Non-Negative Blind Source Separation Algorithm Based on Minimum Aperture Simplicial Cone

Wendyam S. B. Ouedraogo, Antoine Souloumiac, Mériem Jaidane, Christian Jutten

Non-negative Blind Source Separation Algorithm based on Minimum Aperture Simplicial Cone

Wendyam Serge Boris Ouedraogo*, Antoine Souloumiac, Meriem Jaidane, and Christian Jutten, *Fellow, IEEE*

***Abstract*—We address the problem of Blind Source Separation (BSS) when the hidden sources are Nonnegative (N-BSS). In this case, the scatter plot of the mixed data is contained within the simplicial cone generated by the columns of the mixing matrix. The proposed method, termed SCSA-UNS for Simplicial Cone Shrinking Algorithm for Unmixing Non-negative Sources, aims at estimating the mixing matrix and the sources by fitting a Minimum Aperture Simplicial Cone (MASC) to the cloud of mixed data points. SCSA-UNS is evaluated on both independent and correlated synthetic data and compared to other N-BSS methods. Simulations are also performed on real Liquid Chromatography-Mass Spectrum (LC-MS) data for the metabolomic analysis of a chemical sample, and on real dynamic Positron Emission Tomography (PET) images, in order to study the pharmacokinetics of the [18F]-FDG (FluoroDeoxyGlucose) tracer in the brain.**

***Index Terms*—Blind Source Separation, Non-negativity, Simplicial Cone, Aperture, LC-MS, dynamic PET imaging**

I. Introduction

THE well-known problem of Non-negative Blind Source Separation (N-BSS) occurs in many situations of signal and image processing. Solving the N-BSS problem consists in retrieving the hidden non-negative sources and their corresponding mixing profiles from the measured data (also denoted by observations), which are mixtures of the original sources. The use of N-BSS on actual data has continuously increased during the last decades. Applications include chemometrics [51] [45], biomedical image processing [27] [43], spectrometry and spectroscopy [5] [30], remote sensing [29] [18], speech and music power spectra decomposition [47] [15], to name a few. The interested reader can refer to [40] and [11] for more details. The batch noiseless linear instantaneous mixture model of the N-BSS problem is given by:

$$\mathbf{X} = \mathbf{AS} \tag{1}$$

where $\mathbf{X} \in \mathbb{R}_+^{m\times K}$ is the known non-negative matrix of observations, $\mathbf{A} \in \mathbb{R}_+^{m\times n}$ is the unkown non-negative mixing matrix and $\mathbf{S} \in \mathbb{R}_+^{n\times K}$ is the hidden non-negative sources matrix. Integers m, n and K are respectively the number of observations, the number of sources and the number of samples. The scalar $x_i(k)$, $1 \leq i \leq m$ and $1 \leq k \leq K$, is the k-th sample of the observation vector $\mathbf{x}_i$ of size $m \times 1$. The scalar $s_j(k)$, $1 \leq j \leq n$, is the k-th sample of source vector $\mathbf{s}_j$ of size $n \times 1$, and the scalar a_{ij} is the ij-th entry of the mixing matrix $\mathbf{A}$. The task of solving problem (1) consists in recovering the sources $\mathbf{S}$ and their profiles $\mathbf{A}$, given only the observations $\mathbf{X}$, without any prior knowledge on $\mathbf{S}$ and $\mathbf{A}$ except their non-negativity.

Several methods were proposed for solving problem (1). One can distinguish between statistical approaches and deterministic ones.

The methods of the first class use the statistical properties of the sources in order to design the separating algorithms. They include Non-negative Independent Components Analysis (N-ICA) [37] and Bayesian Inference based Non-negative Source Separation (BI-NSS) [20] [30]. N-ICA assumes that the *non-negative* sources, $\mathbf{s}_1$, $\mathbf{s}_2$,$\cdots$, $\mathbf{s}_n$ are random variables, which are mutually statistically *independent* and *well-grounded*[1] [36]. The sources and the mixing matrix are estimated by whitening the observations and by rotating the whitened data to make them non-negative [37] [39]. Unfortunately, N-ICA is not suited to correlated sources, as we will see in the simulations. On the other hand, in the Bayesian approach [20] [30], the sources and the mixing profiles (that is columns of the mixing matrix) are modeled by random variables. The idea is to assign a specific prior probability density to each variable, and to derive the joint posterior probability density. The sources and mixing profiles are obtained by an a posterior estimator (posterior mean or posterior maximum) using Bayes' rule and the Markov Chain Monte Carlo (MCMC) sampling. In the case of non-negative sources, BI-NSS uses a non-negative prior probability density (such as Gamma density) for both sources and profiles. This method can unfortunately be computationally complex and time-consuming, especially for large-scale data.

In order to solve the N-BSS problem, deterministic approaches impose only weak assumptions on the sources distribution. In particular, the independence of the sources is not necessary. These approaches include Non-negative Matrix Factorization (NMF) and geometrical methods. Firstly introduced by Paatero and Tapper [34], NMF was popularized by Lee and Seung [25] and became the prevalent method for solving the N-BSS problem. In NMF, $\mathbf{A}$ and $\mathbf{S}$ are estimated by minimizing a divergence measure between the

Manuscript received December 7, 2012; revised June 13, 2013; Accepted August 22, 2013.

Wendyam Serge Boris Ouedraogo is with the GIPSA-lab/DIS, 11 rue des Mathématiques, BP 46, F-38402 Saint Martin d'Hères Cedex-France, (e-mail: wendyam.ouedraogo@gipsa-lab.grenoble-inp.fr).

Antoine Souloumiac is with CEA LIST/DCSI, Bat. 565, PC 192, F-91191 Gif-sur-Yvette Cedex France, (e-mail: antoine.souloumiac@cea.fr).

Meriem Jaidane is with ENIT/U2S, BP 37, Le Belvedere 1002 Tunis, Tunisie, (e-mail: meriem.jaidane@planet.tn).

Christian Jutten is with the GIPSA-lab/DIS, 11 rue des Mathématiques, BP 46, F-38402 Saint Martin d'Hères Cedex-France, and with Institut Universitaire de France, (e-mail: christian.jutten@gipsa-lab.grenoble-inp.fr).

[1]A ***well-grounded*** random variable $\mathbf{s}_i$, has a nonvanishing probability density function around zero, i.e. $\forall\ \delta > 0$, $Pr(s_i < \delta) > 0$

left and right parts of equation (1) under a non-negativity constraint on both the estimated mixing matrix and sources. See [12] for more details on different divergence measures and optimization algorithms for NMF. Nevertheless, the non-negativity constraint alone (on $\mathbf{S}$ and $\mathbf{A}$) is not sufficient to guarantee the uniqueness of the factorization [14] [31] [23]. Therefore, according to the prior knowledge about the sources (respectively the profiles), some additional constraints, such as the sparsity and/or the smoothness of the sources and/or the mixing profiles, were incorporated in NMF to improve the physical meaning of the estimated sources and/or mixing profiles, and reduce the number of possible solutions [17] [9]. A novel direction has recently been introduced in order to tackle the non-uniqueness of NMF, by adding a minimum determinant constraint on the estimated mixing matrix [22] [44]. The geometrical methods are the other deterministic approaches for solving the Non-negative Blind Source Separation problem. The first geometrical method was introduced by Puntonet et al. [41] for unmixing two sources having bounded probability densities. The mixing matrix is estimated by finding the slopes of the parallelogram containing the scatter plot of mixed data. Babaie-Zadeh et al. [2] have proposed another geometrical method for separating sparse sources. The latter method, which can deal with more than two sources, estimates the mixing matrix by clustering the scatter plot of mixed data and fitting a line (for dimension 2) or hyper-plane (for dimensions greater than 2) to each cluster. Other geometric methods were proposed for sources satisfying certain properties. If for every source there is at least one instance during which the underlying source is active and all the others are not (i.e. $\forall\ 1 \leq i \leq n$, $\exists\ k_i\ (1 \leq k_i \leq K)$ such that $s_i(k_i) \neq 0$ and $s_j(k_i) = 0$ for $j \neq i$), then the sources are termed *locally dominant*. In that case, $\mathbf{A}$ can be estimated by finding the vertices of the convex hull of the scatter plot of mixed data [24] [7] [32]. Unfortunately, the *local dominance* assumption (also termed the pure pixel assumption, in the hyperspectral unmixing) is seldom verified in practice. If the sum on every column of the sources matrix equal to one (i.e. $\forall\ 1 \leq k \leq K,\ \sum_{j=1}^{n} s_j(k) = 1$), then the sources are termed *full additive*, and one can estimate the mixing matrix by looking for the Minimum Volume Simplex (MVS) containing the scatter plot of mixed data [13] [28] [3] [6]. The MVS methods do not require the local dominance of the sources and, in the noiseless case, one can relax the full additivity constraint by normalizing each column of the data matrix to a unit sum as shown in [8] and [16]. However, in the noisy case, enforcing this normalization may amplify the noise and/or lead to a bad estimation of the sources, especially if the number of sources is overestimated, as illustrated in Fig.11.

This paper proposes a new geometrical method for solving the overdetermined N-BSS problem. The proposed method, denoted Simplicial Cone Shrinking Algorithm for Unmixing Non-negative Sources (SCSA-UNS), estimates the mixing matrix and the sources by finding the "Minimum Aperture Simplicial Cone" (MASC) containing the scatter plot of the mixed data. It neither requires the independence of the sources, nor their local dominance, or even their full additivity.

The paper is organized as follows. In section II, we review the geometrical foundations of the N-BSS problem and we derive the Minimum Aperture Simplicial Cone based approach. Section III gives conditions on the sources for the uniqueness of N-BSS, and section IV describes the proposed method. In section V, we proposed a preprocessing for reducing the noise effect. Section VI and section VII present simulation results on synthetic and real data. Finally Section VIII presents our conclusions.

II. Geometrical foundations of the Non-negative Blind Source Separation problem

We restrict to the case where the mixture is determined (i.e. $m = n$). If the mixture is overdetermined ($m > n$), one can reduce to the previous case by a proper dimension reduction, as described in section IV-B.

A. Useful definitions and concepts

In the following, $\mathbb{R}_+^n$ denotes the set of non-negative n-dimensional vectors, and $\mathbb{R}_+^{n\times n}$ denotes the of n-size matrices whose all entries are non-negative.
Given a full rank matrix $\mathbf{U} \in \mathbb{R}_+^{n\times n}$ ($\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n]$, where $\mathbf{u}_i \in \mathbb{R}_+^n$ is the ith column of $\mathbf{U}$), we introduce the following definitions:

Definition 1: Simplicial Cone
The *Simplicial Cone* generated by the columns of $\mathbf{U}$, denoted by $\mathbf{Span}^+(\mathbf{U})$, is defined as:

$$\mathbf{Span}^+(\mathbf{U}) = \left\{\mathbf{z} \,\middle|\, \mathbf{z} = \mathbf{U}\mathbf{y} \text{ with } \mathbf{y} \in \mathbb{R}_+^n\right\} \tag{2}$$

Definition 2: Edge Vector of a *Simplicial Cone*
By abusing the notation, the i-th *Edge Vector* of the simplicial cone $\mathbf{Span}^+(\mathbf{U})$, denoted by $\mathcal{E}_i(\mathbf{Span}^+(\mathbf{U}))$, is defined as:

$$\mathcal{E}_i(\mathbf{Span}^+(\mathbf{U})) = \{\mathbf{z} \mid \mathbf{z} = \alpha\mathbf{u}_i \mid \alpha \geq 0\} \tag{3}$$

Definition 3: Facet of a *Simplicial Cone*
In the same way, the i-th *Facet* of the simplicial cone $\mathbf{Span}^+(\mathbf{U})$, denoted by $\mathcal{F}_i(\mathbf{Span}^+(\mathbf{U}))$, is defined as:

$$\mathcal{F}_i(\mathbf{Span}^+(\mathbf{U})) = \mathbf{Span}^+\left\{\mathbf{U}\backslash\{\mathbf{u}_i\}\right\} \tag{4}$$

Definition 4: Vertex of a Simplicial Cone
The ***Vertex*** of the simplicial cone $\mathbf{Span}^+(\mathbf{U})$, denoted by $\mathcal{V}(\mathbf{Span}^+(\mathbf{U}))$, is defined as:

$$\mathcal{V}(\mathbf{Span}^+(\mathbf{U})) = \bigcap_i \mathcal{E}_i(\mathbf{Span}^+(\mathbf{U})) \tag{5}$$

Remarks :

- For ease of the notations, we set $\mathcal{E}_i(\mathbf{U}) = \mathcal{E}_i(\mathbf{Span}^+(\mathbf{U}))$, $\mathcal{F}_i(\mathbf{U}) = \mathcal{F}_i(\mathbf{Span}^+(\mathbf{U}))$, and $\mathcal{V}(\mathbf{U}) = \mathcal{V}(\mathbf{Span}^+(\mathbf{U}))$
- The simplicial cone, $\mathbf{Span}^+(\mathbf{U})$, has n edge vectors, n facets and 1 vertex.
- The vertex of $\mathbf{Span}^+(\mathbf{U})$ is located at the origin of $\mathbb{R}^n$
- $\mathbb{R}_+^n$, the ***positive orthant***, is the simplicial cone generated by the identity matrix $\mathbf{I}_n$: $\mathbb{R}_+^n = \mathbf{Span}^+(\mathbf{I}_n)$

Given the previous definitions, one can infer the following:

Lemma 1: The scatter plot of the sources $\{\mathbf{s}(k), 1 \leq k \leq K\}$, with $\mathbf{s}(k) = [s_1(k)\ s_2(k)\ \cdots\ s_n(k)]^T$, is contained in the positive orthant, $\mathbf{Span}^+(\mathbf{I}_n)$:

$$\{\mathbf{s}(k), 1 \leq k \leq K\} \subseteq \mathbf{Span}^+(\mathbf{I}_n) \tag{6}$$

Proof: Follow the definition of a simplicial cone. ■

Lemma 2: The scatter plot of the mixed data $\{\mathbf{x}(k), 1 \leq k \leq K\}$, with $\mathbf{x}(k) = [x_1(k)\ x_2(k)\ \cdots\ x_n(k)]^T$, is contained in $\mathbf{Span}^+(\mathbf{A})$, the simplicial cone generated by the columns of the mixing matrix:

$$\{\mathbf{x}(k), 1 \leq k \leq K\} \subseteq \mathbf{Span}^+(\mathbf{A}) \tag{7}$$

Proof: Follow the definition of a simplicial cone. ■

Since $\mathbf{A}$ is also non-negative, one can deduce that:

$$\mathbf{Span}^+(\mathbf{A}) \subseteq \mathbf{Span}^+(\mathbf{I}_n) \tag{8}$$

To illustrate Lemma 2, for $m = n = 3$, we consider the mixing matrix $\mathbf{A}$ given below, and we generate data according to equation (1), where each source is generated following the uniform distribution between 0 and 1.

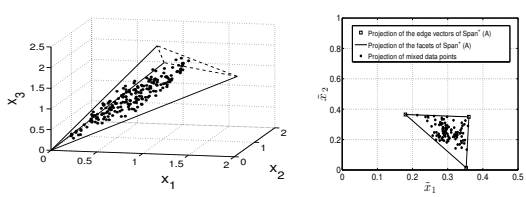
$$\mathbf{A} = \begin{pmatrix} 0.67 & 0.49 & 0.32 \\ 0.65 & 0.02 & 0.65 \\ 0.54 & 0.88 & 0.81 \end{pmatrix}$$

Fig. 1(a) shows the scatter plot of mixed data included in the simplicial cone generated by $\mathbf{A}$. For ease of visualization, we represent on Fig. 1(b) the projection of $\mathbf{Span}^+(\mathbf{A})$ and the mixed data, on the plane $\mathcal{P}$ defined by:

$$\mathcal{P} = \left\{ \mathbf{z} = [z_1, z_2, z_3]^T \in \mathbb{R}^3 \mid \sum_{i=1}^{3} z_i = 1 \right\} \tag{9}$$

The projected data $\tilde{\mathbf{X}}$, and cone $\mathbf{Span}^+(\tilde{\mathbf{A}})$ are obtained by:

$$\tilde{x}_i(k) = x_i(k) / \sum_{l=1}^{3} x_l(k) \text{ and } \tilde{a}_{ij} = a_{ij} / \sum_{l=1}^{3} a_{il} \tag{10}$$



(a) $\{\mathbf{X}(k)\} \subseteq \mathbf{Span}^+(\mathbf{A}_1)$ (b) Projection on $\mathcal{P}$

Fig. 1. Scatter plot of mixed data included in $\mathbf{Span}^+(\mathbf{A})$

Theorem 1: Given two full rank non-negative matrices $\mathbf{U} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{V} \in \mathbb{R}_+^{n\times n}$:

$$\mathbf{Span}^+(\mathbf{U}) \subseteq \mathbf{Span}^+(\mathbf{V}) \Leftrightarrow \mathbf{U} = \mathbf{VM} \quad \text{where } \mathbf{M} \in \mathbb{R}_+^{n\times n} \text{ and } rg(\mathbf{M}) = n \tag{11}$$

Proof: Proof of Theorem 1 is given in Appendix A. ■

Theorem 2: Given two full rank non-negative matrices $\mathbf{U} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{V} \in \mathbb{R}_+^{n\times n}$:

$$\mathbf{Span}^+(\mathbf{U}) = \mathbf{Span}^+(\mathbf{V}) \Leftrightarrow \mathbf{U} = \mathbf{VM} \quad \text{where } \mathbf{M} = \mathbf{DP} \text{ and } rg(\mathbf{M}) = n \tag{12}$$

$\mathbf{D} \in \mathbb{R}_+^{n\times n}$ is a diagonal matrix, and $\mathbf{P} \in \mathbb{R}_+^{n\times n}$ is a permutation one. $\mathbf{M}$ is called a monomial matrix.

Proof: Proof of Theorem 2 is given in Appendix B. ■

B. Minimum Aperture Simplicial Cone based N-BSS

According to equations (7) and (8), one can estimate the mixing matrix (and the sources) by looking for a simplicial cone containing the scatter plot of the mixed data and contained in the positive orthant. But without any additional constraint, there are infinite number of such cones [14] [23] (the positive orthant itself is an example). By assuming that the sources satisfies the full additivity constraint (see Section I), some geometrical methods [13] [28] [3] [6], look for the Minimum Volume Simplex (MVS) containing the scatter plot of mixed data. In this paper, except their non-negativity, no other assumption is made on the sources. In particular, the sources and the mixed data do not have upper bounds, and the volume of a simplicial cone containing the scatter plot of mixed data is undefined. One can normalize each column of the data matrix to a unit sum, and consider the volume of the simplex containing the normalized data, however, as we mentioned in Section I above, enforcing this normalization may amplify the noise and/or lead to a poor estimation of the sources. Therefore, instead of using the volume, we characterize a simplicial cone by its *Aperture* defined below:

Definition 5: Aperture of a Simplicial Cone

The *Aperture* of the given simplicial cone, $\mathbf{Span}^+(\mathbf{U})$, generated by a full rank matrix $\mathbf{U}$, denoted $\mathcal{A}\left(\mathbf{Span}^+(\mathbf{U})\right)$, is defined by:

$$\mathcal{A}\left(\mathbf{Span}^+(\mathbf{U})\right) = \frac{|\det(\mathbf{U})|}{\|\mathbf{u}_1\|_2 \times \|\mathbf{u}_2\|_2 \times \cdots \times \|\mathbf{u}_n\|_2} \tag{13}$$

Remarks :

- For ease of notation, we set: $\mathcal{A}(\mathbf{U}) = \mathcal{A}\left(\mathbf{Span}^+(\mathbf{U})\right)$
- For $n = 2$, $\mathcal{A}(\mathbf{U}) = \sin(\theta)$, where θ is the angle between the two column vectors of $\mathbf{U}$. For $n > 2$, $\mathcal{A}(\mathbf{U}) = \prod_{i=1}^{n-1} \sin(\theta_i)$, where θ_i is the principal angle between $\mathcal{E}_{i+1}(\mathbf{U})$ and $\mathbf{Span}^+\{\mathbf{u}_1, \cdots, \mathbf{u}_i\}$.
- According to Hadamard's inequality which states that $|\det(\mathbf{U})| \leq \prod_{i=1}^{n} \|\mathbf{u}_i\|_2$, we have $0 \leq \mathcal{A}(\mathbf{U}) \leq 1$.
- $\mathcal{A}(\mathbf{I}_n) = 1$

Theorem 3: Given two full rank non-negative matrices $\mathbf{U} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{V} \in \mathbb{R}_+^{n\times n}$:

$$\mathbf{Span}^+(\mathbf{U}) \subseteq \mathbf{Span}^+(\mathbf{V}) \Rightarrow \mathcal{A}(\mathbf{U}) \leq \mathcal{A}(\mathbf{V}) \tag{14}$$

Proof: Proof of Theorem 3 is given in Appendix C. ■

Theorem 4: Given two full rank non-negative matrices $\mathbf{U} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{V} \in \mathbb{R}_+^{n\times n}$:

$$\mathbf{Span}^+(\mathbf{U}) = \mathbf{Span}^+(\mathbf{V}) \Rightarrow \mathcal{A}(\mathbf{U}) = \mathcal{A}(\mathbf{V}) \tag{15}$$

Proof: Proof of Theorem 4 directly comes from Theorem 2 and the definition of the *Aperture* of a simplicial cone. ■

Definition 6: Minimum Aperture Simplicial Cone

We state that $\mathbf{Span}^+(\mathbf{U})$ is a *Minimum Aperture Simplicial*

Cone (MASC) containing the scatter plot of mixed data $\{\mathbf{x}(k), 1 \leq k \leq K\}$, if for any simplicial cone $\mathbf{Span}^+(\mathbf{V})$ containing the scatter plot of mixed data, $\mathcal{A}(\mathbf{U}) \leq \mathcal{A}(\mathbf{V})$.
By misnomer, we also state that $\mathbf{Span}^+(\mathbf{U})$ is a non-negative MASC containing the scatter plot of mixed data, if $\mathbf{Span}^+(\mathbf{U})$ is a MASC and $\mathbf{Span}^+(\mathbf{U}) \subseteq \mathbf{Span}^+(\mathbf{I}_n)$.

Under certain conditions on the sources (which will be discussed in section III), $\mathbf{Span}^+(\mathbf{A})$ is the unique non-negative *Minimum Aperture Simplicial Cone* containing the scatter plot of the mixed data. In this case the mixing matrix can be estimated (up to positive scaling and permutation indeterminations) by finding this simplicial cone. In fact, if $\mathbf{Span}^+(\mathbf{W})$ is another non-negative MASC containing the scatter plot of mixed data, then $\mathbf{Span}^+(\mathbf{W}) = \mathbf{Span}^+(\mathbf{A})$, and therefore $\mathbf{W} = \mathbf{ADP}$, where $\mathbf{D}$ is a diagonal matrix with non-negative entries, and $\mathbf{P}$ is a permutation matrix (Theorem 2).

Before proposing an algorithm for estimating the mixing matrix, we give below conditions on the sources, under which $\mathbf{Span}^+(\mathbf{A})$ is the unique non-negative *Minimum Aperture Simplicial Cone* containing the scatter plot of the mixed data.

III. Conditions on the sources for unique N-BSS by Minimum Aperture Simplicial Cone

To ensure recovering the true mixing matrix, and then the true sources in noiseless case, $\mathbf{Span}^+(\mathbf{A})$ must be the unique non-negative MASC containing the scatter plot of mixed data. The following Theorem 5 transforms this condition into a condition on the sources.

Theorem 5: $\mathbf{Span}^+(\mathbf{A})$ is the unique non-negative MASC containing the scatter plot of the mixed data if and only if $\mathbf{Span}^+(\mathbf{I}_n)$ is the unique non-negative MASC containing the scatter plot of the sources.

Proof: Proof of Theorem 5 is given in Appendix D. ■

A. Necessary condition for unique N-BSS by MASC

Proposition 1: If $\mathbf{Span}^+(\mathbf{I}_n)$ is the unique non-negative MASC containing the scatter plot of the sources, then there is at least one point of the cloud of sources on each facet of $\mathbf{Span}^+(\mathbf{I}_n)$, i.e. $\forall\, 1 \leq i \leq n$, $\exists\, k_i$ such as $s_i(k_i) = 0$

Proof: Assume that there is at least one facet of $\mathbf{Span}^+(\mathbf{I}_n)$ where there is no point of the cloud of sources, i.e $\exists\, 1 \leq i \leq n \mid \forall\, 1 \leq k \leq K,\ s_i(k) > 0$ (Without loss of generality, set $i = 1$ for the demonstration). Let's define the n-size square matrix $\mathbf{U} = [u_{ij}]_{1 \leq i,j \leq n}$ as follow:

$$u_{ij} = \begin{cases} \min\limits_{1 \leq k \leq K} \frac{s_1(k)}{s_2(k)}, s_2(k) \neq 0 & \text{if } i = 1 \text{ and } j = 2 \\ \delta_{ij} & \text{otherwise} \end{cases} \tag{16}$$

Since the sources are non-negative, $u_{12} > 0$ and one can easily verify that $\mathbf{U}^{-1}\mathbf{S} \geq 0$. It comes that $\mathbf{Span}^+(\mathbf{U})$ is a simplicial cone containing the scatter plot of the sources. Since $u_{12} > 0$, then $\mathbf{Span}^+(\mathbf{U}) \subset \mathbf{Span}^+(\mathbf{I}_n)$, hence $\mathcal{A}(\mathbf{U}) < \mathcal{A}(\mathbf{I}_n)$. Therefore $\mathbf{Span}^+(\mathbf{I}_n)$ is not the unique non-negative MASC containing the scatter plot of sources. ■

B. Sufficient conditions for unique N-BSS by MASC

1) Local dominance:

Proposition 2: If the sources are non-negative and *locally dominant* (see section I), then $\mathbf{Span}^+(\mathbf{I}_n)$ is the unique non-negative MASC containing the scatter plot of sources.

Proof: Since sources are non-negative and locally dominant, i.e. $\forall\ 1 \leq i \leq n$, $\exists\ 1 \leq k_i \leq K \mid s_i(k_i) \neq 0$ and $s_j(k_i) = 0$ for $j \neq i$, then the k_ith column of $\mathbf{S}$ is $\mathbf{s}(k_i) = s_i(k_i)\mathbf{e}_i$ where $\mathbf{e}_i$ is the ith column of $\mathbf{I}_n$.
Let $\mathbf{Span}^+(\mathbf{U})$ be a non-negative simplicial cone containing the scatter plot of sources, i.e. $\forall\, 1 \leq i \leq n$, $\mathbf{s}(k_i) \in \mathbf{Span}^+(\mathbf{U})$ then $\mathbf{e}_i \in \mathbf{Span}^+(\mathbf{U})$, so $\mathbf{Span}^+(\mathbf{I}_n) \subseteq \mathbf{Span}^+(\mathbf{U})$ and consequently $\mathcal{A}(\mathbf{I}_n) \leq \mathcal{A}(\mathbf{U})$. Therefore, $\mathbf{Span}^+(\mathbf{I}_n)$ is a non-negative MASC containing the scatter plot of the sources.
If $\mathbf{Span}^+(\mathbf{V})$ is another non-negative MASC containing the scatter plot of the sources, then $\mathbf{V} \geq 0$, $\mathbf{V}^{-1}\mathbf{S} \geq 0$ and $\mathcal{A}(\mathbf{V}) = \mathcal{A}(\mathbf{I}_n) = 1$, which lead us to conclude that $\mathbf{V} = \mathbf{I}_n\mathbf{M}$ where $\mathbf{M}$ is a monomial matrix, and consequently $\mathbf{Span}^+(\mathbf{I}_n)$ is the unique MASC containing the scatter plot of the sources. ■

2) Independence and well-grounded:

Proposition 3: If the sources are *non-negative*, *independent* and *well-grounded*, then asymptotically, $\mathbf{Span}^+(\mathbf{I}_n)$ tends to the unique non-negative MASC containing the scatter plot of the sources.

Proof: On the one hand the sources are well-grounded, then $\forall\ 1 \leq i \leq n$, $\forall\ \delta > 0$, $Pr(\mathbf{s}_i < \delta) > 0$. On the other hand, the sources are non-negative and independent, then for a fixed i:

$$Pr(\mathbf{s}_1 < \delta, \cdots, \mathbf{s}_i > 0, \cdots, \mathbf{s}_n < \delta) = Pr(\mathbf{s}_i > 0) \prod_{j=1, j \neq i}^{n} Pr(\mathbf{s}_j < \delta) \tag{17}$$

By making $\delta \longrightarrow 0$, it follows that asymptotically, there will at least one index k_i such as $s_i(k_i) > 0$ and $s_j(k_i) \longrightarrow 0$ for all $j \neq i$. Therefore, asymptotically, the sources tend to be locally dominant. Hence, asymptotically $\mathbf{Span}^+(\mathbf{I}_n)$ tends to be the unique non-negative MASC containing the scatter plot of the source. ■

3) Sufficiently spread condition:

Proposition 4: If for each facet of $\mathbf{Span}^+(\mathbf{I}_n)$, at least $n-1$ points of the scatter plot of the sources belong to underlined facet, and the vectors corresponding to these points are linearly independent, then $\mathbf{Span}^+(\mathbf{I}_n)$ is the unique non-negative MASC containing the scatter plot of the sources.

Proof: Let $\mathbf{Span}^+(\mathbf{V})$ be a non-negative MASC containing the scatter plot of the sources, then $\mathbf{V} \geq 0$, $\mathbf{V}^{-1}\mathbf{S} \geq 0$ and $\mathcal{A}(\mathbf{V}) \leq 1$. The task is to demonstrate that $\mathbf{V}$ is a monomial matrix. For a fixed i, assume that $\mathbf{s}(k_l^i)$, $l = 1, 2, \cdots, n-1$ belongs to $\mathcal{F}_i(\mathbf{Span}^+(\mathbf{I}_n))$, then $s_i(k_l^i) = 0$, the submatrix $\mathbf{S}^i = [\mathbf{s}(k_1^i)\ \mathbf{s}(k_2^i)\ \cdots\ \mathbf{s}(k_{n-1}^i)]$ is of size $n \times (n-1)$. Since $\left\{\mathbf{s}(k_l^i), 1 \leq l \leq n-1\right\} \subseteq \mathbf{Span}^+(\mathbf{V})$, then there is a non-negative matrix $\mathbf{Y} \geq 0$ of size $n \times (n-1)$, such as $\mathbf{S}^i = \mathbf{VY}$. For $1 \leq j \leq n$, and $1 \leq l \leq n-1$, the jl-th entry of $\mathbf{S}^i$ is:

$$s_j(k_l^i) = \sum_{p=1}^{n} v_{jp} y_{pl} \tag{18}$$

where v_{jp} is the jp-th entry of $\mathbf{V}$ and y_{pl} is the pl-th entry of $\mathbf{Y}$. For $j = i$, we get:

$$s_i(k_l^i) = 0 = \sum_{p=1}^{n} v_{ip} y_{pl} \tag{19}$$

Since $v_{ip} \geq 0$ and $z_{pl} \geq 0$, then $\forall\ 1 \leq p \leq n$, $v_{ip} = 0$ or $y_{pl} = 0$. For a fixed p, if $v_{ip} > 0$ then $\forall\ 1 \leq l \leq n-1$, $y_{pl} = 0$. It follow that each non zero v_{ip} involves that the corresponding row of $\mathbf{Y}$ must have zeros entries. Furthermore, the vectors corresponding to the points $\mathbf{S}(k_l^i)$, $l = 1, 2, \cdots, n-1$ are linearly independent, then $rg(\mathbf{S}^i) = n-1 = \min(rg(\mathbf{V}), rg(\mathbf{Y}))$, since $\mathbf{Y}$ is size $n \times (n-1)$, then there can be no more than one zeros row in $\mathbf{Y}$. Therefore, there is only one non zero entry on the ith row of $\mathbf{V}$. By varying i from 1 to n, one can conclude taht there is only one non zero entry on each row of $\mathbf{V}$. Since $\mathbf{V}$ is non-singular, we deduce that $\mathbf{V}$ is a monomial matrix. ■

One may note that if the source points located on the facets of the positive orthant are distributed such that there is another simplicial cone containing the scatter plot of sources, and whose aperture is lower or equal to the aperture of the positive orthant, then the underlined simplicial cone will not be non-negative (i.e. will not be included in the positive orthant), and cannot be the non-negative MASC containing the scatter plot of the sources. This is illustrated in Fig. 2, for $n = 3$.

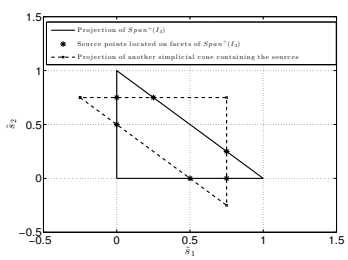


Fig. 2. Illustration of the sufficiently spread condition, for $n = 3$

Propositions 2, 3 and 4, give different sufficient conditions on the sources, under which $\mathbf{Span}^+(\mathbf{I}_n)$ is the unique non-negative MASC containing their scatter plot, and therefore $\mathbf{Span}^+(\mathbf{A})$ is the unique non-negative MASC containing the scatter plot of the mixed data. These conditions cannot all be deduced from each other, therefore we can state that each of this condition is suffucient but not necessary.

C. Conjecture of necessary and sufficient condition on the sources for unique N-BSS by MASC

Based on the necessary condition, and the sufficient conditions described before, we conjecture that a necessary and sufficient condition on sources, under which $\mathbf{Span}^+(\mathbf{I}_n)$ is the unique non-negative MASC containing the scatter plot of the sources can be established as follows:

1) There are at least one point of the scatter plot of sources belonging to each facet of the positive orthant, such as the vectors corresponding to these points are linearly independent: $\forall\ 1 \leq i \leq n$, $\exists\ k_i \mid s_i(k_i) = 0$ and the n-size submatrix $\mathbf{T} = [\mathbf{s}(k_1)\ \mathbf{s}(k_2)\ \cdots\ \mathbf{s}(k_n)]$ is full rank.
2) If $\mathbf{Span}^+(\mathbf{T}) \subset \mathbf{Span}^+(\mathbf{I}_n)$, then there is at least one point of the scatter plot of sources, $\mathbf{s}(l)$ with $1 \leq l \leq K$, which lies outside any simplicial cone involving some or all of the vectors $\mathbf{s}(k_1)$, $\mathbf{s}(k_2)$, $\cdots$, $\mathbf{s}(k_n)$.

Fig. 3 illustrates this intuitive condition on $\mathcal{P}$, in the case where $m = n = 3$. A proof of this conjecture is being studied.

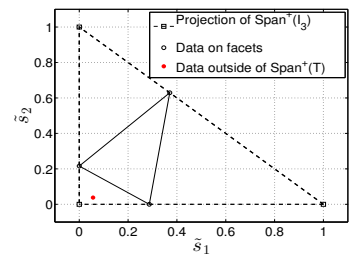


Fig. 3. Illustration of the necessary and sufficient condition, for $n = 3$

IV. Simplicial Cone Shrinking Algorithm for Unmixing Non-negative Sources: SCSA-UNS

Assuming that $\mathbf{Span}^+(\mathbf{I}_n)$ is the unique non-negative MASC containing the scatter plot of the sources, and according to Theorem 5, the task of estimating the mixing matrix is reduced to finding the non-negative MASC containing the scatter plot of mixed data. This can be reduced to solving the following optimization problem (20), where $\hat{\mathbf{A}}$ is the estimated mixing matrix and $\mathcal{A}(\mathbf{A})$ is defined by equation (13):

$$\hat{\mathbf{A}} = \underset{\mathbf{A} \geq 0,\ \mathbf{A}^{-1}\mathbf{X} \geq 0}{\arg\min} \quad \mathcal{A}(\mathbf{A}) \tag{20}$$

For solving (20), a possible direction is a gradient type method. However, it would lead to a non-convex optimization problem, due to the denominator term of the criterion (13). Instead of this approach, we propose an iterative method which starts from an initial simplicial cone containing the scatter plot of the mixed data, and progressively decreases its aperture until its fit the scatter plot of the mixed data. The proposed algorithm, termed SCSA-UNS for Simplicial Cone Shrinking Algorithm for Unmixing Non-negative Sources, is described below.

A. Proposed algorithm

1) Finding a proper initial simplicial cone:

The first step of the proposed SCSA-UNS algorithm is finding an initial simplicial cone $\mathbf{Span}^+(\mathbf{U})$ containing the scatter plot of the mixed data, say $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n]$. For reasons which we will explain in Sec. IV-A2, we require that $\mathbf{u}_p^T \mathbf{u}_q \geq 0$, $\forall\ 1 \leq p, q \leq n$. Since we first restrict to $m = n$, a trivial such simplicial cone is given by $\mathbf{U} = \mathbf{I}_n$. For getting the initial simplicial cone, one can also use an extended version of the solution of a local dominant based method, in this paper, we use extended VCA [32]. If $\mathbf{V}$ is the mixing matrix estimated by VCA, then the i-th column of $\mathbf{U}$, $\mathbf{u}_i$, is computed by $\mathbf{u}_i = \mathbf{v}_i - \alpha\bar{\mathbf{v}}$, where $\mathbf{v}_i$ is the i-th column of $\mathbf{V}$ and $\bar{\mathbf{v}}$ is the mean of the columns of $\mathbf{V}$. The coefficient α should be adjusted in order to satisfy the condition $\mathbf{u}_p^T \mathbf{u}_q \geq 0$, $\forall\ 1 \leq p, q \leq n$.

2) Decreasing the aperture of the current Simplicial Cone:

We define the n-size matrices $\mathbf{R}_i$ by equation (21), where $r_{ii} = 1$, and $r_{ji} \geq 0$ for $1 \leq j \leq n, j \neq i$.

$$\mathbf{R}_i = \begin{pmatrix} 1 & 0 & \cdots & 0 & r_{1i} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & r_{2i} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & r_{(i-1)i} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & r_{(i+1)i} & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & r_{ni} & 0 & \cdots & 1 \end{pmatrix} \quad (21)$$

Proposition 5: Let $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n]$, the n-size matrix (where $\mathbf{u}_p$ is the p-th column of $\mathbf{U}$), and $\mathbf{Span}^+(\mathbf{U})$ the current simplicial cone containing the scatter plot of the mixed data, i.e $\mathbf{U}^{-1}\mathbf{X} \geq 0$. For a fixed i, let $\mathbf{W} = \mathbf{U}\mathbf{R}_i$. If $\forall\ 1 \leq p, q \leq n$, $\mathbf{u}_p^T \mathbf{u}_q \geq 0$, then:

1) $\mathcal{A}(\mathbf{W}) \leq \mathcal{A}(\mathbf{U})$.
2) The coefficients r_{ij} of $\mathbf{R}_i$ can be computed such that $\mathbf{Span}^+(\mathbf{W})$ also contains the scatter plot of the mixed data, i.e $\mathbf{W}^{-1}\mathbf{X} \geq 0$.

Proof: Set $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n]$, where $\mathbf{w}_p$ is the p-th column of $\mathbf{W}$.

i. $|\det(\mathbf{W})| = |\det(\mathbf{U}\mathbf{R}_i)| = |\det(\mathbf{U})|\,|\det(\mathbf{R}_i)|$. Since $\det(\mathbf{R}_i) = 1$ then $|\det(\mathbf{W})| = |\det(\mathbf{U})|$
ii. $\forall\ l \neq i, \mathbf{w}_l = \mathbf{u}_l \Rightarrow \|\mathbf{w}_l\|_2 = \|\mathbf{u}_l\|_2$
iii. $\mathbf{w}_i = \sum_{j=1}^{n} r_{ji}\mathbf{u}_j = \mathbf{u}_i + \sum_{j=1, j\neq i}^{n} r_{ji}\mathbf{u}_j$, then

$$\mathbf{w}_i^T\mathbf{w}_i = \mathbf{u}_i^T\mathbf{u}_i + 2\sum_{\substack{j=1\\ j\neq i}}^{n} r_{ji}\mathbf{u}_i^T\mathbf{u}_j + \sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{p=1\\ p\neq i}}^{n} r_{ji}r_{pi}\mathbf{u}_j^T\mathbf{u}_p \quad (22)$$

Since $\mathbf{u}_p^T\mathbf{u}_q \geq 0$, $\forall\ 1 \leq p, q \leq n$, then $\mathbf{w}_i^T\mathbf{w}_i \geq \mathbf{u}_i^T\mathbf{u}_i$, and therefore $\|\mathbf{w}_i\|_2 \geq \|\mathbf{u}_i\|_2$.

According to the definittion of the aperture of a simplicial cone given by equation (13), and considering items i, ii and iii, one can conclude that $\mathcal{A}(\mathbf{W}) \leq \mathcal{A}(\mathbf{U})$. This proves 1).

We now turn to the second part of the proof of proposition 5. Let $\mathbf{Y} = \mathbf{U}^{-1}\mathbf{X}$ and $\mathbf{Z} = \mathbf{W}^{-1}\mathbf{X}$, then $\mathbf{Z} = [\mathbf{R}_i]^{-1}\mathbf{Y}$. By computing $[\mathbf{R}_i]^{-1}$, one can easy verify that: $\forall\ 1 \leq k \leq K$, $z_i(k) = y_i(k)$ and $z_j(k) = y_j(k) - r_{ji}y_i(k)$ for $j \neq i$. $\mathbf{Span}^+(\mathbf{W})$ contains the scatter plot of the mixed data, if $\mathbf{Z} \geq 0$. It is therefore enough to choose:

$$0 \leq r_{ji} \leq \min_{1\leq k\leq K} \frac{y_j(k)}{y_i(k)}, y_i(k) \neq 0, \text{ for } j \neq i$$

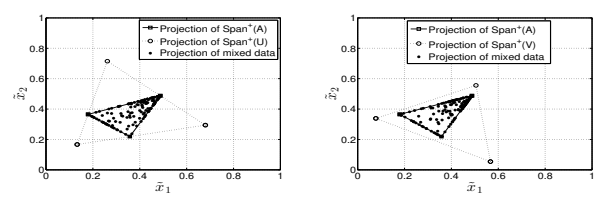
In the algorithm, we set

$$r_{ji} = \frac{1}{2}\min_{1\leq k\leq K} \frac{y_j(k)}{y_i(k)}, y_i(k) \neq 0 \quad (23)$$

This conclude the proof of proposition 5. ■

By starting from a proper initial simplicial cone containing the scatter plot of the mixed data (see Sec.IV-A1), say $\mathbf{Span}^+(\mathbf{U})$, the proposed algorithm iteratively decreases $\mathcal{A}(\mathbf{U})$ by performing several sweeps of n multiplications to the right of $\mathbf{U}$ by $\mathbf{R}_i$, i varying from 1 to n. At each iteration, the matrix $\mathbf{R}_i$ is computed as described in Proposition 5, in order to decrease the aperture of the current simplicial cone, while keeping all the mixed data inside of the new simplicial cone. The algorithm stops when the current simplicial cone fits the scatter plot of the mixed data. In this case, one cannot decrease anymore the aperture of the current simplicial cone by the matrices $\mathbf{R}_i$, while keeping all the mixed data inside of the new cone. This is called "Locking situation" and is detailed in the paragraph below.

3) Locking before convergence and proposed unlocking: If $\mathbf{Span}^+(\mathbf{U})$ is the current estimated simplicial cone, then the algorithm described above stops if for any i between 1 and n, one cannot decrease anymore the aperture of the current simplicial cone by the matrix $\mathbf{R}_i$, while keeping all the mixed data inside of the new cone. In this case, the only matrix $\mathbf{R}_i$ verifying, for $\mathbf{W} = \mathbf{U}\mathbf{R}_i$, $\mathcal{A}(\mathbf{W}) \leq \mathcal{A}(\mathbf{U})$ and $\mathbf{W}^{-1}\mathbf{X} \geq 0$ is $\mathbf{R}_i = \mathbf{I}_n$. This situation often corresponds to the convergence of the current estimated mixing matrix $\mathbf{U}$ to the true mixing matrix $\mathbf{A}$. However, this situation may also occur while $\mathbf{U}$ has not converged yet to $\mathbf{A}$. The latter case is called "*locking before convergence*", and is illustrated in Fig.4(a) for $n = 3$. In this figure, the data and the current simplicial cone are projected on the plane $\mathcal{P}$, as described in section II-A.



(a) Locking before convergence (b) Unlocking process

Fig. 4. Scatter plot of mixed data included in $\mathbf{Span}^+(\mathbf{A})$

In case of locking before convergence, the proposed algorithm finds $\mathbf{R}_i = \mathbf{I}_n$, $\forall\ 1 \leq i \leq n$, and therefore $r_{ji} = \delta_{ji}$. According to equation (23), it follows that $\forall\ 1 \leq i \leq n$, $\exists\ 1 \leq k_i \leq K$ such as $y_j(k_i) = 0$ and $y_i(k_i) \neq 0$. Since a locking before convergence causes zeros values in the current estimated sources $\mathbf{Y} = \mathbf{U}^{-1}\mathbf{X}$, the idea, in order to overcome this situation, is to modify the current simplicial cone $\mathbf{Span}^+(\mathbf{U})$ to make the current estimated sources strictly positive, without increasing $\mathcal{A}(\mathbf{U})$. For this purpose we look for an "*unlocking matrix*", $\mathbf{Q}$, that will rotate the current simplicial cone such that $\mathbf{T} = \mathbf{U}^{-1}\mathbf{Q}\mathbf{X} > 0$, and for $\mathbf{V} = \mathbf{Q}^{-1}\mathbf{U}$ $\mathcal{A}(\mathbf{V}) = \mathcal{A}(\mathbf{U})$. For computing the unlocking matrix, we introduce the criterion J, by equation (24), where t_{ik} is the ik-th entry of $\mathbf{T}$,

$$J(\mathbf{Q}) = \sum_{i=1}^{n}\sum_{k=1}^{K} t_{ik}^{+} \text{ where } t_{ik}^{+} = \begin{cases} 1 & \text{if } t_{ik} = 0 \\ 0 & \text{if } t_{ik} > 0 \\ +\infty & \text{if } t_{ik} < 0 \end{cases} \quad (24)$$

One may note that $J(\mathbf{Q}) = 0$ if and only if $\mathbf{T} > 0$,

and $\mathcal{A}(\mathbf{Q}^{-1}\mathbf{U}) = \mathcal{A}(\mathbf{U})$ if $\mathbf{Q}$ is an orthogonal matrix. A convenient unlocking matrix $\mathbf{Q}$ can be computed by solving the optimization problem (25):

$$\mathbf{Q} = \underset{\mathbf{O}^T\mathbf{O}=\mathbf{I}_n}{\arg\min}\ J(\mathbf{O}) \tag{25}$$

In order to deal with problem (25) in one step by a gradient like method, we slightly modify the original problem by:

- Regularizing the criterion J to avoid Dirac distributions when computing the gradient. We obtain the criterion J_ϵ given by:

$$J_\epsilon(\mathbf{Q}) = \sum_{i=1}^{n}\sum_{k=1}^{K} \exp(-\frac{t_{ik}}{\epsilon}),\ \sigma > 0$$

  One may note that $J_\epsilon(\mathbf{Q}) \longrightarrow J(\mathbf{Q})$ when $\epsilon \longrightarrow 0$.
- Adding a penalty term $J_{orth}(\mathbf{Q}) = \left\|\mathbf{Q}^T\mathbf{Q} - \mathbf{I}_n\right\|_F^2$ to $J_\epsilon(\mathbf{Q})$, in order to penalyze its deviation from orthogonality.

The optimization problem becomes :

$$\mathbf{Q} = \underset{\mathbf{O}}{\arg\min}\ J_\sigma(\mathbf{O}) + \gamma J_{orth}(\mathbf{O}), \text{with } \gamma \geq 0 \tag{26}$$

and can be solved by the iterative gradient algorithm (27):

$$\mathbf{Q}_{(p+1)} = \mathbf{Q}_p - \mu\left[-\frac{(\mathbf{U}^{-1})^T\mathbf{T}^{null}\mathbf{T}^T}{\sigma} + 4\gamma\mathbf{Q}_p\left(\mathbf{Q}_p^T\mathbf{Q}_p - \mathbf{I}_n\right)\right] \tag{27}$$

where $\mathbf{T}_{ij}^{null} = \exp(-\frac{t_{ij}}{\sigma})$, and p denotes the iteration.

The unlocking process is illustrated in Fig.4(b), where $\mathbf{Span}^+(\mathbf{V})$, is obtained by rotating $\mathbf{Span}^+(\mathbf{U})$, with the unlocking matrix $\mathbf{Q}$.

4) Framework of the SCSA-UNS algorithm for noiseless determined mixture: The pseudo-code of the proposed SCSA-UNS algorithm in the noiseless case is given by Algorithm 1.

Algorithm 1 : SCSA-UNS

Require: Mixed data $\mathbf{X} \in \mathbb{R}_+^{n\times K}$
Ensure: Estimated mixing matrix $\hat{\mathbf{A}} \in \mathbb{R}_+^{n\times n}$, and estimated sources $\hat{\mathbf{S}} \in \mathbb{R}_+^{n\times K}$

```
Initialization: Find a proper U by extended VCA, as
  described in Sec. IV-A1, and set Y = U^{-1}X
repeat
   for i = 1 → n do
      Compute R_i as describeb in Sec IV-A2
      U ← UR_i and Y ← [R_i]^{-1} Y
   end for
   if a locking situation occurs then
      Compute Q as described in Section IV-A3
      U ← QU and Y ← U^{-1}Q^{-1}X
   end if
until Q = I_n
Â ← max(U, 0)
Ŝ ← max(Y, 0)
```

5) About the convergence of SCSA-UNS:
The proposed algorithm, SCSA-UNS, is designed to monotonicaly decrease the criterion (i.e the aperture of the current simplicial cone). While the multiplications by the matrices $\mathbf{R}_i$ do not increase the criterion, the unlockings may slightly increase the criterion, since the matrices $\mathbf{Q}$ are not perfectly orthogonal (being computed by a regularized gradient). Although, the monotonic decrease of the criterion is not guaranted, we have noted through simulations that the proposed algorithm does not diverge.

B. Overdetermined mixture ($m > n$)

In case of an overdetermined mixture ($m > n$), one must first perform a dimension reduction before running the SCSA-UNS algorithm. By computing the Singular Value Decomposition of $\mathbf{X}$, we get:

$$\mathbf{X} \approx \mathbf{EFG}^T \tag{28}$$

where $\mathbf{F} \in \mathbb{R}^{n\times K}$ is the diagonal matrix of the n-largest singular values. $\mathbf{E} \in \mathbb{R}^{m\times n}$ (respectively $\mathbf{G} \in \mathbb{R}^{K\times n}$) is the matrix of the corresponding left (respectively right) singular vectors. The second step consists in running the SCSA-UNS algorithm on the reduced data $\mathbf{G}^T$, in order to compute the MASC containing its scatter plot, that is $\mathbf{G}^T = \mathbf{UY}$. The mixing matrix and the source are then estimated by:

$$\hat{\mathbf{A}} = \max(\mathbf{EFU}, 0) \text{ and } \hat{\mathbf{S}} = \max(\mathbf{Y}, 0) \tag{29}$$

V. Noise reduction

This section considers the case where the observations are corrupted by additive noise. The mixed data are then obtained as follows:

$$\mathbf{X} = \mathbf{AS} + \mathbf{B} \tag{30}$$

Due to the additive noise, $\mathbf{B}$, the scatter plot of mixed data may expand outside the simplicial cone generated by the mixing matrix. Therefore running SCSA-UNS on the noisy data, without any pre-processing, will lead to bad a estimation of the mixing matrix (and the sources), as illustrated in Fig. 5, for $m = n = 3$ sources. A pre-processing is necessary to reduce the noise effect, before estimating the mixing matrix.

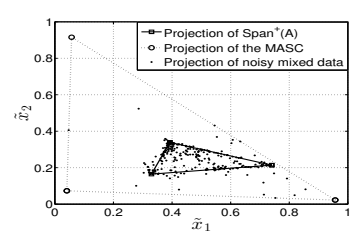


Fig. 5. Projection on $\mathcal{P}$ of the scatter plot of the noisy mixed data, $\mathbf{Span}^+(\mathbf{A})$ and the MASC estimated by SCSA-UNS

Assuming the noise is independent and identically distributed, one can note that the noiseless mixed data points are not affected in the same way. In fact, a noisy mixed data point is given by $\mathbf{x}(k) = \mathbf{As}(k) + \mathbf{b}(k)$, and let's define the "*local signal to noise ratio*" by $\mathbf{snr}(k) = -10\log_{10}\frac{\|\mathbf{As}(k)\|_2}{\|\mathbf{b}(k)\|_2}$. The

small norms noiseless mixed data will be strongly affected by the noise, since their *local signal to noise ratio* will be very low. To reduce the noise effect, we propose to select and discard all the mixed data points whose norm is lower than a certain threshold λ_b, before running the SCSA-UNS algorithm. Finding a proper threshold λ_b is not trivial, and based on experimental result, we propose $\lambda_b = \max_{1\leq K\leq K} \|\mathbf{x}(k)\|_2$. From the remaining mixed data, denoted $\underset{m\times L}{\mathbf{X}_r}$ (where L is the number of remaining data, after removing the small norms data), one can estimate the mixing matrix, $\underset{m\times n}{\hat{\mathbf{A}}}$, using the SCSA-UNS algorithm for overdetermined mixture (see section IV-B). The sources are then estimated by:

$$\hat{\mathbf{S}} = \max\left(\hat{\mathbf{A}}^{\dagger}\mathbf{EFG}^T, 0\right) \tag{31}$$

where $\hat{\mathbf{A}}^{\dagger}$ is the pseudoinverse of $\hat{\mathbf{A}}$, $\mathbf{F}$ is the diagonal matrix of the n-largest singular values of $\mathbf{X}$, and $\mathbf{E}$ (respectively $\mathbf{G}$) is the matrix of the corresponding left (respectively right) singular vectors.

VI. Evaluation on synthetic data

In this section, the proposed method is evaluated on synthetic data, and compared to other Non-negative Blind Source Separation methods. The different methods are evaluated for both, mutually independent and mutually correlated sources.

A. Performance indices and algorithms for comparison

We consider two performance indices for comparison. The first one is the *mixing matrix estimation error*, also called separation error E_{sep}. Given the estimated mixing matrix $\hat{\mathbf{A}}$ and its pseudoinverse $\hat{\mathbf{A}}^{\dagger}$, E_{sep} is defined by equation (32) [1] where $\Delta = \hat{\mathbf{A}}^{\dagger}\mathbf{A}$.

$$E_{sep} = \frac{1}{2n(n-1)}\sum_{i=1}^{n}\left[\sum_{j=1}^{n}\left(\frac{|\Delta_{ij}|}{\max_l|\Delta_{il}|} + \frac{|\Delta_{ji}|}{\max_l|\Delta_{li}|}\right) - 2\right] \tag{32}$$

The smaller is E_{sep}, the better is the separation, and E_{sep} is zero for perfect estimation of the mixing matrix. In this papaer, we'll consider that the mixing matrix is properly estimated if $E_{sep}(dB) < -10dB$.

The second performance index is the *source estimation error*, E_{corr}, defined by equation (33):

$$E_{corr} = -\log_{10}\left[\frac{1}{n}\sum_{i=1}^{n} \underset{1\leq j\leq n}{max} \frac{\left|\hat{\mathbf{s}}_i\mathbf{s}_j^T\right|}{\|\hat{\mathbf{s}}_i\|\,\|\mathbf{s}_j\|}\right] \tag{33}$$

where $\hat{\mathbf{s}}_i$ is the estimate of the i-th source and s_j is the true jth source. E_{corr} is inversely proportional to the correlation between the true sources and the estimated ones. The smaller is E_{corr}, the better is the estimation of the sources, and E_{corr} is zero for perfect separation. In this paper we'll consider that the sources are properly estimated if $E_{corr}(dB) < -20dB$.

The mixed data were generated according to equation (30). The noise matrix, $\mathbf{B}$, entries are generated following the standard Normal distribution, and the Signal to Noise Ratio of the i-th observation, SNR_i, is calculated by:

$$SNR_i = 10\log_{10}\left(\sum_{k}^{K}\left[\sum_{j=1}^{n} a_{ij}s_j(k)\right]^2 / \sum_{k=1}^{K}\left[b_i(k)\right]^2\right) \tag{34}$$

In all the simulations, the SNR are set to the same value for all the observations. Moreover, to keep the non-negativity of the noisy mixed data, we set the negative values to zero.

The proposed method, SCSA-UNS, is compared to six other N-BSS methods, namely SISAL [3], nLCA-IVM [48], VCA [32], N-ICA [38], MVC-NMF [22], and NMF [26]. For each of these methods, the data are pre-normalized to satisfy its required conditions. So, for or nLCA-IVM [48], each row of the data matrix is normalized to unit sum, and for SISAL [3] and VCA [32], each column of the data matrix is normalized to unit sum.

In all simulations, we set the SCSA-UNS parameters to $\mu = 10^{-2}$, $\epsilon = 10^{-3}$, $\gamma = \frac{1}{4}$. The parameters of methods used for comparison follow their original paper.

B. Independent sources

In this simulation, the sources are generated with different sparsity degrees, τ, (i.e number of non-zeros elements). Each non-zeros entry of the sources matrix is generated following the uniform distribution between 0 and 1. Three sparsity degrees are considered, $\tau = 90\%$ (i.e 10% of non-zeros entries), $\tau = 50\%$ (i.e 50% of non-zeros entries) and $\tau \to 0\%$ (i.e nearly 100% of non-zeros entries). The mixing matrix, $\mathbf{A}$, has a uniform random entries. We set the number of sources to $n = 5$, the number of observations to $m = 20$, and the number of samples is $K = 10000$.

Fig. 6, Fig. 7, and Fig. 8, depict the variation of the average performance indices versus the SNR, for the different sparsity degrees. The averages values are computed over 50 independent Monte Carlo runs. The underlined figures show that, for the different sparsity degree, the performance indices of the different methods improve when the SNR increases, N-ICA always presents the best performance indices, followed by SCSA-UNS, for $\tau = 90\%$ and $\tau = 50\%$.

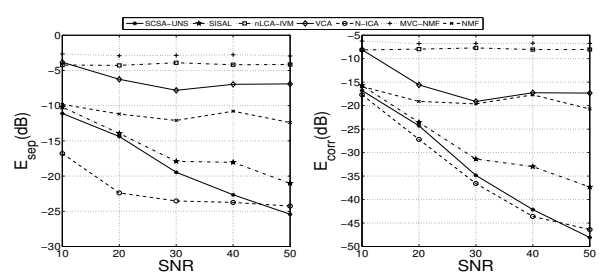


Fig. 6. Average indices vs SNR: $n = 5$, $m = 20$, $K = 10000$, $\tau = 90\%$

C. Correlated sources

Here, we evaluate the ability of the different methods to

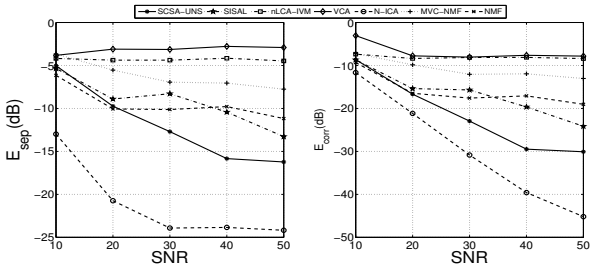


Fig. 7. Average indices vs SNR: $n = 5$, $m = 20$, $K = 10000$, $\tau = 50\%$

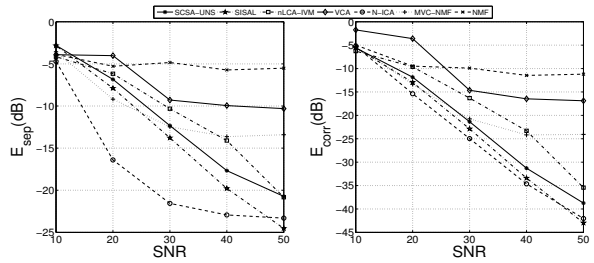


Fig. 8. Average indices vs SNR: $n = 5$, $m = 20$, $K = 10000$, $\tau \to 0\%$

unmix mutually correlated sources. We consider four mutually correlated images of size 350×275 despicted on Fig. 9 [10]. Each image is reshaped to form one row of the source matrix.

Fig. 9. Four correlated images

The estimated empirical correlation matrix, $\hat{C}_{\mathbf{S}}$, defined by equation (35) shows that the four images are highly correlated:

$$\left[\hat{C}_{\mathbf{S}}\right]_{ij} = \frac{\frac{1}{K}\sum_{k=1}^{K}\left[s_i(k) - \mu_{\mathbf{s}_i}\right]\left[s_j(k) - \mu_{\mathbf{s}_j}\right]}{\sigma_{\mathbf{s}_i}\sigma_{\mathbf{s}_j}} \tag{35}$$

$\mu_{\mathbf{s}_l}$ and $\sigma_{\mathbf{s}_l}$ are the mean value and the variance of source l.

$$\hat{C}_{\mathbf{S}} = \begin{pmatrix} 1.00 & 0.91 & 0.84 & 0.84 \\ 0.91 & 1.00 & 0.84 & 0.84 \\ 0.84 & 0.84 & 1.00 & 0.93 \\ 0.84 & 0.84 & 0.93 & 1.00 \end{pmatrix}$$

The number of sources is $n = 4$, the number of samples is $K = 96250$. The mixing matrix entries are randomly generated following a uniform distribution between 0 and 1, we set the number of observations to $m = 20$. Fig. 10 shows the variation of the average over 50 independent Monte Carlo runs, of the performance indices versus the SNR. One can see on Fig.10 that N-ICA has the worst estimated sources, this result is not surprisingly since this method has been designed for independent sources, and then is not suited to mutually correlated ones. The proposed method, SCSA-UNS, presents the best performance indices.

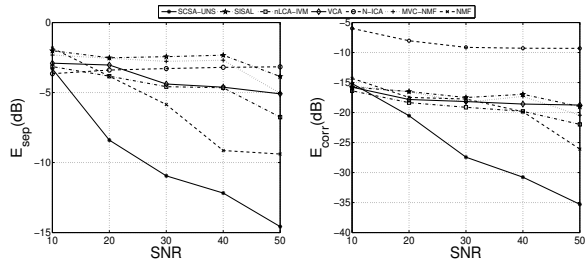


Fig. 10. Average indices vs SNR: $n = 4$, $m = 20$, $K = 96250$

Robustness to overestimation of the number of sources
One of the problem of sources separation in real blind context, is determining the proper number of sources. We compared here the performance of the different methods when the number of sources is overestimated. To perform this evaluation, we consider the previous experiment and we assume that there are five sources. The mixing matrix estimation error is calculated by taking the columns of the estimated mixing matrix most correlated to the columns of the actual mixing matrix, and the sources estimation error is calculated by taking the estimated sources most correlated to actual ones. Fig 11 shows the sources estimated by the different medthod. Visually one can see that only SCSA-UNS and nLCA-IVM properly estimate the four sources. SISAL fails due to the unit sum constraint on the estimated sources, and N-ICA cannot separate correlated sources. The performance indices recorded in Table I show that SCSA-UNS presents the best results.

	Noiseless case	
	$E_{sep}(dB)$	$E_{corr}(dB)$
SCSA-UNS	-9.68	-20.02
SISAL	-3.23	-9.33
nLCA-IVM	-4.45	-17.44
VCA	-3.11	-9.12
N-ICA	-2.61	-8.26
MVC-NMF	-3.82	-9.63
NMF	-6.91	-13.26

TABLE I
AVERAGE PERFORMANCE INDICES FOR CORRELATED SOURCES

VII. RESULTS ON ACTUAL DATA

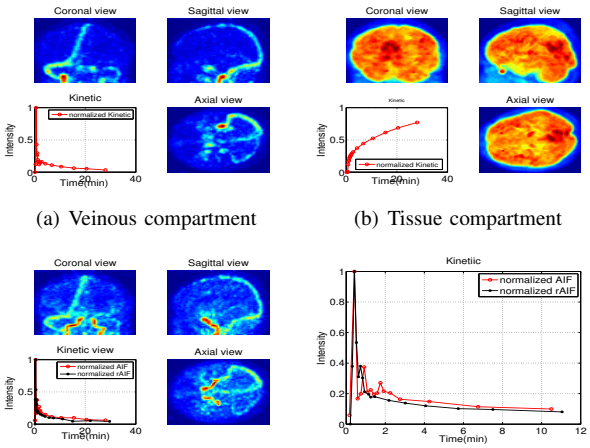
This section presents separation results on two different real data sets: images of positron emission tomography and Mass spectra. The indices E_{sep} and E_{corr}, used before to characterize the separation results on the synthetic data are no longer usable, since the original mixing matrices and sources are unknown. Results of SCSA-UNS are compared to those obtained by SISAL, N-ICA and MVC-NMF.

A. Results on Dynamic Positron Emission Tomography images

In oncology, the effectiveness of an anticancer treatment is often achieved using Positron Emission Tomography (PET) images [33] [42]. But an additional measurement of the Arterial Input Function (AIF), which provides the tracer concentration available for organs, is necessary to evaluate

Fig. 11. Correlated sources serapation for $SNR = 30dB$: (row 1) original sources, and sources estimated by (row 2) SCSA-UNS, (row 3) SISAL, (row 4), nLCA-IVM, (row 5) VCA, (row 6) N-ICA, (row 7) MVC-NMF, (row 8) NMF

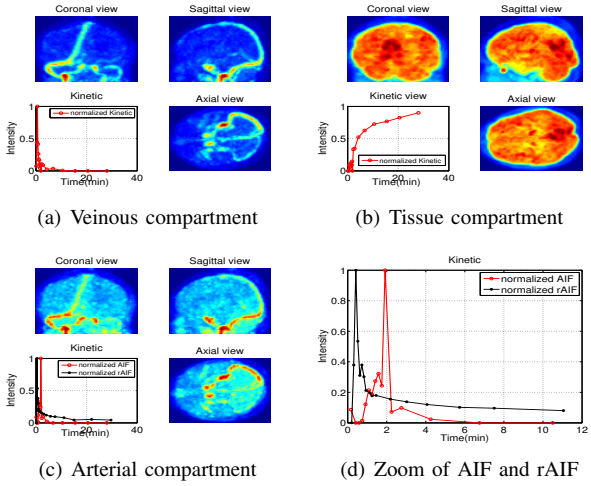
quantitatively the tumor activity. The reference method for AIF estimation is the arterial blood sampling which is unfortunately too invasive for routine clinic use. We perform here, the separation of real Dynamic Positron Emission Tomography (PET) images, to study the pharmacokinetics of the [18F]-FDG (FluoroDeoxyGlucose) tracer on human brain. The main objective is to estimate the arterial pharmacokinetic (Arterial Input Function) using only the dynamic TEP images, without blood sampling [21] [46]. In this experiment, an arterial input fonction obtained by blood sampling (reference Arterial Input Function rAIF) is also considered to assess the accuracy of the AIF estimation by sources separation using PET images. The data were provided by the CEA/I2BM/SHFJ.



(a) Veinous compartment (b) Tissue compartment

(c) Arterial compartment (d) Zoom of AIF and rAIF

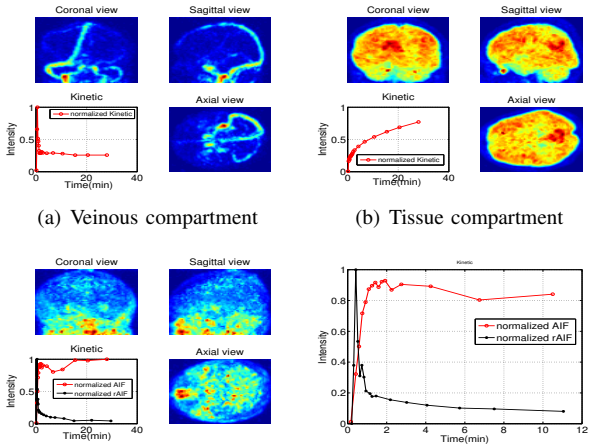
Fig. 12. Pharmacokinetic compartments estimated by SCSA-UNS



(a) Veinous compartment (b) Tissue compartment

(c) Arterial compartment (d) Zoom of AIF and rAIF

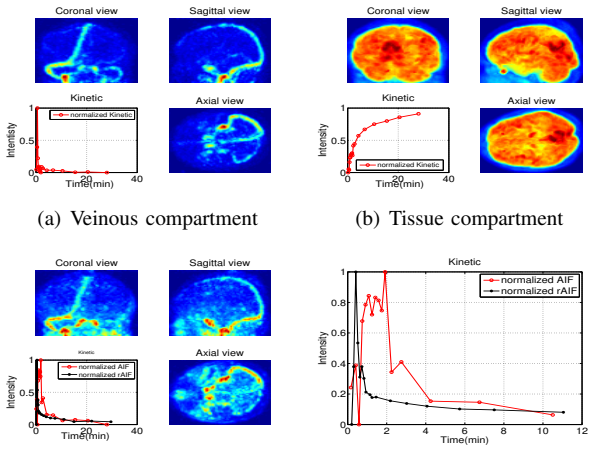
Fig. 13. Pharmacokinetic compartments estimated by SISAL



(a) Veinous compartment (b) Tissue compartment

(c) Arterial compartment (d) Zoom of AIF and rAIF

Fig. 14. Pharmacokinetic compartments estimated by N-ICA



(a) Veinous compartment (b) Tissue compartment

(c) Arterial compartment (d) Zoom of AIF and rAIF

Fig. 15. Pharmacokinetic compartments estimated by MVC-NMF

We have 19 human brain PET images recorded during $30mn$. The original images are size $128 \times 128 \times 63$, but we only considered subimages of size $63 \times 73 \times 58$ for focusing on the voxels with activity. Each 3D PET image is reshaped

to form one row of the observations matrix $\mathbf{X}$. The number of observations is $m = 19$ and the number of samples is $K = 266742$. According to biological consideration, we set the number of sources to $n = 3$.

Figures 12, 13, 14, and 15, shows the pharmacokinetic compartments estimated respectively by SCSA-UNS, SISAL, N-ICA and MVC-NMF. Each subfigure a, b and c, represents, at lower left, the normalized kinetics (columns of the estimated mixing matrix) and the corresponding spatial distributions (rows of the estimated sources) according to the three views, coronal (upper left), sagittal (upper right) and axial (lower right). All the four methods properly estimate the Veinous and the Tissue compartments. SISAL and MVC-NMF fail to recover the Arterial compartment, and one can on Fig 13.d and Fig 15.d that the AIF estimated by these methods are far from the refereence AIF, obtained by blood sampling. The third sources estimated by N-ICA suggests a nose compartment. Unlike the previous methods, SCSA-UNS found the cortid arteries (Fig. 12.c) [21], and Fig. 12.d shows that the AIF estimated by SCSA-UNS closely follows the reference AIF.

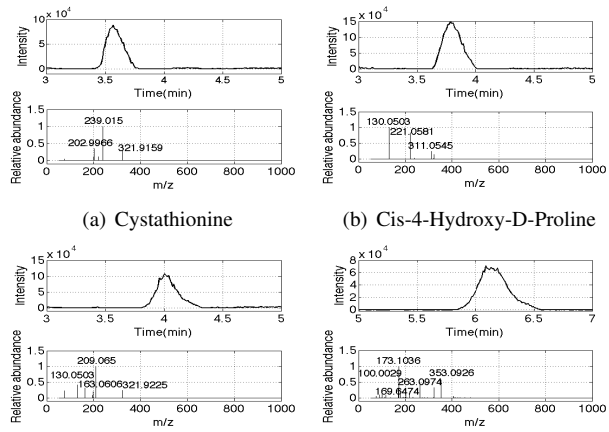
B. Results on Mass spectra for Metabolomics

Liquid Chromatography - Mass Spectrometry (LC-MS) technique is used for identifying and quantifying small organic molecules (also called metabolites) of biological/chemical fluids [50]. The sample to analyse is first introduced into a liquid chromatograph which separates its different metabolites according to their physico-chemical properties. The metabolites thus separated, flow out of the chomatograph during different time intervals, called elution profiles. At the output of the chromatograph, a mass spectrometer measures the mass spectrum of the eluted metabolites. A metabolite is characterized by its elution profile and its mass spectrum. Unfortunately, real biological/chemical fluids (urine, blood, saliva) are complex mixtures of tens or even hundreds of metabolites, therefore the liquid chromatograph cannot completely separate the elution profiles of the differents metabolites which overlap. So, the mass spectra measured can be modelized as linear combinations of the mass spectra of the different metabolites co-eluted (i.e metabolites whose elution profiles overlap). In this experiment, we seek to evaluate the efficiency of SCSA-UNS for separating the elementary elution profiles and mass spectra of differents metabolites co-eluted, given the measured mass spectra [49]. The data is provided by CEA/DSV/LEMM.

Eleven commercial chimical compounds were first individually analyzed by the combination of HPLC and LTQ-Obitrap (mass spectrometer) to built a reference database of their elution profiles and mass spectra. The eleven compounds are mixed and the resulting homogeneous solution is also analyzed by the combination of HPLC and LTQ-Obitrap. The mass spectra were recorded during 20 min and stored in an observation matrix whose rows correspond to mass spectra measured at different times, and columns correspond to mass indices. The observation matrix is prepocessed by removing all the zeros columns and all the rows whose maximum value is smaller than 10^4 (threshold set by our partners biologists). The resulting number of observations is $m = 1453$, and the number of samples is $K = 5638$.

We performed several runs the SCSA-UNS algorithm by incrementing each time the number of sources, starting from $n = 11$. We noticed that all the components are found when $n \geq 16$. Figure 16 shows the four first elution profiles, and corresponding mass spectra estimated by SCSA-UNS. This figure should be compared with the initial components database, give in appendix E. The estimated components do not exactly fit to the components referenced in the database, the greater part of the peaks are retrieved.

The other methods, namely SISAL, MVC-NMF and N-ICA find similar results, which are not shown here.



(a) Cystathionine (b) Cis-4-Hydroxy-D-Proline

(c) Beta-D-Fucose (d) DL-Arginine

Fig. 16. Four first elution profiles and corresponding mass spectra estimated by SCSA-UNS, from HPLC-LTQ Orbitrap data

VIII. Conclusion

In this paper, we propose a geometrical method for separating non-negative sources. The proposal, denoted SCSA-UNS, estimates the mixing matrix and the sources, by first reducing the dimension of the mixed data, followed by fitting a Minimum Aperture Simplicial Cone (MASC) to the scatter plot of the dimension reduced data. SCSA-UNS does not require the independence of sources, neither their local dominance, but the positive orthant must be the unique MASC containing the scatter plot of the sources, to ensure recovering the true mixing matrix and the true sources. In noisy case, the proposed method starts by discarding the points most corrupted by the noise, which can significantly expand the scatter plot of mixed data, before looking for the MASC containing the data. Simulation on synthetic data have showned that the proposed method performs good separation for both independent and mutually correlated sources. The proposal has also been successfully used to estimate the pharmacokinetic compartments of [18F]-FDG tracer on human brain (in particular to estimated the Arterial Input Function) and to separate the elementary mass spectra of differents chemical compounds, from the mass spectra measured at the output of a liquid chromatograph.

Future works include improving the robustness of the proposed method to additive noise, and incorporating multiplicative noise. In fact, in certain application the noise seem to also have a multiplicative part [19] [4], in addition to the additive part. The proof of the necessary and sufficient condition on

the sources, under which the simplicial cone generated by the mixing matrix is the unique non-negative MASC containing the scatter plot of the sources will also be investigated. Evalution on other real data (such as hyperspectral images) will also performed.

ACKNOWLEGDMENT

The authors would like to thank Régine Trébossen and Bertrand Tavitian of the CEA, I2BM, Service Hospitalier Frédéric Joliot, for the dynamic PET images. We also thank Christophe Junot of the CEA, DSV, Laboratoire d'Étude du Métabolisme des Médicaments (LEMM), for the LC-MS data.

APPENDIX A
PROOF OF THEOREM 1

$\mathbf{U} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{V} \in \mathbb{R}_+^{n\times n}$ are full rank matrices, let $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n]$ where $\mathbf{u}_i$ (respectively $\mathbf{v}_i$) the i-th column of $\mathbf{U}$ (respectively $\mathbf{V}$).
If $\mathbf{Span}^+(\mathbf{U}) \subseteq \mathbf{Span}^+(\mathbf{V})$ then $\forall\, 1 \leq i \leq n$, $\mathbf{u}_i \in \mathbf{Span}^+(\mathbf{V})$, therefore there is $\mathbf{m}_i \in \mathbb{R}_+^n$ such as $\mathbf{u}_i = \mathbf{V}\mathbf{m}_i$.
Let $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_n]$, then $\mathbf{M} \in \mathbb{R}_+^{n\times n}$ and $\mathbf{U} = \mathbf{VM}$. Since $rg(\mathbf{U}) = rg(\mathbf{V}) = n$ then $rg(\mathbf{M}) = n$.
Reciprocally, assume that there is a square full column rank non-negative matrix $\mathbf{M} \in \mathbb{R}_+^{n\times n}$ such as $\mathbf{U} = \mathbf{VM}$. For any $\mathbf{z} \in \mathbf{Span}^+(\mathbf{U})$, there is $\mathbf{y} \in \mathbb{R}_+^n$ such as $\mathbf{z} = \mathbf{U}\mathbf{y}$. Since $\mathbf{U} = \mathbf{VM}$, then $\mathbf{z} = \mathbf{VMy}$. It follows that $\mathbf{z} \in \mathbf{Span}^+(\mathbf{V})$ (because $\mathbf{My} \in \mathbb{R}_+^n$), and therefore $\mathbf{Span}^+(\mathbf{U}) \subseteq \mathbf{Span}^+(\mathbf{V})$.

APPENDIX B
PROOF OF THEOREM 2

Proof of the forward sense: Assume that $\mathbf{Span}^+(\mathbf{U}) = \mathbf{Span}^+(\mathbf{V})$, then $\mathbf{Span}^+(\mathbf{U}) \subseteq \mathbf{Span}^+(\mathbf{V})$, and there is a square full column rank non-negative matrix $\mathbf{M} \geq 0$ such as $\mathbf{U} = \mathbf{VM}$ (due to Theorem 1), so $\mathbf{V} = \mathbf{UM}^{-1}$. For any $\mathbf{y} \geq 0$, $\mathbf{Vy} \in \mathbf{Span}^+(\mathbf{V})$, then $\mathbf{Vy} \in \mathbf{Span}^+(\mathbf{U})$ (since $\mathbf{Span}^+(\mathbf{V}) = \mathbf{Span}^+(\mathbf{U})$). Furthermore $\mathbf{Vy} = \mathbf{UM}^{-1}\mathbf{y}$ $\Rightarrow$ $\mathbf{UM}^{-1}\mathbf{y} \in \mathbf{Span}^+(\mathbf{U})$, then $\mathbf{M}^{-1}\mathbf{y} \geq 0$, and therefore $\mathbf{M}^{-1} \geq 0$. Since $\mathbf{M} \geq 0$ and $\mathbf{M}^{-1} \geq 0$, one can conclude that $\mathbf{M}$ is a monomial matrix [35].

Proof of the reverse sense: Assume that $\mathbf{U} = \mathbf{VM}$ where $\mathbf{M}$ is a monomial matrix. For any $\mathbf{z} \in \mathbf{Span}^+(\mathbf{U})$, there is $\mathbf{y} \geq 0$ such as $\mathbf{z} = \mathbf{U}\mathbf{y}$. Then $\mathbf{z} = \mathbf{VMy}$, since $\mathbf{M}$ is a monomial matrix then $\mathbf{My} \geq 0$, therefore $\mathbf{z} \in \mathbf{Span}^+(\mathbf{V})$, and one can conclude that $\mathbf{Span}^+(\mathbf{U}) \subseteq \mathbf{Span}^+(\mathbf{V})$. In the same way, and using the fact that $\mathbf{M}$ is a monomial matrix, one can easily show that $\mathbf{Span}^+(\mathbf{V}) \subseteq \mathbf{Span}^+(\mathbf{U})$, which lead us to conclude that $\mathbf{Span}^+(\mathbf{U}) = \mathbf{Span}^+(\mathbf{V})$.

APPENDIX C
PROOF OF THEOREM 3

Let $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_n]$ and $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_n]$ with $\mathbf{g}_i = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|_2}$ and $\mathbf{h}_i = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|_2}$ then $\mathbf{Span}^+(\mathbf{G}) = \mathbf{Span}^+(\mathbf{U})$ and $\mathbf{Span}^+(\mathbf{H}) = \mathbf{Span}^+(\mathbf{V})$ (Theorem 2). It is enough to prove that $\mathcal{A}(\mathbf{G}) \leq \mathcal{A}(\mathbf{H})$.
$\mathbf{Span}^+(\mathbf{U}) \subseteq \mathbf{Span}^+(\mathbf{V}) \Rightarrow \mathbf{Span}^+(\mathbf{G}) \subseteq \mathbf{Span}^+(\mathbf{H})$ there is a full column rank non-negative matrix $\mathbf{M}$ such as $\mathbf{G} = \mathbf{HM}$ (Theorem 1). Let $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_n]$, where $\mathbf{m}_i$ is the ith column of $\mathbf{M}$. For fixed i: $g_{li} = \sum_{r=1}^{n} h_{lr} m_{ri}$, then

$$g_{li}^2 = \left(\sum_{r=1}^{n} h_{lr} m_{ri}\right)^2 \geq \sum_{r=1}^{n} h_{lr}^2 m_{ri}^2$$

$$\text{thus } \sum_{l=1}^{n} g_{li}^2 \geq \sum_{l=1}^{n}\left(\sum_{r=1}^{n} h_{lr}^2 m_{ri}^2\right) = \sum_{r=1}^{n} m_{ri}^2 \left(\sum_{l=1}^{n} h_{lr}^2\right).$$

Since $\sum_{l=1}^{n} g_{li}^2 = 1$ and $\sum_{l=1}^{n} h_{li}^2 = 1$, then $1 \geq \sum_{r=1}^{n} m_{ri}^2$, so $1 \geq \|\mathbf{m}_i\|_2$. It follows that:

$$1 \geq \prod_{i=1}^{n} \|\mathbf{m}_i\|_2 \;\Rightarrow\; 1 \leq \frac{1}{\|\mathbf{m}_1\| \times \|\mathbf{m}_2\| \times \cdots \times \|\mathbf{m}_n\|}$$

Furthermore, since $|\det(\mathbf{G})| = |\det(\mathbf{H})|\,|\det(\mathbf{M})|$ then

$$|\det(\mathbf{G})| \leq |\det(\mathbf{H})| \frac{|\det(\mathbf{M})|}{\|\mathbf{m}_1\|_2 \times \|\mathbf{m}_2\|_2 \times \cdots \times \|\mathbf{m}_n\|_2}$$

On the other hand, $\frac{|\det(\mathbf{M})|}{\|\mathbf{m}_1\|_2 \times \|\mathbf{m}_2\|_2 \times \cdots \times \|\mathbf{m}_n\|_2} \leq 1$ due to Hadamard's inequality, then $|\det(\mathbf{G})| \leq |\det(\mathbf{H})|$. Moreover, since $\prod_{i=1}^{n} \|\mathbf{g}_i\|_2 = 1$ and $\prod_{i=1}^{n} \|\mathbf{h}_i\|_2 = 1$, then $\mathcal{A}(\mathbf{G}) \leq \mathcal{A}(\mathbf{H})$.

APPENDIX D
PROOF OF THEOREM 5

We proceed by contradiction.

Proof of the forward sense: Suppose that $\mathbf{Span}^+(\mathbf{I}_n)$ is not the unique non-negative MASC containing the scatter plot of sources, there is thus a non-negative matrix $\mathbf{U}$ such as $\mathbf{U}^{-1}\mathbf{S} \geq 0$ and $\mathcal{A}(\mathbf{U}) \leq \mathcal{A}(\mathbf{I}_n)$.
$\mathbf{S} = \mathbf{A}^{-1}\mathbf{X} \Rightarrow \mathbf{U}^{-1}\mathbf{A}^{-1}\mathbf{X} = (\mathbf{AU})^{-1}\mathbf{X} \geq 0$, therefore $\mathbf{Span}^+(\mathbf{AU})$ is another simplicial cone containing the scatter plot of the mixed data. Since $\mathbf{A} \geq 0$ and $\mathbf{U} \geq 0$ then $\mathbf{Span}^+(\mathbf{AU}) \subseteq \mathbf{Span}^+(\mathbf{I}_n)$. Furthermore $\mathcal{A}(\mathbf{AU}) \leq \mathcal{A}(\mathbf{A})$ (because $\mathbf{U}$ is non-negative and due to Theorems 1 and 3), then $\mathbf{Span}^+(\mathbf{A})$ is not the unique non-negative MASC containing the scatter plot of the mixed data.

Proof of the reverse sense: Assume that $\mathbf{Span}^+(\mathbf{A})$ is not the unique non-negative MASC containing the scatter plot of the mixed data, there is thus another non-negative MASC, say $\mathbf{Span}^+(\mathbf{W})$, containing the scatter plot of the mixed data (i.e $\mathbf{X} = \mathbf{WY}$, where $\mathbf{Y} \geq 0$), and $\mathcal{A}(\mathbf{W}) \leq \mathcal{A}(\mathbf{A})$.
$\mathbf{S} = \mathbf{A}^{-1}\mathbf{X} \Rightarrow \mathbf{S} = \mathbf{A}^{-1}\mathbf{WY}$ with $\mathbf{Y} \geq 0$, then $\mathbf{Span}^+(\mathbf{A}^{-1}\mathbf{W})$ is another simplicial cone containing the scatter plot of the sources. $\forall\ \mathbf{y} \geq 0$, $\mathbf{Wy}$ is in the scatter plot of the mixed data, then $\mathbf{A}^{-1}\mathbf{Wy} \geq 0$, and it follow that $\mathbf{A}^{-1}\mathbf{W} \geq 0$. Since $\mathcal{A}\left(\mathbf{A}^{-1}\mathbf{W}\right) \leq 1$ (due to Hadamard's inequality), and $\mathcal{A}(\mathbf{I}_n) = 1$, then $\mathcal{A}\left(\mathbf{A}^{-1}\mathbf{W}\right) \leq \mathcal{A}(\mathbf{I}_n)$, and one can conclude that $\mathbf{Span}^+(\mathbf{I}_n)$ is not the unique non-negative MASC containing the scatter plot of the sources.

Chimical Compound	Times (min)	ID	m/z
Cystathionine	3.75	C4 H8 O2 N S	134,02812
		C7 H13 O4 N2 S	221.059
Cis-4-Hydroxy-D-Proline	3.83	C5H8NO3	130.05097
		C6 H8 N Na2 O5	220.020338
Beta-D-Fucose	4.01	C3H5O3	89.02442
		C6H11O5	163.0612
		C7H13O7	209.06668
DL-Arginine	6.23	C5H11O2N2	131.0826
		C6H13O2N4	173.1044
		C6H13O2N3[15]N	174.10143
Carnitine	8.87	C4H5N0O3	101,02442
		C10H18ClNNa3 O9	400.03687
DL Vanillomandelic Acid	11.48	C9 H9 O5	197.04555
		C18 H19 O10	395.09837
P-Amino Hippuric Acid	11.8	C8H9ON2	149.07204
		C9H9O3N2	193.06187
		C12H12O9N2Na3	397.02414
Adipic Acid	11.92	C5[13]CH7O3	128.04342
		C6H9O4	145.05063
		C5[13]CH9O4	146.05399
Folic Acid	12.22	C5H6O3N	128.03532
		C14H10N6O3	155.04127
		C18H17N7O4	197.56765
		C19H18O6N7	440.1324
Hydroxybenzoic Acid	13.69	C7H5O3	137.02442
		C7H5ClNaO3	194.98304
Naringenin	16.92	C15 H11 O5	271.0612
		C30 H23 O10	543.12967

Fig. 17. Database of the eleven commercial chimical compounds

APPENDIX E
DATABASE OF THE COMMERCIAL CHIMICAL COMPOUNDS

REFERENCES

[1] S. Amari, A. Cichocki, and H. Yang, "A New Learning Algorithm for Blind Signal Separation," in *Advances in Neural Information Processing Systems*. MIT Press, 1996, pp. 757–763.

[2] M. Babaie-Zadeh, A. Mansour, C. Jutten, and F. Marvasti, "A Geometric Approach for Separating Several Speech Signals," in *Independent Component Analysis and Blind Signal Separation*, ser. Lecture Notes in Computer Science, 2004, vol. 3195, pp. 798–806.

[3] J. Bioucas-Dias, "A Variable Splitting Augmented Lagrangian Approach to Linear Spectral Unmixing," in *First IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2009, pp. 1–4.

[4] F. P. Breitwieser, A. Müller, L. Dayon, T. Köcher, A. Hainard, P. Pichler, U. Schmidt-Erfurth, G. Superti-Furga, J.-C. Sanchez, K. Mechtler, and n. J. C. K. L. Bennett, "General Statistical Modeling of Data from Protein Relative Expression Isobaric Tags," *Journal of Proteome Research*, vol. 10, no. 6, pp. 2758–2766, 2011.

[5] R. H. C.Gobinet, E. Perrin, "Application of Non-negative Matrix Factorization to Fluorescence Spectroscopy," in *12th European Signal Processing Conference*, 2004, pp. 1095–1098.

[6] T. H. Chan, C. Y. Chi, Y. M.Huang, and W. K. Ma, "A Convex Analysis-Based Minimum-Volume Enclosing Simplex Algorithm for Hyperspectral Unmixing," *IEEE Transactions on Signal Processing*, vol. 57, pp. 4418–4432, 2009.

[7] T. H. Chan, W. K. Ma, C. Y. Chi, and Y. Wang, "A Convex Analysis Framework for Blind Separation of Non-Negative Sources," *IEEE Transactions on Signal Processing*, vol. 56, pp. 5120–5134, 2008.

[8] T.-H. Chan, "Convex Analysis Based Non-negative Blind Source Separation for Biomedical and Hyperspectral Image Analysis," Ph.D. dissertation, Institute of Communications Engineering, NTHU, 2009.

[9] Z. Chen and A. Cichocki, "Nonnegative matrix factorization with temporal smoothness and/or spatial decorrelation constraints," in *Laboratory for Advanced Brain Signal Processing, RIKEN, Tech. Rep*, 2005.

[10] A. Cichocki and R. Zdunek. NMFLAB MATLAB Toolbox for Non-Negative Matrix Factorization. [Online]. Available: http://www.bsp.brain.riken.jp/ICALAB/nmflab.html

[11] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, "8- Selected Applications," in *Nonnegative Matrix and Tensor Factorizations Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. John Wiley and Sons, Ltd, 2009.

[12] ——, *Nonnegative Matrix and Tensor Factorizations, Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, Wiley, Ed. John Wiley and Sons, Ltd, 2009.

[13] M. D. Craig, "Minimum-Volume Transforms for Remotely Sensed Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, pp. 542–552, 1994.

[14] D. Donoho and V. Stodden, "When Does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts?" in *Advances in Neural Information Processing Systems 16*. MIT Press, 2004.

[15] J.-L. Durrieu and J.-P. Thiran, "Sparse Non-negative Decomposition of Speech Power Spectra for Formant Tracking," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 5260 – 5263.

[16] N. Gillis, "Sparse and Unique Nonnegative Matrix Factorization Through Data Preprocessing," *Journal of Machine Learning Research*, vol. 13, pp. 3349–3386, 2012.

[17] P. O. Hoyer, "Non-negative Matrix Factorization with Sparseness Constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[18] A. Huck and M. Guillaume, "Robust Hyperspectral Data Unmixing with Spatial and Spectral Regularized NMF," in *2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2010, pp. 1–4.

[19] C. Hundertmark, R. Fischer, T. Reinl, S. May, F. Klawonn, and L. Jnsch, "MS-specific noise model reveals the potential of iTRAQ in quantitative proteomics," *Bioinformatics*, vol. 25, no. 8, pp. 1004–1011, 2009.

[20] M. M. Ichir and A. Mohammad-Djafari, "Bayesian Blind Source Separation of Positive Non Stationary Sources," in *American Institute of Physics Conference Series*, vol. 735, 2004, pp. 493–500.

[21] K. Kim, H. W. M., Shidahara, J. Ahn, S. Choi, N. Kudomi, K. Hayashida, Y. Miyake, and H. Iida, "Noninvasive estimation of cerebral blood flow using image-derived carotid input function in H215O dynamic PET," in *IEEE Nuclear Science Symposium Conference Record*, vol. 3, 2001, pp. 1282–1285.

[22] H. Q. L. Miao, "Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.

[23] H. Laurberg, M. G. Christensen, M. D. Plumbley, . K. Hansen, and S. H. Jensen, "Theorems on Positive Data: On the Uniqueness of NMF," *Computational Intelligence and Neuroscience*, vol. 2008, 2008.

[24] C. Lazar, D. Nuzillard, and A. Nowé, "A New Geometrical BSS Approach for Non Negative Sources," in *Latent Variable Analysis and Signal Separation*, ser. Lecture Notes in Computer Science, 2010, vol. 6365, pp. 530–537.

[25] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[26] ——, "Algorithms for Non-negative Matrix Factorization," *Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.

[27] J. S. Lee, D. D. Lee, S. Choi, and D. S. Lee, "Application Of Non-Negative Matrix Factorization To Dynamic Positron Emission Tomography," in *3rd International Conference on Independent Component Analysis and Blind Signal Separation*, 2001, pp. 629–632.

[28] J. Li and J. M. Bioucas-Dias, "Minimum Volume Simplex Analysis : A Fast Algorithm to Unmix Hypersectral Data," in *IEEE International Symposium on Geoscience and Remote Sensing Symposium*, vol. 3, 2008, pp. III – 250–III – 253.

[29] L. Miao and H. Qi, "A Constrained Non-Negative Matrix Factorization Approach to Unmix Highly Mixed Hyperspectral data," in *IEEE International Conference on Image Processing*, 2007, pp. 185–188.

[30] S. Moussaoui, D. Brie, A. M. Djafari, and C. Carteret, "Separation of Non-Negative Mixture of Non-Negative Sources Using a Bayesian Approach and MCMC Sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4133–4145, 2006.

[31] S. Moussaoui, D. Brie, and J. Idier, "Non-Negative Source Separation: Range of Admissible Solutions and Conditions for the Uniqueness of the Solution," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005, pp. 289–292.

[32] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, pp. 898–910, 2005.

[33] J. Okada, H. Oonishi, K. Yoshikawa, K. Imaseki, K. Uno, J. Itami, and N. Arimizu, "FDG-PET for the evaluation of tumor viability after anticancer therapy," *Annals of Nuclear Medicine*, vol. 8, no. 2, pp. 109–113, 1994.

[34] 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.

[35] R. J. Plemmons and R. E. Cline, "The Generalized Inverse of a Nonnegative Matrix," in *American Mathematical Society*, vol. 31, no. 1, 1972, pp. 46–50.

[36] M. Plumbley, "Conditions for Nonnegative Independent Component Analysis," *IEEE Signal Processing Letters*, vol. 9, pp. 177–180, 2002.

[37] ——, "Algorithms for Nonnegative Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol. 14, no. 3, pp. 534–543, 2003.

[38] ——, "Optimization using Fourier Expansion over a Geodesic for Non-Negative ICA," *Lecture Notes in Computer Science*, vol. 3195, pp. 44–56, 2004.

[39] ——, "Geometrical methods for non-negative ICA: Manifolds, Lie groups and toral subalgebras," *Neurocomputing*, vol. 67, pp. 161–197, 2005.

[40] M. Plumbley, A. Cichocki, and R. Bro, "Chapter 13 -Non-negative mixtures," in *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, P. Comon and C. Jutten, Eds. Academic Press, 2010, pp. 515–547.

[41] C. Puntonet, A. Mansour, and C. Jutten, "A Geometrical Algorithm for Blind Source Separation," in *GRETSI*, Juan-les-Pins, September 1995, pp. 273–276.

[42] P. Rigo, P. Paulus, B. J. Kaschten, R. Hustinx, T. Bury, G. Jerusalem, T. Benoit, and J. Foidart-Willems, "Oncological applications of positron emission tomography with fluorine-18 fluorodeoxyglucose," *European Journal of Nuclear Medicine*, vol. 23, no. 23, pp. 1641–1674, 1996.

[43] P. Sajda, S. Du, T. Brown, R. Stoyanova, D. Shungu, and L. P. X. Mao, "Nonnegative Matrix Factorization for Rapid Recovery of Constituent Spectra in Magnetic Resonance Chemical Shift Imaging of the Brain," *IEEE Transactions on Medical Imaging*, vol. 23, no. 12, pp. 1453–1465, 2004.

[44] R. Schachtner, G. Pöppel, and E. Lang, "Towards unique solutions of non-negative matrix factorization problems by a determinant criterion," *Digital Signal Processing*, vol. 21, pp. 528–534, 2011.

[45] X. Shao, Z. Zhichao, and W. Cai, "Extraction of chemical information from complex analytical signals by a non-negative indepependent component analysis," *Analyst*, vol. 134, pp. 2095–2099, 2009.

[46] Y. Su, A. M. Arbelaez, T. Benzinger, A. Z. Snyder, A. G. Vlassenko, M. A. Mintun, and M. E. Raichle, "Noninvasive estimation of the arterial input function in positron emission tomoraphy imaging of cerebral blood flow," *Journal of Cerebral Blood Flow & Metabolism*, vol. 33, pp. 115–121, 2012.

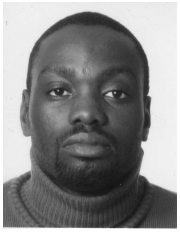
[47] T. Virtanen, "Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, 2007.

[48] F. Y. Wang, C. Y. Chi, T. H. Chan, and Y. Wang, "Nonnegative Least-Correlated Component Analysis for Separation of Dependent Sources by Volume Maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 875–888, 2010.

[49] X.-C. Xiong, X. Fang, Z. Ouyang, Y. Jiang, Z.-J. Huang, and Y.-K. Zhang, "Feature Extraction Approach for Mass Spectrometry Imaging Data Using Non-negative Matrix Factorization," *Chinese Journal of Analytical Chemistry*, vol. 40, no. 5, p. 663?669, 2012.

[50] B. Zhou, J. Xiao, L. Tuli, and H. Ressom, "LC-MS-based metabolomics," *Molecular BioSystems*, vol. 8, no. 2, pp. 470–481, 2012.

[51] Y. Zhu, T. H. Chan, E. P. Hoffman, and Y. Wang, "Gene Expression Dissection by Non-negative Well-Grounded Source Separation," in *IEEE Workshop on Machine Learning for Signal Processing*, 2008, pp. 255–260.



Wendyam Serge Boris Ouedraogo was born in Ouagadougou, Burkina Faso, in 1984. In 2009, following a double-degree program, he received the M.S. degree in Information Processing and Complexity, from École Nationale d'Ingénieurs de Tunis, Tunis, Tunisia, and the M.S. degree in Mathematics and Computer Science, from Université Paris Descartes, Paris, France. After a joint degree program, he earned the Ph.D. degree in Signal Processing, from Grenoble Institute of Technology (Grenoble INP), Grenoble, France, as well as the Ph.D. degree in Telecommunications, from École Nationale d'Ingénieurs de Tunis, Tunis, Tunisia, in 2012. He conducted his Ph.D. research with the Laboratoire d'Outils pour l'Analyse de Données (LOAD), of the CEA LIST, Saclay, France. Currently, he is postdoctoral researcher at GIPSA-lab, Grenoble, France. His research interests are in statistical signal processing, with an emphasis on blind source separation, independent component analysis, non-negative matrix factorization, and application in biomedical signal processing, chemical spectra analysis, and underwater acoustic.

Antoine Souloumiac was born in Bordeaux, France, in 1964. He received the M.S. degree and the Ph.D. degree in signal processing from the cole Nationale Suprieure des Tlcommunications, Paris, France, in 1987 and 1993, respectively. From 1993 until 2001, he was a Research Scientist with Schlumberger, Montrouge, France. He is currently with the CEA LIST. His research interests are in the area of statistical signal processing and its applications, with emphasis on point processes, biomedical signal processing, and independent component analysis (ICA) or blind source separation (BSS).

Mériem Jaïdane received the M.Sc. degree in electrical engineering from École Nationale d'Ingénieurs de Tunis (ENIT), Tunisia, in 1980. From 1980 to 1987, she has worked as research engineer at the Laboratoire des Signaux et Systèmes, CNRS/Ecole Supérieure d'Electricité, France. She received the Doctorat d'Etat degree in 1987. Since 1987, she has been with ENIT where she is currently a full Professor at the Information and Communications Technologies Department. She is a researcher at the Signals and Systems Lab at ENIT, Université Tunis El Manar. Her teaching and research interests are in adaptive systems for digital communications and audio processing.

Christian Jutten is full professor in University Joseph Fourier of Grenoble. He has been deputy director of the Grenoble images, speech, signal and control laboratory (GIPSA, 300 people) and director of the Department Images-Signal (DIS, 100 people) from 2007 to 2010. For 30 years, his research interests are blind source separation, independent component analysis and learning in neural networks, including theoretical aspects (separability, source separation in nonlinear mixtures, sparsity) and applications in signal processing (biomedical, seismic, hyperspectral imaging, speech). He is co-author of more than 75 papers in international journals, 4 books, 24 invited plenary talks and 150 communications in international conferences. He is currently deputy director at Institute for Information Sciences and Technologies of CNRS. For his contributions in source separation and independent component analysis, he received the Medal Blondel (1997) from the French Electrical Engineering society, was elevated as a Fellow IEEE and as a senior member of Institut Universitaire de France in 2008. In 2012, he was awarded by an ERC Advanced Grant CHESS. In 2013, he has been elevated as EURASIP Fellow and reconducted for 5 years as a senior member of Institut Universitaire de France.