Semi-Blind Source Separation with Learned Physic-Driven Constraints

Rémi Carloni Gertosio, Jérôme Bobin, Fabio Acero

To cite this version:

Rémi Carloni Gertosio, Jérôme Bobin, Fabio Acero. Semi-Blind Source Separation with Learned Physic-Driven Constraints. 2021. ⟨hal-03270406⟩

HAL Id: hal-03270406
https://hal.science/hal-03270406v1

Preprint submitted on 24 Jun 2021

Semi-Blind Source Separation with Learned Physic-Driven Constraints

Rémi Carloni Gertosio*, Jérôme Bobin, Fabio Acero

***Abstract*—Blind source separation (BSS) algorithms are unsupervised methods, which are the cornerstone of multispectral data analysis, allowing for a physically meaningful data decomposition. Being ill-posed inverse problems, BSS algorithms must rely on efficient regularization schemes to better distinguish between the sources and yield interpretable solutions. For that purpose, we investigate a semi-blind source separation approach in which we combine a projected alternate least-squares algorithm with a learning-based regularization scheme. In this article, we focus on constraining the mixing matrix to belong to a learned physics-based manifold; to that end, we propose to make use of the interpolatory autoencoder (IAE). Altogether, we show that this allows for an innovative BSS algorithm, with improved accuracy, which yet provides physically interpretable solutions. The proposed method, coined sGMCA, is tested on realistic multispectral astrophysical data in challenging scenarios involving strong noise, highly correlated spectra and unbalanced sources. The results highlight the significant benefit of the IAE prior to reduce the leakages between the sources, which allows a better disentanglement between the sources.**

***Index Terms*—Blind source separation (BSS), manifold learning.**

I. INTRODUCTION

MULTISPECTRAL data are found in various fields, ranging from astrophysics to biology or remote sensing for instance. Such data are constituted of measurements of the same scene that are acquired at different wavelengths across the electromagnetic spectrum. More precisely, under the linear mixture model, a multispectral measurement $\mathbf{X}_j \in \mathbb{R}^P$ at a channel $j \in [1 \,..\, J]$ of a given wavelength can be expressed as the weighted sum of elementary sources $\{\mathbf{S}_i\}_{i\in[1\,..\,I]}$:

$$\mathbf{X}_j = \sum_{i=1}^{I} \mathbf{A}_j^i \mathbf{S}_i + \mathbf{N}_j, \tag{1}$$

where $\mathbf{A}_j^i$ is the contribution of source i at the wavelength of channel j and $\mathbf{N}_j$ is some additive noise. Equation (1) can be rewritten with matrices, yielding:

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}, \tag{2}$$

where $\mathbf{X} \in \mathbb{R}^{J\times P}$, $\mathbf{A} \in \mathbb{R}^{J\times I}$ and $\mathbf{S} \in \mathbb{R}^{I\times P}$. The columns of the so-called mixing matrix $\mathbf{A}$ include the electromagnetic spectra of the sources.

Analyzing multispectral data can be tackled by factorizing the data $\mathbf{X}$ into a source matrix $\mathbf{S}$ and a mixing matrix $\mathbf{A}$. It is a very challenging task, which is ill-posed and that requires the use of prior information. Different kinds of priors have led to diverse classes of matrix factorization algorithms. One can cite the methods based on independent component analysis (ICA) [1], which assume the statistical independence of the sources, or the extensive class of nonnegative matrix factorization (NMF) algorithms [2], which make the hypothesis that both the sources and the mixing matrix have nonnegative elements. The method presented in this paper is part of the family of the sparse blind source separation (BSS) methods, which exploits the sparsity of the sources in a given representation.

When applied to real data, the first challenge of BSS is to build efficient priors to better discriminate between the sources and limit leakage, that is when the estimate of a component is contaminated by another component in the source or the spectrum domain. In physical applications, this is fundamental, especially to disentangle complex mixtures such as those involving partially correlated sources [3]. If regularization schemes are well suited for the sources, it is much less the case for the mixing matrices, which generally exhibit strong structures (*e.g.* in multispectral imaging, the mixing matrix corresponds to physical spectra). NMF does impose the non-negativity of the mixing weights, but it is a very loose constraint which does not particularly aim at limiting leakage. Secondly, the regularizations should provide physically exploitable information to produce interpretable results.

Altogether, these challenges mandate the design of regularizations that inject physics-driven information. As stated by the manifold hypothesis [4], natural signals generally lie on a low-dimensional manifold embedded in a large space. This is also the case for physically relevant signals in a given application. One approach would amount to switch to end-to-end learning approaches to tackle BSS problems [5]–[7]. However, such methods are not well adapted to account for the exact mixture model and the noise statistics, which is key in scientific applications. Consequently, an hybrid approach that combines standard variational approaches [8] and prior learning is better suited. In this framework, designing an effective BSS algorithm, with interpretable solutions requires building regularizations, for the mixing matrix and/or the sources, that precisely account for the manifold property. Since the underlying manifold structure of the data is generally unknown, one has to learn the regularizations from data.

In the literature, combining variational approaches for inverse problems with regularization learning procedures comes in different flavors. In [9]–[12], the proposed learning architecture builds upon a residual network to mimic standard proximal algorithms, in which regularization learning is one

---

The authors are with IRFU, CEA, Université Paris-Saclay, F-91191, Gif-sur-Yvette, France (e-mail: remi.carlonigertosio@cea.fr; jerome.bobin@cea.fr; fabio.acero@cea.fr)

element of the learning scheme. So far, it is unclear whether this approach is well-suited for multiconvex problems such as BSS. A different way of solving inverse problems with learned regularizations consists in first inverting the observation operator (*e.g.* the mixing matrix, when the sources alone are estimated) with a fast and simple procedure (*e.g.* least-squares solution) and then clean inversion artifacts thanks to a learned denoiser. This technique has been investigated both with standard convolutional networks denoisers [13]–[17] or with generative models with adversarial learning [18]. However, the underlying denoiser depends on the linear operator to be inverted. In the case of BSS, this means that it should depend on the sources and the mixing matrix that are the variables to be estimated. A DNN-based denoiser on the sources is used in [19] in the context of audio source separation, but the denoising is likely not robust with respect to the prevailing leakages in BSS.

So far, we identify two major limitations of these methods: *(i)* they generally require a significant number of training samples, which might make it not well adapted to most physical applications, and *(ii)* in the case of BSS, learning a denoising procedure might not be constraining enough to efficiently discriminate between the sources and limit leakages of unknown morphologies or statistical distributions.

The interpolatory autoencoder: To that purpose, we introduced in [20] a novel framework, coined interpolatory autoencoder (IAE), to build physics-driven regularizations for general multi-linear inverse problems, which BSS is part of. Rather than learning the underlying manifold structure, we propose to learn to travel on it by interpolation between so-called "anchor points", which are samples that belong to the manifold. To do so, we make use of a particular neural-network-based autoencoder that builds maps between the sample space and a code space. It is trained so that the code of each sample can be expressed as a linear combination, or barycenter, of the codes of the chosen anchor points; in doing so, the manifold tends to be linearized in the code space.

More formally, let $\mathcal{M}$ be the manifold to be modeled, from which N anchor points $\{\alpha_n\}_{n\in[1..N]}$ are picked, and let ϕ and ψ be a forward encoder and a backward decoder, respectively. One has:

$$\forall x \in \mathcal{M}, \exists \{\lambda_n\}_{n\in[1..N]} \in \mathbb{R}^N, \phi(x) = \sum_{n=1}^{N} \lambda_n \phi(\alpha_n). \quad (3)$$

The autoencoder is optimized so as to minimize the reconstruction error between some training data $\{x_t\}_t$ and the interpolants in the code space backprojected in the sample space, *i.e.*:

$$\underset{\phi,\psi}{\operatorname{argmin}} \sum_t \left\| x_t - \psi\left(\sum_{n=1}^{N} \lambda_{n,t} \phi(\alpha_n) \right) \right\|_2^2 \quad (4)$$

where $\{\lambda_{n,t}\}_{n\in[1..N]}$ are the weights of the linear interpolation of $\phi(x_t)$ in the code space.

Once the autoencoder is learned, the manifold can be approximated as the barycentric span of the anchor points in the code domain backprojected in the sample domain:

$$\mathcal{M} \approx \left\{ x, \exists \{\lambda_n\}_{n\in[1..N]} \in \mathbb{R}^N, x = \psi\left(\sum_{n=1}^{N} \lambda_n \phi(\alpha_n) \right) \right\}. \quad (5)$$

The IAE allows to build an efficient regularization scheme in a variational inference approach by constraining signals to belong to the resulting barycentric span, that is the modeled manifold. In contrast to the aforementioned data-driven methods, we show in [20] that the resulting interpolatory scheme can perform well even when training samples are scarce. The detailed architecture and a discussion of the hyper-parameters of this autoencoder are presented in [20].

In this article, our main contributions are:

- We adapt and apply the IAE framework in the context of BSS to constrain the spectra constituting the mixing matrix. The resulting regularization is expected to greatly constrain the sources by rejecting efficiently the leakages, allowing an improved disentanglement between the sources.
- We combine the IAE regularization with the sparse BSS algorithm GMCA (generalized morphological component analysis). The resulting procedure, which associates a standard variational approach with learned priors on the spectra, is named semi-blind GMCA (sGMCA). While the known components are better constrained, the components whose emission spectra are poorly or not known can still be identified blindly, which is essential in physical applications.
- We evaluate and compare the proposed method with several BSS algorithms on realistic multispectral astrophysical data in various challenging experimental settings.

Notations: A vector is written in bold lowercase; $\mathbf{x}_n$ denotes the nth entry of vector $\mathbf{x}$. A matrix is written in bold uppercase; $\mathbf{X}_n$ and $\mathbf{X}^m$ are respectively the nth row and the mth column of matrix $\mathbf{X}$. $\|\cdot\|_p$ designates the p-norm for vectors and the entrywise p-norm for matrices. $\odot$ denotes the entrywise multiplication. $\mathbf{X}^\top$ and $\mathbf{X}^+$ designates respectively the transpose and the Moore–Penrose inverse of $\mathbf{X}$. Finally, $\iota_{(\cdot)}$ and $\Pi_{(\cdot)}$ denote the characteristic function and the orthogonal projection on the set in subscript, respectively.

II. THE sGMCA ALGORITHM

A. Principle

The proposed source separation algorithm is based on a variational approach. In order to build the corresponding cost function, we first state the different assumptions on the sources, the mixing matrix and the noise.

Data-fidelity term: In the following, the noise is assumed to be additive, Gaussian, and independent and identically distributed. Thus, in accordance with the forward model in

(2), the data-fidelity term is chosen as being the Frobenius norm between the data $\mathbf{X}$ and the product $\mathbf{AS}$:

$$f(\mathbf{A},\mathbf{S}) = \frac{1}{2}\left\|\mathbf{X}-\mathbf{AS}\right\|_2^2 . \tag{6}$$

Sources constraint: The sources are assumed to be sparse in a given domain $\mathbf{W}$, such as a wavelet representation. The sparsity is enforced by a ℓ_1-penalization term:

$$g(\mathbf{S}) = \left\|\mathbf{\Lambda}\odot\left(\mathbf{S}\mathbf{W}^\top\right)\right\|_1 . \tag{7}$$

The hyperparameter $\mathbf{\Lambda}$ allows to tune the effect of the source regularization; being a matrix, it can depend on the source and the sample. For the sake of simplicity, we will assume thereafter that $\mathbf{W}$ is orthogonal.

Mixing matrix constraint: In a *semi-blind* approach, we suppose that among the I sources, M have a spectrum modeled by a IAE and $I-M$ are fully unknown. Let $\mathbb{M}\subset[1\mathinner{..}I]$ be the indices of the modeled components. The regularization term on the mixing matrix is:

$$h(\mathbf{A}) = \sum_{i\in\mathbb{M}} \iota_{\mathcal{B}_{m_i}}\left(\mathbf{A}^i\right) + \sum_{i\notin\mathbb{M}} \iota_{\mathcal{O}}\left(\mathbf{A}^i\right). \tag{8}$$

Depending whether a spectrum $\mathbf{A}^i$ is modeled or not, the applied constraint differs. If not (*i.e.* $i\notin\mathbb{M}$), the spectrum is constrained to belong to $\mathcal{O}=\{\mathbf{a}\in\mathbb{R}^J, \|\mathbf{a}\|_2=1, \forall j\in[1\mathinner{..}J], \mathbf{a}_j\geq 0\}$, that is the intersection of the unit Euclidean sphere with the nonnegative orthant; the unit-norm constraint prevents scale degeneracies such as $\|\mathbf{A}^i\|_2\to\infty$ and $\|\mathbf{S}_i\|_2\to 0$. If so (*i.e.* $i\in\mathbb{M}$), the spectrum is enforced to belong to the barycentric span $\mathcal{B}_{m_i}$ associated to the corresponding IAE model m_i, which is defined as:

$$\mathcal{B}_{m_i} = \left\{\mathbf{a}\in\mathbb{R}^J, \exists\boldsymbol{\lambda}\in\mathcal{S}^{N_{m_i}}, \mathbf{a}=\psi_{m_i}(\mathbf{\Phi}_{\boldsymbol{m_i}}\boldsymbol{\lambda})\right\} \tag{9}$$

where:

- $\mathbf{\Phi}_{\boldsymbol{m_i}}\in\mathbb{R}^{J\times N_{m_i}}$ is the column-wise application of a forward encoder $\phi_{m_i}:\mathbb{R}^J\to\mathbb{R}^J$ to N_{m_i} anchor points,
- $\boldsymbol{\lambda}\in\mathcal{S}^{N_{m_i}}=\{\boldsymbol{\lambda}\in\mathbb{R}^{N_{m_i}}, \sum_{n=1}^{N_{m_i}}\boldsymbol{\lambda}_n=1\}$ are the barycentric coefficients; they are enforced to sum to one for the sake of robustness to potential contaminants [20],
- $\psi_{m_i}:\mathbb{R}^J\to\mathbb{R}^J$ is a backward decoder.

The autoencoders are learned beforehand, from the data sets of each family of spectra. Thereafter, in order to match the constraint $\iota_{\mathcal{O}}$ applied to the unknown spectra, we will assume that the IAE models reconstruct spectra of unit Euclidean norm.

To sum up, sGMCA seeks to solve the following problem:

$$\min_{\mathbf{A},\mathbf{S}} \frac{1}{2}\left\|\mathbf{X}-\mathbf{AS}\right\|_2^2 + \left\|\mathbf{\Lambda}\odot\left(\mathbf{S}\mathbf{W}^\top\right)\right\|_1 + \sum_{i\in\mathbb{M}} \iota_{\mathcal{B}_{m_i}}\left(\mathbf{A}^i\right) + \sum_{i\notin\mathbb{M}} \iota_{\mathcal{O}}\left(\mathbf{A}^i\right). \tag{10}$$

Equation (10) is non-convex. It is however convex with respect to $\mathbf{S}$ when $\mathbf{A}$ is fixed. This is however not the case for the update with respect to $\mathbf{A}$ since the projection onto the barycentric span is likely not convex. Fortunately, if the first guess obtained by the least-squares estimate is decent, and not completely drowned by leakages from other sources, the projection tends to converge to a stationary point (see Section III-A). Though it is not convex, it generally leads to stable projections that allow using traditional proximal minimization schemes.

B. Algorithm and implementation

```
Input: data $\mathbf{X}$, number of sources $I$, IAE models
$\mathbf{A},\mathbf{S}\leftarrow\mathrm{GMCA}(\mathbf{X},I)$
while convergence not reached do
    Sources update:
    Least squares: $\mathbf{S}\leftarrow\mathbf{A}^+\mathbf{X}$
    Thresholding: $\mathbf{S}\leftarrow\mathcal{T}_{\mathbf{\Lambda}}\left(\mathbf{S}\mathbf{W}^\top\right)\mathbf{W}$
    Mixing matrix update:
    Least squares: $\mathbf{A}\leftarrow\mathbf{X}\mathbf{S}^+$
    Model-to-source mapping & starting points init.
    for $i\in\mathbb{M}$ do
        Barycentric span projection: $\mathbf{A}^i\leftarrow\Pi_{\mathcal{B}_{m_i}}(\mathbf{A}^i)$
    end for
    for $i\notin\mathbb{M}$ do
        Projection on $\mathcal{O}$: $\mathbf{A}^i\leftarrow\Pi_{\mathcal{O}}\left(\mathbf{A}^i\right)\approx\frac{[\mathbf{A}^i]_+}{\|[\mathbf{A}^i]_+\|_2}$
    end for
end while
Output: mixing matrix $\mathbf{A}$, sources $\mathbf{S}$
```

Fig. 1. The sGMCA algorithm

The sGMCA method is described in Fig. 1. It is based on GMCA [21], [22], which is a BSS scheme built upon a projected alternate least-squares (pALS) minimization procedure. The sources and the mixing matrix are initialized with GMCA. At this point the solution is approximate; the mixing matrix and the sources are likely to be contaminated by residuals of other components. $\mathbf{S}$ and $\mathbf{A}$ are updated alternatively and iteratively until convergence is reached. Each update comprises a least-squares estimate, so as to minimize the data-fidelity term f, followed by the application of the proximal operator of the corresponding regularization term [23]. The procedure is stopped when either the estimated sources have stabilized ($\|\mathbf{S}^{(k)}-\mathbf{S}^{(k-1)}\|_2/\|\mathbf{S}^{(k)}\|_2\leq\epsilon$ with $\epsilon=1\mathrm{e}{-6}$ in practice) or a maximal number of iterations have been reached (50 in practice).

1) Sources update: The proximal operator of the sources constraint g is the soft-thresholding operator in the representation $\mathbf{W}$, with threshold $\mathbf{\Lambda}$. Following the pALS minimization scheme, the sources update reads as:

$$\mathbf{S}\leftarrow\mathrm{prox}_g(\mathbf{A}^+\mathbf{X}) = \mathcal{T}_{\mathbf{\Lambda}}\left(\left(\mathbf{A}^+\mathbf{X}\right)\mathbf{W}^\top\right)\mathbf{W} \tag{11}$$

with $\mathcal{T}_{\mathbf{\Lambda}}(\cdot)=\max\left(|\cdot|-\mathbf{\Lambda},\mathbf{0}\right)\odot\mathrm{sign}(\cdot)$, where the absolute, max and sign operators are intended element-wise and $\mathbf{0}$ is a matrix of zeros. The thresholding allows to refine the least-squares estimate of the sources by removing the contaminating noise as well as the potential residuals of other sources, which could come from a misestimation of $\mathbf{A}$. The automatic tuning strategy of the thresholds proposed in GMCA, which is detailed in [22], is implemented similarly in sGMCA. Briefly, the threshold of a source i is equal to $3\sigma_i$, where σ_i is the standard deviation of the noise which contaminates source i. σ_i can either be calculated analytically from the noise level of the data, if it is available, or be estimated with the median absolute deviation estimator. Moreover, the thresholds are adapted to each sample thanks to a ℓ_1-reweighting scheme, which allows to reduce the artifacts induced by the thresholding [24].

2) Mixing matrix update: Let $\mathbf{A_{LS}} = \mathbf{X}\mathbf{S}^+$ be the least-squares estimate of the mixing matrix. Matrix factorization is subject to permutation indeterminacy, which is why the sources need to be mapped with the right models prior to the application of the constraints. The mapping procedure is detailed in section II-B3. Hereinafter, the sources are considered already mapped.

It is recalled that the proximal operator of the characteristic function of a set is the orthogonal projection on the aforementioned set. The orthogonal projection on $\mathcal{O}$ is the composition of the projections on the nonnegative orthant and the unit Euclidean sphere[1]. Thus, if a spectrum is unknown (*i.e.* $i \notin \mathbb{M}$), the update simply reads as:

$$\mathbf{A}^i \leftarrow \Pi_{\mathcal{O}}\left(\mathbf{A}^i\right) \approx \frac{\left[\mathbf{A_{LS}}^i\right]_+}{\left\|\left[\mathbf{A_{LS}}^i\right]_+\right\|_2}. \tag{12}$$

where $[\cdot]_+$ is the projection on the nonnegative orthant.
If a spectrum is modeled (*i.e.* $i \in \mathbb{M}$), one has:

$$\mathbf{A}^i \leftarrow \Pi_{\mathcal{B}_{m_i}}\left(\mathbf{A_{LS}}^i\right). \tag{13}$$

The projection on $\mathcal{B}_{m_i}$ is detailed bellow.

3) Projection on the barycentric span: The projection of a spectrum $\mathbf{a}$ on a barycentric span $\mathcal{B}_m$ is given by:

$$\Pi_{\mathcal{B}_m}(\mathbf{a}) = \psi_m\left(\boldsymbol{\Phi_m}\hat{\boldsymbol{\lambda}}\right), \tag{14}$$

with

$$\hat{\boldsymbol{\lambda}}, \hat{\rho} = \underset{\boldsymbol{\lambda}\in\mathcal{S}^{N_m}, \rho\in\mathbb{R}^+}{\operatorname{argmin}} \left\|\mathbf{a} - \rho\ \psi_m(\boldsymbol{\Phi_m}\boldsymbol{\lambda})\right\|_2^2. \tag{15}$$

The work of [20] is extended by estimating jointly the amplitude ρ of the input spectrum. It is indeed unknown because of the preceding least-squares update and the potential presence of leakages of other sources. The reconstructed spectrum is enforced to be of unit norm (by construction of ψ_m).

The projection on the barycentric span boils down to a non-linear constrained least-squares optimization problem, that does not have a closed-form solution. In practice, it is estimated with the AdaGrad iterative algorithm [25]; the sum-to-one constraint $\boldsymbol{\lambda} \in \mathcal{S}^{N_m}$ is applied by defining a coefficient equal to one minus the sum of the other coefficients.

Problem (15) being non-convex, the quality of the solution greatly depends on the initialization of $\boldsymbol{\lambda}$ and ρ. Drawing inspiration from [20], we propose to initialize $\boldsymbol{\lambda}$ by a fast interpolation in the code domain:

$$\begin{aligned}\tilde{\boldsymbol{\lambda}} &= \underset{\boldsymbol{\lambda}\in\mathcal{S}^{N_m}}{\operatorname{argmin}} \left\|\phi_m(\Pi_{\mathcal{O}}(\mathbf{a})) - \boldsymbol{\Phi_m}\boldsymbol{\lambda}\right\|_2^2 \\ &\approx \Pi_{\mathcal{S}^{N_m}}(\boldsymbol{\Phi_m}^+\ \phi_m(\Pi_{\mathcal{O}}(\mathbf{a}))).\end{aligned} \tag{16}$$

The input spectrum is projected on $\mathcal{O}$ to fulfill the unit-norm hypothesis of the autoencoder. The solution is approximated by a projected least squares, in which $\Pi_{\mathcal{S}^{N_m}}$ is further approximated by a rescaling ($\Pi_{\mathcal{S}^{N_m}}(\boldsymbol{\lambda}) \approx \boldsymbol{\lambda}/\sum_{n=1}^{N_m}\boldsymbol{\lambda}_n$); this is enough for the sake of initialization. Let $\widetilde{\Pi_{\mathcal{B}_m}}(\mathbf{a}) = \psi_m(\boldsymbol{\Phi_m}\tilde{\boldsymbol{\lambda}})$ be the reconstructed spectrum using $\tilde{\boldsymbol{\lambda}}$. ρ is then initialized so as to minimize $\|\mathbf{a} - \rho\ \widetilde{\Pi_{\mathcal{B}_m}}(\mathbf{a})\|_2$, yielding:

$$\tilde{\rho} = \frac{\mathbf{a}^\top \widetilde{\Pi_{\mathcal{B}_m}}(\mathbf{a})}{\left\|\widetilde{\Pi_{\mathcal{B}_m}}(\mathbf{a})\right\|_2^2}. \tag{17}$$

The initialization proposed above is not robust to the potential interferences which are likely to contaminate the input spectrum. Indeed, the calculation of $\tilde{\boldsymbol{\lambda}}$ supposes that the interferences are orthogonal to the input spectrum, which is in general not the case. To that end, we use a greedy initialization procedure, which draws inspiration from the Orthogonal Matching Pursuit algorithm. Succinctly, it associates iteratively the source and the model which minimizes the Euclidean distance between the input spectrum *free of the interference* and its approximate projection. In practice, the interferences are removed by subtracting a linear combination of the identified spectra up to the current iteration. This is performed with a coarse grid-search, which is enough for the initialization. In addition, it is noted that this initialization procedure also allows to map the sources with the right models.

III. NUMERICAL EXPERIMENTS

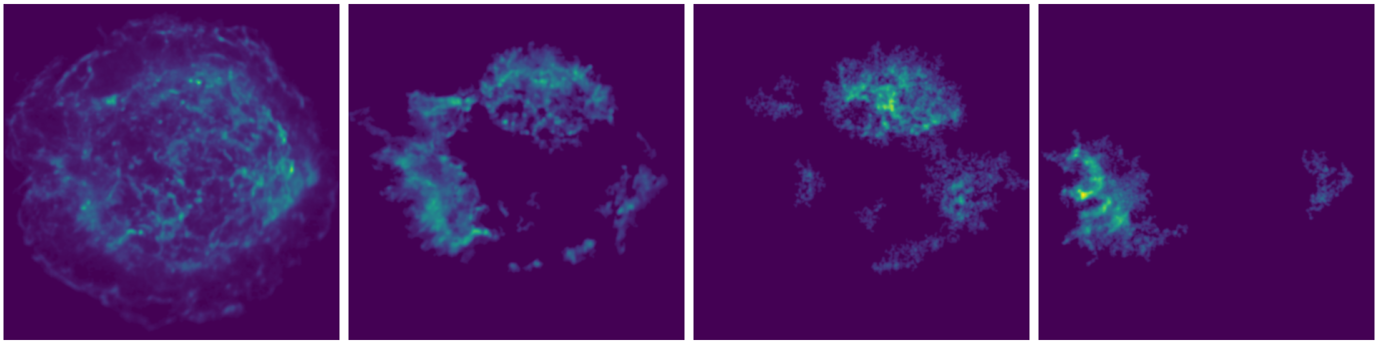
In this section, the proposed method is evaluated on a realistic toy model of the Cassiopeia A supernova remnant observed by the X-ray space telescope Chandra[2] (see [3] for more details about these data). The data are composed of $I = 4$ sources of size $P = 346 \times 346$ (see Fig. 2), specifically one synchrotron source (radiation of the energetic charged particles in the supernova remnant by the synchrotron process), one thermal source (emission from a 10^7 K plasma including continuum and lines emissions) and two Gaussian line emissions. The mixtures are observed over $J = 75$ channels. An absorbed power-law model and a hot plasma emission model produced using the *Astrophysical Plasma Emission Code* (APEC [26]) convolved with the spectral response of the Chandra telescope are used to generate the sets of synchrotron and thermal spectra (see Fig. 3a and 3b). The emission line spectra are modeled as Gaussian kernels, with widths proportional to the center as shown in Fig. 3c.

The sGMCA code that is used is open source (see Appendix). To the best of our knowledge, sGMCA is the first

---

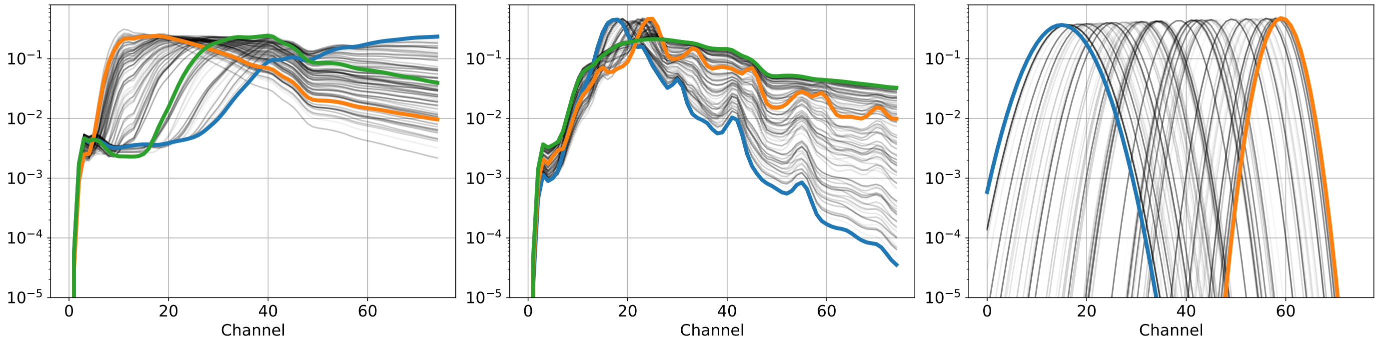
[1] This result holds if at least one coefficient of the input vector is positive, which is in practice the case.

[2] `chandra.harvard.edu`



(a) (b) (c) (d)

Fig. 2. Spatial templates obtained from Chandra X-ray observations of the Cassiopeia A supernova remnant (logarithmic scale). (a) Synchrotron. (b) Thermal. (c) Gaussian I. (d) Gaussian II.



(a) (b) (c)

Fig. 3. Ensemble of (a) absorbed synchrotron model, (b) absorbed thermal and (c) Gaussian lines. The colored thick spectra are the chosen anchor points.

BSS method that makes use of prior physical information. We compare sGMCA with two benchmark experiments:

- an oracle version of GMCA, where $\mathbf{S}$ (respectively $\mathbf{A}$) is estimated with the ground-truth $\mathbf{A}$ (respectively $\mathbf{S}$); it provides an upper-bound for the reconstruction performances,
- an alternate version of sGMCA in which the spectra are regularized with a nearest-neighbor search among the spectra of the training sets; this experiment allows to better highlight the benefit of the IAE regularization.

The sGMCA algorithm is also compared to three standard BSS algorithms, which have no prior physical information on the spectra, namely GMCA and:

- HALS [27], which is a NMF algorithm designed to solve $\min_{\mathbf{A}\geq 0,\mathbf{S}\geq 0} \|\mathbf{X}-\mathbf{AS}\|_2^2$ (the relationship $\geq$ is intended element-wise). The resolution is built upon a block-coordinate descent. The non-negativity of the solution is guaranteed thanks to multiplicative updates.
- SNMF [28], which is a NMF algorithm enforcing the sparsity of the sources in the direct domain by solving $\min_{\mathbf{A}\geq 0,\mathbf{S}\geq 0} \|\mathbf{X}-\mathbf{AS}\|_2^2 + \lambda\|\mathbf{S}\|_1$. Similarly to HALS, multiplicative updates ensure the non-negativity of the solution. To the best of our knowledge, there is no extension of SNMF in a transformed domain, which would better suit the data.

For the GMCA-based algorithms, the sparsity of the sources is enforced in the starlet (an isotropic undecimated wavelet) representation with two details scales [29].

The estimated spectra are assessed with the spectral angular distance (SAD), recast in a logarithmic scale for the sake of precision:

$$\mathrm{SAD}_i = -10\log_{10}\left(\arccos\left(\frac{{\mathbf{A}^{i}}^{\top}\mathbf{A}^{*i}}{\left\|\mathbf{A}^{i}\right\|_2 \left\|\mathbf{A}^{*i}\right\|_2}\right)\right) \tag{18}$$

with $\mathbf{A}^*$ the ground truth mixing matrix. Consequently, the greater the SAD, the more accurate the estimation. We will consider the overall SAD, that we define as the mean SAD over all spectra.

The sources are evaluated with the signal-to-distortion, interference and noise ratios (SDR, SIR, SNR respectively), which are classical performance metrics in BSS [30]. It is reminded that the SDR evaluates the overall reconstruction error, while the SIR, the SNR and the SAR assess more particularly the residuals from other sources, the noise contamination and the remaining artifacts (*e.g.* the ones due to ℓ_1 regularization), respectively. Similarly to the SAD, we will use these metrics in the logarithmic scale.

TABLE I
PARAMETERS OF THE LEARNING STAGE

Parameter	Value		
	Sync.	Therm.	Gauss.
Anchor points	3	3	2
Training samples	541	601	600
Validation and test samples	178	198	200
Number of epochs	10 000	10 000	100 000
Regularization parameter	0.1		
Number of layers	4		
Residual parameter	0.1		
Learning rate	1e−4		
Solver	Adam		
Batch size	25		

TABLE II
MEDIAN RECONSTRUCTION SAD OVER THE TEST SETS

Spectra	Sync.	Therm.	Gauss.	Overall
SAD (dB)	25.26	25.92	24.81	25.18

A. Learning of the IAE models

Before tackling the source separation problem, the IAE models of the three families of spectra need to be learned.
The first step is to choose the anchor points and their numbers. The minimum number of anchor points required by the IAE is the dimension of the underlying manifold plus one. The thermal and synchrotron spectra depend on two physical parameters, therefore we will select three anchor points. The Gaussian spectra depend on one parameter only, thus two anchor points are needed. The choice of the anchor points is in practice not particularly critical [20]; as long as they are not colinear the reconstruction performances are satisfying. The anchor points are selected by hand, so that they tend to maximize their contrasts (see Fig. 3).
Once the anchor points are removed from the data sets, the three sets of spectra are decomposed into training, validation and test sets. The autoencoders are trained using the Adam optimizer, with the largest learning rates which allow to converge. The IAE is recalled to be equipped with skip connections between each layers to avert vanishing-gradient issues. For the same reason, greedy layer-wise pretrainings are performed. The reconstruction performances are found to stabilize as of four layers, which is therefore the number of layers chosen for all three models. The values of the hyper-parameters of the IAE models are summarized in Table I.

Figure 4 shows examples of reconstructed spectra by the IAE, that is the results of the projections on the barycentric spans, as well as the reconstruction errors. The spectra are well reproduced on a reasonably wide dynamic, except for the smallest samples. To correct this, the spectra could be learned in a logarithmic scale, but it is out of the scope of the present work. The median SAR of each type of spectra and in overall are reported in Table II. These results give the IAE modeling error bounds, that sGMCA cannot exceed.

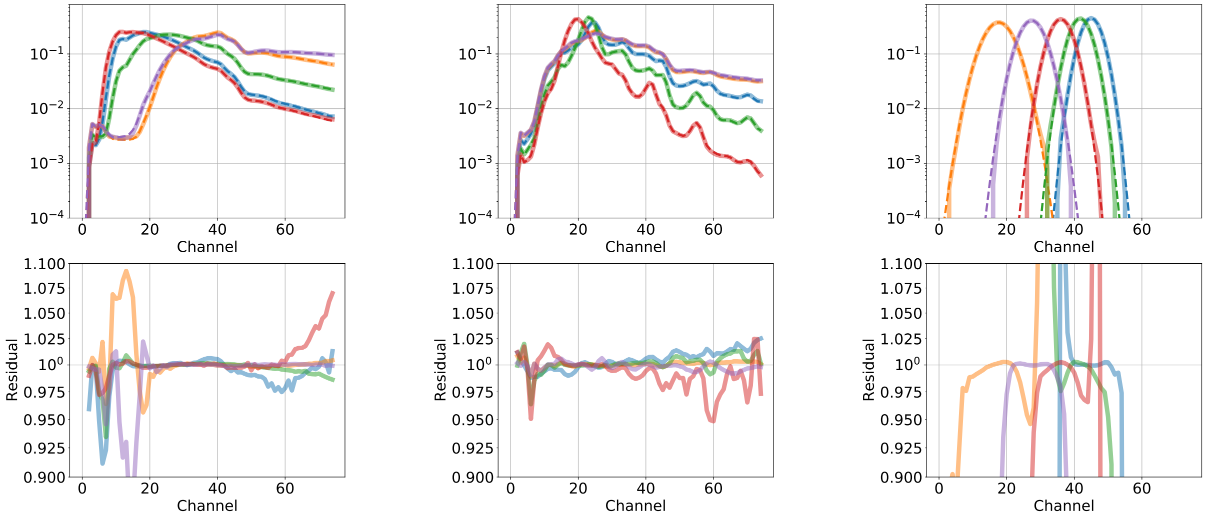
Figure 5 shows two optimization landscapes of barycentric span projections, that is Eq. (15) as a function of $\boldsymbol{\lambda}$. As stated earlier, the optimization landscape is quite convex near the solution, which allows using proximal minimization schemes as long as the initialization is decent.

B. Results

1) Overall results: Firstly, let us compare qualitatively sGMCA to the BSS algorithms on a typical run. Figure 6 shows the estimated spectra along with the estimation errors. The three BSS methods are particularly prone to interferences, which result in either portions of spectra misestimated at zero or with leakages of other components, which is problematic for astrophysical interpretations. On the contrary, sGMCA manages to remove most interferences and recovers satisfactory spectra. Figures 7 and 8 show the estimates of respectively the synchrotron source and the Gaussian II source, as well as the associated estimation errors. Residuals of other sources and/or reconstruction artifacts are clearly visible in the estimation errors of the three blind methods. The sGMCA algorithm provides a more accurate source, whose error equally originates from interference, noise contamination and artifacts.

2) Impact of the experimental settings: In order to confirm quantitatively these results, three experiments, each described in a paragraph below, are performed. These are constituted of Monte-Carlo trials with varying spectrum (taken from the test sets) and noise realizations. The median performance metrics are reported in Fig. 9.
The sparsity hyperparameter of SNMF is found to be very sensitive to the noise level, noise realization and mixing matrix. In the reported results, a fixed mean ”good” hyperparameter $\lambda = 1\mathrm{e}{-7}$ is taken.

a) Noise level: In this experiment, we test the four algorithms with a zero-mean Gaussian noise of varying standard deviation.
The overall tendencies are consistent; the higher the SNR, the better the estimations. As we had observed previously on the example, sGMCA estimates particularly well the spectra; the gain in SAD is from a few to 10 dB compared to GMCA, depending on the noise level. Since the sources are better disentangled, this results in a significant gain of the SIR and thus the SDR.
The additional prior information to which the nearest-neighbor benchmark algorithm has access leads to improved performances compared to GMCA, but not as much as sGMCA. Indeed, the Euclidean distance of the nearest-neighbor search is not adapted for spectra; this emphasizes the advantage of the regularization provided by the IAE, which learns in particular a better-suited metric.
It is noted that the performance metrics of sGMCA and GMCA reach a plateau at a high SNR. According to the hyperparameter tuning strategy of GMCA and sGMCA, the thresholds applied to the sources are low at a low noise level, inducing an underregularization. Interestingly, the sGMCA is less sensitive to this effect, most likely because

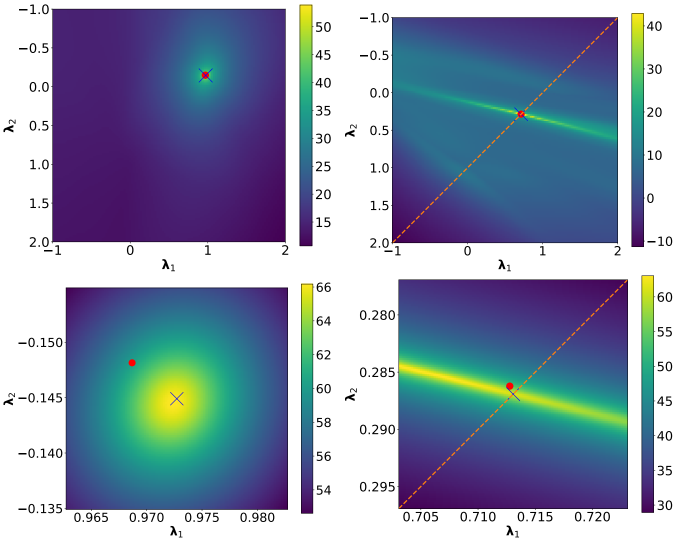


(a)

(b)

(c)

Fig. 4. Examples of reconstructed spectra by the IAE. Top row: spectra (solid lines: reconstructions, dashed lines: ground truth). Bottom row: reconstructed spectra over ground-truth spectra (plotted if ground truth greater than $1\mathrm{e}{-4}$). The figures on a same row share the same ordinate range. (a) Synchrotron. (b) Thermal. (c) Gaussian line emission.



(a)

(b)

Fig. 5. Optimization landscapes of the barycentric span projections, that is $\|\mathbf{a} - \rho^* \ \psi_m(\boldsymbol{\Phi}_{\boldsymbol{m}}\boldsymbol{\lambda})\|_2^2$ as a function of $\boldsymbol{\lambda}$ with the ground truth ρ^*, in an inverse logarithmic scale. Blue cross: barycentric span projection, red circle: initialization. Top row: overview. Bottom row: zoom on the solution. (a) Synchrotron model, with $\boldsymbol{\lambda}_3 = 1 - \boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2$ to fulfill the sum-to-one constraint. (b) Gaussian model, the dashed line is the sum-to-one constraint $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 = 1$.

the components are better separated.

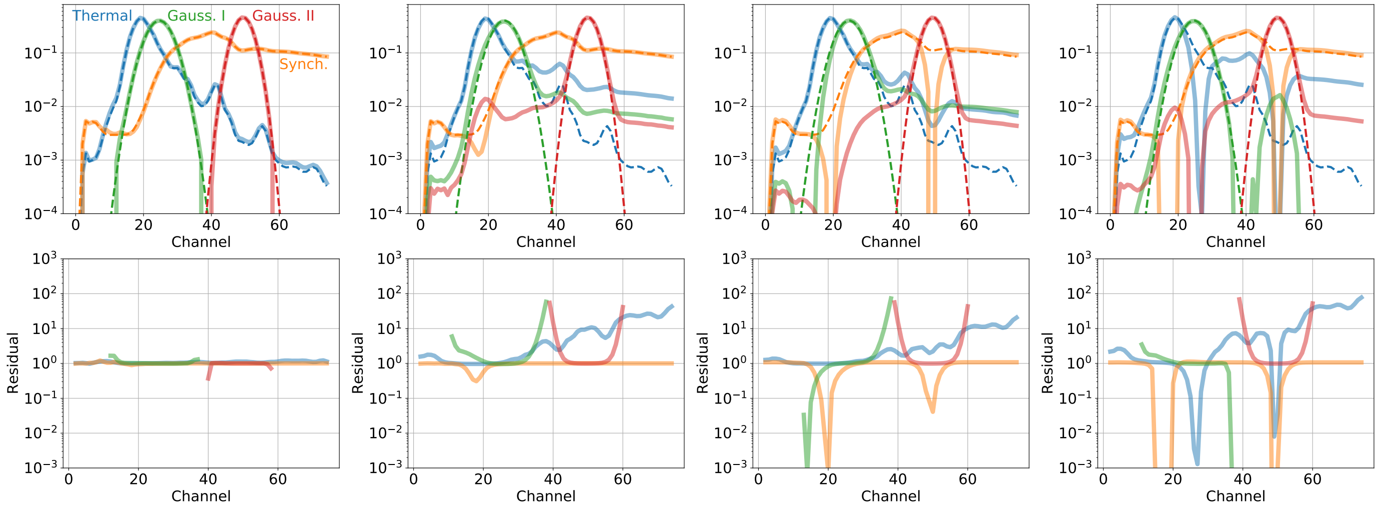
b) Collinearity of the spectra: Two sources originating from a similar physical process can share a similar spectrum, as our two Gaussian line emissions. In this experiment, we aim at testing to which extent the proposed method can separate such sources. To do so, we define a dimensionless variable *distance*; starting from an extreme position where the two Gaussian spectra are disjointed (*distance* equal to one), we progressively bring them closer until superposition (*distance* equal to zero).

Similarly to the previous experiment, sGMCA outperforms the BSS algorithms, with appreciable gains in SAD by 5 to 9 dB and in SDR by 7 to 18 dB compared to GMCA. When the two Gaussian spectra tend to coincide, the performances of sGMCA and the oracle decline because the least-squares update of the sources in the minimization scheme becomes ill-conditioned. Moreover, the performance metrics of GMCA do not depend of the distance because the errors of GMCA are dominated by the thresholding artifacts.

c) Unbalance of the sources: Until now, the four sources were normalized. However, real physical sources are likely to be unbalanced and BSS algorithms may misestimate or fail to identify the least energetic components. In order to evaluate this phenomenon, the sources are unbalanced by multiplying the synchrotron source by a varying scalar while keeping the other sources constant. The reported SDR concerns only the thermal and Gaussian components, since we want to assess the impact on the hidden sources. The SIR is not shown because it is delicate to compare two SIR from data with different amplitudes.

Again, sGMCA allows to recover more precisely the spectra, with a considerable gain in SAD ranging from 8 dB to 12 dB compared to GMCA, and up to 5 dB for the SDR.

When the sources are unbalanced, GMCA estimates precisely the synchrotron spectrum but very poorly the three other spectra (they are contaminated by the synchrotron component). The accurate estimation of the synchrotron spectrum makes it possible to separate the synchrotron source from the data in the least-squares update of the sources, which allows for the retrieval of the hidden sources, hence the acceptable SDR (see examples of estimated sources when the synchrotron source is a hundred times brighter in Fig. 10 and 11).



(a) (b) (c) (d)

Fig. 6. Example of estimated spectra, with SNR = 40 dB, *distance* = 1 and *amplitude ratio* = 1. Top row: spectra (solid lines: estimation, dashed lines: ground truth). Bottom row: estimated spectra over ground truth spectra (plotted if ground truth greater than 1e−4). The figures on a same row share the same ordinate range. (a) sGMCA. (b) GMCA. (c) HALS. (d) SNMF.

On the contrary, the nearest-neighbor benchmark algorithm performs particularly poorly; it generally fails to identify the spectra of the three hidden sources, because at the first iteration (which we recall to be the output of GMCA) the nearest neighbors of all four spectra are synchrotron spectra.

3) Impact of the IAE prior modeling: In the previous experiments, all four components were constrained by IAE models. In this subsection, we focus on the impact of having one or more components fully unknown. To that end, Monte Carlo experiments are performed where the IAE models of the synchrotron, thermal and/or Gaussian spectra are removed, both in the balanced and an unbalanced case. The results are reported in Table III.

In the balanced case, the performance metrics are particularly sensitive to the presence or not of the thermal model. The thermal source strongly correlates with the two Gaussian line sources, it is therefore not surprising that constraining the thermal spectrum notably improves the unmixing.

In the unbalanced case, where the synchrotron source has a norm 100 times greater than the other sources, it is also beneficial to include the synchrotron model. It allows to reduce the leakages of the synchrotron source in the estimates of the hidden sources.

TABLE III

MEDIAN METRICS OVER 100 REALIZATIONS. *Amplitude ratio* IS THE RATIO BETWEEN THE NORM OF THE THERMAL OR A GAUSSIAN SOURCE (THEY HAVE THE SAME NORM) AND THE NORM OF THE SYNCHROTRON SOURCE

		Components modeled with an IAE				
Amplitude ratio	Metric (dB)	All	Therm. & Gauss.	Sync. & Gauss.	Gauss.	None
1	SAD	21.67	20.71	15.59	15.32	12.54
	SDR	38.73	37.31	22.09	21.97	19.45
	SIR	38.54	35.36	18.81	18.52	16.61
0.01	SAD	17.73	12.27	8.83	8.27	5.78
	SDR	33.74	26.32	23.13	22.80	19.86
	SIR[a]	39.15	31.66	24.40	23.87	17.77

[a] calculated over the thermal and Gaussian sources

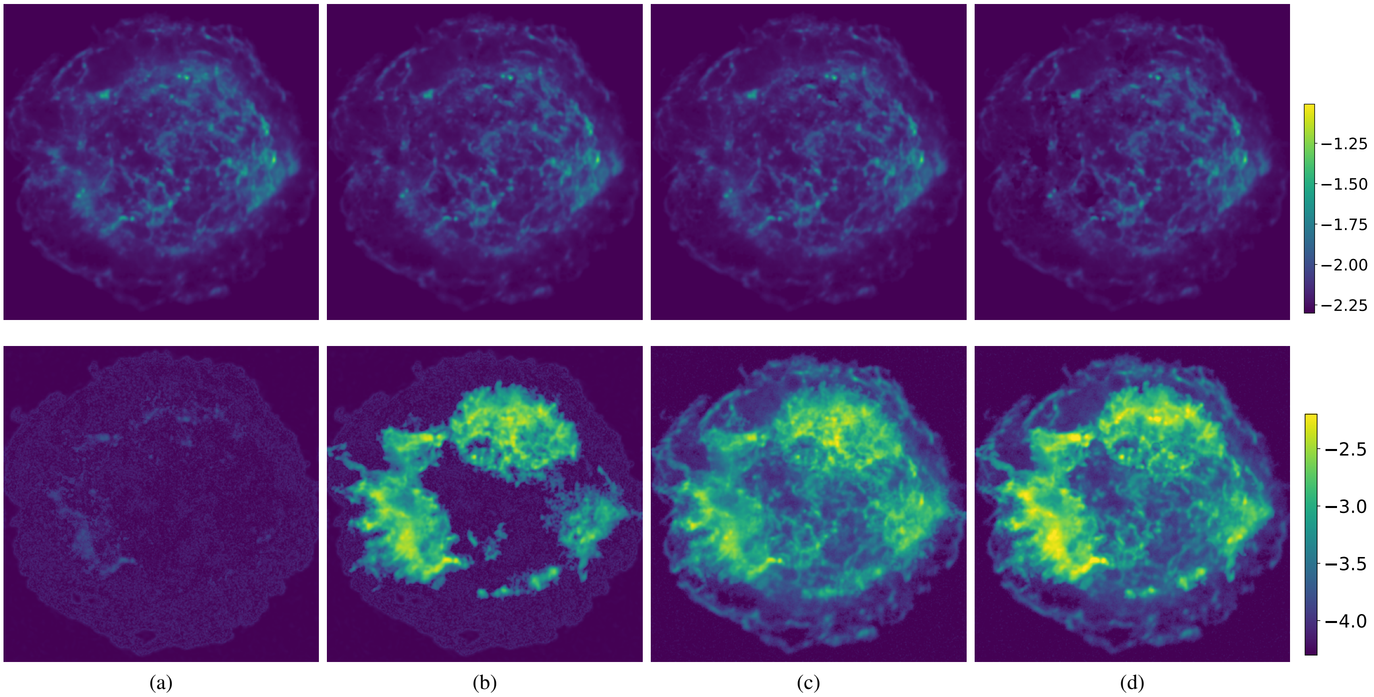
IV. CONCLUSION

We introduce a novel source separation approach to tackle physical multispectral data. Compared to standard blind source separation methods, the objective is twofold: to better discriminate between sources and to ensure the provision of physically relevant information. To do so, we make use of a learned, physic-driven, prior on the spectra of the sought-after sources in a standard variational framework. Extensive numerical experiments on realistic astrophysical data show that the introduced regularization efficiently rejects inter-source leakages, thus improving significantly the estimations of both the sources and the spectra, including in challenging settings.

We choose to specifically illustrate the application of the proposed sGMCA algorithm in the context of multispectral imaging. If remote sensing is a natural application of such kind of methods, it can be applied to a large number of matrix factorization problems with manifold-valued/manifold-constrained factors. One can think of biomedical signal processing (*e.g.* LC-MS data analysis) or EEG and MEG data analysis in neurosciences. In the former, the temporal response of the device (*e.g.* retention time) naturally belongs to an unknown manifold. In the latter, the sGMCA algorithm could be applied to extract certain brain activity patterns from MEG data.

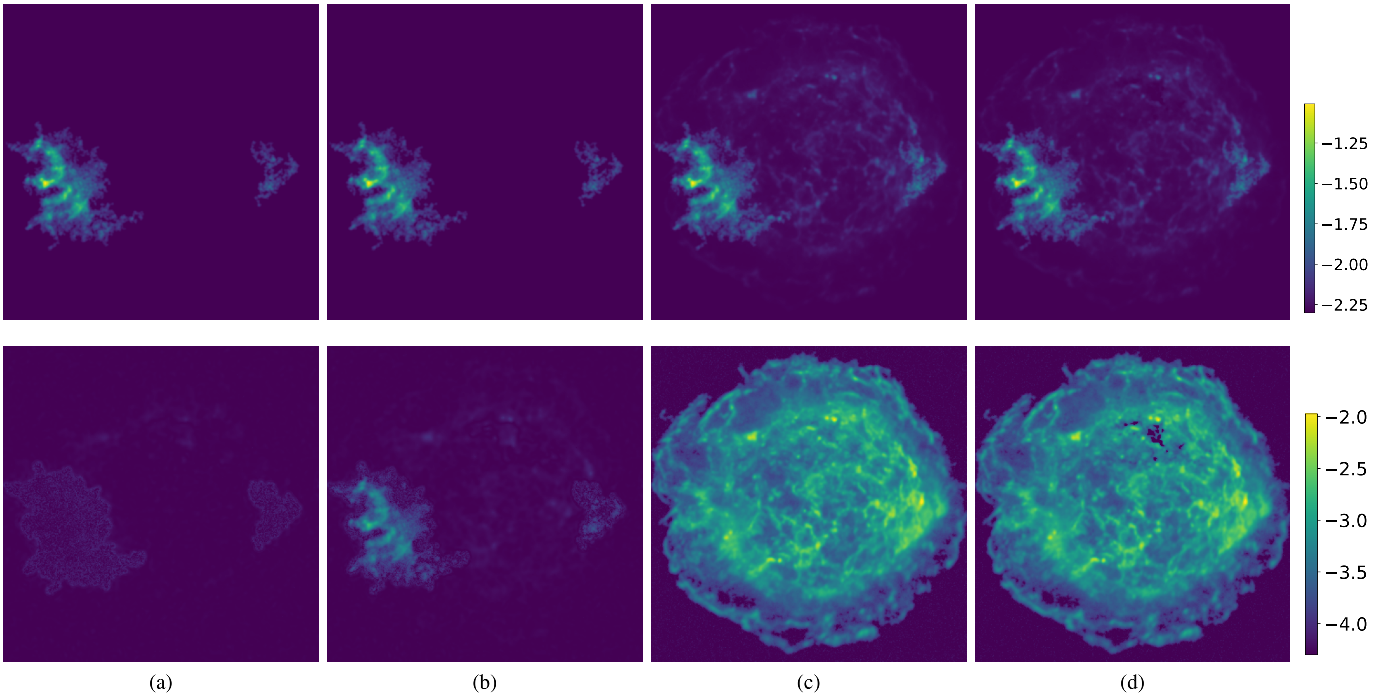
APPENDIX

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



(a) (b) (c) (d)

Fig. 7. Example of the estimated synchrotron source, with SNR = 40 dB, *distance* = 1 and *amplitude ratio* = 1. Top: estimations (logarithmic scale), bottom: absolute error (logarithmic scale). The figures on a same row share the same color scale. (a) sGMCA. (b) GMCA. (c) HALS. (d) SNMF.



(a) (b) (c) (d)

Fig. 8. Example of the estimated Gaussian II source, with SNR = 40 dB, *distance* = 1 and *amplitude ratio* = 1. Top: estimations (logarithmic scale), bottom: absolute error (logarithmic scale). The figures on a same row share the same color scale. (a) sGMCA. (b) GMCA. (c) HALS. (d) SNMF.

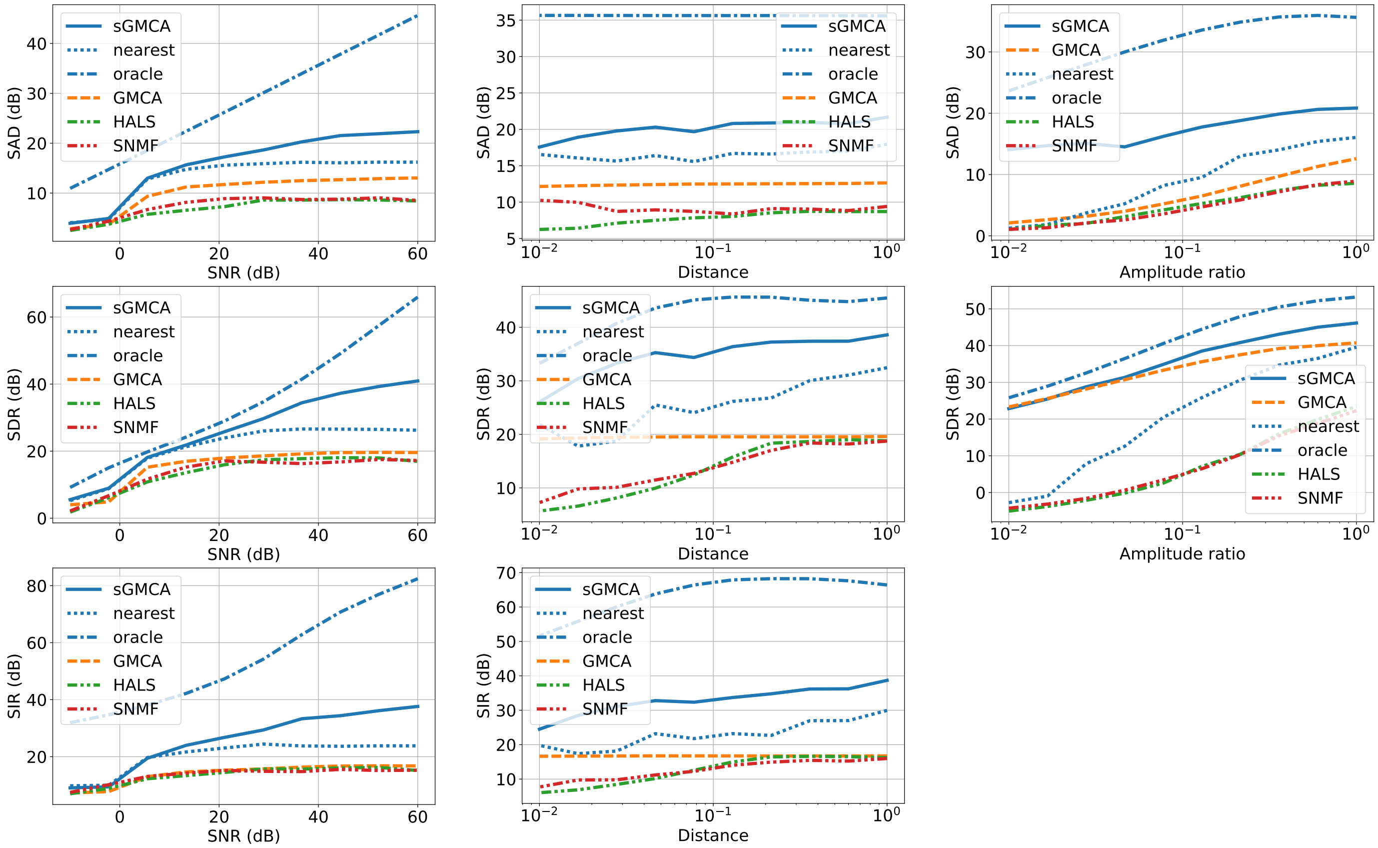


Fig. 9. Median metrics over 100 realizations. The top row (SAD) concerns only the spectra. The middle and bottom rows (SDR and SIR) concern only the sources. In these experiments, the SNR and SAR are poorly informative and thus not reported. *Distance* denotes a dimensionless distance between the two Gaussian line emissions – at 1 the latter are approximately disjointed and at 0 they coincide. *Amplitude ratio* is the ratio between the norm of the thermal or a Gaussian source (they have the same norm) and the norm of the synchrotron source; in this experiment, the reported SDR is calculated over the thermal and Gaussian sources only and the SIR is not shown because it is delicate to compare two SIR from data with different amplitudes.

REFERENCES

[1] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, Inc., 2010.

[2] N. Gillis, *Nonnegative Matrix Factorization*. Society for Industrial and Applied Mathematics, 2020.

[3] A. Picquenot, F. Acero, J. Bobin, P. Maggi, J. Ballet, and G. Pratt, “Novel method for component separation of extended sources in X-ray astronomy,” *A&A*, vol. A139, no. 627, 2019.

[4] C. Fefferman, S. Mitter, and H. Narayanan, “Testing the manifold hypothesis,” *Journal of the American Mathematical Society*, vol. 29, pp. 983–1049, 2013.

[5] L. B. Almeida, “MISEP – linear and nonlinear ICA based on mutual information,” *Journal of Machine Learning Research*, vol. 4, no. Dec, pp. 1297–1318, 2003.

[6] P. Brakel and Y. Bengio, “Learning independent features with adversarial nets for non-linear ICA,” *arXiv preprint arXiv:1710.05050*, 2017.

[7] H. Kameoka, L. Li, S. Inoue, and S. Makino, “Supervised determined source separation with multichannel variational autoencoder,” *Neural Computation*, vol. 31, no. 9, pp. 1891–1914, 2019.

[8] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, 1st ed. Springer Publishing Company, Incorporated, 2008.

[9] J. Adler and O. Öktem, “Solving ill-posed inverse problems using iterative deep neural networks,” *Inverse Problems*, vol. 33, no. 12, p. 124007, Nov. 2017.

[10] J. Adler and O. Öktem, “Learned primal-dual reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.

[11] D. Gilton, G. Ongie, and R. Willett, “Neumann networks for linear inverse problems in imaging,” *IEEE Transactions on Computational Imaging*, vol. 6, pp. 328–343, 2020.

[12] Y. Yang, J. Sun, H. Li, and Z. Xu, “Deep ADMM-net for compressive sensing MRI,” in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ser. NIPS’16. Red Hook, NY, USA: Curran Associates Inc., 2016, p. 10–18.

[13] J. Xie, L. Xu, and E. Chen, “Image denoising and inpainting with deep neural networks,” in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012, pp. 341–349.

[14] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, “Learning a variational network for reconstruction of accelerated MRI data,” *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.

[15] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, “Deep convolutional neural network for inverse problems in imaging,” *IEEE Transactions on Image Processing*, vol. 26, 2017.

[16] Y. Romano, M. Elad, and P. Milanfar, “The little engine that could: Regularization by denoising (RED),” *SIAM J. Imaging Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.

[17] F. Sureau, A. Lechat, and J.-L. Starck, “Deep learning for a space-variant deconvolution in galaxy surveys,” *Astronomy and astrophysics*, vol. 11, no. A67, 2020.

[18] S. Lunz, O. Öktem, and C.-B. Schönlieb, “Adversarial regularizers in inverse problems,” in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018, pp. 8507–8516.

[19] A. A. Nugraha, A. Liutkus, and E. Vincent, “Multichannel audio source separation with deep neural networks,” *IEEE/ACM Transactions on*

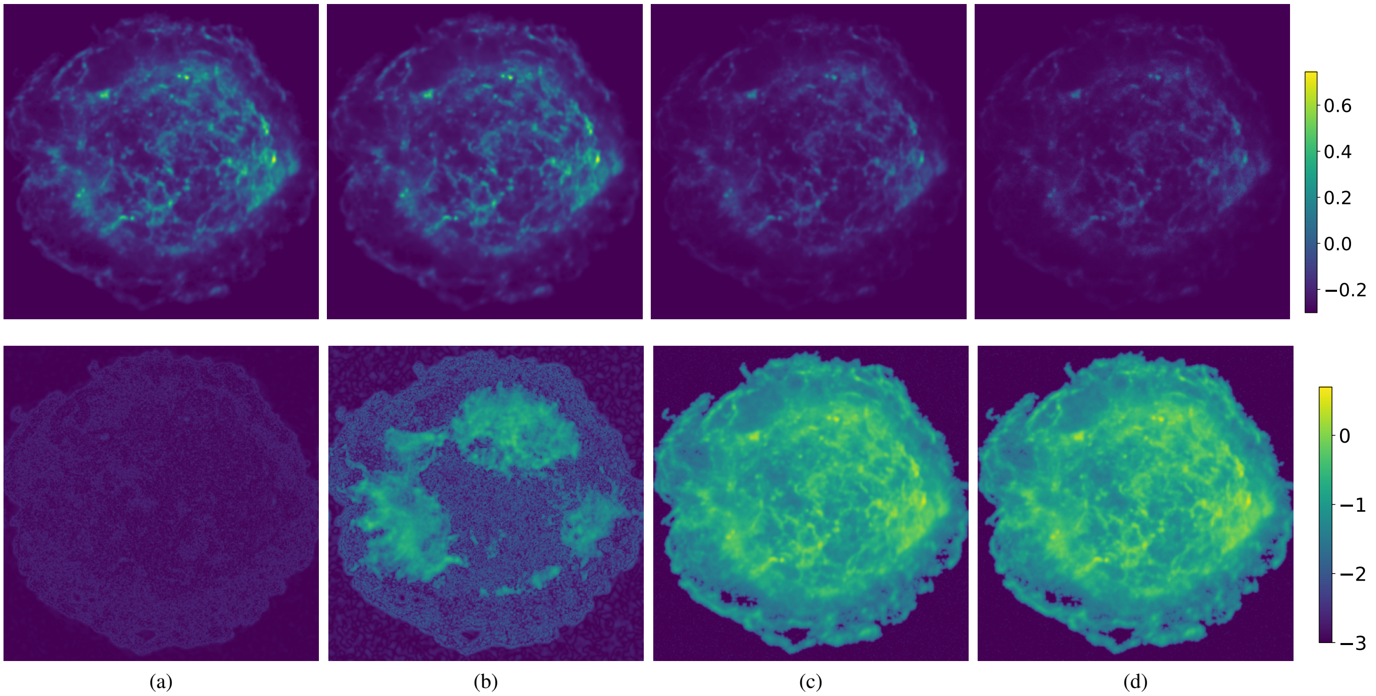


Fig. 10. Example of the estimated synchrotron source, with SNR $= 40$ dB, *distance* $= 1$ and *amplitude ratio* $= 0.01$. Top: estimations (logarithmic scale), bottom: absolute error (logarithmic scale). The figures on a same row share the same color scale. (a) sGMCA. (b) GMCA. (c) HALS. (d) SNMF.

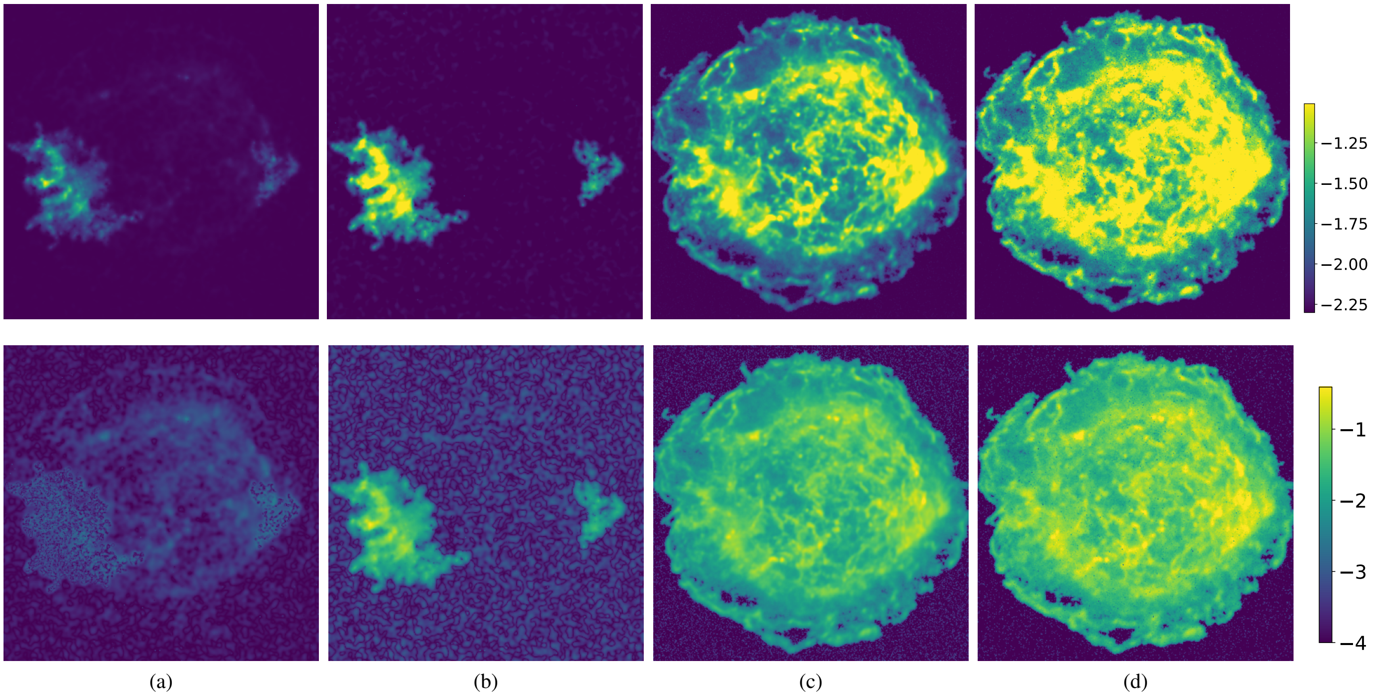


Fig. 11. Example of the estimated Gaussian II source, with SNR $= 40$ dB, *distance* $= 1$ and *amplitude ratio* $= 0.01$. Top: estimations (logarithmic scale), bottom: absolute error (logarithmic scale). The figures on a same row share the same color scale. (a) sGMCA. (b) GMCA. (c) HALS. (d) SNMF.

Audio, Speech, and Language Processing, vol. 24, no. 9, pp. 1652–1664, 2016.

[20] J. Bobin, R. Carloni Gertosio, C. Bobin, and C. Thiam, "Non-linear interpolation learning for example-based inverse problem regularization," Jun. 2021, working paper or preprint. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03265254

[21] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *Image Processing, IEEE Transactions on*, vol. 16, no. 11, pp. 2662 –2674, Nov. 2007.

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

[23] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, 2014.

[24] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2007.

[25] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, p. 2121–2159, Jul. 2011.

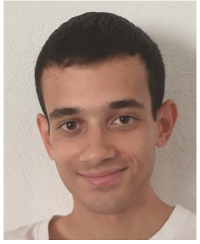
[26] A. R. Foster and K. Heuer, "PyAtomDB: Extending the AtomDB Atomic Database to Model New Plasma Processes and Uncertainties," *Atoms*, vol. 8, no. 3, p. 49, Aug. 2020.

[27] N. Gillis and F. Glineur, "Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization," *Neural Computation*, vol. 24, no. 4, pp. 1085–1105, 2012.

[28] J. Le Roux, F. Weninger, and J. Hershey, "Sparse NMF – half-baked or well done?" Mitsubishi Electric Research Laboratories, Cambridge MA, USA, Tech. Rep. TR2015-023, Mar. 2015.

[29] J. Starck, F. Murtagh, and M. Bertero, *Starlet transform in astronomical data processing*. United States: Springer New York, May 2015, vol. 1, pp. 2053–2098.

[30] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.



Rémi Carloni Gertosio graduated from Ecole Normale Supérieure (ENS) Paris-Saclay, France, in 2019. He received a M.Sc. degree in signal and image processing from Université Paris-Saclay, France. He also passed the Agrégation de Sciences Industrielles option Ingénierie Electrique in 2018.

He is currently pursuing a Ph.D. degree at CEA Paris-Saclay, France, under the supervision of J. Bobin, where he is developing blind source separation methods to tackle widefield and wideband multiwavelength measurements. His research interests focus on signal processing, inverse problems, blind source separation, machine learning.

Jérôme Bobin graduated from the Ecole Normale Supérieure (ENS) de Cachan, France, in 2005. He received the M.Sc. degree in signal and image processing from ENS Cachan and Paris-Sud University, the Agregation degree in physique and the Ph.D. degree in electrical engineering from University of Paris-Saclay in 2004 and 2008, respectively, and the Habilitation degree in 2015.

He has been a Postdoctoral Scholar in collaboration with Prof. Candes in Applied Mathematics at Caltech from 2008 to 2009 and with the Mathematics Department, Stanford University, from 2009 to 2010. He joined the CEA, France, in 2010. His research interests include statistical signal processing, multiscale methods and sparse representations in signal/image processing, machine learning, and their applications in physics and astrophysics.

Fabio Acero received his Ph.D. degree in astronomy in 2008 studying the remnants of exploded stars (supernovae remnants) with X-ray telescopes and the Habilitation degree in 2020. He has been a Postdoctoral Scholar at the University of Montpellier and at the NASA Goddard Space Flight Center.

He is currently a Staff Researcher in the Astrophysics Department, CEA-Saclay, where he studies our Galaxy with ground or space-based telescopes in the quest for the sources of cosmic-rays (highly-energetic particles). The potential particle accelerators include supernovae remnants, pulsars, and pulsar wind nebulae. He is an expert in high-energy data analysis in X- and gamma-rays with a special emphasis on the analysis of extended sources.