Explicit-Implicit Mapping Approach to Nonlinear Blind Separation of Sparse Nonnegative Dependent Sources from a Single-Mixture: Pure Component Extraction from Nonlinear Mixture Mass Spectra

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

The nonlinear, nonnegative single-mixture blind source separation (BSS) problem consists of decomposing observed nonlinearly mixed multicomponent signal into nonnegative dependent component (source) signals. The problem is difficult and is a special case of the underdetermined BSS problem. However, it is practically relevant for

the contemporary metabolic profiling of biological samples when only one sample is available for acquiring mass spectra; afterwards, the pure components are extracted. Herein, we present a method for the blind separation of nonnegative dependent sources from a single, nonlinear mixture. First, an explicit feature map is used to map a single mixture into a pseudo multi-mixture. Second, an empirical kernel map is used for implicit mapping of a pseudo multi-mixture into a high-dimensional reproducible kernel Hilbert space (RKHS). Under sparse probabilistic conditions that were previously imposed on sources, the single-mixture nonlinear problem is converted into an equivalent linear, multiple-mixture problem that consists of the original sources and their higher order monomials. These monomials are suppressed by robust principal component analysis, hard-, soft- and trimmed thresholding. Sparseness constrained nonnegative matrix factorizations in RKHS yield sets of separated components. Afterwards, separated components are annotated with the pure components from the library using the maximal correlation criterion. The proposed method is depicted with a numerical example that is related to the extraction of 8 dependent components from 1 nonlinear mixture. The method is further demonstrated on 3 nonlinear chemical reactions of peptide synthesis in which 25, 19 and 28 dependent analytes are extracted from 1 nonlinear mixture mass spectra. The goal application of the proposed method is, in combination with other separation techniques, mass spectrometry-based non-targeted metabolic profiling, such as biomarker identification studies.

1. INTRODUCTION

Metabolomics is the measurement and analysis of metabolites, such as amino acids, carbohydrates and lipids from biofluids, and cellular extracts [1]. This technique has been used to understand the pathology of disease [2] and/or to diagnose disease [3-5]. Its unique value in disease diagnosis and drug discovery is based on the fact that changes in the metabolome are direct outcomes of the perturbations in cellular activity [1]. The two major instrument platforms for measuring the metabolite levels in biological samples are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) [1]. In particular, the non-targeted approach to metabolite profiling[1] involves the use of NMR and/or MS to simultaneously measure as many metabolites as possible in a biological specimen. Thereby, candidates for biomarkers are typically obtained through matching the acquired spectra with metabolites from the library [1, 6-8], such as the BioMagResBank metabolomics database [9] or, in the case of mass spectrometry, the NIST 14 Mass Spectral Library [10]. Biomarkers have great potential to improve medical treatment and reduce the costs of healthcare; however, few biomarkers have been validated for routine clinical practice [11]. Numerous pitfalls hamper the smooth transfer from biomarker discovery to everyday clinical utility. While the ideal biomarker should be reproducible, specific and sensitive, the number of samples necessary for biomarker validation remains questionable. The small number of analyzed samples is one of the factors contributing to failure in finding highly specific, sensitive biomarkers. However, metabolomic experiments are expensive, and because

---

[1] Metabolic profiling aims to identify and quantify small-molecule metabolites (a.k.a. analytes, pure components or sources) present in biological samples (typically urine, serum or biological tissue extract). Non-targeted approach to metabolic profiling implies that no specific metabolite is targeted, i.e. that all metabolites present in a sample are possible targets.

sampling many subjects is time-consuming, the determination of an optimal sample size is of high importance [12]. The quality of annotation, measured by the occurrence of false positives and/or false negatives, depends on how well pure components are separated from one another. In this respect, the presence of a large number of metabolites in the studied samples makes metabolic profiling, which is one of the most challenging tasks in chemical biology [13], notoriously difficult. When hundreds of metabolites are present in the spectra of biological specimens, the peak overlap is high, making them difficult to separate. Progress in mass spectrometry-based technologies has uncovered the complexity of biological samples. For example, analysis of the human adult urinary metabolome by liquid chromatography-mass spectrometry (LC-MS) has revealed the presence of 1484 components, while 384 of them were characterized by matching their spectra with the references stored in libraries [6]. Therefore, separation techniques are crucial for reducing the false alarm rate during the annotation process, which increases the accuracy of biomarker identification studies. In this respect, blind source separation (BSS) methods that aim to extract the component spectra using the recorded mixtures mass spectra alone are of special interest [14-18]. While the great majority of BSS methods assume that multi-component mixtures are linear combinations of, at most, the same number of components [19-22], recent work in [18] has addressed the underdetermined nonlinear nonnegative BSS (uNNBSS) problem related to the extraction of a greater number of sparse dependent nonnegative components from a smaller number of nonlinear mixture mass spectra. That work has been motivated by the aforementioned complexity of biological samples as well as by the fact that a nonlinear model offers a more accurate description of the processes and interactions occurring in biological systems [23]. The development of BSS methods for

underdetermined linear [17] and nonlinear [18] problems has been motivated by the practical reasons stemming from the fact that the number of biological samples available is small compared with the number of components expected to be present in them. The number of samples is limited by the availability, laboratory practice—particularly when complex protocols are required—and ethical issues [24]. In this respect, one can speculate further regarding whether the multiple biological samples that are required by multivariate uBSS methods are always available. The present work addresses the problem of blind separation of pure components from a *single* multicomponent *nonlinear* mixture mass spectrum. Thereby, it is assumed that the components are nonnegative and, as experimentally validated for the pure component mass spectra [18], sparse in support and amplitude.[2] Given that linear [17] and nonlinear [18] underdetermined BSS problems are difficult, the single-mixture nonlinear BSS problem is exceptionally difficult. Even linear BSS from a single mixture is a highly ill-posed problem, and hard constraints must be imposed on the source signals to enable their separation [25-35]. To the best of our knowledge, there is no other method proposed for the nonlinear BSS from a single mixture. Under assumed sparse prior conditions on the sources, a nonlinear mixture is approximated with a linear one that consists of the original sources and their second order monomials. As in [18], the influence of higher order monomials, which represent error terms, is reduced by preprocessing the mixture spectra using robust principal component analysis (RPCA) [36, 37] as well as hard- (HT), soft- (ST) [38] and trimmed thresholding (TT) [39].

---

[2] Even though nonlinear mixing system in assumption A4 is required to be nonnegative it is not strictly required by proposed method. That is, in equivalent linear blind source separation problem (4) and explicit feature mapped problem (9)/(10) the mixing coefficients can be even negative. That is because empirical kernel map-based mapping in (12)-(15) ensures nonnegativity of the mixtures in mapped space and enables application of nonnegativity constrained matrix factorization to separate pure components.

Algorithms for single-mixture BSS first have to transform the single- to the pseudo multi-mixture BSS problem [25-35]. Subsequently, some existing multivariate algorithms are used to perform BSS. Herein, we used an explicit feature map (EFM) for observation-wise nonlinear mapping of the recorded mixture mass spectra into pseudo multiple mixtures spectra[3]. That procedure was followed by the reduction of EFM-induced higher order monomials using RPCA, HT, ST and TT. Preprocessed pseudo multiple mixture data are mapped observation-wise in high-dimensional RKHS using an empirical kernel map (EKM). This step is the same as that used in [18] to address the uNNBSS problem. The combination of EFM and EKM based nonlinear mappings with sparse prior imposed on source components enables nonlinear BSS from a single mixture and represents the original contribution of the paper. Due to RPCA, HT, ST and TT methods used to reduce the high order monomial terms, one nonlinear nonnegative single-mixture BSS problem is actually converted into four nonnegative pseudo multiple mixture linear BSS problems in RKHS. Sparseness constrained NMF is performed in RKHS to solve these nonnegative linear BSS problems. Accordingly, the components separated by NMF are annotated with the pure components from the library using the maximal correlation criterion.

The proposed single-mixture nonlinear BSS algorithm differs from the existing single-mixture BSS algorithms in the following aspects: (*i*) algorithms [25-35] address the linear BSS problem, while the proposed method addresses the nonlinear BSS problem and (*ii*) the hard constraints imposed on the source signals by single-mixture

---

[3] Within the scope of this paper mixture refers to mass spectrum of a sample. Thus, observation index corresponds with the mass-charge ratio (*m/z*) index of the mass spectra. Sometimes we use terminology sample as replacement for mixture. As opposed to that, the BSS community assumes each complete measured mixed spectrum as a mixture, observation or channel. Furthermore, the BSS community assumes value of one mixed signal or a vector at particular independent variable as a sample.

BSS algorithms [25-35] do not apply to the pure component mass spectra that are of interest in this study. This statement is supported through the following analysis. The method [25] assumed that the source signals have disjoint support. The method [25] partitions single-channel time series to yield a pseudo multichannel mixture, and an independent component analysis (ICA) algorithm was then applied to extract the sources. The disjoint support assumption does not hold for the overlapped pure component mass spectra. The algorithm [26] used empirical mode decomposition [41] to decompose the single-channel mixture into intrinsic mode functions (IMFs) that represent the pseudo multichannel mixture. For separation by ICA algorithms, sources of interest are required to be IMFs, which does not hold for the pure component mass spectra. In [27], the wavelet transform is used to generate a pseudo multichannel mixture from a single-channel version. In this way, mother wavelets have to be non-orthogonal and have to match the shapes to the sources of interest. Thus, this wavelet-ICA method is applicable to the separation of the specific source signals, such as vibration signals [28, 29]. Many of single-channel BSS algorithms are derived to separate acoustic signals by factorizing a nonnegative spectrogram (magnitude of the short time Fourier transform) [30-35].

The remainder of the paper is organized as follows. Section 2 presents the theory upon which the proposed single-mixture nonlinear BSS method is built. Section 3 describes the experiments performed on computational and experimental data. Section 4 presents and discusses the results of comparative performance analysis between the proposed single-mixture nonlinear BSS method and state-of-the-art multiple-mixture linear NMF [40] and nonlinear BSS [18] algorithms. Concluding remarks are given in Section 5.

2. THEORY AND ALGORITHM

The goal application of the proposed single-mixture nonlinear BSS method is the extraction of analytes from a single multicomponent nonlinear mixture mass spectrum. In this respect, the method is complementary to the methods developed for linear [17] and nonlinear [18] underdetermined BSS problems. We justify the application of this method to mass spectrometry with the previously discussed reasons. Together with NMR spectroscopy, it is a major platform for measuring the metabolite levels in biological samples [1, 6]. As in [6, 17, 18], we assume that the library of reference mass spectra is available to annotate the components extracted by the proposed method.

It has already been discussed in [42, 17, 18] that the performance of many NMF algorithms depends on the optimal usage of parameters that are required to be known *a priori*, such as a balance parameter that regulates the influence of the sparseness constraint [43], or the number of overlapping components in mixtures [44]. Because these parameters are difficult to optimally select in practice, we have selected, as in [17, 18], the nonnegative matrix underapproximation (NMU) algorithm [40] to solve nonnegative BSS problems in RKHS. As such, the NMU algorithm does not require *a priori* information from the user. In summary, we propose the combination of RPCA, HT, ST and TT preprocessing transforms; EFM and EKM based nonlinear mappings with the NMU algorithm in mapping-induced high-dimensional RKHS, resulting in the PTs-EFM-EKM-NMU algorithm. The performance of the PTs-EFM-EKM-NMU is compared on one numerical and three experimental problems with the linear multivariate NMU algorithm and nonlinear multivariate PTs-EKM-NMU algorithm [18].

2.1 Single-mixture nonlinear blind separation of nonnegative dependent sources

The single-mixture NNBSS problem with dependent sources is described as:

$$x_t = f(\mathbf{s}_t) \qquad t = 1,...,T \tag{1}$$

where $x_t \in R_{0+}$ stands for the nonnegative intensity acquired at some of the T mass-to-charge (m/z) channels and $\mathbf{s}_t \in R_{0+}^{M \times 1}$ is the unknown vector that consists of the intensities of M nonnegative sources. $f : R_{0+}^{M} \to R_{0+}$ is an unknown multivariate mapping. Problem (1) can be applied to the vector-matrix framework:

$$\mathbf{x} = f(\mathbf{S}) \tag{2}$$

such that $\mathbf{x} \in R_{0+}^{1 \times T}$ and $\mathbf{S} \in R_{0+}^{M \times T}$, where $\{\mathbf{s}_t\}_{t=1}^{T}$ are column vectors of the matrix $\mathbf{S}$ and $f(\mathbf{S})$ implies that nonlinear mapping is performed column wise, such as in (1). It is further assumed that: $\{\|\mathbf{s}_t\|_0 \leq L\}_{t=1}^{T}$ where $\|\mathbf{s}_t\|_0$ stands for the ℓ_0 quasi-norm that counts the number of non-zero coefficients of $\mathbf{s}_t$ and $L = \max_{t=1,...,T} \|\mathbf{s}_t\|_0$. Evidently, it applies: $L \leq M$, where L denotes the maximal number of sources that can simultaneously be present at any coordinate t. Hence, L is related to a degree of dependence (correlation) between the sources. To avoid confusion between the column and row vectors, they will be indexed by lowercase letters that correspond to the uppercase letters, which are related to the dimensions of the corresponding matrix. For example, $\mathbf{s}_t$ refers to the column- and $\mathbf{s}_m$ to the row vector of matrix $\mathbf{S} \in R_{0+}^{M \times T}$. Evidently, uppercase bold letters denote matrices, lowercase bold letters denote vectors and italic lowercase letters denote scalars. The

single-mixture NNBSS problem implies that components mass spectra, $\left\{ \mathbf{s}_m \in R_{0+}^{1 \times T} \right\}_{m=1}^M$, should be inferred from the mixture data vector **x** only. Analogously with [17, 18], we make the following assumptions for the nonlinear mixture model (1)/(2):

A1) $0 \le x_t \le 1$ $\quad \forall t = 1, \ldots, T,$

A2) $0 \le s_{mt} \le 1$ $\quad \forall m = 1, \ldots, M$ and $\forall t = 1, \ldots, T,$

A3) Amplitude s_{mt} obeys an exponential distribution on the (0, 1] interval and a discrete distribution at zero, see also eq.(3),

A4) The vector valued function $f(\mathbf{s}_t)$: $f(\mathbf{s}_t) : R_{0+}^{M \times 1} \mapsto R_{0+}$ is differentiable up to an unknown order K,

A5) $M \!<\!< T.$

To be useful, the solution of the single-mixture NNBSS problem is expected to be essentially unique; as such, the estimated matrix of pure components (sources) $\hat{\mathbf{S}}$ and the true matrix of pure components $\mathbf{S}$ should be related through $\hat{\mathbf{S}} = \mathbf{P \Lambda S}$, where $\mathbf{P}$ and $\mathbf{\Lambda}$ stand, respectively, for the $M \times M$ permutation and diagonal matrices. Such a requirement was difficult to satisfy even for linear [17] and nonlinear [18] underdetermined but multivariate BSS problems. We obtained an estimate that satisfied the essential uniqueness condition only approximately as: $\hat{\mathbf{S}} \approx \mathbf{P \Lambda S}$. As discussed in [18], this finding has been ensured by imposing a hard constraint on pure components $\left\{ \mathbf{s}_m \right\}_{m=1}^M$ to comply with the sparse probabilistic model imposed by A3 [45, 46, 18, 17]:

$$p(s_{mt}) = \rho_m \delta(s_{mt}) + (1 - \rho_m) \delta^*(s_{mt}) f(s_{mt}) \quad \forall m = 1, \ldots, M \quad \forall t = 1, \ldots, T \qquad (3)$$

where $\delta(s_{mt})$ is an indicator function and $\delta^*(s_{mt}) = 1 - \delta(s_{mt})$ is its complementary function,

$\rho_m = \{P(\mathbf{s}_{mt} = 0)\}_{t=1}^{T}$. Hence, $\{P(s_{mt} > 0) = 1 - \rho_m\}_{t=1}^{T}$. The nonzero state of s_{mt} is

distributed according to $p(s_{mt})$. Exponential distribution $p(s_{mt}) = (1/\mu_m) \exp(-s_{mt}/\mu_m)$

was chosen in [18] to model the sparse distribution of the nonzero states; in this case,

the most probable outcomes are equal to μ_m. It has been verified in [18] that model (3)

accurately describes the mass spectra of the pure components and it has been estimated,

using mass spectra of 25 pure components, that $\hat{\rho}_m \in [0.27, 0.74]$ and

$\hat{\mu}_m \in [0.0012, 0.0014]$. Constraint A3, i.e., probabilistic model (3) with estimates for ρ_m

and μ_m, implies that each component will be zero at the large part of its support (number

of m/z channels T) as well as that the non-zero intensity will be distributed according to

the exponential distribution with a small expected value. These should ensure that, in

probability, the maximal number of analytes L present at the particular m/z coordinate is

sufficiently small compared to the number of sources M and number of mixtures N (that

for single-mixture scenario equals $N=1$). Obviously, the last requirement, $L<N$, is

impossible to satisfy in a single-mixture (univariate) scenario. In case of linear [17] and

nonlinear [18] underdetermined BSS problems, the number of measurements (samples)

N was increased by mapping the original problem into RKHS using EKM. However,

EKM-based mapping is not directly applicable to the single-mixture nonlinear BSS

problem implied by (1)/(2). For that to be achieved, (1)/(2) has to be approximated by

an equivalent linear single-mixture BSS problem and, afterwards, mapped to a pseudo

multivariate BSS problem. Under the experimentally verified sparse prior A3 nonlinear

mixture model (2) will after Taylor expansion around reference point $\mathbf{S}=\mathbf{0}$ retain second

order terms only [18]:

$$\mathbf{x} = \mathbf{g}^1 \mathbf{S} + \frac{1}{2} \mathbf{g}^2_{(1)} \begin{bmatrix} \mathbf{s}_1^2 \\ ... \\ \mathbf{s}_M^2 \\ .... \\ \left\{ \mathbf{s}_{m_1} \mathbf{s}_{m_2} \right\}_{m_1,m_2=1}^{M} \end{bmatrix} + HOT \qquad (4)$$

where $\mathbf{g}^1$ stands for the vector of first order partial derivatives $\partial f / \partial s_m$, $m=1,...,M$ and

$\mathbf{g}^2_{(1)}$ stands for the vectorized tensor of the second order partial derivatives $\partial^2 f / \partial s_i \partial s_j$,

$i, j =1,..,M$, while *HOT* stands for the higher-order terms. As analyzed in [18], the

contribution of third order terms in (4) is on the order of 3.43×10^{-7}. To reduce *HOT*

entry-wise, thresholding of $\mathbf{x}$ can be performed. By neglecting fourth- and higher-order

terms we have, in [18], empirically arrived at the threshold value of: $\tau \in [10^{-6}, 10^{-4}]$. As

in [16], we apply RPCA, HT, ST and TT on $\mathbf{x}$ in (2)/(4) to suppress *HOT*, where the

threshold value for HT, ST and TT was set to $\tau=10^{-5}$, exponent value for TT to $\alpha=3.5$

and regularization constant for RPCA to $\lambda=0.05$. Thereby, RPCA, HT, ST and TT are

applied either sequentially or in parallel to $\mathbf{x}$. We refer the interested reader to section

2.3 in [18], [36-39] as well as in Supporting Information for further details related to the

suppression of *HOT*. The suppression of *HOT* in (2)/(4) yields:

$$\mathbf{x} \overset{RPCA}{\longmapsto} \mathbf{a}, \quad \mathbf{x} \overset{HT}{\longmapsto} \mathbf{b}, \quad \mathbf{x} \overset{ST}{\longmapsto} \mathbf{c}, \quad \mathbf{x} \overset{TT}{\longmapsto} \mathbf{d} \tag{5}$$

where:

$$\mathbf{a} \approx \mathbf{g}^1 \mathbf{S} + \frac{1}{2} \mathbf{g}^2_{(1)} \begin{bmatrix} \mathbf{s}_1^2 \\ ... \\ \mathbf{s}_M^2 \\ .... \\ \left\{ \mathbf{s}_{m_1} \mathbf{s}_{m_2} \right\}_{m_1, m_2 = 1}^{M} \end{bmatrix} \tag{6}$$

and structurally, the same model applies for $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$ in (5).

2.2 Explicit feature map-based pseudo multivariate BSS problem

To enable extraction of pure components by the multivariate BSS method, the *HOT* suppressed single-mixture data $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{d}$ in (6) need to be mapped to pseudo multivariate data. Unlike the methods in [25-35], we used an explicit feature map of some order D. In what follows, we illustrate it for the *HOT*-suppressed vector $\mathbf{a}$ (6), but the same methodology applies to vectors $\mathbf{b}, \mathbf{c}$ and $\mathbf{d}$. Therefore, vector $\mathbf{a}$ is mapped observation-wise according to: $\left\{ a_t \mapsto \phi_D\left(a_t\right) \in R_{0+}^{D+1 \times 1} \right\}_{t=1}^{T}$ such that[4]:

$$\phi_D\left(a_t\right) = \begin{bmatrix} 1 & a_t & a_t^2 ... a_t^D \end{bmatrix}^{\mathrm{T}} \quad \forall t = 1, ..., T \tag{7}$$

---

[4] In addition to nonlinear mapping (7) other types of nonlinear mappings can be used for EFM as well. One way is to construct approximate EFM through factorization of chosen kernel function using reproducibility condition, a.k.a. *kernel trick*, $\kappa\left(\mathbf{x}, \mathbf{y}\right) = \left\langle \phi\left(\mathbf{x}\right), \phi\left(\mathbf{y}\right) \right\rangle$ where $\left\langle \circ, \circ \right\rangle$ stands for inner product in induced RKHS.

Using (6) and the multinomial theorem, we obtain:

$$a_t^d = \left(\sum_{m=1}^M g_m s_{mt} + \sum_{m=1}^M g_m^2 s_{mt}^2 + \sum_{\substack{m,n=1\\m\neq n}}^M g_{i(m,n)} s_{mt} s_{nt} \right)^d$$

$$= \sum_{k_1+k_2+...+k_M=d} \frac{d!}{k_1! k_2!...k_M!} \prod_{1\le i\le M} \left(\sum_{m=1}^M g_m s_{mt} + \sum_{m=1}^M g_m^2 s_{mt}^2 + \sum_{\substack{m,n=1\\m\neq n}}^M g_{i(m,n)} s_{mt} s_{nt} \right)^{k_i} \quad d=0,...,D \tag{8}$$

If the sparseness assumption A3, eq (3), is activated, we may approximate (8) by using monomials $\{s_{mt}^{k_m} s_{nt}^{k_n}\}$ up to the second order only, i.e., $0\le k_m+k_n\le 2$.[5] Therefore, we can write (7) on the observation (m/z) index level as:

$$\phi_D(a_t) = \begin{bmatrix} 1 \\ \mathbf{A} \end{bmatrix} \begin{bmatrix} 1 \\ s_{1t} \\ ... \\ s_{Mt} \\ s_{1t}^2 \\ ... \\ s_{Mt}^2 \\ \{s_{mt} s_{nt}\}_{m\neq n=1}^M \end{bmatrix} + HOT \ \forall t=1,...,T \tag{9}$$

where $\mathbf{A}$ is matrix of appropriate dimensions. The model (9) applies analogously to $\phi_D(b_t)$, $\phi_D(c_t)$ and $\phi_D(d_t)$. We can also write (9) in matrix version as:

---

[5] Sparsity according to A3, eq.(3), is both in term of support and in term of amplitude. Due to first term, products of the order higher than 2 will mostly vanish because it is highly probable that at least one component will be zero. If that is not the case, then product of 3 or more components with small amplitude values will be very small.

$$\phi_D(\mathbf{a}) = \begin{bmatrix} 1 \\ \mathbf{A} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{s}_1 \\ \dots \\ \mathbf{s}_M \\ \mathbf{s}_1^2 \\ \dots \\ \mathbf{s}_M^2 \\ \{\mathbf{s}_m \mathbf{s}_n\}_{m \neq n=1}^{M} \end{bmatrix} + HOT \quad \forall t = 1,\dots,T \tag{10}$$

such that $\phi_D(\mathbf{a}) \in R_{0+}^{D+1 \times T}$. Matrices $\phi_D(\mathbf{b})$, $\phi_D(\mathbf{c})$ and $\phi_D(\mathbf{d})$ are structurally equivalent to $\phi_D(\mathbf{a})$ in (10). Due to difference in methods used for suppressing *HOT* in (4)/(5), the corresponding basis matrices will be different for structurally equivalent models of $\phi_D(\mathbf{a})$, $\phi_D(\mathbf{b})$, $\phi_D(\mathbf{c})$ and $\phi_D(\mathbf{d})$. To suppress *HOT*, we apply RPCA, HT, ST and TT methods, respectively, to $\phi_D(\mathbf{a})$, $\phi_D(\mathbf{b})$, $\phi_D(\mathbf{c})$ and $\phi_D(\mathbf{d})$. To keep the notational complexity low, we use the same notation for *HOT*-suppressed matrices, i.e.:

$$\phi_D(\mathbf{a}) \overset{RPCA}{\longmapsto} \phi_D(\mathbf{a}), \phi_D(\mathbf{b}) \overset{HT}{\longmapsto} \phi_D(\mathbf{b}), \phi_D(\mathbf{c}) \overset{ST}{\longmapsto} \phi_D(\mathbf{c}), \ \phi_D(\mathbf{d}) \overset{TT}{\longmapsto} \phi_D(\mathbf{d}) \tag{11}$$

$\phi_D(\mathbf{a})$, $\phi_D(\mathbf{b})$, $\phi_D(\mathbf{c})$ and $\phi_D(\mathbf{d})$ now stand for pseudo multivariate versions of the single-mixture nonlinear BSS problem (1)/(2).

2.3 Empirical kernel map based mapping and sparseness constrained NMF

We can use now the EKM-based methodology in the same way as described in section 2.4 in [17]. Nevertheless, to have the present paper self-contained, we shall briefly describe the EKM-based mapping step. Let us illustrate the EKM concept on $\phi_D(\mathbf{a})$. The procedure is the same for $\phi_D(\mathbf{b})$, $\phi_D(\mathbf{c})$ and $\phi_D(\mathbf{d})$ in (11). We select each observation $\phi_D(a_t)$ as a basis vector, i.e., $\{\mathbf{v}_t = \phi_D(a_t)\}_{t=1}^{T}$. Therefore, $span\{\mathbf{v}_t\}_{t=1}^{T} = span\{\phi_D(a_t)\}_{t=1}^{T}$. We now assume that EFM, possibly different from $\phi_D(a_t)$, is applied to $\phi_D(a_t)$, i.e., $\varphi(a_t) \doteq \varphi(\phi_D(a_t))$. Thereby, symbol $\doteq$ means "by definition". We assume that $\varphi(a_t)$ is associated with its kernel function through the reproducibility condition, a.k.a. *kernel trick*: $\kappa(\mathbf{x},\mathbf{y}) = \langle \varphi(\mathbf{x}), \varphi(\mathbf{y}) \rangle$. Then, the EKM with respect to the basis $\mathbf{V}$ is defined as[6]:

$$\Psi(\phi_D(a_t)) = \begin{bmatrix} \kappa(\varphi(a_t), \mathbf{v}_1) \\ ... \\ \kappa(\varphi(a_t), \mathbf{v}_T) \end{bmatrix} \quad \forall t = 1,...,T \tag{12}$$

or, on the matrix level:

$$\Psi(\phi_D(\mathbf{a})) = \begin{bmatrix} \kappa(\varphi(a_1), \mathbf{v}_1) & ... & \kappa(\varphi(a_T), \mathbf{v}_1) \\ ... & ... & ... \\ \kappa(\varphi(a_1), \mathbf{v}_T) & ... & \kappa(\varphi(a_T), \mathbf{v}_T) \end{bmatrix} \tag{13}$$

---

[6] For formal definition of empirical kernel map we refer to definition 2.2.3 in [17], i.e. to definition 2.15 in [47].

where $\Psi\left(\phi_D\left(\mathbf{a}\right)\right) \in R_{0+}^{T \times T}$. We use the same procedure to obtain $\Psi\left(\phi_D\left(\mathbf{b}\right)\right)$, $\Psi\left(\phi_D\left(\mathbf{c}\right)\right)$

and $\Psi\left(\phi_D\left(\mathbf{d}\right)\right)$. Herein, as in [17, 18], we choose

$\kappa\left(\varphi\left(a_t\right), \mathbf{v}_d\right) = \exp\left(-\left\|\phi_D\left(a_t\right) - \mathbf{v}_d\right\|^2 / \sigma^2\right)$.[7] When assumption A1 holds, we can set

$\sigma^2 \approx 1$. Using the sparseness assumption A3 and eq. (3), we arrive at [18, 17]:

$$\Psi\left(\phi_D\left(a_t\right)\right) = \mathbf{Z} + \overline{\mathbf{A}}\begin{bmatrix} \mathbf{0}_{1 \times T} \\ \overline{\mathbf{S}} \end{bmatrix} + HOT \tag{14}$$

where $\mathbf{Z}$ is a bias term and does not play a role in the parts based decomposition that

follows, $\mathbf{0}_{1 \times T}$ is a row vector of zeros and $\overline{\mathbf{S}} \in R_{0+}^{P \times T}$ is a matrix with $P \approx 2M + M(M-1)/2$

rows that contain the original source components and their second order monomials. $\overline{\mathbf{A}}$

---

[7] We have selected Gaussian kernel since it is used dominantly in machine learning. On the more fundamental (theoretical) level it is expected from the chosen kernel to be capable for approximation of the arbitrary target function g^* on some compact subset $\mathbf{X}$ of the input pattern space $\mathcal{X}$. In that case $\lim_{T \to \infty}\left(g\left(\mathbf{x}\right) := \sum_{t=1}^{T} c_t \kappa\left(\mathbf{x}, \mathbf{x}_t\right)\right) = g^*\left(\mathbf{x}\right)$. Such kernels are said to have universal approximation property. Examples of universal kernels are Gaussian $\kappa\left(\mathbf{x}, \mathbf{y}\right) = \exp\left(-\left\|\mathbf{x} - \mathbf{y}\right\|^2 / \sigma^2\right)$, Laplacian $\kappa\left(\mathbf{x}, \mathbf{y}\right) = \exp\left(-\left\|\mathbf{x} - \mathbf{y}\right\| / \sigma^2\right)$ and inverse multiquadratic $\kappa\left(\mathbf{x}, \mathbf{y}\right) = 1 / \sqrt{\left\|\mathbf{x} - \mathbf{y}\right\|^2 + c}$, $c > 0$, [48, 49]. Universal kernels belong to translation invariant group of kernels $\kappa\left(\mathbf{x}, \mathbf{y}\right) = \kappa\left(\mathbf{x} - \mathbf{y}\right)$. In this case induced RKHS $\mathcal{Y}$ is composed of functions $g \in L^2\left(\mathbb{R}^N\right)$ such that: $\int_{\mathbb{R}^N} \frac{\left|\hat{g}\left(\omega\right)\right|^2}{\hat{k}\left(\omega\right)} d\omega < \infty$ where $\hat{g}\left(\omega\right)$ if Fourier transform of g and $\hat{\kappa}\left(\omega\right)$ is Fourier transform of the kernel κ, [50]. This interpretation shows that functions in the RKHS $\mathcal{Y}$ induced by translation invariant kernels have a Fourier transform that decays very rapidly. Thus, RKHS $\mathcal{Y}$ is a rich space of smooth functions. As an example, Fourier transform of Gaussian kernel is given by $\hat{\kappa}\left(\omega\right) = \left(2\pi\sigma^2\right)^{N/2} \exp\left(-\frac{\sigma^2}{2}\left|\omega\right|^2\right)$, and RKHS $\mathcal{Y}$ induced by Gaussian kernel contains smooth functions with Fourier transform decaying very rapidly.

is a matrix of appropriate dimensions. EKM-mapped matrices $\Psi\big(\phi_D(\mathbf{b})\big)$, $\Psi\big(\phi_D(\mathbf{c})\big)$

and $\Psi\big(\phi_D(\mathbf{d})\big)$ structurally have the same approximation as (14). To suppress *HOT*, we

apply RPCA, HT, ST and TT methods, respectively, to $\Psi\big(\phi_D(\mathbf{a})\big)$, $\Psi\big(\phi_D(\mathbf{b})\big)$,

$\Psi\big(\phi_D(\mathbf{c})\big)$ and $\Psi\big(\phi_D(\mathbf{d})\big)$. To keep the notational complexity low, we use the same

notation for *HOT*-suppressed matrices, i.e.:

$$\Psi\big(\phi_D(\mathbf{a})\big)\overset{RPCA}{\longmapsto}\Psi\big(\phi_D(\mathbf{a})\big), \ \ \Psi\big(\phi_D(\mathbf{b})\big)\overset{HT}{\longmapsto}\Psi\big(\phi_D(\mathbf{b})\big),$$
$$\Psi\big(\phi_D(\mathbf{c})\big)\overset{ST}{\longmapsto}\Psi\big(\phi_D(\mathbf{c})\big), \ \ \Psi\big(\phi_D(\mathbf{d})\big)\overset{TT}{\longmapsto}\Psi\big(\phi_D(\mathbf{d})\big) \tag{15}$$

The sparseness constrained NMF is now applied to (15) to estimate $\overline{\mathbf{S}}^{\mathbf{a}}$, $\overline{\mathbf{S}}^{\mathbf{b}}$, $\overline{\mathbf{S}}^{\mathbf{c}}$ and $\overline{\mathbf{S}}^{\mathbf{d}}$,

respectively, from $\Psi\big(\phi_D(\mathbf{a})\big)$, $\Psi\big(\phi_D(\mathbf{b})\big)$, $\Psi\big(\phi_D(\mathbf{c})\big)$ and $\Psi\big(\phi_D(\mathbf{d})\big)$. This step is

equivalent to the one described in section 2.5 in [18]:

$$\overline{\mathbf{S}}^{\mathbf{a}} = sNMF\big(\Psi\big(\phi_D(\mathbf{a})\big)\big) \tag{16}$$

$$\overline{\mathbf{S}}^{\mathbf{b}} = sNMF\big(\Psi\big(\phi_D(\mathbf{b})\big)\big) \tag{17}$$

$$\overline{\mathbf{S}}^{\mathbf{c}} = sNMF\big(\Psi\big(\phi_D(\mathbf{c})\big)\big) \tag{18}$$

$$\overline{\mathbf{S}}^{\mathbf{d}} = sNMF\big(\Psi\big(\phi_D(\mathbf{d})\big)\big) \tag{19}$$

Hence, we have converted the original single mixture nonlinear BSS problem (1)/(2)

characterized by a triplet (*N*=1, *M*, *L*) to four linear BSS problems (16) to (19)

characterized by a triplet (*N*=*T*, *P*, *Q*), where *P* stands for the unknown number of

components present in $\Psi\big(\phi_D(\mathbf{a})\big)$, $\Psi\big(\phi_D(\mathbf{b})\big)$, $\Psi\big(\phi_D(\mathbf{c})\big)$ and/or $\Psi\big(\phi_D(\mathbf{d})\big)$.

Furthermore, *Q*>*L* stands for the maximum number of components that is

simultaneously present at some observation index $t=1,...,T$ in $\Psi\left(\phi_D(\mathbf{a})\right)$, $\Psi\left(\phi_D(\mathbf{b})\right)$, $\Psi\left(\phi_D(\mathbf{c})\right)$ and/or $\Psi\left(\phi_D(\mathbf{d})\right)$. Under the sparseness assumption A3, eq (3), it is justified to expect that compared to approximation (6) of the single-mixture BSS problem (1)/(2) the new BSS problems (16) to (19) satisfy:

$$(N=T)\gg(P/M) \text{ and } (N=T)\gg(Q/L) . \tag{20}$$

Therefore, in the light of the uniqueness conditions [51], the sparseness constrained factorization of (16) to (19) should be able to separate components $\overline{\mathbf{S}}^{\mathbf{a}}$, $\overline{\mathbf{S}}^{\mathbf{b}}$, $\overline{\mathbf{S}}^{\mathbf{c}}$ and $\overline{\mathbf{S}}^{\mathbf{d}}$. However, the quality of separation does not simply depend on the sparseness Q of the components in (16) to (19). It also depends on the mutual coherence $0 \leq \mu(\mathbf{A}) \leq 1$ of the basis matrix $\mathbf{A}$ in (10), i.e., $0 \leq \mu(\overline{\mathbf{A}}) \leq 1$ of $\overline{\mathbf{A}}$ in (14), [52]. The level of $\mu(\overline{\mathbf{A}})$, that is the maximal normalized correlation between a pair of basis vectors, is important in the presence of noise and/or modeling errors. These modeling errors are represented by *HOT* in (14). If δ denotes the ℓ_2-norm of *HOT* at the observation index *t,* the error in the estimation of $\overline{\mathbf{s}}_t$ by means of basis pursuit denoising algorithm is limited from above as [52]:

$$\left\|\hat{\overline{\mathbf{s}}}_t - \overline{\mathbf{s}}_t\right\|_2^2 \leq \frac{4\delta^2}{1+\mu(\overline{\mathbf{A}})(4Q-1)} .$$

Thereby, the required amount of sparseness Q of $\bar{\mathbf{s}}_t$ has to satisfy:

$$Q < \frac{1}{4}\left(1 + \frac{1}{\mu(\overline{\mathbf{A}})}\right)$$

Therefore, it is important that the value of $\mu(\overline{\mathbf{A}})$ is not too close to 1. In this respect, the value of $\mu(\overline{\mathbf{A}})$ depends in part on how distinct (different) the coefficients of $\mathbf{g}^1$ and $\mathbf{g}^2_{(1)}$ are in the approximation (6) of the original single-mixture BSS problem (1)/(2). This value depends on the properties of nonlinear mapping f in (1)/(2) as well as on the similarity of concentrations of pure components $\{\mathbf{s}_m\}_{m=1}^{M}$ in mixture $\mathbf{x}$. The value of $\mu(\overline{\mathbf{A}})$ also depends on the EFM mapping (7) and EKM mapping (12), i.e., they are expected to yield the basis matrix $\overline{\mathbf{A}}$ in (14) with $\mu(\overline{\mathbf{A}})$, which, hopefully, is not too close to 1. Due the structural equivalence of the approximation such as (14) for EKM-mapped matrices $\Psi(\phi_D(\mathbf{b}))$, $\Psi(\phi_D(\mathbf{c}))$ and $\Psi(\phi_D(\mathbf{d}))$, the conclusions presented in the above analysis also apply to problems (17) to (19).

The NMU algorithm [40] with a MATLAB code available at [53] was used for the sparseness constrained NMF in (16) to (19). A main reason for that is that there are no regularization constants in the NMU algorithm that require a tuning procedure. When performing NMU-based factorizations in (16) to (19), the unknown number P of components present in $\Psi(\phi_D(\mathbf{a}))$, $\Psi(\phi_D(\mathbf{b}))$, $\Psi(\phi_D(\mathbf{c}))$ and $\Psi(\phi_D(\mathbf{d}))$ needs to be used in the algorithm as an input. As in [18, 17] we set: $P=T$. To avoid losing a component, we prefer to extract all T rank-one factors.

Four sets of separated components are compared with the pure components stored in the library using the normalized correlation coefficient. Each pure component from the library $\{\mathbf{s}_m\}_{m=1}^{M}$ is annotated with the most similar separated component in $\{\overline{\mathbf{s}}_m^{\mathbf{a}}\}_{m=1}^{T}$, $\{\overline{\mathbf{s}}_m^{\mathbf{b}}\}_{m=1}^{T}$, $\{\overline{\mathbf{s}}_m^{\mathbf{c}}\}_{m=1}^{T}$ and $\{\overline{\mathbf{s}}_m^{\mathbf{d}}\}_{m=1}^{T}$. The normalized correlation coefficient is used as a similarity measure, yielding four sets of ordered pairs: $\{(\hat{\mathbf{s}}_m^{\mathbf{a}}, c_m^{\mathbf{a}})\}_{m=1}^{M}$, $\{(\hat{\mathbf{s}}_m^{\mathbf{b}}, c_m^{\mathbf{b}})\}_{m=1}^{M}$, $\{(\hat{\mathbf{s}}_m^{\mathbf{c}}, c_m^{\mathbf{c}})\}_{m=1}^{M}$ and $\{(\hat{\mathbf{s}}_m^{\mathbf{d}}, c_m^{\mathbf{d}})\}_{m=1}^{M}$. Thereby, $c_m^{\mathbf{a}}$, $c_m^{\mathbf{b}}$, $c_m^{\mathbf{c}}$, and $c_m^{\mathbf{d}}$, respectively, stand for the normalized correlation coefficients between the pure component $\mathbf{s}_m$ and the annotated components $\hat{\mathbf{s}}_m^{\mathbf{a}} \in \{\overline{\mathbf{s}}_m^{\mathbf{a}}\}_{m=1}^{T}$, $\hat{\mathbf{s}}_m^{\mathbf{b}} \in \{\overline{\mathbf{s}}_m^{\mathbf{b}}\}_{m=1}^{T}$, $\hat{\mathbf{s}}_m^{\mathbf{c}} \in \{\overline{\mathbf{s}}_m^{\mathbf{c}}\}_{m=1}^{T}$ and $\hat{\mathbf{s}}_m^{\mathbf{d}} \in \{\overline{\mathbf{s}}_m^{\mathbf{d}}\}_{m=1}^{T}$. The final annotation $\{(\hat{\mathbf{s}}_m^{\mathbf{f}}, c_m^{\mathbf{f}})\}_{m=1}^{M}$ is obtained such that:

$$\mathbf{f} = \underset{\{\mathbf{a,b,c,d}\}}{\arg\max} \{c_m^{\mathbf{a}}, c_m^{\mathbf{b}}, c_m^{\mathbf{c}}, c_m^{\mathbf{d}}\} \quad m = 1, \ldots, M \tag{21}$$

We summarize the PTs-EFM-EKM-NMU algorithm in Algorithm 1.

3. EXPERIMENTS

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

3.1 Numerical study

In a numerical study, we simulate a single-mixture NNBSS problem (2) with number of mixtures $N=1$, number of components $M=8$, number of overlapping components $L=3$ and number of realizations $T=1000$:

$$x_1 = f_1(\mathbf{s}) = s_1^3 + s_2^2 + \tan^{-1}(s_3) + s_4^2 + s_5^3 + s_6^3 + \tanh(s_7) + \sin(s_8)$$

Source signals were generated according to a mixed state probabilistic model (3) with an exponential prior. Thereby, $\mu_m=1.5\times10^{-3}$ $\forall m=1,...,M$. We generated a scenario with $\rho_m=0.8$ $\forall m=1,...,M$. The values for μ_m and ρ_m are equivalent to those obtained by fitting the probabilistic model (3) to an experimental mass spectra of 25 pure components. For more details, see section 3.2 and Figure 4 in [18]. As the performance of the PTs-EFM-EKM-NMU algorithm is compared with the performance of the multivariate NMU [40] and PTs-EKM-NMU [18] algorithms, we have, in addition to the single mixture required by (2), generated two more nonlinear mixtures as in [18]:

$$x_2 = f_2(\mathbf{s}) = \tanh(s_1) + s_2^3 + s_3^3 + \tan^{-1}(s_4) + \tanh(s_5) + \sin(s_6) + s_7^2 + s_8^2$$

$$x_3 = f_3(\mathbf{s}) = \sin(s_1) + \tan^{-1}(s_2) + s_3^2 + s_4^3 + \tanh(s_5) + \sin(s_6) + s_7^3 + \tan^{-1}(s_8)$$

Thus, the proposed PTs-EFM-EKM-NMU method was applied on mixture $\mathbf{x}_1$, while the NMU and PTs-EKM-NMU algorithms were applied on mixtures $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. The HT, ST and TT operators used in steps 1, 3 and 5 in Algorithm 1 were implemented with $\tau=10^{-5}$; $\alpha=3.5$ has been used for the TT operator and $\lambda=0.05$ has been used for the RPCA operator. The order of the EFM in (7) has been set to $D=8$. A Gaussian kernel based EKM in (13) has been used with $\sigma^2=1$ and $T=1000$.

3.2 Experimental data

Three chemical reactions for the formation of peptides were used as a source of experimental data. Generally, amino acids react under basic conditions in the presence of an activating agent, yielding di-, tri-, and tetrapeptides with different structures. The schematic presentation of chemical reactions and formation of new components as the reaction proceeds is given in Figure 2. Chemical reactions were analyzed with a HPLC-MS triple quadrupole instrument with an electrospray ionization (ESI) that was operated in a positive ion mode. Mixture mass spectra were acquired from aliquots of the reaction mixture that was withdrawn at different time points. The libraries of compounds required for the validation of the PTs-EFM-EKM-NMU algorithm were built by integrating each peak in the chromatogram, corresponding to the last acquired mixture mass spectrum. The first reaction consisted of 9 mixtures with 25 pure components. It has been used and described in section 3.2.1 in [18]. Therefore, we refer the reader to [18] as well as to the Supporting Information for a more detailed description of the chemical reactions. The structures of the possible components present in reaction mixture are schematically described in Figure S-1 in the Supporting Information as well as in Fig. 2 in [18]. Figure S-2 shows 9 chromatograms that correspond to the reaction mixtures recorded at 9 time instants (t_0-t_8) during the reaction. The mass spectra of 9 mixtures ($\mathbf{x}_1$ to $\mathbf{x}_9$), obtained by full integration of chromatograms, and mass spectra of 25 pure components ($\mathbf{s}_1$ to $\mathbf{s}_{25}$) arising during the reaction are, respectively, shown in Figures S-3 and S-4 in the Supporting Information. Mixture $\mathbf{x}_9$ was used by the proposed PTs-EFM-EKM-NMU algorithm. The mass spectra of pure components 1, 4, 8 and 11 are also shown in Figure 3. Table S-1 in the Supporting Information shows normalized correlations between 25 pure components,

where pairs of components with a correlation above 0.1 are displayed. The second reaction consists of 12 mixtures with 19 pure components. The structures of the possible components present in the reaction mixture are schematically described in Figure S-7 in the Supporting Information. Table S-2 in the Supporting Information shows the normalized correlations between 19 pure components, where pairs of components with correlations above 0.1 are displayed. Mixture $\mathbf{x}_{12}$ was used by the proposed PTs-EFM-EKM-NMU algorithm. The third reaction consists of 12 mixtures with 28 pure components identified. The structures of the possible components present in the reaction mixture are schematically drawn in Figure S-8 in the Supporting Information. Table S-3 in Supporting Information shows the normalized correlations between 28 pure components, where pairs of components with a correlation above 0.1 are displayed. Mixture $\mathbf{x}_{12}$ was used by the proposed PTs-EFM-EKM-NMU algorithm. The mass spectrometry measurements procedure is described in the Supporting Information as well as in section 3.2.2 in [18].

4. RESULTS AND DISCUSSION

4.1 Numerical study

Table 1 and Figure 1 show the results of comparative analysis for the case of ρ_m=0.8, which was obtained by NMU, PTs-NMU-EFM-EKM and PTs-EKM-NMU algorithms. We have used two performance measures, (i) the mean value of the normalized correlations between the extracted and correctly annotated pure components from the library. (Thereby, the sum of correlation coefficients of correctly annotated components was divided by the overall number of components, i.e., M=8, which penalizes incorrect

annotations) and (*ii*) the number of incorrect annotations that occur due to the lack of purity in the separated components. Despite of being applied on single-mixture alone, the proposed PTs-EFM-EKM-NMU method had the smallest number of incorrect annotations averaged over 10 runs, which was 0.6, compared to the 0.9 incorrect annotations by PTs-EKM-NMU and 3 incorrect annotations by NMU. Such result for the NMU algorithm was expected because it assumes a linear mixture model. The PTs-EFM-EKM-NMU method had a significantly smaller mean value of the normalized correlation coefficient, 0.38, which reflects the reliability of annotation, than the multivariate PTs-EKM-NMU algorithm [18], 0.58, which was applied on three mixtures. Overall, given the degree of difficulty of the problem of blind separation of 8 overlapping components from a single nonlinear mixture, the PTs-EFM-EKM-NMU algorithm yielded a useful result in terms of the small number of incorrect annotations.

4.2 Experimental data

The correlation structures of the pure components mass spectra for three chemical reactions are given, respectively, in Tables S-1, S-2 and S-3 in the Supporting Information, where the normalized cross-correlations above 0.1 were displayed. For the first reaction, 30 pairs of mass spectra of 25 pure components have a cross-correlation above 0.1. Therefore, the mass spectra of pure components significantly overlap, which makes it difficult to separate them. Furthermore, pure components 1 and 2, 16 and 17 as 19 and 21 have normalized correlation coefficients above 0.97 and, consequently, it is impossible to differentiate between them. In addition, pure components 5 and 7 have a normalized correlation coefficient above 0.78. Therefore, we do not expect from the

proposed PTs-EFM-EKM-NMU method, PTs-EKM-NMU method [18] or NMU method [40] to discriminate between these pure components. However, in spite of using one mixture ($\mathbf{x}_9$) only, the PTs-EFM-EKM-NMU method was able to correctly annotate the remainder of the components as well as PTs-EKM-NMU method could achieve with 9 mixtures. Nevertheless, the purity of components, measured by the mean normalized correlation, which was separated by the PTs-EFM-EKM-NMU, was significantly lower than the purity of components separated by the PTs-EKM-NMU method (0.467 vs. 0.702); see Table 2 for more details. Due to the inadequacy of the assumed linear mixture model, the NMU algorithm had 15 incorrect annotations even though it used 9 mixtures. Figure 3 shows the mass spectra of pure components 1, 4, 8 and 11 as well as the estimates obtained by the PTs-EFM-EKM-NMU method. The mass spectra of all 25 separated components are shown in Figure S-6 in the Supporting Information (Figure S-5 shows the mass spectra of the components separated by the PTs-EKM-NMU method [18]). Regarding the second nonlinear reaction, 42 pairs of mass spectra of 19 pure components have a cross-correlation above 0.1, making the mass spectra of pure components very difficult to separate. In addition, pure components 1 and 2 as well as 15 and 18 have a normalized correlation coefficient above 0.8. Therefore, we do not expect from the proposed PTs-EFM-EKM-NMU method, PTs-EKM-NMU method [18] or NMU method [40] to discriminate these pure components. As seen in Table 2, despite using one mixture ($\mathbf{x}_{12}$), the PTs-EFM-EKM-NMU method had only one incorrect annotation compared to all correct annotations by the PTs-EKM-NMU method, which used 12 mixtures, and 15 incorrect annotations by the NMU method, which also used 12 mixtures. Again, the purity of the components separated by PTs-EFM-EKM-NMU was significantly lower than the purity of the components separated

by the PTs-EKM-NMU method (0.439 vs. 0.702). Regarding the third nonlinear reaction, 31 pairs of mass spectra of 28 pure components have a cross-correlation above 0.1. Thereby, pure components 2, 3 and 27, 6 and 21, as well as 26 and 28, have normalized correlation coefficients above 0.7. Therefore, we do not expect from the proposed PTs-EFM-EKM-NMU method, PTs-EKM-NMU method [18] or NMU method [40] to discriminate these pure components. As seen in Table 2, despite using one mixture ($\mathbf{x}_{12}$) only, the PTs-EFM-EKM-NMU method had two incorrect annotations in the remaining components compared to all correct annotations by the PTs-EKM-NMU method, which used 12 mixtures, and 10 incorrect annotations by the NMU method, which also used 12 mixtures. As before, the purity of components separated by PTs-EFM-EKM-NMU was significantly lower than the purity of the components separated by the PTs-EKM-NMU method (0.368 vs. 0.656).

5. CONCLUSION

The blind source separation approach to pure component extraction is most often based on a linear mixture model in which the number of mixtures spectra is required to be greater than the number of pure components, which is not realistic. Herein, we present the method for separating pure components from single multicomponent nonlinear mixture mass spectra. Although the problem is very difficult from the mathematical point of view, it includes a special (degenerative) case of the underdetermined BSS problem, it is practically relevant for contemporary metabolic profiling of biological samples. In many situations, only one sample is available from which to acquire mass spectra; afterwards, the components are extracted. In particular, we have combined

explicit feature maps-based and implicit kernel-based mappings to formulate a method for the blind separation of sources from a single nonlinear mixture. The sources were constrained by the sparsity prior on the support and amplitude, which was experimentally verified on the mass spectra of pure components because the main aim of the proposed method is non-targeted metabolic profiling using mass spectrometry. For one numerical and three experimental nonlinear chemical reactions, we demonstrated that the proposed method could separate components with a degree of purity that was sufficient to yield a small number of incorrect annotations. It is evident from the results presented in Table 2 that proposed method is characterized by: (*i*) high computational complexity, and (*ii*) low separation quality. The first issue can be addressed by implementing proposed method in either C language or on graphical processing units (GPU). As demonstrated in [53], GPU-based implementation can bring 100-fold increase of speed of computation. Further decrease in computational complexity can be obtained by using subspace dimension in EKM induced RKHS (12) and (13) to be significantly less than the number of observations T. Improvement of the separation quality can be possibly achieved by executing sparseness constrained NMF in more than one RKHS and, afterwards, combining results obtained from multiple RKHSs.

REFERENCES

1. Gebregiworgis T, Powers R. Application of NMR Metabolomics to Search for Human Disease Biomarkers. *Comb. Chem. & High Throug. Screen*. 2012; **15**: 595-610.

2. Vinayavekin, N, Homan, E A, Saghatelian, A. Exploring disease through metabolomics. *ACS Chem. Biol.* 2010; **5** (1): 91-103.

3. Gowda G A, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D . Metabolomics-based methods for early disease diagnostics. *Expert Rev. Mol. Diag.* 2008; **8** (5): 617-633.

4. Bain J R, Stevens R D, Wenner B R, Ilkayeva O, Muoio D M, Newgard C B. Metabolomics Applied to Diabetes Research - Moving From Information to Knowledge. *Diabetes.* 2009; **58**: 2429-2443.

5. Wang T J, *et al.* 2-Aminoadipic acid is a biomarker for diabetes risk. *The J. Clinical Invest*. 2013; **123** (10): 4309-4317.

6. 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*. 2102; **84**: 6429−6437.

7. 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.

8. 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.

9. Seavey B R, Farr E A, Westler W M, Markley J L. A relational database for sequence- specific protein NMR data. *J. Biomol. NMR*. 1991; **1**: 217-236.

10. The NIST 11 Mass Spectral Library web site: http://www.sisweb.com/software/ms/nist.htm.

11. Drucker E, Krapfenbauer K. Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. *The EPMA Journal* 2013, **4**:7.

12. Nyamundanda G, Gormley I C, Fan Y, Gallagher W M, Brennan L. MetSizeR: selecting the optimal sample size for metabolomic studies using an analysis based approach. *BMC Bioinformatics*. 2013, **14**: 338.

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

14. 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.

15. 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.

16. 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.

17. Kopriva I, Jerić I, Brkljačić L. Nonlinear mixture-wise expansion approach to underdetermined blind separation of nonnegative dependent sources. *J. Chemometrics* 2013; **27**: 189-197.

18. Kopriva I, Jerić I, Filipović M, Brkljačić L. Empirical Kernel Map Approach to Nonlinear Underdetermined Blind Separation of Sparse Nonnegative Dependent Sources: Pure Components Extraction from Nonlinear Mixtures Mass Spectra. *J. Chemometrics* 2014; **28**: 704-715.

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

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

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

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

23. Walleczek J(ed). *Self-organized biological dynamics and non-linear control*. Cambridge University Press: Cambridge, UK. 2000.

24. Franceschi P, Vrhovsek U, Mattivi F, Wehrens R. Metabolic Biomarker Identification with Few Samples, in: *Chemometrics in Practical Applications*, Varmuza K. Ed., InTech, 2012.

25. Davies M E, James C I. Source separation using single channel ICA. *Sig. Proc.*, 2007; **87**: 1819-1832.

26. Ouyang Y C, Chen H M, Chai J W, Chen C C C, Poon S K, Yang C W, Lee S K, Chang C I. Band Expansion-Based Over-Complete Independent Component Analysis for Multispectral Processing of Magnetic Resonance Image. *IEEE Trans. Biomed. Eng.* 2008; **55** (6): 1666-1677.

27. Mijović B, Vos M De, Gligorijević I, Taelman J, Van Huffel S. Source Separation from Single-Channel Recordings by Combining Empirical Mode Decomposition and Independent Component Analysis. *IEEE Trans. on Biomed. Eng.* 2010; **57**(9): 2188-2196.

28. Lin J, Zhang A. Fault feature separation using wavelet-ICA filter. *NDT&E International.* 2005; **38**: 421-427.

29. He Q, Su S, Du R. Separating mixed multi-component signal with an application in mechanical watch movement. *Dig. Sig. Proc.* 2008; **18**: 1013-1028.

30. Gunawan D, Sen D. Iterative Phase Estimation for the Synthesis of Separated Sources From Single-Channel Mixtures. *IEEE Sig. Proc*. *Let*. 2010; **17**(5): 421-424.

31. Parry R M, Essa I. Phase-Aware Non-negative Spectrogram Factorization. *Lect. Notes Comp. Sci.* 2007; **4666**: 536-543.

32. Gao B, Woo W L, Ling B W K. Machine Learning Source Separation Using Maximum A Posteriori Nonnegative Matrix Factorization. *IEEE Trans. on Cybernetics*. 2014; **44** (7): 1169-1179.

33. Jang G J, Lee T W. A Maximum Likelihood Approach to Single-channel Source Separation. *J. Machine Learn. Res.* 2003; **4**: 1365-1392.

34. Roweis S T. "One microphone source separation," in: *Advance in Neural Information Processing Systems 13*. The MIT Press: 2000, pp. 793-799.

35. Grais E M, Erdogan H. Source separation using regularized NMF with MMSE estimates under GMM priors with online learning from uncertainties. *Dig. Sig. Proc.* 2014; **29**: 20-34.

36. Candès E J, Li X, Ma Y, Wright H. Robust Principal Component Analysis? *J. ACM* 2011; **58**: Article 11 (37 pages).

37. Chandrasekaran V, Sanghavi S, Paririlo P A,   Wilsky A S. Rank-Sparsity Incoherence for Matrix Decomposition. *SIAM J. Opt.* 2011; **21**: 572-596.

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

39. Fang H T, Huang D S. Wavelet de-noising by means of trimmed thresholding. in: *Proc. of the 5th World Congress on Intelligent Control and Automation*, June 15-19, 2004, Hangzhou, P. R. China, pp. 1621-1624.

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

41. Huang N E, Wu M L, Long S R, Shen S S, Qu W D, Gloersen P, Fan K L. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Royal Soc. London*. 1998; **454A** (1971): 903–993.

42. 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 Net. and Learn. Sys.* 2012; **23** (10): 1601-1610.

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

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

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

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

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

48. Micchelli C A, Xu Y, Zhang H. Universal Kernels. *J. Mach. Learn. Res.* 2006; **7**: 2651-2667.

49. Sriperumbudur B K, Fukumizu K, Lanckriet G R G. Universality, Characteristic Kernels and RKHS Embedding of Measures. *J. Mach. Learn. Res.* 2011; **12**: 2389-2410.

49. Girosi F, Jones M, Poggio T. Regularization theory and neural network architectures. *Neural Comput.* 1995; **7**: 219-269.

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

51. Bruckstein A M, Donoho D L, Elad M. From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Review* 2009; **51** (1): 34-81.

52. The Nicolas Gillis Website. https://sites.google.com/site/nicolasgillis/code [20 January 2015].

53. Preibisch S, Amat F, Stamataki E, Sarov M, Singer R H, Myers E, Tomancak P. Efficient Bayesian-based multiview deconvolution. *Nat. Methods* 2014; **11**: 645-648.

Figure Captions

Figure 1 (color online). Numerical study of the EFM-EKM-NMU method. The probability of state zero is equal to 0.8. Top row: number of wrong annotations vs. Monte Carlo run index. Left: RPCA- (triangles), HT- (squares), ST- (stars), TT- (circles) and PTs-EFM-EKM-NMU (diamonds) methods. Right: PTs-EFM-EKM-NMU (squares), PPTs-EKM-NMU (stars) and NMU (circles). Bottom row: normalized correlation coefficient vs. Monte Carlo run index between the true and extracted sources according to the same legend as for the first row. The NMU and PTs-EKM-NMU methods were applied on three mixtures, while the PTs-EFM-EKM-NMU method was only applied on one mixture.
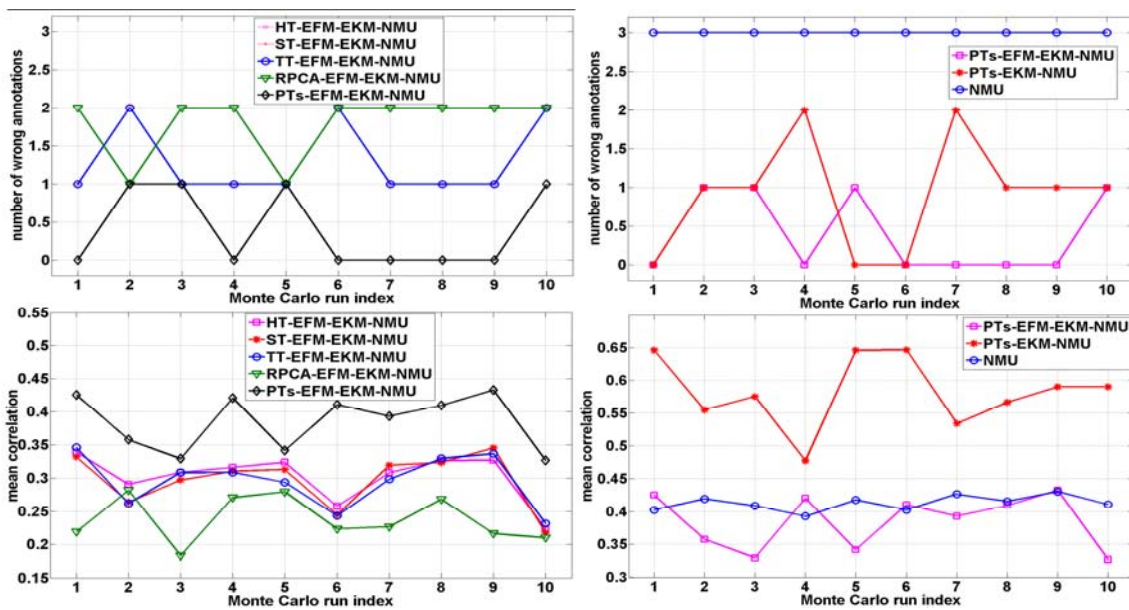
Fig. 1, Kopriva, Jerić & Brkljačić

Figure 2. Schematic presentation of the chemical reactions and formation of new components as the reaction proceeds.
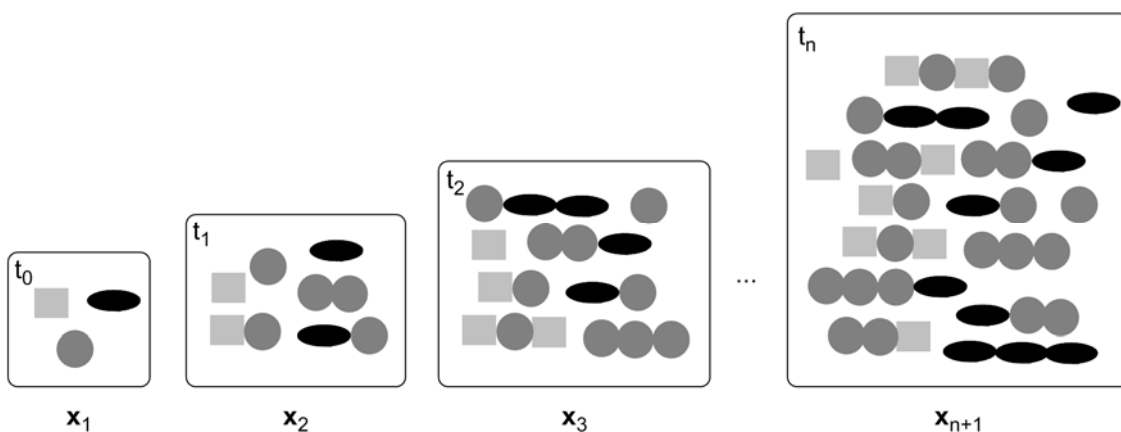

Fig. 2, Kopriva, Jerić & Brkljačić

Figure 3. Two top rows: mass spectra of the pure components s_1, s_4, s_8 and s_{11}. Two bottom rows: mass spectra of the pure components s_1, s_4, s_8 and s_{11} estimated by the PTs-EFM-EKM-NMU algorithm. Information on the value of the highest normalized correlation coefficient and associated error reduction method (RPCA, HT, ST and TT) is also displayed.
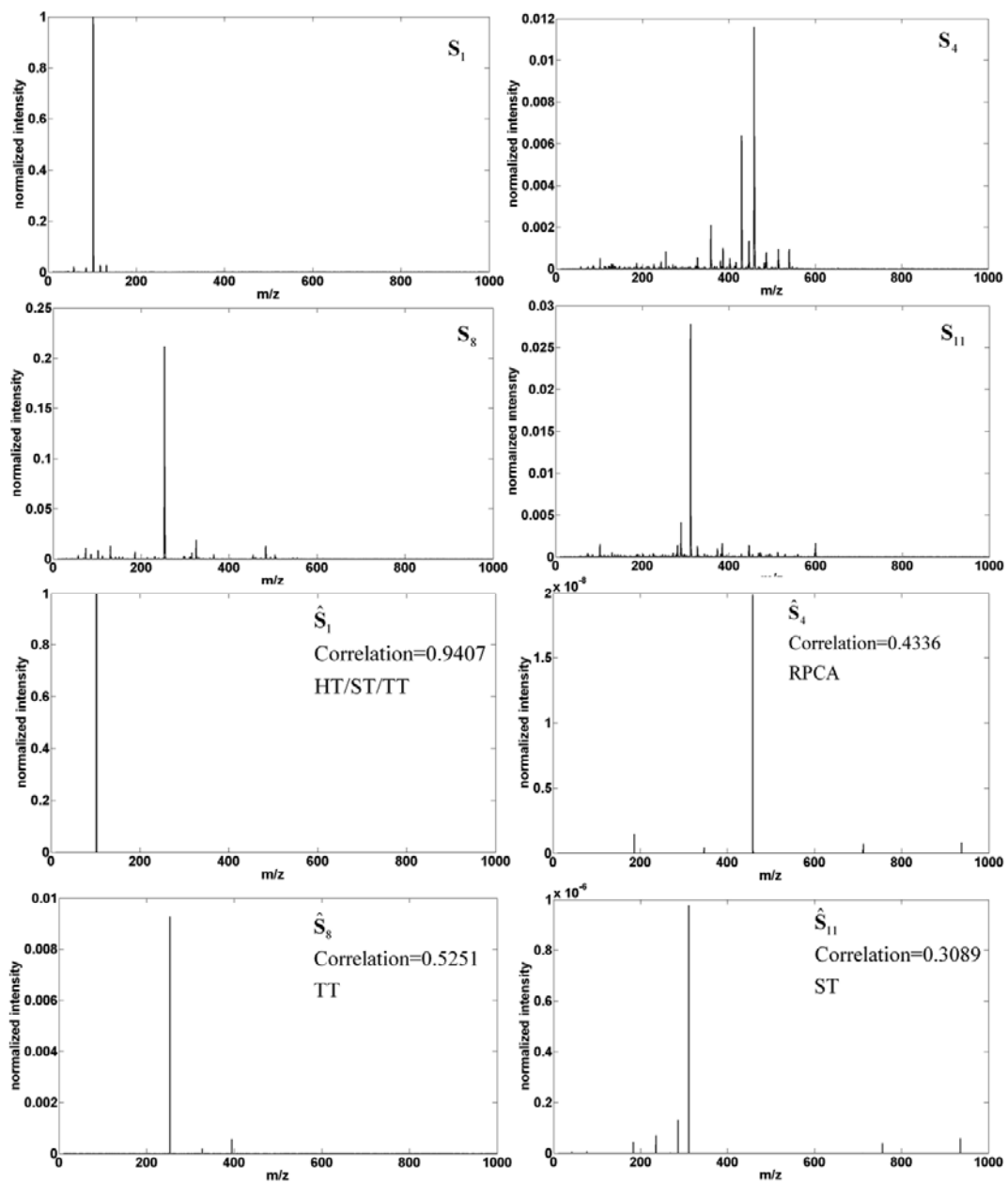
Table Captions

Algorithm 1. The PTs-EFM-EKM-NMU algorithm.

Required:

$\mathbf{x} \in R_{0+}^{1 \times T}$. If A1) is not satisfied, perform scaling $\mathbf{x} \to \mathbf{x}/\max\{x_t\}_{t=1}^{T}$.

1. Perform *HOT* suppression (5): $\mathbf{x} \overset{RPCA}{\longmapsto} \mathbf{a}$, $\mathbf{x} \overset{HT}{\longmapsto} \mathbf{b}$, $\mathbf{x} \overset{ST}{\longmapsto} \mathbf{c}$, $\mathbf{x} \overset{TT}{\longmapsto} \mathbf{d}$ on $\mathbf{x}$ in (2)/(4). For RPCA, set $\lambda \approx 1/\sqrt{T}$. For HT, ST and TT, set $\tau \approx 10^{-5}$.

2. Perform EFM mapping (7) $\mathbf{a} \to \phi_D(\mathbf{a})$, $\mathbf{b} \to \phi_D(\mathbf{b})$, $\mathbf{c} \to \phi_D(\mathbf{c})$, and $\mathbf{d} \to \phi_D(\mathbf{d})$ with $D \approx 8$.

3. Perform *HOT* suppression (11): $\phi_D(\mathbf{a}) \overset{RPCA}{\longmapsto} \phi_D(\mathbf{a}), \phi_D(\mathbf{b}) \overset{HT}{\longmapsto} \phi_D(\mathbf{b})$, $\phi_D(\mathbf{c}) \overset{ST}{\longmapsto} \phi_D(\mathbf{c})$, $\phi_D(\mathbf{d}) \overset{TT}{\longmapsto} \phi_D(\mathbf{d})$ with λ and τ set as in step **1**.

4. Perform EKM mappings $\phi_D(\mathbf{a}) \to \Psi(\phi_D(\mathbf{a}))$, $\phi_D(\mathbf{b}) \to \Psi(\phi_D(\mathbf{b}))$, $\phi_D(\mathbf{c}) \to \Psi(\phi_D(\mathbf{c}))$ and $\phi_D(\mathbf{d}) \to \Psi(\phi_D(\mathbf{d}))$ according to (13). Use a Gaussian kernel with $\sigma^2 = 1$.

5. Perform *HOT* suppression (15):
   $\Psi(\phi_D(\mathbf{a})) \overset{RPCA}{\longmapsto} \Psi(\phi_D(\mathbf{a}))$, $\Psi(\phi_D(\mathbf{b})) \overset{HT}{\longmapsto} \Psi(\phi_D(\mathbf{b}))$,
   $\Psi(\phi_D(\mathbf{c})) \overset{ST}{\longmapsto} \Psi(\phi_D(\mathbf{c}))$ and $\Psi(\phi_D(\mathbf{d})) \overset{TT}{\longmapsto} \Psi(\phi_D(\mathbf{d}))$ with λ and τ set as in step **1**.

6. Perform sparseness constrained factorization, preferably with the NMU algorithm, of matrices $\Psi(\phi_D(\mathbf{a}))$, $\Psi(\phi_D(\mathbf{b}))$, $\Psi(\phi_D(\mathbf{c}))$ and $\Psi(\phi_D(\mathbf{d}))$ to obtain the separated components $\overline{\mathbf{S}}^{\mathbf{a}}$, $\overline{\mathbf{S}}^{\mathbf{b}}$, $\overline{\mathbf{S}}^{\mathbf{c}}$ and $\overline{\mathbf{S}}^{\mathbf{d}}$.

7. Annotate, using the pure components from the library $\{\mathbf{s}_m\}_{m=1}^{M}$, those separated components $\overline{\mathbf{S}}^{\mathbf{a}}$, $\overline{\mathbf{S}}^{\mathbf{b}}$, $\overline{\mathbf{S}}^{\mathbf{c}}$ and $\overline{\mathbf{S}}^{\mathbf{d}}$ with the highest normalized correlation coefficient: $\left\{\left(\hat{\mathbf{s}}_m^{\mathbf{a}}, c_m^{\mathbf{a}}\right)\right\}_{m=1}^{M}$, $\left\{\left(\hat{\mathbf{s}}_m^{\mathbf{b}}, c_m^{\mathbf{b}}\right)\right\}_{m=1}^{M}$, $\left\{\left(\hat{\mathbf{s}}_m^{\mathbf{c}}, c_m^{\mathbf{c}}\right)\right\}_{m=1}^{M}$ and $\left\{\left(\hat{\mathbf{s}}_m^{\mathbf{d}}, c_m^{\mathbf{d}}\right)\right\}_{m=1}^{M}$.

8. Obtain a final annotation $\left(\hat{\mathbf{s}}_m^{\mathbf{f}}, c_m^{\mathbf{f}}\right)$ according to (21): $\mathbf{f} = \arg\max_{\{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}\}} \{c_m^{\mathbf{a}}, c_m^{\mathbf{b}}, c_m^{\mathbf{c}}, c_m^{\mathbf{d}}\} \quad \forall m = 1,...,M$.

Table 1. Comparative performance analysis of the NMU, PTs-EKM-NMU, and PTs-EFM-EKM-NMU algorithms. The probability of the zero state was $\rho_m=0.8$. The two metrics used in the comparative performance analysis were the "mean" value of the correlation coefficient calculated for the correctly annotated components, but averaged over all components, and the number of pure components that were incorrectly assigned (annotated) due to poor separation. Two metrics were calculated with respect to the predefined labeling of the pure components stored in the library. Incorrect assignment implies that, based on the maximal correlation criterion, two or more pure components are assigned to the same separated component. The mean values and variance are reported and estimated over 10 Monte Carlo runs. The NMU and EKM-NMU algorithms were applied on three mixtures, while the EFM-EKM-NMU algorithm was only applied on one mixture. The best result in each metric is given in bold.

	NMU [40] mixtures x_1, x_2, x_3	PTs-EKM-NMU [18] mixtures x_1, x_2, x_3	PTs-EFM-EKM-NMU mixture x_1
mean correlation	0.41±0.01	**0.58±0.05**	0.38±0.04
number of incorrect annotations	3.0±0.0	0.9±0.74	**0.6±0.52**

Table 2. Comparative performance analysis of the NMU, PTs-EKM-NMU and PTs-EFM-EKM-NMU (D=8) algorithms for 3 nonlinear reactions. The two metrics used in comparative performance analysis were the mean value of the correlation coefficient over all associated components and the numbers of the pure components that were incorrectly assigned (due to poor separation). The best result in each metric is given in bold. The first metric was only calculated for correctly assigned components.

	NMU [40]	PTs-EKM-NMU [18]	PTs-EFM-EKM-NMU
First nonlinear reaction: 9 mixtures and 25 pure components. PTs-EFM-EKM-NMU was applied to mixture $\mathbf{x}_9$. NMU and PTs-EKM-NMU were applied to mixtures $\mathbf{x}_1$ to $\mathbf{x}_9$.			
mean correlation	0.342	**0.702**	0.467
incorrect annotations	15	**0**	**0**
CPU time	$\approx$1.3 s	$\approx$4$\times$78 h	$\approx$4$\times$45 h
Second nonlinear reaction: 12 mixtures and 19 pure components. PTs-EFM-EKM-NMU was applied to mixture $\mathbf{x}_{12}$. NMU and PTs-EKM-NMU were applied to mixtures $\mathbf{x}_1$ to $\mathbf{x}_{12}$.			
mean correlation	0.460	**0.702**	0.439
incorrect annotations	15	**0**	1
CPU time	$\approx$1.3 s	$\approx$4$\times$78 h	$\approx$4$\times$78 h
Third nonlinear reaction: 12 mixtures and 28 pure components. PTs-EFM-EKM-NMU was applied to mixture $\mathbf{x}_{12}$. NMU and PTs-EKM-NMU were applied to mixtures $\mathbf{x}_1$ to $\mathbf{x}_{12}$.			
mean correlation	0.614	**0.656**	0.368
incorrect annotations	10	**0**	2
CPU time	$\approx$1.3 s	$\approx$4$\times$72.4 h	$\approx$4$\times$78 h

Explicit-Implicit Mapping Approach to Nonlinear Blind Separation of Sparse Nonnegative Sources from a Single-Mixture: Pure Component Extraction from Nonlinear Mixture Mass Spectra

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 is proposed for the nonlinear, blind separation of nonnegative sparse dependent sources from a single mixture. This method combines explicit feature maps-based and implicit kernel-based mappings to convert a single-mixture nonlinear blind source separation problem into an approximately linear pseudo multiple-mixture problem in mapping induced reproducible kernel Hilbert space (RKHS). Therein, the sparseness constrained nonnegative matrix factorization is used to separate the sources. The goal application of the proposed method is mass spectrometry-based non-targeted metabolic profiling.