A review of blind source separation in NMR spectroscopy

Ichrak Toumi, Stefano Caldarelli, Bruno Torrésani

A review of Blind Source Separation in NMR Spectroscopy

Ichrak Toumi, Stefano Caldarelli

iSm2, UMR 7313, Aix Marseille Université CNRS Marseille, France

Bruno Torrésani

Aix-Marseille Université CNRS, Centrale Marseille
I2M, UMR 7373, 13453 Marseille, France

Abstract

Fourier transform is the data processing naturally associated to most NMR *experiments. Notable exceptions are Pulse Field Gradient and relaxation analysis, the structure of which is only partially suitable for FT.*
With the revamp of NMR *of complex mixtures, fueled by analytical challenges such as metabolomics, alternative and more apt mathematical methods for data processing have been sought, with the aim of decomposing the* NMR *signal into simpler bits. Blind Source Separation is a very broad definition regrouping several classes of mathematical methods for complex signal decomposition that use no hypothesis on the form of the data. Developed outside* NMR *, these algorithms have been increasingly tested on spectra of mixtures. In this review, we shall provide an historical overview of the application of Blind Source Separation methodologies to* NMR *, including methods specifically designed for the specificity of this spectroscopy.*

Keywords: NMR Spectroscopy, BSS, Non Negative Matrix Factorization, Independent Component Analysis, sparsity.

Contents

1 Introduction **3**

2 The BSS Paradigm **7**

2.1 Introduction to BSS 7

2.2 Mathematical overview of the approach and application domains 9

3 Application of BSS to NMR Spectroscopy **11**

3.1 BSS Methods in NMR Spectroscopy 12

3.1.1 Methods based on statistical modelling 13

3.1.2 Methods based on sparsity 23

3.1.3 Variational approaches 43

3.2 Tensor based methods (PARAFAC) 48

4 Validation process: **59**

4.1 Impact of noise: artificial mixtures and additional artificial noise 62

4.2 Case of real-world mixtures and real noise 66

5 Conclusion **69**

1. Introduction

This review concerns the application to *NMR* of a specific class of algorithms, collectively known as the Blind Source Separation (**BSS**) approach, which has been used in areas as different as multichannel audio signal separation, speech recognition, multispectral image processing or bio-medical signal processing to quote only a few (see [1, 2, 3] and references therein).

Indeed, the very high resolution of solution-state *NMR* spectroscopy has led towards its application in cases of very high spectral complexity, such as proteins or liquid crystals, both of which can present hundreds of resonances. However, one obvious and widespread alternative utilization of the resolving power of *NMR* is analytical, the identification and quantification of the components of a mixture.

The challenge here is two-fold: either to detect selected and interesting compounds (for instance new natural products or elusive metabolites) or to extract cumulative spectral features descriptive of a sample properties, such as biomarkers [4]. Indeed, while an analytical application of *NMR* has been in use since the earliest times, it tooks a whole new dimension with the inception and blooming of multivariate analysis studies of the kind that became common in metabolomics or food science, among others. Here, tens to hundreds of compounds of moderate molecular size are within the detection limit of *NMR* (of the order of nM to μM for classical *NMR*).

Remarkably, the resolution of 2D *NMR* spectra is such that even by visual comparison it has been possible to identify even features related to original natural products [5, 6]. At any rate, the identification of the section of a 2D spectrum of a mixture that belongs to a pure compound relies on exploring the peak connectivities and the comparison with databases. Such

an identification process would be simplified if a higher degree of "spectral purity" can be achieved, for instance by mathematical un-mixing. This latter may take many forms, **BSS** being just one of them. To illustrate the context, we provide below a quick overview of specific but not-BSS processing.

In some instances, a certain degree of specialized information can be successfully extracted even from monodimensional *NMR* spectra with a high degree of overlapping. Fitting to known metabolites or deconvolution using Bayesian analysis has been demonstrated, for example in [7, 8]. Spiking with a known molecule has been proposed as a way of identifying and removing the specific signal of uninformative molecules [9].

Nonetheless, spreading of the resonances through classical multidimensional experiments, albeit time-consuming, is one of the typical solutions to the lack of resolution of simple 1D spectra. Thus, *COSY*, *TOCSY*, *HSQC*, *HMQC* and more rarely *HMBC* are common spectral tools employed to unravel the composition of complex mixtures via *NMR* , particularly for assignment [10]. For unlabeled molecules, the use of the simplified spectra associated to multiple-quantum transitions has also allowed a very high discrimination [11, 12, 13, 14, 15, 16].

First attempts at introducing 2D experiments directly as metabolomics tools have been performed, as reviewed in [17].

As the number of mixture components increase, some degree of overlap of the signals even in the 2D experiments becomes inevitable, so that it is all natural to try to further improve their resolution by data processing, covariance analysis [18, 19, 20, 21, 22, 23, 24, 25, 26, 27] or pure-shift spectroscopy [28, 29, 30, 31, 32, 33, 34, 35] being notable examples.

Identification of signals or of group of signals can be recognized, by spe-

cialized statistics, according to their variations along a series of spectra, for instance due to changes in the molecular concentration of the sample constituents. Thus, the peak intensity constitutes an additional dimension, thanks to which the spectra are partially decomposed as seen in the Statistical Total Correlation Spectroscopy (STOCSY) [36]. This approach has been explored in some depth, with a number of published variants, reviewed recently [37].

The variation in the intensity of single molecular components in pseudo-2D *NMR* experiments correlating a molecular spectrum and the molecular diffusion can also provide sufficient variance to be analyzed according a similar scheme [38].

Indeed, *NMR* diffusometry has attracted considerable attention for mixture analysis. Specifically, the *DOSY* layout of the PFG-NMR experiment, with its conceptual proximity to chromatography, has been a favorite method for mixture analysis since its inception [39, 40, 41].

However, besides some attempts to add this technique to the metabolomics toolset, *DOSY* performs best so far with less than ten components. Indeed, *DOSY* suffers from limitations in the achievable resolution linked to the instability of common algorithms for inverting sum of exponential decaying functions, which limits the resolution along the molecular mobility dimension. While differences in mobility in a multicomponent sample can be amplified by interaction with a suitable matrix with selective affinity towards some of the mixture compounds [11, 39, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52] or by simplifying the *NMR* dimension [50], significant efforts have been invested in developing better performing *DOSY* processing.

These experiments provide the ideal playground for testing data processing aimed at demixing the *NMR* spectrum. Indeed, the amplitude variations

expected for a *DOSY* experiment follow regular laws and all stem from a unique dataset, so that spurious source of signal variations are minimal and the mathematical treatment facilitated [53, 54, 55, 56, 57, 58, 59].

Finally, it should be noted that linewidth and relaxation differences have found a limited number of applications to resolve the spectra of mixtures [60, 61].

This short overview provides the context that justifies the introduction of alternative mathematical analysis to better describe the *NMR* spectra of mixtures compared to the classical FT approach.

The underlying problems illustrated above for decomposing *NMR* spectra echo those encountered in the processing of other families of multichannel signals (for example acoustic, or biomedical), and methods developed in this context can be borrowed and adapted. Some of the classical approaches here are parametric with respect to the sources, namely they rely on a model for the experimental data and set the separation problem as a parameter identification problem, that is to say the sum spectrum is decomposed with respect to those of known samples. This solution is only partially viable for typical *NMR* of mixtures, as often the relevant compounds are unknown. Thus, **BSS** appears to be a sound place to start.

In the following we shall cover the definition and underlying principles of the declinations of the **BSS** approach that have been applied so far in *NMR* spectroscopy, along with a discussion of the original examples and a comparative discussion of the possible limitations. As this is a very active domain of research in applied mathematics and signal processing, the reported literature deals sometimes with tests of established methods but also with algorithms designed for the specificity of *NMR* . At any rate, the review is organized according to grand classes of algorithms, as they share similar

computational setups (and thus problems), in order to provide the most consistent view to date of the experimentation that has been performed in the field.

2. The BSS Paradigm

2.1. Introduction to BSS

Blind Source Separation aims at recovering a set of pure signals starting from linear mixtures of these latter without prior information about the source signals, whence the use of the word blind. This concept is so broad that under **BSS** one may include a large variety of approaches and algorithms, adapted to various application domains. We are interested here in the so-called *instantaneous* **BSS** problem, in which no extra transformation is performed on the sources prior to mixing.

More precisely, the *instantaneous* **BSS** model supposes the existence of r unobserved source signals $S(t), \dots S_r(t)$ giving rise to n observations (i.e. mixtures) $X_1(t), \dots X_n(t)$, written as linear combinations of the source signals in the form:

$$X_i(t) \approx \sum_{k=1}^{r} A_{ik}\, S_k(t), \quad i = 1, \dots n, \quad t = 1, \dots p \tag{1}$$

The numbers A_{ik} are called the mixing coefficients, and form a matrix A called *mixing matrix*. In matrix form, this brings us to the general **BSS** equation:

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

where $X, N \in \mathbb{R}^{n\times p}$, $A \in \mathbb{R}^{n\times r}$, $S \in \mathbb{R}^{r\times p}$. N represents additive noise. The rows of X represent the observations, and the rows of S are the source signals. Both the sources and mixing matrix are assumed to be unknown,

and the goal of **BSS** is precisely to identify them from the observations. A **BSS** problem is called *determined* if the number of observations n is greater than or equal to the number of sources r and *undetermined* otherwise.
Based on equation (2) and given a matrix of measurements X, the objective of the **BSS** approach is to estimate the matrices A and S. Existence and uniqueness of the solutions are often not guaranteed, and additional assumptions and/or constraints are generally necessary. In particular, two types of indeterminacies have to be taken into account:

- Sources are defined up to a *normalization factor*: multiplying a row of S by a nonzero value, and dividing the corresponding column of A by the same does not modify X.
- Sources are defined up to permutation: exchanging two rows of S and the corresponding two columns column of A does not change X.

As a consequence of the normalization indeterminacy, the independent sources can be assumed to have unit variance, without loss of generality. The consequence of the permutation indeterminacy is the fact that **BSS** does not allow one to order sources, without any additional assumption.

whitening: *Assuming as above that the sources are white (i.e. have unit variance) simplifies the estimation problem. First, notice that the observations can also be whitened as follows. Starting from an estimate C_X of the covariance matrix of X, assume that C_X is non-degenerate and denote by $W = C_X^{-\frac{1}{2}}$ the inverse square root of C_X (which can be well defined, as C_X is an Hermitian matrix). Then $X' = W\,X$ is white, and the model $X = A\,S$ can be written in the form $X' = W\,A\,S = A'\,S$. It can be shown that the matrix $A' = W\,A$ is now a unitary matrix, so that the search of A' amounts to a change of (orthonormal) basis. Once A' has been estimated, $A = W^{-1}A'$ is readily obtained. In practice, the covariance matrix C_X has to be estimated from data, and is not the true covariance.*

2.2. Mathematical overview of the approach and application domains

Blind source separation and more generally blind signal processing, have attracted significant attention during the last twenty years, because of the numerous potential applications in many areas of signal and image processing.

As alluded to above, the **BSS** problem can be tackled using various approaches that exploit different assumptions. The interested reader can find thorough descriptions of general principles and the main approaches in textbooks [1, 2]. Early works on **BSS** relied on statistical modelling, and assumed the sources to consist in i.i.d. (independent, identically distributed, i.e. white) realizations of independent random variables. It was shown that in such situations identifiability implies that all (but one) sources must have non-Gaussian distributions. Such approaches led to algorithms (the so-called **ICA**, *Independent Component Analysis*) aiming at optimizing some specific independence criteria: find an un-mixing matrix B such that the corresponding estimated sources $\hat{S} = B\,X$ are *maximally independent*, with

respect to the chosen criterion. Criteria include mutual information, negentropy, specific properties of cumulants of order larger than 2, and several *contrast functions*, which are often connected to non-Gaussianity measures [62]. The case of coloured (i.e. correlated) sources also attracted significant attention; in this case, it was shown that when the source spectra are different enough, the separation can be performed using only order two statistics, for example auto-covariance matrices, while most approaches rely on joint diagonalization of these latter [63, 64].

Besides statistics based approaches, paradigms have been developed that lead to **BSS** approaches exploiting different basic principles. Among these, sparsity has recently emerged as a powerful generic principle: the rationale is the fact that in a suitable representation space, sources are sparse, i.e. characterized by a very small set of nonzero (or non-negligible) coefficient values. Such relevant coefficients being mostly different for all sources, a given coefficient can be assumed to belong to a single source (or a very small number of sources), which leads to simpler estimation procedures. This approach, which originates in the seminal paper [65], has stimulated an important activity since then, and many **BSS** algorithms exploiting sparsity in a way or another have been proposed in the literature.

Without trying to be exhaustive, let us conclude this short overview by mentioning a third road to blind signal separation that has gradually emerged during the last 10 years, namely the variational approaches which involvs a fidelity term and try to minimize an objective function exploiting a number of constraints like non-negativity. Two methods are distinguished in this group: **PARAFAC** (parallel factor analysis) approaches which are three mode factor analytic methods and **NMF** (Non-negative matrix factorization) approaches which are often based upon simple and efficient opti-

mization algorithms, and easy to implement. We notice that in the context we are interested in here, non-negativity is quite a natural requirement (1D and 2D *NMR* spectra, as well as concentrations are non-negative), which makes the variational based methods very appealing. Finally, let us point out that we have only mentioned here three main generic approaches to **BSS** . Obviously, the latter can be combined to yield still other algorithms, which can prove efficient in various contexts. For instance, we shall be discussing in some details a combination of **NMF** and sparsity constrained method.

3. Application of BSS to NMR Spectroscopy

In the mixture case, the *NMR* spectrum is a linear combination of the spectra of the underlying individual components, which is the appropriate situation for using instantaneous **BSS** .

One of the main problems in ^1HNMR spectroscopy of mixtures is signal overlapping, which tends to increase with the number of components, their complexity, and/or similarity. Spreading the spectrum to a second dimension can significantly overcome this shortcoming. 2D *NMR* techniques have been used for mixture analysis, the most popular being: *DQF-COSY*, *J-RES*, *TOCSY*, *HSQC* and *DOSY* . Particularly, because of the notorious instabilities of the Inverse Laplace Transform (ILT) originally proposed for the popular *DOSY* processing of Pulse Field Gradient *NMR* [39], this experiment has been the focus of many alternative processing schemes, including **BSS** ones [66, 67, 68, 57, 53]. Since this experiment (and more precisely the analysis of PFG-NMR decays) will be the object of a number of examples in the following, we summarize briefly the underlying mathematics. For a mixture, a PFG-NMR dataset acquired using a series of n variable gradients,

G_i, produces a signal rationalized by the Stejskal-Tanner equation [69]:

$$X_i = \sum_k S_k \exp(-D_k \gamma^2 G_i^2 \delta^2 (\Delta - \delta/3)) \ , \tag{3}$$

where S_k is a source *NMR* spectrum, D_k is the corresponding diffusion coefficient, γ is the gyromagnetic ratio, δ is the duration of the pulse gradient field and Δ is the time in which diffusion is allowed to take place. Transporting equation (3) to **BSS** , the mixing coefficients A_{ik} are therefore all positive and represent the scaling factor of the signals from molecules when submitted to gradient G_i. The spectra of the individual compounds that define the matrix S are also positive-valued functions. The linear mixing model described by (2) is guaranteed and therefore satisfies the **BSS** condition.

As it will become clear in the following, **BSS** applications to *NMR* have been mostly evaluated qualitatively. A clear assessment of the conditions for which one can expect a good separation have not been established. Particularly, in the case of PFG-NMR, the limitations in terms of number of overalapping species and the required intensity variations along the series of experiments remains to be determined. However, this is hardly an issue confined to **BSS** , but rather a general one for un-mixing problems. At any rate, the examples discussed below will rely on visual appreciation of the separation performance in some test cases. Attempt at predicting the resolving power of a few selected algorithms will also be illustrated later on.

3.1. BSS Methods in NMR Spectroscopy

In this section we will review the **BSS** approaches proposed in the literature for unmixing of 1D/2D *NMR* spectra. We will also detail the methods that we have selected and more systematically tested. As stressed before,

the main differences between the various methods mentioned in the literature lie in the assumptions that were made to perform the separation. We can divide the methods in two different groups: those which are based on explicit statistical assumptions and those which rely on the minimisation of some specific criterion, involving a data fidelity term and sometimes some regularization term, incorporating prior knowledge. Methods involving statistics can be further subdivided into two families, those for which a statistical independence assumption between the rows of matrix S (spectra of components) is made and those for which the assumption is applied for the columns of matrix S (acquisition variable: time, frequency, etc.)

3.1.1. Methods based on statistical modelling

In these approaches, the observation and source matrices are modelled in such a way that their columns are realizations of identically distributed random vectors. When **BSS** is tackled from a statistical point of view, the sources are assumed to be mutually decorrelated. Two main families of approaches have been proposed, developed and studied thoroughly. The first one assumes that sources are indeed mutually decorrelated, but that each individual source is correlated, in such a way that the individual correlation matrices differ significantly. The separation thus rests on these differences. The second one assumes that the source decorrelation is replaced with the (stronger) assumption of source independence.

Second Order methods . The second-order **BSS** methods are based on calculating a second order criterion of independence between the sources to be separated. The criterion is usually characterized by the covariance function.

We can mention as example the **SOBI** (Second Order Blind Identification) algorithm, which Nuzillard *et al.* [64] applied to *NMR* spectroscopy.

SOBI exploits the time coherence of the source signals, in the case of ^{13}C *NMR* spectra since there the resonance lines are generally narrow enough to limit the probability of peak superimposition, which fulfill the orthogonality constraint, so that the sources are pairwise decorrelated. Moreover, modeling of *NMR* time-domain signals as sums of decaying exponential functions provides a time-correlation property required for **SOBI** .

The proposed approach relies only on stationary second-order statistics, and is based on a joint diagonalization of a set of covariance matrices [63]. We outline below the basic principles. The method supposes that the sources are mutually decorrelated, each individual source being correlated.

Assumption (SOBI)**.** *The rows of the source matrix are decorrelated realizations of correlated random sequences.*

This can be expressed mathematically by introducing the families of fixed lag covariance matrices $R_X(\tau)$ and $R_S(\tau)$, defined as follows. For each pair of rows x and x' of X, introduce the corresponding sample covariances, defined by:

$$R_{xx'}(\tau) = \sum_j x(j)\ x'\ (j+\tau) \tag{4}$$

For each value of the lag τ this generates a square matrix $R_X(\tau)$, which can be seen as a sample estimate of the true covariance matrix. The fixed lag covariance matrices of the sources $R_S(\tau)$ will be defined likewise. According to the above assumption, the source covariance matrices are expected to be diagonal. The basic principle of the corresponding approaches is to search for a linear transformation $X \to Y = BX$ such that a suitably chosen set of matrices $R_Y(\tau)$ becomes (at least approximately) diagonal. The corresponding Y will be the estimate for the source matrix S, and B will be the estimate for the un-mixing matrix. Notice that $R_Y(\tau)$=$BR_X(\tau)B^T$, so

that finding Y amounts to simultaneously diagonalize the fixed lag sample covariance matrices. Let us recall that a set of matrices $M_1, M_2, \ldots$ can be diagonalized simultaneously if and only if all matrices commute, i.e. if $M_i M_j = M_j M_i$ for all i, j. Otherwise, an approximate joint diagonalization can be performed numerically, by optimizing a suitable criterion, for example minimizing the sum of squares of off-diagonal elements, i.e. the quantity:

$$\mathsf{Off}(M) = \sum_{k \neq l} M_{kl} \ . \tag{5}$$

Joint diagonalization thus amounts in this case to searching for a unitary matrix U that solves the problem:

$$\hat{U} = \min_{U} \sum_{i} \mathsf{Off}(U M_i U^{-1}) \tag{6}$$

The first proposed algorithm following these principles, named **AMUSE**, exploits joint diagonalization of two fixed covariance matrices, namely $R_S(0)$ and a suitably chosen $R_S(\tau)$. In the context of *NMR* spectroscopy, A.M. Tomé and *al.* [70] developed a new version of **AMUSE** called *dAMUSE* which offers a fast and efficient way of removing the water artifact from the spectra and allows a denoising of a reconstructed artifact-free protein spectra to achieve noise levels comparable to those of the experimental spectra. The tool was tested on the 2D *NOESY* 1H *NMR* spectra of aqueous solutions of proteins.

SOBI exploits approximate joint diagonalization of a larger set of covariance matrices, according to the above criterion. The approximate joint diagonalization is performed numerically, using Jacobi transformations.

An application of this **BSS** method was done in 1D and 2D *NMR* Spectroscopy, as we briefly discuss below.

A first demonstration of **SOBI** *NMR* un-mixing followed the isomerization of $\alpha-$glucose into $\beta-$Glucose in D_2O. The spectra of mixtures consisted in five 1D ^{13}C *NMR* spectra, shown in Figure 1 (left plot) along with the estimated ^{13}C *NMR* spectra of the sources (right plot). Some cross-talk artifacts are visible, especially for the $\beta-$Glucose spectrum, which were interpreted as arising from small frequency misalignment due to concentration effect.[64].

Typically, *HSQC* is presented as a frequency correlation plot while **SOBI** was designed to deal with 1D time domain signals, therefore some pre- and post-processing steps were required. First, rectangular zones were defined around the cross peaks volumes to locate signals from all the sources. These regions were extracted and subjected to an inverse Fourier Transformation to produce time correlated data. The **SOBI** algorithm was then applied to obtain the mixing matrix and therefore the pseudo FIDs of the sources, which served to reconstruct a 2D frequency-domain presentation, from *HSQC* spectra of three mixtures of three components: sorbitol, mannitol and xylitol in D_2O (Figure 2 (left panels)). Some spurious residues can be noticed, in Figure 2 (right panels), in the sorbitol spectrum. The authors tentatively justified the imperfect demixing as a consequence of variations in the position of overlapping peaks caused by temperature fluctuations.

Independent component analysis (JADE, fastICA and variants). In generic ICA approaches, no correlation structure is assumed on the individual sources, but compared to **SOBI** assumption, the decorrelation hypothesis is replaced with the (stronger) hypothesis of mutual independence of the sources. Statistical independence in such models is a way of describing the differences

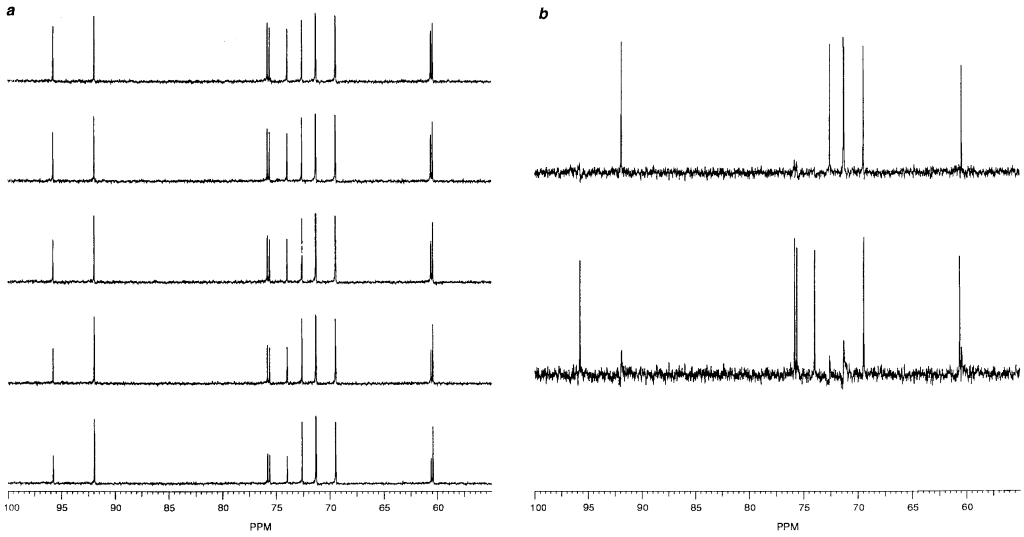

Figure 1: Demonstration of the SOBI algorithm. (a) Five ^{13}C spectra recorded during the isomerization of α-glucose to β-glucose in D_2O. (b) The separated spectra of α-glucose (upper trace) and β-glucose (lower trace). Reprinted from Journal of magnetic resonance, vol 133, D. Nuzillard, S. Bourg and J.-M. Nuzillard, Model-Free Analysis of Mixtures by NMR Using Blind Source Separation, p 358-363.Copyright 1998, with permission from Elsevier

between source spectra: the stronger the independence, the less similar the component spectra.

Assumption (ICA)**.** *The rows of the source matrix are independent realizations of independent identically distributed random sequences.*

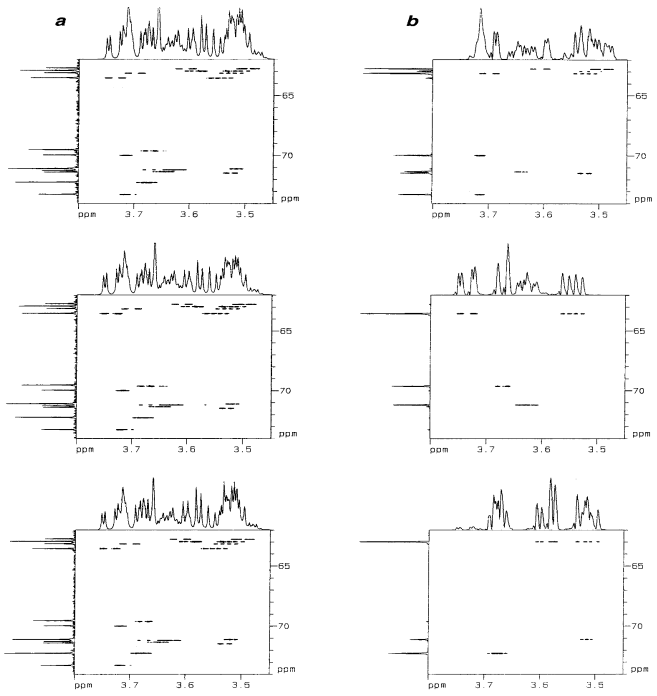

Figure 2: Demonstration of the SOBI algorithm. (a) The HSQC spectra of three mixtures of sorbitol, mannitol, and xylitol in D_2O. (b) The separated HSQC spectra of the components of the mixtures. Reprinted from Journal of magnetic resonance, vol 133, D. Nuzillard, S. Bourg and J.-M. Nuzillard, Model-Free Analysis of Mixtures by NMR Using Blind Source Separation, p 358-363.Copyright 1998, with permission from Elsevier

The ICA **BSS** problem is thus formulated as follows: find a (un-mixing) matrix B such that the rows of the corresponding un-mixed source matrix

$$Y = BX \tag{7}$$

are maximally independent, according to a given criterion. In general, one

obtains an estimate for the the un-mixing matrix B, from which an estimate for the mixing matrix is obtained (using pseudo-inverse, or some more sophisticated inversion procedure).

Let us briefly recall some basic probabilistic principles. Given two random variables y_1 and y_2, denote by $p(y_1, y_2)$ their joint probability density function (*pdf* for short), and by $p_1(y_1) = \int p(y_1, y_2)dy_2$ and $p_2(y_2) = \int p(y_1, y_2)dy_1$ the marginal *pdf*s. The two random variables are independent if $p(y_1, y_2) = p_1(y_1)p_2(y_2)$. This definition can be extended to any number n of random variables, in which case independence means that the joint *pdf* equals the product of the n marginal *pdf*s.

As it is well known, independence implies decorrelation (which only involves first and second order moments). In many ICA methods, mixture data are first decorrelated (using standard techniques, based upon principal component analysis) prior to ***BSS*** *. This generally simplifies the independent sources estimation, as already alluded to in the whitening remark.*

To solve the ICA problem and estimate the independent sources and the mixing matrix, one generally relies on optimization procedures, and search for an un-mixing matrix that minimizes the dependence of corresponding unmixed signals. Given some generic dependence criterion (also called *contrast function* $Y \rightarrow \mathsf{DepCrit}(Y)$, the optimization problem is formulated as

$$\hat{B} = \arg\min_B \mathsf{DepCrit}(BX) \ , \tag{8}$$

and solved numerically.

A classical and often advocated dependence criterion is the so-called *mutual information*, which measures the divergence between the *pdf* of a

random vector and the product of marginal *pdf*s of its components:

$$I(Y) = \int p(y_1, \dots y_n) \log \left(\frac{p(y_1, \dots y_n)}{p_1(y_1) \dots p_n(y_n)} \right) \, dy_1 \dots dy_n \ . \tag{9}$$

The mutual information is always non-negative, and vanishes if and only if the components of the random vector are mutually independent.

In practice, the mutual information cannot be computed explicitly, as the *pdf*s are not available (only sample estimates for *pdf*s can be available).

Many algorithms have been proposed, based upon the optimization of substitutes for the mutual information. The latter can indeed be based upon sample estimates (defined as in (9)), but also more general contrast functions, that measure some specific types of departures from independence.

Among these, the FastICA family of algorithms, described in the review paper of Hyvarinen, [71] [72] is among the simplest, and has been used for many applications like audio signal processing, genomics, EEG/MEG data analysis and *DOSY NMR* Spectroscopy [73]. FastICA relies on an approximation of the mutual information by a contrast function which can be regarded as a measure of non-gaussianity, and is optimized through a simple projected gradient method. The method proposes several choices for the contrast function, and two different optimization strategies: a global optimization, and an iterative method (called deflation and introduced in [74]) in which the sources are estimated one after the nother.

Among variants, let us quote the Efica algorithm which is an improved version of FastICA presented by Koldovsky and Tichavsky in [75], the MILCA (*Mutual Information based Least dependent Component Analysis*) which estimates the mutual information based on a nearest neighbors algorithm [76] and SNICA [77] (*stochastic non-negative independent component analysis*), a method dedicated to the analysis of non-negative signals that

performs best on signals with intensity distributions peaked at zero (like in spectroscopy) ; those algorithms have been used in [78] for the quantitative and qualitative analysis of UV absorption spectra of complex mixtures.

An example of application of these algorithms to *NMR* spectroscopy was made by J. Zhong et al in [73]. They proposed a new method called "**DIFFICA**"' which combines the Fast ICA algorithm and "*DOSY* " (1D & 2D) to perform the separation. According to the authors, based on the expression of intensity for *DOSY* eq. (3), the unmixing matrix $B = A^{-1}$ is initialized and the independence is ensured by the difference of the diffusion coefficients of the sources, which has to be large, to ensure a good separation.

JADE. Fourth-order cumulants can be more robust than the (MI) criterion for measuring the independence between the sources. The algorithm related to this named **JADE** , which stands for Joint Approximate Diagonalization of Eigen-matrices algorithm and we account for it in some details below. As most ICA algorithms, **JADE** consists to an estimate of the optimal un-mixing matrix B that restitutes an un-mixed signal matrix Y whose rows are the most statistically independent. **JADE** exploits higher-order statistics to perform the identification of the un-mixing matrix. As mentioned above, the covariance matrix is used to whiten the observations. After whitening, the covariance matrix of the observed mixtures is diagonal (and even equal to the identity). Independence, which is a stronger assumption than decorrelation, implies that all cumulants tensors are diagonal. Without going into the abstract definition of cumulants, let us simply mention that the cumulants tensors are higher order generalization of covariance matrices. For example, the n-th order joint cumulant of random variables $x_1, \dots x_n$ is obtained from the corresponding joint moment (the expectation of the

product $x_1 \dots x_n$) by substracting some corrective terms (mainly symmetric products of moments of lower order). Given a random $1 \times n$ vector, the second order joint cumulants form an $n \times n$ matrix, and the fourth order joint cumulants form an $n \times n \times n \times n$ tensor (i.e. a 4 entries hypercubic table).

JADE rests on the fact that given a random vector with independent components, all the corresponding cumulants tensors are diagonal. After whitening (that diagonalizes the second order cumulants tensor, i.e. the covariance matrix), **JADE** therefore seeks numerically a change of basis that (approximately) diagonalizes the fourth order cumulant tensor, by optimizing some contrast function. The latter is chosen to be the sum of the values of the $\mathsf{Off}(M_i)$ (sum of squares of off-diagonal elements, see the section on **SOBI** above) of the order two slices M_i of the order four cumulant tensor. Again, it is worth mentioning that the actual cumulant tensors are not available, only sample estimates can be used. We refer to chapter 5 of [1] for details.

Fig. 3, Fig. 4 and Fig. 5 illustrate the separation results by **JADE** for PFG-NMR experiments of three mixtures: **SM** (Sucrose, Maltotriose), **QGC** (Quinine, Geraniol and camphene) and **DENET** (Dextran, Ethanol, Nicotinic acid, Ephedrine and Tartrazine) datasets respectively [79].

In Fig. 6, Fig. 7 and Fig. 8, we show a *DOSY* reconstructed figures from the obtained matrices A and S. In fact, this analysis allowed the construction of a *DOSY* chart, by fitting matrix A to Stejskal-Tanner equation [69] to obtain the diffusion coefficients and subsequently locating the sources S on the chart at their corresponding values, broadened by a gaussian uncertainty as indicated by the error of the fit.

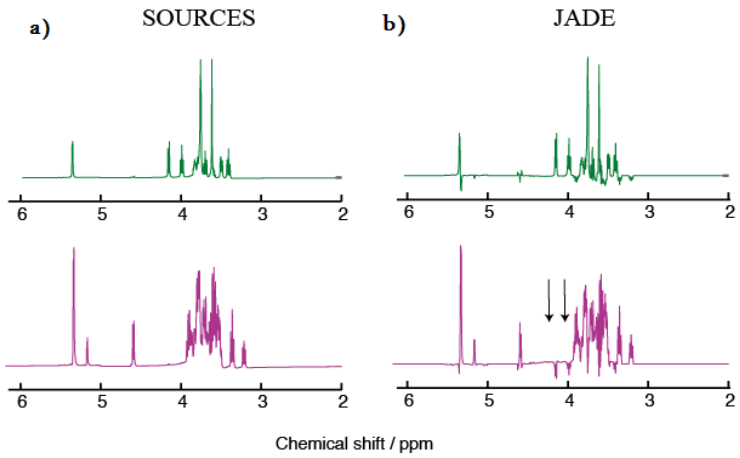

Figure 3: The ground truth sources (a) and the recovered sources by JADE (b) for a mixsture of Maltotriose and Sucrose, based on the analysis of a series of PFG-NMR experiments. Reprinted with permission from Anal. Chem, Vol 85, Toumi. I, Torrésani.B and Caldarelli.S, Effective Processing of Pulse Field Gradient NMR of Mixtures by Blind Source Separation,p 11344-11351. Copyright 2013 American Chemical Society

3.1.2. Methods based on sparsity

It has been noticed by several authors that the estimated sources provided by ICA often satisfy some sparsity property: they are characterized by probability distributions that are often sharply peaked at the origin. During the last ten years, sparsity has emerged as a new generic paradigm for signal processing (see for example the book [80]), and has found many applications in various areas. Sparsity can be understood in various ways, including peakyness of pdf, or in a stricter sense as follows:

Sparsity: *A vector y in n-dimensional space is k-sparse ($k \leq n$) in a transformed domain if its corresponding transform Tx involves no more than k non-zero coefficients.*

The rationale for the application of the sparsity concept to **BSS** is the following: suppose that one is given several mixtures $x_1(t), \ldots x_n(t)$ of

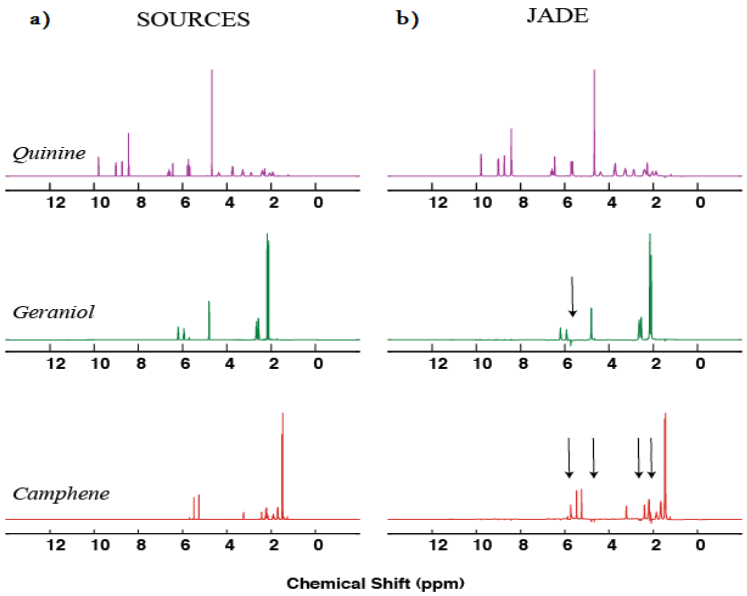

Figure 4: The recovered sources by JADE (a) and the ground truth sources (b) for a mixture of Quinine, Camphene and Geraniol, based on the analysis of a series of PFG-NMR experiments. Reprinted with permission from Anal. Chem, Vol 85, Toumi. I, Torrésani.B and Caldarelli.S, Effective Processing of Pulse Field Gradient NMR of Mixtures by Blind Source Separation, p 11344-11351. Copyright ©2013 American Chemical Society

sparse source signals $s_1(t), \dots s_m(t)$, and assume that the sources are different enough. If for some value of $t = t_0$ a given source $s_i(t)$ takes a significant value, it is very likely that the other sources $s_j(t)$ will take negligible values at $t = t_0$.

Therefore, finding values of t where only a single source is active can yield simple estimates of the mixing matrices, and thus the sources. This is the basic idea of the so-called sparse component analysis (**SCA**, see [81]), which has been exploited successfully in various domains. Sparsity is generally searched for in a transformed domain (for example, short time Fourier transform for audio source separation, wavelet transform for applications to

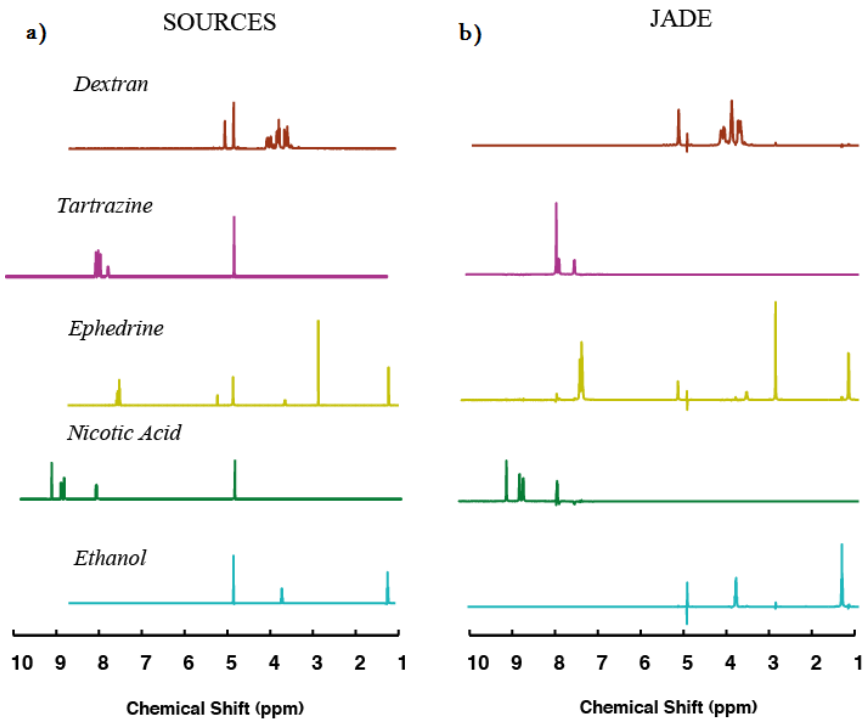

Figure 5: The recovered sources by JADE (a) and the ground truth sources (b) for a mixture of Dextran, Tartrazine, Ephedrine, Nicotinic Acid and Ethanol, based on the analysis of a series of PFG-NMR experiments.Reprinted with permission from Anal. Chem, Vol 85, Toumi. I, Torrésani.B and Caldarelli.S, Effective Processing of Pulse Field Gradient NMR of Mixtures by Blind Source Separation, p 11344-11351. Copyright ©2013 American Chemical Society

image processing). We give below a short account of various implementations of these ideas to *NMR* . Note that for all presented methods, the mixing matrix A was estimated at first in different ways then the sources matrix S was estimated using sparsity, pseudo-inverse of A or some post-processing steps.

Sparsity Based Robust Multicomponent Analysis. A recent work was presented by Kopriva and Jeric in [82, 81] very much in the spirit of **SCA** .

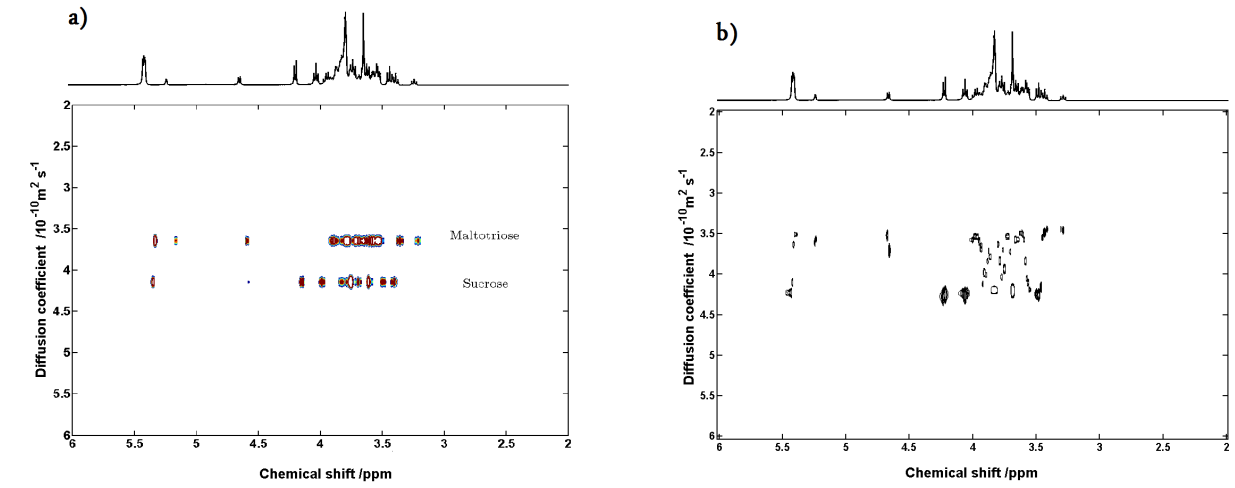

Figure 6: The reconstructed DOSY from JADE separation (a) and the from a monexponential fitting of the peaks in the PFG NMR experiment (b) for the SM mixture

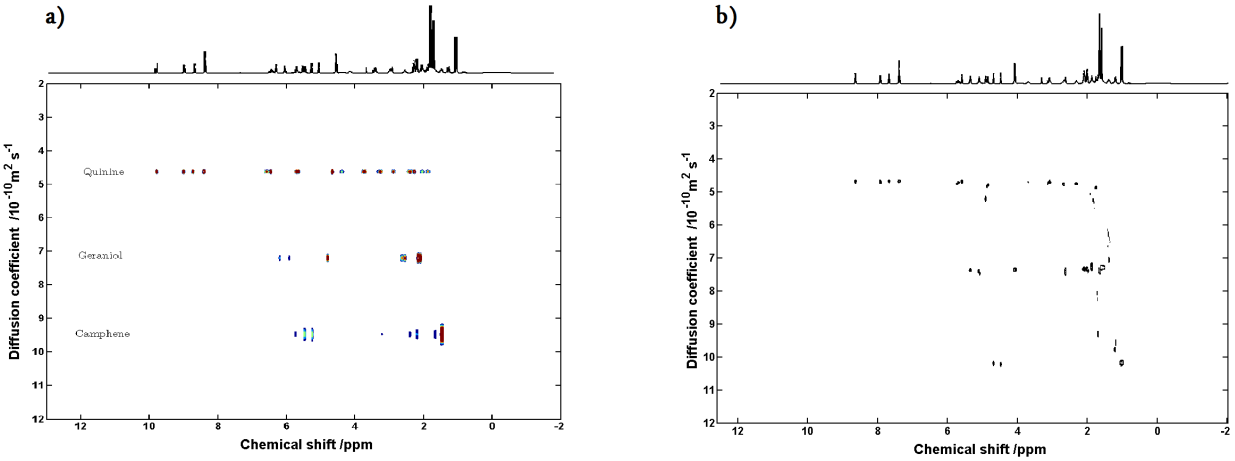

Figure 7: The reconstructed DOSY from JADE separation (a) and the from a monexponential fitting of the peaks in the PFG NMR experiment (b) for the QGC mixture

In addition to **BSS** , the method also features a simple rule for estimating the number k of analytes, no matter if k is less than, equal to, or greater than the number of mixture spectra. To cope with the problem of signal

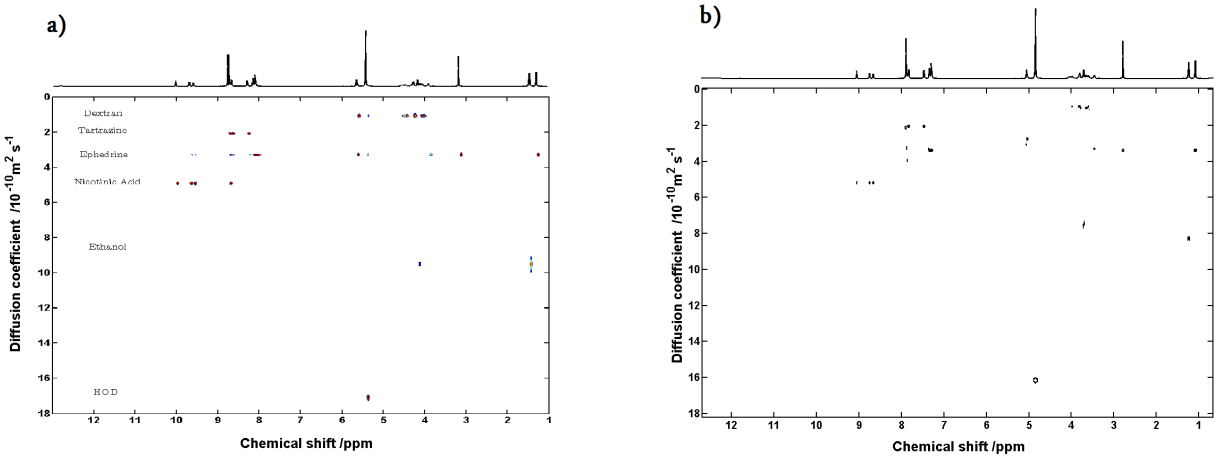

Figure 8: The reconstructed DOSY from JADE separation (a) and the from a monexponential fitting of the peaks in the PFG NMR experiment (b) of the DETENET mixture

overlapping, notoriously difficult in *NMR* spectroscopy, the method relies on the assumption that a specific set of points exists in the representation domain where components to be estimated are mutually sparse.

According to this assumption the authors suggested to rewrite eq. (1) in a new representation domain using a linear transform T so that it becomes:

$$T(X) = A\,T(S) \quad , \tag{10}$$

For example, the linear transform T could be wavelet or Fourier transforms and is applied row-rise to the observation matrix X. The method is then based on three steps:

1. Sparse Representation and Single-Component-Analysis (SAPs):
   After a suitable transformation, determine the points involving only one active analyte (i.e. sample points where analytes are 1-sparse): the so-called *Single Analyte Points*, SAPs for short. The detection can be based upon various techniques, and the authors of [81, 82] focus on a

specific approach suitable for *NMR* spectra.

The SAPs should verify a common assumption which was firstly introduced in *NMR* spectroscopy by Nuzillard [83]: For each source, there is at least one value of the acquisition variable for which only this source presents a non-zero response. More formally, this could be written as:

Assumption (SAP)**.** *For each source* S_i *where* $i \in \{1, \ldots, r\}$*, there exists an* $j_i \in \{1, \ldots, p\}$ *such that* $s_{i,j_i} > 0$ *and* $s_{k,j_i} = 0$ *for* $k = 1, \ldots, i-1, i+1, \ldots, r$*.*

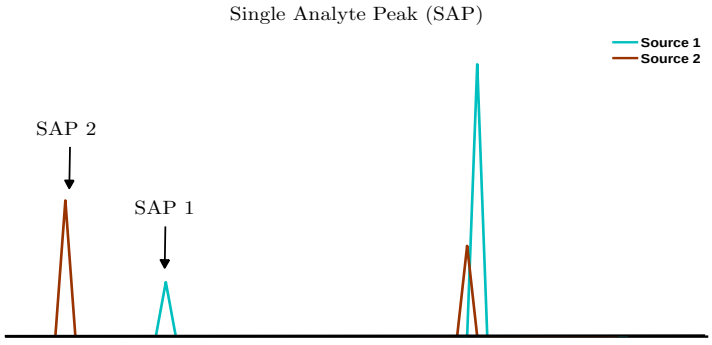

Figure 9: Schematic representation of two sources satisfying the SAP condition

The approach relies on the geometric concept of direction to detect points where single analytes are present. The detection criterion requires complex representation of signals and in the case of *NMR* signals it is applied in the Fourier basis.

2. Data clustering based on estimation of k and mixing matrix A:

Once the set of the *SAPs* is identified, an accurate estimation of the number of analytes k and the mixing matrix A is possible. Seeing that the set of points are 1-sparse, this guarantees that the estimation of A

is unique (if there is no noise) up to permutation and scale. In order to estimate the number of analytes, a clustering function was proposed in [82]:

$$f(a) = \sum_{i=1}^{P} \exp\left(-\frac{d^2(x_i, a)}{2\sigma^2}\right) , \tag{11}$$

where

d is a distance function, defined by $d(x_i, a) = [1 - (x_i \cdot a)^2]^{1/2}$, σ is a scale parameter that defines the resolving power of the function $f(a)$ and $x_i \cdot a$ denotes the inner product and a is the mixing vector in a two-dimensional subspace parameterized as:

$a = [\cos(\phi) \sin(\phi)]^T$ where $\phi \in [0, \frac{\pi}{2}]$ is the mixing angle.

The number of peaks of the function $f(a)$ in the interval $[0, \frac{\pi}{2}]$ provides the desired estimate of the number of analytes k present in the mixture. Once this is done, a mixing matrix $\hat{A}$ is estimated on the same set $SAPs$ using data clustering methods.

3. Estimation of Analytes (source matrix S):

   To estimate the analytes two cases are considered:

   Determined case: $k \leq n$. In this case the matrix of analytes S can be estimated through a simple matrix pseudo inverse: $\hat{S} = \hat{A}^{\sharp} X$.

   Undetermined case: $k > n$. In this case there are more sources than mixtures, and some regularization is needed. In the proposed approach, sparsity assumptions are made again, and the estimation of S is performed via a ℓ^1-regularized least-squares problem, or by linear programming.

The separation method was tested first on ^1H and ^{13}C *NMR* spectroscopy by extracting three pure components from two mixtures.[82] The method

was further validated on more complex cases of study: 2D *COSY* experiments to decompose three mixture of four analytes (glycopeptides) and on mass spectrometry by separating two analytes from the spectra of five mixtures.

In Figure 10, the plots of the *COSY NMR* spectra of pure analytes are shown, to be compared to the estimated ones in Figure 11 [82]. Visual analysis of these spectra reveals that the sources were indeed well sparse and that the estimate correct. In order to show the complexity of the considered case, the authors measured the likeness between the different pure analytes by calculating the correlation between their spectra. The same measure was used to compare the spectra of pure analytes and the spectra of the estimated ones. This indicator proved the effectiveness of the method which turned out to give better results compared to the **JADE ICA** algorithm. This was justified by the author by the fact that the **ICA** model is not really appropriate to these data since the significant correlation between spectra of the pure analytes violates the statistical independence assumption required by **ICA**.

LPBSS Algorithm. The **LPBSS** (Linear Programming BSS) method, also known as the "NN" (for Naanaa and Nuzillard) method, and introduced in [83], exploits non-negativity constraints and the *local orthogonality* principle (SAP), introduced above, to better cope with real life problems. In fact, statistical independence requires uncorrelated source signals, which is not the case all the time seen that is exists molecules whose spectra are known to be correlated in *NMR* Spectroscopy. Therefore, there was a need to use a blind separation methods which integrate more flexible and adequate constraints depending on the physical and chemical origin of the signals.

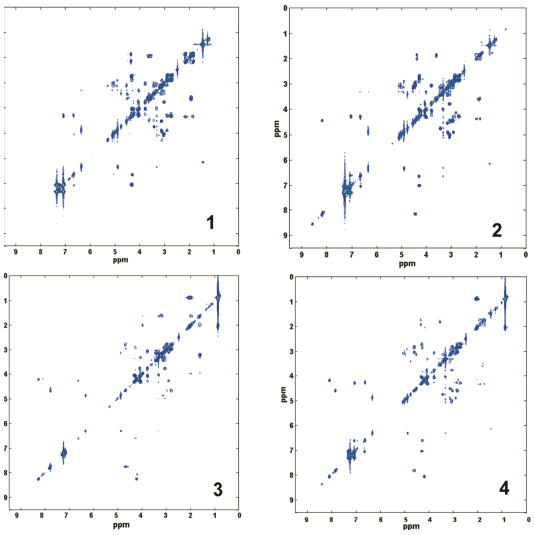

Figure 10: Sparse-based blind source separation. COSY NMR spectra of four glycopeptides, for comparison with the estimated sources in figure 11. Reprinted with permission from Anal. Chem, Vol 82, I. Kopriva and I. Jeric, Blind Separation of Analytes in Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry: Sparseness-Based Robust Multicomponent Analysis, p 1911-1920. Copyright 2010 American Chemical Society

The non-negativity constraint applied to the matrix of source signals S and the local orthogonality constraint is provided through the (SAP) assumption. In this work, only the determined case was considered.

To describe the method and give details on the mathematical steps, it is necessary to use some notations. Given a matrix A, we denote by A^j its j^{th} column, and by $A^{\setminus j}$ the submatrix of A consisting of all columns, but A^j. With these notations, equation (2) reads:

$$X^j = \sum_{k=1}^{r} s_{k,j} A^k, \qquad j = 1, \dots p \ . \tag{12}$$

For the particular subscripts $j_i \in (1 \dots r)$, and based on the (SAP) assumption, the equation collapses to:

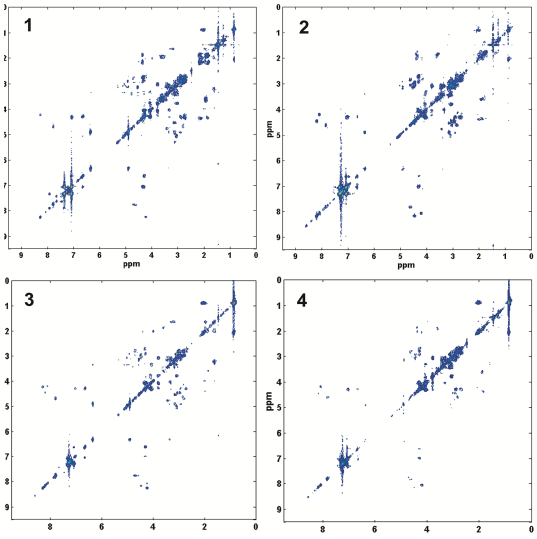

Figure 11: Demonstration of the Sparsity Based Robust Multicomponent Analysis. COSY NMR spectra of the estimated analytes, the spectra of which are shown in figure 10 . Reprinted with permission from Anal. Chem, Vol 82, I. Kopriva and I. Jeric, Blind Separation of Analytes in Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry: Sparseness-Based Robust Multicomponent Analysis, p 1911-1920. Copyright 2010 American Chemical Society

$$X^{j_i} = s_{i,j_i} A^i, \qquad i = 1, \ldots r \tag{13}$$

That means that every column of A is colinear to a column of X locally, as one source only is present in this frequency range. By replacing each A^k in (12) from (13) one obtains:

$$X^j = \sum_{i=1}^{r} \frac{s_{i,j}}{s_{i,j_i}} X^{j_i}, \qquad (1 \leqslant i \leqslant r, 1 \leqslant j \leqslant p) \ . \tag{14}$$

Assume that $\hat{X}$ consists of all the mutually non-colinear columns of X then we note $\hat{A}$, the submatrix of $\hat{X}$ consisting of r columns each of them is colinear to a particular column of A.

According to one property, a column of $\hat{X}$ is selected to form $\hat{A}$ if it is not a non negative linear combination of the other columns of $\hat{X}$. This identification may be achieved by considering the following equations system:

$$X^{\setminus j}\alpha(j) = \hat{X}^j, \quad \alpha(j) \geq 0 \ , \tag{15}$$

where $\alpha(j)$ denotes an unknown column vector. The algorithm consists in solving the following optimization problem by using a linear programming technique:

$$\hat{X} = \arg\min_{\alpha_i(j)} \left\| \hat{X}^{\setminus j}\alpha(j) - \hat{X}^j \right\|, \qquad i = 1, \ldots n, j = 1, \ldots p \ . \tag{16}$$

where $\alpha_i(j)$ denotes one of the components of the vector $\alpha(j)$.

Hence, a score is computed for each $\hat{X}^j$ in order to find the columns from $\hat{X}$ that will form $\hat{A}$:

$$\mathsf{score}_j = \left\| \hat{X}^{\setminus j}\alpha^*(j) - \hat{X}^j \right\| \tag{17}$$

If the score is low, it is unlikely that the considered column is a non-negative linear combination of the other columns forming the $X^{\setminus j}$ and therefore it is unlikely a column of $\hat{A}$. The inverse means that the involved column may be a column of $\hat{A}$. The $\hat{A}$ is formed from the n columns of $\hat{X}$ associated to high calculated scores.

Once the matrix $\hat{A}$ is formed, each column of it is replaced by the average of all columns in X that are approximately colinear to it.

Finally an estimate $\hat{S}$ of S is obtained as before using the Moore-Penrose pseudo-inverse $\hat{A}^\sharp$ of $\hat{A}$, via $\hat{S} = \hat{A}^{\#}X$.

The method was tested on two different datasets. The first a PFG-NMR

experiment realized on a mixture of two organic compounds, menthol and β-sitosterol.

As illustrated in Figure 12, the separation was achieved but with the persistence of some small artifacts that can be singled out by comparison with the reference spectra of the two pure components.

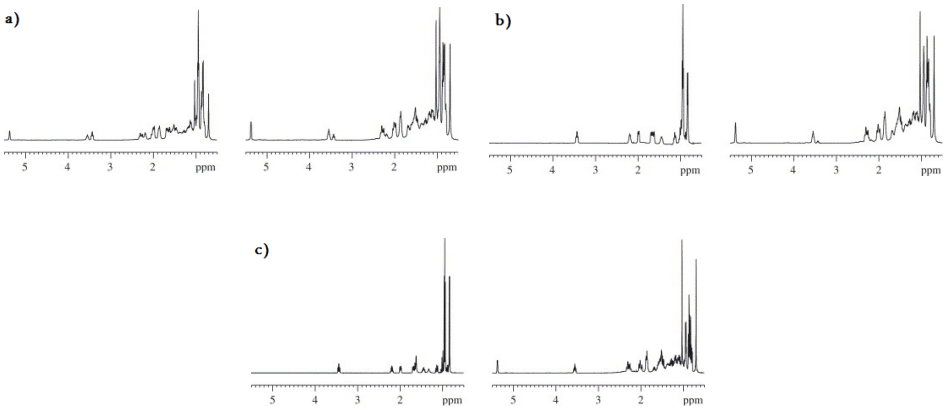

Figure 12: Demonstration of LPBSS. (a): Diffusion-modulated spectra of a menthol-β-sitosterol mixture obtained for two magnetic field gradient strengths; (b): Calculated source spectra: menthol (left),β-sitosterol (right); (c): Reference spectra. Left: menthol,right: β-sitosterol. Adapted from Signal Processing, vol. 85, W. Naanaa and J.-M. Nuzillard, Blind source separation of positive and partially correlated data, p 1711-1722. Copyright 2005, with permission from Elsevier

Further tests were performed on four synthetic mixtures obtained from the spectra of menthol, β-sitosterol, mannitol and β-cyclodextrine with addition of white Gaussian noise with $SNR = 15\,dB$ (see figure 13). To gain insight into the separating power of **BSS** , a comparison was done between **LPBSS** , **SOBI** , **Fast ICA** and **JADE** , on the basis of two performance measures: *Comon*'s [84] and *Choi*'s [3] indexes, showing a predominance of the **LPBSS** algorithm. The definition of these indexes, along with their use for an estimation of the performance of the methods and their comparison

will be discussed more in details later on.

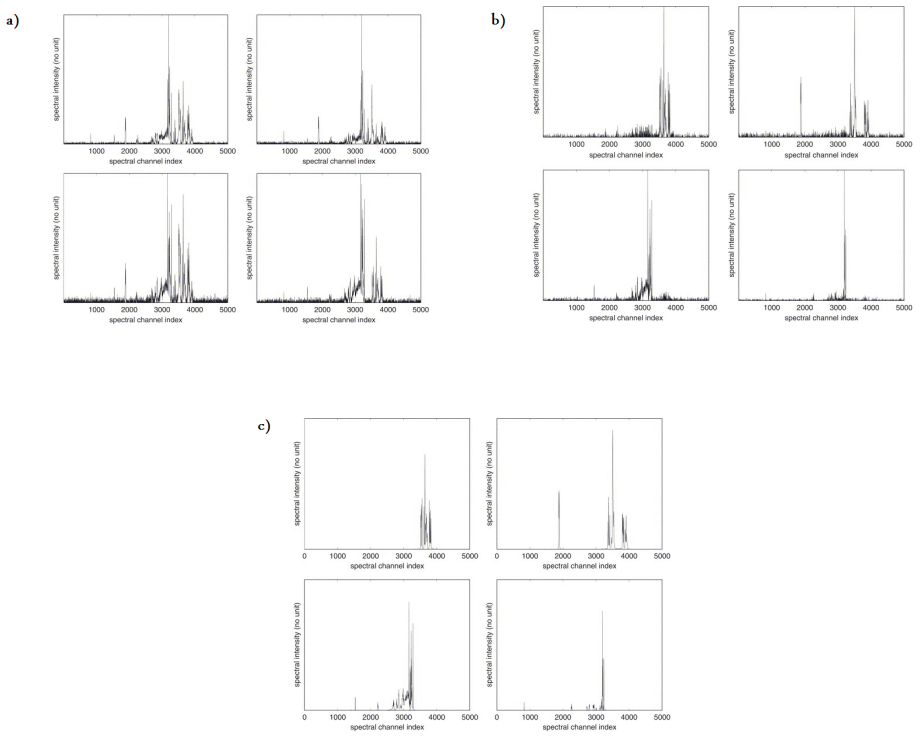

Figure 13: Demonstration of LPBSS. (a): Four simulated mixtures obtained by combining the spectra of menthol, β-sitosterol,mannitol,and β-cyclodextrine and adding noise for a $SNR = 15dB$; (b): Source spectra computed by the **LPBSS** algorithm; (c): Reference spectra: mannitol, β-cyclodextrine, β-sitosterol,and menthol. Adapted from Signal Processing, vol. 85, W. Naanaa and J.-M. Nuzillard, Blind source separation of positive and partially correlated data, p 1711-1722. Copyright 2005, with permission from Elsevier

An improvement of this method was developed later by Y.Sun et al in [85], who introduced a relaxed SAP condition which basically assumed the existence of points where a given source dominates all the others:

Assumption (rSAP)**.** *For each Source* S_i *where* $i \in \{1, \ldots, r\}$ *there exists an* $j_i \in \{1, \ldots, p\}$ *such that* $s_{i,j_i} > 0$ *and for* $k = \{1, \ldots, r\}$ *and* $k \neq i$,

$s_{k,j_i} \ll s_{i,j_i}$.

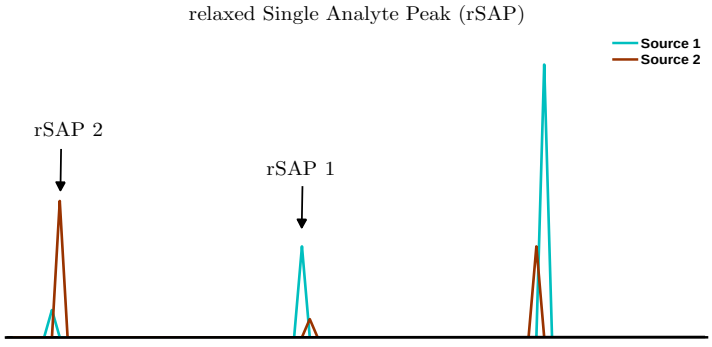

Figure 14: Schematic representation of two sources satisfying the rSAP condition

In a nutshell, each source signal has a dominant peak at one acquisition position where the other sources are small (instead of zero as in the *SAP* condition).

Hence, it is considered as a generalization of the **LPBSS** method for more complicated cases where the *SAP* condition does not hold. The method consists of applying the **LPBSS** algorithm first and then post processing the output to reduce its errors. The post processing is done by using:

- Random error detection method to perform the output source matrix S by discarding the incoherent components [85],
- Peak-based correction step which aims to extract a better estimation of the mixing matrix A by imposing a pairwise overlap condition (POC) on the source signals as follows:

Assumption (POC)**.** *Each source signal has a dominant peak at some acquisition location where other source signals are allowed to be nonzero. Fur-*

thermore, there exist different acquisition regions where the source signals overlap each other pairwise.

In order to test the effectiveness of the proposed enhancements, two synthetic *NMR* datasets and one real mixture were used. For both synthetic mixtures, the source spectra were mixed according to the model $X = AS$. The first dataset included two mixtures issued from two sources and the second included three mixtures issued from three sources. The used real world-data was a mixture of Camphor and Quinine analysed in the PFG-NMR.

The results for the real *NMR* data are shown in Fig. 15 where Fig. 15.a is corresponding to the mixtures, the reference spectra of camphor and quinine are shown in 15.d and the recovered source spectra by **LPBSS** method (referred here as NN) and **PBC** (Peak-based correction) method are exposed respectively in Fig. 15.b and Fig. 15.c.

The **LPBSS** separation results were rather good, especially the spectrum of Quinine (Fig. 15.b), but one could notice the presence of remarkable residues in both spectra. The presence of these residues is due, according to the authors, to the large peaks of Camphor. However, in the figure (15.c) we can see that with the peak-based correction (PBC) the artifact is reduced considerably.

rBSS method. In many *NMR* spectra, most particularly biologically relevant samples such as biofluids, broad signals from macromolecules are typically coexisting and overlapping with narrow resonances from smaller metabolites. In this condition, the dominance of a source in a frequency interval must be characterized in a more subtle fashion. Hence, Sun and Xin addressed this issue [86] by relaxing further the assumption of *SAP* to take care of these

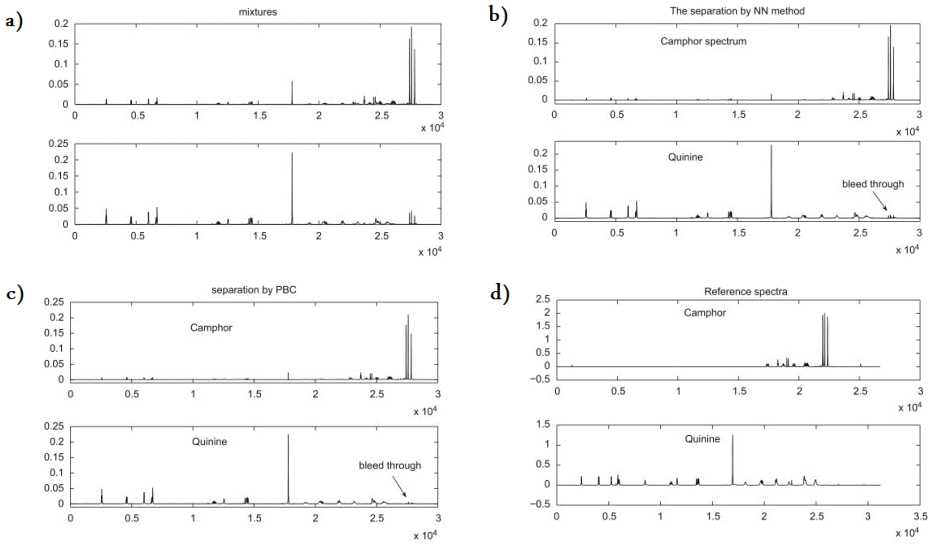

Figure 15: Demonstration of LPBSS (a): Mixtures of Camphor and Quinine, (b): Camphor and Quinine spectra recovered by LPBSS method; (c): Camphor and Quinine spectra recovered by PBC, (d): Reference spectra of camphor and quinine. Adapted from Signal Processing, vol.91, Y. Sun, C. Ridge, F. del Rio, A. J. Shaka and J. Xin, Postprocessing and sparse blind source separation of positive and partially overlapped data, p 1838-1851. Copyright 2011, with permission from Elsevier

specific sets of problems.

This is formulated by the following assumption called the *Dominant Interval condition* (*DI*) which basically states that each source S_i where $i = 2, 3, \ldots, n$, is allowed to have dominant interval(s) over $S_{i-1}, \ldots, S_2, S_1$, while other part of S_i may overlap with $S_{i-1}, \ldots, S_2, S_1$:

Assumption (DI)**.** *For each* $k \in 1, \ldots, r$ *there is a set* $\mathcal{L}_k \subset 2, 3, \ldots, p$ *such that for each* $l \in \mathcal{L}_k s_{il} \gg s_{jl}$ *for* $(i = k, k+1, \ldots, r, j = 1, \ldots, k-1)$.

A schematic representation of a two sources example is given in Figure 16 where we notice that source 1 has a dominant region R1 while source 2 dominates in region R2.

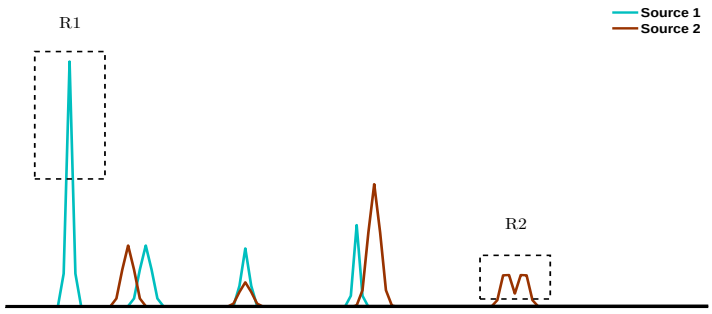

Figure 16: Schematic representation of two sources satisfying the DI condition

The method consists of two major steps: a the backward step in order to reduce the separation problem to a series of sub-**BSS** problems and a forward step to recover the sources. The number of mixtures is supposed to be equal to the number of sources to estimate. More explicitly in the backward step, the columns of X were written based on the the DI assumption as:

$$X^k = s_{r,k}A^r + \sum_{i=1}^{r-1} o_{i,k}A^i \ , \tag{18}$$

where $s_{r,k} \gg o_{i,k}$ for $i = 1, \ldots r-1$.

From this equation, it was noticed that A^r is equivalent to finding a cluster formed by these X^k's in $\mathbb{R}^r$. So to estimate A^r, it was obvious to determine the set of vector columns X^k that cannot be written as linear combinations of the other vectors, containing in $= X^1, X^2, \ldots, X^P$. The set of the X^k vectors are contained in a frame and among all the elements of the frame, A^r is the one attracting a cluster. To solve this, linear programming can be used. Once the A^r is obtained, eliminating S_r from X reduces the model so that a new mixture matrix is formed as:

$$X_{1,2,\ldots,r-1} = \begin{pmatrix} X_1 - \frac{A_{1r}}{A_{rr}} X_r \\ X_2 - \frac{A_{2r}}{A_{rr}} X_r \\ \vdots \\ X_{r-1} - \frac{A_{r-1,r}}{A_{rr}} X_r \end{pmatrix} \in \mathbb{R}^{(r-1)\times p} \tag{19}$$

The reduced **BSS** model could be written as:

$$X_{1,2,\ldots,r-1} = \tilde{A}^{(1,2,\ldots,r-1)} S_{(1,2,\ldots,r-1)} \tag{20}$$

In this new set of mixtures $X_{(1,2\ldots r-1)}$, source S_{r-1} has dominant intervals over other sources. Therefore, the data clustering and linear programming could be used to recover the mixing coefficients of S_{r-1} from $X_{(1,2\ldots r-1)}$.

Then for $k \leqslant r-1$, this procedure combined with mixtures reduction is repeated in a recursive manner until source S_1 is obtained.

In summary, the backward step allows the extraction the source signal S_1 as well as a series of reduced mixtures $X_{1,2}, X_{1,2,3}, \ldots X_{1,2\ldots k}, X_{1,2\ldots r-1}$.

The forward step comes at a second moment to recover the rest of sources from S_2 to S_r. In order to simplify the problem, the source signals are supposed to be sparse in some transformed domain. Therefore the *NMR* spectrum is considered as a linear convolution of a Lorentzian kernel with some sparse function consisting in a few peaks. The source signal could be written as follows:

$$S = \hat{S} * \mathcal{L}_\omega \ , \tag{21}$$

where $\hat{S}$ is the sparse function and $\mathcal{L}_\omega$ is the Lorentzian function with width ω.

According to this, to recover S_k sources for $k = 2 \dots r-1$, the authors proposed to resolve the following l^1 minimization problem:

$$\min_{\substack{0 \leqslant A \in \mathbb{R}^{k \times (k-1)} \\ \hat{S} \in \mathbb{R}^{k \times p}, \hat{S} \geqslant 0}} \left[\mu \left\| \hat{S} \right\|_1 + \frac{1}{2} \left\| X_{(1,2,\dots,k)} - A^{(1,2,\dots,k-1)} S_{(1,2,\dots,k-1)} - \hat{S} * \mathcal{L}_{\omega_k} \right\|_2^2 \right], \tag{22}$$

where the rows of $\hat{S} * \mathcal{L}_{\omega_k}$ are the multiples of source S_k in $X_{(1,2,\dots,k)}$ and ω_k is the peak width of S_k.

The equation is solved by using a projected gradient descent approach for its simplicity and then sources S_k for $k = 1 \dots r-1$ are retrieved. Finally the last source S_r is separated by minimizing the same equation problem but with replacing 'k' by 'r'.

The method was tested on three datasets: two synthetic and one real world *NMR* spectroscopy. The first example includes the separation of three sources from three mixtures. knowing that the shape of the peaks differs between the different sources (narrow, wide, very wide), we can see from the illustration in Fig.17 that the spectra of the three sources were well separated. However, the linewidth for the sources was estimated directly on the spectra.

More examples were produced on simple mixtures. Recently, the same authors expanded the method [87] to separate non-negative and correlated data in mixtures. The motivation was the separation of *NMR* spectra of biofluids such as urine and blood for metabolic fingerprinting and disease diagnosis. They considered the following assumption:

Assumption. *Consider the over-determined case where n sources are to be separated from $m \geq n$ mixtures. Among the n source signals, there are $n-1$*

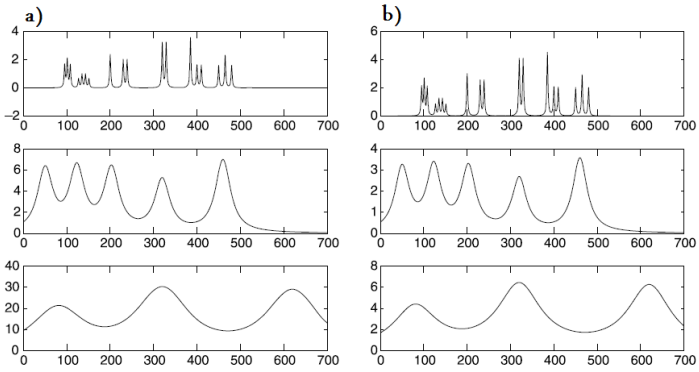

Figure 17: Demonstraiton of the forward step of rBSS on an artificail mixture (a): the recovered sources by $\ell_1$minimization. (b): is the reference spectra. Reprinted from Journal of Scientific Computing, vol. 51, Y. Sun and J. Xin, A Recursive Sparse Blind Source Separation Method and its Application to Correlated Data in NMR Spectroscopy of Biofluids, p 733-753. Copyright 2012 Springer-Verlag Berlin Heidelberg

partially overlapping (PO) sources assumed to satisfy (SAP) and one positive everywhere (Pe) source which is required to have dominant interval(s) (DI). Consider the over-determined case where n sources are to be separated from $m \geq n$ mixtures. Among the n source signals, there are $n-1$ partially overlapping (PO) sources assumed to satisfy (SAP) and one positive everywhere (Pe) source which is required to have dominant interval(s) (DI).

The mathematical challenge of the problem here is that the ideal standalone peak (SAP) [83] is again not satisfied since the *NMR* spectra of biofluids contain both wide-peak (e.g. proteins) and narrow-peak sources and that the latter ones could dominate the wide-peak signal in intensity. The method consists on three steps:

- Identifying the mixing coefficients of the (Pe) source (the broad one) by exploiting geometry in data clustering so that the (Pe) sources is

eliminated for the next step.

- New mixtures containing only the (PO) sources (the narrow ones) are constructed from the previous step, for which the convex cone method and related linear programming are applied.

- Solving a convex ℓ^1 minimization problem to extract the (Pe) source signals.

The method was applied to three synthetic datasets and to real-world data produced by $DOSY$, a mixture of quinine, geraniol and camphor. The separation was satisfactory especially on the three synthetic datasets but on the real-world dataset.

3.1.3. Variational approaches

The last family of approaches we shall account for here relies on the joint numerical optimization of some objective function (i.e. with respect to both source and mixing matrices), that generally involves both a data fidelity term $D(X\backslash AS)$ and a regularization term $\Psi(A, S)$, implementing some prior information about the sources:

$$\Phi(A, S) = D(X\backslash AS) + \Psi(A, S) \ , \tag{23}$$

complemented by additional constraints (such as non-negativity).

Several approaches have been proposed in the literature, that involve various choices for the data fidelity term D and the prior term Ψ, as well as different numerical optimization strategies. We note in passing that, in the context of chemometrics, the specific approach called Multivariate Curve Resolution (MCR) has essentially the same goals as **BSS** . Indeed,

its version based on ALS (MCR-ALS) [88] or on gradient descents (MCR-NLR) [89, 90, 91] have been tested as an alternative *DOSY* processing.

Non-Negative Matrix Factorization (NMF). **NMF** refers to a category of approaches for decomposing a matrix with non-negative coefficients as a product of two matrices with non negative coefficients. **NMF** has been proved to be a useful multivariate data decomposition technique in various contexts where one has to deal with nonnegative data. It is therefore a relevant approach for instantaneous **BSS** when observations, sources and mixing matrices are non-negative, which is the case for *NMR* spectroscopy.

Mathematically speaking, the **NMF** problem can be written as follows: Given a non negative $m \times n$ matrix X as in model (2), compute a non negative $m \times r$ matrix A and a nn negative $r \times n$ matrix S where $r \ll m, n$ such that: $X \approx A\, S$.

NMF is generally formulated as a minimization problem with bound constraints since it attempts to minimize an objective function representing the difference between the original data X and the approximation AS:

$$(\hat{A}, \hat{S}) = \arg\min_{A \geq 0, S \geq 0} D(X \backslash AS) \ , \tag{24}$$

where $D(X \backslash AS)$ is a separable measure of fit (often called a divergence) of the form:

$$D(X \backslash AS) = \sum_{i=1}^{m} \sum_{j=1}^{n} d([X]_{ij} \backslash [AS]_{ij}) \tag{25}$$

and $d(x \backslash y)$ is a scalar cost function.

The most frequently used divergence measure is the so-called quadratic loss:

$$D(X\backslash AS) = \frac{1}{2}\left\|X - AS\right\|_F^2 = \frac{1}{2}\sum_i\sum_j (X_{ij} - [AS]_{ij})^2 \ , \qquad (26)$$

but several other choices have been proposed in the literature, which we will also discuss below.

It is important to stress that criteria such as the criterion in (26) are generally non-convex (even though they are can be convex with respect to A and S separately, they are not convex with respect to the pair (A, S)). Therefore, most optimization techniques cannot guarantee to yield a global optimum, and care is needed with initialization.

Most approaches rely on alternate optimization with respect to A and S, that therefore update alternatively the mixing and the source matrices. There are several possible approaches, that exploit different updates rules. A simple example is the so-called ALS (alternating least square) method proposed by Paatero in [92] for the quadratic loss function in (26). The optimization with respect to both A and S has a closed form solution, which is used in an iterative algorithm, together with a projection step to enforce non-negativity. ALS-type approaches are considered computationally expensive but seem to be quite robust. They are also limited to situations where a closed form expression for the updates of A and S are available.

A second class of **NMF** algorithms exploit classical gradient descent techniques, as discussed and used by Chih in [93], still in the case of the quadratic loss. Gradient-based methods are perhaps the simplest techniques to implement but the convergence is often somewhat slow compared to the other methods. A standard shortcoming of gradient based methods is their sensitivity to the choice of stepsize. Adaptive stepsize techniques can be developed, but these can be hard to tune. For these reasons, other approaches

are often preferred.

An application of these algorithms to a *TOCSY* spectrum of a mixture of seven common metabolites was done by Snyder *et al* in [94]. They proposed to use the (**PCA**) Principal Component Analysis method to estimate an approximate number of components and then **NMF** algorithm is applied with variations of this number. According to the paper, if the estimated number of components was less than the real one, the peaks coming from different components and that overlapped partially in the spectrum are represented by a single component and if it was the inverse then duplicate components occur with closely related peak patterns representing the same source.

The so-called multiplicative algorithms have enjoyed significant popularity since their introduction in the seminal paper of Lee and Seung [95]. They are a very good compromise between speed and ease of implementation, and have the advantage of automatically satisfying the constraint. Also, they have recently been shown to rely on particular cases of the so-called *majoration-minimization* (MM) algorithms, a fairly classical family of methods in non-convex optimization. Still, multiplicative algorithms exploit alternate MM-type optimizations with respect to A and S, and there is no proof showing that any limit point is a stationary point of the objective function.

Modifications of the original **NMF** algorithm have been proposed, for example by Lin in [96] and Sajda *et al* in [97] who introduced the "**cNMF**" algorithm to deal with negative observations by assuming that they arise from the noise distribution. A quadratic cost function was used and a threshold constraint is added by forcing the negative values of S to be approximately zero and such the mixing matrix A will be treated symmetrically in the same

manner.

The **NMF** algorithms discussed above can be generalized also to involve different choices of the cost functions Φ and Ψ in eq. (23), that may be better suited for real world data. In the standard approach, $\Psi = 0$ and Φ is a quadratic function, which implicitely assumes white Gaussian noise. To account for other noise models, the function $d(x\backslash y)$ in eq. (25) can be replaced with other divergence functions.

Prior information on the sources S and the mixing matrix A can also be introduced in the regularization term Ψ. We shall describe in some details a variant (*sparse* ***NMF***) enforcing sparsity on the sources by taking for Ψ some ℓ^1 norm of the sources. Non-negativity condtrains introduced in MCR *DOSY* processing did produce somewhat better separations [89, 90, 91].

Sparse NMF To enhance the decomposition of multivariate data, prior information about the sources and mixing matrix can be exploited. We have seen already that sparsity seems to be a relevant paradigm for *NMR* spectra. Sparsity can be introduced in different ways into the **NMF** approach, corresponding algorithms are generically termed *sparse-**NMF*** .

The sparse **NMF** was introduced by P. Hoyer in [98], where the sparsity is enforced by taking for Ψ in (23) an ℓ^1 prior term:

$$\Psi(A, S) = \sum_r \|S_r\|_1 \ .$$

The algorithm therefore looks for a minimizer of an objective function composed by a quadratic data fidelity term, and an ℓ^1 prior:

$$\Phi(A, S) = \|X - AS\|_F^2 + \lambda \sum_r \|S_r\|_1 \ , \tag{27}$$

where λ is a positive regularization constant which allows controlling the sparsity rate for the components to be estimated. The proposed algorithm, termed **NNSC** "Non-Negative Sparse Coding", combines a projected gradient step for updating A with a MM-based multiplicative step for updating the sources S.

As an alternative, sparsity can also be introduced as a strict constraint. The sparse **NMF** introduced by P. Hoyer in [99] introduces a sparsity measure defined by:

$$\mathsf{sparsity}(a) = \frac{\sqrt{n} - (\sum |a_i|)/\sqrt{\sum a_i^2})}{\sqrt{n} - 1} \ , \tag{28}$$

where n is the dimensionality of a. The sparsity constraint can be imposed on either A or S, or both. The optimization algorithm goes along the same lines as **NNSC** .

Since the use of this approach requires some specific knowledge about the sparsity of the sources, it is probably more complex for *NMR* applications, and was never applied to this problem.
We applied recently **NNSC** to the unmixing of mixtures, using PFG-NMR datasets described in [79]. The results are shown in figures: 18 for **SM** , figure 19 for **QGC** and figure 20 for **DENET** , using the Matlab code published in [100]:

The reconstructed *DOSY* (following the procedure described before in the JADE section) made from estimated matrices S and A by **NNSC** for all datasets are presented in figures 21, 22 and 23.

3.2. Tensor based methods (PARAFAC)

Introduction to PARAFAC model:. **PARAFAC** (parallel factor analysis) is considered as a generalization of **PCA** to higher order arrays. In the case of

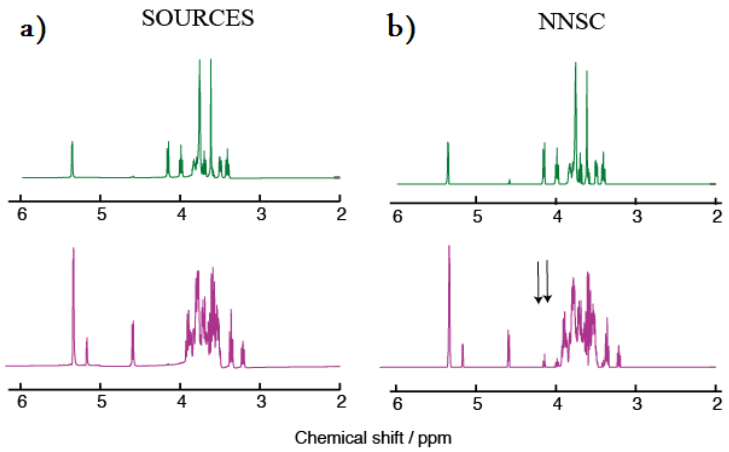

Figure 18: The sources recovered by NNSC with gradient stepsize $\delta = 510 - 06$ and $\lambda = 30$ (a) and the ground truth sources (b) forthe mixture of Maltotriose and Sucrose, analyzed with a PFG-NMR series. Reprinted with permission from Anal. Chem, Vol 85, Toumi. I, Torrésani.B and Caldarelli.S, Effective Processing of Pulse Field Gradient NMR of Mixtures by Blind Source Separation, p 11344-11351. Copyright 2013 American Chemical Society

an three-way data analysis, a decomposition of the data is made into triads or trilinear components, but instead of one score and one loading vector as in bilinear **PCA**, each component consists of one score and two loading vectors [101] [102]. According to the nature of data, additional restrictions, such as non negativity and orthogonality can be applied for all/some of the modes.

The model was proposed earlier by Harshman [103] and Carroll & Chang [104] who named the model CANDECOMP (canonical decomposition) which is a generalization of the matrix singular value decomposition (SVD) to tensors [1]. Mathematically, it is a straighforward generalization of the bilinear model of factor (or component) analysis to a trilinear one following this expression:

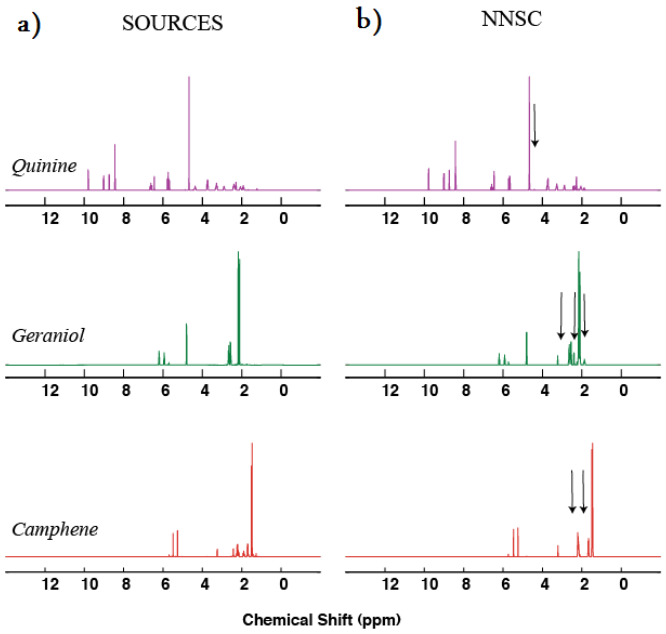

Figure 19: The recovered sources by NNSC with gradient stepsize $\delta = 5.e-05$ and $\lambda = 40$ (a) and the ground truth sources (b) for for Quinine, Camphene and Geraniol. Reprinted with permission from Anal. Chem, Vol 85, Toumi. I, Torrésani.B and Caldarelli.S, Effective Processing of Pulse Field Gradient NMR of Mixtures by Blind Source Separation, p 11344-11351. Copyright 2013 American Chemical Society

$$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk} \ , \tag{29}$$

with an associated sum-of-squares loss:

$$\min_{A,B,C} \sum_{ijk} \left[x_{ijk} - \sum_{r}^{R} a_{ir} b_{jr} c_{kr} \right]^2 \ . \tag{30}$$

Here, x_{ijk} is an entry of a three-way array X with modes A, B and C. the a_{ir} gives the weight or loading of factor r on level i of mode A; b_{ir} and c_{kr} give the weight or loading of the same factor on level j of mode B and

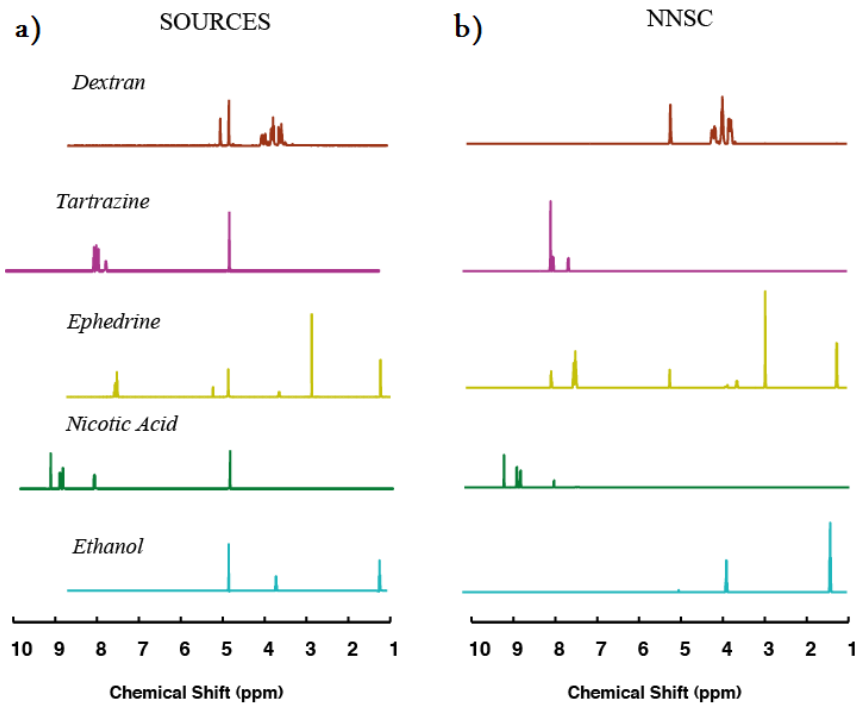

Figure 20: The sources recovered by NNSC with gradient stepsize $\delta = 510-05$ and $\lambda = 20$ (for the regions between 1 and 6 ppm) and $\lambda = 40$ (for the regions between 6 and 10 ppm) (a) and the ground truth sources (b) for the mixture of Dextran, Tartrazine, Ephedrine, Nicotinic Acid and Ethanol, recorded as a series of PFG-NMR spectra. Reprinted with permission from Anal. Chem, Vol 85, Toumi. I, Torrésani.B and Caldarelli.S, Effective Processing of Pulse Field Gradient NMR of Mixtures by Blind Source Separation, p 11344-11351. Copyright 2013 American Chemical Society

level k of mode C, respectively; e_{ijk} is the residual or error term.

The model can be directly fitted to a three-way array of observations with factorial structure, or it can be indirectly fit to the original observations by using a set of covariance matrices computed from the observations, with each matrix corresponding to a two-way subset of the data [105].

The fitting method used for **PARAFAC** is again the Alternating Least

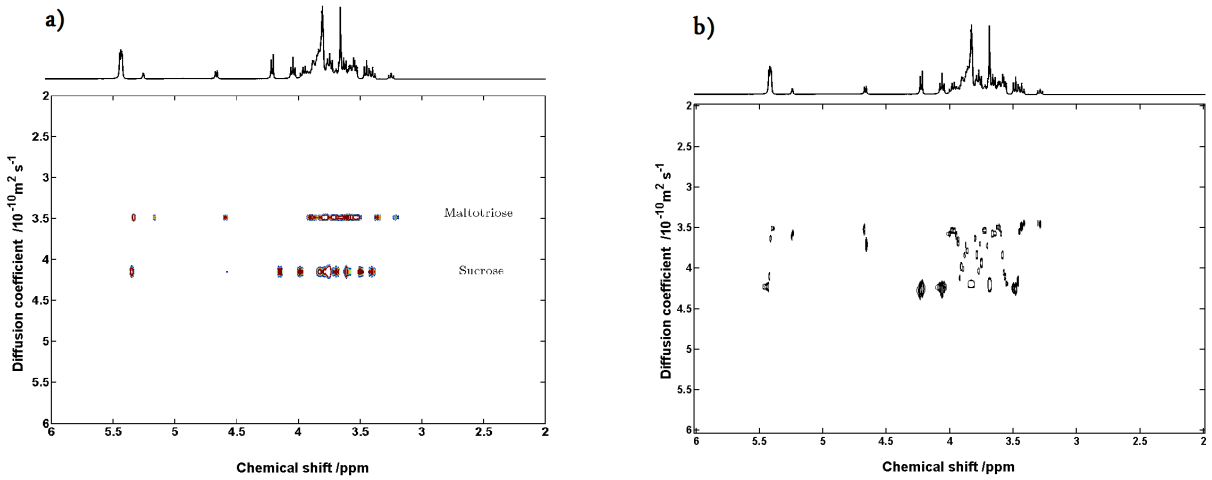

Figure 21: The DOSY reconstructed from NNSC separation (left) and the equivalent monoexponential fitting (right) of sugars mixtures (gradient stepsize $\delta = 5.e-06$ and $\lambda = 30$)

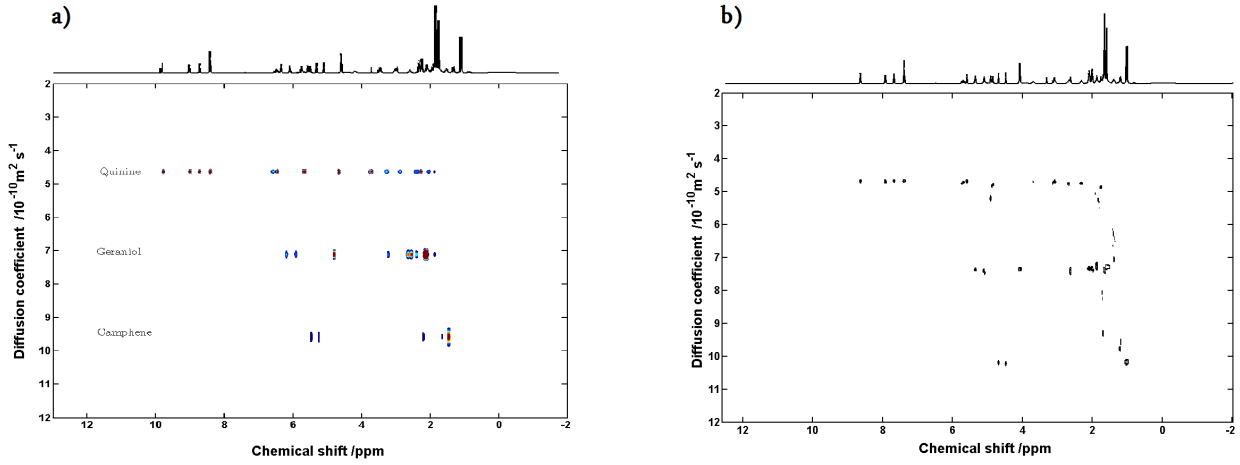

Figure 22: The reconstructed DOSY from NNSC separation (left) and the equivalent monoexponential fitting (right) of the QGC mixture (gradient stepsize $\delta = 5.e-05$ and $\lambda = 40$)

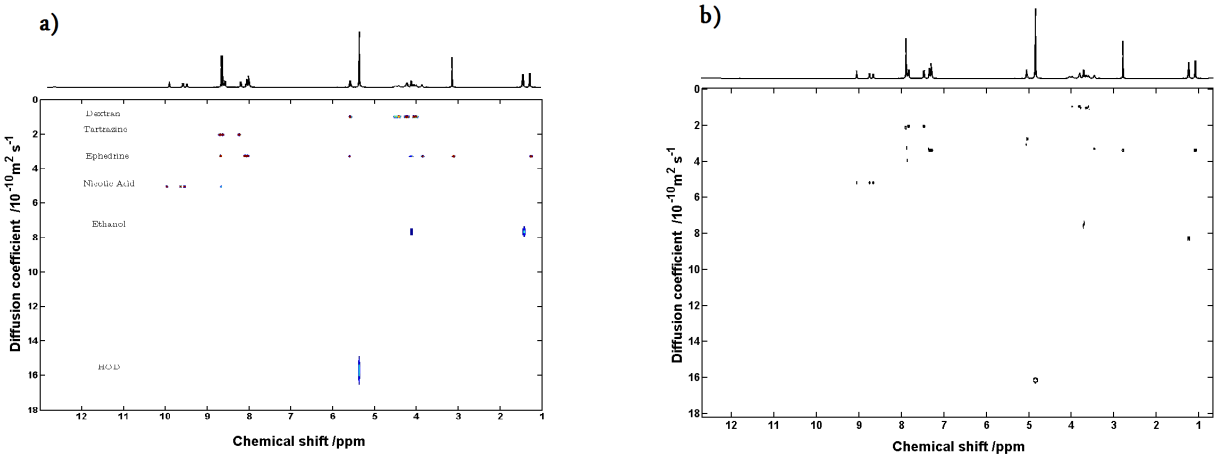

Figure 23: The DOSY reconstructed from NNSC separation (left) and the one obtained with a monoexponential fitting (right) ofthe DENET mixture (gradient stepsize $\delta = 5.e-05$ and $\lambda = 20$ (for the region between 1 and 6 ppm) and $\lambda = 40$ (for the region between 6 and 10 ppm))

Squares. The trilinear model is broken up into three sets of parameters, such that it is linear in each set given fixed values for the other two sets. An obvious advantage of the **PARAFAC** model is the uniqueness of the solution for the reasons explained in [105].

The general **PARAFAC** ALS algorithm follows these steps:

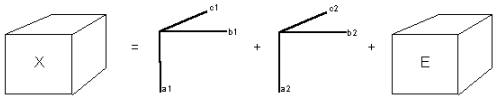

Figure 24: A graphical representation of a two-component PARAFAC model of the data array X (R=2). Reprinted from Analytica Chimica Acta, vol.531, M. Dyrby et al, Analysis of lipoproteins using 2D diffusion-edited NMR spectroscopy and multi-way chemometrics, p 209-216. Copyright 2005, with permission from Elsevier

1. Decide on the number of components, R
2. Initialize B and C
3. Estimate A from X, B and C by least squares regression
4. Estimate B likewise
5. Estimate C likewise
6. Continue from 3 until convergence (little change in fit or loadings).

More details of the algorithm can be found in [101].

Application of PARAFAC to NMR spectrosopy:. The first application of multi-way calibration by N-PLS (*N-way partial least squares*) and multi-way curve resolution by PARAFAC to 2D diffusion-edited ^1H *NMR* spectra was presented in the paper of Dyrby *et al* [106]. The aim of the analysis was to evaluate the potential for quantification of lipoprotein in human plasma samples using these methods since the lipoprotein spectrum presents many overlapping signals and very small differences in diffusion coefficients, which make the full separation with 2D diffusion-edited *NMR* spectroscopy almost impossible.

PARAFAC was used on 2D diffusion-edited *NMR* data of a plasma sample containing 24 spectra. **PARAFAC** models using two to four components were generally informative and provided a good fit to the data. Non negativity constraints were considered for the anaysis on all modes. The next figure shows the best result obtained in this work, which was based on the methylene signal (1.31-1.20 ppm) only and using four **PARAFAC** components:

The figure 25 (A) shows four smooth spectral loadings that are very similar NMR spectra but have different diffusion coefficients, corresponding tentatively to lipoproteins of four different sizes. The four diffusion loadings

showed in Fig. 25 (B) correspond to the diffusion curves of the four spectral loadings in the 25 (A). Although the separation looked correct, the corresponding concentrations of the four **PARAFAC** components did not match the reference concentrations as determined by ultracentrifugation, which was tentatively ascribed to the continous density profiles of lipoproteins.

Forshed *et al* [107] came later to present a method to enhance the multivariate data interpretation of metabolic profiles which was done by correlation scaling of ^1H *NMR* data by the time pattern of drug metabolite peaks identified by LC/MS, followed by **PARAFAC**. A different application of PARAFAC in order to do the metabolic profiling based on the two-Dimensional J-resolved ^1H *NMR* was presented in [108].

Montoliu *et al* [109] applied unsupervised chemometrics for integrating ^1H *NMR* metabolic profiles from mouse plasma, liver, pancreas, adrenal gland and kidney cortex matrices in order to infer intercompartments functional links. Since (**PCA**) and multiway **PCA** do not offer enough information on intercompartment metabolic relationships, integration of metabolic

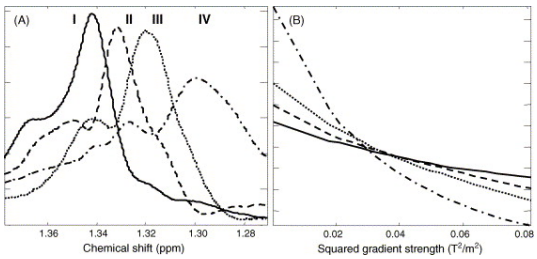

Figure 25: Result of a PARAFAC model with four components on the 2D diffusion-edited NMR spectrum of the methylene peak of lipoprotein lipids: (A) spectral loadings and (B) diffusion loadings. Reprinted from Analytica Chimica Acta, vol.531, M. Dyrby et al., Analysis of lipoproteins using 2D diffusion-edited NMR spectroscopy and multi-way chemometrics, p 209-216. Copyright 2005, with permission from Elsevier

profiles using (**MCR**) and (**PARAFAC**) enabled the characterization of compartment-specific metabolite signatures. This was the first application of these methods in a metabonomic description of intercompartmental functional relationships.

Trilinear analysis was applied in the case of diffusion *NMR* spectroscopy by Mathias Nilsson *et al* in [110], by using concentration variations in a ongoing reaction as the third dimension, which allowed to describe the reaction kinetics. In fact, *DOSY* / timecourse spectra are bilinear data where the signal intensity I is measured as a function of two variables, frequency and gradient amplitude, and frequency and time, respectively. So, in case of spectral overlap, it is common to use multivariate method to help to resolve the component spectra (diffusion/ kinetics). For bilinear analysis, it was necessary to apply constraints such as non negativity and/or known/hypothesised kinetic models, in order to avoid the problem of rotational analysis and allow the true solutions to be selected out from the infinite range of linear combinations. According to the authors, this problem can be avoided by using **PARAFAC** and therefore the experimental data to be used should be represented according to the model in Eq. (29):

$$I = \sum_{i=1}^{N} S_i A_i C_i + E \tag{31}$$

Where S are spectra as a function of frequency f, A are diffusional attenuations as a function of gradient g, C are the concentrations profile as a function of time t and E is the noise. Here, the only requirement is that $S_i(f)$, $A_i(g)$ and $C_i(t)$ of each species be independent of each other.

PARAFAC fitting was carried out for a spectral region of a reacting mixture well known as acid hydrolysis of maltose to glucose, with one as-

sumption, that there were two components.

According to the obtained results, the decomposition proved to be robust and efficient when it was used for experimental timecourse combined with diffusion information.

A second application to *DOSY* (diffusion-ordered spectroscopy) was done by the same authors in [111] but this time relaxation was incorporated as the third dimension. The experiment was named T_1-*DOSY* .
In order to combine relaxation encoding with diffusion encoding, three sequences were investigated, all which are based on the standard diffusion encoding *DOSY* oneshot sequence. The first two sequences were constructed by concatenating a relaxation encoding segment with the *DOSY* sequence and the third by incorporating relaxation encoding whithin the existing diffusion delay. The T_1-*DOSY* experiments were tested on a mixture of 1-propanol and 3-methyl-pentanol for each of the three pulse sequences.

The figure 26 shows good separation for the three different pulse sequences which proved that adding a third dimension based on relaxation to diffusion experiments can help in decomposing the overlapping spectrum of a discrete mixture into the spectra of its individual components when combined with appropriate multiway data processing methods like **PARAFAC**.

Recently, Bjorneras. J *et al.* published a successful application of T_1-*DOSY* to a mixture of 5 components (quinine, camphene, geraniol, residual OH signals from methanol and water) [112].

Rasmus Bro *et al* [113] continued to exploit the **PARAFAC** model for 2D spectra in order to resolve the signals from a signal analyte in a complex mixture with diffusion, *NMR* spectrum and analyte concentration being the three factors in eq.(31). The approach was named "mathematical chromatography". As an example, it was applied to a series of diffusion-

edited 2D *NMR* spectra of mixtures of glucose, maltose and maltotriose. The figure 27 shows the PARAFAC solution which includes three parts: estimated relative concentrations (scores) together with estimated spectra (loadings) and estimated diffusion profiles (loadings)for each of the three compounds.

Despite that the diffusion coefficients of the three compounds were close (around 7, 5 and $4.10^{-6} cm^2/s$ for glucose, maltose and maltotriose respectively) and that their individual spectra have highly overlapping regions,

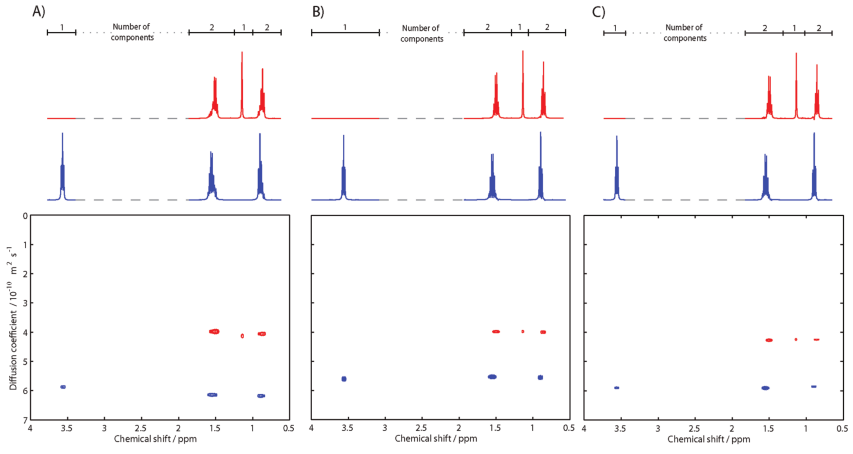

Figure 26: ^1H spectra obtained by PARAFAC decomposition of the results of different T_1-*DOSY* experiments on the mixture of 3-methyl-3-pentanol and 1-propanol. The component spectra (top) constructed from the results of PARAFAC processing using the specified number of components for the four spectral segments indicated, and (bottom) the *DOSY* spectrum constructed from the component spectra and diffusion coefficients obtained for the individual spectral segments: A, B and C refered to the three considered pulse sequences. Reprinted with permission from Analytical Chemistry, Vol. 81, M. Nilsson et al., T1-Diffusion-Ordered Spectroscopy: Nuclear Magnetic Resonance Mixture Analysis Using Parallel Factor Analysis, p 8119-8125. Copyright 2009, American Chemical Society

PARAFAC provided a good separation which confirmed that it may provide a successful method of identification of individual components in highly overlapping 2D *NMR* spectra.

4. Validation process:

Although all **BSS** methods discussed above have successfully been demonstrated in selected cases, a proper assessment of their general applicability

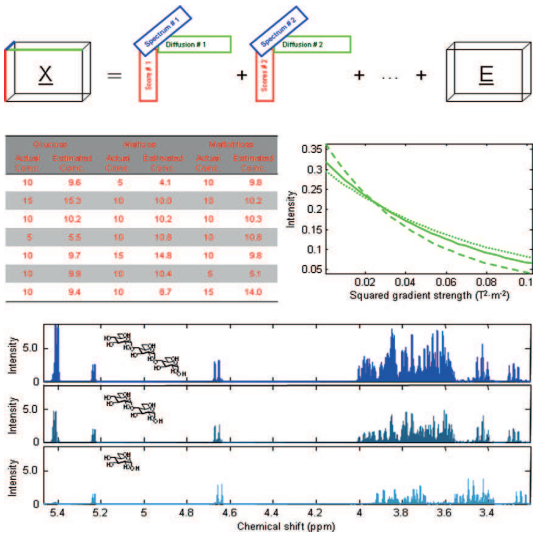

Figure 27: The three-component PARAFAC solution for the NMR data from mixtures of glucose, maltose and maltotriose using diffusion-edited 2D NMR data. The PARAFAC solution (above with color codes) provides the estimated relative concentrations (scores) of each component, which are to be scaled only to provide the true concentrations of each of the three compounds (inserted table). Furthermore, the resolved diffusion profiles related to each of the three compounds and the resolved pure NMR spectra of each of the three different compounds are estimated. Reprinted from Trends in Analytical Chemistry, Vol. 29, R. Bro et al., Mathematical chromatography solves the cocktail party effect in mixtures using 2D spectra and PARAFAC, p 281-284 .Copyright 2010, with permission from Elsevier

remains elusive. This point is a general one and not only restricted to *NMR* applications. Not all methods presented in the literature have been accompanied by an attempt at estimating their limits in terms of resolving power. Nuzillard proposed first to estimate the quality of separation of **SOBI** , **Fast-ICA** and **JADE** declinations of **ICA** and **LPBSS** , using two fidelity indexes, both of which focused on a measure of the distance between the estimated and real mixing matrix, A. This kind of analysis is possible only on data in which the mixing matrix is constructed artificially , so that the sources and mixing matrices, as well as the noise are known in advance.

In the following we illustrate an original similar performance test, which includes additional comparison of the estimated and real spectral source. Indeed, while a faithful reproduction of the mixing matrix is important for quantitative analysis, it is rather on the aspect of the estimated spectra that the attention of the spectroscopist focuses first, since it allows the assignment of the spectral features to a given compounds precisely.
While these tests provide useful insight on the intrinsic separation power of a given method, one must bare in mind that any experimental aspect that should induce variations of a signal shape or position (e.g. phase, baseline correction etc.) would have an additional impact on the quality of the separation.

Various **BSS** algorithms were tested on the **SM** dataset but considering two cases:

- Artificial **SM** mixtures which were generated from the real ^1H *NMR* spectra of pure components are $0.06(mol/l)$ and $0.04(mol/l)$ respectively. The mixture is arranged in a pseudo PFG-NMR experiment corresponding to diffusion constants of $4.10^{-10}m.s^{-1}$ for the sucrose

and $3.10^{-10} m.s^{-1}$ for the maltotriose. To test the robustness of the methods under study in the presence of the noise, a matrix N with random values of maximum fixed amplitude was added to mixture signals, with various values of Peak Signal-to-Noise Ratio (PSNR).
The standard SNR was also computed. We used the following definitions for SNR and PSNR:

$$\mathsf{SNR} = 20\ \log_{10}\left(\frac{\mathrm{std}(signal)}{\mathrm{std}(noise)}\right) \tag{32}$$

$$\mathsf{PSNR} = 20\ \log_{10}\left(\frac{\max(signal)}{\mathrm{std}(noise)}\right) \tag{33}$$

where "std" stands for *standard deviation*.

- Real-world noised **SM** *DOSY* dataset.

A systematic comparison between **JADE** , **NNSC** , **LPBSS** and **SOBI** was achieved on the basis of three performance indices:
The *Comon index* and the *Choi index*: which evaluate some specific distance measures between the estimated mixing matrix $\hat{A}$ and the real mixing matrix A [83, 85]:

$$\epsilon_{Choi}(A, \hat{A}) = \frac{1}{2(n-1)} \sum_{i=1}^{n} \left(\sum_{k=1}^{n} \frac{|g_{ik}|^2}{(\max_j |g_i j|^2)} - 1 + \sum_{k=1}^{n} \frac{|g_{ki}|^2}{\max_j |g_{ji}|^2} - 1 \right) \tag{34}$$

$$\epsilon_{Comon}(A, \hat{A}) = \sum_i \left| \sum_j |d_{ij}| - 1 \right|^2 + \sum_j \left| \sum_i |d_{ij}| - 1 \right|^2 + \sum_i \left| \sum_j |d_{ij}|^2 - 1 \right| + \sum_j \left| \sum_i |d_{ij}|^2 - 1 \right| \tag{35}$$

where d_{ij} are the elements of $D = \underline{A}^{-1}\hat{\underline{A}}$ where the notation $\underline{A}$ designates the matrix obtained from A by multiplying each column A^j by $\left\|A^j\right\|^{-1}$.
g_{ij} are the elements of $G = \hat{A}^{-1}A$.
We developed furthermore an error on the sources index, which estimates the similarity degree between the estimated spectra of components $\hat{S}$ and the real ones $\hat{S}$. the expression of the error is illustrated below:

$$\epsilon_S(S, \hat{S}) = \log \frac{\left\|S - \hat{S}\right\|_p}{\|S\|} \tag{36}$$

with $p = 4$,in order to get more interest in the regions where there is more information (peaks).

4.1. Impact of noise: artificial mixtures and additional artificial noise

The following Fig. 28, Fig. 29 and Fig. 30 show the behavior of algorithms **JADE** , **SOBI** , **LPBSS** and **NNSC** according to the variation of the **SNR**.

The Figs 28 and 29 represent the performance accroding to Choi and Comon fidelity indices, respectively. Although these two do not produce totally coinciding results, some general trends can be inferred. Here, the **SOBI** approach is confirmed to be the least effective one. The remaining methods perform best for low noise content, starting from around $60dB$. The **NNSC** method appears to perform best overall, as it is able to produce acceptable results even for slightly lower signal-to-noise levels. On the other hand, for little or no noise content (i.e. $S/N > 64dB$), **LPBSS** is predicted to be the most faithful algorithm. Indeed, the regularization factor introduced in **NNSC** induces a toll on the similarity between real and

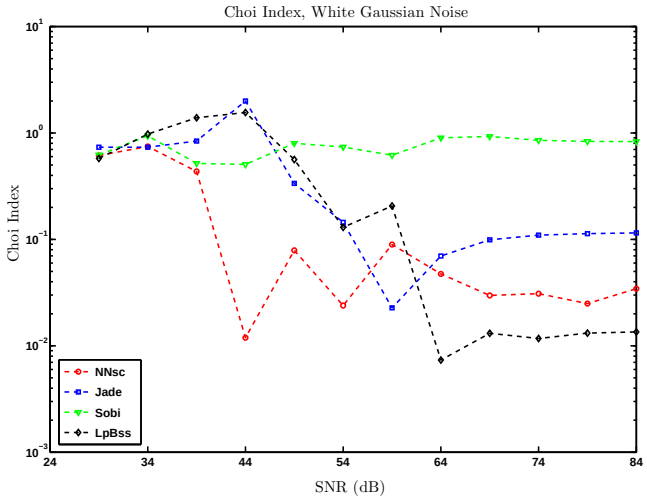

Figure 28: The evolution of the Choi fidelity index on the A matrixindex according to the SNR variation for SOBI, LPBSS, NNSC and JADE on a SM mixture. See text for details.

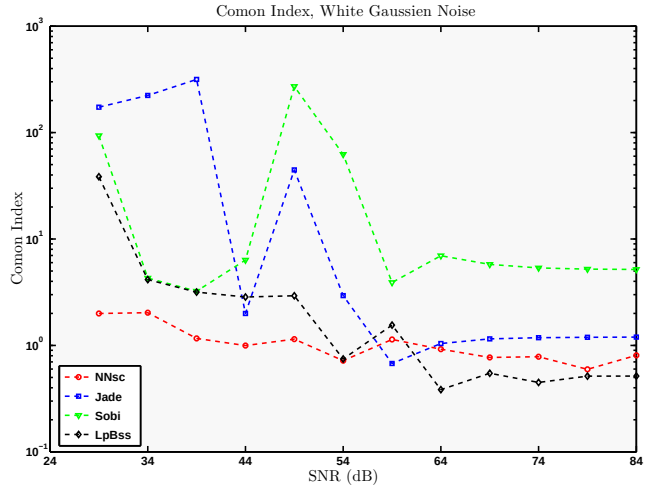

Figure 29: The evolution of the Comon fidelity index on the matrix A according to the SNR variation for SOBI, LPBSS, NNSC and JADE on a SM mixture. See text for details.

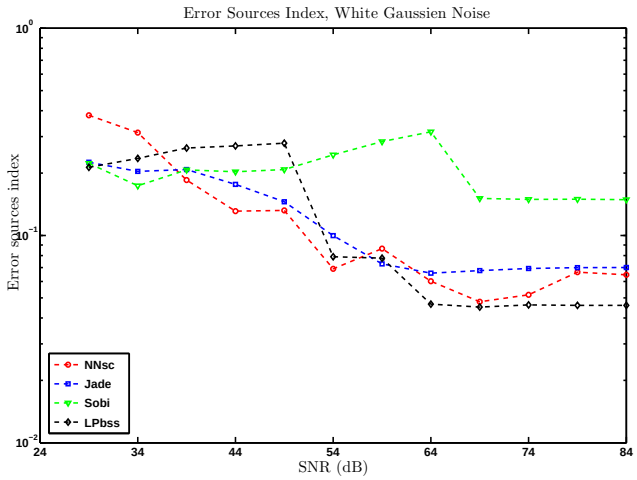

Figure 30: The evolution of the Error on the source spectra according to the SNR variation

estimated sources, since all calculated peaks are reduced of an amount proportional to this parameter. Note that the S/N ratio will vary for peaks of different intensity in these spectra, and thus the less intense peak are the one ones that will suffer the highest relative error. These indexes provide a global estimate, so that visual inspection (or point by point estimation) can reveal significant distortions in the estimated sources that can go unnoticed.

We further analysed the separation behaviour of the algorithm by displaying in details two cases: $SNR = 54dB$ and $SNR = 64dB$.

According to these results, the algorithm that provides the worst separation is **SOBI** which is due to the absence of constraints that highlight the nature of *NMR* data, such as sparsity and nonnegativity.

JADE performed badly in the low range of SNR but eventually became more stable providing a good separation. This can be understood since the estimated sources contain negative residuals from the other sources which increases the values of the error on the estimation of the source matrix S

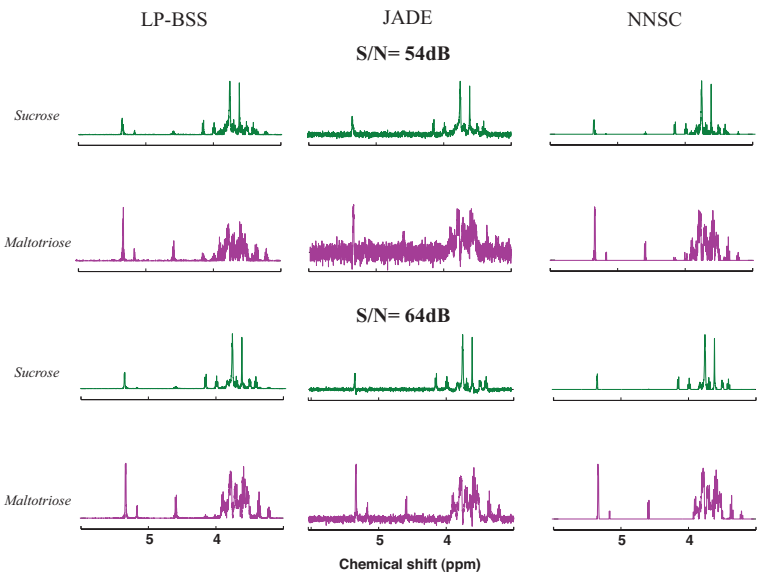

Figure 31: Calculated sources with different methods for the demixing test performed on artificial mixtures made of sucrose and maltotriose signal with a variable amount of noise. The spectra were calculated for a value of S/N of 54 dB (top panels) and 64 dB (bottom panels). Left panels correspond to LPBSS, middle ones to JADE, and right ones to NNSC. Reprinted with permission from Anal. Chem, Vol 85, Toumi. I, Torrsani.B and Caldarelli.S, Effective Processing of Pulse Field Gradient NMR of Mixtures by Blind Source Separation, p 11344-11351. Copyright 2013 American Chemical Society

and on the other hand, **JADE** estimates a matrix B supposed to be the pseudo-inverse of A (mixing matrix) so the recovered elements of A are not very precise.

The performance of **LPBSS** (which involves nonnegativity constraints and local sparsity) follows a sudden increase for high values of this parameter indicating that this algorithm is rather efficient in the absence of noise.

The **NNSC** algorithm is the one that seems to be the most stable regardless the nature of the noise. Since the algorithm is based on sparseness and nonnegativity constraints, it was required to estimate the number of

sources that are actually present in the mixtures without considering the noise as a source, showing that the algorithm is less affected by noise than others.

A weakness with the **NNSC** algorithm is the necessity of guessing λ, which on one hand requires testing to find the value that provides a good separation, moreover the intensities of the signals in the estimated sources are not the same as in the ground truth but reduced by a factor proportional to λ.

4.2. Case of real-world mixtures and real noise

The previous study was done in the case of artificial **SM** mixtures with artificial additive gaussian noise. In order to get a more realistic estimate of the separation in the case of noisy mixtures, we studied three different cases of real-world noisy mixtures [114], corresponding to PSNR equals to 70, 72 and 74 dB.

The separation was done using **JADE** , **LPBSS** and **NNSC** algorithms for all these cases (Fig. 32, Fig. 33 and Fig. 34).

We can see from Fig. 32 that **JADE** separated better the Sucrose spectrum but with some negative residues in the region between 3 and 4 ppm for all cases. The Maltotriose spectrum was badly estimated with highly significant residues which spread over the region 3-6 ppm.

This confirms that **JADE** only has a chance at working for very high SNR. For instance, we proposed to use **ICA** (rather than PCA as in [94]) to estimate the number of components to submit to a more accurate but slower algorithm as **NNSC** .

Figure 33 shows that the separation with the **LPBSS** algorithm was done only in one case ($PSNR = 74\ dB$) with presence of few artifacts between 3 and 4 ppm. This suggests that the method is not able to separate

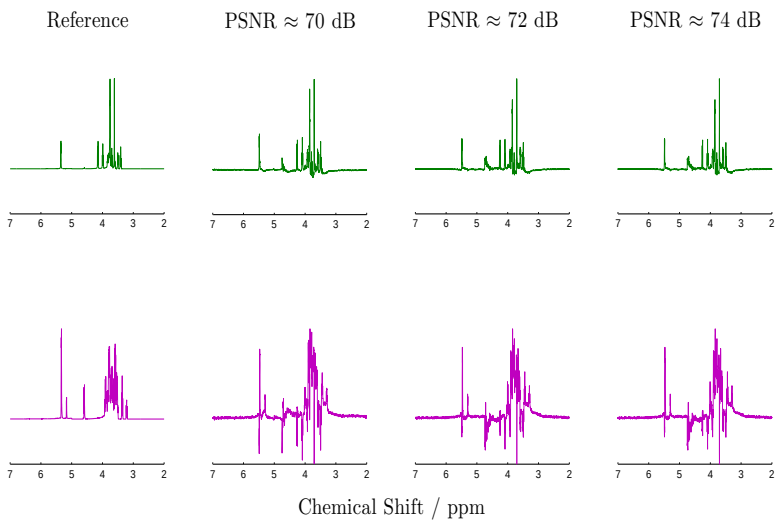

Figure 32: The sources recovered by JADE for an real-world SM mixture for different noise levels, compared to the ground truth sources

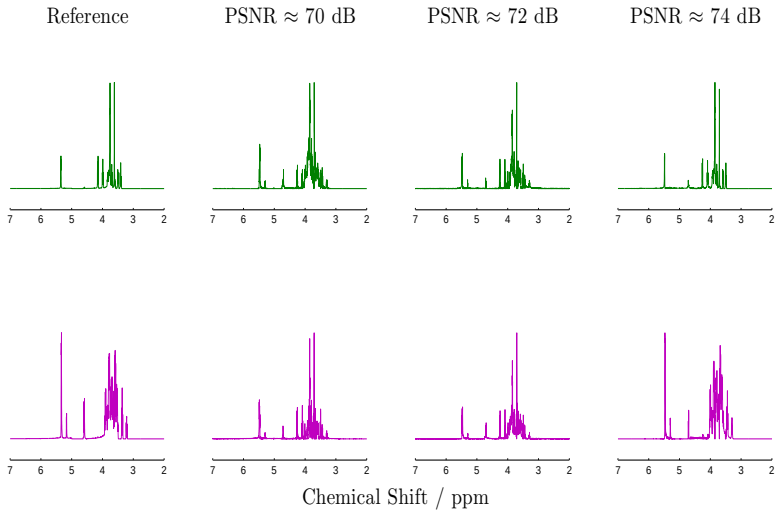

Figure 33: The sources recovered by LPBSS for an real-world SM mixture for different noise levels, compared to the ground truth sources

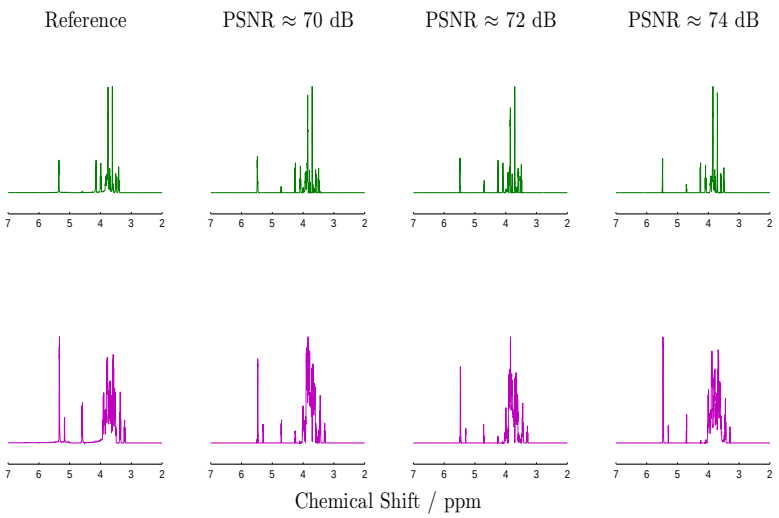

Figure 34: The sources recovered by NNSC for an real-world SM mixture for different noise levels, compared to the ground truth sources

noisy mixtures and hence it requires very high quality of *NMR* data to perform adequately.

Unlike **JADE** and **LPBSS** , **NNSC** performed well in all three cases (Figure 34). The separation is enhanced going from the $PSNR = 70\ dB$ and $PSNR = 72\ dB$, where the separated spectra presented a few residues between 3.5 and 4.5 ppm, to $PSNR = 74\ dB$, where residues were noticed just around 4 ppm.

These results reinforce the outcome from synthetic data and additive noise and prove that both **JADE** and **LPBSS** have a good performance in the case of low noise *NMR* spectra and that the **NNSC** algorithm is the most robust algorithm for the case of *NMR* spectroscopy even when it consists of high overlapped spectra from noised mixtures.

5. Conclusion

The examples shown in this first review of Blind Source Separation algorithms applied to *NMR* of mixtures illustrate as this concept is still in its infancy, and is expected to develop considerably in the coming years as one of the alternatives to plain Fourier Transform.

Thus far, most demonstrations have been dealing with PFG-NMR decays, which is understandable as this experiment is still in strong need for an effective and robust processing toolset. A main point that requires better understanding is the prediction of the separation capabilities of a particular **BSS** algorithm. However, although the superiority with respect to current methods has been illustrated for specific datasets, a clear description of the resolving power of **BSS** methods for *NMR* has not been provided, and the evaluation of **BSS** performances remains thus far very qualitative.

Finally, in the verge of the fast expansion of *NMR* of complex mixtures of small molecules, **BSS** is likely to be further tested in this context, for which just a few but promising examples exist to date.

References:

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

[2] A. Cichocki, S.-i. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, John Wiley & Sons, Inc., New York, NY, USA, 2002.

[3] S. Choi, A. Cichocki, H.-M. Park, S.-Y. Lee, Blind source separation and independent component analysis: A review, Neural Information Processing - Letters and Reviews 6 (1) (2005) 1–57.

[4] S. L. Robinette, R. Bruschweiler, F. C. Schroeder, A. S. Edison, Nmr in metabolomics and natural products research: Two sides of the same coin, Accounts of Chemical Research 45 (2) (2012) 288–297.

[5] A. E. Taggi, J. Meinwald, F. C. Schroeder, A new approach to natural products discovery exemplified by the identification of sulfated nucleosides in spider venom, Journal of the American Chemical Society 126 (33) (2004) 10364–10369.

[6] F. C. Schroeder, D. M. Gibson, A. C. L. Churchill, P. Sojikul, E. J. Wursthorn, S. B. Krasnoff, J. Clardy, Differential analysis of 2d nmr spectra: New natural products from a pilot-scale fungal extract library, Angewandte Chemie International Edition 46 (6) (2007) 901–904.

[7] D. V. Rubtsov, J. L. Griffin, Time-domain bayesian detection and estimation of noisy damped sinusoidal signals applied to nmr spectroscopy, Journal of Magnetic Resonance 188 (2) (2007) 367–379.

[8] D. V. Rubtsov, C. Waterman, R. A. Currie, C. Waterfield, J. D. Salazar, J. Wright, J. L. Griffin, Application of a bayesian deconvolution approach for high-resolution h-1 nmr spectra to assessing the metabolic effects of acute

phenobarbital exposure in liver tissue, Analytical Chemistry 82 (11) (2010) 4479–4485.

[9] T. Ye, C. Zheng, S. Zhang, G. A. N. Gowda, O. Vitek, D. Raftery, "add to subtract": A simple method to remove complex background signals from the h-1 nuclear magnetic resonance spectra of mixtures, Analytical Chemistry 84 (2) (2012) 994–1002.

[10] O. Beckonert, H. C. Keun, T. M. Ebbels, J. Bundy, E. Holmes, J. C. Lindon, J. K. Nicholson, Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts, Nat Protoc 2 (11) (2007) 2692–2703.

[11] M. Reddy G. N, R. Ballesteros G, J. Lacour, S. Caldarelli, Determination of labile chiral supramolecular ion pairs by chromatographic nmr spetroscopy, Angewandte Chemie International Edition 52 (11) (2013) 3255–3258.

[12] M. R. G. N., S. Caldarelli, Demixing of severely overlapping nmr spectra through multiple-quantum nmr, Analytical Chemistry 82 (8) (2010) 3266–3269.

[13] G. N. Manjunatha Reddy, S. Caldarelli, Improved excitation uniformity in multiple-quantum nmr experiments of mixtures., Magn Reson Chem 51 (4) (2013) 240–244.

[14] G. N. M. Reddy, S. Caldarelli, Identification and quantification of epa 16 priority polycyclic aromatic hydrocarbon pollutants by maximum-quantum nmr, Analyst 137 (2012) 741–746.

[15] G. N. M. Reddy, S. Caldarelli, Maximum-quantum (maxq) nmr for the speciation of mixtures of phenolic molecules, Chem. Commun. 47 (2011) 4297–4299.

[16] M. Piotto, G. N. Manjunatha Reddy, S. Caldarelli, Non-uniformly sampled maximum quantum spectroscopy., J Magn Reson 213 (1) (2011) 107–111.

[17] K. Bingol, R. Bruschweiler, Multidimensional approaches to nmr-based metabolomics., Anal Chem 86 (1) (2014) 47–57.

[18] R. Bruschweiler, Theory of covariance nuclear magnetic resonance spectroscopy, Journal of Chemical Physics 121 (1) (2004) 409–414.

[19] R. Bruschweiler, F. L. Zhang, Covariance nuclear magnetic resonance spectroscopy, Journal of Chemical Physics 120 (11) (2004) 5253–5260.

[20] N. Trbovic, S. Smirnov, F. L. Zhang, R. Bruschweiler, Covariance nmr spectroscopy by singular value decomposition, Journal of Magnetic Resonance 171 (2) (2004) 277–283.

[21] F. L. Zhang, R. Bruschweiler, Spectral deconvolution of chemical mixtures by covariance nmr, Chemphyschem 5 (6) (2004) 794–796.

[22] K. Bingol, R. Brueschweiler, Deconvolution of chemical mixtures with high complexity by nmr consensus trace clustering, Analytical Chemistry 83 (19) (2011) 7412–7417.

[23] S. L. Robinette, F. Zhang, L. Bruschweiler-Li, R. Bruschweiler, Web server based complex mixture analysis by nmr, Analytical Chemistry 80 (10) (2008) 3606–3611.

[24] D. A. Snyder, R. Brueschweiler, Generalized indirect covariance nmr formalism for establishment of multidimensional spin correlations, Journal of Physical Chemistry A 113 (46) (2009) 12898–12903.

[25] F. Zhang, S. L. Robinette, L. Bruschweiler-Li, R. Brueschweiler, Web server suite for complex mixture analysis by covariance nmr, Magnetic Resonance in Chemistry 47 (2009) S118–S122.

[26] F. Zhang, L. Bruschweiler-Li, R. Brueschweiler, Simultaneous de novo identification of molecules in chemical mixtures by doubly indirect covariance

nmr spectroscopy, Journal of the American Chemical Society 132 (47) (2010) 16922–16927.

[27] F. Zhang, R. Brueschweiler, Robust deconvolution of complex mixtures by covariance tocsy spectroscopy, Angewandte Chemie-International Edition 46 (15) (2007) 2639–2642.

[28] K. Zangger, H. Sterk, Homonuclear broadband-decoupled nmr spectra, Journal of Magnetic Resonance 124 (2) (1997) 486 – 489.

[29] M. Nilsson, G. A. Morris, Pure shift proton dosy: diffusion-ordered 1h spectra without multiplet structure, Chem. Commun. (2007) 933–935.

[30] J. Aguilar, S. Faulkner, M. Nilsson, G. Morris, Pure shift 1h nmr: A resolution of the resolution problem?, Angewandte Chemie International Edition 49 (23) (2010) 3901–3903.

[31] G. A. Morris, J. A. Aguilar, R. Evans, S. Haiber, M. Nilsson, True chemical shift correlation maps: A tocsy experiment with pure shifts in both dimensions, Journal of the American Chemical Society 132 (37) (2010) 12770–12772.

[32] J. A. Aguilar, M. Nilsson, G. A. Morris, Simple proton spectra from complex spin systems: Pure shift nmr spectroscopy using bird, Angewandte Chemie International Edition 50 (41) (2011) 9716–9717.

[33] J. A. Aguilar, A. A. Colbourne, J. Cassani, M. Nilsson, G. A. Morris, Decoupling two-dimensional nmr spectroscopy in both dimensions: Pure shift noesy and cosy, Angewandte Chemie International Edition 51 (26) (2012) 6460–6463.

[34] S. Islam, J. A. Aguilar, M. W. Powner, M. Nilsson, G. A. Morris, J. D. Sutherland, Detection of potential tna and rna nucleoside precursors in a prebiotic mixture by pure shift diffusion-ordered nmr spectroscopy, Chemistry A European Journal 19 (14) (2013) 4586–4595.

[35] L. Paudel, R. W. Adams, P. Kirly, J. A. Aguilar, M. Foroozandeh, M. J. Cliff, M. Nilsson, P. Sndor, J. P. Waltho, G. A. Morris, Simultaneously enhancing spectral resolution and sensitivity in heteronuclear correlation nmr spectroscopy, Angewandte Chemie International Edition 52 (44) (2013) 11616–11619.

[36] O. Cloarec, M.-E. Dumas, A. Craig, R. H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J. C. Lindon, E. Holmes, J. Nicholson, Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic 1h nmr data sets, Analytical Chemistry 77 (5) (2005) 1282–1289.

[37] S. L. Robinette, J. C. Lindon, J. K. Nicholson, Statistical spectroscopic tools for biomarker discovery and systems medicine, Analytical Chemistry 85 (11) (2013) 5297–5303.

[38] L. M. Smith, A. D. Maher, O. Cloarec, M. Rantalainen, H. Tang, P. Elliott, J. Stamler, J. C. Lindon, E. Holmes, J. K. Nicholson, Statistical correlation and projection methods for improved information recovery from diffusion-edited nmr spectra of biological samples, Analytical Chemistry 79 (15) (2007) 5682–5689.

[39] C. J. Jr., Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications, Progress in Nuclear Magnetic Resonance Spectroscopy 34 (3-4) (1999) 203 – 256.

[40] K. F. Morris, C. S. Johnson, Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy, Journal of the American Chemical Society 114 (8) (1992) 3139–3141.

[41] B. R. Martini, V. A. Mandelshtam, G. A. Morris, A. A. Colbourne, M. Nilsson, Filter diagonalization method for processing pfg nmr data, Journal of Magnetic Resonance 234 (0) (2013) 125 – 134.

[42] G. Pages, C. Delaurent, S. Caldarelli, Investigation of the chromatographic process via pulsed-gradient spin-echo nuclear magnetic resonance. role of the solvent composition in partitioning chromatography, Analytical Chemistry 78 (2) (2006) 561–566.

[43] S. Viel, F. Ziarelli, S. Caldarelli, Enhanced diffusion-edited nmr spectroscopy of mixtures using chromatographic stationary phases, Proceedings of the National Academy of Sciences of the United States of America 100 (17) (2003) 9696–9698.

[44] D. W. Armstrong, T. J. Ward, A. Berthod, Micellar effects on molecular-diffusion - theoretical and chromatographic considerations, Analytical Chemistry 58 (3) (1986) 579–582.

[45] R. E. Hoffman, H. Arzuan, C. Pemberton, A. Aserin, N. Garti, High-resolution nmr chromatography using a liquids spectrometer, Journal of Magnetic Resonance 194 (2) (2008) 295–299.

[46] R. Evans, S. Haiber, M. Nilsson, G. A. Morris, Isomer resolution by micelle-assisted diffusion-ordered spectroscopy, Analytical Chemistry 81 (11) (2009) 4548–4550.

[47] C. F. Tormena, R. Evans, S. Haiber, M. Nilsson, G. A. Morris, Matrix-assisted diffusion-ordered spectroscopy: mixture resolution by nmr using sds micelles, Magnetic Resonance in Chemistry 48 (7) (2010) 550–553.

[48] R. W. Adams, J. A. Aguilar, J. Cassani, G. A. Morris, M. Nilsson, Resolving natural product epimer spectra by matrix-assisted dosy, Organic & Biomolecular Chemistry 9 (20) (2011) 7062–7064.

[49] F. Asaro, N. Savko, Resolution of a nonionic surfactant oligomeric mixture by means of dosy with inverse micelle assistance, Magnetic Resonance in Chemistry 49 (4) (2011) 195–198.

[50] A. K. Rogerson, J. A. Aguilar, M. Nilsson, G. A. Morris, Simultaneous enhancement of chemical shift dispersion and diffusion resolution in mixture analysis by diffusion-ordered nmr spectroscopy, Chemical Communications 47 (25) (2011) 7063–7064.

[51] C. Pemberton, R. Hoffman, A. Aserin, N. Garti, New insights into silica-based nmr "chromatography", Journal of Magnetic Resonance 208 (2) (2011) 262–269.

[52] C. Pemberton, R. E. Hoffman, A. Aserin, N. Garti, Nmr chromatography using microemulsion systems, Langmuir 27 (8) (2011) 4497–4504.

[53] P. Stilbs, Automated core, record, and grecord processing of multi-component pgse nmr diffusometry data, European Biophysics Journal 42 (1) (2013) 25–32.

[54] P. Stilbs, K. Paulsen, P. C. Griffiths, Global least-squares analysis of large, correlated spectral data sets: Application to component-resolved ft-pgse nmr spectroscopy, The Journal of Physical Chemistry 100 (20) (1996) 8180–8189.

[55] P. Stilbs, Record processing - a robust pathway to component-resolved hr-pgse nmr diffusometry, Journal of Magnetic Resonance 207 (2) (2010) 332 – 336.

[56] M. A. Delsuc, T. E. Malliavin, Maximum entropy processing of dosy nmr spectra, Analytical Chemistry 70 (10) (1998) 2146–2148.

[57] M. Nilsson, G. A. Morris, Speedy component resolution: An improved tool for processing diffusion-ordered spectroscopy data, Analytical Chemistry 80 (10) (2008) 3777–3782.

[58] M. Nilsson, The DOSY Toolbox: A new tool for processing PFG NMR diffusion data, Journal of Magnetic Resonance 200 (2) (2009) 296–302.

[59] A. A. Colbourne, G. A. Morris, M. Nilsson, Local covariance order diffusion-ordered spectroscopy: A powerful tool for mixture analysis, Journal of the American Chemical Society 133 (20) (2011) 7640–7643.

[60] R. M. Maria, T. B. Moraes, C. J. Magon, T. Venancio, W. F. Altei, A. D. Andricopulob, L. A. Colnago, Processing of high resolution magic angle spinning spectra of breast cancer cells by the filter diagonalization method, Analyst 137 (2012) 4546–4551.

[61] Y. Nishiyama, M. H. Frey, S. Mukasa, H. Utsumi, C-13 solid-state nmr chromatography by magic angle spinning h-1 t-1 relaxation ordered spectroscopy, Journal of Magnetic Resonance 202 (2) (2010) 135–139.

[62] J. F. Cardoso, Blind signal separation: statistical principles, Proceedings of the IEEE 86 (10) (1998) 2009–2025.

[63] A. Belouchrani, K. A. Meraim, J. F. Cardoso, E. Moulines, A blind source separation technique using second order statistics, IEEE Trans. on Sig. Proc. 45 (1997) 434–444.

[64] D. Nuzillard, S. Bourg, J.-M. Nuzillard, Model-free analysis of mixtures by {NMR} using blind source separation, Journal of Magnetic Resonance 133 (2) (1998) 358 – 363.

[65] M. Zibulevsky, B. A. Pearlmutter, Blind source separation by sparse decomposition, Neural Computation 13 (1999) 863–882.

[66] W. Windig, B. Antalek, Direct exponential curve resolution algorithm (decra): A novel application of the generalized rank annihilation method for a single spectral mixture data set with exponentially decaying contribution profiles, Chemometrics and Intelligent Laboratory Systems 37 (2) (1997) 241–254.

[67] B. Antalek, W. Windig, Generalized rank annihilation method applied to a single multicomponent pulsed gradient spin echo nmr data set, Journal of the American Chemical Society 118 (42) (1996) 10331–10332.

[68] G. S. Armstrong, N. M. Loening, J. E. Curtis, A. Shaka, V. A. Mandelshtam, Processing dosy spectra using the regularized resolvent transform, Journal of Magnetic Resonance 163 (1) (2003) 139 – 148.

[69] E. O. Stejskal, J. E. Tanner, Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient, The Journal of Chemical Physics 42 (1) (1965) 288–292.

[70] damuse-a new tool for denoising and blind source separation, Digital Signal Processing 15 (4) (2005) 400 – 421.

[71] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Netw. 13 (4-5) (2000) 411–430.

[72] A.Hyvärinen, E.Oja, Handbook of Independent Component Analysis, 1st Edition, Wiley-Interscience, 2001.

[73] J. Zhong, N. DiDonato, P. G. Hatcher, Independent component analysis applied to diffusion-ordered spectroscopy: separating nuclear magnetic resonance spectra of analytes in mixtures, Journal of Chemometrics 26 (5) (2012) 150–157.

[74] A. Hyvärinen, Fast and Robust Fixed-Point Algorithms for Independent Component Analysis, IEEE Transactions on Neural Networks 10 (3) (1999) 626–634.

[75] Z. Koldovsky, P. Tichavsky, E. Oja, Efficient variant of algorithm fastica for independent component analysis attaining the cramér-rao lower bound, Trans. Neur. Netw. 17 (5) (2006) 1265–1277.

[76] H. Stogbauer, A. Kraskov, S. A. Astakhov, P. Grassberger, Least dependent component analysis based on mutual information, Tech. rep. (2004).

[77] S. A. Astakhov, H. Stogbauer, A. Kraskov, P. Grassberger, Monte carlo algorithm for least dependent non-negative mixture decomposition, CoRR abs/physics/0601161.

[78] Y. B. Monakhova, S. A. Astakhov, A. Kraskov, S. P. Mushtakova, Independent components in spectroscopic analysis of complex mixtures, Chemometrics and Intelligent Laboratory Systems 3 (2010) 5. 22 p.

[79] I. Toumi, B. Torrsani, S. Caldarelli, Effective processing of pulse field gradient nmr of mixtures by blind source separation, Analytical Chemistry 85 (23) (2013) 11344–11351.

[80] S. Mallat, A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way, 3rd Edition, Academic Press, 2008.

[81] I. Kopriva, I. Jeric, V. Smrecki, Extraction of multiple pure component 1h and 13c nmr spectra from two mixtures: Novel solution obtained by sparse component analysis-based blind decomposition, Analytica Chimica Acta 653 (2) (2009) 143 – 153.

[82] I. Kopriva, I. Jeric, Blind separation of analytes in nuclear magnetic resonance spectroscopy and mass spectrometry: Sparseness-based robust multicomponent analysis, Analytical Chemistry 82 (5) (2010) 1911–1920.

[83] W. Naanaa, J.-M. Nuzillard, Blind source separation of positive and partially correlated data, Signal Processing 85 (9) (2005) 1711 – 1722.

[84] P. Comon, Independent component analysis, a new concept?, Signal Processing 36 (3) (1994) 287 – 314.

[85] Y. Sun, C. Ridge, F. del Rio, A. J. Shaka, J. Xin, Postprocessing and sparse

blind source separation of positive and partially overlapped data, Signal Process. 91 (8) (2011) 1838–1851.

[86] Y. Sun, J. Xin, A recursive sparse blind source separation method and its application to correlated data in nmr spectroscopy of biofluids, J. Sci. Comput. 51 (3) (2012) 733–753.

[87] Y. Sun, J. Xin, Nonnegative Sparse Blind Source Separation for NMR Spectroscopy by Data Clustering, Model Reduction, and l1 Minimization, SIAM Journal on Imaging Sciences 5 (3) (2012) 886–911.

[88] R. Huo, R. Wehrens, J. Duynhoven, L. Buydens, Assessment of techniques for {DOSY} {NMR} data processing, Analytica Chimica Acta 490 (12) (2003) 231 – 251.

[89] R. Huo, R. Wehrens, L. M. C. Buydens, Improved dosy nmr data processing by data enhancement and combination of multivariate curve resolution with non-linear least square fitting., J Magn Reson 169 (2) (2004) 257–269.

[90] R. Huo, van de Molengraaf, P. R.A, R. J.A, Wehrens, L. Buydens, Diagnostic analysis of experimental artefacts in dosy nmr data by covariance matrix of the residuals, Journal of Magnetic Resonance 172 (2) (2005) 346–358.

[91] R. Huo, R. Wehrens, L. Buydens, Robust dosy nmr data analysis, Chemometrics and Intelligent Laboratory Systems 85 (1) (2007) 9 – 19.

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

[93] C.-J. Lin, Projected gradient methods for nonnegative matrix factorization, Neural Comput. 19 (10) (2007) 2756–2779.

[94] D. A. Snyder, F. Zhang, S. L. Robinette, L. Bruschweiler-Li, R. Bruschweiler,

Non-negative matrix factorization of two-dimensional nmr spectra: application to complex mixture analysis., J Chem Phys 128 (5).

[95] D. D. Lee, H. S. Seung, Algorithms for Non-negative Matrix Factorization, in: T. K. Leen, T. G. Dietterich, V. Tresp (Eds.), Advances in Neural Information Processing Systems 13, MIT Press, The MIT Press, 55 Hayward Street Cambridge, MA 02142-1493 USA, 2001, pp. 556–562.

[96] C.-J. Lin, On the Convergence of Multiplicative Update Algorithms for Nonnegative Matrix Factorization, Neural Networks, IEEE Transactions on 18 (6) (2007) 1589–1596.

[97] P. Sajda, S. Du, L. C. Parra, Recovery of constituent spectra using nonnegative matrix factorization (2003) 321–331.

[98] P. O. Hoyer, Non-negative sparse coding, in: IN NEURAL NETWORKS FOR SIGNAL PROCESSING XII (PROC. IEEE WORKSHOP ON NEURAL NETWORKS FOR SIGNAL PROCESSING, 2002, pp. 557–565.

[99] P. O. Hoyer, Non-negative matrix factorization with sparseness constraints, Journal of Machine Learning Research 5 (2004) 1457–1469.

[100] P. O. Hoyer, The NNSC algorithm, `http://www.cs.helsinki.fi/u/phoyer/software.html`.

[101] R. Bro, PARAFAC. Tutorial and applications, Chemometrics and Intelligent Laboratory Systems.

[102] C. F. Beckmann, S. M. Smith, Tensorial extensions of independent component analysis for multisubject FMRI analysis, Neuroimage 25 (1) (2005) 294–311.

[103] R. A. Harshman, Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multi-modal factor analysis, UCLA Working Papers in Phonetics 16 (1) (1970) 84.

[104] J. Carroll, J.-J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition, Psychometrika 35 (3) (1970) 283–319.

[105] R. A. Harshman, M. E. Lundy, Parafac: Parallel factor analysis, Computational Statistics & Data Analysis 18 (1) (1994) 39 – 72.

[106] D. M. and. all, Analysis of lipoproteins using 2d diffusion-edited nmr spectroscopy and multi-way chemometrics, Analytica Chimica Acta 531 (2005) 209–216.

[107] J. Forshed, R. Stolt, H. Idborg, S. P. Jacobsson, Enhanced multivariate analysis by correlation scaling and fusion of lc/ms and 1h nmr data, Chemometrics and Intelligent Laboratory Systems 85 (2) (2007) 179 – 185.

[108] A. Yilmaz, N. Nyberg, J. Jaroszewski, Metabolic profiling based on two-dimensional j-resolved 1h nmr data and parallel factor analysis., Anal Chem 83 (21) (2011) 8278–85.

[109] I. Montoliu, F.-P. J. Martin, S. Collino, S. Rezzi, S. Kochhar, Multivariate modeling strategy for intercompartmental analysis of tissue and plasma 1h nmr spectrotypes. 8 (5) (2009) 2397–406.

[110] M. Nilsson, M. Khajeh, A. Botana, M. A. Bernstein, G. A. Morris, Diffusion nmr and trilinear analysis in the study of reaction kinetics, Chem. Commun. 0 (2009) 1252–1254.

[111] M. Nilsson, A. Botana, G. A. Morris, T1-diffusion-ordered spectroscopy: Nuclear magnetic resonance mixture analysis using parallel factor analysis, Analytical Chemistry 81 (2009) 8119 – 8125.

[112] J. Bjorneras, A. Botana, G. A. Morris, M. Nilsson, Resolving complex mixtures: trilinear diffusion data, Journal of Biomolecular NMR (2013) 1–7.

[113] R. Bro, N. Viereck, M. Toft, H. Toft, P. I. Hansen, S. B. Engelsen, Mathematical chromatography solves the cocktail party effect in mixtures using 2d spectra and parafac, TrAC Trends in Analytical Chemistry 29 (4) (2010) 281 – 284.

[114] I. Toumi, Decomposition methods of nmr signal of complex mixtures: Models and applications, Ph.D. thesis, Université Aix Marseille (2013).

Glossary:

BSS: Blind Source Separation, 7

DENET: Mixture of Dextran,Ethanol,Nicotinic acid,Ephedrine and Tartrazine, 22

DI: Dominant Interval, 38

DOSY: Diffusion Ordered NMR Spectroscopy, 5

ICA: independent Component Analysis, 16

JADE: Joint Approximate Diagonalization of Eigen-matrices, 21

LPBSS: Linear Programming BSS, 30

NMF: Non Negative Matrix Factorization, 10

NNSC: Non Negative Sparse Coding, 48

PARAFAC: Parallel Factor Analysis, 10

PSNR: Peak Signal-to-Noise Ratio, 61

QGC: Mixture of Quinine, Geraniol and Camphene, 22

rSAP: relaxed SAP, 35

SAP: Single Analyte Point, 28

SCA: Sparse Component Analysis, 25

SM: Mixture of Sucrose and maltotriose, 22

SNR: Signal-to-Noise Ratio, 61

SOBI: Second Order Blind Identification, 13