



HAL
open science

METRIC LEARNING FOR SEMI-SUPERVISED SPARSE SOURCE SEPARATION WITH SPECTRAL EXAMPLES

Jerome Bobin, Fabio Acero, Adrien Picquenot

► **To cite this version:**

Jerome Bobin, Fabio Acero, Adrien Picquenot. METRIC LEARNING FOR SEMI-SUPERVISED SPARSE SOURCE SEPARATION WITH SPECTRAL EXAMPLES. CAMSAP 2019, Dec 2019, Guadeloupe, France. hal-02426994

HAL Id: hal-02426994

<https://hal.science/hal-02426994>

Submitted on 3 Jan 2020

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.

METRIC LEARNING FOR SEMI-SUPERVISED SPARSE SOURCE SEPARATION WITH SPECTRAL EXAMPLES

Jerome Bobin, Fabio Acero and Adrien Picquenot

CEA, IRFU, DaP/DEDIP
F-91191 Gif-sur-Yvette

Index Terms— blind source separation, sparsity, metric learning, physical constraint

ABSTRACT

Sparse Blind source separation (sBSS) is an unsupervised matrix factorization problem, which is now a key tool to analyse multispectral data, especially in astrophysics. Being an ill-posed problem, designing priors is crucial. In the present paper, we investigate how the prior knowledge based on examples of physical spectra can be exploited in sBSS, based on the projection onto the barycentric span of these examples. For that purpose, we investigate different metrics to build such projections, and further introduce a novel machine learning approach to build physically relevant reconstruction. Secondly, we show how this can be deployed to design a semi-blind sparse BSS method coined. Preliminary numerical results on realistic astrophysical X-ray images show very promising separation results.

1. INTRODUCTION

Multi/hyperspectral data are formed of m observations \mathbf{X}_i , each of which is made of t samples. The standard mixture model defines each observation as a linear combination of n sources \mathbf{S}_j to which Gaussian noise is added: $\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}$, where $\mathbf{X} \in \mathbb{R}^{m \times t}$ is the data matrix, $\mathbf{S} \in \mathbb{R}^{n \times t}$ the source matrix, $\mathbf{A} \in \mathbb{R}^{m \times n}$ the mixing matrix and $\mathbf{N} \in \mathbb{R}^{m \times t}$ for the noise contribution. BSS aims at recovering both the mixing matrix \mathbf{A} and the sources \mathbf{S} from the data \mathbf{X} only, which is an unsupervised matrix factorization. Since it is ill-posed, it requires additional assumptions about the sources and/or the mixing matrix, such as statistical independence [1], non-negativity [2] or sparsity [3] to only name three.

In a very large number of physical applications, examples of spectra (*i.e.* columns of \mathbf{A}) are available that can originate from simulations of parametric models or measurements. *The goal of this article is to investigate how these examples can be exploited to constrain the estimated mixing matrix.* To the best of our knowledge, this approach has never been investigated so far in the scope of BSS. The proposed approach first

builds upon constraining the estimated mixing matrix to lie in the barycentric span of the available examples, according to given metric. Several metrics will be studied. Different physical examples will be considered, for which fixed metrics do not yield physically relevant estimates. For that purpose we introduce in Section 2 a machine learning approach that allows learning such relevant metric based on some training set. In Section 3, we show how this example-based constraint can be exploited in the semi-blind BSS setting.

2. BUILDING MODEL-BASED CONSTRAINTS

Barycentric span and orthogonal projection: In the next, it is assumed that some column \mathbf{a} of the mixing matrix can be described by some model \mathcal{M} , for which d examples $\{\varphi_{\mathcal{M}}^e\}_{e=1, \dots, d}$ are known. Each example is normalized to have unit ℓ_1 norm. If the model is parametric, it could be possible to directly estimate the underlying parameters. However, this has two major drawbacks: i) the model can be rather complex to cope with as it can include highly non-linear physical effects as well as instrumental responses, which general leads to highly challenging non-linear optimization problems, and ii) the model may not be known and only some examples of spectra are available. To alleviate these caveats, we propose to *constrain \mathbf{a} to belong to the barycentric span of the examples* according to some metric ϕ .

Formally, this assumption means that there exists some positive weights $\{\lambda_e\}_{e=1, \dots, d}$ and $\sum_e \lambda_e = 1$ so that \mathbf{a} is the solution to:

$$\min_{\mathbf{a}} \sum_{e=1}^d \lambda_e \phi(\mathbf{a}, \varphi_{\mathcal{M}}^e), \quad (1)$$

Several metrics (distances or divergences) can be considered. We will investigate the following three standard metrics: i) the Euclidean distance, which leads to barycenters that are the weighted sum of the examples, ii) the Kullback-Leibler (KL) divergence and iii) the Wasserstein or Earth Mover's distance [4], which has recently attracted a lot of interest (see [5] and references therein). With the Euclidean and KL metrics, computing barycenters, as defined in Eq. 1, can be tackled using a proximal gradient descent algorithm [6]. Following [5], Wasserstein barycenters can be efficiently evaluated thanks to

This work was supported by the European Research Council - Contract 678282

an entropic regularization of the original Wasserstein distance and the use of the Sinkhorn iterative algorithm [7].

In the sequel, enforcing \mathbf{a} to belong to some barycentric span requires computing its orthogonal projection onto the barycentric span $\mathcal{B}_\phi(\{\varphi_e^{\mathcal{M}}\})$, which requires finding the weights $\{\lambda_e\}_{e=1,\dots,d}$ so that:

$$\min_{\{\lambda_e\}} \mathcal{J}(\{\lambda_e\}) = \|\mathbf{a} - \mathbf{b}^{\mathcal{M}}(\{\lambda_e\})\|_{\ell_2}^2, \quad (2)$$

where $\mathbf{b}^{\mathcal{M}}(\{\lambda_e\})$ is the barycenter that is the solution to Eq. 1. The function $\mathcal{J}(\{\lambda_e\})$ being a smooth, the orthogonal projection $P_{\mathcal{B}_\phi(\{\varphi_e^{\mathcal{M}}\})}(\mathbf{a})$ can be evaluated using a standard gradient descent algorithm with fixed stepsize τ . In practice, evaluating the gradient with respect to the barycentric weights $\{\lambda_e\}_e$ can be tackled using Theano’s auto-differentiation function ¹ as done for the Wasserstein distance in [5]. A similar strategy also applies to the Euclidean distance and the Kullback-Leibler divergence. In this article, preliminary results have been obtained with fixed gradient stepsize τ (in the range $[1e - 4, 1e - 3]$), whose value has been adapted to the metric.

2.1. Learning a data-driven metric

The ability of fixed metrics to build physically relevant estimations of spectra will highly depend on the shape of the spectra as well as the evolution of their shape when their underlying physical characteristics evolve. In this article, two distinct meaningful cases will be investigated in the field of X-ray astrophysical imaging: i) *emission lines* are the instrumental response to a single Dirac at a fixed wavelength, which are generally well described by diracs convolved with a Gaussian-shaped kernel, with a wavelength-dependent width ii) thermal emissions, which lead to observed spectra with several Gaussian-shaped emission lines on top of a continuum. In astrophysical X-ray imaging, the evolution of the continuum yields spectra with different slopes and maximum positions. The amplitude of the emission lines may vary. An example is featured by the purple solid line of Figure 1.

In these cases, improving the accuracy of the proposed barycentric constraint requires *designing data-driven metrics, which are more adapted to capture the physically meaningful variations of the sought-after spectra*. For that purpose, **we introduce a new approach that consists in learning metrics that are adapted to build physically relevant barycenters from examples**, which makes perfect sense in a large number of applications where training/validation sets can be built based on physical/instrumental models.

Let $\Omega^{\mathcal{M}}$ be some training set made of T samples for a given model \mathcal{M} . The proposed metric learning method builds upon the structure of a standard autoencoder [8], which is defined by an encoder Φ that maps the input to the feature space, and a decoder Ψ that maps back to the input space. In contrast

to standard autoencoders, metric learning is performed by minimizing the error between the input and the reconstructed projection (according to the Euclidean distance) onto the subspace spanned by the examples *i.e. in the learnt/encoded feature space*.

More precisely, for some encoder Φ , the representation of some input $\omega \in \Omega^{\mathcal{M}}$ (*resp.* some example φ_e) is denoted by $\Phi(\omega)$ (*resp.* $\Phi(\varphi_e)$). One can then define the orthogonal projection of $\Phi(\omega)$ onto the span of $\{\Phi(\varphi_e)\}_e$. To that end, we resort to a simple scheme that further ease the learning process:

- The reconstructed spectra are not enforced to belong to the simplex (*i.e.* with positive entries and unit ℓ_1 norm). Since the reconstruction error in the input domain is minimized, it should be close to be the case.
- The interpolation in the feature space is not enforced to be a convex combination of the transformed examples. It is substituted by the orthogonal projection onto the subspace spanned by $\{\Phi(\varphi_e)\}_e$, which admits a closed-form expression. The weights are then obtained by minimizing the Euclidean distance between the input spectrum and the examples in the feature space:

$$\{\hat{\lambda}_e(\omega)\}_e = \operatorname{argmin}_{\{\lambda_e\}_e} \left\| \Phi(\omega) - \sum_e \lambda_e \Phi(\varphi_e) \right\|_{\ell_2}, \quad (3)$$

which is a standard least-square problem for fixed Φ . The reconstruction is then defined as $\mathbf{b}^{\mathcal{M}}(\hat{\lambda}_e(\omega)) = \Psi\left(\sum_e \hat{\lambda}_e(\omega) \Phi(\varphi_e)\right)$

Learning the metric means estimating Φ and Ψ as follows:

$$\min_{\Phi, \Psi} \sum_{\omega \in \Omega^{\mathcal{M}}} \left\| \omega - \Psi\left(\sum_e \hat{\lambda}_e(\omega) \Phi(\varphi_e)\right) \right\|_{\ell_2}^2, \quad (4)$$

where the weights $\{\hat{\lambda}_e(\omega)\}_e$ are defined in Eq. 3. In this article both the encoder and decoder are implemented with multilayer neural networks, such as in standard autoencoders [8]. The number of layers is \mathcal{L} , the number of hidden units per layer is fixed to m and activation functions are hyperbolic tangents. The learning stage described in Eq. 4 then boils down to estimating the weight matrices and biases of each layer of both Φ and Ψ . For that purpose, we use the autograd python module ² along with the Adam optimizer.

Illustration In this section, fixed metrics as well as the proposed metric learning approach are evaluated on two distinct cases of spectra that are generally used to describe astrophysical X-ray spectra in the energy range 2 – 6keV. These two types of spectra are the Gaussian-shaped emission lines and

¹<https://github.com/Theano/Theano>

²<https://pypi.org/project/autograd/>

thermal emissions (see Fig. 1). In these experiments, the number of channels is fixed to $m = 50$ and $d = 2$ examples have been selected. For both models 24 samples are available. Amongst them 2 significantly different spectra have been selected at random as examples, 15 have been chosen at random to compose the training set and the remaining 7 form the test set. The same number of training and test samples have been chosen for the two types of spectra. For the fixed metrics (*i.e.* Euclidean, KL and Wasserstein), the optimization parameters are fixed as follows: i) *barycenter evaluation*: the step-size is fixed to 10^{-3} and the number of iterations to 500, and ii) *computation of the orthogonal projection*: the gradient step-size is set to 10^{-4} and the number of iterations to 2000. These quantities have been set empirically so that they provide good results for both types of spectra. As for the metric learning method, the number of iterations is fixed to $2 \cdot 10^4$ and the step-size to $1e - 4$. These values numerically led to solutions with stable values of the cost function for all tested number of layers and for the various kinds of spectra. The number of layers is set to $\mathcal{L} = 4$ and the number of hidden units per layer to $m = 50$. In the case of the thermal emission (see Figure 1), none of the fixed metrics are able to capture with accuracy the peak height/position, the slope of the spectrum at high energy and the emission lines. Interestingly, as testified by the reconstruction error (right panel), the metric learning approach leads to a very precise reconstruction of these different structures from only $d = 2$ examples. More quantitatively, the mean-squared error of the reconstruction from the test set is reported in Table 1. It includes the results obtained for the metric learning method for different numbers of layers. This confirms that learning the metric significantly improves the reconstruction results. Additionally, the accuracy increases with the number of layers up to $\mathcal{L} = 4$; it then collapses for a deeper network. This could very likely be fixed by resorting to standard techniques such as batch normalization or residual network structure, which are known to ease the learning of deep networks. These investigations are left to future work.

3. SEMI-BLIND SPARSE BSS

The sGMCA algorithm ‘ In this section, we investigate how the proposed example-based constraint can be combined with sparse BSS to provide more physical and more accurate separation results. To that end, let assume that amongst the n sought-after sources some are perfectly unknown and should be treated blindly and others, say n' admit spectra that can be described with n' different physical models $\{\mathcal{M}_\eta\}_{\eta=1, \dots, n'}$. The indices of these sources and their corresponding mixing matrix column is denoted by \mathcal{I} . In this article, we will only consider cases where a single model applies to a single source; relaxing this limitation is left for future work. Each spectrum \mathbf{A}^j for $j \in \mathcal{I}$ is then enforced to belong to the barycentric span of the examples related to model \mathcal{M}_j as

follows:

$$\operatorname{argmin}_{\mathbf{A}, \mathbf{S}} \|\mathbf{A} \odot \mathbf{S}\mathbf{W}\|_1 + \sum_{j \in \mathcal{I}} \iota_{\mathcal{S}_m}(\mathbf{A}^j) + \dots \dots \sum_{j \in \mathcal{I}} \iota_{\mathcal{B}_{\phi_j}(\{\varphi_{\mathcal{M}_j}^e\})}(\mathbf{A}^j) + \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2, \quad (5)$$

where $\iota_{\mathcal{S}_m}$ is the characteristic function of the m -dimensional hypersphere (*i.e.* the corresponding columns of \mathbf{A} are enforced to have unit ℓ_2 norm), $\iota_{\mathcal{B}_{\phi_j}(\{\varphi_{\mathcal{M}_j}^e\})}$ is the characteristic function of the barycentric span of the examples $\{\varphi_{\mathcal{M}_j}^e\}$ according to the metric ϕ_j and \mathbf{W} is the operator that is associated with some sparsifying transform.

The problem in Eq. 5 is globally nonconvex but is convex with respect to \mathbf{S} . Convexity with respect to \mathbf{A} is not ensured since the constraint onto the hypersphere is not rigorously convex and, equipped with the learnt metric, the convexity of the example-based constraint is not guaranteed as well. However, orthogonal projections onto these constraints can be defined and used in an optimization procedure. While several methods can be considered optimization, we opted for a projected alternate least-squares method such as GMCA [3], which allows designing algorithms that have a low computational burden, with easy-to-tune parameters. Extending the GMCA algorithm to the proposed semi-blind setting leads to a two-step algorithm semi-blind GMCA (sGMCA), which is detailed in Alg. 1. More precisely, step 1 is composed of two stages: i) the least-squares estimation of the sources, which boils down to applying the left pseudo-inverse of the current estimate of the mixing matrix to the data as in step 1 – *a*, ii) apply the soft-thresholding operator \mathcal{T} with parameter Λ_i to each source bfS_i . The determination of the values for Λ are described in the next paragraph. The second step copes with the update of the mixing matrix, which is formed of three stages: i) computing the least-squares estimate of \mathbf{A} as in step 2 – *a*, ii) projection each column of \mathbf{A} with indices in \mathcal{I} onto the barycentric span $\mathcal{B}_{\phi_j}(\{\varphi_{\mathcal{M}_j}^e\})$ (described in 2) and iii) normalize the remaining columns to have unit ℓ_2 norm.

Parameter setting The initialization is performed using the GMCA algorithm [9]. The number of sources n is assumed to be known and we assume in this article that only a single column of the mixing matrix is described by a single model. Hence, constraining each model to apply to only a single source helps avoiding extra degeneracies. This assumption already allows to tackle a wide range of application scenarios and can be generalized. Then, initialization includes identifying the indices \mathcal{I} that associate each model to a single source from the first guess. For each model, this is carried out by finding the column of mixing matrix that provides the minimum Euclidean distance with respect to its projection onto the barycentric span.

As emphasized in [3], the role of sparsity in BSS is to improve the separation process by increasing the contrast between the sought-after sources, and to increase its robustness with respect to noise. At a given iteration, the error that contaminates a given source estimate is mainly comprised of noise

and interferences between the sources, which comes from an imperfect estimation of the mixing matrix. The goal of the thresholding step is then to remove a large part of these two terms. Assuming that they are morphology dense and therefore non-sparse, the standard deviation of this combination of errors can be empirically evaluated using the Median Absolute Deviation estimator (MAD). The regularization parameters Λ can include both a noise-related value based on the MAD and a signal-related term that limits the thresholding bias in the spirit of re-weighted ℓ_1 regularization. In this article, the threshold for each source \mathbf{S}_j will be constant and equal to $3\sigma_j$, where σ_j is the MAD-based estimate of the contamination standard deviation (see [10] for more details). The maximum number of iterations is fixed to 1000. The stopping criterion is based on the spectral angular distance between consecutive iterates of the mixing matrix. More precisely, for normalized columns of \mathbf{A} with unit ℓ_2 norm, the algorithm stops when $\max_j \left| \arcsin \left| \langle \mathbf{A}^{j(k)}, \mathbf{A}^{j(k+1)} \rangle \right| \right| \leq \epsilon$, where $\epsilon = 1e - 6$

4. NUMERICAL EXPERIMENTS

Experimental setting In this section, comparisons are carried out on realistic simulations of X-ray images provided by the Chandra space telescope³. These data correspond to observations of the Cassiopea A supernova remnants [11]. They are comprised of a linear combination of 3 astrophysical components (displayed in Fig. 3): synchrotron emission, a thermal emission and a redshifted iron (Fe) emission lines (described by the Gaussian-shape emission line) as shown in Fig. 2. These components are representative of typical supernovae remnants in the energy band 5000 – 6000 eV (electron-volt) - see [11]. In the next experiments, the data are composed of 50 observations of size 128×128 pixels. In addition to the sGMCA algorithm, comparisons will be performed with the HALS algorithm [2], which is a NMF algorithm that minimizes a standard quadratic data fidelity term under positivity constraints. In both the GMCA and sGMCA algorithms, the signal representation \mathbf{W} is chosen as the isotropic undecimated wavelet transform [12]. The sGMCA algorithm is tested with two distinct metrics: i) the KL divergence, and ii) the metric learnt with the machine learning approach introduced in Section 2. The metric-based constraint applies to the thermal and line emissions. In the following comparisons are carried out based on the mixing matrix criterion C_A , which is defined as: $C_A = \|\mathbf{Q}\mathbf{A}^+ \mathbf{A}^* - \mathbf{I}\|_{\ell_1}$, where $\mathbf{Q}\mathbf{A}^+$ is the pseudo-inverse of the estimated mixing matrix that has been corrected for the scale/permutation indeterminacies with the matrix \mathbf{Q} . \mathbf{A}^* is the true mixing matrix. Low values of C_A then relates to better separation performances. Each result has been obtained given as the mean of 10 Monte-Carlo simulations with

different noise realisations.

Results The contribution of the synchrotron emission is estimated blindly. Furthermore, it is often a dominant emission that hampers the extraction of dimmer emissions. In this paragraph, the robustness of the methods is assessed when the relative flux (ratio with respect to the total ℓ_2 norm of the contribution of the other sources) of the synchrotron emission increases from 1 to 20; the SNR is fixed to 15dB. As displayed in Fig. 4, the sGMCA algorithm leads to an improvement of about one order of magnitude, even when the synchrotron emission is highly dominant.

5. CONCLUSION

We introduce a novel semi-blind source separation method, based on examples. In the proposed approach, semi-blindness is exploited by enforcing the sought-after mixing matrix to have columns that belong to some example-based barycentric span. While fixed metric do not generally yield physically relevant spectra, we further introduce a machine learning approach – inspired by the autoencoders – to learn a metric that is adapted to build accurate and relevant estimations from examples. This constraint is employed in the proposed sGMCA algorithm to tackle sparse semi-blind source separation. Preliminary results based on realistic X-ray astrophysical data, show very promising results, which will be further detailed during the conference.

	$\mathcal{L} = 3$	4	5	Eucl.	Wass.	KL
G-shaped	6.7	39.3	4.6	-0.5	13.5	25.4
Thermal	40.5	47.2	21.8	13.5	10.6	13.7

Table 1: MSE in dB of the reconstructed spectra from different metrics.

Algorithm 1 semi-blind GMCA

- 1: **procedure**
 - 2: Initialization based on GMCA
 - 3: **while** Convergence is not reached **do**
 - 4: (1-a) Least-square estimation: $\mathbf{S}' = \mathbf{A}^{(k)+} \mathbf{X}$
 - 5: (1-b) Thresholding: $\mathbf{S}^{(k+1)} = \mathcal{T}_\Lambda(\mathbf{S}'\mathbf{W}) \mathbf{W}^T$
 - 6: (2-a) Least-square estimation: $\mathbf{A}' = \mathbf{X}\mathbf{S}^{(k+1)+}$
 - 7: (2-b) Projection onto the barycentric span:
 - 8: $\forall j \in \mathcal{I}, \mathbf{A}^{j(k+1)} = P_{\mathcal{B}_{\phi_j}(\{\varphi_{\mathcal{M}_j}^e\})}(\mathbf{A}'^j)$
 - 9: (2-c) Normalize all columns \mathbf{A}'^j with $j \notin \mathcal{I}$
 - 10: **end while**
 - 11: $\mathbf{b}^{\mathcal{M}}(\{\lambda_e\}) = \operatorname{argmin}_{\mathbf{a}} \sum_{e=1}^d \lambda_e \phi(\mathbf{a}, \varphi_{\mathcal{M}}^e)$
 - 12: **end procedure**
-

³<http://chandra.harvard.edu/>

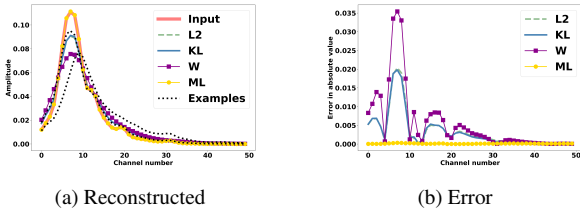


Fig. 1: Left: reconstructed thermal emission spectra of the KL, Wasserstein, Euclidean and learnt metrics. Right: estimation errors.

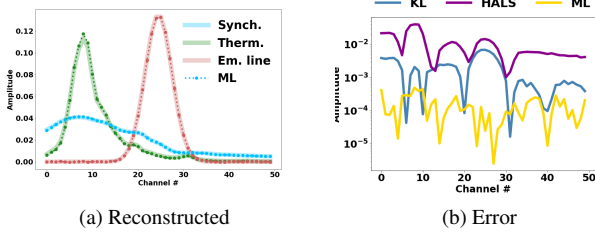


Fig. 2: Left: input and estimated spectra with sGMCA. Right: Estimation errors for HALS and dGMCA with the KL and the learnt metric.

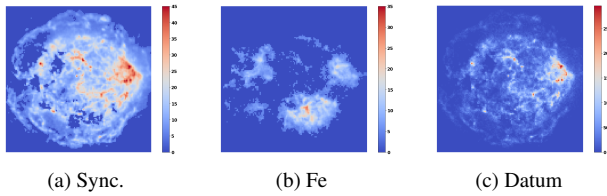


Fig. 3: Left: input synchrotron source. Middle: input iron (with emission line spectrum). Right: observation at 5548eV.

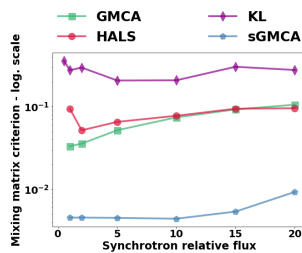


Fig. 4: Evolution of the mixing matrix criterion as a function of the synchrotron relative amplitude.

6. REFERENCES

- [1] P. Comon and C. Jutten, *Handbook of blind source separation*, Academic Press, 2010.
- [2] N. Gillis and F. Glineur, “Accelerated Multiplicative Updates and Hierarchical ALS Algorithms for Nonnegative Matrix Factorization,” *Neural Computation*, vol. 24, no. 4, pp. 1085–1105, 2012.
- [3] Jerome Bobin, Jeremy Rapin, Anthony Larue, and Jean-Luc Starck, “Sparsity and adaptivity for the blind separation of partially correlated sources,” *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1199–1213, 2015.
- [4] C. Villani, *Optimal Transport: Old and New*, Grundlehren Math. Wiss. 338, Springer, 2008.
- [5] M. Schmitz, M. Heitz, N. Bonneel, F. Ngolè, D. Coeurjolly, M. Cuturi, G. Peyré, and J.-L. Starck, “Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning,” *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 643–678, 2018.
- [6] Neal Parikh, Stephen Boyd, et al., “Proximal algorithms,” *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [7] M. Cuturi and A. Doucet, “Fast computation of wasserstein barycenters,” in *Proceedings of the 31st International Conference on Machine Learning*, 2014.
- [8] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, “Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion,” *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.
- [9] Jérôme Bobin, Jean-Luc Starck, Yassir Moudden, and Jalal Fadili, “Blind Source Separation : the Sparsity Revolution,” *Advances in Imaging and Electron Physics*, vol. 152, no. January 2008, pp. 1–82, 2008.
- [10] C. Kervazo, J. Bobin, and C. Chenot, “Heuristics for efficient sparse blind source separation,” in *8th International Conference on New Computational Methods for Inverse Problems 25 May 2018, Cachan, France*, 2018, vol. 1131.
- [11] A. Picquenot, F. Acero, J. Bobin, P. Maggi, J. Ballet, and G.W. Pratt, “Novel method for component separation of extended sources in x-ray astronomy,” *A&A - accepted - ArXiv:1905.10175*, 2019.
- [12] J.-L. Starck, J. Fadili, and F. Murtagh, “The undecimated wavelet decomposition and its reconstruction,” *IEEE Transactions on Image Processing*, vol. 16, pp. 297–309, 2007.