Received: 21 January 2014, Revised: 25 April 2014, Accepted: 2 May 2014, Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/cem.2635

Empirical kernel map approach to nonlinear underdetermined blind separation of sparse nonnegative dependent sources: pure component extraction from nonlinear mixture mass spectra

Ivica Kopriva[a]*, Ivanka Jerić[b], Marko Filipović[a] and Lidija Brkljačić[b]

Nonlinear underdetermined blind separation of nonnegative dependent sources consists in decomposing a set of observed nonlinearly mixed signals into a greater number of original nonnegative and dependent component (source) signals. This hard problem is practically relevant for contemporary metabolic profiling of biological samples, where sources (a.k.a. pure components or analytes) are aimed to be extracted from mass spectra of nonlinear multicomponent mixtures. This paper presents a method for nonlinear underdetermined blind separation of nonnegative dependent sources that comply with a sparse probabilistic model, that is, sources are constrained to be sparse in support and amplitude. This model is validated on experimental pure component mass spectra. Under a sparse prior, a nonlinear problem is converted into an equivalent linear one comprised of original sources and their higher-order, mostly second-order, monomials. The influence of these monomials, which stand for error terms, is reduced by preprocessing a matrix of mixtures by means of robust principal component analysis and hard, soft and trimmed thresholding. Preprocessed data matrices are mapped in high-dimensional reproducible kernel Hilbert space (RKHS) of functions by means of an empirical kernel map. Sparseness-constrained nonnegative matrix factorizations in RKHS yield sets of separated components. They are assigned to pure components from the library using a maximal correlation criterion. The methodology is exemplified on demanding numerical and experimental examples related respectively to extraction of eight dependent components from three nonlinear mixtures and to extraction of 25 dependent analytes from nine nonlinear mixture mass spectra recorded in nonlinear chemical reaction of peptide synthesis. Copyright © 2014 John Wiley & Sons, Ltd.

Additional supporting information may be found in the online version of this article at the publisher's web site.

Keywords: nonlinear underdetermined blind source separation; robust principal component analysis; thresholding; empirical kernel maps; nonnegative matrix factorization

1. INTRODUCTION

Identification of pure components present in mixtures is a traditional problem in spectroscopy (nuclear magnetic resonance, infrared and Raman) and mass spectrometry [1–4]. Identification proceeds often by matching separated component spectra with a library of reference compounds [5–7], whereas the degree of correlation depends on how well pure components are separated from each other. Thereby, of interest are blind source separation (BSS) methods that use only the matrix with recorded mixture spectra as input information [8–11]. In majority of scenarios, separation of pure components is performed by assuming that mixture spectra are linear combinations of pure components [1–4]. While a linear mixture model is adequate for many scenarios, a nonlinear model offers a more accurate description of processes and interactions occurring in biological systems. Living organisms are the best examples of complex nonlinear systems that function far from equilibrium. Internal and external stimuli (disease, drug treatment and environmental changes) cause perturbations in the system as a result of highly synchronized molecular interactions [12]. As opposed to many BSS methods developed for linear problems, the number of methods that address nonlinear BSS problem is considerably smaller; see for example chapter 14 in [11]. This number is reduced further when a related nonlinear BSS problem is underdetermined, that is, when the number of pure components is greater than the number of mixtures. That is why metabolic profiling, which aims to identify and quantify small-molecule analytes (a.k.a. pure components or sources) present in biological samples (typically urine, serum or biological tissue extract), is seen as one of the most challenging tasks in systems biology [13]. Therefore, the underdetermined problem is of practical relevance.

The aim of the paper is to present a method for blind separation of pure components from a smaller number of multicomponent

* *Correspondence to: Ivica Kopriva, Division of Laser and Atomic Research and Development, Ruđer Bošković Institute, Bijenička cesta 54, HR-10000, Zagreb, Croatia. E-mail: ikopriva@irb.hr*

a *I. Kopriva, M. Filipović*
Division of Laser and Atomic Research and Development, Ruđer Bošković Institute, Bijenička cesta 54, HR-10000, Zagreb, Croatia

b *I. Jerić, L. Brkljačić*
Division of Organic Chemistry and Biochemistry, Ruđer Bošković Institute, Bijenička cesta 54, HR-10000, Zagreb, Croatia

nonlinear mixtures mass spectra. Therefore, it is assumed that components are nonnegative and sparse. To this end, we address an underdetermined nonlinear nonnegative BSS (uNNBSS) problem with sparse and dependent sources. As has been discussed at great length in [4], even a linear underdetermined BSS problem comprised of dependent sources is challenging with only few algorithms addressing it. There is basically no method proposed for a uNNBSS problem. Herein, we propose a method for a uNNBSS problem that can be considered as a generalization of the method developed in [4] for underdetermined linear nonnegative BSS (uLNBSS) problem comprised of dependent sources. The proposed method constrains sources to be nonnegative and to comply with a sparse probabilistic model [14,15], that is, sources are assumed to be sparse in support and amplitude. The model is validated on experimental mass spectrometry data and is therefore practically relevant (Section 3.2). This represents the first original contribution of the paper. Under this sparse prior, a nonlinear problem is approximated by a linear one comprised of original sources and their second-order monomials. This follows from analytical derivations based on Taylor expansion of a nonlinear mixture model (i.e., the vector function with vector argument) up to an arbitrary order. Analytical derivation of the Taylor expansion based on the Tucker model of tensor derivatives represents, arguably, the second original contribution of the paper. The key contribution of this paper is the reduction of the influence of higher-order monomials that stand for error terms. This is achieved by preprocessing a matrix of mixtures by means of robust principal component analysis (RPCA) [16,17] and hard (HT), soft (ST) [18] and trimmed thresholding (TT) [19]. Preprocessed data matrices are mapped observation-wise in high-dimensional RKHS by means of an empirical kernel map (EKM). Thus, one uNNBSS problem is converted into four nonnegative BSS problems in RKHS with the same number of observations but an increased number of mixtures. Sparseness-constrained nonnegative matrix factorization (NMF) is performed in RKHS to solve these nonnegative BSS problems. Thereby, components separated by NMF are assigned to pure components from the library using a maximal correlation criterion.

The rest of the paper is organized as follows. Section 2 gives an overview of nonlinear BSS methods and presents the theory upon which the proposed uNNBSS is built. Section 3 describes experiments performed on computational and experimental data. Section 4 presents and discusses the results of comparative performance analysis between the proposed uNNBSS and some state-of-the-art NMF algorithms. Concluding remarks are given in Section 5.

2. THEORY AND ALGORITHM

The aimed application of the proposed uNNBSS method is the extraction of analytes from multicomponent nonlinear mixtures of mass spectra. As emphasized in [4], mass spectrometry is chosen because of its increasing importance in clinical chemistry, safety and quality control as well as biomarker discovery and validation. As in [4,5], we assume that a library of reference mass spectra is available to evaluate the quality of components extracted by the proposed method.[1] For an example, the National Institute of Science and Technology and Wiley-Interscience universal spectral library [7] contains more than 800,000 mass spectra (corresponding to more than 680,000 compounds). As opposed to [4], where a linear mixture model is assumed, a nonlinear model is assumed herein. Thereby, a linear model is implicitly included as a special case.

From the viewpoint of a uNNBSS problem with dependent sources, existing algorithms for a nonlinear BSS problem have at least one of the following deficiencies: (i) they assume that the number of mixtures is equal to or greater than the unknown number of sources [21–29]; (ii) they do not take into account the nonnegativity constraint that is present when sources are pure component mass spectra [21–32]; and (iii) they assume that source signals are statistically independent [22–24,27–32] and, sometimes, individually correlated [28,30,31]. None of these assumptions holds true for the uNNBSS problem considered herein. The algorithm described in [33] is developed for a uNNBSS problem composed of nonnegative sources. However, the assumption made by the algorithm is that a set of observation indexes exists such that each source is present alone in at least one of these observations. This assumption seems too strong for the considered uNNBSS problem where the mass spectra of structurally similar pure components are expected to overlap. This is especially the case if the resolution of the mass spectrometer is low. Algorithms [34–36] execute nonlinear nonnegative BSS by means of NMF in a reproducible kernel Hilbert space (RKHS). Nevertheless, unlike the uNNBSS method proposed herein, they do not do the following: (i) enforce the sparseness constraint that is shown herein to be an enabling condition for solving otherwise intractable uNNBSS problems; (ii) reduce the influence of higher-order monomials of the original sources (error terms) induced by a nonlinear mixing process and that is shown herein to be crucial for obtaining a reasonably accurate solution of the uNNBSS problem. As seen in Section 2.2, the uNNBSS problem is converted into an equivalent uLNBSS problem with a large number of sources: the original ones and their higher-order monomials induced by a nonlinear mixing process. Without activation of a sparse probabilistic prior, an equivalent uLNBSS problem is intractable.

As seen in Sections 3.1 and 3.2, the proposed methodology significantly improves accuracy relative to the case when the NMF algorithm is performed on an empirically kernel-mapped matrix of mixture data without suppression of higher-order monomials. It has already been discussed in [4,37] that the performance of many NMF algorithms depends on optimal usage of parameters required to be known *a priori*, such as the balance parameter that regulates the influence of the sparseness constraint [38] or the number of overlapping components that exist in mixtures [39]. Often, these parameters are difficult to select optimally in practice. That is why the nonnegative matrix underapproximation (NMU) algorithm [40] is proposed to solve nonnegative BSS problems in RKHS. That is, it does not require *a priori* information from the user. Thus, we propose herein to combine RPCA, HT, ST and TT preprocessing transforms and EKM-based nonlinear mapping with the NMU algorithm in mapping-induced high-dimensional RKHS, hence the PTs-EKM-NMU algorithm. The PTs-EKM-NMU is exemplified on numerical and experimental problems. Nevertheless, proposed preprocessing transforms can also be used in combination with other sparseness-constrained NMF (sNMF) algorithms. Provided that the number of overlapping components can be reasonably accurately inferred, an NMF algorithm with ℓ_0-constraints (NMF_L0) [39] is a good choice.

[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 used in hyperspectral image analysis where identification of minerals proceeds by comparison of estimated endmembers with spectral profiles stored in the library; see for an example the ASTER spectral library in [20].

2.1. Underdetermined nonlinear nonnegative blind source separation with dependent sources

The uNNBSS problem with dependent sources is described as

$$\mathbf{x}_t = \mathbf{f}(\mathbf{s}_t) \quad t = 1, \ldots, T \tag{1}$$

where $\mathbf{x}_t \in R_{0+}^{N\times 1}$ stands for the nonnegative measurement vector comprised of intensities acquired at some of T mass-to-charge (m/z) channels and $\mathbf{s}_t \in R_{0+}^{M\times 1}$ stands for an unknown vector comprised of intensities of M nonnegative sources. $\mathbf{f} : R_{0+}^{M} \rightarrow R_{0+}^{N}$ is an unknown multivariate mapping such that $\mathbf{f}(\mathbf{s}_t) = [f_1(\mathbf{s}_t) \ldots f_N(\mathbf{s}_t)]^T$ and $\left\{f_n : R_{0+}^{M} \rightarrow R_{0+}\right\}_{n=1}^{N}$. Problem (1) can be casted in the matrix framework

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

such that $\mathbf{X} \in R_{0+}^{N\times T}$ and $\mathbf{S} \in R_{0+}^{M\times T}$, where $\{\mathbf{x}_t\}_{t=1}^{T}$ and $\{\mathbf{s}_t\}_{t=1}^{T}$ are column vectors of matrices $\mathbf{X}$ and $\mathbf{S}$, respectively, and $\mathbf{f}(\mathbf{S})$ implies that nonlinear mapping is performed column-wise such as in (1). It is further assumed that $\left\{\|\mathbf{s}_t\|_0 \leq L\right\}_{t=1}^{T}$, where $\|\mathbf{s}_t\|_0$ stands for the ℓ_0 quasi-norm that counts the number of nonzero coefficients of $\mathbf{s}_t$ and $L = \max_{t=1,\ldots,T} \|\mathbf{s}_t\|_0$. Evidently, it applies that $L \leq M$, where L denotes the maximal number of sources that can be present at any coordinate t. The uNNBSS problem implies that component mass spectra, $\left\{\mathbf{s}_m \in R_{0+}^{1\times T}\right\}_{m=1}^{M}$, ought to be inferred from mixture data matrix $\mathbf{X}$ only. In this paper, the following assumptions are made on the nonlinear mixture model (1)/(2):

A1) $0 \leq x_{nt} \leq 1$ $\forall$ $n = 1, \ldots, N$ and $\forall$ $t = 1, \ldots, T$,
A2) $0 \leq s_{mt} \leq 1$ $\forall$ $m = 1, \ldots, M$ and $\forall$ $t = 1, \ldots, T$,
A3) $M > N$,
A4) Amplitude s_{mt} obeys exponential distribution on (0, 1] interval and discrete distribution at zero, see also Eqn (3),
A5) Components of the vector-valued function $\mathbf{f}(\mathbf{s}) : f_n(\mathbf{s}) : R_{0+}^{M\times 1} \mapsto R_{0+}, \forall n = 1, \ldots, N$ are differentiable up to an unknown order K and
A6) $M \ll T$.

To avoid confusion between column and row vectors, they will be indexed by lowercase letters that correspond with uppercase letters related to dimensions of the corresponding matrix. As an 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. In order to be useful, the solution of the uNNBSS problem is expected to be essentially unique; that is, the estimated matrix of pure components (sources) $\hat{\mathbf{S}}$ and the true matrix of pure components $\mathbf{S}$ have to be related through $\hat{\mathbf{S}} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{S}$, where $\mathbf{P}$ and $\boldsymbol{\Lambda}$ stand respectively for $M \times M$ permutation and diagonal matrices. As discussed at great length in [4], even a linear underdetermined BSS problem requires constraints to be imposed on sources in order to ensure an essentially unique solution. A nonlinear BSS problem is more difficult. Herein, we assume that pure components $\{\mathbf{s}_m\}_{m=1}^{M}$ comply with the sparse probabilistic model imposed by A4. It implies that each component will be zero at a great part of its support (number of m/z channels T) and that nonzero intensity will be distributed according to exponential distribution with a small expected value. These two constraints are expected to ensure that, in probability, compared with N and M, the maximal number of analytes L present at the particular m/z coordinate is small enough. However, N stands for the number of biological samples available, and it is expected to be small. Thus, it can virtually be impossible to satisfy the aforementioned requirement. That is why, as in [4], in order to increase the number of measurements (samples), the original uNNBSS problem (1) has to be mapped into RKHS by using EKM. Before that, we need to approximate the uNNBSS problem (1)/(2) by an equivalent uLNBSS problem.

2.2. Sparse probabilistic model of source signals

The Taylor expansion of the nonlinear model (1) up to an arbitrary order K is derived in the Supporting Information. It is shown that the uNNBSS problem (1) can be represented by an equivalent uLNBSS problem, Eqn (7) in the Supporting Information, comprised of M original sources and $\sum_{k=2}^{K} M^{(k)}$ higher-order monomials, where $M^{(k)} = \begin{pmatrix} M+k-1 \\ k \end{pmatrix}$. Thus, without further constraints, the uNNBSS problem (1) is computationally intractable. That is why, according to A4, we assume that sources $\mathbf{s}$ comply with the sparse probabilistic model comprised of mixed state distribution [4,14,15]:

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

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 $f(s_{mt})$. We have chosen the exponential distribution $f(s_{mt}) = (1/\mu_m)\exp(-s_{mt}/\mu_m)$ to model the sparse distribution of the nonzero states, in which case, the most probable outcomes are equal to μ_m. It has been verified in [4] that model (3) describes well the mass spectra of the pure components. Herein, by using the mass spectra of 25 pure components, we have estimated $\hat{\rho}_m \in [0.27, 0.74]$ and $\hat{\mu}_m \in [0.0012, 0.0014]$; see Section 3.2 and Figure 4 for more details.[2] Under the exponential prior, the probability that amplitude $s_{mt} \in [\varepsilon, \mu_m]$, for $0 < \varepsilon \ll 1$, is 0.632. Thus, in 36.8% of the cases, a random realization of s_{mt} will have amplitudes greater than the most probable value μ_m. For a given μ_m and given probability $p(\varepsilon < s_{mt} \leq s)$, the value of s is obtained as: $s \approx -\mu_m \ln(1-p)$. Thus, for $p = 0.99$ and $\mu_m = 1.5 \times 10^{-3}$, it follows that $s = 7 \times 10^{-3}$. Hence, we may approximate the equivalent uLNBSS model, Eqn (7) in Supporting Information, by retaining the second-order terms only:

$$\mathbf{X} = \mathbf{G}_{(1)}^{1}\mathbf{S} + \frac{1}{2}\mathbf{G}_{(1)}^{2}\begin{bmatrix} \mathbf{s}_1^2 \\ \ldots \\ \mathbf{s}_M^2 \\ \ldots. \\ \{\mathbf{s}_{m_1}\mathbf{s}_{m_2}\}_{m_1,m_2=1}^{M} \end{bmatrix} + HOT \tag{4}$$

where $\mathbf{G}_{(1)}^{1}$ and $\mathbf{G}_{(1)}^{2}$ stand for unfolded versions of the tensor of first-order and second-order derivatives, respectively, and HOT

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

stands for higher-order terms. The contribution of third-order terms in (4) is of the order $(7\times10^{-3})^3 = 3.43\times10^{-7}$. In order to reduce *HOT*, entry-wise thresholding of **X** can be performed. By neglecting fourth-order and higher-order terms, we have empirically arrived at the threshold value of $\tau\in[10^{-6}, 10^{-4}]$.[3]

2.3. Suppression of higher-order (error) terms

The mass spectra of 25 pure components recorded in the nonlinear chemical reaction of peptide bond formation (Section 3.2 and Figures 3 and S-4 in Supporting Information) illustrate the diversity of morphologies. Some have few very dominant (large) peaks (see spectra of pure components 1, 2, 8, 13, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25), and some have intensities distributed on several *m/z* values, whereas intensities can be small (see spectra of pure components 3, 4, 5, 6, 7, 9, 10, 11, 12, 14 and 15). It is thus hard to propose one preprocessing (thresholding) transform for suppression of higher-order terms induced by nonlinear mixing process. We, therefore, propose a combination of methods for this purpose.

2.3.1. *Robust principal component analysis*

Robust principal component analysis has been proposed in [16,17] to decompose data matrix **X** into sum of two matrices: $\mathbf{X} = \mathbf{A} + \mathbf{E}$. Provided that **A** is a low-rank matrix and **E** is a sparse matrix, decomposition is unique, and it is obtained as a solution of the optimization problem:

$$\text{minimize } \|\mathbf{A}\|_* + \lambda\|\mathbf{E}\|_1 \text{ subject to } \mathbf{A} + \mathbf{E} = \mathbf{X} \tag{5}$$

Thereby, $\|\mathbf{A}\|_* = \sum_{i=1}^{I\le N}\sigma_i$ denotes the nuclear norm (sum of singular values) and $I\le N$ is a rank of matrix **A**, $\|\mathbf{E}\|_1 = \sum_{n=1}^{N}\sum_{t=1}^{T} e_{nt}$ denotes the ℓ_1-norm of **E** and $\lambda\approx1/\sqrt{T}$ is a regularization constant. In terms of the equivalent uLNBSS problem (4), **A** is associated with first-order and second-order terms, and **E** is associated with *HOT*. **A** is actually represented by a linear mixture model composed of $2M + M(M-1)/2$ sources and N mixtures. Because both N and $2M + M(M-1)/2$ are small compared with T, the rank of **A** equals $\min(N, 2M + M(M-1)/2) = N$. Thus, it is low. **E** is comprised of monomials (products of the original source components) of the order 3 or higher. Because by assumption, A4, source components are sparse in support and amplitude, their three-order and higher-order products are either zero or very small. Thus, **E** is sparse. Therefore, it is justified to use RPCA decomposition of **X** in (4) to suppress higher-order terms induced by a nonlinear mixing process. This yields approximation of **X**, that is, **A**, with suppressed higher-order terms. In the experiments reported in Section 3, we have used an accelerated proximal gradient algorithm [41], available with a MATLAB code at [42], to solve (5).

2.3.2. *Hard thresholding*

An HT operator [18] can be applied entry-wise to **X** in (4) according to $b_{nt} = HT(x_{nt}) = \begin{cases} x_{nt} \ \ if\ \ x_{nt}\ge\tau_1 \\ 0\ if\ \ x_{nt} < \tau_1 \end{cases}$, $n = 1, \ldots, N$, $t = 1, \ldots, T$ and $\tau_1\in[10^{-6}, 10^{-4}]$, which stands for a threshold. HT preprocessing transform of **X** yields matrix **B** that is expected to have the same structure as **A** in (5).

2.3.3. *Soft thresholding*

An ST operator [18] can be applied entry-wise to **X** in (4) according to $c_{nt} = ST(x_{nt}) = \max(0, x_{nt} - \tau_2)$, $n = 1, \ldots, N$, $t = 1, \ldots, T$ and $\tau_2\in[10^{-6}, 10^{-4}]$. ST preprocessing transform of **X** yields matrix **C** that, same as **B** obtained by HT, is also expected to have the same structure as **A** in (5).

2.3.4. *Trimmed thresholding*

A TT operator [19] is applied entry-wise to **X** in (4) according to

$$d_{nt} = TT(x_{nt}) = \begin{cases} x_{nt}\dfrac{x_{nt}^{\alpha} - \tau_3^{\alpha}}{x_{nt}^{\alpha}} \ \ if\ \ x_{nt}\ge\tau_3 \\ 0\ if\ \ x_{nt} < \tau_3 \end{cases}, n = 1,\ \ldots,\ N,\ t = 1,\ \ldots,\ T$$

and $\tau_3\in[10^{-6}, 10^{-4}]$. α is a trade-off parameter between HT and ST. When $\alpha = 1$, TT equals ST. When $\alpha\to\infty$, TT is equivalent to HT. Herein, we set $\alpha = 3.5$ because this value yields TT to operate between ST and HT [19]. TT preprocessing transform of **X** yields matrix **D** that, same as **B** obtained by HT and **C** obtained by ST, is also expected to have the same structure as **A** in (5).

2.4. Empirical kernel map-based nonlinear mapping of preprocessed mixture matrix

So far, we have substituted uNNBSS problem (1)/(2) by four uLNBSS problems in the form of (4). While the original uNNBSS problem is characterized by nonlinear multivariate mapping **f** and triplet (N, M, L), the uLNBSS problems are characterized by (N, P, Q), where $P\approx 2M + M(M-1)/2$ stands for the number of dependent sources in (4) and $Q\approx 2L + L(L-1)/2$ stands for the maximal number of sources at particular *m/z* coordinates. Because by assumption A3, $M > N$, it follows that $P\gg N$. Thus, even with the activation of sparseness constraints imposed by A4, it will be virtually impossible to ensure an essentially unique solution of these uLNBSS problems. To this end, as in [4], we apply the EKM-based nonlinear mapping of uLNBSS problems represented by preprocessed mixture matrices **A**, **B**, **C** and **D** to RKHS in order to increase number of samples/mixtures from N to $D\gg N$. The theory and discussion related to it have been presented in great detail in Section 2.2 in [4]. We therefore present it in a concise form herein. EKM Ψ of column vectors $\{\mathbf{a}_t\}_{t=1}^{T}$ in (4) with respect to a basis $\{\mathbf{v}_d\}_{d=1}^{D}$ is $\psi: R^N\to R^D$, such that $\mathbf{a}_t\mapsto\kappa(\circ,\mathbf{a}_t)\big|_{\{\mathbf{v}_d\}_{d=1}^{D}} = [\kappa(\mathbf{v}_1,,\mathbf{a}_t),\ldots,\kappa(\mathbf{v}_D,,\mathbf{a}_t)]^T\forall t = 1,\ldots,T$. Thereby, $\kappa(\mathbf{v}_d, \mathbf{a}_t)$ is a positive definite symmetric function. The basis $\{\mathbf{v}_d\}_{d=1}^{D}$ has

[3] These threshold values can be justified by the following analysis. Because of A1 and A2, elements of **G** in (7) in the Supporting Information are less than 1. In pursuing the worst-case analysis of third-order effects, we assume that third-order derivative coefficients in **G** are less than some value g_3. Thus, the contribution of third-order terms is limited above by $x^{(3)} = M^{(3)}g_3 s$. If the mixture value x_{nt} is greater than $x^{(3)}$, then it is probably due to first-order and second-order terms. The threshold value evidently depends on values of $M^{(3)}$, g_3 and s. For example, assuming $M = 100$ ($M^{(3)} = 171{,}700$), $g_3 = 0.1$ and $s = 3.4\times10^{-7}$, we obtain $x^{(3)} = 5.8\times10^{-3}$. However, this is overly pessimistic given the fact that most of the third-order cross-products will, owing to sparseness, vanish. Thus, the optimal threshold value is somewhere in the interval 10^{-6}, 10^{-4}.

to span the empirical set of patterns $\{\mathbf{a}_t\}_{t=1}^{T}$ such that $span\{\mathbf{v}_d\}_{d=1}^{D} \approx span\{\mathbf{a}_t\}_{t=1}^{T}$. In this case, $span\{\phi(\mathbf{v}_d)\}_{d=1}^{D} \approx span\{\phi(\mathbf{a}_t)\}_{t=1}^{T}$, where $\left\{\mathbf{a}_t \mapsto \phi(\mathbf{a}_t) \in R_{0+}^{\overline{N}}\right\}_{t=1}^{T}$, that is, $\left\{\mathbf{v}_d \mapsto \phi(\mathbf{v}_d) \in R_{0+}^{\overline{N}}\right\}_{d=1}^{D}$, is in principle an infinite-dimensional nonlinear mapping. If $\phi(\mathbf{a}_t) = \kappa(\circ, \mathbf{a}_t)$ and $\phi(\mathbf{v}_d) = \kappa(\circ, \mathbf{v}_d)$, projection of $\{\phi(\mathbf{a}_t)\}_{t=1}^{T}$ onto $\{\phi(\mathbf{v}_d)\}_{d=1}^{D}$ yields in matrix form

$$\Psi(\mathbf{A}) = \begin{bmatrix} \kappa(\mathbf{a}_1, \mathbf{v}_1) & \dots & \kappa(\mathbf{a}_T, \mathbf{v}_1) \\ \dots & \dots & \dots \\ \kappa(\mathbf{a}_1, \mathbf{v}_D) & \dots & \kappa(\mathbf{a}_T, \mathbf{v}_D) \end{bmatrix} \tag{6}$$

Herein, as in [4], we choose $\kappa(\mathbf{a}_t, \mathbf{v}_d) = \exp(-\|\mathbf{a}_t - \mathbf{v}_d\|^2/\sigma^2)$. When assumption A1 holds, we can set $\sigma^2 \approx 1$. We analogously obtain empirical kernel mappings of matrices **B**, **C** and **D**, which respectively yields $D \times T$ matrices $\Psi(\mathbf{B})$, $\Psi(\mathbf{C})$ and $\Psi(\mathbf{D})$. Likewise, as in [4], we use a *k*-means data clustering algorithm to estimate basis **V** by clustering $\{\mathbf{a}_t\}_{t=1}^{T}$ in *D* clusters. Thereby, by setting $D = T$, clustering is unnecessary because each empirical pattern is a basis vector. This, however, comes at an increased computing cost. By using sparseness assumption A4, it is shown in [4] that

$$\Psi(\mathbf{A}) = \mathbf{Z} + \overline{\mathbf{G}} \begin{bmatrix} \mathbf{0}_{1 \times T} \\ \overline{\mathbf{S}} \end{bmatrix} + HOT \tag{7}$$

where **Z** is a bias term and does not play a role in 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 original source components and their second-order monomials. $\overline{\mathbf{G}}$ is a matrix of appropriate dimensions. Empirical kernel-mapped matrices $\Psi(\mathbf{B})$, $\Psi(\mathbf{C})$ and $\Psi(\mathbf{D})$ follow the same approximation as $\Psi(\mathbf{A})$ in (7). It is important to emphasize that in (4), higher-order (error) terms are induced by nonlinear mixing process $\mathbf{f}(\mathbf{S})$, while in (7), they are induced by the nonlinear character of the EKM. That is, an increase of the number of mixtures from *N* to *D* in $\Psi(\mathbf{A})$, $\Psi(\mathbf{B})$, $\Psi(\mathbf{C})$ and $\Psi(\mathbf{D})$ comes at the cost of errors induced by the EKM. However, as in [4] and (4), we can again apply preprocessing transforms to suppress *HOT*. Because matrices **B**, **C** and **D** were obtained by respectively applying *HT*, *ST* and *TT* operators on **X** in (4), we apply these operators in the same order on $\Psi(\mathbf{B})$, $\Psi(\mathbf{C})$ and $\Psi(\mathbf{D})$. In order to keep the level of notational complexity as low as possible, we keep the same notation for thresholded versions of matrices $\Psi(\mathbf{B})$, $\Psi(\mathbf{C})$ and $\Psi(\mathbf{D})$. We do not apply RPCA decomposition on $\Psi(\mathbf{A})$ because its rank is dictated by **Z** and is equal to $\min(D, T) = D$, which is not low. The final effect of EKM-based mappings is to ensure that sparseness-constrained factorization of $\Psi(\mathbf{A})$, $\Psi(\mathbf{B})$, $\Psi(\mathbf{C})$ and $\Psi(\mathbf{D})$ yields, with greater probability, more accurate solution compared with decomposition by the same method of **A**, **B**, **C** and **D**. This will be the case when the following condition holds:

$$(D/N) >> (P/M) \text{ and } (D/N) >> (Q/L) \tag{8}$$

Because $P \approx 2M + M(M-1)/2$ and $Q \approx 2L + L(L-1)/2$, condition (8) becomes $(D/N) \gg (M/2 - 3/2)$ and $(D/N) \gg (L/2 - 3/2)$. The numerical problem studied in Section 3 is characterized by $N = 3$, $M = 8$, $L = 3$ and $D = T = 1000$. Evidently, the preceding condition is fulfilled.

2.5. Sparseness-constrained factorization

To increase the accuracy of the pure component extraction, we apply sNMF in RKHS to matrices $\Psi(\mathbf{A})$, $\Psi(\mathbf{B})$, $\Psi(\mathbf{C})$ and $\Psi(\mathbf{D})$.[4] This yields four sets of separated components:

$$\left\{\overline{\mathbf{s}}_m^{\mathbf{A}}\right\}_{m=1}^{P} = sNMF(\Psi(\mathbf{A})) \tag{9}$$

$$\left\{\overline{\mathbf{s}}_m^{\mathbf{B}}\right\}_{m=1}^{P} = sNMF(\Psi(\mathbf{B})) \tag{10}$$

$$\left\{\overline{\mathbf{s}}_m^{\mathbf{C}}\right\}_{m=1}^{P} = sNMF(\Psi(\mathbf{C})) \tag{11}$$

$$\left\{\overline{\mathbf{s}}_m^{\mathbf{D}}\right\}_{m=1}^{P} = sNMF(\Psi(\mathbf{D})) \tag{12}$$

When it comes to implementation of the sNMF algorithms, we use, as in [4], the NMU algorithm [40] with a MATLAB code available at [44] and the NMF_L0 algorithm [39] with a MATLAB code available at [45]. The NMF_L0 algorithm was run with the following parameter setup: reverse sparse nonnegative least square sparse coder and alternating nonnegative least square for dictionary update stage. A main reason for preferring the NMU algorithm over other sNMF algorithms is that there are no regularization constants that require a tuning procedure. When performing NMU-based factorizations in (9)-(12), the unknown number of pure components *P* needs to be given to the algorithm as an input. As in [4], we set $P = D = T$. That is, in order not to lose some component, we prefer to extract all *T* rank-one factors.[5] These four sets of separated components are compared with the pure components stored in the library using normalized correlation coefficients. Each pure component is associated with the separated component by which it has the highest correlation. As a reference, in the benchmark numerical study, we have used the solution obtained by applying the NMF_L0 algorithm to (9)-(12). Afterwards, the maximal correlation criterion has been used to assign separated components to pure components in the library. NMF_L0 is based on a natural sparseness measure, the ℓ_0-pseudo-norm of the component matrix $\overline{\mathbf{S}}$, and this is known from compressed sensing theory [47] to yield the best results when sparseness of $\overline{\mathbf{S}}$ decreases. The NMF_L0 when applied in (9)-(12) requires *a priori* information on the number of components *P* and number of overlapping components *Q*, and they are related to *M* and *L* through $P \approx 2M + M(M-1)/2$ and $Q \approx 2L + L(L-1)/2$. In a numerical scenario, both *M* and *L* are known, while in an experimental scenario, the selection of the optimal (true) value of *L* is hard. We summarize the PTs-EKM-NMU/NMF_L0 algorithm in Algorithm 1.

[4] To ensure essentially unique decomposition, sNMF algorithms have been formulated such as in [38–40]. However, only very recently is it proven in [43] (Theorem 4 and Corollary 2) that the uniqueness of some asymmetric NMF $\mathbf{S} = \mathbf{WH}$ implies that each column of **W** (row of **H**) contains at least $M - 1$ zeros, where *M* is a nonnegative rank of **S**.

[5] The factorization problems (9–12) are related to the determination of the nonnegative rank of a nonnegative matrix, which is defined as the smallest number of rank 1 matrices into which the original matrix can be decomposed [46]. For some matrix $\mathbf{\Psi} \in R_{0+}^{D \times T}$ with $D \leq T$, a nonnegative rank equals the smallest positive integer *P* for which there exist nonnegative column vectors $\left\{\mathbf{g}_p\right\}_{p=1}^{P}$ such that each column vector of $\mathbf{\Psi}$ can be represented as a linear combination with nonnegative coefficients of the column vectors $\left\{\mathbf{g}_p\right\}_{p=1}^{P}$.

3. EXPERIMENTS

Studies on numerical and experimental data reported in the following 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. A MATLAB 2012b environment has been used for programming.

Algorithm 1. The PTs-EKM-NMF (preferably NMU) algorithm.

```
Required:

    $\mathbf{X}\in R_{0+}^{N\times T}$. If A1 is not satisfied, perform scaling
    $\mathbf{X}\to\mathbf{X}/\ \arg\max_t\{\|\mathbf{x}_t\|_1\}_{t=1}^{T}$ or $\mathbf{X}\to\mathbf{X}/\ \arg\max_{nt}\{\mathbf{X}_{nt}\}_{n,t=1}^{N,T}$.

1. Perform RPCA (5) on X in (2)/(4) with $\lambda\approx 1/\sqrt{T}$. It yields
   approximation A in (4).
2. Perform HT on X in (2)/(4) with $\tau_1\in[10^{-6}, 10^{-4}]$. It yields
   approximation B.
3. Perform ST on X in (2)/(4) with $\tau_2\in[10^{-6}, 10^{-4}]$. It yields
   approximation C.
4. Perform TT on X in (2)/(4) with $\tau_3\in[10^{-6}, 10^{-4}]$ and $\alpha=3.5$.
   It yields approximation D.
5. Perform empirical kernel mappings $\mathbf{A}\to\Psi(\mathbf{A})$, $\mathbf{B}\to\Psi(\mathbf{B})$,
   $\mathbf{C}\to\Psi(\mathbf{C})$ and $\mathbf{D}\to\Psi(\mathbf{D})$ according to (6). Use Gaussian kernel
   with $\sigma^2=1$.
6. Perform HT, ST and TT, respectively, of matrices $\Psi(\mathbf{B})$, $\Psi(\mathbf{C})$ and
   $\Psi(\mathbf{D})$.
7. Perform sparseness-constrained factorization, preferably by
   the NMU algorithm, of matrices $\Psi(\mathbf{A})$, $\Psi(\mathbf{B})$, $(\Psi\mathbf{C})$ and $\Psi(\mathbf{D})$ to
   obtain separated components $\bar{\mathbf{S}}^{\mathbf{A}}$,$\bar{\mathbf{S}}^{\mathbf{B}}$, $\bar{\mathbf{S}}^{\mathbf{C}}$ and $\bar{\mathbf{S}}^{\mathbf{D}}$.
8. Assign to pure components from the library those separated
   components $\bar{\mathbf{S}}^{\mathbf{A}}$, $\bar{\mathbf{S}}^{\mathbf{B}}$, $\bar{\mathbf{S}}^{\mathbf{C}}$ and $\bar{\mathbf{S}}^{\mathbf{D}}$ with the highest normalized
   correlation coefficient.
```

3.1. Numerical study

In a numerical study, we simulate uNNBSS problem (2) with $N=3$, $M=8$, $L=3$ and $T=1000$. Source signals were generated according to mixed state probabilistic model (3) with exponential prior. Thereby, $\mu_m=1.5\times10^{-3}\ \forall\ m=1, \ldots, M$. We have generated two scenarios with $\rho_m=0.5$ and $\rho_m=0.8\ \forall\ m=1, \ldots, M$. Values for μ_m and ρ_m are equivalent to those obtained by fitting probabilistic model (3) to experimental mass spectra of 25 pure components; see Section 3.2 and Figure 4 for details. The uNNBSS problem (2) has been simulated using nonlinear mixtures:

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

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

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

Nonlinear mixtures are chosen arbitrarily to demonstrate the capability of the proposed algorithm to solve the uNNBSS problem comprised of unknown nonlinear mixtures. HT, ST and TT operators used in steps 2–4 and 6 in Algorithm 1 were implemented with $\tau=10^{-5}$, and $\alpha=3.5$ has been used for TT operator. Gaussian kernel-based EKM has been used with $\sigma^2=1$ and $D=T=1000$. Table I shows the results of the comparative analysis, for the case of $\rho_m=0.8$, obtained by NMU and NMF_L0 applied to uNNBSS (1)/(2); NMU and NMF_L0 applied in (9–12) without suppression of higher-order monomials (EKM-NMU and EKM-NMF_L0); and NMU and NMF_L0 applied in (9–12) after RPCA, HT, ST and TT preprocessing transforms (PTs-EKM-NMU and PTs-EKM-NMF_L0). Because of sparse prior imposed on sources, it was reasonable to expect that useful results can be obtained by direct factorization of uNNBSS problem (2). Results for $\rho_m=0.5$ are shown in Table S1 in the Supporting Information, while results for $\rho_m=0.8$ and $\rho_m=0.5$ as a function of the Monte Carlo index are shown in Figure 1. For the value of a normalized correlation coefficient between a pure component and an assigned separated component, we evaluate the performance in terms of four metrics described in the notes of Table I. They are defined with respect to predefined labeling of the pure components stored in the library. The first three metrics are calculated for correctly assigned components only. That is why NMU and NMF_L0 appear to have comparable performance in

Table I. Comparative performance analysis of NMU, NMF_L0, EKM-NMU, EKM-NMF_L0, PTs-EKM-NMU and PTs-EKM-NMF_L0 algorithms

	NMU	NMF_L0	EKM-NMU	EKM-NMF_L0	PTs-EKM-NMU	PTs-EKM-NMF_L0
Correlation ≥ 0.6	2.8 ± 0.92	2.3 ± 1.34	3.7 ± 0.48	3.2 ± 0.63	**3.8 ± 0.42**	3.7 ± 0.48
Mean correlation	**0.70 ± 0.03**	0.61 ± 0.11	0.69 ± 0.02	0.64 ± 0.03	**0.70 ± 0.03**	0.69 ± 0.04
Minimal correlation	**0.53 ± 0.04**	0.42 ± 0.08	0.51 ± 0.03	0.45 ± 0.04	0.52 ± .04	0.49 ± 0.06
Incorrect assignments	3.4 ± 0.70	3.1 ± 0.57	2.4 ± 0.97	2.2 ± 0.63	2.0 ± 0.88	**1.5 ± 1.43**

Probability of zero state was $\rho_m=0.8$. The four metrics used in comparative performance analysis were the number of associated components with normalized correlation coefficient greater than or equal to 0.6, mean value of correlation coefficient over all associated components, minimal value of correlation coefficient and number of pure components assigned incorrectly (which occurs because of poor separation). All four metrics were calculated with respect to 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. Mean values and variance are reported and estimated over 10 Monte Carlo runs. The best result in each metric is in bold. The first three metrics are calculated only for correctly assigned components. That is why NMU and NMF_L0 appear to have comparable performance.

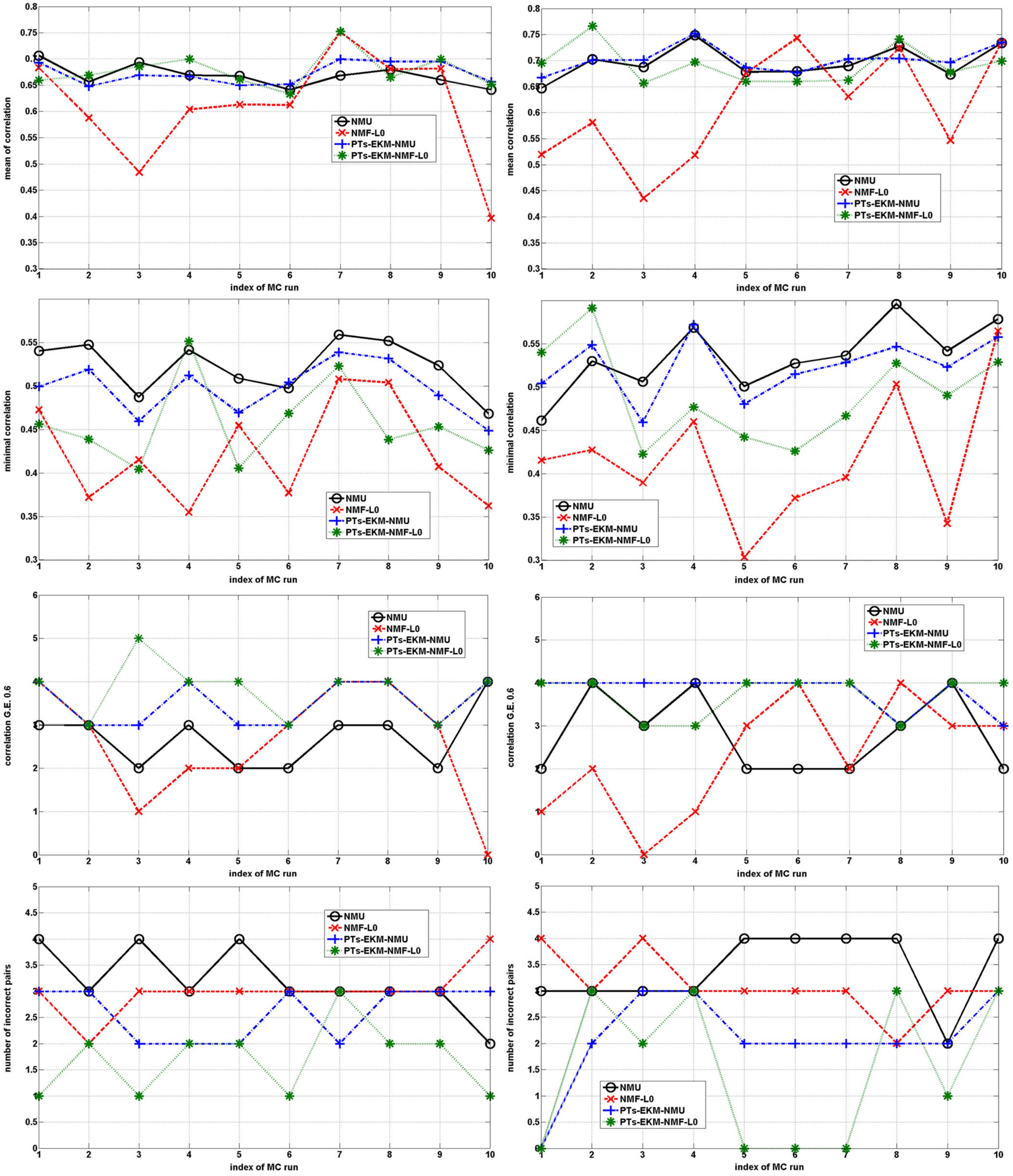

Figure 1. Numerical study. Normalized correlation coefficient versus Monte Carlo run index between true and extracted sources by algorithms NMF_L0 (crosses), NMU (circles) and PTs-EKM-NMU (pluses) and PTs-EKM-NMF_L0 (stars). Mean value (first row), minimal value (second row), number of values greater than or equal to 0.6 (third row) and number of incorrect pairs (fourth row). Probability of state 0 equal to 0.5 (left column) and 0.8 (right column).

terms of mean and minimal correlation metrics. But they are inferior in the number of separated components correlated with pure components with a correlation greater than or equal to 0.6 as well as in the number of (in)correctly assigned separated components (due to poor separation). Thereby, an incorrect assignment implies that two or more pure components are assigned to the same separated component. We also can see that preprocessing transforms improve performance compared with factorizations of mixture data without preprocessing related to suppression of higher-order monomials.

3.2. Experimental data on chemical reaction comprising peptide synthesis

3.2.1. *Chemicals*

Chemical reaction has been performed according to the following procedure: L-leucine (200 mg, 1.52 mmol) was dissolved in 5 mL of dry dimethylformamide, and the solution was cooled to 0 °C. *N*-methylmorpholine (3.05 mmol, 337 μL) and isobutyl chloroformate (3.34 mmol, 458 μL) were added. Aliquots of the reaction mixture (100 μL) were withdrawn every 30 min (t_0–t_8), and the solvent was evaporated and the residue dissolved in 1 mL of 0.1% formic acid (FA) in 50% MeOH. Aliquots (100 μL) were diluted with 400 μL of 0.1% FA in 50% MeOH, and 10 μL was injected through an autosampler on a column (Zorbax XDB C18, 3.5 μm, 4.7 mm) at a flow rate of 0.5 mL/min. The mobile phase was 0.1% FA in water (solvent A) and 0.1% FA in MeOH (solvent B). The gradient was applied as follows: 0 min 40% B, 0–15 min 90% B, 12–15 min 90% B, 17.1 min 40% B and 17.1–20 min 40% B. Figure S2 in the Supporting Information shows nine chromatograms corresponding to the reaction mixture recorded at nine time instants (t_0–t_8) during the reaction. The mass spectra of nine mixtures ($\mathbf{x}_1$–$\mathbf{x}_9$), obtained by full integration of chromatograms, and mass spectra of 25 pure components ($\mathbf{s}_1$–$\mathbf{s}_{25}$) arising during the reaction are respectively shown in Figures S3 and S4 in the Supporting Information. The mass spectra of pure components 1, 4, 8 and 11 are also shown in Figure 3.

3.2.2. *Mass spectroscopy measurements*

Electrospray ionization–mass spectrometry measurements operating in a positive ion mode were performed on an high-performance liquid chromatography–mass spectrometry triple quadrupole instrument equipped with an autosampler (Agilent Technologies, Palo Alto, CA, USA). The desolvation gas temperature was 300 °C with a flow rate of 8.0 L/min. The fragmentor voltage was 135 V, and the capillary voltage was 4.0 kV. Mass spectra were recorded in the *m/z* segment of 10–2000. All data acquisition and processing were performed using Agilent MassHunter software. Acquired mass spectra are composed of intensities at $T = 9901$ *m/z* coordinates.

3.2.3. *Setting up an experiment*

Peptides and proteins are compounds involved in numerous biological processes of key importance, like cell–cell communication, immune response, cell growth and proliferation, and hormonal and enzymatic activity. They are, therefore of ever-increasing interest as tools in studies of biological systems and modulators of biological functions. Chemical synthesis of peptides involves condensation of two suitably protected parts (amino acids or peptides) in order to obtain a single, desirable product. However, for the purpose of this work, a different approach was undertaken. Non-protected amino acid, L-leucine, was allowed to react under basic conditions (*N*-methylmorpholine) in the presence of isobutyl chloroformate, giving various products: dipeptides, tripeptides, tetrapeptides and corresponding intermediates. The nonlinearity of the described reaction was assured based on the following: (i) the concentration of individual components does not change linearly with time and (ii) as the reaction proceeds, new components appear that were not present at the beginning of the reaction. Figure 2 schematically describes the possible components present in the reaction mixture. It is important to note that the aim of this experiment was not to determine the structure of all components, but to provide reliable experimental data on nonlinear reaction. A library of compounds required for the validation of the algorithm was built by integration of each peak in the chromatogram corresponding to the mixture $\mathbf{x}_9$ and subsequent extraction of mass spectrum. During the library generation, no discrimination based on the intensity of peaks was made. Therefore, all peaks were treated as pure components.

4. RESULTS AND DISCUSSION

Inspection of pure component mass spectra shown in Figure S4 in the Supporting Information shows significant overlapping, resulting from the similarity of the chemical structure of components. Pure components 1 and 2, 16 and 17 as well as 19 and 21 have normalized correlation coefficients above 0.97, and consequently, they are impossible to distinguish. In addition to that, pure components 5 and 7 have normalized correlation coefficients above 0.78. Thus, they are also expected to be very hard to discriminate. However, we expect from the proposed PTs-EKM-NMU method to be able to discriminate the rest of the components. This is not trivial given the fact that normalized correlation coefficients

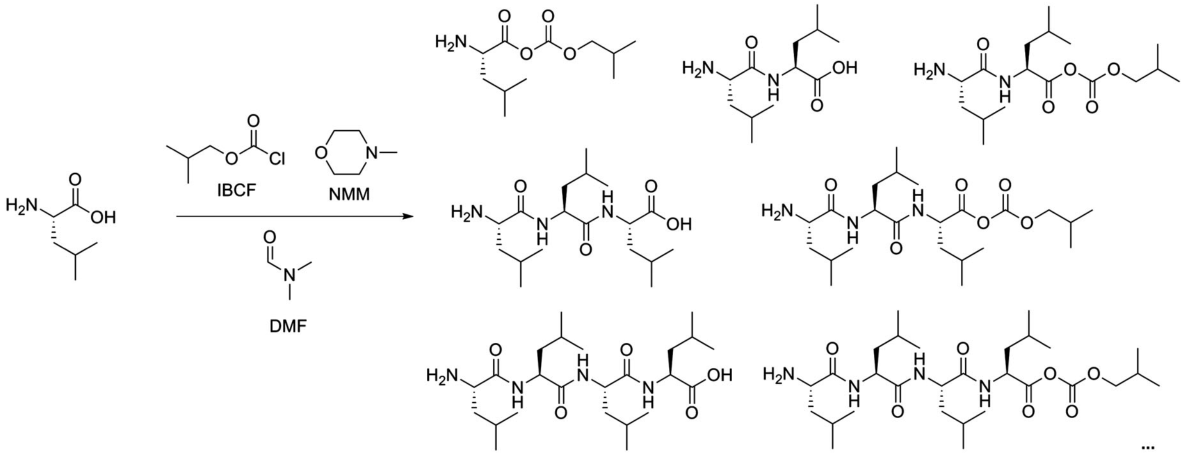

Figure 2. Structures of possible components present in the reaction mixture.

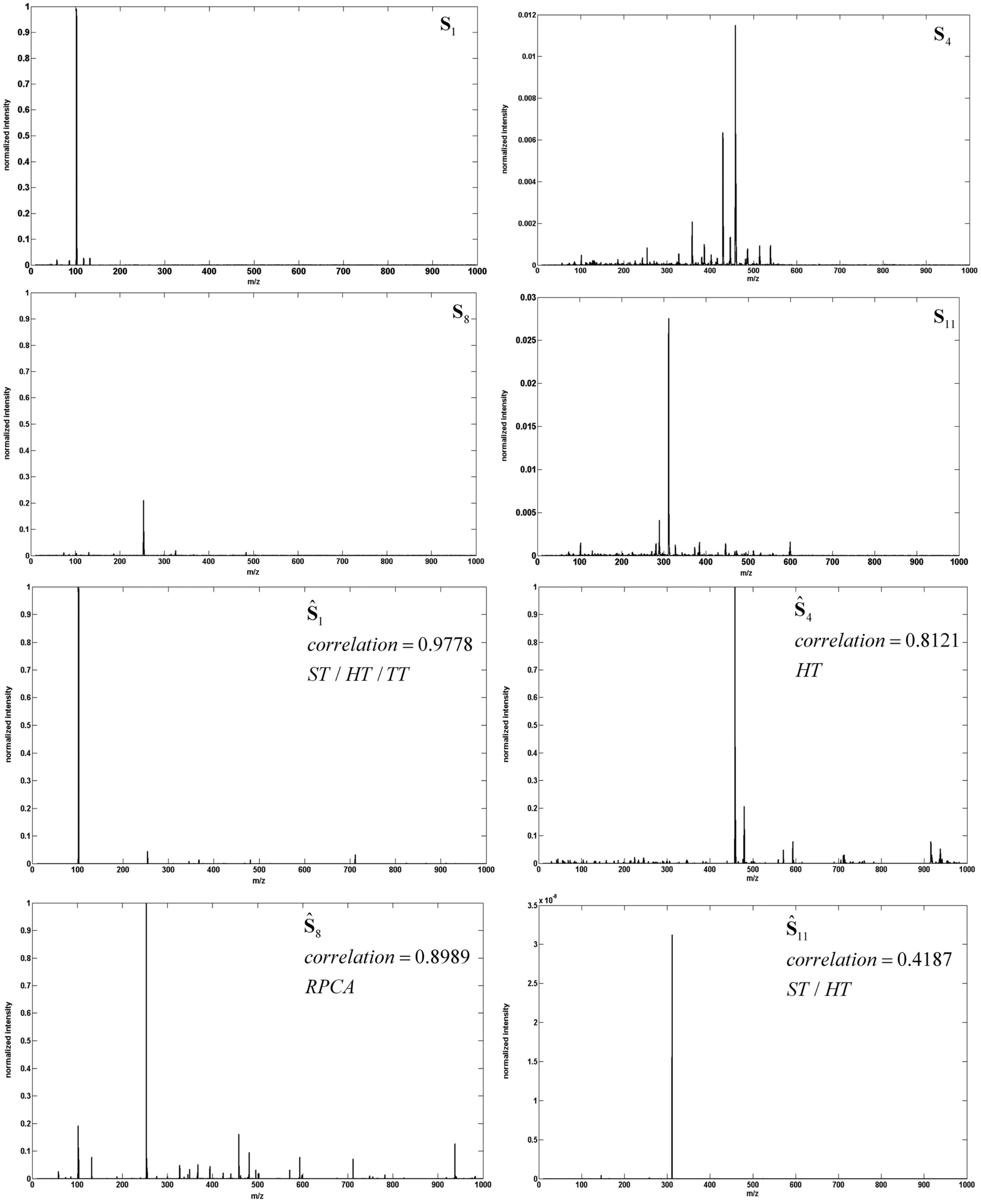

Figure 3. Two top rows: mass spectra of pure components $\mathbf{s}_1$, $\mathbf{s}_4$, $\mathbf{s}_8$ and $\mathbf{s}_{11}$. Two bottom rows: estimated mass spectra of pure components $\mathbf{s}_1$, $\mathbf{s}_4$, $\mathbf{s}_8$ and $\mathbf{s}_{11}$ by the proposed PTs-EKM-NMU algorithm. Information on the value of the highest normalized correlation coefficient and associated error reduction method (RPCA, HT, ST and TT) is also displayed.

for 26 combinations of pure components vary between 0.1 and 0.44. This makes the uNNBSS problem comprised of correlated pure components very hard. The correlation matrix of the pure component mass spectra, where pairs of pure components are identified with a normalized correlation coefficient above 0.1, is shown in Table S2 in the Supporting Information. As emphasized previously, it is the sparseness of the pure component mass spectra in support and amplitude that is expected to enable the solution of related uNNBSS. To this end, mixed state probabilistic model (3) with exponential prior on continuous distribution of the nonzero amplitude has been fitted to experimental pure component mass spectra (they are shown in Figure S4 in the Supporting Information as well as in Figure 3 for pure components 1, 4, 8 and 11). Even though these pure components are correlated with others and some (4 and 11) have small intensity, they are uniquely assigned to the true pure components from the library. Figure 4 (left), also Figure S5 in the Supporting Information, shows the estimated probability that the value of the pure component mass spectra is zero. As can be seen, 22 out of 25 pure components have zero amplitudes at 40–75% of their support. Figure 4 (right), also Figure S6 in the Supporting Information, shows most expected values (mean) of exponential distribution estimated by fitting exponential distribution to amplitude histograms. They were estimated for 25 pure components in the range (0, 1] within intervals of the 0.01 width. It can be seen that $\hat{\mu}_m \in [0.0012, 0.0014]$ for $m = 1, \ldots, 25$. Figure S7 shows the probability that the amplitude of the pure component mass spectra occurs in the interval [0, A], such that $0.01 \leq A \leq 1$, that is, an average estimate over 25 pure components. It is seen that $0.01 \leq A \leq 0.08$ occurs with a probability of 0.97. Reported results confirm that sparse probabilistic model (3) is experimentally well grounded. This is further confirmed by Figure S8 in the Supporting Information, which shows estimated histograms (stars) and exponential probability density functions (squares) calculated with the mean values from Figure 4 (right). It is seen that approximation is very good. Estimated histograms versus exponential probability density functions for pure components 1, 4, 8 and 11 are also shown in Figure 5.

Table II presents the results of the comparative performance analysis using the four metrics as in Section 3.1 for NMU, EKM-NMU, PTs-EKM-NMU for $D = T = 9901$ and PTs-EKM-NMU for $D = 4000$. Thus, in the last case, k-means clustering has been used to find a basis $\{\mathbf{v}_d\}_{d=1}^{4000}$ in the input space of patterns $\{\mathbf{x}_t\}_{t=1}^{9901}$. Provided that it retains accuracy, the subspace approximation is very important from computational reasons. This is because when four preprocessing transforms are combined, sNMF in (9)-(12) has to be performed four times. This can be carried out in parallel. Nevertheless, one factorization of the 9901 × 9901 matrix by the NMU algorithm takes approximately 79 h on an earlier specified machine, while factorization of the 4000 × 9901 matrix by the same algorithm takes approximately 13.7 h. For NMF_L0, the number of overlapping components, L, has to be reported to the algorithm as input information. For the PTs-EKM-NMF_L0 algorithm, the optimal value of L can be inferred by running the NMF_L0 algorithm multiple times on a problem such as (9). This, however, would result in high computational costs. That is why NMF_L0 has not been used in RKHS on problems (9)-(12). It is seen from Table II that linear sparseness-constrained matrix factorization yields poor quality of separation compared with linear factorization in the RKHS. This is especially the case with the number of incorrectly assigned components and is a direct consequence of the low purity of separated components. This, indirectly, also confirms the nonlinear character of the mixture mass spectra of the desired chemical reaction. It is also seen the that combination of four preprocessing transforms for suppression of higher-order monomials and sparseness-constrained factorization in RKHS significantly improves the quality of separation. In this regard, Figure S9 in the Supporting Information shows the mass spectra of 25 separated components assigned to pure components according to the maximal correlation criterion. Separated pure components 1, 4, 8 and 11 are also shown in Figure 3. Thereby, the value of the normalized correlation coefficient and preprocessing transform (RPCA, HT, ST or TT) that yielded the best result are also reported. Because of the diversity of morphologies of mass spectra, all four preprocessing transforms yielded the best results at some cases. It is also important to notice that subspace approximation of proposed method with $D = 4000$ yields results very comparable with those obtained by $D = 9901$ but with a much shorter computation time. Thus, the proposed approach to pure component extraction can, when implemented on a state-of-the-art multiprocessor (grid) platform, be executed in an even shorter time, which makes it practically relevant.

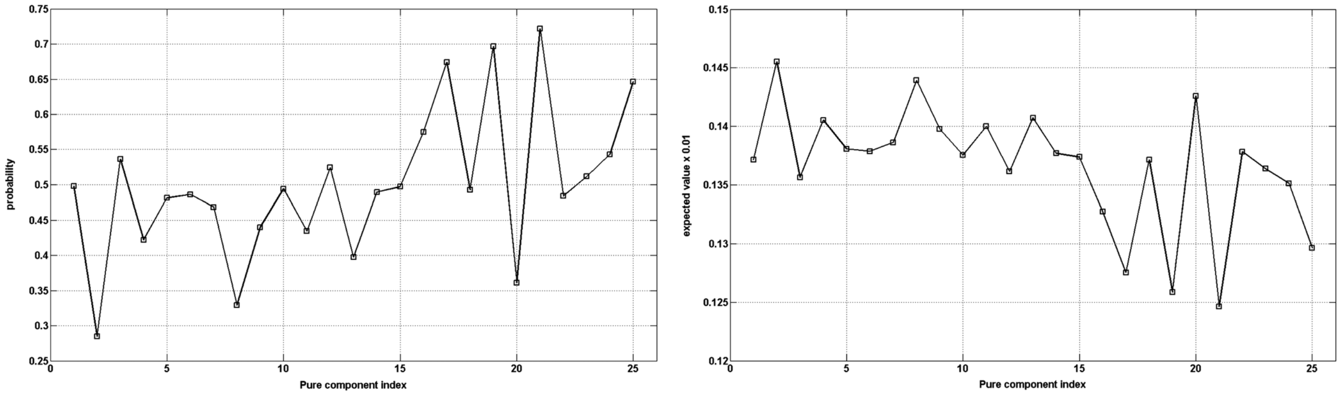

Figure 4. Experimental study. Left: estimated probability that the value of the pure component mass spectra is zero, that is, estimate of ρ_m, $m = 1, \ldots, 25$. Right: estimates of most expected values (means) of exponential distribution obtained by fitting the exponential distribution to amplitude histograms. They were estimated for 25 pure components in the range (0, 1] within intervals of the 0.01 width.

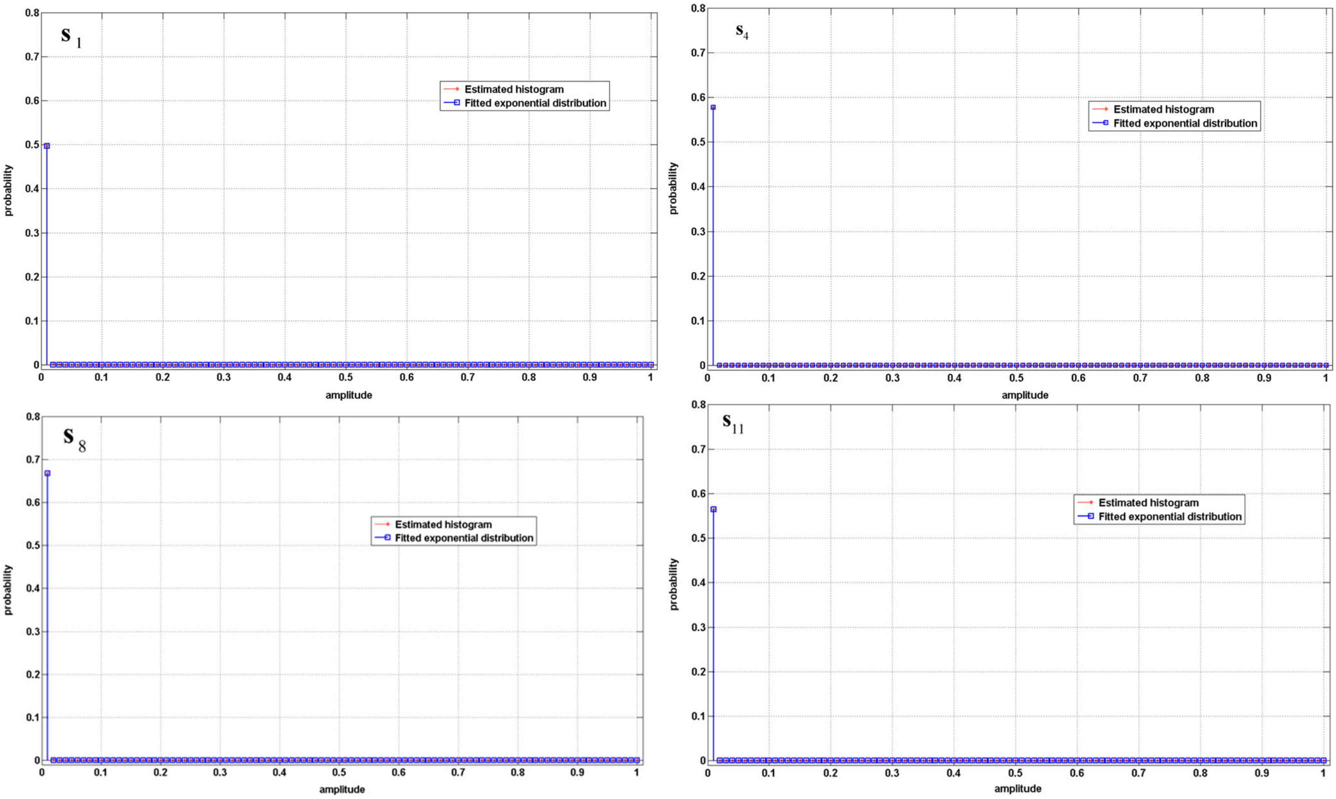

Figure 5. Experimental study for pure components 1, 4, 8 and 11. Estimated histograms (stars) versus exponential probability density functions (squares), calculated with the estimates of mean values shown in Figure 4 (right), fitted to amplitude histograms.

Table II. Comparative performance analysis of NMU, NMF_L0, EKM-NMU, PTs-EKM-NMU ($D = T = 9901$) and PTs-EKM-NMU ($D = 4000$) algorithms of nine experimental nonlinear mixture mass spectra related to peptide synthesis

	NMU	NMF_L0	EKM-NMU	PTs-EKM-NMU	PTs-EKM-NMU
				$D = T = 9901$	$D = 4000$
Correlation ≥ 0.6	8	14	16	**18**	**18**
Mean correlation	0.342	0.518	0.673	**0.702**	**0.708**
Minimal correlation	0.038	0.039	0.267	**0.419**	0.283
Incorrect assignments	15	7	**0**	**0**	1
CPU time	1.3 s	40 s	78.78 h	4 × 78 h	4 × 13.7 h

The number of pure components equals 25. The four metrics used in comparative performance analysis were number of associated components with normalized correlation coefficient greater than or equal to 0.6, mean value of correlation coefficient over all associated components, minimal value of correlation coefficient and number of pure components assigned incorrectly (which occurs because of poor separation). The best result in each metric is in bold. The first three metrics are calculated only for correctly assigned components.

5. CONCLUSION

A blind source separation approach to pure component extraction is most often based on a linear mixture model. That is, mixture spectra are assumed to be the unknown weighted linear combination of pure component spectra. Herein, we have addressed the problem related to extraction of pure components from nonlinear mixtures of mass spectra. Thereby, the number of mixtures is assumed to be (significantly) less than the number of pure components. We propose an approach that combines four preprocessing methods for suppression of higher-order monomials induced by nonlinear mixing process and sNMF in RKHS induced by EKM. Two practically important properties of the proposed approach are that no information about the character of the nonlinear mixing process is required and that the linear mixing problem is contained implicitly as a special case. It is believed that these properties make the proposed approach practically relevant for contemporary metabolic profiling of biological samples, that is, pure component extraction in biomarker identification studies. The proposed approach is demonstrated on demanding numerical and experimental scenarios. In the last case, related to chemical reaction of synthesis of peptides, components separated from nine nonlinear mixture mass spectra are assigned uniquely to 25 pure components from the library. On the same problem, separation by linear NMF algorithms yielded 15 (NMU) and 7 (NMF_L0) incorrectly assigned components.

Acknowledgement

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. Nuzillard D, Bourg S, Nuzilard JM. Model-free analysis of mixtures by NMR using blind source separation. *J. Magn. Reson.* 1998; **133**: 358–363.
2. Visser E, Lee TW. 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.
3. 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.
4. 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.
5. 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.
6. Abu-Farha M, Elisma F, Zhou H, Tian R, Asmer MS, Figeys D. Proteomics: from technology developments to biological applications. *Anal. Chem.* 2009; **81**: 4585–4599.
7. McLafferty FW, Stauffer DA, Loh SY, Wesdemiotis C. Unknown identification using reference mass spectra. Quality evaluation of databases. *J. Am. Soc. Mass Spectrom.* 1999; **10**: 1229–1240.
8. Hyvärinen A, Karhunen J, Oja E. *Independent Component Analysis*, John Wiley & Sons, Inc.: New York,2001.
9. Cichocki A, Amari S. *Adaptive Blind Signal and Image Processing*, John Wiley: New York,2002.
10. Cichocki A, Zdunek R, Phan AH, Amari SI. *Nonnegative Matrix and Tensor Factorizations*, John Wiley: Chichester, UK,2009.
11. Comon P, Jutten C (eds). *Handbook of Blind Source Separation*. Academic Press: Oxford, UK, 2010.
12. Walleczek J. (ed). *Self-organized Biological Dynamics and Non-linear Control*. Cambridge University Press: Cambridge, UK, 2000.
13. Nicholson JK, Lindon JC. Systems biology: metabonomics. *Nature* 2008; **455**(7216): 1054–1056.
14. Bouthemy P, Piriou CHG, Yao J. Mixed-state auto-models and motion texture modeling. *J. Math Imaging Vision* 2006; **25**: 387–402.
15. Caifa C, Cichocki A. Estimation of sparse nonnegative sources from noisy overcomplete mixtures using MAP. *Neural Comput.* 2009; **21**: 3487–3518.
16. Candès EJ, Li X, Ma Y, Wright H. Robust principal component analysis? *J. ACM* 2011; **58**: Article 11 (37 pages).
17. Chandrasekaran V, Sanghavi S, Paririlo PA. Wilsky AS Rank-sparsity incoherence for matrix decomposition. *SIAM J. Opt.* 2011; **21**: 572–596.
18. Donoho DL. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* 1995; **41**(3): 613–627.
19. Fang HT, Huang DS. 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, China; 1621–1624.
20. The website of the ASTER spectral library. http://speclib.jpl.nasa.gov [13 January 2014]
21. Zhang K, Chan L. Minimal nonlinear distortion principle for nonlinear independent component analysis. *J. Mach. Learn. Res.* 2008; **9**: 2455–2487.
22. Levin DN. Using state space differential geometry for nonlinear blind source separation. *J. Appl. Phys.* 2008; **103**: 044906:1–12.
23. Levin DN. Performing nonlinear blind source separation with signal invariants. *IEEE Trans. Sig. Proc.* 2010; **58**: 2131–2140.
24. Taleb A, Jutten C. Source separation in post-nonlinear mixtures. *IEEE Trans. Sig. Proc.* 1999; **47**: 2807–2820.
25. Duarte LT, Suyama R, Rivet B, Attux R, Romano JMT, Jutten C. Blind compensation of nonlinear distortions: applications to source separation of post-nonlinear mixtures. *IEEE Trans. Sig. Proc.* 2012; **60**: 5832–5844.
26. Filho EFS, de Seixas JM Calôba LP Modified post-nonlinear ICA model for online neural discrimination. *Neurocomputing* 2010; **73**: 2820–2828.
27. Nguyen TV, Patra JC, Das A. A post nonlinear geometric algorithm for independent component analysis. *Digital Sig. Proc.* 2005; **15**: 276–294.
28. Ziehe A, Kawanabe M, Harmeling S, Müller KR. Blind separation of post-nonlinear mixtures using Gaussianizing transformations and temporal decorrelation. *J. Mach. Learn. Res.* 2003; **4**: 1319–1338.
29. Zhang K, Chan LW. Extended Gaussianization method for blind separation of post-nonlinear mixtures. *Neural Comput.* 2005; **17**: 425–452.
30. Harmeling S, Ziehe A, Kawanabe M. Kernel-based nonlinear blind source separation. *Neural Comput.* 2003; **15**: 1089–1124.
31. Martinez D, Bray A. Nonlinear blind source separation using kernels. *IEEE Tr. Neural Net.* 2003; **14**: 228–235.
32. Almeida L. MISEP-linear and nonlinear ICA based on mutual information. *J. Mach. Learn. Res.* 2003; **4**: 1297–1318.
33. Vaerenbergh SV, Santamaria IA. Spectral clustering approach to underdetermined postnonlinear blind source separation of sparse sources. *IEEE Trans. Neural Net.* 2006; **17**: 811–814.
34. Buciu I, Nikolaidis N, Pitas I. Nonnegative matrix factorization in polynomial feature space. *IEEE Trans. Neural Net.* 2007; **19**: 1090–1100.
35. Zafeiriou S, Petrou M. Non-linear non-negative component analysis. *IEEE Trans. Image Proc.* 2010; **19**: 1050–1066.
36. Pan B, Lai J, Chen WS. Nonlinear nonnegative matrix factorization based on Mercer kernel construction. *Pattern Rec.* 2011; **44**: 2800–2810.
37. 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.
38. Cichocki A, Zdunek R, Amari SI, Hierarchical ALS. Algorithms for nonnegative matrix factorization and 3D tensor factorization. *LNCS* 2007; **4666**: 169–176.
39. Peharz R, Pernkopf F. Sparse nonnegative matrix factorization with ℓ^0-constraints. *Neurocomputing* 2012; **80**: 38–46.
40. Gillis N, Glineur F. Using underapproximations for sparse nonnegative matrix factorization. *Pattern Rec.* 2010; **43**: 1676–1687.
41. Lin Z, Ganesh A, Wright J, Wu L, Chen M, Ma Y. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. *UIUC Technical Report UILU-ENG-09-2214*, August 2009.
42. The website on low-rank matrix recovery and completion via convex optimization. http://perception.csl.illinois.edu/matrix-rank/sample_code.html [13 January 2014]
43. Huang K, Sidiropoulos ND Swami A Non-negative matrix factorization revisited: uniqueness and algorithm for symmetric decomposition. *IEEE Trans. Sig. Proc.* 2014; **62**: 211–224.
44. The Nicolas Gillis Website. https://sites.google.com/site/nicolasgillis/code [13 January 2014].
45. The Robert Peharz Website. http://www3.spsc.tugraz.at/people/robert-peharz [13 January 2014].
46. Cohen JE, Rothblum UC. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra and Its Applications* 1993; **190**: 149–168.
47. Chartran R, Staneva V. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems* 2008; **24**: 035020 (14 pages).

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.