

Hamburger Beiträge

zur Angewandten Mathematik

On the Construction of Non-Negative Dimensionality Reduction Methods

Sara Krause-Solberg, Armin Iske

Nr. 2016-04
March 2016

On the Construction of Non-Negative Dimensionality Reduction Methods

Sara Krause-Solberg

Faculty of Mathematics, Munich University of Technology
Boltzmannstraße 3, 85748 Garching, Germany
sara.krause-solberg@ma.tum.de

Armin Iske

Department of Mathematics, University of Hamburg
Bundesstraße 55, 20146 Hamburg, Germany
armin.iske@uni-hamburg.de

Abstract

Many relevant applications of signal processing rely on the separation of sources from a mixture of signals without prior knowledge about the mixing process. Given a mixture of signals, the task of signal separation is to estimate the mixture components by using specific assumptions on their time-frequency behaviour or statistical characteristics. Time-frequency data is often very high-dimensional, which reduces the performance of signal separation methods quite significantly. Therefore, the embedding dimension of the signal's time-frequency representation should be reduced prior to the application of a decomposition strategy, such as independent component analysis (ICA) or non-negative matrix factorization (NNMF). In other words, a suitable dimensionality reduction method should be applied, before the data is decomposed and then back-projected. But the choice of the dimensionality reduction method requires particular care, especially in combination with NNMF and certain types of ICA, since they require non-negative input data. In this paper, we introduce a generic concept for the construction of suitable non-negative dimensionality reduction methods. Furthermore, we discuss the two different decomposition strategies ICA and NNMF for single channel signal separation. We apply the resulting methods to the separation of acoustic signals with transitory components.

Key words and phrases: Non-negative dimensionality reduction, signal separation, independent component analysis, non-negative matrix factorization

1 Introduction

In relevant applications of signal processing there is an increasing demand for efficient signal separation methods. Given a mixture of signals, the basic task of signal separation is the estimation of the signal's components without prior knowledge about the mixing process. Different decomposition techniques were recently developed to do so, e.g. *independent subspace analysis* (ISA) [2, 4, 5, 19], based on *independent component analysis* (ICA) or *non-negative matrix factorization* (NNMF) [6, 18, 20]. On the down side, the computational complexity of these decomposition methods may be prohibitively large.

To reduce the computational complexity, dimensionality reduction methods can be employed. Many dimensionality reduction methods rely on projections, e.g. the linear *principal component analysis* (PCA) and *multidimensional scaling* (MDS). More recently, *nonlinear* dimensionality reduction methods (see [14] for an overview) were developed, e.g. *isomap* and *local linear embedding* (LLE). In [8], dimensionality reduction is combined with ISA for audio signal separation.

In signal separation, dimensionality reduction is first applied to reduce the dimension of the data obtained from a time-frequency transform (e.g. STFT), before the reduced data is decomposed into different components, each assigned to one of the sources. In combination with decomposition methods as NNMF, *non-negative* dimensionality reduction methods are essentially required to output non-negative data from non-negative input data. Therefore, the construction of *non-negative dimensionality reduction* (NNDR) methods is of particular interest.

In this paper we develop a novel concept for the construction of non-negative dimensionality reduction methods for signal separation. We remark that dimensionality reduction for signal separation is studied in previous work [5, 8, 19], but without using any non-negativity constraints. Therefore, the idea to apply NNDR methods in the context of signal separation is new.

For the construction of NNDR methods we characterize the dimensionality reduction as an optimization problem with a suitable cost functional. To generate a NNDR method, we include a non-negativity constraint, which leads us to a constrained optimization problem. To solve this problem, we consider splitting the problem of NNDR into two subproblems that can (under suitable assumptions) be decoupled: (a) the problem of dimensionality reduction without non-negativity constraints; (b) the construction of a rotation for mapping the reduced data to the positive orthant. Details on problems (a) and (b) and their solutions are explained later in this paper. Moreover, for the purpose of illustration, we compare the two decomposition methods ICA and NNMF in combination with a suitable NNDR method, whose construction relies on PCA.

Our concept of NNDR is explained in Section 2, and its coupling with NNMF and ICA is subject of Section 3. Numerical examples and comparisons between the decomposition methods ICA and NNMF are finally presented in Section 4.

2 Non-Negative Dimensionality Reduction

On given data

$$X = \{x_i\}_{i=1}^n \subset \mathbb{R}^D,$$

and for large dimension D , dimensionality reduction aims to find a representation Y of X , so that Y can be embedded in a Euclidean space of much smaller dimension $d \ll D$, i.e., $Y \subset \mathbb{R}^d$. For convenience, we will from now use the matrix notation

$$X = (x_1, \dots, x_n) \in \mathbb{R}^{D \times n}.$$

More precisely, with assuming the data X to lie on a (smooth) manifold $\mathcal{M} \subset \mathbb{R}^D$, we wish to compute a low-dimensional representation Ω of \mathcal{M} , where $\Omega \subset \mathbb{R}^d$ is embedded in \mathbb{R}^d , for $d \ll D$. This way we obtain $Y \subset \Omega$ representing $X \subset \mathcal{M}$, as illustrated in the following diagram.

$$\begin{array}{ccc} X \subset \mathcal{M} \subset \mathbb{R}^D & & \\ \downarrow P & \downarrow B & \\ Y \subset \Omega \subset \mathbb{R}^d & & \end{array} \quad (1)$$

In this paper, we work with dimensionality reduction methods that can be characterized as an optimization problem

$$\min_{P \in \mathcal{U}} g(P) \quad (2)$$

with $\mathcal{U} \subset \{f: X \rightarrow \mathbb{R}^d\}$ and a cost functional $g: \mathcal{U} \rightarrow \mathbb{R}$, where g is a measure for the *distance* between the projection P and the homeomorphism B in (1). Note that the pair (\mathcal{U}, g) entirely determines the dimensionality reduction method. We remark that most dimensionality reduction methods, including PCA, MDS, isomap and LLE, can be formulated as (2) and, moreover, efficient algorithms for solving their corresponding optimization problem (2) are available (see [14]).

In many relevant applications, dimensionality reduction is applied in a pre-processing step to reduce the data complexity for making subsequent applications of signal analysis tools feasible. Quite often, subsequent operations essentially require *non-negative* input data. This motivates us to construct *non-negative* dimensionality reduction (NNDR) methods. To this end, it is straightforward to add a non-negativity constraint $P(X) \geq 0$ to (2), and so we obtain a constrained optimization problem

$$\min_{\substack{P \in \mathcal{U} \\ P(X) \geq 0}} g(P). \quad (3)$$

Note that the non-negativity constraint makes the optimization problem (3) rather difficult to solve. This is in contrast to the unconstrained optimization problem (2), for whose solution standard algorithms are readily available [14].

For solving optimization problem (3), we consider a splitting approach, where in a first step the dimensionality reduction problem (2) is solved, before in a second step the reduced data is rotated into the positive orthant of \mathbb{R}^d . To be more precise, in the second step we compute a rotation matrix $W \in SO(d)$ satisfying $WP(X) \geq 0$, where P is a solution of (2).

We remark that the splitting approach does not work for all dimensionality reduction methods. But it turns out, that, whenever such a rotation matrix W exists, the rotational invariance of g , i.e.,

$$W \in SO(d) \quad \implies \quad WP \in \mathcal{U} \text{ and } g(WP) = g(P),$$

guarantees the solvability of problem (3) by the proposed splitting approach. We can show that both PCA and MDS fit in this framework, so that the existence of a rotation matrix W is guaranteed, if the input (high-dimensional) data is non-negative, see [12] for more details.

2.1 Non-Negative PCA (NNPCA)

Principal component analysis (PCA) is a classical linear dimensionality reduction method, whose construction relies on the singular value decomposition. Here, the mapping P is an orthogonal projection, satisfying $Y = P(X) = U^T X$, with $U \in \mathbb{R}^{D \times d}$. The projection matrix U is obtained by solving the minimization problem

$$\min_{U^T U = \mathbf{1}} \sum_{k=1}^n \|x_k - U U^T x_k\|^2. \quad (4)$$

We remark that the solution of minimization problem (4) is given by the maximizer of the variance $\text{var}(Y)$ of Y , which in turn is given by the trace of $Y Y^T$. This allows us to reformulate (4) as an equivalent maximization problem,

$$\max_{U^T U = \mathbf{1}} \text{tr}(U^T X X^T U), \quad (5)$$

where the maximizer U of $\text{var}(Y)$ is given by a matrix U whose d columns contain the eigenvectors of the d largest eigenvalues of the covariance matrix $X X^T$.

In this context, the pair (\mathcal{U}, g) in problem (2) is given by

$$\begin{aligned} \mathcal{U} &= \{U^T \in \mathbb{R}^{d \times D} \mid U^T U = \mathbf{1}\}, \\ g(U^T) &= -\text{tr}(U^T X X^T U). \end{aligned}$$

According to (3), we reformulate the maximization problem in (5) by adding a non-negativity constraint:

$$\max_{\substack{U^T U = \mathbf{1} \\ U^T X \geq 0}} \text{tr}(U^T X X^T U). \quad (6)$$

Note that this additional restriction transforms the simple PCA problem in (5) into a much more difficult non-convex optimization problem (6) with many local solutions, for which (in general) none of the solutions is known analytically.

Let us further discuss this special case of PCA in more detail. First note that the cost functional g is rotational invariant, due to the cyclic invariance of the trace. Indeed, for $W \in SO(d)$ we find the relation

$$g(WU^T) = -\text{tr}(WU^T X X^T U W^T) = -\text{tr}(W^T W U^T X X^T U) = g(U^T).$$

Furthermore, the orthogonal projection U^T solving (5) is an angle preserving mapping from \mathcal{M} to \mathbb{R}^d , i.e., data lying in the positive orthant of \mathbb{R}^D is mapped into a pointed linear cone of angle 90° in \mathbb{R}^d . Thus, the splitting approach, as described above, can be applied. Indeed, there exists a matrix $W \in SO(d)$ rotating the reduced data into the positive orthant of the low-dimensional space \mathbb{R}^d . The value of the cost functional g in (6) is not changed by this rotation.

The crucial point, however, is to compute such a rotation matrix W . In order to do so, we consider solving an auxiliary optimization problem on the set of orthogonal matrices $SO(d)$ for the cost functional

$$J(W) = \frac{1}{2} \sum_{i,j} \left[(WU^T X)_- \right]_{ij}^2, \quad (7)$$

where we let

$$[Y_-]_{ij} = \begin{cases} y_{ij} & \text{if } y_{ij} < 0, \\ 0 & \text{otherwise,} \end{cases}$$

as this was proposed in [17] in the context of ICA. Note that the cost functional J in (7) penalizes negative entries in the low-dimensional representation $WU^T X$, and it attains its minimal value zero for a suitable rotation.

However, the minimization of (7) can not be solved directly by an additive update algorithm, since the set of rotation matrices $SO(d)$ is not invariant under summation, i.e., updating with ΔW does in general not imply $W + \Delta W \in SO(d)$. There are several possibilities to overcome this difficulty. After each update, a projection to the space of orthogonal matrices $SO(d)$ as in [16] would do. But it is not clear if this would lead to a viable descent strategy. Another approach is to include a relaxation term and to consider an unconstrained optimization problem, see [22].

A more elegant way to minimize the cost functional J in (7) exploits the Lie-group structure of $SO(d)$, leading to a corresponding optimization problem in the associated Lie algebra of skew-symmetric matrices $\mathfrak{so}(d)$ (see [9] for a comprehensive account to Lie groups and Lie algebras). In the set of skew-symmetric matrices an additive update is computed and mapped to the group of special orthogonal matrices via the matrix exponential. In this way, a multiplicative update in $SO(d)$ is induced and the orthogonality of the updated matrix

is guaranteed in a rather natural way. Due to the vector space property of $\mathfrak{so}(d)$, standard methods, e.g. line search, can be applied to find the minimum [12, 17].

We finally remark that the NNPCA, as specified in [22] and used in several papers (e.g. [10, 15, 21]), essentially differs from the one introduced here, since they require $U \geq 0$ rather than $U^T X \geq 0$, which is a more restrictive constraint.

3 Application of NNDR to Signal Separation

In recent years, techniques for blind source signal separation, such as *independent component analysis* (ICA) or *independent subspace analysis* (ISA), were coupled with dimensionality reduction methods, e.g. PCA, LLE or isomap [5, 8, 19]. Moreover, *non-negative matrix factorization* (NNMF) has gained enormous interest in signal separation (see e.g. [11, 18, 20]). Often, the utilized decomposition, in particular NNMF, essentially requires non-negative input. Therefore, the construction of *non-negative* dimensionality reduction (NNDR) methods is of primary interest in signal separation.

Let us briefly introduce the basic concept of signal separation, with focussing on the interaction between dimensionality reduction and decompositions. For a band-limited signal $f \in L^2([0, T])$ and a segmentation of its domain $[0, T]$ into small patches, we can analyze f on each of these patches, as this is usually done by short-time Fourier-transform (STFT) or wavelet analysis. In this way, we obtain a dataset X of vectors in \mathbb{R}^D by drawing n samples from f :

$$X = \{x_i\}_{i=1}^n \subset \mathbb{R}^D.$$

With using the matrix notation $X = (x_1, \dots, x_n) \in \mathbb{R}^{D \times n}$, the output of the time-frequency analysis is a high-dimensional and (componentwise) non-negative matrix X , called the *spectrogram* of signal f . The columns of X represent frequencies which are present in f at a certain time. Now, a suitable dimensionality reduction method is first applied to the data matrix X , before a decomposition of the reduced data is performed. The reduced data is represented by a matrix $Y \in \mathbb{R}^{d \times n}$ of much smaller dimension, $d \ll D$. With the subsequent decomposition of Y , each extracted component of Y is assigned to one source signal. The basic steps of the overall procedure are shown in Figure 1.

Recall that we assume the data points in X to lie on a smooth manifold. The dimensionality reduction map P which maps X on Y is required to be non-negativity preserving according to (3). Furthermore, for the reconstruction of the source signals we need a suitable left-inverse of the map P , giving an approximation of B^{-1} in diagram (1). This assumption is in general very restrictive, but in the special case of PCA it is satisfied by $P^{-1}(y) = Uy$.

There are different methods for the decomposition of the (reduced) spectrogram Y . Among them, ICA and NNMF are commonly used. In either case, for

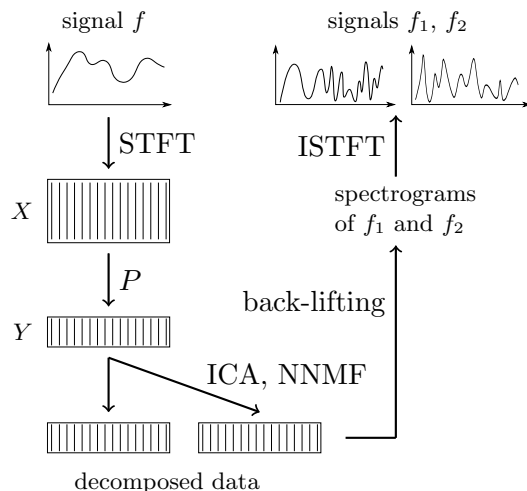


Figure 1: Signal separation with dimensionality reduction.

the application of ICA or NNMF, we assume the input data Y to be a linear mixture of source terms s_i , i.e.,

$$Y = AS, \quad (8)$$

where the matrices $A \in \mathbb{R}^{d \times d}$ and $S \in \mathbb{R}^{d \times n}$ are unknown. We assume $S \subset \Omega$ as this is the case for PCA. For the estimation of A and S we need specific additional assumptions to balance the disproportion of equations and unknowns in the factorization problem (8).

1. *Independent Component Analysis (ICA)*. The basic assumption of ICA is that the source signals are statistically independent. Furthermore, the data matrix Y is assumed to result from n realizations of a d -dimensional random vector. In order to estimate S , a random variable \mathcal{S} is constructed, whose n realizations yield the columns of the source matrix S . The components of \mathcal{S} are chosen to be as stochastically independent as possible, where the stochastic independence can be measured by the Kullback-Leibler distance (cf. [3]).

In practice, the number of sources is usually not known. In consequence, we may detect more independent components than the true number of sources. In this case, two or more of the separated components belong to the same source. Thus, the sources are combinations of the independent components. In a subsequent step, the sources are grouped (partitioned) into independent subspaces, each corresponding to one source. Finally, the sources are reconstructed from these multi-component subspaces (see [2]). This procedure is called *independent subspace analysis (ISA)*. The main difficulty of ISA is to identify components belonging to the same multi-component subspace.

2. *Non-negative Matrix Factorization (NNMF)*. Note that the factorization of the given data set Y into a mixing matrix A and the source signals (source components) S , i.e., $Y = AS$, could also be done by a matrix factorization. The data which we use for signal separation are obtained by taking the modulus of the signal's STFT, and so the input data is non-negative. Since the source components are assumed to be spectrograms, too, we assume them to be non-negative as well. Therefore, non-negative matrix factorizations (NNMF) are suitable tools for decomposition.

Different algorithms for NNMF are available, each of which relies on the non-negativity $Y, A, S \geq 0$, where different measures $d(Y, AS)$ for the reconstruction error were proposed (see e.g. [6, 18, 20]). Here, we consider using the generalized *Kullback-Leibler divergence*

$$d(Y, AS) = \sum_{i,j} Y_{ij} \log \frac{Y_{ij}}{(AS)_{ij}} - Y_{ij} + (AS)_{ij},$$

proposed by Lee & Seung [13], as used for decomposition of signal data in [20].

We summarize the steps of the resulting signal separation as follows.

Algorithm (Signal Separation by NNDR).	
input	signal f as time series
compute	spectrogram X and phase Φ of f (e.g. STFT)
reduce	dimension of X (by e.g. PCA, $Y = U^T X$)
rotate	low dimensional data set Y by $W \in SO(d)$,
	such that $WY \geq 0$, where
└	$W = \arg \min_{WY \geq 0} J(W)$ with Lie Group Method
decompose	WY by ICA (jade [1]) or NNMF (toolbox [7])
separate	$WY = Z_1 + Z_2$, see [2]
invert	dimensionality reduction ($X_1 = UW^T Z_1, X_2 = UW^T Z_2$)
reconstruct	time series of sources f_i using phase Φ (ISTFT)

4 Numerical Results

For the purpose of illustration, we present one numerical example, where we compare the decompositions by ICA and NNMF. To this end, we consider a mixture $f = f_1 + f_2$ of acoustic transient signals, where f_1 is a sequence of castanets and f_2 a cymbal signal, see Figure 2.

The combination $f = f_1 + f_2$ of the castanets and the cymbal is shown in Figure 3. Their spectrograms were generated by a STFT using a Hamm-window. Since f_2 is a high-energy signal, f has a complex frequency characteristic. Therefore, the task of extracting the castanets signal f_1 , being active only at a few time steps, is quite challenging. For the ISTFT we have used the phase information from the original signal.

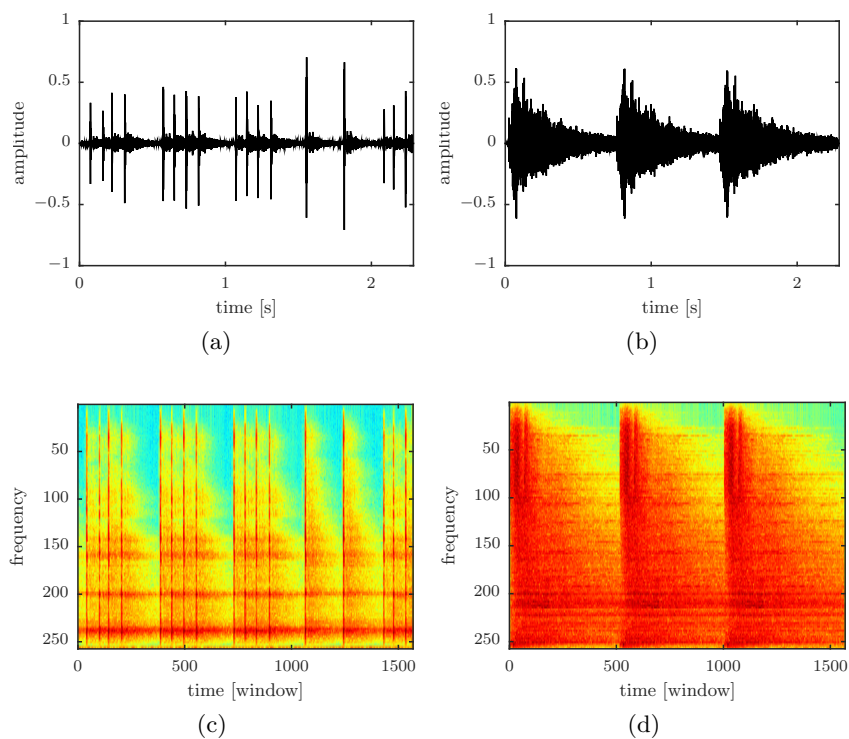


Figure 2: Two acoustic signals: (a) castanets f_1 , (b) cymbal f_2 with their spectrograms: (c) for f_1 and (d) for f_2 .

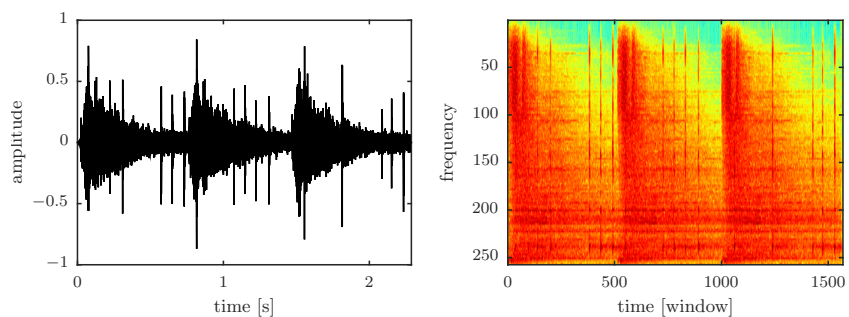


Figure 3: Signal $f = f_1 + f_2$ and its spectrogram X .

The separations obtained from NNPCA and the decomposition methods, NNMF and ICA, are shown in Figure 4. Note that either method, ICA or NNMF, reproduces the characteristic peaks of the castanets quite well. But for the separation by NNMF, a few more artifacts are visible in the cymbal signal, in contrast to the separation by ICA (see Figure 4). Merely at time steps, where

a high amplitude of the cymbal exactly matches the peaks of the castanets, a correct separation is not quite achieved. For the reconstruction of the reduced signal, either method, ICA or NNMF, yields an almost complete reconstruction of f , as their reconstructions are nearly identical (cf. Figure 5).

We finally remark that for signal separation *without* dimensionality reduction, NNMF is competitive to ICA [20]. This suggests that the separation by NNPCA-NNMF could be improved. In fact, this could be achieved by more sophisticated (*nonlinear*) dimensionality reduction methods, which, however, would lead to a much more complicated back-projection of the data. Nevertheless, although NNPCA is only a *linear* projection method, our numerical results show that its performance in combination with ICA is already quite promising.

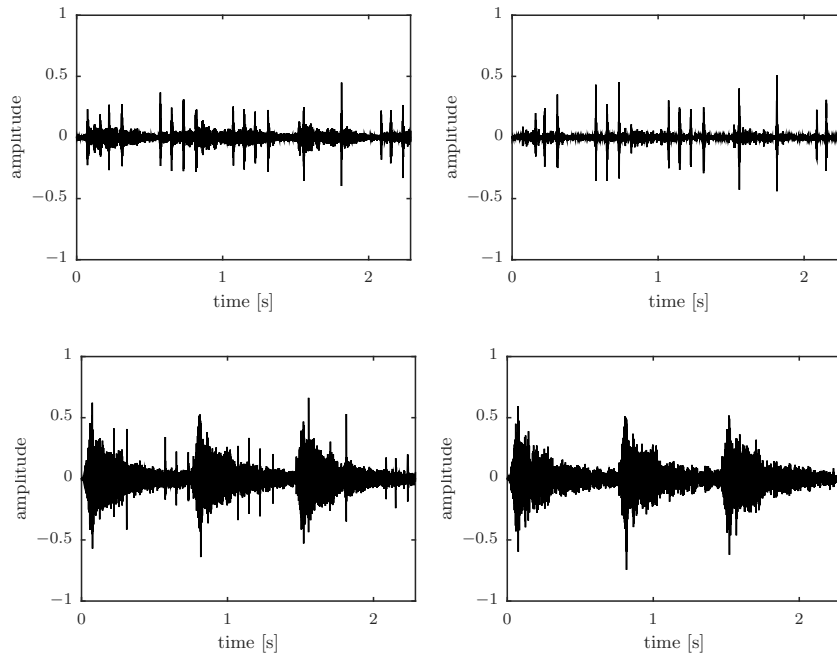


Figure 4: Signal separation by NNPCA-NNMF (left column); separation by NNPCA-ICA (right column).

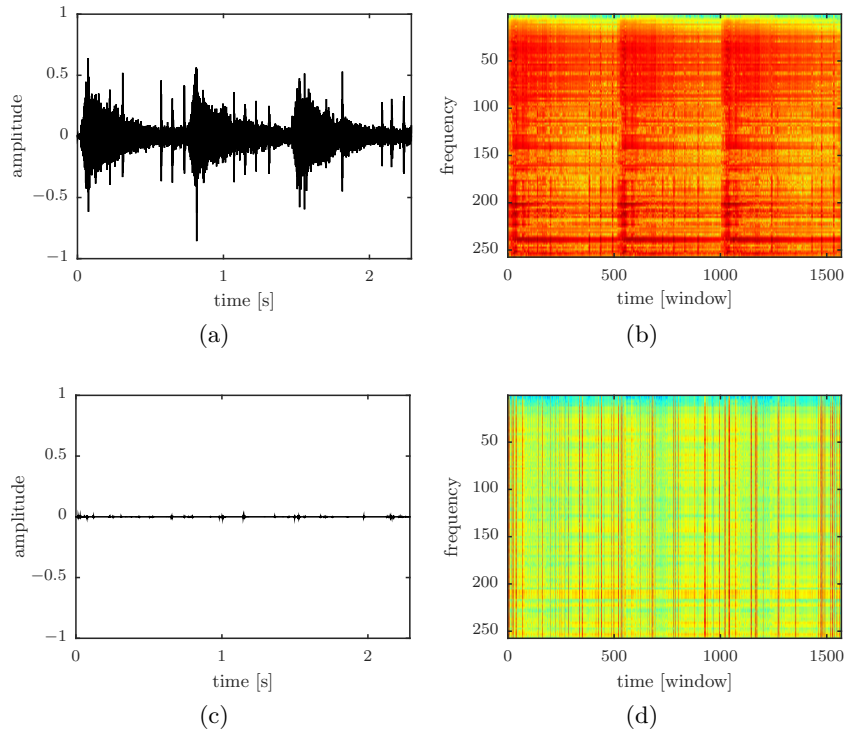


Figure 5: (a) Reconstruction of f by NNPCA and NNMF and (b) spectrogram; (c) difference to reconstruction by NNPCA and ICA and (d) spectrogram.

References

- [1] J.-F. Cardoso: perso.telecom-paristech.fr/~cardoso/guidesepsou.html
- [2] M.A. Casey and A. Westner: Separation of mixed audio sources by independent subspace analysis. *Proceedings of the International Computer Music Conference*, Berlin, 2000.
- [3] P. Comon: Independent component analysis, a new concept? *Signal Processing* **36**(3), 1994, 287–314.
- [4] D. Fitzgerald, E. Coyle, and B. Lawlor: Sub-band Independent subspace analysis for drum transcription. *Proceedings of the 5th International Conference on Digital Audio Effects (DAFX-02)*. Hamburg, Germany, 2002.
- [5] D. Fitzgerald, E. Coyle, and B. Lawlor: Independent subspace analysis using locally linear embedding. *Proceedings of the 6th International Conference on Digital Audio Effects (DAFX-03)*, London, UK, 2003, 13–17.

- [6] D. Fitzgerald, M. Cranitch, and E. Coyle: Non-negative tensor factorisation for sound source separation. *Proceedings of Irish Signals and Systems Conference*, Dublin, Ireland, 2005, 8–12.
- [7] G. Grindlay: <http://www.ee.columbia.edu/~grindlay/code.html>.
- [8] M. Guillemard, A. Iske, and S. Krause-Solberg: Dimensionality reduction methods in independent subspace analysis for signal detection. *Proceedings of the 9th International Conference on Sampling Theory and Applications (SampTA2011)*, Singapore, 2011.
- [9] B.C. Hall: *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Graduate Texts in Mathematics, vol. 222, Springer, New York, 2004.
- [10] X. Han: Nonnegative principal component analysis for proteomic tumor profiles. *Proceedings of the SIAM International Conference on Data Mining*, USA, 2010, 269–280.
- [11] M. Helén and T. Virtanen: Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine. *Proceedings of 13th European Signal Processing Conference*, Istanbul, Turkey, 2005, 1091–1094.
- [12] S. Krause-Solberg: *Non-Negative Dimensionality Reduction in Signal Separation*. Ph.D. thesis, Universität Hamburg, 2016.
- [13] D.D. Lee and H.S. Seung: Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, vol. 13, MIT Press, 2000, 556–562.
- [14] J.A. Lee and M. Verleysen: *Nonlinear Dimensionality Reduction*. Information Science and Statistics Series. Springer, London, 2010.
- [15] Y. Panagakis, C. Kotropoulos, and G.R. Arce: Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification. *IEEE Transactions on Audio, Speech, and Language Processing* **18**(3), 2010, 576–588.
- [16] M. Plumbley and E. Oja: A nonnegative PCA algorithm for independent component analysis. *IEEE Transactions on Neural Networks* **15**(1), 2004, 66–76.
- [17] M.D. Plumbley: Geometrical methods for non-negative ICA: manifolds, Lie groups and toral subalgebras. *Neurocomputing* **67**, 2005, 161–197.

- [18] P. Smaragdis and J.C. Brown: Non-negative matrix factorization for polyphonic music transcription. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2003, 177–180.
- [19] C. Uhle, C. Dittmar and T. Sporer: Extraction of drum tracks from polyphonic music using independent subspace analysis. *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japan, 2003, 843–848.
- [20] T. Virtanen: Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria. *IEEE Transactions on Audio, Speech, and Language Processing* **15**(3), 2007, 1066–1074.
- [21] S. Zafeiriou and N.A. Laskaris: Nonnegative embeddings and projections for dimensionality reduction and information visualization. *Proceedings of the 20th International Conference on Pattern Recognition*, 2010, 726–729.
- [22] R. Zass and A. Shashua: Nonnegative sparse PCA. *Advances in Neural Information Processing Systems*, vol. 19, MIT Press, 2007, 1561–1568.