Nonlinear Mixture-wise Expansion Approach to Underdetermined Blind Separation of Nonnegative Dependent Sources

Journal:	*Journal of Chemometrics*
Manuscript ID:	CEM-13-0069.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Kopriva, Ivica; Ruđer Bo koviæ Institute, Laser and Atomic Research and Development Jerić, Ivanka; Ruđer Bošković Institute, Organic Chemistry and Biochemistry Brkljačić, Lidija; Ruđer Bošković Institute, Organic Chemistry and Biochemistry
Keyword:	Underdetermined blind source separation, Dependent sources, Reproducible kernel Hilbert spaces, Empirical kernel maps, Nonnegative matrix factorization

Nonlinear Mixture-wise Expansion Approach to Underdetermined Blind Separation of Nonnegative Dependent Sources

Ivica Kopriva[1*], Ivanka Jerić[2], and Lidija Brkljačić[2]

Ruđer Bošković Institute, Bijenička cesta 54, HR-10000, Zagreb, Croatia

[1]Division of Laser and Atomic Research and Development

phone: +385-1-4571-286, fax:+385-1-4680-104

e-mail: ikopriva@irb.hr

[2]Division of Organic Chemistry and Biochemistry

e-mail: ijeric@irb.hr, Lidija.Brkljacic@irb.hr

Abstract

Underdetermined blind separation of nonnegative dependent sources consists in decomposing set of observed mixed signals into greater number of original nonnegative and dependent component (source) signals. That is an important problem for which very few algorithms exist. It is also practically relevant for contemporary metabolic profiling of biological samples, such as biomarker identification studies, where sources (a.k.a. pure components or analytes) are aimed to be extracted from mass spectra of complex

multicomponent mixtures. This paper presents method for underdetermined blind separation of nonnegative dependent sources. The method performs nonlinear mixture-wise mapping of observed data in high-dimensional reproducible kernel Hilbert space (RKHS) of functions and sparseness constrained nonnegative matrix factorization (NMF) therein. Thus, original problem is converted into new one with increased number of mixtures, increased number of dependent sources and higher-order (error) terms generated by nonlinear mapping. Provided that amplitudes of original components are sparsely distributed, that is the case for mass spectra of analytes, sparseness constrained NMF in RKHS yields, with significant probability, improved accuracy relative to the case when the same NMF algorithm is performed on original problem. The method is exemplified on numerical and experimental examples related respectively to extraction of ten dependent components from five mixtures and to extraction of ten dependent analytes from mass spectra of two to five mixtures. Thereby, analytes mimic complexity of components expected to be found in biological samples.

Key words: Underdetermined blind source separation, Dependent sources, Reproducible kernel Hilbert spaces, Empirical kernel maps, Nonnegative matrix factorization.

1. INTRODUCTION

Blind source separation (BSS) refers to extraction of unknown source signals from observed mixture signals only [1-4]. Within BSS framework a nonnegative BSS (NBSS), where both mixing and source matrix are nonnegative, has drawn significant

attention recently yielding algorithms such as nonnegative independent component analysis (NICA) [5], nonnegative matrix factorization (NMF) [3, 6-8], convex analysis/geometry [9-11], nonnegative least correlated component analysis (nLCA) [12], determinant based sparseness measure approach to NBSS [13], and sparse component analysis (SCA) that combines data clustering and ℓ_1-minimization [14, 15]. A challenge for NBSS algorithms set by real world problems is characterized by more sources than mixtures available, i.e. NBSS problem is underdetermined (uNBSS), whereas sources are dependent. Such problems, associated with research related to health, food and environment, set motivation for development of the uNBSS algorithm to be presented herein. For example, 326 analytes were quantified in extracts of *Arabidopsis thaliana* leaf tissue [16], while the independent gas chromatography-mass spectrometry (GC-MS) study of *Arabidopsis thaliana* leaves detected 497 unique chemical components [17]. Metabolic profiling, that is seen as one of the most challenging tasks in chemical biology [18], aims to identify and quantify small-molecule analytes (a.k.a. pure components or sources) present in biological samples, typically urine, serum or tissue extract. Thereby, number of analytes can be large. For example, analysis of human adult urinary metabolome by liquid chromatography-mass spectrometry (LC-MS) revealed presence of 1484 components, while 384 of them were characterized by matching their spectra with references stored in libraries [19]. Great majority of algorithms developed for separation of dependent sources are incapable to deal with uNBSS problem, [5, 6, 9-14]. As opposed to them, few algorithms capable to handle uNBSS problem with dependent sources include [7, 8, 15, 20]. Hence, we propose new method for uNBSS problem with nonnegative dependent sources. It is a preprocessing method that performs nonlinear mixture-wise mapping of observed data

and sparseness constrained NMF in high-dimensional mapped space. As properly pointed out in [13], performances of many algorithms depends on optimal usage of parameters required to be known *a priori*, such as balance parameter that regulates influence of sparseness constraint [15, 20], or number of overlapping components that exist in mixtures [8]. These parameters are difficult to select optimally in practice. To the best of our knowledge the nonnegative matrix underapproximation (NMU) algorithm is the only one that can handle uNBSS problem with dependent sources and does not require *a priori* information from the user. Therefore, we propose herein to combine nonlinear preprocessing transform (NPT) with the NMU algorithm in mapped high-dimensional space. Hence, the NPT-NMU algorithm. The NPT-NMU is exemplified on numerical and experimental problems. Nevertheless, proposed preprocessing method can be used in combination with other sparseness constrained NMF algorithms such as NMF algorithm with ℓ_0-constraints (NMF_L0) [8].

The rest of the paper is organized as follows. Section 2 introduces instantaneous (memoryless) linear mixture model, commonly used in chemometrics, defines uNBSS problem and presents theory upon which proposed NPT approach is based. Section 3 describes experiments performed on synthetic and MS mixtures. Results of comparative performance analysis between NMU, NMF-L0, NPT-NMU and NPT-NMF-L0 algorithms are discussed in Section 4. The NMF-L0 algorithm has been used as a reference since it is known that ℓ_0-constraints yield best results in the case of dependent (overlapping) sources [21, 22]. In numerical, and even experimental, examples it was possible to set optimally parameter related to number of overlapping sources. Concluding remarks are given in Section 5.

2. THEORY AND ALGORITHM

Aimed application of proposed method is in extraction of analytes from multicomponent mixtures of mass spectra. MS is chosen due to its increasing importance in clinical chemistry, safety and quality control as well as biomarker discovery and validation. Identification of analytes is often achieved by matching experimental spectra to the ones stored in the library [23]. For an example the NIST and Wiley-Interscience universal spectral library [24], contains more than 800 000 mass spectra (corresponding to more than 680 000 compounds). Thus, we also assume that library of reference mass spectra is available to evaluate quality of components extracted by the proposed method.[1] Although various analytical methods are available for the separation of individual compounds from mixtures, ideal separation cannot be always accomplished, especially when dealing with complex samples [19]. There are also analytes that are prone to chemical decomposition and thus cannot be isolated [25]. Furthermore, when two or more analytes elute from chromatography column close to each other in time their peaks overlap partially or completely [26]. Thus, instead of analytes, their mixture will be compared with the reference pure components in the library. This sets motivation for development of algorithm for uNBSS problem with dependent sources.

2.1. Underdetermined nonnegative blind source separation with dependent sources

[1] Please note that any BSS algorithm when applied to experimental data requires some kind of expert knowledge to evaluate the separation results. Herein the library of pure components is such an "expert". The same concept is also in use in hyperspectral image analysis.

Linear mixture model (LMM) is commonly used in chemometrics [27-30] in general, and in MS in particular [29, 30]. It is the model upon which linear instantaneous BSS methods are based [1-4]. In the absence of additive noise the model reads as:

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

where $\mathbf{X} \in \mathbb{R}_{0+}^{N \times T}$ represents matrix of acquired nonnegative mass spectra such that each row $\{\mathbf{x}_{n\bullet}\}_{n=1}^{N}$ of $\mathbf{X}$ contains one recorded multicomponent mixture mass spectra comprised of intensity values at T m/z channels. $\mathbf{A} \in \mathbb{R}_{0+}^{N \times M}$ represents mixture matrix, whereas each column vector $\{\mathbf{a}_{\bullet m}\}_{m=1}^{M}$ represents concentration profile of the corresponding analyte across the N mixture spectra. $\mathbf{S} \in \mathbb{R}_{0+}^{M \times T}$ is a matrix with the rows $\{\mathbf{s}_{m\bullet}\}_{m=1}^{M}$ representing mass spectra of the unknown number of M analytes present in the mixture spectra $\mathbf{X}$. Thereby, the number of analytes M can be less than, equal to and greater than the number of recorded mixtures spectra N. This, respectively, leads to over-, even- and under-determined BSS problems in which case it is assumed that information about concentration of analytes (stored in the mixing matrix $\mathbf{A}$) is not known to the BSS algorithm. That is, it is expected from BSS method to estimate matrix of analytes $\mathbf{S}$ by having at disposal matrix with recorded mixtures spectra $\mathbf{X}$ only. Due to high complexity of spectra of biological samples and, quite often, small number of recorded spectra available, it is certain that in such analyses corresponding BSS problem will be (highly) underdetermined: $M > N$. Thus, in this paper the following assumptions are made on LMM (1):

A1) $0 \le s_{mt} < 1 \;\; \forall m = 1,...,M \;\; t = 1,...,T$, [2]

A2) $a_{nm} \ge 0 \;\; \forall n = 1,...,N \;\; m = 1,...,M$ and $\left\|\mathbf{a}_{\bullet m}\right\|_2 = 1 \; \forall m = 1,...,M$, [3]

A3) $M > N$

A4) $M << T$, [4]

Due to A1) and A2) it is clear that $\mathbf{X} \ge \mathbf{0}$ as well. Furthermore, components spectra will overlap implying that at some *m*/*z* coordinates multiple components will be present. This implies for column vectors of $\mathbf{S}$: $\left\{\left\|\mathbf{s}_{\bullet t}\right\|_0 \le K\right\}_{t=1}^{T}$, where $\left\|\mathbf{s}_{\bullet t}\right\|_0$ denotes ℓ_0 quasi-norm that counts number of nonzero entries of $\mathbf{s}_{\bullet t}$. Thus, K stands for maximal number of analytes that can be present at the particular *m*/*z* coordinate. Hence, sources $\left\{\mathbf{s}_{m\bullet}\right\}_{m=1}^{M}$ will be statistically dependent. The uNBSS problem (1) is ill-posed due to the fact that matrix factorization suffers from indeterminacies: $\mathbf{X}=\mathbf{AS}=\mathbf{AB}^{-1}\mathbf{BS}$ for some invertible $M \times M$ square matrix $\mathbf{B}$. Hence, it has an infinite number of solutions. Meaningful solutions are characterized by the permutation and scaling indeterminacies in which case $\mathbf{B}=\mathbf{P}\mathbf{\Lambda}$, where $\mathbf{P}$ represents permutation and $\mathbf{\Lambda}$ represents diagonal scaling matrix.

[2] Provided that A1) is not satisfied it can be satisfied by scaling $\mathbf{X}$ with a constant *c*: $\mathbf{X} \rightarrow \mathbf{X}/c$. The conservative scaling strategy that always guarantees A1) is given with: $c = \arg\max_t \left\{\left\|\mathbf{x}_{\bullet t}\right\|_1\right\}_{t=1}^{T}$. However, scaling by $c = \arg\max_{n,t} \left\{\mathbf{X}_{nt}\right\}_{n,t=1}^{N,T}$ will satisfy A1) in great majority of occasions.

[3] Due to the scaling indeterminacy that is inherent to the BSS problem magnitude of the mixing vectors cannot be guaranteed. Therefore, A2) constraint is assumed commonly in BSS.

[4] This technical assumption is necessary to ensure that resolution of the spectrometer is high enough to enable discrimination between components the number of which is expected to be large.

However, constraints are necessary to be imposed on **A** and/or **S** to obtain solution of uNBSS problem (1) that is unique up to permutation and scaling indeterminacies. The necessary constraint is sparseness of analytes spectra $\{\mathbf{s}_{m\bullet}\}_{m=1}^{M}$. Sparseness constraint implies that in relation to N and M the maximal number of analytes K present at the particular m/z coordinate is small enough. However, K is application dependent. When number of sources present in the mixture is large, K will grow. Compressive sensing theory has established condition between N, M and K necessary to obtain unique solution for underdetermined system of linear equations: $\{\mathbf{x}_{\bullet t} = \mathbf{A}\mathbf{s}_{\bullet t}\}_{t=1}^{T}$ assuming that **A** is known and random with the entries distributed according to Gaussian or Bernoulli distributions. For ℓ_1-constrained solutions $\{\mathbf{s}_{\bullet t}\}_{t=1}^{T}$ number of measurements N necessary to obtain unique solution with probability one is given with: $N \approx K\log(M/K)$ [31]. When ℓ_p-constraint, $0 \le p \le 1$, is used instead, condition on number of measurements N is given with: $N \ge C_1(p)K + pC_2(p)K\log(M/K)$ [21], where C_1 and C_2 are constants that depend on choice of the norm p. Hence, $\lim_{p\to 0} N \ge C_1(0)K$, i.e. when $p=0$ number of measurements N does not depend on M. That explains good results of ℓ_0-constrained algorithms for solving (1) [8, 23], when compared against ℓ_1-constrained algorithms when K is increasing. However, when (1) is associated with uNBSS problem in chemometrics **A** is not random but deterministic, i.e. it is a concentration matrix. To the best of our knowledge there is only one result related to condition necessary for unique solution of the underdetermined system of equations: $\{\mathbf{x}_{\bullet t} = \mathbf{A}\mathbf{s}_{\bullet t}\}_{t=1}^{T}$ when **A** is deterministic. It is shown in [32] that for cyclic polynomial matrix **A** it applies: $N = O(K^2)$. That is significantly worse than $N \approx K\log(M/K)$ [31], for random **A**. When

analysis of complex mixtures takes place, where number of sources can be large and consequently K will grow, it can be necessary to ensure large number of mixtures spectra N in order to obtain solution of the uNBSS problem (1) that is, possibly, unique up to scaling and permutation indeterminacies. However, when N is associated with biological samples it can virtually be impossible to satisfy this requirement. Therefore, we propose nonlinear transformation of LMM (1) into quasi-linear model with increased number of measurements.

2.2. Nonlinear transform of linear mixture model

We propose mixture-wise nonlinear transform of LMM (1): $\left\{\mathbf{x}_{\bullet t} \mapsto \phi\left(\mathbf{x}_{\bullet t}\right) \in \mathbb{R}_{0+}^{\bar{N}}\right\}_{t=1}^{T}$, such that $\bar{N} \gg N$. We would like $\bar{N}$ to be (very) large and possibly even infinite. The mapping has the following structure:

$$\phi\left(\mathbf{x}_{\bullet t}\right)=\left[\left\{c_{q_1 \ldots q_N} x_{1t}^{q_1} \ldots x_{Nt}^{q_N}\right\}_{q_1,\ldots,q_N=0}^{\bar{N}}\right]^{T} \quad such\ that \sum_{n=1}^{N} q_n \leq \bar{N}, \quad \forall t=1,\ldots,T. \qquad (2)$$

In (2) $\left\{c_{q_1 \ldots q_N}\right\}$ are real constants that are mapping dependent. By taking into account that $x_{nt}=\sum_{m=1}^{M} a_{nm} s_{mt}$, (2) can be written as:

$$\phi\left(\mathbf{x}_{\bullet t}\right)=c_0 \mathbf{e}_1+\mathbf{B}\begin{bmatrix}0\\ \mathbf{s}_{\bullet t}\end{bmatrix}+\mathbf{B}_{HOT}\begin{bmatrix}0\\ \mathbf{0}_{M\times 1}\\ \mathbf{s}_{\bullet t\,HOT}\end{bmatrix} \quad \forall t=1,\ldots,T \qquad (3)$$

where *HOT* stands for higher order (nonlinear) terms introduced by mapping $\phi\left(\mathbf{x}_{\bullet t}\right)$, $\mathbf{e}_1$ is a unit vector in $\mathbb{R}^{\bar{N}}$, $\mathbf{0}_{M\times 1}$ is column vector with zero entries and $\mathbf{s}_{\bullet tHOT}$ is $\bar{N}-M-1$ column vector comprised of $\left\{s_{1t}^{q_1} \times .. \times s_{Mt}^{q_M}\right\}_{q_1,\ldots,q_M=2}^{\bar{N}}$ and $\sum_{m=1}^{M} q_m \leq \bar{N}$. Provided that

LMM (1) is related to MS data analysis, $\{s_{mt}\}_{m=1}^{M}$ represent analytes in mixtures mass spectra at the particular *m/z* coordinate, i.e. *t* corresponds to *m/z*. Then, all the cross-terms $s_{1t}^{q_1} \times ... \times s_{Mt}^{q_M}$ will be zero if only one analyte is not present at this coordinate. Thus, $\mathbf{s}_{\bullet tHOT}$ in (3) will simplify to $\mathbf{s}_{\bullet tHOT} \approx \left[s_{1t}^2 ... s_{Mt}^2 ... s_{1t}^{\bar{N}} ... s_{Mt}^{\bar{N}} \right]^T$. Due to assumption A1), $\{0 \le s_{mt} < 1\}_{m=1}^{M}$, many higher order terms in $\mathbf{s}_{\bullet tHOT}$ will go to zero as power term increases. Speed of decay depends on distribution of amplitudes. For sparse distributions, such as those encountered in MS, it is reasonable to expect that only several HOT of each source will be significantly greater than zero. For an example, for amplitude s_{mt}=0.5, the 10th order power is 9.7×10^{-4}. Nevertheless, powers of $\{\mathbf{s}_{m\bullet}\}_{m=1}^{M}$ will represent new sources that are statistically dependent with the original ones. Thus we can write (3) as:

$$\phi(\mathbf{x}_{\bullet t}) \approx c_0 \mathbf{e}_1 + \bar{\mathbf{B}} \begin{bmatrix} 0 \\ \bar{\mathbf{s}}_{\bullet t} \end{bmatrix} \quad \forall t = 1, ..., T \tag{4}$$

where $\bar{\mathbf{s}}_{\bullet t} = [\mathbf{s}_{\bullet t} \ \mathbf{s}_{\bullet tHOT}]^T$ and $\bar{\mathbf{B}}$ combines on appropriate way $\mathbf{B}$ and $\mathbf{B}_{HOT}$. Model (4) can be written in matrix format yielding:

$$\phi(\mathbf{X}) \approx \left[\underbrace{c_0 \mathbf{e}_1 ... c_0 \mathbf{e}_1}_{\times T \ times} \right] + \bar{\mathbf{B}} \begin{bmatrix} 0 \\ \bar{\mathbf{S}} \end{bmatrix} \tag{5}$$

where $\phi(\mathbf{X}) \in \mathbb{R}_{0+}^{\bar{N} \times T}$, $\bar{\mathbf{B}} \in \mathbb{R}_{0+}^{\bar{N} \times P+1}$ and $\bar{\mathbf{S}} \in \mathbb{R}_{0+}^{P \times T}$. Hence, the uNBSS problem (1) characterized by triplet (*N*, *M*, *K*) is converted into new problem (5) characterized by triplet ($\bar{N}$, *P*, *Q*) where *P*>*M* stands for number of dependent sources in (5) and *Q*>*K*

stands for number of overlapping sources in (5). Provided that amplitudes of the sources are sparsely distributed it is justified to expect that:

$$(\bar{N}/N) >> (P/M) \text{ as well as } (\bar{N}/N) >> (Q/K). \quad (6)$$

In the light of the uniqueness condition related analysis presented in [32], sparseness constrained factorization of (5) will with significant probability yield, depending on fulfillment of (6), increased accuracy when compared against the same factorization method used for the uNBSS problem (1). The difficulty with factorization of problem (5) is that $\bar{N}$ can be large or even infinite, in which case factorization becomes computationally intractable. To alleviate this difficulty a special type of nonlinear mapping ϕ is selected such that space induced by it is reproducible kernel Hilbert space (RKHS) of functions. To this end we introduce the following definitions and theorems.

Definition 2.2.1. A real function $\kappa: \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is positive semi-definite if it is symmetric and satisfies for any finite set of points $\{\mathbf{x}_{\bullet t}\}_{t=1}^{T}$ in $\mathbb{R}^N$ and real numbers $\{\alpha_t\}_{t=1}^{T}$: $\sum_{i,j=1}^{T} \alpha_i \alpha_j \kappa(\mathbf{x}_{\bullet i}, \mathbf{x}_{\bullet j}) \geq 0$.

Theorem 2.2.1. The Moore-Aronszajn theorem, [33]. Given any nonnegative definite function $\kappa(\mathbf{x}, \mathbf{y})$ there exists a uniquely determined RKHS H_κ consisting of real valued functions on set $\mathbf{X} \subset \mathbb{R}^N$ $f: \mathbf{X} \to \mathbb{R}$ such that: (i) $\forall \mathbf{x} \in \mathbf{X}$, $\kappa(\circ, \mathbf{x}) \in H_\kappa$; (ii) $\forall \mathbf{x} \in \mathbf{X}$, $\forall f \in H_\kappa$, $f(\mathbf{x}) = \langle f, \kappa(\circ, \mathbf{x}) \rangle_{H_\kappa}$. Here, $\langle \circ, \circ \rangle$ denotes the inner product associated with H_κ.

Definition 2.2.2. Replacing $f(\mathbf{x})$ in (ii) in Theorem 2.2.1. by $\kappa(\circ,\mathbf{x})$, it follows $\kappa(\mathbf{x}_{\bullet t},\mathbf{x})=\langle\kappa(\circ,\mathbf{x}_{\bullet t}),\kappa(\circ,\mathbf{x})\rangle_{H_\kappa}$. By selecting the nonlinear map as $\phi(\mathbf{x})=\kappa(\circ,\mathbf{x})$ it follows $\kappa(\mathbf{x}_{\bullet t},\mathbf{x})=\langle\phi(\mathbf{x}_{\bullet t}),\phi(\mathbf{x})\rangle_{H_\kappa}$. That is known as *kernel trick*. The nonlinear mapping $\phi(\mathbf{x}_{\bullet t})$ is called explicit feature map (EFM).

Practical importance of the *kernel trick* is enormous since it substitutes evaluation of inner product of possibly infinite dimensional mappings $\langle\phi(\mathbf{x}_{\bullet t}),\phi(\mathbf{x})\rangle_{H_\kappa}$ in H_κ by evaluation of kernel function $\kappa(\mathbf{x}_{\bullet t},\mathbf{x})$ in the space spanned by empirical set of patterns **X**. To substitute EFM-based nonlinear mappings in (5) by implicit kernel-based mappings we need to define empirical kernel map (EKM). To this end we use the following definition, see also definition 2.15 in [34].

Definition 2.2.3. Empirical kernel map. For a given set of patterns $\{\mathbf{v}_{\bullet d}\in\mathbb{R}^N\}_{d=1}^D\subset\mathbf{X}$, $D\in\mathbb{N}$, we call $\psi:\mathbb{R}^N\to\mathbb{R}^D$, where $\mathbf{x}_{\bullet t}\mapsto\kappa(\circ,\mathbf{x}_{\bullet t})\Big|_{\{\mathbf{v}_{\bullet d}\}_{d=1}^D}=\left[\kappa(\mathbf{v}_{\bullet 1},\mathbf{x}_{\bullet t}),...,\kappa(\mathbf{v}_{\bullet D},\mathbf{x}_{\bullet t})\right]^T\ \forall t=1,...,T$ the empirical kernel map with respect to $\{\mathbf{v}_{\bullet d}\}_{d=1}^D$.

Hence, EKM $\psi(\mathbf{x}_{\bullet t})$ is obtained by projecting EFM $\phi(\mathbf{x}_{\bullet t})$ associated with kernel $\kappa(\circ,\mathbf{x}_{\bullet t})$ on a *D*-dimensional subspace in RKHS spanned by $\{\phi(\mathbf{v}_{\bullet d})\in\mathbb{R}^{\bar{N}}\}_{d=1}^D$:

$$\psi(\mathbf{x}_{\bullet t}) = \left[\phi(\mathbf{v}_{\bullet 1}) \ldots \phi(\mathbf{v}_{\bullet D})\right]^T \phi(\mathbf{x}_{\bullet t}) = \begin{bmatrix} \kappa(\mathbf{x}_{\bullet t}, \mathbf{v}_{\bullet 1}) \\ .... \\ \kappa(\mathbf{x}_{\bullet t}, \mathbf{v}_{\bullet D}) \end{bmatrix} \quad \forall t=1,...,T\,. \tag{7}$$

If (4) is substituted in (7) we obtain:

$$\psi(\mathbf{x}_{\bullet t}) = \left[\phi(\mathbf{v}_{\bullet 1}) \ldots \phi(\mathbf{v}_{\bullet D})\right]^T \phi(\mathbf{x}_{\bullet t}) \approx \left[\phi(\mathbf{v}_{\bullet 1}) \ldots \phi(\mathbf{v}_{\bullet D})\right]^T \left[c_0 \mathbf{e}_1 + \bar{\mathbf{B}} \begin{bmatrix} 0 \\ \bar{\mathbf{s}}_{\bullet t} \end{bmatrix} \right]$$

$$\approx \underbrace{c_0 \begin{bmatrix} \kappa(\mathbf{e}_1, \mathbf{v}_{\bullet 1}) \\ \ldots \\ \kappa(\mathbf{e}_1, \mathbf{v}_{\bullet D}) \end{bmatrix}}_{\mathbf{c}} + \underbrace{\left[\phi(\mathbf{v}_{\bullet 1}) \ldots \phi(\mathbf{v}_{\bullet D})\right]^T \bar{\mathbf{B}}}_{\bar{\bar{\mathbf{B}}}} \begin{bmatrix} 0 \\ \bar{\mathbf{s}}_{\bullet t} \end{bmatrix} \approx \mathbf{c} + \bar{\bar{\mathbf{B}}} \begin{bmatrix} 0 \\ \bar{\mathbf{s}}_{\bullet t} \end{bmatrix} \quad \forall t=1,...,T$$

Hence, we can write (7) in the matrix version as:

$$\psi(\mathbf{X}) \approx \mathbf{C} + \bar{\bar{\mathbf{B}}} \begin{bmatrix} \mathbf{0}_{1\times T} \\ \bar{\mathbf{S}} \end{bmatrix} \tag{8}$$

where $\psi(\mathbf{X}) \in \mathbb{R}_{0+}^{D\times T}$, $\mathbf{C} \in \mathbb{R}_{0+}^{D\times T}$, $\bar{\bar{\mathbf{B}}} \in \mathbb{R}_{0+}^{D\times P+1}$ and $\bar{\mathbf{S}} \in \mathbb{R}_{0+}^{P\times T}$. $\mathbf{C}$ in (8) represents bias term and does not play a role in parts based decomposition of $\Psi(\mathbf{X})$ that is enforced by sparseness constrained NMF. Hence, the uNBSS problem (1) characterized by triplet (*N*, *M*, *K*) is converted into new problem (8) characterized by triplet (*D*, *P*, *Q*) where *P*>*M* stands for number of dependent sources in (8) and *Q*>*K* stands for number of overlapping sources in (8). Analogously to (6), provided that amplitudes of the sources are sparsely distributed it is justified to expect that:

$$(D/N) >> (P/M) \text{ as well as } (D/N) >> (Q/K). \tag{9}$$

In the light of the uniqueness condition related analysis presented in [32], sparseness constrained factorization of (8) will with significant probability yield, depending on

fulfillment of (9), increased accuracy when compared against the same factorization method used for the uNBSS problem (1). Thus nonnegativity and sparseness constrained factorization of (8) should extract original sources $\{\mathbf{s}_{m\bullet}\}_{m=1}^{M}$ as well as their powers that actually are new sources that are dependent with the original ones. While in (5) $\bar{N}$ is large or even infinite, *D* in (8) is finite. To perform projection implied by (7) a basis in the original empirical data set **X** has to be constructed $\mathbf{V}=\{\mathbf{v}_{\bullet d}\in\mathbb{R}_{0+}^{N}\}_{d=1}^{D}$ such that

$$span\{\phi(\mathbf{v}_{\bullet d})\}_{d=1}^{D}\approx span\{\phi(\mathbf{x}_{\bullet t})\}_{t=1}^{T} \tag{10}$$

where *span* denotes a vector space spanned by particular set of vectors, i.e. it is expected that basis vectors span the same vector space that is spanned by empirical set of patterns. The basis **V** can be constructed on several ways, for example using data clustering whereas cluster centers represent basis vectors. Hence, basis construction can be computationally challenging problem for itself. This, however, can be avoided if each pattern vector $\{\mathbf{x}_{\bullet t}\}_{t=1}^{T}$ is chosen as a basis vector, i.e. **V**=**X**. Then condition (10) is satisfied perfectly. In this case, however, dimension of the projected sub-space *D* equals the number of the *m*/*z* channels *T*. Hence, the matrix $\psi(\mathbf{X})$ implied by (8) will have dimensions $T\times T$. For low-resolution mass spectrometry *T* is of the order of several thousands and matrix factorization problems implied by (8) are computationally tractable even on today's personal computers. When it comes to the kernel function $\kappa(\circ,\mathbf{x}_{\bullet t})$ necessary to compute $\psi(\mathbf{x}_{\bullet t})$ in (7), respectively (8) for matrix formulation, it is important that induced RKHS is high-dimensional. Although there are many kernel functions that satisfy this requirement we restrict ourselves herein to the one, arguably,

most often used kernel, [34]: the Gaussian kernel $\kappa\left(\mathbf{x}_{\bullet i},\mathbf{x}_{\bullet j}\right)=\exp\left(-\left\|\mathbf{x}_{\bullet i}-\mathbf{x}_{\bullet j}\right\|^2/\sigma^2\right)$ where σ^2 denotes kernel bandwidth or variance. In principle, optimal value of σ^2 depends on dimension N and has to be estimated through cross-validation. When A1) and A2) apply we have found empirically that variance is in the interval $\sigma^2\in[0.1, 1]$. Higher order terms (HOT) present in $\bar{\mathbf{S}}$ in (8) can also be seen as a noise that is generated by nonlinear transformation. Thus, error introduced by HOT can be reduced by applying entry-wise either soft thresholding nonlinearity on $\left\{\psi(\mathbf{X})_{dt}\right\}_{d,t=1}^{D,T}:\left\{\psi(\mathbf{X})_{dt}\rightarrow\eta_\tau\left(\psi(\mathbf{X})_{dt}\right)=\max\left(0,\psi(\mathbf{X})_{dt}-\tau\right)\right\}_{d,t=1}^{D,T}$, [35], or hard thresholding nonlinearity: $\left\{\psi(\mathbf{X})_{dt}\rightarrow\nu_\tau\left(\psi(\mathbf{X})_{dt}\right)=\psi(\mathbf{X})_{dt}1_{\psi(\mathbf{X})_{dt}>\tau}\right\}_{d,t=1}^{D,T}$ where $1_{\psi(\mathbf{X})_{dt}>\tau}$ represents indicator function. Through numerical experiments under assumptions A1) and A2) we have empirically found that, if "de-noising" operator is applied, thresholding parameter ought to be set to $\tau\approx10^{-7}$. Hence, from the user perspective the NPT-NMU algorithm is virtually almost parameter free.

2.3. Sparseness constrained factorization

The NMU algorithm [7], with a MATAB code available at [36], has been used to evaluate effectiveness of proposed nonlinear mixture expansion method. The NMU method performs factorization of (8) in a recursive manner extracting one component at a time. After identifying optimal rank-one solution $\left(\bar{\bar{\mathbf{b}}}_{\bullet 1},\bar{\mathbf{s}}_{1\bullet}\right)$ the rank-one factorization is performed on the residue matrix $\psi(\mathbf{X})\leftarrow\psi(\mathbf{X})-\bar{\bar{\mathbf{b}}}_{\bullet 1}\bar{\mathbf{s}}_{1\bullet}$. To preserve non-negativity of $\psi(\mathbf{X})$ an underapproximation constraint is imposed on $\bar{\bar{\mathbf{B}}}$ and $\bar{\mathbf{S}}$: $\bar{\bar{\mathbf{B}}}\bar{\mathbf{S}}\le\psi(\mathbf{X})$. This constraint yields localized parts-based decomposition where different basis elements

describe disjoint parts of the input data $\psi(\mathbf{X})$. It has been proven in theorem 1 in [7] that sparseness (number of non-zero entries) of $\bar{\bar{\mathbf{B}}}$ and $\bar{\mathbf{S}}$ is less than sparseness of $\psi(\mathbf{X})$. A main reason for preferring the NMU algorithm over other sparseness constrained NMF algorithms is that there are no regularization constants that require a tuning procedure. When performing NMU-based factorization of matrix $\psi(\mathbf{X})$ in (8), the unknown number of analytes *P* needs to be given to the algorithm as an input. *P* represents nonnegative rank of $\psi(\mathbf{X})$ in (8) and its optimal selection is related the model order selection. However, the well-known difficulty with model selection methods is that very often, due to different theoretical assumptions, they yield (significantly) different result when applied to the same data. Due to this reason and also due to the reason of not to lose some of the components our strategy was to set an assumed value of *P* as: $\hat{P} = D = T$. Then, due to (8), $P < \hat{P}$. Then, the NMU algorithm will extract all the $\hat{P}$ components contained in $\psi(\mathbf{X})$ at an increased computational complexity, i.e. the *T*-*P* rank-one factors will be computed unnecessarily. Nevertheless, that is the price worth being paid in order not to lose some components that potentially can lead to biomarker discovery. When all *T* components are extracted they are compared with the reference components from the library. We identify analytes candidates as those that are maximally correlated with components in the library. As a reference solution in the benchmark problem we have used solution obtained by applying the NMF_L0 algorithm [8], to the original problem (1). The MATLAB code of the NMF_L0 algorithm is available at [37]. NMF_L0 is based on natural sparseness measure, the ℓ_0-pseudo-norm of the source or component matrix $\mathbf{S}$. The NMF_L0 when applied to (1) requires *a priori* information on the number of components *M* and number of overlapping components *K*. In

numerical scenario both M and K are known while in experimental scenario selecting optimal (true) value of K can be difficult. Nevertheless, the NMF_L0 can provide a good reference in validating worst-case and average performance of the NMU algorithm when applied on uNBSS problem (1) and as well as the NPT-NMU algorithm, that is the NMU algorithm applied on uNBSS problem (8). We summarize the NPT-NMU algorithm in the Algorithm 1.

3. EXPERIMENTS

Studies on numerical and experimental data reported below were executed on personal computer running under Windows 64-bit operating system with 64GB of RAM using Intel Core i7-3930K processor and operating with a clock speed of 3.2 GHz. Matlab 2012b environment has been used for programming.

3.1. Numerical simulation

In numerical study we simulate LMM (1) with N=5, M=10, T=1000 and $K \in \{2, 3, 4\}$. Each source is generated according to A1) and distributed according to probability density function of mixed state random variable, [38, 39]:

$$p(s_{mt}) = \rho\delta(s_{mt}) + (1-\rho)\delta^*(s_{mt})\, f(s_{mt}) \;\forall m = 1,...,M \;\forall t = 1,...,T \qquad (11)$$

where $\delta(s_{mt})$ is an indicator function and $\delta^*(s_{mt})$=1-$\delta(s_{mt})$ is its complementary function, $\rho = \{P(\mathbf{s}_{mt} = 0)\}_{t=1}^{T}$. Hence, $\{P(s_{mt} > 0) = 1-\rho\}_{t=1}^{T}$. We have generated sources with probability of being zero ρ=0.9. The nonzero state of s_{mt} is distributed according to $f(s_{mt})$. We have chosen exponential distribution $f(s_{mt}) = (1/\mu)\exp(-s_{mt}/\mu)$ to model

sparse distribution of the nonzero states such that the most probable outcomes were equal to μ=0.1 and μ=0.01.[5] For these outcomes figures 1 and 2 respectively show values of normalized correlation coefficients

$$c_{mm} = \langle \mathbf{s}_{m.}, \hat{\mathbf{s}}_{m.} \rangle / \|\mathbf{s}_{m.}\| / \|\mathbf{s}_{m.}\|, \ \forall m = 1, ..., M \qquad (12)$$

between true and separated sources versus Monte Carlo run index. Sources were separated by the NMU algorithm [7], the NMF_L0 algorithm [8] and proposed NPT-NMU algorithm. Since for the NPT-NMU algorithm one run took roughly two hours, only 10 Monte Carlo runs were executed for each simulation scenario. Left column shows minimal value of the correlation coefficient attained by each of the algorithms while right column shows average value of the correlation coefficients for ten sources. The true values for *M* and *K* were reported to the NMF_L0 algorithm and true value for *M* was reported to the NMU algorithm. The NPT-NMU algorithm assumed that *M*=*T*. The NMF_L0 algorithm was run with the following parameter setup: reverse sparse nonnegative least square (rsNNLS) sparse coder and alternating nonnnegative least square (ANLS) for dictionary update stage. Careful inspection of results presented in figures 1 and 2 suggests that NPT-NMU algorithm yields better accuracy than NMU algorithm in 30% of the runs, while NMU is better in 60% of the runs. While average performance of the NMF_L0 algorithm is the best it yields the worse value of the minimal correlation coefficient.

3.2. MS measurements

[5] Even though the exponential distribution has support on the [0,∞) interval setting μ=0.1 implies that with probability 0.9999546 realizations will be contained in [0, 1] interval. For μ=0.01 realizations will be contained in [0, 1] interval with a probability that is close to 1 with an error of 3.72×10^{-44}. Thus, this justifies a choice of exponential distribution to model sparse distribution of amplitudes s_{mt} on interval [0,1].

3.2.1. Chemicals

A library composed of mass spectra of ten amino acids, namely Ala, Asn, Asp, Gln, Glu, Leu, Lys, Phe, Pro and Val ($\mathbf{C}_1$-$\mathbf{C}_{10}$), was constructed. All amino acids and solvents were commercially available. Stock solutions of these amino acids (1 mg/mL) were prepared in 10% methanol. Working solutions (0.16 mg/mL) were prepared by diluting the stock solutions with 10% methanol. Five mixtures ($\mathbf{X}_1$-$\mathbf{X}_5$) were prepared by mixing different volumes of amino acid stock solutions according to Table S-1 given in Supplemental Information. Mass spectra of analytes were recorded by injection of 5 µl of amino acid working solutions and mass spectra of five mixtures were obtained by injection of 15 µl of mixture solutions prepared as described above, to the ion source. Mass spectra of analytes $\mathbf{C}_1$-$\mathbf{C}_{10}$ and mixtures $\mathbf{X}_1$-$\mathbf{X}_5$ are given in Supplemental Information (Figures S-1 and S-2).

3.2.2. Mass spectroscopy measurements

Electrospray ionization-mass spectrometry (ESI-MS) measurements operating in a positive ion mode were performed on a HPLC-MS triple quadrupole instrument equipped with an autosampler (Agilent Technologies, Palo Alto, CA, USA). The desolvation gas temperature was 300 ^{o}C with flow rate of 6.0 L/min. The fragmentor voltage was 135 V and capillary voltage was 4.0 kV. Mobile phase was 0.1% formic acid in 50% methanol and a flow rate of mobile phase was 0.2 ml/min. Mass spectra as total ion current spectra were recorded in *m*/*z* segment of 10-300. All data acquisition and processing was performed using Agilent MassHunter software.

3.2.3. Setting up an experiment

Naturally occurring L-amino acids Ala, Asn, Asp, Gln, Glu, Leu, Lys, Phe, Pro and Val, Figure 3, were chosen for the construction of mass spectra library and preparation of five mixtures used in validation of the proposed method. Using amino acids as testing compounds is rationalized as follows: (a) they are metabolites often followed in metabolomic studies, [40, 41], (b) their mass profile falls into relatively narrow *m/z* window thus mimicking complexity (overlapping) expected to be found in spectra of real biological samples and (c) owing to the fragmentation often taking place in the MS ion source, mass spectra of amino acids are enriched with numerous fragment ions making separation problem even more challenging. Mass spectra of components $\mathbf{C}_1$-$\mathbf{C}_{10}$ are given in Figure S-1 in Supporting Information together with the assignment of the most abundant fragment ions. Inspection of mass spectra clearly shows that some fragment ions are present in spectra of different components. For example, the fragment ion at *m/z* 84 share components $\mathbf{C}_4$, $\mathbf{C}_5$ and $\mathbf{C}_7$, while that at *m/z* 116 components $\mathbf{C}_2$, $\mathbf{C}_3$ and $\mathbf{C}_9$. Moreover, difference in mass spectra of components $\mathbf{C}_4$ (Gln) and $\mathbf{C}_7$ (Lys) is only in the intensity of fragment ions. Normalized cross-correlation coefficients of components are shown in Table S-2 given in Supporting Information. As seen from mass spectra, many of them ($\mathbf{C}_4$ and $\mathbf{C}_5$, $\mathbf{C}_5$ and $\mathbf{C}_7$, $\mathbf{C}_3$ and $\mathbf{C}_9$, $\mathbf{C}_2$ and $\mathbf{C}_3$ as well as $\mathbf{C}_2$ and $\mathbf{C}_9$) are significantly correlated, while correlation between $\mathbf{C}_4$ and $\mathbf{C}_7$ is 0.9539. Thus, blind extraction of these analytes from small number of given mixtures is a (very) hard problem. It is also important to emphasize that mass spectra of mixtures were obtained by direct injection of sample to the ion source (one minute run), *without* chromatographic separation prior to MS analysis.

4. RESULTS AND DISCUSSION

Sparseness constrained matrix factorization methods such as those used in our previous work [30] and the NMU method [7] failed to extract analytes from mixtures. That is explained by significant correlation (overlapping) between the analytes. The NMF_L0 method [8] when applied to LMM (1) yielded decent results in extracting components from five mixtures. Correlation structure discussed in subsection 3.2.3 suggested that optimal value for K could be 4, 5 or 6. Thus, NMF_L0 was cross-validated for values of $K \in \{3,4,5,6\}$. Note, however, that in truly experimental scenario correlation structure of analytes is unknown and selection of optimal K would require extensive cross-validation. Table S-3 shows the best results, in terms of maximal normalized correlation coefficients between extracted components and components in the library, obtained by applying the NMU and NMF_L0 algorithms on mixture spectra according to the LMM (1). Arguably, the best result by NMF_L0 is obtained for K=5. Since recorded mass spectra were composed of 2901 m/z points, extraction of analytes according to model (8) has been reduced to NMU-, respectively NMF_L0, based factorization of the 2901×2901 matrix in mapped space. Before mapping the mixing matrix $\mathbf{X}$ was scaled by $\arg\max_t \left\{ \|\mathbf{x}_{\cdot t}\|_1 \right\}_{t=1}^{T}$ as well as by $\arg\max_{n,t} \left\{ \mathbf{X}_{nt} \right\}_{n,t=1}^{N,T}$. Gaussian kernel with $\sigma^2 \in \{0.1,1,10\}$ has been used. Proposed NPT-NMU method managed to extract ten analytes with a reasonable accuracy from even two mixtures only. The accuracy improves by increasing the number of mixtures. In Table 1 selected results are shown in terms of maximal normalized correlation coefficient (12) obtained by the proposed NPT-NMU method as well as by NMF_L0, while more comprehensive results are reported in Table S-4 in the Supporting Information. In took roughly two hours in described software environment

to perform decomposition of each particular combination of mixture spectra. It was especially demanding to cross-validate number of overlapping components Q in model (8), and that is required by the NMF_L0 algorithm, since value of Q depends on how fast power terms of the original sources decay toward zero. The "best" result was obtained for $Q \approx 50$, but that is significantly worse than obtained by the NMF_L0 based factorization of (1) for $K=5$. On the other side the NPT-NMU method yielded much better result than NMF_L0 method. Although quality of components extracted by the NPT-NMU and NMF_L0 methods was not perfect they were assigned uniquely to the true ones in the library. This aspect is of practical importance in different areas such as disease diagnosis, food quality control, environmental related studies that depend on library matching. Mass spectra of analytes extracted by the proposed NPT-NMU method from five mixtures are shown in Figure S-3 in Supporting Information. To take into account scaling indeterminacy extracted analytes were scaled to 0-100 range (dividing each extracted analyte by its maximal value and multiplying by 100).

5. CONCLUSION

Problems such as metabolic profiling of biological samples aim to extract many dependent (overlapping) analytes from small number of multicomponent mixtures mass spectra. That results in underdetermined nonnegative blind source separation problem (uNBSS) with dependent sources for which an algorithm is proposed. It performs nonlinear mixture-wise mapping of observed data into reproducible kernel Hilbert space (RKHS) of functions and sparseness constrained nonnegative matrix factorization (NMF) in RKHS. For sparse signals such as those encountered in mass spectroscopy the method yields, with significant probability, improved accuracy relative to the case when

the same NMF algorithm is performed on the original uNBSS problem. On demanding numerical and experimental problems the algorithm demonstrated capability to extract ten dependent analytes from two to five mixtures. Thereby, extracted components were assigned uniquely to the true ones in the library. That is practically important for biomarker identification studies.

ACKNOWLEDGMENT

This work has been supported through grant 9.01/232 "Nonlinear component analysis with applications in chemometrics and pathology" funded by the Croatian Science Foundation.

REFERENCES

1. Hyvärinen A, Karhunen J, Oja E. *Independent Component Analysis*. John Wiley & Sons, Inc.: New York, US, 2001.

2. Cichocki A, Amari S. *Adaptive Blind Signal and Image Processing*. John Wiley: New York, 2002.

3. Cichocki A, Zdunek R, Phan A H, Amari, S I. *Nonnegative Matrix and Tensor Factorizations*. John Wiley: Chichester, UK, 2009.

4. Comon P, Jutten C (eds). *Handbook of Blind Source Separation*. Academic Press: Oxford, UK, 2010.

5. Plumbley M, Oja E. A 'nonnegative PCA' algorithm for independent component analysis. *IEEE Trans. Neural Netw.* 2004; **15** (1): 66-76.

6. Lee DD, Seunng HS. Learning the parts of objects by nonnegative matrix factorization. *Nature* 1999; **401**: 788-791.

7. Gillis N, Glineur F. Using underapproximations for sparse nonnegative matrix factorization. *Pattern Rec.* 2010; **43**: 1676-1687.

8. Peharz R, Pernkopf, F. Sparse nonnegative matrix factorization with ℓ^0-constraints. *Neurocomputing*. 2012; **80**: 38-46.

9. Chan T H, Ma W K, Chi C Y, Wang Y. Convex Analysis Framework for blind separation of nonnegative sources. *IEEE Trans. Sig. Proc.* 2008; **56**(10): 5120-5134.

10. Ambikapathi, A M, Chan T H, Ma W K, Chi C Y. Chance-Constrained Robust Minimum-Volume Enclosing Simplex Algorithm for Hyperspectral Unmixing. *IEEE Trans. Geosc. Remote Sens.* 2011; **49** (11): 4194-4209.

11. Naanaa W, Nuzillarrd J M. A geometric approach to blind separation of nonnegative and dependent sources. *Sig. Proc.* 2012; **92**: 2775-2784.

12. Wand F Y, Chi C Y, Chan T H, Wang Y. Nonnegative least correlated component analysis for separation of dependent sources by volume maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* 2010; **32** (5): 875-888.

13. Yang Z, Xiang Y, Xie S, Ding S, Rong Y. Nonnegative Blind Source Separation by Sparse Component Analysis Based on Determinant Measure. *IEEE Trans. Neural Netw. and Learn. Sys.* 2012; **23** (10): 1601-1610.

14. Sun Y, Xin J. Nonnegative Sparse Blind Source Separation for NMR Spectroscopy by Data Clustering, Model Reduction, and ℓ_1 Minimization. *SIAM J. Imaging Sci.* 2012; **5** (3): 886-911.

15. Kopriva I, Cichocki A. Blind decomposition of low-dimensional multi-spectral image by sparse component analysis. *J. Chemometrics* 2009; **23** (11): 590-597.

16. Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey R N, Willmitzer L. Metabolite profiling for plant functional genomics. *Nature Biotechnology* 2000; **18**: 1157-1161.

17. Jonsson P, Johansson A I, Gullberg J, Trygg J, Jiye A, Grung B, Marklund S, Sjöström M, Antti H, Moritz T. High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses," *Analytical Chem.* 2005; **77**: 5635-5642.

18. Nicholson J K, Lindon J C. Systems biology: Metabonomics. *Nature* 2008; **455** (7216): 1054-1056.

19. Roux A, Xu Y, Heilier J-F, Olivier M-F, Ezan E, Tabet J-C, Junot C. Annotation of the human adult urinary metabolome and metabolite identification using ultra high performance liquid chromatography coupled to a linear quadrupole ion trap-orbitrap mass spectrometer. *Anal. Chem*. 2012; **84**: 6429−6437.

20. Cichocki A, Zdunek R, Amari S I. Hierarchical ALS Algorithms for Nonnegative Matrix Factorization and 3D Tensor Factorization. *LNCS* 2007; **4666**: 169-176.

21. Chartran R, Staneva V. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems* 2008; **24**: 035020 (14 pages).

22. Mohimani H, Babie-Zadeh B, Jutten C. A Fast Approach for Overcomplete Sparse Decomposition Based on Smoothed ℓ_0 Norm. *IEEE Trans. Sig. Proc.* 2009; **57**(1): 289-301.

23. Abu-Farha M, Elisma F, Zhou H, Tian R, Asmer M S, Figeys D. Proteomics: from technology developments to biological applications. *Anal. Chem.* 2009; **81**: 4585-4599.

24. McLafferty F W, Stauffer D A, Loh S Y, Wesdemiotis C. Unknown Identification Using Reference Mass Spectra. Quality Evaluation of Databases. *J. Am. Soc. Mass. Spectrom.* 1999; **10**: 1229-1240.

25. Forseth R R, Schroeder F C. NMR-spectroscopic analysis of mixtures: from structure to function. *Curr. Opin. Chem. Biol.* 2011; **15** (1): 38-47.

26. Ni Y, Qiu Y, Jiang W, Suttlemyre K, Su M, Zhang W, Jia W, Du X. ADAP-GC 2.0 Deconvolution of Coeluting Metabolites from GC/TOF-MS Data for Metabolomic Studies. *Anal. Chem.* 2012; **84**: 6619-6629.

27. Nuzillard D, Bourg S, Nuzilard J M. Model-Free Analysis of Mixtures by NMR Using Blind Source Separation *J. Magn. Reson.* 1998; **133**: 358-363.

28. Visser E, Lee T W. An information-theoretic methodology for the resolution of pure component spectra without prior information using spectroscopic measurements. *Chemom. Int. Lab. Syst.* 2004; **70**: 147-155.

29. Kopriva I, Jerić I. Multi-component Analysis: Blind Extraction of Pure Components Mass Spectra using Sparse Component Analysis. *J. Mass Spectrom.* 2009; **44** (9): 1378-1388.

30. Kopriva I, Jerić I. Blind separation of analytes in nuclear magnetic resonance spectroscopy and mass spectrometry: sparseness-based robust multicomponent analysis. *Anal. Chem.* 2010; **82**: 1911-1920.

31. Candès E, Tao T. Near optimal signal recovery from random projections: universal encoding strategy?. *IEEE Trans. Information Theory* 2006; **52**: 5406-5425.

32. DeVore R A. Deterministic constructions of compressed sensing matrices. *Journal of Complexity* 2007; **23**: 918-925.

33. Aronszajn N. The theory of reproducing kernels. *Trans. of the Amer. Math. Soc*. 1950; **68**: 337-404.

34. Schölkopf B, Smola A. *Learning with kernels*. MIT Press, 2002.

35. Donoho D L. De-Noising by Soft-Thresholding. *IEEE Trans. Inf. Theory* 1995; **41** (3): 613-627.

36. The Nicolas Gillis Website. https://sites.google.com/site/nicolasgillis/code [7 March 2013].

37. The Robert Peharz Website. http://www3.spsc.tugraz.at/people/robert-peharz [7 March 2013].

38. Bouthemy P, Piriou C H G, Yao J. Mixed-state auto-models and motion texture modeling. *J. Math Imaging Vision* 2006; **25**: 387-402.

39. Caifa C, Cichocki A. Estimation of Sparse Nonnegative Sources from Noisy Overcomplete Mixtures Using MAP. *Neural Comput.* 2009; **21**: 3487-3518.

40. Hisamatsu T, Okamoto S, Hashimoto M, Muramatsu T, Andou A, Uo M, Kitazume M T, Matsuoka K, Yajima T, Inoue N, Kanai T, Ogata H, Iwao Y, Yamakado M, Sakai R, Ono N, Ando T, Suzuki M, Hibi T. Novel, objective, multivariate biomarkers composed of plasma amino acid profiles for the diagnosis and assessment of inflammatory bowel disease. *PLoS ONE* 2012; **7**: 1-10.

41. Xu Y, Yang L, Yang F, Xiong Y, Wang Z, Hu Z. Metabolic profiling of fifteen amino acids in serum of chemical-induced liver injured rats by hydrophilic interaction liquid chromatography coupled with tandem mass spectrometry. *Metabolomics* 2012; **8**(3): 475-483.

Figure Captions

Figure 1. Normalized correlation coefficient vs. Monte Carlo run index between true and extracted sources by algorithms: NMF_L0 (squares), NMU (stars) and NPT-NMU (circles). Left: minimal (worst) value; (right) mean value for ten sources. From top to bottom - number of overlapping sources *K*: 2, 3, and 4. Most probable value of the nonzero state, generated according to exponential distribution, equal to 0.1.

Figure 2. Normalized correlation coefficient vs. Monte Carlo run index between true and extracted sources by algorithms: NMF_L0 (squares), NMU (stars) and NPT-NMU (circles). Left: minimal (worst) value; (right) mean value for ten sources. From top to bottom - number of overlapping sources *K*: 2, 3, and 4. Most probable value of the nonzero state, generated according to exponential distribution, equal to 0.01.

Figure 3. Structures of components $\mathbf{C}_1$-$\mathbf{C}_{10}$.

Table Captions

Algorithm 1. The NPT-NMF (preferably NMU) algorithm.

Required:

$\mathbf{X} \in \mathbb{R}_{0+}^{N \times T}$. If A1) is not satisfied perform scaling

$\mathbf{X} \to \mathbf{X} \Big/ \arg\max_{t} \left\{ \left\| \mathbf{x}_{\cdot t} \right\|_1 \right\}_{t=1}^{T}$ or $\mathbf{X} \to \mathbf{X} \Big/ \arg\max_{nt} \left\{ \mathbf{X}_{nt} \right\}_{n,t=1}^{N,T}$.

1. Perform mapping $\psi(\mathbf{X})$ in (7)/(8).
2. Optionally, apply "de-noising" by soft- or hard thresholding operator entry-wise on $\psi(\mathbf{X})$ in (8) with $\tau=10^{-7}$.
3. Apply selected NMF algorithm to $\mathbf{X}$ in (1) and $\psi(\mathbf{X})$ in (8) to estimate, respectively, $\mathbf{S}$ and $\bar{\mathbf{S}}$.
4. Compare estimated $\mathbf{S}$ and $\bar{\mathbf{S}}$ with the reference components in the library to obtain the final estimate of $\mathbf{S}$.

Table 1. Maximal normalized correlation coefficients between analytes $\mathbf{C}_1$ to $\mathbf{C}_{10}$ and components extracted by the proposed NPT-NMU method. Columns from left to right: correlation coefficients; combinations of mixture spectra. The star '*' denotes analytes in the library associated with the same extracted component. As expected highly correlated analytes $\mathbf{C}_4$ and $\mathbf{C}_7$ were associated with the same extracted component.

	$\mathbf{X}_1$ to $\mathbf{X}_5$ NMF_L0	$\mathbf{X}_1$ to $\mathbf{X}_5$ NPT-NMU	$\mathbf{X}_{(1,3,4,5)}$ NPT-NMU	$\mathbf{X}_{(1,2,3)}$ NPT-NMU	$\mathbf{X}_{(3,4)}$ NPT-NMU
$c_{1,1}$	0.7269	0.8792	0.8486	0.7194	0.6232
$c_{2,2}$	0.9567	0.9370	0.8484	0.8855	0.8479
$c_{3,3}$	0.7448	0.9160	0.9142	0.6495	0.6889
$c_{4,4}$	0.8595	0.9816*	0.9008*	0.7474*	0.7308*
$c_{5,5}$	0.5616	0.6994	0.6107	0.5863	0.6461
$c_{6,6}$	0.9922	0.9844	0.9160	0.7958	0.9386
$c_{7,7}$	0.7117	0.9684*	0.8993*	0.7830*	0.7621*
$c_{8,8}$	0.6401	0.9869	0.9826	0.9671	0.9318
$c_{9,9}$	0.9924	0.9194	0.8746	0.9413	0.7998
$c_{10,10}$	0.9880	0.9398	0.9359	0.9826	0.8085
Kernel variance	Does not apply	1.0	1.0	1.0	0.1
Scaled by	$\max_{nt}\{\mathbf{X}_{nt}\}_{n,t}^{N,}$	$\max_{nt}\{\mathbf{X}_{nt}\}_{n,t=1}^{N,T}$	$\max_{nt}\{\mathbf{X}_{nt}\}_{n,t=1}^{N,T}$	$\max_{nt}\{\mathbf{X}_{nt}\}_{n,t=1}^{N,T}$	$\max_{t}\{\lVert\mathbf{x}_{\bullet t}\rVert_1\}_{t=1}^{T}$
"Denoising"	Does not apply	Hard thresholding $\tau=10^{-7}$	Hard thresholding $\tau=10^{-7}$	Hard thresholding $\tau=10^{-7}$	Hard thresholding $\tau=10^{-7}$

Nonlinear Mixture-wise Expansion Approach to Underdetermined Blind Separation of Nonnegative Dependent Sources

Ivica Kopriva[1*], Ivanka Jerić[2], and Lidija Brkljačić[2]

Ruđer Bošković Institute, Bijenička cesta 54, HR-10000, Zagreb, Croatia

[1]Division of Laser and Atomic Research and Development

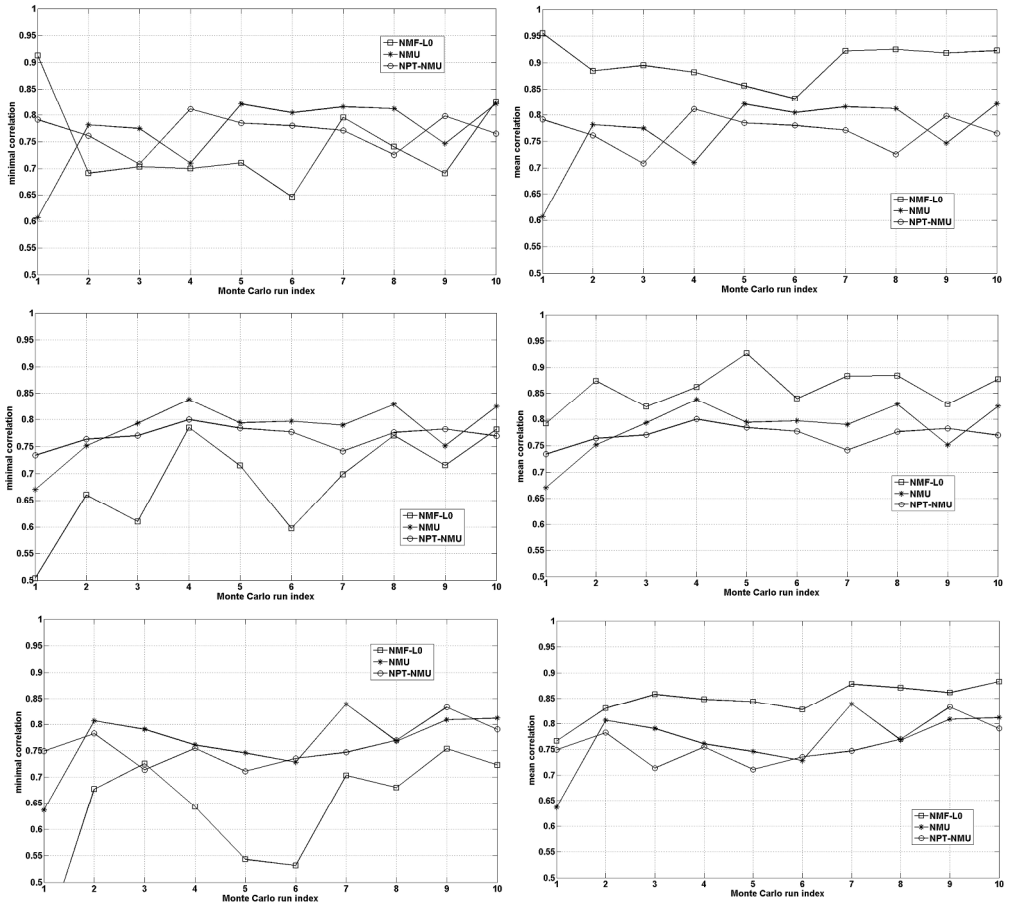
phone: +385-1-4571-286, fax:+385-1-4680-104

e-mail: ikopriva@irb.hr

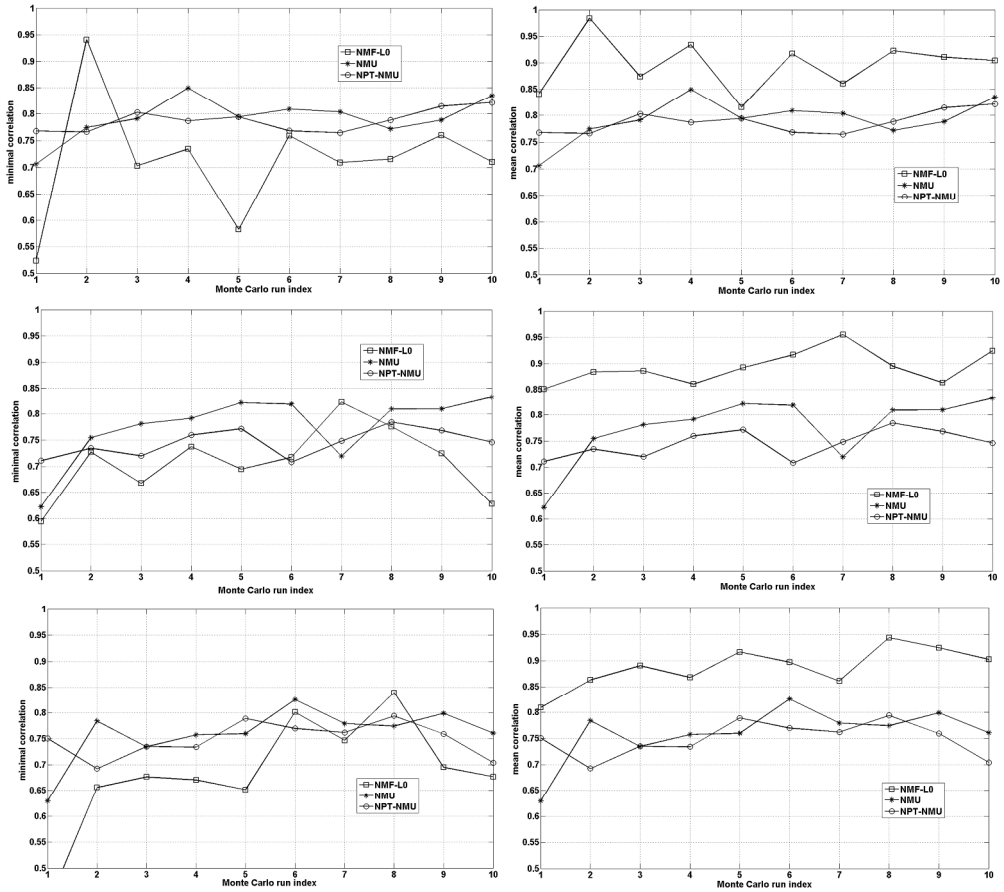
[2]Division of Organic Chemistry and Biochemistry

e-mail: ijeric@irb.hr, Lidija.Brkljacic@irb.hr

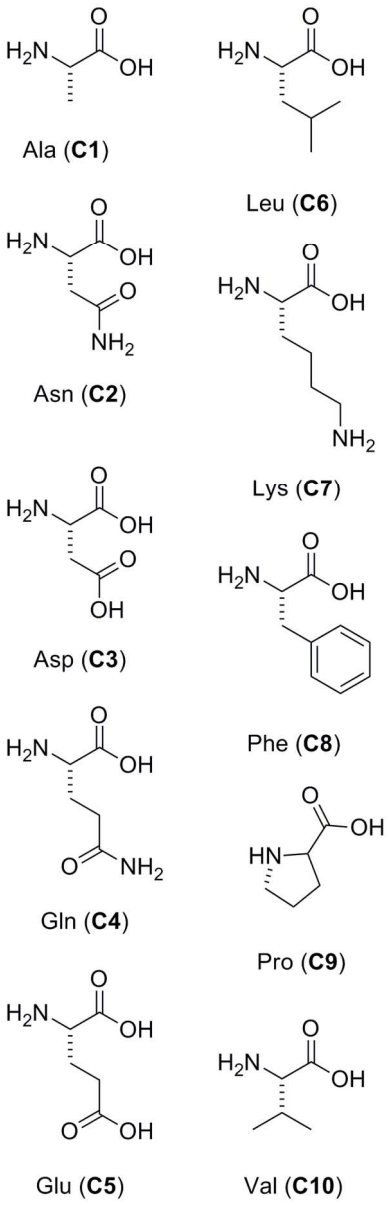
Summary abstract. A method for underdetermined blind separation of nonnegative dependent sources is proposed. The method performs nonlinear mixture-wise mapping of observed data and sparseness constrained nonnegative matrix factorization (NMF) in high-dimensional mapped space. Proposed method can be applied with existing NMF algorithms to extract analytes from mass spectra of multicomponent mixtures in biomarker related studies of biological samples.



Normalized correlation coefficient vs. Monte Carlo run index between true and extracted sources by algorithms: NMF_L0 (squares), NMU (stars) and NPT-NMU (circles). Left: minimal (worst) value; (right) mean value for ten sources. From top to bottom - number of overlapping sources K: 2, 3, and 4. Most probable value of the nonzero state, generated according to exponential distribution, equal to 0.1.
169x152mm (600 x 600 DPI)



Normalized correlation coefficient vs. Monte Carlo run index between true and extracted sources by algorithms: NMF_L0 (squares), NMU (stars) and NPT-NMU (circles). Left: minimal (worst) value; (right) mean value for ten sources. From top to bottom - number of overlapping sources K: 2, 3, and 4. Most probable value of the nonzero state, generated according to exponential distribution, equal to 0.01. 173x154mm (600 x 600 DPI)



Structures of components C1 to C10.
58x179mm (300 x 300 DPI)