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

Mijail Guillemard

Institute of Mathematics, Hamburg University of Technology

Am Schwarzenberg-Campus 3, 21073 Hamburg, Germany

mijail.guillemard@tuhh.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.

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) [4, 7, 8, 26], based on *independent component analysis* (ICA) or *non-negative matrix factorization* (NNMF) [9, 25, 27]. 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 [20] for an overview) were developed, e.g. *isomap* and *local linear embedding* (LLE). In [14], 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 [8, 14, 26], 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

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

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.

$$
\begin{array}{ccccc}
X & \subset & \mathcal{M} & \subset & \mathbb{R}^D \\
\Big\downarrow P & & \Big\downarrow B & & \\
Y & \subset & \Omega & \subset & \mathbb{R}^d
\end{array}
\tag{1}
$$

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

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

with $\mathcal{U} \subset \{f \colon X \to \mathbb{R}^d\}$ and a cost functional $g \colon \mathcal{U} \to \mathbb{R}$, 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 [20]). 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

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

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 [20].

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 expected rotational invariance of g, i.e.,

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

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 [18] 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

$$\min_{U^T U = \mathbf{1}} \sum_{k=1}^{n} \left\| x_k - UU^T x_k \right\|^2. \tag{4}$$

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 YY^T. This allows us to reformulate (4) as an equivalent maximization problem,

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

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 XX^T.

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

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

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

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

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

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

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

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

where we let

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

as this was proposed in [24] 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 [23] 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 [29].

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 [15] 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 [18, 24].

We finally remark that the NNPCA, as specified in [29] and used in several papers (e.g. [16, 22, 28]), 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 [8, 14, 26]. Moreover, *non-negative matrix factorization* (NNMF) has gained enormous interest in signal separation (see e.g. [17, 25, 27]). 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 $X = (x_1, ..., x_n) \in \mathbb{R}^{D \times n}$, 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
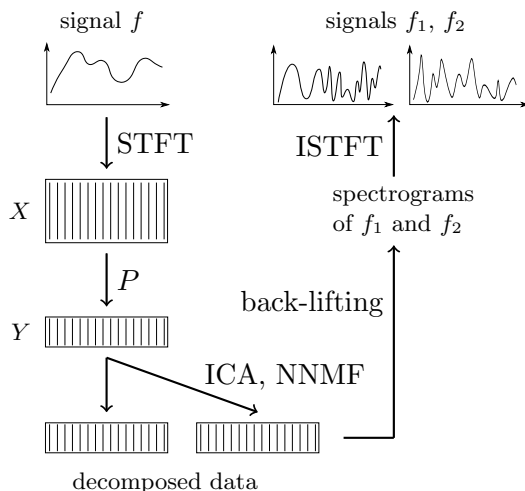
Figure 1: Signal separation with dimensionality reduction.

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

$$Y = AS, \tag{8}$$

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 stochastical independence can be measured by the Kullback-Leibler distance (cf. [5]).

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 [4]). 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. [9, 25, 27]). 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 [19], as used for decomposition of signal data in [27].

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 [3]) or NNMF (toolbox [12])
separate	$WY = Z_1 + Z_2$, see [4]
invert	dimensionality reduction ($X_1 = UW^T Z_1, X_2 = UW^T Z_2$)
reconstruct	time series of sources s_i using phase Φ (ISTFT)

We remark that the consecutive steps of our proposed signal separation by NNDR can be performed by the application of the computational methods, as they are suggested for the different stages in the above algorithm. But we wish to discuss three main steps of our algorithm in more detail: (a) reduction (b) decomposition (c) inversion. To this end, we discuss theoretical properties of the algorithm's ingredients to further justify our signal separation framework.

In particular, we explain basic hypotheses concerning the performance of the three steps (a), (b), and (c). In the case of ICA, statistical independence is the main requirement for a successful signal separation procedure. In the context of our setting, it needs to be verified whether the dimensionality reduction map P preserves the statistical independence of the input signals. In the case of NNMF, we require suitable criteria that can be formulated by using a *geometrical* interpretation of NNMF. Indeed, as indicated for instance in [6, 10], the procedure of NNMF can be described as a *nested polytopes problem.*

Moreover, a decomposition of $Y = AS$ in the sense of NNMF requires an analysis of the *pullback of Y, $\phi(Y) = YD(Y)$,* where $D(Y)$ is a diagonal matrix whose diagonal elements are the inverse of the ℓ_1-norms of the columns of Y. Now, the geometrical interpretation of NNMF can be described as the search of a polytope $T = \mathrm{conv}(\phi(A))$ (the convex hull of the columns of the matrix $\phi(A)$), such that $\mathrm{conv}(\phi(Y)) \subseteq T \subseteq \Delta^n$, where Δ^n is the unit simplex of dimension $n - 1$. The theoretical analysis for our framework can therefore be described as the adequate dimensionality reduction and reconstruction of the polytopes involved in the geometrical interpretation of NNMF.

The reconstruction part (i.e, the inversion in the penultimate line of our algorithm) can also be formalized. Especially, the rich theory on approximation methods for high-dimensional scattered data data by using radial kernels, or, *radial basis functions* (RBF) is particularly useful. For a more general discussion on theoretical aspects of multivariate scattered data approximation by RBF we refer to the textbooks [2, 30], whereas a more recent account to the interactions between kernels and signal analysis tools can be found in our work [13].

As previously described, the application of (nonlinear) dimensionality reduction methods requires suitable strategies for mapping data between high and low dimensional Euclidean spaces. In fact, various (nonlinear) dimensionality reduction methods incorporate intrinsic interpolation strategies. In the general case, however, the reconstruction of high-dimensional scattered data from low-dimensional representations is a difficult problem. The general concept of reproducing kernels in their native Hilbert spaces has already provided powerful tools for multivariate scattered data approximation [2, 30]. A more basic description of the main tools can be found in [1]. We finally remark that additional comparisons between ICA and NNMF can be found in [21].

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.
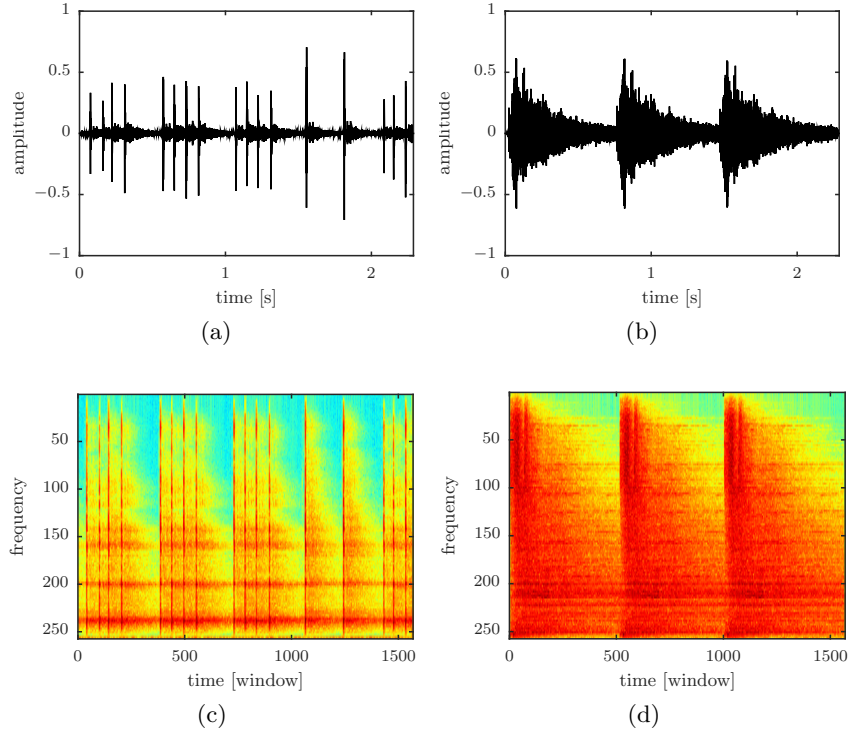
Figure 2: Two acoustic transient signals: (a) castanets f_1, (b) cymbal signal f_2; (c) spectrogram of f_1, (d) spectrogram of f_2.

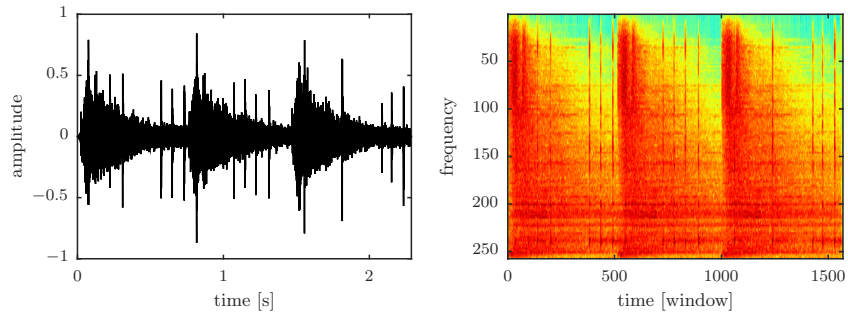

Figure 3: Signal $f = f_1 + f_2$ and its spectrogram X.

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 Hann-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
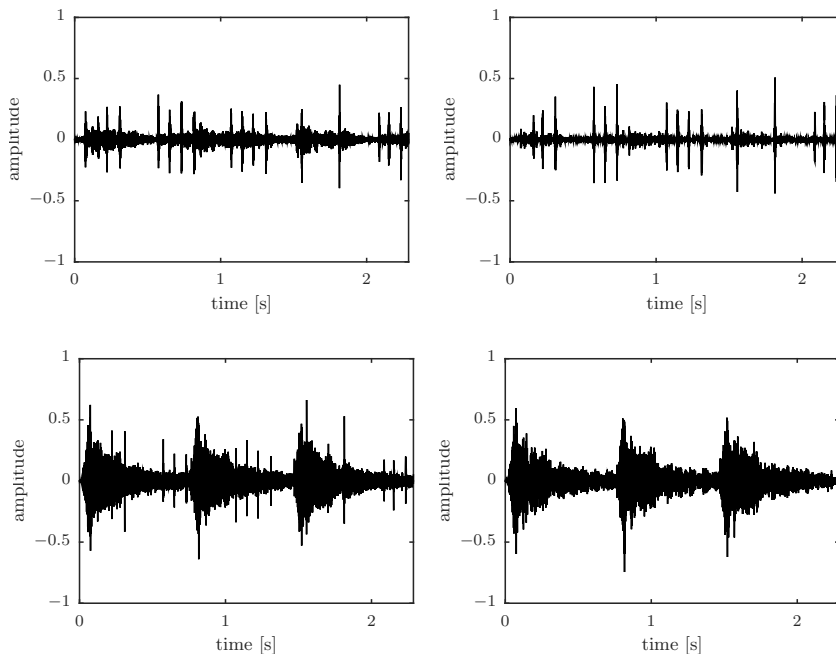
from the original signal.

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

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). Given our numerical results, we can conclude that NNPCA-ICA outperforms NNPCA-NNMF, as shown in waveform for intermediate low-energy regions of the cymbal signal (see Figure 4).

We can explain the effects that are visible in Figure 4 as follows.

First note that the statistical independence between the attack in the castanets and the cymbal signal may be more pronounced in the low-energy areas of the cymbal signal, in contrast to the high energy areas. Secondly, by application of the proposed signal separation method, phase information of the frequency data is being removed. But phase information is of importance, when it comes to the reconstruction of attack components in acoustic signals. However, in our current implementation of NNMF, only the energy levels in the spectrogram are considered. For similar comparisons between ICA and NNMF, we refer to [21].

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 (see Fig-

ure 4). This, in fact, is not too surprising since matching energy levels of sources are rather critical, not only from a technical viewpoint, but also from a numerical viewpoint. Yet a more systematic tradeoff study between overall recovery error versus relative energy levels of the various sources remains to be developed.

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).
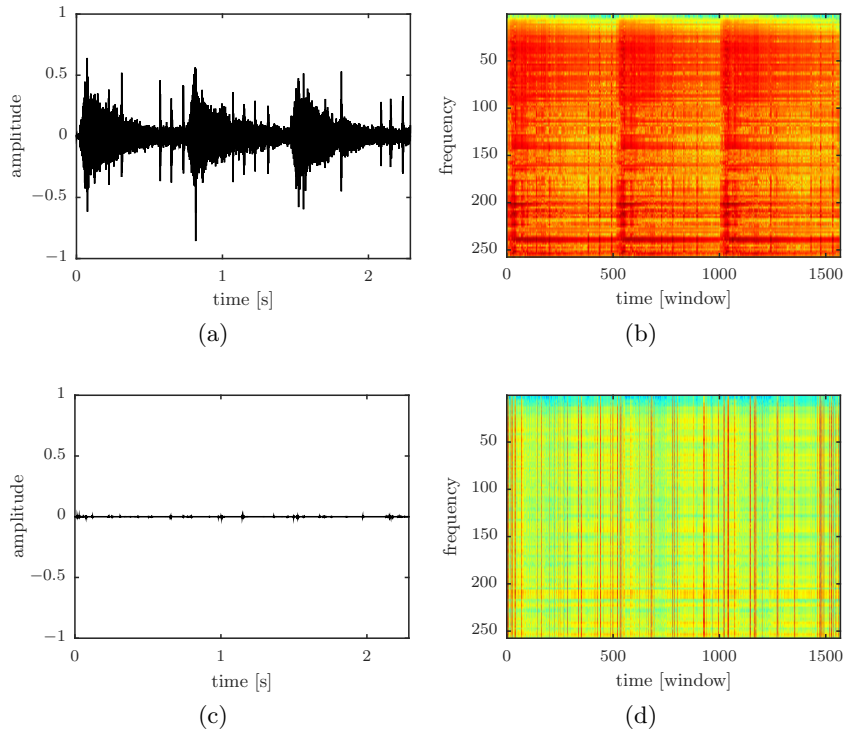

Figure 5: (a) Reconstruction of f by NNPCA and NNMF and (b) spectrogram; (c) difference to reconstruction by NNPCA and ICA and (d) spectrogram.

For the purpose of further evaluation and comparison between the utilized signal separation algorithms, we have recorded their L^∞-error ε_∞ and their *signal to noise ratio* (SNR) for the reconstructed audio signal, $\mathrm{SNR}_{\mathrm{rec}}$, for the cymbal signal, $\mathrm{SNR}_{\mathrm{cym}}$, and for the castanet signal, $\mathrm{SNR}_{\mathrm{cas}}$, respectively. Our numerical results are shown in Table 1.

We can explain the utilized quality indicators as follows.

The L^∞-error in the time-amplitude domain is the maximum in time for the difference between the amplitudes of f_i and those of the reconstructed source

Table 1: Numerical comparison by L^∞-error ε_∞ in (9) and signal to noise ratio (SNR) in (10) for the reconstructed signal, SNR_{rec}, for the cymbal, SNR_{cym}, and for the castanet signal, SNR_{cas}. The overall length of the signal is 2.27 seconds [s] (elapsed time) and the size of the spectrogram is 256×1569.

# components	time [s]	ε_∞ in (9)	SNR_{rec}	SNR_{cym}	SNR_{cas}
NNPCA-NNMF					
3	14.8	0.57	2.899	2.339	0.978
10	35.6	0.55	2.239	1.262	0.553
20	248	0.68	2.170	1.942	0.237
NNPCA-ICA					
3	8	0.70	3.212	2.421	1.114
10	20	0.58	2.227	1.368	1.888
20	201	0.64	2.207	2.082	0.109

signals s_i, i.e.,

$$\varepsilon_\infty(f_i, s_i) = \max_k |f_i(t_k) - s_i(t_k)|. \tag{9}$$

For the reconstruction we only used the phase of the original mixture rather than separating also the phase spectrogram. Therefore, the extracted sources may have different phases than the original input sources, which typically leads to an increase of the L^∞-error.

Nevertheless, L^∞-error ε_∞ can be used to quantify the effect incurred by the dimensionality reduction. To this end, we just reduced the dimension and lifted the data back to the high-dimensional space to obtain a reconstruction of the mixture. This reconstruction does not suffer from a possible phase shift as the phase remains unchanged. Thus, with using the L^∞-error the reconstructed mixture can be compared to the original input signal.

The signal to noise ratio (SNR) is a standard error measure for estimating the portion of noise in a signal, and is measured in decibel (dB). Here, the error $f_i(t_k) - s_i(t_k)$ can be considered as noise and thus, the SNR error [11] is

$$\varepsilon_{\text{SNR}}(f_i, s_i) = 10 \log_{10} \left(\frac{\sum_k |f_i(t_k)|^2}{\sum_k |f_i(t_k) - s_i(t_k)|^2} \right). \tag{10}$$

The obtained SNR values are interpreted as follows. The larger the SNR, the smaller the noise level, i.e., the better the reconstruction.

Our numerical results indicate that for signal separation *without* dimensionality reduction, NNMF is quite competitive to ICA, as further supported by [27]. We remark that this complies with our numerical results in [18] and with earlier of our numerical experiments which we omitted here for the sake of brevity. Although it remains to further strengthen the theory on this, we feel comfortable enough to state that the separation by NNPCA-NNMF can 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.

References

[1] D.S. Broomhead and M. Kirby: A new approach to dimensionality reduction: theory and algorithms. *SIAM Journal on Applied Mathematics* **60**(6), 2000, 2114–2142.

[2] M.D. Buhmann: *Radial Basis Functions.* Cambridge University Press, Cambridge, UK, 2003.

[3] J.-F. Cardoso: `perso.telecom-paristech.fr/~cardoso/guidesepsou.html`

[4] M.A. Casey and A. Westner: Separation of mixed audio sources by independent subspace analysis. *Proceedings of the International Computer Music Conference*, Berlin, 2000.

[5] P. Comon: Independent component analysis, a new concept? *Signal Processing* **36**(3), 1994, 287–314.

[6] D. Donoho and V. Stodden: When does non-negative matrix factorization give correct decomposition into parts? *Proceedings of Neural Information Processing Systems*, 2003.

[7] 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.

[8] 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.

[9] 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.

[10] N. Gillis: Sparse and unique nonnegative matrix factorization through data preprocessing. *Journal of Machine Learning Research* **13**, 2012, 3349–3386.

[11] R.C. Gonzalez and R.E. Woods: *Digital Image Processing.* Third Edition, Prentice Hall, 2008.

[12] G. Grindlay: `http://www.ee.columbia.edu/~grindlay/code.html`.

[13] M. Guillemard and A. Iske: Interactions between kernels, frames and persistent homology. In: *Recent Applications of Harmonic Analysis to Function Spaces, Differential Equations, and Data Science. Volume 2: Novel Methods in Harmonic Analysis*, I. Pesenson, Q.T. Le Gia, A. Mayeli, H. Mhaskar, and D.-X. Zhou (eds.), Birkhäuser, 2017, 861–888.

[14] 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 SamplingTheory and Applications (SampTA2011)*, Singapore, 2011.

[15] B.C. Hall: *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Graduate Texts in Mathematics, vol. 222, Springer, New York, 2004.

[16] X. Han: Nonnegative principal component analysis for proteomic tumor profiles. *Proceedings of the SIAM International Conference on Data Mining*, USA, 2010, 269–280.

[17] 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.

[18] S. Krause-Solberg: *Non-Negative Dimensionality Reduction in Signal Separation*. Ph.D. thesis, Universität Hamburg, 2016.

[19] 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.

[20] J.A. Lee and M. Verleysen: *Nonlinear Dimensionality Reduction*. Information Science and Statistics Series. Springer, London, 2010.

[21] A. Mirzal: NMF versus ICA for blind source separation. *Advances in Data Analysis and Classification* **11**(1), 2017, 25–48.

[22] 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.

[23] M. Plumbley and E. Oja: A nonnegative PCA algorithm for independent component analysis. *IEEE Transactions on Neural Networks* **15**(1), 2004, 66–76.

[24] M.D. Plumbley: Geometrical methods for non-negative ICA: manifolds, Lie groups and toral subalgebras. *Neurocomputing* **67**, 2005, 161–197.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] R. Zass and A. Shashua: Nonnegative sparse PCA. *Advances in Neural Information Processing Systems*, vol. 19, MIT Press, 2007, 1561–1568.

[30] H. Wendland: *Scattered Data Approximation*. Cambridge University Press, Cambridge, UK, 2005.