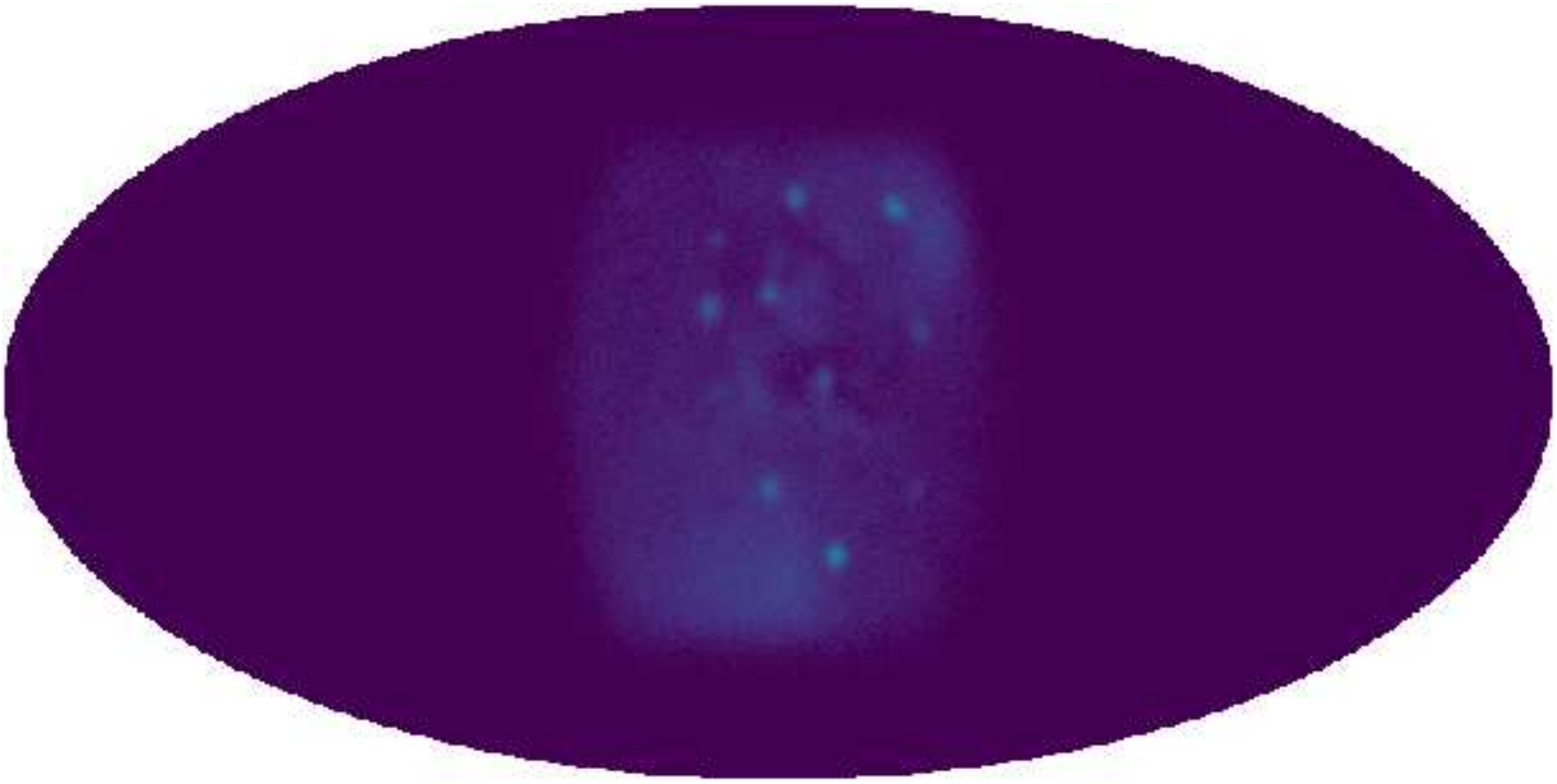
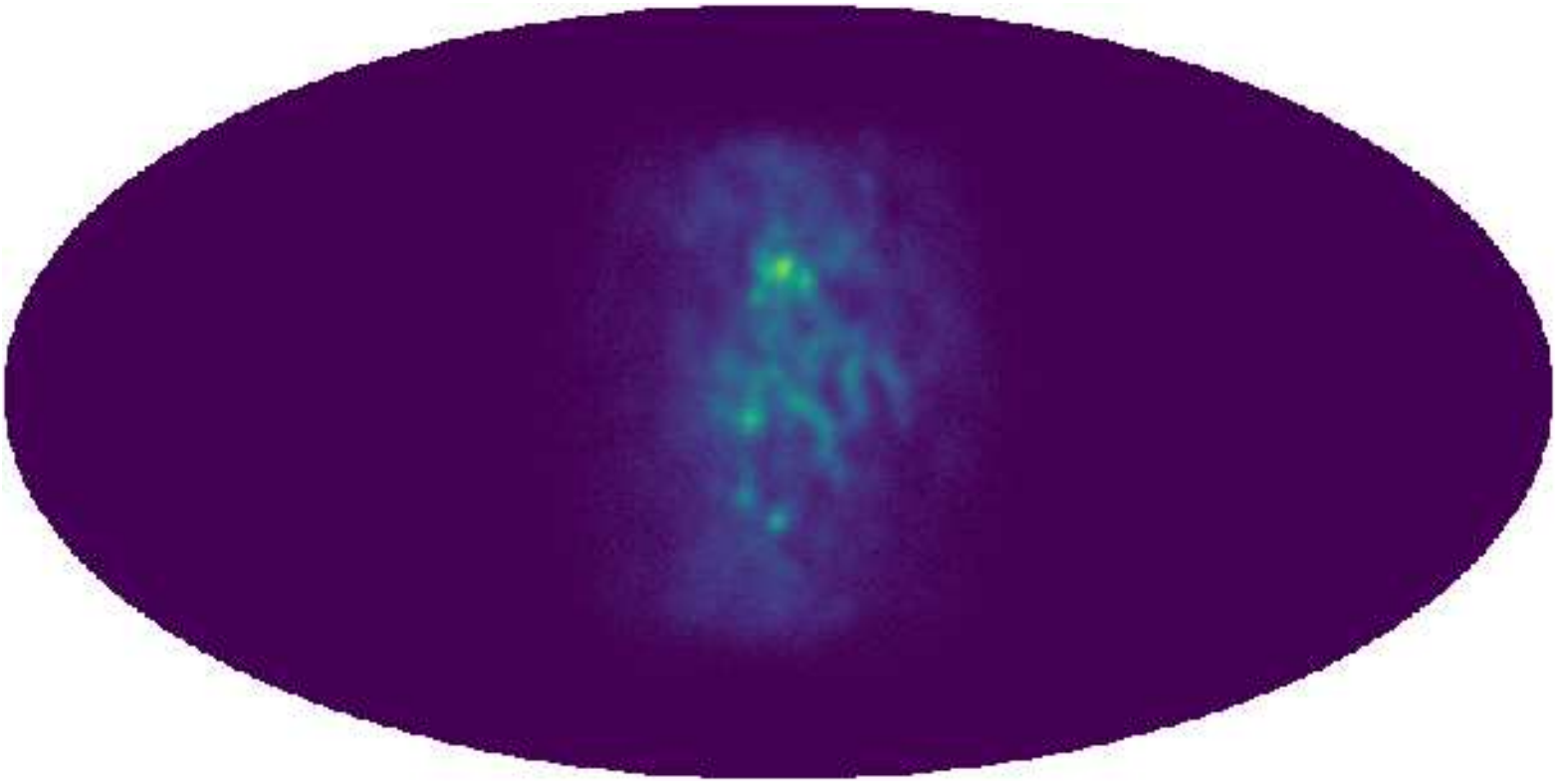



Modified synchrotron
Free-free
Dust
Spectrum of emission (arb. unit)
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
Frequency (arb. unit)


Modified synchrotron
Free-free
Dust
Spectrum of emission (arb. unit)
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
Frequency (arb. unit)




Modified synchrotron
Free-free
Dust
Spectrum of emission (arb. unit)
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
Frequency (arb. unit)


Modified synchrotron
Free-free
Dust
Spectrum of emission (arb. unit)
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
Frequency (arb. unit)