

Non-Negative Dimensionality Reduction for Audio Signal Separation by NNMF and ICA

Sara Krause-Solberg and Armin Iske

Department of Mathematics, University of Hamburg, Germany

{sara.krause-solberg, iske}@math.uni-hamburg.de

Abstract—Many relevant applications of signal processing rely on the separation of sources from a mixture of signals without a prior knowledge about the mixing process. Given a mixture of signals $f = \sum_i f_i$, the task of signal separation is to estimate the components f_i by using specific assumptions on their time-frequency behaviour or statistical characteristics. Time-frequency data is often very high-dimensional which affects the performance of signal separation methods quite significantly. Therefore, the embedding dimension of the time-frequency representation of f 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 ICA and NNMF, since non-negative input data are required. 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 NNMF and ICA for single channel signal separation in combination with non-negative principal component analysis (NNPCA), where our main interest is in acoustic signals with transitory components.

I. INTRODUCTION

In many relevant applications of signal processing there is an increasing demand for effective methods to estimate the components from a mixture of acoustic signals. In recent years, different decomposition techniques were developed to do so, including *independent subspace analysis* (ISA), based on *independent component analysis* (ICA), see [1], [3], [18], and *non-negative matrix factorization* (NNMF), see [5], [17], [19]. The computational complexity of these methods, however, may be prohibitively large, in particular for real-time computations on audio signals, which is one application class that we wish to address in this paper. In our previous work [6], we have combined dimensionality reduction techniques with decomposition methods to reduce the large complexity in signal detection from acoustic signals.

In recent years, several different projection methods were used for dimensionality reduction, among them the classical linear *principal component analysis* (PCA) and *multidimensional scaling* (MDS), as well as nonlinear methods, such as *isomap* and *local linear embedding* (LLE). For a more comprehensive account to nonlinear dimensionality reduction

methods we refer to [13]. In applications of signal separation, dimensionality reduction methods are used to first 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 source signals. This combination was shown to be a promising tool in various applications of signal separation.

However, for the application of dimensionality reduction in combination with NNMF and ICA, respectively, *non-negative* dimensionality reduction methods are essentially required to guarantee non-negative output data from non-negative input data. Therefore, the construction of *non-negative* dimensionality reduction methods is of particular interest.

In this paper we introduce a novel concept for the construction of *non-negative dimensionality reduction* (NNDR) methods, to be used in combination with signal separation. We remark that dimensionality reduction for signal separation has been studied in previous work [4], [6], [18], but without any non-negativity constraints on the dimensionality reduction method. Therefore, the idea to apply NNDR methods in the context of signal separation is new. For the construction of suitable NNDR methods we first consider dimensionality reduction methods which can be written as an optimization problem with a suitable cost functional. Then, in order to generate NNDR methods, we reformulate that optimization problem by adding a non-negativity constraint. This reformulation allows us to split the problem of NNDR into two decoupled subproblems: (a) the problem of dimensionality reduction without non-negativity constraint; (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 projections (see also [11]).

Our concept of non-negative dimensionality reduction is explained in Section II, and its coupling with NNMF and ICA is subject of Section III. Numerical examples comparing both decomposition methods are finally presented in Section IV.

II. NON-NEGATIVE DIMENSIONALITY REDUCTION

On given data

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

This work is partly supported by the priority program DFG-SPP 1324 of the Deutsche Forschungsgemeinschaft (DFG).

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 , 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 consider dimensionality reduction methods which can be written as an optimization problem

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

with $\mathcal{U} \subset \{f: X \rightarrow \mathbb{R}^d\}$ and 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 [13]).

In many relevant applications, dimensionality reduction is applied in a preprocessing 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.

Now the idea is to split the optimization problem (3) into two steps: In the first step the dimensionality reduction problem (2) is solved, before in the 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, however, that the splitting does not work for all dimensionality reduction methods. But it turns out, that, whenever such a rotation W exists, the rotational invariance of cost functional g , i.e., $WP \in \mathcal{U}$ with $g(WP) = g(P)$, guarantees

the solution of the splitted problem to also solve problem (3). For more details on this, we refer to our work [11]. But it can be shown that PCA, and also MDS, fit in this framework (cf. [11]). In particular, the existence of the rotation matrix W is guaranteed, if the high-dimensional data is non-negative.

A. Non-Negative PCA (NNPCA)

Classical PCA is a linear dimensionality reduction method, whose construction relies on the singular value decomposition (SVD). 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

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

We remark that the solution of this minimization problem 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 the minimization problem in (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

$$\mathcal{U} = \{U^T \in \mathbb{R}^{d \times D} \mid U^T U = \mathbf{1}\}$$

and

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

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 have

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

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, it is obvious that the splitting ansatz as described above can be applied: Clearly, there exists a matrix $W \in SO(d)$ rotating the reduced data set 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 . To do so, we consider solving an auxiliary optimization problem on the set of orthogonal matrices $SO(d)$ for the cost functional

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

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

as this was proposed in [16] in the context of ICA. Note that the cost functional J penalizes negative entries in the low-dimensional representation $WU^T X$ and 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 a ΔW does not necessarily 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 [15] would ensure the desired. However, one drawback of the approach in [15] is that the determination of the decent direction and the orthogonalization are performed separately in each step. Therefore, it is not clear if the combination of both yields indeed a decent step. Another approach to solve the optimization problem is to include a relaxation term and to consider an unconstrained optimization problem, see [21].

But the most elegant way to minimize the cost functional J in (7) uses the Lie-group structure of $SO(d)$ to transfer the optimization into an optimization on the associated Lie-algebra of skew-symmetric matrices $\mathfrak{so}(d)$. 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 natural way. Due to the vector space property of $\mathfrak{so}(d)$, standard methods, such as line search, can be applied to find the minimum (see [7], [10], [11], [16]).

We finally remark that the NNPCA, as specified in [21] and used in several papers (e.g. [8], [14], [20]), 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.

III. APPLICATION OF NNDR TO SIGNAL SEPARATION

Non-negative dimensionality reduction (NNDR) has several applications to signal processing. NNDR can e.g. be used for signal detection as well as for blind source separation. In recent years, blind source separation techniques as *independent component analysis* (ICA) or *independent subspace analysis* (ISA) were coupled with different dimensionality reduction methods such as PCA, LLE or isomap, to mention just a few (see e.g. [1], [6], [18]). But the required non-negativity of the input data for the separation algorithm has not been discussed. On the other hand, *non-negative matrix factorization* (NNMF) has gained enormous interest in signal separation (see [9], [17],

[19]), but has not been used in combination with dimensionality reduction tools so far.

Let us briefly introduce the basic concept of signal separation, with focussing on the interaction between dimensionality reduction and decomposition techniques. 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 applied to the data matrix X to then decompose the reduced data, being represented by a matrix Y of smaller dimension, into different components, each of which is to be assigned to one of the source signals. The basic steps of this procedure are shown in Fig. 1.

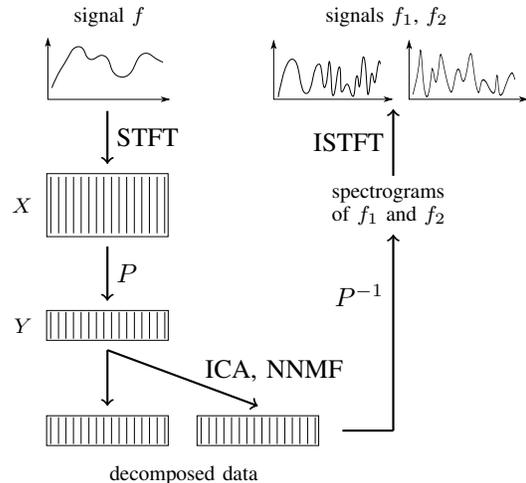


Fig. 1: Signal separation with dimensionality reduction.

Recall that we assume the data points in X to lie on a smooth manifold (see the diagram (1)). The dimensionality reduction map P which maps X on Y is non-negativity preserving in the sense of (3). Furthermore, for the reconstruction of the source signals we need a suitable inverse mapping P^{-1} , where P^{-1} is an approximation of B^{-1} in diagram (1). This assumption is in general very restrictive, but in the special case of PCA it is fulfilled 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.,

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

where $A \in \mathbb{R}^{d \times d}$ and $S \in \mathbb{R}^{d \times n}$ are unknown. We assume that $S \subset \Omega$ as it 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 (for details see [2]).

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 [1]). This procedure is called *independent subspace analysis (ISA)*. The main difficulty of ISA is - beside the ICA - to identify components belonging to the same multi-component subspace. This problem, however, is far beyond the aims and scopes of this work.

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 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.

There are different NNMF algorithms available, all of which are relying on the non-negativity $Y, A, S \geq 0$, where different measures $d(Y, AS)$ for the reconstruction error were proposed (see e.g. [5], [17], [19]). Here, we consider using the generalized *Kullback-Leibler divergence*

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

as proposed by Lee and Seung in [12] and used for the decomposition of signal data in [19].

IV. NUMERICAL RESULTS

For the purpose of illustration, we present one numerical example, where we compare the different decomposition strategies 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 Fig. 2.

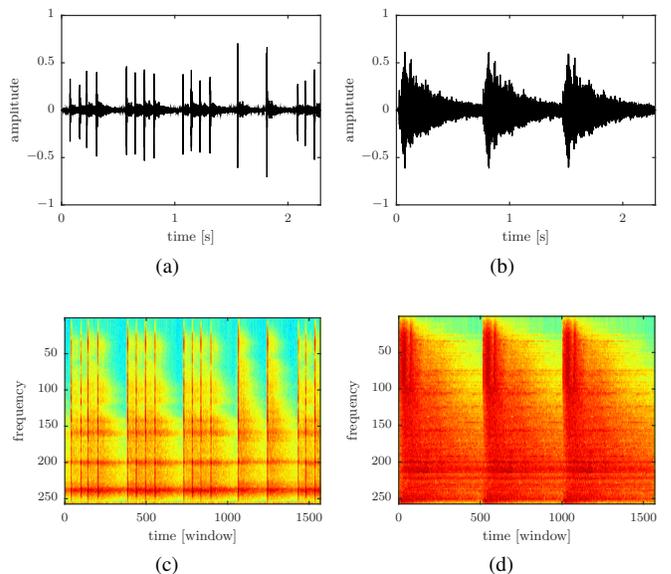


Fig. 2: Two acoustic signals: (a) castanets f_1 , (b) cymbal f_2 ; their spectrograms: (c) for f_1 and (d) for f_2 .

The combination $f = f_1 + f_2$ of the castanets and the cymbal is shown in Fig. 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.

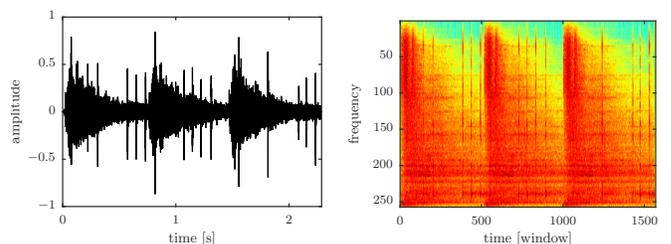


Fig. 3: Signal $f = f_1 + f_2$ and its spectrogram X .

The separations obtained from NNPCA and the decomposition methods, NNMF and ICA, are shown in Fig. 4. Note that either method, NNMF or ICA, reproduces the characteristic peaks of the castanets quite well. But for the separation by NNMF, a few artefacts are visible in the cymbal signal, in contrast to the separation by ICA (see Fig. 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, NNMF or ICA, provides an almost complete reconstruction of the input signal f . In fact, their reconstructions are nearly identical (cf. Fig. 5).

We finally remark that for signal separation *without* dimensionality reduction, NNMF is competitive to ICA [19]. This suggests that the separation by NNPCA-NNMF could be

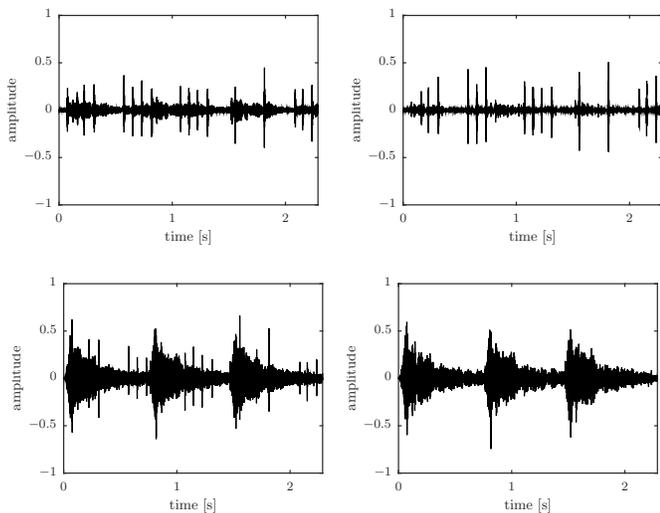


Fig. 4: Left col.: Signal separation by NNPCA-NNMF; Right column: separation by NNPCA-ICA.

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. We defer these points to future research. 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] M. A. CASEY and A. WESTNER. *Separation of Mixed Audio Sources by Independent Subspace Analysis*. In *Proceedings of the International Computer Music Conference*. Berlin, 2000.

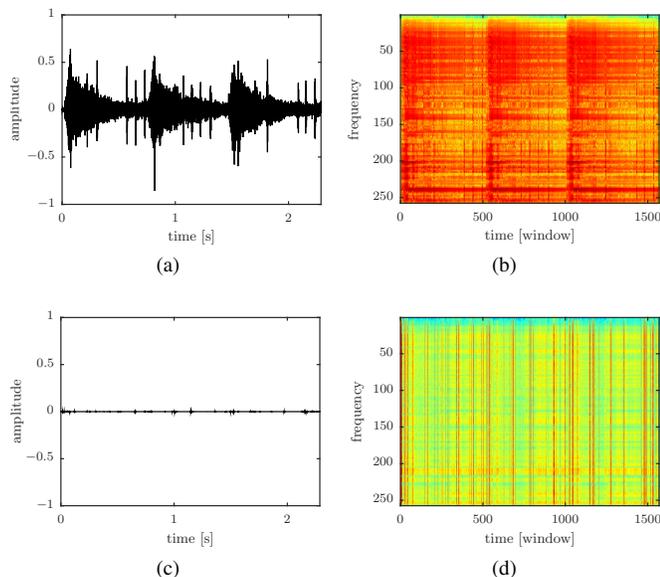


Fig. 5: (a) Reconstruction of f by NNPCA and NNMF and (b) spectrogram; (c) difference to reconstruction by NNPCA and ICA and (d) spectrogram.

[2] P. COMON. *Independent component analysis, A new concept?* In *Signal Processing*, **36**(3), pp. 287–314, 1994.

[3] D. FITZGERALD, E. COYLE and B. LAWLOR. *Sub-band Independent Subspace Analysis for Drum Transcription*. In *Proceedings of the 5th International Conference on Digital Audio Effects (DAFX-02)*. Hamburg, Germany, 2002.

[4] D. FITZGERALD, E. COYLE and B. LAWLOR. *Independent subspace analysis using locally linear embedding*. In *Proceedings of the 6th International Conference on Digital Audio Effects (DAFX-03)*, pp. 13–17. London, UK, 2003.

[5] D. FITZGERALD, M. CRANITCH and E. COYLE. *Non-negative tensor factorisation for sound source separation*. In *Proceedings of Irish Signals and Systems Conference*, pp. 8–12. Dublin, Ireland, 2005.

[6] M. GUILLEMARD, A. ISKE and S. KRAUSE-SOLBERG. *Dimensionality Reduction Methods in Independent Subspace Analysis for Signal Detection*. In *Proceedings of the 9th International Conference on Sampling Theory and Applications (SampTA2011)*. Singapore, 2011.

[7] B. C. HALL. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, volume 222 of *Graduate Texts in Mathematics*. Springer, New York, 2004.

[8] X. HAN. *Nonnegative Principal Component Analysis for Proteomic Tumor Profiles*. In *Proceedings of the SIAM International Conference on Data Mining*, pp. 269–280. USA, 2010.

[9] M. HELÉN and T. VIRTANEN. *Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine*. In *Proceedings of 13th European Signal Processing Conference*, pp. 1091–1094. Istanbul, Turkey, 2005.

[10] A. ISERLES, H. Z. MUNTHE-KAAS, S. NØRSETT and A. ZANNA. *Lie-group methods*. In *Acta Numerica*, pp. 215–365, 2000.

[11] S. KRAUSE-SOLBERG and A. ISKE. *Non-negative dimensionality reduction in signal separation*. Technical report, Department of Mathematics, University of Hamburg, 2015. URL www.math.uni-hamburg.de/home/iske/papers/hbam.pdf.

[12] D. D. LEE and H. S. SEUNG. *Algorithms for Non-negative Matrix Factorization*. In *Advances in Neural Information Processing Systems*, volume 13, pp. 556–562. MIT Press, 2000.

[13] J. A. LEE and M. VERLEYSEN. *Nonlinear Dimensionality Reduction*. Information Science and Statistics Series. Springer, London, 2010.

[14] Y. PANAGAKIS, C. KOTROPOULOS and G. R. ARCE. *Non-Negative Multilinear Principal Component Analysis of Auditory Temporal Modulations for Music Genre Classification*. In *IEEE Transactions on Audio, Speech, and Language Processing*, **18**(3), pp. 576–588, 2010.

[15] M. PLUMBLEY and E. OJA. *A ‘nonnegative PCA’ algorithm for independent component analysis*. In *IEEE Transactions on Neural Networks*, **15**(1), pp. 66–76, 2004.

[16] M. D. PLUMBLEY. *Geometrical methods for non-negative ICA: Manifolds, Lie groups and toral subalgebras*. In *Neurocomputing*, **67**, pp. 161–197, 2005.

[17] P. SMARAGDIS and J. C. BROWN. *Non-Negative Matrix Factorization for Polyphonic Music Transcription*. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 177–180, 2003.

[18] C. UHLE, C. DITTMAR and T. SPORER. *Extraction of drum tracks from polyphonic music using Independent Subspace Analysis*. In *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pp. 843–848. Nara, Japan, 2003.

[19] T. VIRTANEN. *Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria*. In *IEEE Transactions on Audio, Speech, and Language Processing*, **15**(3), pp. 1066–1074, 2007.

[20] S. ZAFEIRIOU and N. A. LASKARIS. *Nonnegative Embeddings and Projections for Dimensionality Reduction and Information Visualization*. In *Proceedings of the 20th International Conference on Pattern Recognition*, pp. 726–729, 2010.

[21] R. ZASS and A. SHASHUA. *Nonnegative Sparse PCA*. In *Advances in Neural Information Processing Systems*, volume 19, pp. 1561–1568. MIT Press, 2007.