β-NMF AND SPARSITY PROMOTING REGULARIZATIONS FOR COMPLEX MIXTURE UNMIXING. APPLICATION TO 2D HSQC NMR

Afef Cherni, Sandrine Anthoine, Caroline Chaux

β-NMF AND SPARSITY PROMOTING REGULARIZATIONS FOR COMPLEX MIXTURE UNMIXING. APPLICATION TO 2D HSQC NMR.

Afef CHERNI, Sandrine ANTHOINE, Caroline CHAUX

Aix-Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France.

ABSTRACT

In Nuclear Magnetic Resonance (NMR) spectroscopy, an efficient analysis and a relevant extraction of different molecule properties from a given chemical mixture are important tasks, especially when processing bidimensional NMR data. To that end, using a blind source separation approach based on a variational formulation seems to be a good strategy. However, the poor resolution of NMR spectra and their large dimension require a new and modern blind source separation method. In this work, we propose a new variational formulation for blind source separation (BSS) based on a β-divergence data fidelity term combined with sparsity promoting regularization functions. An application to 2D HSQC NMR experiments illustrates the interest and the effectiveness of the proposed method whether in simulated or real cases.

Index Terms— BSS, sparsity, β-divergence, majorization-minimization (MM), multiplicative algorithm, 2D NMR.

1. INTRODUCTION

Blind Source Separation (BSS) consists in estimating N sources from M mixtures (in this work we consider $M > N$) without knowing the mixing operator. It was efficiently studied in many applications such as biology, chemistry, astronomy, telecommunications, etc. [1]. In this paper, we are interested in nuclear magnetic resonance (NMR) bidimensional data. It is worth noticing that NMR is a powerful tool used to characterize and determine properties of molecules present in a given chemical mixture. NMR 2D data are nonnegative and characterized by a high sparsity level presenting crowded spectra with an important spectral overlap and poor resolution (see Fig. 1). Thus, a robust BSS approach is still an open question in the 2D NMR context.

Nonnegative matrix factorization (NMF) introduced by Lee and Seung [2] constitutes one of the most popular approaches used to estimate nonnegative objects such as in audio source separation [3]. This concept was exploited in different applications either based on the classical Frobenius distance [4, 2] or based on the β-divergence family of cost functions [5, 6]. Moreover, different works showed that the Frobenius distance associated with regularization functions is an efficient framework enabling to solve the BSS problem. Recently, in [7] the Frobenius norm combined with various regularization functions was proposed and demonstrated its effectiveness to unmix complex NMR mixtures. In this work, we propose to investigate a β-NMF approach in which a β-divergence is associated with regularization functions that favour sparsity.

This paper is organized as follows: section 2 presents the problem statement and the adopted methodology based on an MM strategy. Section 3 is devoted to the proposed algorithm where multiplicative update rules are derived. The proposed algorithm is applied to blind source separation of 2D NMR spectra and results are presented in section 4. Some conclusions are finally drawn in section 5.

2. BLIND SOURCE SEPARATION

2.1. Problem statement

For N sources composed of L samples, represented by the matrix $\mathbf{S} \in \mathbb{R}^{N\times L}$, let us consider M mixtures represented by the matrix $\mathbf{X} \in \mathbb{R}^{M\times L}$ that follow the model

$$\mathbf{X} = \mathcal{D}(\mathbf{AS}), \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{M\times N}$ is a linear operator and $\mathcal{D}$ is the degradation model that depends on the application. The problem considered here is the estimation of $\mathbf{A}$ and $\mathbf{S}$ from $\mathbf{X}$. It corresponds to an ill-posed inverse problem. In the particular case of additive noise, model (1) can be re-written as

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

where $\mathbf{N} \in \mathbb{R}^{M\times L}$ corresponds to the acquisition noise. The aim of blind source separation is to identify jointly the sources $\mathbf{S}$ and the mixing matrix $\mathbf{A}$ from $\mathbf{X}$ in the presence of noise $\mathbf{N}$. One popular strategy aims to use a variational formulation in order to make the estimated product $\mathbf{AS}$ closed to the given mixtures $\mathbf{X}$. The corresponding mathematical formulation is given by

$$\underset{\mathbf{A},\,\mathbf{S}}{\text{minimize}} \;\; \Phi(\mathbf{A}, \mathbf{S}), \tag{3}$$

where Φ is the objective function. The choice of Φ is the main lock of our problem.

The authors would like to thank E. Piersanti, L. Shintu and M. Yemloul from iSm2, Aix-Marseille Univ., for their collaboration.

In the NMR application, the sources correspond to the Fourier Transform of a Free Induction Decay which explains the nonnegativity and the sparsity of $\mathbf{S}$. The mixing matrix $\mathbf{A}$ represents the concentration of each source $(\mathbf{S}_i)_{1\leq i\leq N}$ in the mixture $(\mathbf{X}_j)_{1\leq j\leq M}$, hence the nonnegativity of $\mathbf{A}$.

The NMF approach is a standard technique [2, 4] used to solve the BSS problem. It is given by

$$\underset{\mathbf{A},\mathbf{S}}{\text{minimize}}\ \ \Phi(\mathbf{A},\mathbf{S}) := \tfrac{1}{2}\|\mathbf{X}-\mathbf{A}\mathbf{S}\|_F^2+\lambda_{\mathbf{A}}\Psi_{\mathbf{A}}(\mathbf{A})+\lambda_{\mathbf{S}}\Psi_{\mathbf{S}}(\mathbf{S}), \tag{4}$$

where the fidelity term is based on the squared Frobenius norm and $\Psi_{\mathbf{A}}$ and $\Psi_{\mathbf{S}}$ are the regularization functions used to promote nonnegativity constraint on $\mathbf{A}$ and/or $\mathbf{S}$ or some other priors such as sparsity, energy, etc. $\lambda_{\mathbf{A}} \geq 0$ and $\lambda_{\mathbf{S}} \geq 0$ are the regularization parameters. This method has been successfully used in various works [2, 4]. However, it is related to the linear model given by Eq. (2). Another objective function based on the so-called β-divergence (noted β-**div**) was proposed in [8, 6] to ensure the blind source separation. It is defined as

$$\underset{\mathbf{A},\mathbf{S}}{\text{minimize}}\ \ \Phi(\mathbf{A},\mathbf{S}) := \beta\text{-}\mathbf{div}(\mathbf{X},\mathbf{A}\mathbf{S})+\lambda_{\mathbf{A}}\Psi_{\mathbf{A}}(\mathbf{A})+\lambda_{\mathbf{S}}\Psi_{\mathbf{S}}(\mathbf{S}). \tag{5}$$

2.2. Definition of β-divergence

The β-divergence is a measure of similarity between two objects u and v [9], denoted as β-*div* and defined for all $(u,v) \in (\mathbb{R}_+)^2$ and $\beta \in \mathbb{R}\setminus\{0,1\}$ as follows

$$\beta\text{-}div(u|v) = \tfrac{1}{\beta(\beta-1)}\left(u^\beta + (\beta-1)v^\beta - \beta u v^{\beta-1}\right). \tag{6}$$

An extension of Eq. (6) for $\beta = 1$ called the Kullback-Leibler divergence is defined as $\beta\text{-}div(u|v) = u\log(\frac{u}{v}) - u + v$ [10]. This divergence is commonly used when the problem statement involves a Poisson noise. When $\beta = 0$, the β-divergence is called the Itakura-Saito divergence, and is defined as $\beta\text{-}div(u|v) = \frac{u}{v} - \log(\frac{u}{v}) - 1$. This divergence was used especially in audio processing [5]. Note that the case of $\beta = 2$ in Eq. (6) corresponds to the Frobenius distance $\frac{1}{2}\|u-v\|^2$. The choice of β varies generally according to the context and the problem characteristics (type of noise for example). Let us define the β-divergence on $\mathbb{R}^L$ as a separable function β-**div**, that can be written for all $\mathbf{u} = (u_i)_{1\leq i\leq L} \in \mathbb{R}_+^L$ and $\mathbf{v} = (v_i)_{1\leq i\leq L} \in \mathbb{R}_+^L$ as

$$\beta\text{-}\mathbf{div}(\mathbf{u},\mathbf{v}) = \sum_{i=1}^{L} \beta\text{-}div(u_i|v_i). \tag{7}$$

2.3. Regularization function

Using a NMF strategy based on the squared Frobenius norm (Eq. (4)) or a β-divergence (Eq. (5)) with nonnegativity constraints, the solution of BSS problem is not unique: multiplying $\mathbf{S}$ by a diagonal or a permutation matrix $\Lambda \in \mathbb{R}^{M\times M}$ and $\mathbf{A}$ by its inverse defines new solutions ($\mathbf{S}' = \Lambda\mathbf{S}$ and $\mathbf{A}' = \mathbf{A}\Lambda^{-1}$). Therefore, adding an efficient prior information on $\mathbf{S}$ and $\mathbf{A}$ helps stabilizing the solution.

In the NMR application, only the nonnegativity constraint is applied on the mixing entries of $\mathbf{A}$. Therefore, the regularization on $\mathbf{A}$ can be defined as the indicator function of the nonnegative set denoted by ι_+ and given by

$$(\forall\, \mathbf{u} = (u_i)_{1\leq i\leq L}) \quad \iota_+(\mathbf{u}) = \begin{cases} 0 & \text{if } u_i \geq 0\ \forall i \\ +\infty & \text{otherwise.} \end{cases} \tag{8}$$

On the other hand, NMR spectra present not only nonnegative values but also a high sparsity level. Thus, the regularization function on $\mathbf{S}$ should promote the positivity and the sparsity of the solution. In the long history of sparse signal restoration, the ℓ_p norm defined for all $p \in]0,1]$ as

$$(\forall\, \mathbf{u} = (u_i)_{1\leq i\leq L} \in \mathbb{R}^L) \quad \ell_p(\mathbf{u}) = \left(\sum_{i=1}^{L} |u_i|^p\right)^{\frac{1}{p}} \tag{9}$$

is probably the most standard regularization function used to recover sparse signal especially with $p = 1$. It was used in various applications such as Compressive Sensing [11] and image restoration [12]. Recently, the Shannon entropy given for every $\mathbf{u} = (u_i)_{1\leq i\leq L} \in \mathbb{R}^L$ as $\mathbf{Ent}(\mathbf{u}) = \sum_{i=1}^{L} \mathrm{ent}(u_i)$ where

$$(\forall\, u \in \mathbb{R}) \quad \mathrm{ent}(u) = \begin{cases} u\log(u) & \text{if } u > 0 \\ 0 & \text{if } u = 0 \\ +\infty & \text{otherwise,} \end{cases} \tag{10}$$

was proposed in [13, 14] as a sparsity promoting penalty in the NMR context. In this paper, we propose to resolve Eq. (5) for $\beta > 2$ with the aforementioned regularization functions. In other words, we choose the β-divergence as the data fidelity term and the ℓ_1 or **Ent** function as the regularization term in addition to the nonnegativity constraint. Note that the choice of $\beta > 2$ is related to our optimization method.

3. ALTERNATING ALGORITHM FOR BSS

Many algorithms that adopt an alternating minimization procedure were proposed so as to solve the BSS problem such as ICA, SOBI, NMF, etc. A state-of-the-art of these algorithms used in the NMR context can be found in [15]. Based on this alternating procedure, we propose to solve Eq. (5) for $\beta > 2$ iteratively as presented in Algo. 1. To build a multiplicative

Algorithm 1 Generic alternating minimization strategy

For $k = 0, 1, \ldots$

$\mathbf{A}_{k+1} = \underset{\mathbf{A}}{\text{argmin}}\ \ \beta\text{-}\mathbf{div}(\mathbf{X},\mathbf{A}\mathbf{S}_k) + \lambda_{\mathbf{A}}\Psi_{\mathbf{A}}(\mathbf{A})$ **(I)**

$\mathbf{S}_{k+1} = \underset{\mathbf{S}}{\text{argmin}}\ \ \beta\text{-}\mathbf{div}(\mathbf{X},\mathbf{A}_{k+1}\mathbf{S}) + \lambda_{\mathbf{S}}\Psi_{\mathbf{S}}(\mathbf{S})$ **(II)**

algorithm that solves Eq. (5), a Majorization-Minimization

strategy [16] can be efficiently used. For that, the objective function Φ is split into the sum of a convex and a concave function, both terms being majorized independently. Let us mention that in order to majorize the convex term, Jensen's inequality will be applied, while the concave term will be locally majorized by its tangent. This strategy provides the following update rules for $\beta > 2$:

(**I**) We optimize $\mathbf{A}$ and we assume that $\mathbf{S}$ is fixed: choosing $\Psi_{\mathbf{A}} = \iota_+$, the update rule for the mixing matrix $\mathbf{A}$ will have the following expression

$$\mathbf{A}_{k+1} = \left(\frac{\left(\mathbf{X} \odot (\mathbf{A}_k\mathbf{S})^{\odot(\beta-2)}\right)\mathbf{S}^T}{(\mathbf{A}_k\mathbf{S})^{\odot(\beta-1)}\mathbf{S}^T} \right)_+^{\odot\frac{1}{\beta-1}} \odot \mathbf{A}_k. \quad (11)$$

We denote by $\odot$ the Hadamard product between two matrices (i.e. element-wise) and $(.)_+$ the projection onto the nonnegative set.

(**II**) We optimize $\mathbf{S}$ assuming that $\mathbf{A}$ is fixed: as mentioned previously, we propose to apply the nonnegativity constraint, the ℓ_1 norm and the **Ent** function as penalties on $\mathbf{S}$. Therefore, the update rule of $\mathbf{S}$ will be given as

a) $\Psi_{\mathbf{S}} = \iota_+$

$$\mathbf{S}_{k+1} = \left(\frac{\mathbf{A}^T(\mathbf{X} \odot (\mathbf{A}\mathbf{S}_k)^{\odot(\beta-2)})}{\mathbf{A}^T(\mathbf{A}\mathbf{S}_k)^{\odot(\beta-1)}} \right)_+^{\odot\frac{1}{\beta-1}} \odot \mathbf{S}_k \quad (12)$$

b) $\Psi_{\mathbf{S}} = \ell_1 + \iota_+$

$$\mathbf{S}_{k+1} = \left(\frac{\mathbf{A}^T(\mathbf{X} \odot (\mathbf{A}\mathbf{S}_k)^{\odot(\beta-2)}) - \lambda_{\mathbf{S}}}{\mathbf{A}^T(\mathbf{A}\mathbf{S}_k)^{\odot(\beta-1)}} \right)_+^{\odot\frac{1}{\beta-1}} \odot \mathbf{S}_k \quad (13)$$

c) $\Psi_{\mathbf{S}} = \mathbf{Ent} + \iota_+$

$$\mathbf{S}_{k+1} = \left(\frac{\gamma}{\alpha} \mathcal{W}\left(\frac{\alpha}{\gamma} \exp(-\frac{\delta}{\gamma}) \right) \right)_+^{\odot\frac{1}{\beta-1}} \odot \mathbf{S}_k \quad (14)$$

where $\mathcal{W}$ denotes the Lambert function [17], and

$\alpha = \mathbf{A}^T(\mathbf{A}\mathbf{S}_k)^{\odot(\beta-1)} \odot \mathbf{S}_k$,
$\gamma = \frac{\lambda_{\mathbf{S}}}{\beta-1}\mathbf{S}_k$,
$\delta = \lambda_{\mathbf{S}}(\mathbf{S}_k + \mathbf{S}_k \odot \log(\mathbf{S}_k)) - \mathbf{A}^T(\mathbf{X} \odot (\mathbf{A}\mathbf{S}_k)^{\odot(\beta-2)}) \odot \mathbf{S}_k$.

These updates are computed for $\beta > 2$. For some $\beta \leq 2$, solving Eq. (5) with ℓ_1 or **Ent** regularization function may be complicated and the MM strategy can not always be applied.

4. EXPERIMENTAL RESULTS

4.1. 2D HSQC NMR data

We process here 2D Heteronuclear Single Quantum Coherence (HSQC) data where 5 mixtures $\mathbf{X} \in \mathbb{R}^{5\times1024\times2048}$ and 4 pure sources (Limonene, Nerol, Terpinolene and Caryophyllene) noted $\mathbf{S} \in \mathbb{R}^{4\times1024\times2048}$ are acquired on a Bruker Avance III 600 MHz spectrometer. The real matrix $\mathbf{A}$ is given to us by the chemists who have acquired the data. We propose to apply the matricization technique of tensors. Therefore, we will have matrices $\mathbf{X} \in \mathbb{R}^{5\times2097152}$ and $\mathbf{S} \in \mathbb{R}^{4\times2097152}$.

In the synthetic case, we use the 4 sources and mixing matrix described above, we simulate the model (2) with a noise $\mathbf{N}$ assumed to be zero-mean Gaussian, i.i.d, with a known standard deviation $\sigma = 1.9713 \times 10^4$. We apply our algorithm to the obtained synthetic measures $\mathbf{X}$. In the real data case, we apply it to the measured data $\mathbf{X}$ described above.

4.2. Processing

We apply Algorithm 1 with the different updates proposed in Section 3 (Eq. (11) to (14)) to estimate the 4 sources from the 5 mixtures. The performances of the proposed approach are compared to the case when the popular data fidelity term ($\frac{1}{2}\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2$) is used. In this case, we implemented the Block-Coordinate Variable Metric Forward Backward (BC-VMFB) algorithm [18] as it was used in [15]. Moreover, both of these algorithms are initialized with a projection of the JADE [19] result onto the nonnegative space, and run for a maximum of 15000 iterations. The stopping criterion here is defined by $(\|\mathbf{S}_{k+1} - \mathbf{S}_k\|_F / \|\mathbf{S}_k\|_F) \leq 10^{-6}$ and $(\|\mathbf{A}_{k+1} - \mathbf{A}_k\|_F / \|\mathbf{A}_k\|_F) \leq 10^{-6}$.

To evaluate efficiently the quality of estimated sources $\mathbf{S}$, we use the SDR (Source to Distortion Ratio), SIR (Source to Interferences Ratio) and SAR (Source to Artefacts Ratio) [20] measures expressed in dB. Besides, we compute the Moreau-Amari index [21] to evaluate the estimation of the mixing matrix $\mathbf{A}$. Note that our results were obtained with Matlab R2018b running on Ubuntu 7.4.0-1.

4.3. Results

We present in Table 1 (simulated mixtures) the SDR, SIR and SAR computed for each source ($\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \mathbf{S}_4$) and the Amari-index for different objective functions Φ based on β-divergence ($\beta = 3$) and Frobenius norm, with various regularization parameters $\lambda_{\mathbf{S}}$. As we can see, when using Eq. (4), the highest SDR and SAR values are obtained with the ℓ_1 norm. However, it is the **Ent** regularization function that ensures almost the greatest SIR values with specific regularization parameter $\lambda_{\mathbf{S}}$. It is clear that the β-divergence improves SDR, SAR and SIR for both proposed regularization functions. Comparing the Amari-indexes, the best estimation of mixing matrix $\mathbf{A}$ corresponding to the smallest Amari-index is obtained using the β-divergence function. However, it seems that the choice of the regularization function $\Psi_{\mathbf{S}}$ and the regularization parameter $\lambda_{\mathbf{S}}$ still be related to each source $\mathbf{S}_i$ for $i = 1, ..., 4$ and should be adapted. Moreover, the NMF based on β-divergence presents an efficient choice of data fidelity term to solve the BSS problem in the 2D NMR.

Let us now turn our attention to the real case. Table 2 shows the average criterion on the 4 sources with the op-

Data fidelity term	$\lambda_{\mathbf{S}}$	$\Psi_{\mathbf{S}}$	SDR				SIR				SAR				Amari index
			$\mathbf{S}_1$	$\mathbf{S}_2$	$\mathbf{S}_3$	$\mathbf{S}_4$	$\mathbf{S}_1$	$\mathbf{S}_2$	$\mathbf{S}_3$	$\mathbf{S}_4$	$\mathbf{S}_1$	$\mathbf{S}_2$	$\mathbf{S}_3$	$\mathbf{S}_4$	
Squared Frobenius		ι_+	26.376	18.712	15.445	11.759	40.244	29.701	25.512	19.961	40.244	29.701	25.512	19.961	0.01210
	0.1σ	$\ell_1+\iota_+$	**33.530**	**34.235**	**25.526**	**27.906**	33.616	36.779	25.867	29.641	**50.577**	**37.769**	**36.766**	**32.735**	0.02722
		Ent $+\iota_+$	26.401	18.866	15.662	12.220	43.238	37.364	**33.223**	33.612	26.492	18.928	15.741	12.254	0.00901
	σ	$\ell_1+\iota_+$	32.406	19.773	16.850	15.531	33.159	20.530	17.293	16.171	40.391	27.771	27.063	24.262	0.04925
		Ent $+\iota_+$	25.732	18.141	14.826	10.633	39.248	**43.159**	30.675	34.554	25.931	18.155	14.944	10.652	0.01979
	10σ	$\ell_1+\iota_+$	32.544	15.969	10.304	9.3487	**63.139**	17.351	10.953	10.880	32.548	21.694	19.215	14.960	0.01892
		Ent $+\iota_+$	18.692	18.500	15.211	11.683	19.445	27.105	31.901	**44.047**	26.718	19.153	15.308	11.686	0.08606
β-divergence		ι_+	**52.074**	**38.083**	28.571	**28.115**	59.300	46.037	**28.933**	29.147	**59.300**	**46.037**	28.933	29.147	0.00543
	0.1σ	$\ell_1+\iota_+$	51.658	37.855	28.558	28.053	59.296	46.038	**28.933**	29.147	52.479	38.700	39.391	34.579	0.00543
		Ent $+\iota_+$	**52.074**	**38.083**	**28.574**	**28.115**	59.300	46.037	**28.933**	29.147	52.986	38.841	39.588	34.866	0.00543
	σ	$\ell_1+\iota_+$	42.803	32.294	27.523	25.547	59.235	46.171	28.929	29.138	42.903	32.475	33.110	28.050	0.00543
		Ent $+\iota_+$	52.071	**38.083**	**28.574**	**28.115**	59.297	46.003	**28.933**	29.214	52.983	38.841	**39.589**	**34.868**	0.00543
	10σ	$\ell_1+\iota_+$	33.682	22.722	19.549	15.759	55.521	**50.985**	28.764	**29.293**	33.711	22.729	20.007	15.961	0.00544
		Ent $+\iota_+$	52.055	38.036	28.570	28.091	**59.302**	46.025	28.932	29.146	52.962	38.788	39.543	34.760	0.00543

Table 1. SDR, SAR, SIR and Amari-index obtained in 2D simulated NMR spectra with various $\lambda_{\mathbf{S}}$.

timal regularization parameter $\lambda_{\mathbf{S}}$. The β-divergence combined with ℓ_1 norm or **Ent** function ensures the BSS of the 2D HSQC NMR data (see Fig. 1). However, compared with simulated data, we have a significant decrease of the SDR, SIR and SAR values which can probably be explained by a wrong assumption on $\mathcal{D}$ and possibly the linearity of the model. This raises the question about the choice of the objective function Φ and requires further investigations to characterize adequately the model in the 2D NMR context.

5. CONCLUSION

In this work, we have presented an NMF strategy based on a β-divergence to solve the BSS problem for 2D NMR spectroscopy data. We proposed to use the ℓ_1 and **Ent** regularization functions to favour the sparsity of the sources. For both sparsity promoting regularizations, our proposition ensures a good separation quality in the context of simulated data. However, in the real case even better results could be obtained. To do so, it will be very important to study the noise model $\mathcal{D}$ for 2D HSQC NMR data.

Data fidelity term	$\Psi_{\mathbf{S}}$	SDR	SIR	SAR	Amari-index
Squared Frobenius	$\ell_1+\iota_+$	04.984	13.956	07.951	0.18037
	Ent $+\iota_+$	05.755	14.434	08.446	0.17926
β-divergence	$\ell_1+\iota_+$	07.240	11.487	10.574	0.16098
	Ent $+\iota_+$	07.220	11.396	10.632	0.16526

Table 2. Average SDR, SIR and SAR on the 4 sources, with Amari-index computed for 2D real NMR data with optimal $\lambda_{\mathbf{S}}$ found empirically equal to 10σ.

6. ACKNOWLEDGEMENTS

The project leading to this publication has received funding from the Excellence Initiative of Aix-Marseille University - A*Midex, a French "Investissements d'Avenir" program.

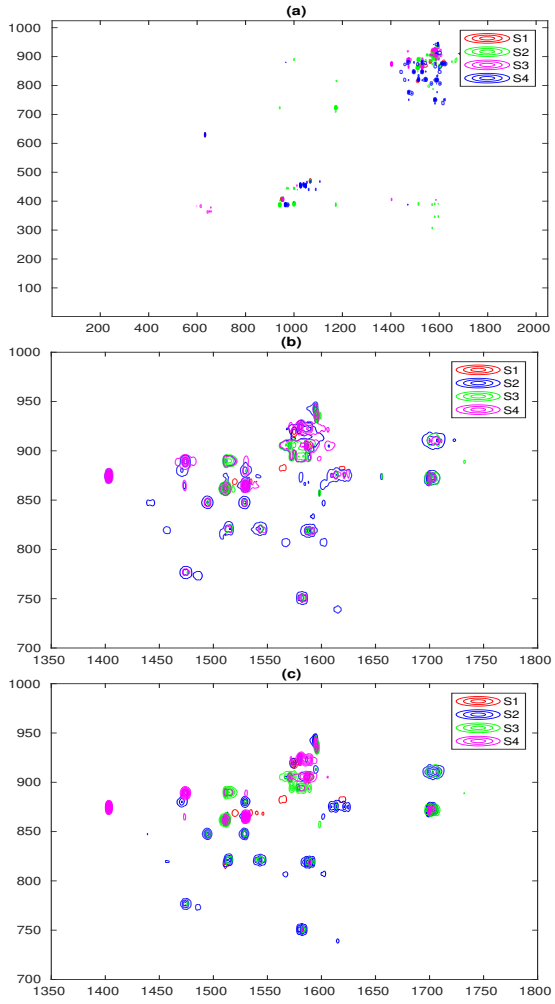

Fig. 1. 2D HSQC sources (Limonene (red), Nerol (blue), Terpinolene (magenta) and Caryophyllene (green)): pure sources (a), zoom on the most important terpene zone $[1300:1850, 700:1000]$ of the estimated sources: using Eq. (4) with ℓ_1 norm (b), or Eq. (5) with **Ent** regularization function (c).

7. REFERENCES

[1] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic press, 2010.

[2] D.-D Lee and H.-S Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788, 1999.

[3] E. Vincent, T. Virtanen, and S. Gannot, *Audio source separation and speech enhancement*, John Wiley & Sons, 2018.

[4] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.

[5] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.

[6] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

[7] A. Cherni, E. Piersanti, and C. Chaux, "Nmf-based sparse unmixing of complex mixtures," 2019.

[8] R. Kompass, "A generalized divergence measure for nonnegative matrix factorization," *Neural computation*, vol. 19, no. 3, pp. 780–791, 2007.

[9] A. Basu, I.-R Harris, N.-L Hjort, and M.-C Jones, "Robust and efficient estimation by minimising a density power divergence," *Biometrika*, vol. 85, no. 3, pp. 549–559, 1998.

[10] T. Hoffman, "Probabilistic latent semantic indexing," in *Proceedings of the 22nd Annual ACM Conference on Research and Development in Information Retrieval*, 1999, pp. 50–57.

[11] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[12] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[13] A. Cherni, E. Chouzenoux, and M.-A Delsuc, "Proximity operators for a class of hybrid sparsity+ entropy priors application to dosy nmr signal reconstruction," in *2016 International Symposium on Signal, Image, Video and Communications (ISIVC)*. IEEE, 2016, pp. 120–125.

[14] A. Cherni, E. Chouzenoux, and M.-A Delsuc, "PALMA, an improved algorithm for dosy signal processing," *Analyst*, vol. 142, no. 5, pp. 772–779, 2017.

[15] A. Cherni, E. Piersanti, S. Anthoine, C. Chaux, L. Shintu, M. Yemloul, and B. Torrésani, "Challenges in the decomposition of 2d nmr spectra of mixtures of small molecules," *Faraday discussions*, vol. 218, pp. 459–480, 2019.

[16] D.-R Hunter and K. Lange, "Rejoinder," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 52–59, 2000.

[17] R.-M. Corless, G.-H. Gonnet, D.-E. Hare, D.-J. Jeffrey, and D.-E. Knuth, "On the lambert W function," *Advances in Applied Mathematics*, vol. 5, no. 1, pp. 329–359, 1996.

[18] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function," *Journal of Optimization Theory and Applications*, vol. 162, no. 1, pp. 107–132, 2014.

[19] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proceedings F - Radar and Signal Processing*, vol. 140, no. 6, pp. 362–370, 1993.

[20] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE transactions on audio, speech, and language processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

[21] E. Moreau and O. Macchi, "A one stage self-adaptive algorithm for source separation," in *Proceedings of ICASSP 94. IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 1994, vol. 3, pp. III–49.