

Metadata of the chapter that will be visualized in SpringerLink

Book Title	Latent Variable Analysis and Signal Separation	
Series Title		
Chapter Title	Joint Nonnegative Matrix Factorization for Underdetermined Blind Source Separation in Nonlinear Mixtures	
Copyright Year	2018	
Copyright HolderName	Springer International Publishing AG, part of Springer Nature	
Corresponding Author	Family Name	Kopriva
	Particle	
	Given Name	Ivica
	Prefix	
	Suffix	
	Role	
	Division	Division of Electronics
	Organization	Ruder Bošković Institute
	Address	Bijenička cesta 54, 10000, Zagreb, Croatia
	Email	Ivica.Kopriva@irb.hr
Abstract	<p>An approach is proposed for underdetermined blind separation of nonnegative dependent (overlapped) sources from their nonlinear mixtures. The method performs empirical kernel maps based mappings of original data matrix onto reproducible kernel Hilbert spaces (RKHSs). Provided that sources comply with probabilistic model that is sparse in support and amplitude nonlinear underdetermined mixture model in the input space becomes overdetermined linear mixture model in RKHS comprised of original sources and their mostly second-order monomials. It is assumed that linear mixture models in different RKHSs share the same representation, i.e. the matrix of sources. Thus, we propose novel sparseness regularized joint nonnegative matrix factorization method to separate sources shared across different RKHSs. The method is validated comparatively on numerical problem related to extraction of eight overlapped sources from three nonlinear mixtures.</p>	
Keywords (separated by '-')	Underdetermined blind source separation - Nonlinear mixtures - Empirical kernel map - Joint nonnegative matrix factorization - Sparseness	



Joint Nonnegative Matrix Factorization for Underdetermined Blind Source Separation in Nonlinear Mixtures

Ivica Kopriva^(✉)

Division of Electronics, Ruder Bošković Institute, Bijenička cesta 54,
10000 Zagreb, Croatia
Ivica.Kopriva@irb.hr

Abstract. An approach is proposed for underdetermined blind separation of nonnegative dependent (overlapped) sources from their nonlinear mixtures. The method performs empirical kernel maps based mappings of original data matrix onto reproducible kernel Hilbert spaces (RKHSs). Provided that sources comply with probabilistic model that is sparse in support and amplitude nonlinear underdetermined mixture model in the input space becomes overdetermined linear mixture model in RKHS comprised of original sources and their mostly second-order monomials. It is assumed that linear mixture models in different RKHSs share the same representation, i.e. the matrix of sources. Thus, we propose novel sparseness regularized joint nonnegative matrix factorization method to separate sources shared across different RKHSs. The method is validated comparatively on numerical problem related to extraction of eight overlapped sources from three nonlinear mixtures.

Keywords: Underdetermined blind source separation · Nonlinear mixtures
Empirical kernel map · Joint nonnegative matrix factorization · Sparseness

1 Introduction

Blind source separation (BSS) refers to extraction of source signals from observed mixture signals only [1]. When the sources and mixing matrix are nonnegative algorithms of nonnegative matrix factorization (NMF) are shown to be effective solving the BSS problem [2–4]. In particular, when nonnegativity is combined with sparseness underdetermined BSS problems, characterized with more sources than mixtures available, can be solved [5, 6]. That, as an example, is relevant to mass spectrometry [7] or nuclear magnetic resonance (NMR) spectroscopy [8] based metabolic profiling where large number of sources (a.k.a. pure components or analytes) needs to be separated from the small number of available mixture spectra [7]. However, in a large number of cases algorithms for BSS address separation of sources from their linear mixtures. As opposed to them the number of methods that address the nonlinear BSS problem is rather small, see chapter 14 in [1]. Thus, we propose a method for separation of nonnegative mutually dependent (overlapped) but individually independent and identically distributed (i.i.d.) sources from smaller number of their nonlinear mixtures. Compared with proposed

method existing methods either: (i) address determined case, where the number of sources equals the number of mixtures [9–18]; (ii) do not take into account nonnegativity constraint [9–21]; (iii) assume that sources [10–12, 15–17, 19–22] or their derivatives [18] are statistically independent or that sources are individually correlated [16, 19–21]. In particular, we map data matrix from the input space onto reproducible kernel Hilbert spaces (RKHSs) by means of empirical kernel maps (EKM) [23]. We treat mapped data as they are coming from different views and propose a linear mixture model (LMM)-based representation such that all models have different mixing matrices but share the same source (representation) matrix. Thus, we propose an algorithm for joint NMF such that LMMs of mixture data mapped in multiple RKHSs share the same source matrix. That is different from joint NMF approach to multi-view clustering [24], where LMMs comprised of view dependent mixing and source matrices are assumed such that source matrices are forced to converge towards common consensus. We introduce nonlinear BSS problem in Sect. 2. Section 3 presents new joint NMF-based approach to nonlinear underdetermined BSS problem. Results of comparative performance analysis on numerical problem are presented in Sect. 4. Section 5 concludes the paper.

2 Problem Formulation

Nonlinear BSS problem with nonnegative dependent sources is formulated as:

$$\mathbf{X} = \mathbf{f}(\mathbf{S}) + \mathbf{E} \quad (1)$$

where $\mathbf{X} \in \mathbb{R}_{0+}^{N \times T}$ stands for nonnegative matrix of N nonlinear mixtures at T observations, $\mathbf{S} \in \mathbb{R}_{0+}^{M \times T}$ stands for matrix of M unknown nonnegative sources, $\mathbf{f} : \mathbb{R}_{0+}^M \rightarrow \mathbb{R}_{0+}^N$ stands for unknown nonlinear mapping $\mathbf{f} := [f_1(\mathbf{S}) \dots f_N(\mathbf{S})]^T$ acting observation-wise, $\mathbf{E} \in \mathbb{R}_{0+}^{N \times T}$ stands for an error term and \mathbb{R}_{0+} stands for the set of real nonnegative numbers.

The symbol “:=” means “by definition”. We also assume that $\left\{ \|\mathbf{s}_t\|_0 \leq K \right\}_{t=1}^T$, where $\|\mathbf{s}_t\|_0$ is indicator function that counts number of non-zero entries of \mathbf{s}_t , and K denotes maximal number of sources that can be present (active) at any observation coordinate t . The nonlinear BSS problem implies that sources $\{\mathbf{s}_m \in \mathbb{R}_{0+}^{1 \times T}\}_{m=1}^M$ have to be inferred from the mixture data matrix \mathbf{X} only. Herein, we impose assumptions on the sources and nonlinear mixture model (1):

- (A1) $0 \leq s_{mt} \leq 1 \forall m = 1, \dots, M$ and $\forall t = 1, \dots, T$,
- (A2) Amplitude s_{mt} is i.i.d. random variable that obeys exponential distribution on $(0, 1]$ interval and discrete distribution at zero, see Eq. (2),
- (A3) Components of the vector-valued function $\mathbf{f}(\mathbf{S}) : \mathbb{R}_{0+}^{M \times T} \rightarrow \mathbb{R}_{0+}^{1 \times T}$ are differentiable up to second-order,
- (A4) $M > N$.

Assumptions A1 to A4 are shown in [7] to be relevant for separation of pure components from nonlinear mixtures of mass spectra. They are expected to hold for

separation of pure components from amplitude NMR spectra as well [8]. To be useful solution of any BSS problem is expected to be essentially unique [1]. However, even for linear underdetermined BSS problem hard (sparseness) constraints ought to be imposed on sources [7, 25] to obtain essentially unique solution. The quality of separation heavily depends on degree of sparseness, i.e. the value of K . To make nonlinear underdetermined BSS problem tractable we assume, as in [27], that amplitudes of the source signals comply with sparse probabilistic model [25, 26]:

$$p(s_{mt}) = \rho_m \delta(s_{mt}) + (1 - \rho_m) \delta^*(s_{mt}) g(s_{mt}) \quad \forall m = 1, \dots, M \text{ and } \forall t = 1, \dots, T \quad (2)$$

where $\delta(s_{mt})$ is an indicator function and $\delta^*(s_{mt}) = 1 - \delta(s_{mt})$ is its complementary function, $\rho_m = \{P(s_{mt} = 0)\}_{t=1}^T$. Thus, $\{P(s_{mt} > 0)\} = 1 - \rho_m\}_{t=1}^T$. The nonzero state of s_{mt} is distributed according to probability density function $g(s_{mt})$. Exponential distribution $g(s_{mt}) = (1/\mu_m) \exp(-s_{mt}/\mu_m)$ is selected in which case the most probable outcome is equal to μ_m . To emphasize practical relevance of probabilistic model (2) we point out [7]. Another modality to which model (2) can be relevant is NMR spectroscopy [8]. It has been verified in [7] that mass spectra of pure components obey (2) with exponential distribution selected for $g(s_{mt})$. Thereby $\hat{\rho}_m \in [0.27, 0.74]$ and $\hat{\mu}_m \in [0.0012, 0.0014]$. Under such priors the nonlinear mixture model (1) simplifies to [27]:

$$\mathbf{X} = \mathbf{J}\mathbf{S} + \frac{1}{2}\mathbf{H}_{(1)} \begin{bmatrix} \mathbf{s}_1^2 \\ \vdots \\ \mathbf{s}_M^2 \\ \vdots \\ \{\mathbf{s}_i \mathbf{s}_j\}_{i,j=1}^M \end{bmatrix} + HOT = \mathbf{B} \begin{bmatrix} \mathbf{S} \\ \mathbf{s}_1^2 \\ \vdots \\ \mathbf{s}_M^2 \\ \vdots \\ \{\mathbf{s}_i \mathbf{s}_j\}_{i,j=1}^M \end{bmatrix} + HOT \quad (3)$$

where \mathbf{J} stands for Jacobian matrix, $\mathbf{H}_{(1)}$ stands for mode-1 unfolded third-order Hessian tensor, $\mathbf{B} = [\mathbf{J} \frac{1}{2} \mathbf{H}_{(1)}]$ stands for the overall mixing matrix and HOT stands for higher order terms. Since original nonlinear problem (1) is underdetermined the equivalent linear problem (3) is even more underdetermined because it is comprised of the same number of mixtures, N , but of the $P = 2M + M(M - 1)/2$ dependent sources. When degree of the overlap of the sources in (1) is K degree of the overlap of new sources in (3) is $Q \approx 2K + K(K - 1)/2$. Uniqueness of the solution of (3) depends on the triplet (N, P, Q) . For deterministic mixing matrix \mathbf{B} the necessary condition for uniqueness is $N = O(Q^2)$ [28]. Thus, it becomes virtually impossible to obtain an essentially unique solution of the underdetermined nonlinear BSS problem (1) with overlapped sources. Separation quality can, however, be increased through nonlinear mapping of mixture data $\{\mathbf{x}_t \in \mathbb{R}_{0+}^{N \times 1} \rightarrow \phi(\mathbf{x}_t) \in \mathbb{R}_{0+}^{\bar{N} \times 1}\}_{t=1}^T$ where explicit feature map (EFM) $\phi(\mathbf{x}_t)$ maps data into, in principle, infinite dimensional feature space. To make calculations in mapped space computationally tractable, $\phi(\mathbf{X}) := \{\phi(\mathbf{x}_t)\}_{t=1}^T$ needs to be projected to a low-dimensional subspace of induced space spanned by $\phi(\mathbf{V}) := \{\phi(\mathbf{v}_d)\}_{d=1}^D$. Thereby, the basis $\mathbf{V} := \{\mathbf{v}_d \in \mathbb{R}_{0+}^{N \times 1}\}_{d=1}^D$ spans the input space: $span\{\mathbf{v}_d\}_{d=1}^D \approx span\{\mathbf{x}_t\}_{t=1}^T$ and it is estimated from \mathbf{X} by k -means

clustering algorithm. Projection known as EKM, see Definition 2.15 in [23], maps data from input space onto RKHS:

$$\Psi(\mathbf{V}, \mathbf{X}) = \phi(\mathbf{V})^T \phi(\mathbf{X}) = \mathbf{K}(\mathbf{V}, \mathbf{X}) \quad (4)$$

where $\mathbf{K}(\mathbf{V}, \mathbf{X}) \in \mathbb{R}_{0+}^{D \times T}$ denotes Gram or kernel matrix with the elements $\{\kappa(\mathbf{v}_d, \mathbf{x}_t) = \phi(\mathbf{v}_d)^T \phi(\mathbf{x}_t)\}_{d,t=1}^{D,T}$. It is shown in [7] that under sparse probabilistic prior (2) Eq. (4) becomes:

$$\Psi(\mathbf{V}, \mathbf{X}) \approx \mathbf{A} \begin{bmatrix} \mathbf{0}_{1 \times T} \\ \mathbf{S} \\ \{\mathbf{s}_i \mathbf{s}_j\}_{i,j=1}^M \end{bmatrix} + \bar{\mathbf{E}} \quad (5)$$

where \mathbf{A} denotes a nonnegative mixing matrix of appropriate dimensions, $\mathbf{0}_{1 \times T}$ stands for row vector of zeros and $\bar{\mathbf{E}}$ stands for approximation error. The uniqueness condition for system (5) becomes: $D = O(Q^2)$, [28]. When $D \gg N$ that can be fulfilled with greater probability than uniqueness condition for system (3): $N = O(Q^2)$, [7, 27]. Thus, the role of nonlinear EKM-based mapping is to “increase number of mixtures”.

3 Joint Nonnegative Matrix Factorization in Reproducible Kernel Hilbert Spaces

It has been demonstrated that sparseness constrained NMF in an EKM-induced RKHS enables separation of nonnegative dependent sources from smaller number of their nonlinear mixtures [27]. However, the fundamental issue is how to select the kernel function, i.e. its parameters in (4)/(5). The common choice is a Gaussian kernel $\kappa(\mathbf{v}_d, \mathbf{x}_t) = \exp\left(-\|\mathbf{v}_d - \mathbf{x}_t\|^2 / \sigma^2\right)$. That is justified by its universal approximation property [29]. However, proper selection of the kernel variance σ^2 requires a priori knowledge of the signal-to-noise (SNR) ratio. When dealing with experimental data that is often hard to know in practice. Herein, we propose to map data \mathbf{X} into multiple RKHSs using EKMs with Gaussian kernel with the values for variance that cover wide enough range: $\sigma^2 \in \{\sigma_1^2, \dots, \sigma_{n_v}^2\}$. Hence, we obtain n_v data matrices in induced RKHSs with representations as follow:

$$\Psi_i(\mathbf{V}, \mathbf{X}) = \mathbf{A}_i \bar{\mathbf{S}} + \bar{\mathbf{E}}_i \quad i = 1, \dots, n_v \quad (6)$$

where meaning of $\bar{\mathbf{S}}$ is clear from direct comparison between (6) and (5). To establish weak analogy with the multi-view clustering, [24], we denoted mixture matrices in RKHSs as data arising from multiple views. Also, without loss of generality, to enable fair comparison with multi-view NMF algorithm [24] we assume that mixture matrices in each “view i ” satisfy $\{\|\Psi_i(\mathbf{V}, \mathbf{X})\|_1 = 1\}_{i=1}^{n_v}$. The difference between our model (6) and multi-view NMF model [24] is that our model (6) assumes that all the views share the same source matrix $\bar{\mathbf{S}}$, while in [25] source matrices are different for each view and

are enforced to converge towards a common consensus. To derive the NMF update rule on the level of “view i ” we assume Gaussian distribution for the error term in (6) and minimize the loss function under constrains $\mathbf{A}_i \geq \mathbf{0}$, $\bar{\mathbf{S}} \geq \mathbf{0}$:

$$L(\mathbf{A}_i, \bar{\mathbf{S}}) = \frac{1}{2} \|\Psi_i(\mathbf{V}, \mathbf{X}) - \mathbf{A}_i \bar{\mathbf{S}}\|_2^2 + \alpha \|\bar{\mathbf{S}}\|_1 \quad (7)$$

where α stands for sparseness regularization constant. Minimization yields the following update rules for \mathbf{A}_i and $\bar{\mathbf{S}}$, see also Table 1 in [3]:

$$\begin{aligned} \mathbf{A}_i &= \mathbf{A}_i \otimes \frac{\Psi_i(\mathbf{V}, \mathbf{X}) \bar{\mathbf{S}}^T}{\mathbf{A}_i \bar{\mathbf{S}} \bar{\mathbf{S}}^T + \varepsilon \mathbf{1}_{DP}} \\ \bar{\mathbf{S}} &= \bar{\mathbf{S}} \otimes \frac{[\mathbf{A}_i^T \Psi_i(\mathbf{V}, \mathbf{X}) - \alpha \mathbf{1}_{PT}]_+}{\mathbf{A}_i^T \mathbf{A}_i \bar{\mathbf{S}} + \varepsilon \mathbf{1}_{PT}} \end{aligned} \quad (8)$$

In (8) \otimes denotes entry-wise multiplication, $\mathbf{1}_{DP}$ and $\mathbf{1}_{PT}$ stand for matrices of all ones, ε is a small constant and $[x]_+$ stands for $\max(0, x)$ operator. At each iteration the algorithm cycles through all the views $1, \dots, n_v$. It is clear that representation (6) automatically resolves the permutation indeterminacy issue that is problematic for joint NMF across multiple views [24]. We coin our method multi-view NMF (mvNMF). The joint NMF method [24] is coined multi-view consensus NMF (mvCNMF). Even though our method is developed for separation of sources from nonlinear underdetermined mixtures it can be applied directly to multi-view clustering in the same spirit as joint NMF method in [24]. In that case $\Psi_i(\mathbf{V}, \mathbf{X})$ ought to be replaced with the data matrix at view i : \mathbf{X}_i . Furthermore, when BSS problem is linear, $\mathbf{X} = \mathbf{A}\mathbf{S}$, with one view only, i.e. $n_v = 1$, Eq. (8) with the appropriate substitutions represents standard sparseness constrained NMF [3].

4 Numerical Evaluation

To validate proposed mvNMF method we generate three nonlinear mixtures of eight overlapped sources according to:

$$\begin{aligned} f_1(\mathbf{s}) &= s_1^3 + s_2^2 + \tan^{-1}(s_3) + s_4^2 + s_5^3 + s_6^3 + \tanh(s_7) + \sin(s_8) + \mathbf{e}_1 \\ f_2(\mathbf{s}) &= \tanh(s_1) + s_2^3 + s_3^3 + \tan^{-1}(s_4) + \tanh(s_5) + \sin(s_6) + s_7^2 + s_8^2 + \mathbf{e}_2 \\ f_3(\mathbf{s}) &= \sin(s_1) + \tan^{-1}(s_2) + s_3^2 + s_4^3 + \tanh(s_5) + \sin(s_6) + s_7^3 + \tan^{-1}(s_8) + \mathbf{e}_3 \end{aligned}$$

We generated eight source signals in $T = 1000$ observations with degrees of overlap equal to $K \in \{1, 3, 5\}$. According to probabilistic prior (2) we set $\{\rho_m = 0.6\}_{m=1}^8$ and $\{\mu_m = 0.15\}_{m=1}^8$. Thus, generated sources correspond with the real world signals such as mass spectra. Furthermore, noise was added to the mixtures with SNR = 0 dB. We use the Gaussian kernel with the variance $\sigma^2 \in \{1000, 100, 10, 1, 0.1, 0.01, 0.001\}$. That covers wide range of possible SNRs. The basis matrix \mathbf{V} for

EKM-based mappings was estimated from \mathbf{X} by *k-means* algorithm with $D = 100$ cluster centers. We compare the proposed mvNMF algorithm with the mvCNMF algorithm [24], with ordinary NMF algorithm [3] applied directly to mixture data matrix \mathbf{X} and with algorithm (8) applied to each “view”, mapped data matrix $\{\Psi_i(\mathbf{V}, \mathbf{X})\}_{i=1}^{n_v}$, separately. We coined the last algorithm as single view NMF (svNMF) and point out that it coincides with the algorithm [27] applied in each RKHS separately. We set sparseness related regularization constant in (8) to $\alpha = 0.2$. In case of mvCNMF algorithm we use the result for consensus matrix to be compared with the result of mvNMF. For each value of K we repeated the comparison 100 times. In each experiment we separated eight sources from the mixtures and annotated them with the true sources using mean normalized correlation as criterion:

$$\text{mean correlation} = \left(\sum_{i \in I_c} c_i(\hat{s}_i, s_i) \right) / M \quad (9)$$

where I_c denotes index set of correctly assigned sources, \hat{s}_i denotes the separated and s_i the true source and $0 \leq c_i(\hat{s}_i, s_i) \leq 1$ stands for the normalized correlation coefficient. Thus, if more than one separated source was assigned to the same true source that was counted as assignment error and reduced value of the mean correlation.

Figure 1a shows mean values of assignment errors (with the variance as error bar) for NMF, mvNMF and mvCNMF algorithms. Figures 1b shows assignment errors for the svNMF algorithm. Corresponding correlation coefficients (9) are shown in Fig. 2a and b. The largest mean value of assignment error is 36% for mvNMF, 34.63% for NMF, 45.37% for mvCNMF and 35.37% for svNMF. The largest values of mean correlation coefficient for the algorithms in respective order are 8.12%, 4.23%, 7.44% and 4.36%. Thus, proposed mvNMF method increased correlation coefficient in comparison with NMF and svNMF methods having similar assignment error. In comparison with mvCNMF the mvNMF method has similar correlation coefficient but smaller assignment error. The mvCNMF method extracted typically two or three unique sources with the “highest” value of correlation coefficient. That explains its

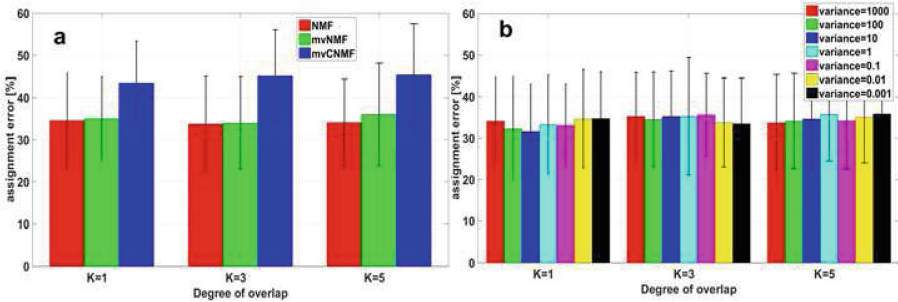


Fig. 1. Assignment error. (a) NMF, mvNMF and mvCNMF algorithms. (b) svNMF algorithm, i.e. NMF algorithm applied to each “view” separately. (Color figure online)

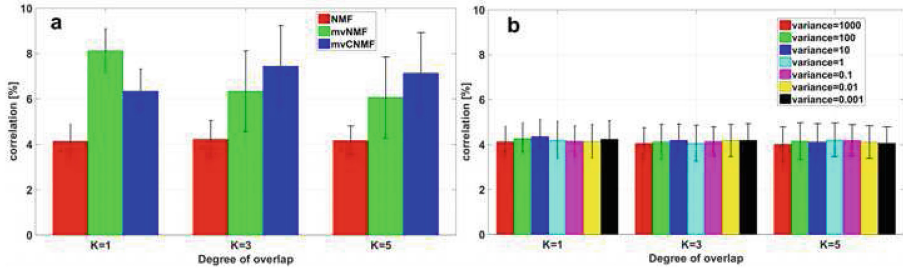


Fig. 2. Mean correlation coefficients (9) of correctly assigned sources separated with: (a) NMF, mvNMF and mvCNMF algorithms. (b) svNMF algorithm, i.e. NMF algorithm applied to each “view” separately. (Color figure online)

“good” performance in terms of correlation and poor in terms of assignment error. Figures 1b and 2b show that performing source separation in each induced RKHS separately yields results worse than when all the RKHSs are used. Although the separation quality of proposed mvNMF method could be considered low we comment that described nonlinear BSS problem is hard and for it, to the best of our knowledge, no method is developed yet.

5 Conclusion

Blind separation of nonnegative dependent (overlapped) sources from smaller number of nonlinear mixtures represents a hard problem with, arguably, no algorithm proposed to solve it. Herein, we propose method for separation of sparse dependent sources by joint NMF on mixture matrices mapped in multiple RKHSs. RKHSs were induced by mappings based on Gaussian kernel with variances that cover a wide range of possible SNR values. Mixtures in induced RKHSs were represented with the linear mixture models comprised of different mixing matrices and common matrix of sources. That is justified by the fact that mixtures in mapped data space are obtained from the same mixture matrix in input data space. Thus, a novel joint NMF method is proposed to separate common source matrix from multiple mixtures. On numerical experiment the proposed method achieved competitive performance. In addition for nonlinear BSS proposed joint NMF method could be also used for clustering data from multiple views in the spirit of [24].

Acknowledgments. This work has been supported in part by the Grant IP-2016-06-5253 funded by Croatian Science Foundation and in part by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS).

References

1. Comon, P., Jutten, C. (eds.): Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press, New York, NY, USA (2010)
2. Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.I.: Nonnegative Matrix and Tensor Factorizations. Wiley, Chichester, UK (2009)
3. Cichocki, A., Zdunek, R., Amari, S.: Csiszár's divergences for non-negative matrix factorization: family of new algorithms. In: Rosca, J., Erdogmus, D., Principe, J.C., Haykin, S. (eds.) ICA 2006. LNCS, vol. 3889, pp. 32–39. Springer, Heidelberg (2006). https://doi.org/10.1007/11679363_5
4. Lee, D.D., Seung, H.S.: Learning the parts of objects by nonnegative matrix factorization. *Nature* **401**, 788–791 (1999)
5. Peharz, R., Pernkopf, F.: Sparse nonnegative matrix factorization with ℓ^0 -constraints. *Neurocomputing*, **80**, 38–46 (2012)
6. Gillis, N., Glineur, F.: Using underapproximations for sparse nonnegative matrix factorization. *Pattern Rec.* **43**, 1676–1687 (2010)
7. Kopriva, I., Jerić, I., Brkljačić, L.: Nonlinear mixture-wise expansion approach to underdetermined blind separation of nonnegative dependent sources. *J. of Chemometr.* **27**, 189–197 (2013)
8. Kopriva, I., Jerić, I.: Blind separation of analytes in nuclear magnetic resonance spectroscopy: improved model for nonnegative matrix factorization. *Chemometr. Int. Lab. Syst.* **137**, 47–56 (2014)
9. Zhang, K., Chan, L.: Minimal nonlinear distortion principle for nonlinear independent component analysis. *J. Mach. Learn. Res.* **9**, 2455–2487 (2008)
10. Levin, D.N.: Using state space differential geometry for nonlinear blind source separation. *J. Appl. Phys.* **103**(044906), 1–12 (2008)
11. Levin, D.N.: Performing nonlinear blind source separation with signal invariants. *IEEE Trans. Sig. Proc.* **58**, 2132–2140 (2010)
12. Taleb, A., Jutten, C.: Source separation in post-nonlinear mixtures. *IEEE Trans. Sig. Proc.* **47**, 2807–2820 (1999)
13. Duarte, L.T., Suyama, R., Rivet, B., Attux, R., Romano, J.M.T., Jutten, C.: Blind compensation of nonlinear distortions: applications to source separation of post-nonlinear mixtures. *IEEE Trans. Sig. Proc.* **60**, 5832–5844 (2012)
14. Filho, E.F.S., de Seixas, J.M., Calôba, L.P.: Modified post-nonlinear ICA model for online neural discrimination. *Neurocomputing* **73**, 2820–2828 (2010)
15. Nguyen, V.T., Patra, J.C., Das, A.: A post nonlinear geometric algorithm for independent component analysis. *Digit. Sig. Proc.* **15**, 276–294 (2005)
16. Ziehe, A., Kawanabe, M., Harmeling, S., Müller, K.R.: Blind separation of post-nonlinear mixtures using gaussianizing transformations and temporal decorrelation. *J. Mach. Learn. Res.* **4**, 1319–1338 (2003)
17. Zhang, K., Chan, L.W.: Extended gaussianization method for blind separation of post-nonlinear mixtures. *Neural Comput.* **17**, 425–452 (2005)
18. Ehsandoust, B., Babaie-Zadeh, M., Rivet, B., Jutten, C.: Blind source separation in nonlinear mixtures: separability and a basic algorithm. *IEEE Trans. Sig. Proc.* **65**, 4352–4399 (2017)
19. Harmeling, S., Ziehe, A., Kawanabe, M.: Kernel-based nonlinear blind source separation. *Neural Comput.* **15**, 1089–1124 (2003)
20. Martinez, D., Bray, A.: Nonlinear blind source separation using kernels. *IEEE Trans. Neural Net.* **14**, 228–235 (2003)

21. Yu, H.-G., Huang, G.-M., Gao, J.: Nonlinear blind source separation using kernel multi-set canonical correlation analysis. *Int. J. Comput. Netw. Inf. Secur.* **1**, 1–8 (2010)
22. Almeida, L.: MISEP-linear and nonlinear ICA based on mutual information. *J. Mach. Learn. Res.* **4**, 1297–1318 (2003)
23. Schölkopf, B., Smola, A.: *Learning With Kernels*. The MIT Press, Cambridge, MA, USA (2002)
24. Liu, L., Wang, C., Gao, J., Han, J.: Multi-view clustering via joint nonnegative matrix factorization. In: 2013 Proceedings of SIAM International Conference on Data Mining (SDM 2013), pp. 252–260 (2013). <https://doi.org/10.1137/1.9781611972832.28>
25. Caifa, C., Cichocki, A.: Estimation of sparse nonnegative sources from noisy overcomplete mixtures using MAP. *Neural Comput.* **21**, 3487–3518 (2009)
26. Bouthemy, P., Piriou, C.H.G., Yao, J.: Mixed-state auto-models and motion texture modeling. *J. Math Imag. Vis.* **25**, 387–402 (2006)
27. 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. of Chemometr.* **28**, 704–715 (2014)
28. DeVore, R.A.: Deterministic constructions of compressed sensing matrices. *J. Complex.* **27**, 918–925 (2007)
29. Micchelli, C.A., Xu, Y., Zhang, H.: Universal kernels. *J. Mach. Learn. Res.* **7**, 2651–2667 (2006)