

Non-negative Matrix Factorization, A New Tool for Feature Extraction: Theory and Applications

Workshop invited key lecture

Ioan Buciu

Abstract: Despite its relative novelty, non-negative matrix factorization (NMF) method knew a huge interest from the scientific community, due to its simplicity and intuitive decomposition. Plenty of applications benefited from it, including image processing (face, medical, etc.), audio data processing or text mining and decomposition. This paper briefly describes the underlying mathematical NMF theory along with some extensions. Several relevant applications from different scientific areas are also presented. NMF shortcomings and conclusions are considered.

Keywords: Non-negative matrix factorization, image decomposition, applications.

1 Introduction

Data decomposition has multiple meanings and goals, arising from many applications, such as data compression, transmission or storage. An important application comes from the pattern recognition field, where the purpose is to automatically cluster the data samples into distinct classes. This task usually requires the extraction of discriminant latent features from the initial data prior to classification. Through the feature extraction issue the computational load is reduced and possible discovery of task-relevant hidden variables can be performed. Feature extraction is possible as the data we perceive may contain significant redundant information. Another reason is that we do not need the full information. Rather, the extracted information is application-dependent. For example, in a facial expression recognition task, we do not care about the whole face. We can discard information related to the cheek or nose, which does not contribute to discriminate among expressions, but, certainly the eyes or mouth region encapsulate valuable information which is highly relevant for this task and can not be neglected.

Non-negative Matrix Factorization is a relatively recent approach to decompose data into two factors with non-negative entries. At least two reasons exist to constraint the entries being non-negative. The first reason comes from neurophysiology where the firing rate of visual perception neurons are non-negative. The second reason comes from the image processing field, where the intensity images have nonnegative values. Given a matrix \mathbf{X} of size $m \times n$ whose columns contain data samples, the data decomposition task can be described by factoring \mathbf{X} into two terms \mathbf{W} and \mathbf{H} of size $m \times p$ and $p \times n$, respectively, where $p < \min(m, n)$. The decomposition is performed so that the product \mathbf{WH} should approximate as best as possible the original data \mathbf{X} , for $p < \min(m, n)$. The columns of \mathbf{W} are usually called basis vectors and the rows of \mathbf{H} are called decomposition (or encoding) coefficients. Thus, the original data are represented as linear combinations of these basis vectors. Contrary to other decomposition techniques which allow some factors to vanish (due to their equal values and opposite signs, as in Principal Component Analysis, for instance), NMF provides a more intuitive and meaningful decomposition allowing only additive operations.

This paper briefly describes the underlying mathematical NMF theory along with some extensions. Several relevant applications from different scientific areas are also presented. NMF shortcomings and conclusions end the paper.

2 Non-negative Matrix Factorization Approach

Through the decomposition process, each element x_{ij} of the matrix \mathbf{X} can be written as $x_{ij} \approx \sum_k w_{ik} h_{kj}$. The quality of approximation depends on the cost function used. Two cost functions were proposed by Lee and Seung in [1]: the Euclidean distance between \mathbf{X} and \mathbf{WH} and KL divergence. KL based cost function is expressed as:

$$D_{NMF}(\mathbf{X} \| \mathbf{WH}) \triangleq \sum_{i,j} \left(x_{ij} \ln \frac{x_{ij}}{\sum_k w_{ik} h_{kj}} + \sum_k w_{ik} h_{kj} - x_{ij} \right), \quad (1)$$

This expression can be minimized by applying multiplicative update rules subject to $\mathbf{W}, \mathbf{H} \geq 0$. Since both matrices \mathbf{W} and \mathbf{H} are unknown, we need an algorithm which is able to find these matrices by minimizing the divergence

(1). By using an auxiliary function and the Expectation Maximization (EM) algorithm [2], the following update rule for computing h_{kj} is found to minimize the KL divergence at each iteration t [1]:

$$h_{kj}^{(t)} = h_{kj}^{(t-1)} \frac{\sum_i w_{ki}^{(t)} \frac{x_{ij}}{\sum_k w_{ik}^{(t)} h_{kj}^{(t-1)}}}{\sum_i w_{ik}^{(t)}}. \quad (2)$$

By reversing the roles of \mathbf{W} and \mathbf{H} in (2), a similar update rule for each element w_{ik} of \mathbf{W} is obtained:

$$w_{ik}^{(t)} = w_{ik}^{(t-1)} \frac{\sum_j \frac{x_{ij}}{\sum_k w_{ik}^{(t-1)} h_{kj}^{(t)}} h_{jk}^{(t)}}{\sum_j h_{kj}^{(t)}}. \quad (3)$$

Both updating rules are applied alternatively in an EM manner and they guarantee a nonincreasing behavior of the KL divergence.

2.1 Non-negative Matrix Factorization Extensions

Several NMF variants have been proposed in the literature for tailoring the standard NMF approach to specific applications or rationales. Li et al [3] have developed a variant, termed Local Non-negative Matrix Factorization (LNMF) algorithm, imposing more constraints to the KL cost function to get more localized image features. The associated cost function is then given by:

$$f_{LNMF}^{KL}(\mathbf{X}||\mathbf{WH}) \triangleq f_{NMF}^{KL}(\mathbf{X}||\mathbf{WH}) + \alpha \sum_{ij} u_{ij} - \beta \sum_i v_{ii}, \quad (4)$$

where $[\mathbf{u}_{ij}] = \mathbf{U} = \mathbf{W}^T \mathbf{W}$ and $[\mathbf{v}_{ij}] = \mathbf{V} = \mathbf{H} \mathbf{H}^T$. Here, α and $\beta > 0$ are constants. By maximizing the third term in (4), the total squared projection coefficients over all training images is maximized. The second term can be further split into two parts:

1. $\sum_i u_{ii} \rightarrow \min$. This term guarantees the generation of more localized features on the basis images \mathbf{Z} , than those resulting from NMF, since the basis image elements are constrained to be as small as possible.
2. $\sum_{i \neq j} u_{ij} \rightarrow \min$. This enforces basis orthogonality, in order to minimize the redundancy between image bases.

The following factors updating rules were found for the KL -based LNMF cost function:

$$h_{kj}^t = \sqrt{h_{kj}^{t-1} \frac{\sum_i w_{ki} \frac{x_{ij}}{\sum_k w_{ik} h_{kj}^{t-1}}}{h_{kj}^{t-1}}} \quad (5)$$

$$w_{ik}^t = \frac{w_{ik}^{t-1} \sum_j \frac{x_{ij}}{\sum_k w_{ik}^{t-1} h_{kj}} h_{jk}}{\sum_j h_{kj}}. \quad (6)$$

LNMF was further extended by Buciu and Pitas [4], who developed a NMF variant that takes into account class information. Their algorithm, termed Discriminant Non-negative Matrix Factorization (DNMF), leads to a class-dependent image representation. The KL -based DNMF cost function is given by:

$$f_{DNMF}^{KL}(\mathbf{X}||\mathbf{WH}) \triangleq f_{LNMF}^{KL}(\mathbf{X}||\mathbf{WH}) + \gamma \mathbf{S}_w - \delta \mathbf{S}_b, \quad (7)$$

where γ and δ are constants. The new terms are the within-class \mathbf{S}_w and the between-class \mathbf{S}_b scatter matrix, respectively, expressed as following:

1. $\mathbf{S}_w = \sum_{c=1}^Q \sum_{l=1}^{n_c} (\mathbf{h}_{cl} - \mu_c)(\mathbf{h}_{cl} - \mu_c)^T$. Here, Q are the image classes and n_c is the number of training samples in class $c = 1, \dots, Q$. Each column of the $p \times n$ matrix \mathbf{H} is viewed as image representation coefficients vector \mathbf{h}_{cl} , where $c = 1, \dots, Q$ and $l = 1, \dots, n_c$. The total number of coefficient vectors is $n = \sum_{c=1}^Q n_c$. Further, $\mu_c = \frac{1}{n_c} \sum_{l=1}^{n_c} \mathbf{h}_{cl}$ is the mean coefficient vector of class c , and $\mu = \frac{1}{n} \sum_{c=1}^Q \sum_{l=1}^{n_c} \mathbf{h}_{cl}$ is the global mean coefficient vector.
2. $\mathbf{S}_b = \sum_{c=1}^Q (\mu_c - \mu)(\mu_c - \mu)^T$ defines the scatter of the class mean around the global mean μ .

By imposing these terms in the cost function, the decomposition coefficients now encode class information and they are updated according to the following expression:

$$h_{kl(c)}^{(t)} = \frac{2\mu_c - 1 + \sqrt{(1 - 2\mu_c)^2 + 8\xi h_{kl(c)}^{(t-1)} \sum_i w_{ki}^{(t)} \frac{x_{ij}}{\sum_k w_{ik}^{(t)} h_{kl(c)}^{(t-1)}}}}{4\xi} \quad (8)$$

where $\xi = \gamma - \beta$ is a constant. The elements h_{kl} are then concatenated for all Q classes as:

$$h_{kj}^{(t)} = [h_{kl(1)}^{(t)} | h_{kl(2)}^{(t)} | \dots | h_{kl(Q)}^{(t)}] \quad (9)$$

where “|” denotes concatenation. The expression for updating the basis vectors remains the same as in the LNMF approach.

Sparseness is an important issue for image decomposition and representation in the Human Visual System (HVS). Many theoretical papers and experiments brought evidences that the response of the mammalian primary visual cortex (know also as V1 neurons) can be described by localized, oriented and bandpass filters (also known as receptive fields). When applied to natural images, these filters decompose the images into features that are very similar to those obtained by HVS receptive fields. Viewed within this light, Hoyer [5] proposed a new NMF version called Non-negative Sparse Coding (NNSC) where auxiliary constraints are used to impose factor sparseness. The sparseness measure is based on the relation between the L_1 norm and L_2 norm, i.e., $\text{sparseness}(\mathbf{x}) = \frac{\sqrt{m} - \|\mathbf{x}\|_1 / \|\mathbf{x}\|_2}{\sqrt{m} - 1}$. Furthermore, a penalty term of the form $J(\mathbf{W}) = (\zeta \|\text{resh}(\mathbf{W})\|_2 - \|\text{resh}(\mathbf{W})\|_1)^2$ is introduced in the standard NMF problem, where $\zeta = \sqrt{km} - (\sqrt{km} - 1)\eta$, and $\text{resh}(\cdot)$ is the operator which transforms a matrix into a column vector in a column-wise fashion. Here, the desired sparseness for the basis vectors is controlled by η , which can vary from 0 to 1. By replacing \mathbf{W} with \mathbf{H} , the sparseness control can be applied to the encoding coefficients.

A sparse NMF variant called nonsmooth NMF (nsNMF) was proposed by Montano et al. in [6], which allows a controlled sparseness degree in both factors. Like LNMF and NNSC methods, nsNMF also leads to a local image decomposition. However, unlike the LNMF approach, nsNMF explicitly modifies the sparseness degree. Also, unlike NNSC, this variant applies the sparseness concept directly to the model, achieving global sparseness. Imposing sparseness in one of the NMF factors (as NNSC does) will almost certainly force smoothness in the other in an attempt to reproduce the data as best as possible. Additionally, forcing sparseness constraints on both the basis and the encoding vectors decreases the data variance explained by the model. The new variant nsNMF seems to be more robust to this effect. The nsNMF decomposition is given by $\mathbf{X} = \mathbf{W}\mathbf{O}\mathbf{H}$. The matrix $\mathbf{O}_{p \times p}$ is a square positive symmetric “smoothing” matrix defined as:

$$\mathbf{O} = (1 - v)\mathbf{I} + \frac{v}{p}\mathbf{1}\mathbf{1}^T, \quad (10)$$

with \mathbf{I} the identity matrix and $\mathbf{1}$ is a vector of ones. The parameter $0 \leq v \leq 1$ controls the extent of smoothness of the matrix operator \mathbf{O} . However, strong smoothing in \mathbf{O} will force strong sparseness in both the basis and the encoding vectors, in order to maintain faithfulness of the model to the data. Accordingly, the parameter v controls the model sparseness. The suitability of the proposed method over NMF, NNSC and LNMF is investigated with respect to the deterioration of the goodness of fit between the data and the model. The nsNMF model maintained almost perfect faithfulness to the data, expressed by a variance (of goodness) value greater than 99 % for a wider range of sparseness degree, compared with the other NMF variants whose variance decreases with sparseness degree modification.

Other several NMF extensions exist. Due to space limitation we only mention some of them, including: projective NMF [7], temporal NMF with spatial decorrelation constraints [8], shifted NMF [9], incremental NMF [10], sparse higher order NMF [11], and polynomial NMF [12].

3 Non-negative Matrix Factorization Applications

The standard NMF approach and its variants have been extensively used as feature extraction techniques for various applications, especially for high dimensional data analysis. The newly formed low dimensionality subspace represented by the basis vectors should capture the essential structure of the input data as best as possible. Although, theoretically, NMF could be applied to data compression, not much work was carried out in this regard.

Rather, the computer vision community focused its attention to the application of NMF to pattern recognition applications, where the extracted NMF features are subsequently clustered or classified using classifiers. The NMF applications can be characterized according to several criteria. We provide the following application classes:

- *1D signal* applications (including sounds and EEG data), where the input matrix \mathbf{X} contains in its columns one-dimensional data varying over time.
- *2D signal* applications (face object images, etc.), where the input matrix \mathbf{X} contains in its columns a vectorized version of the 2D signals (basically 2D images) obtained by lexicographically concatenating the rows of the two-dimensional images.
- *Other applications*, including text or e-mail classification.

3.1 1D signal applications

The separation of pitched musical instruments and drums from polyphonic music is one application, where NMF was considered by Helén and Virtanen in [13]. The NMF splits the input data spectrogram into components which are further classified by an SVM to be associated to either pitched instruments or drums. Within this application, each column of the input matrix \mathbf{X} represents a short-time spectrum vector \mathbf{x}_t . The non-negative decomposition takes the form $\mathbf{x}_t = \sum_{i=1}^n \mathbf{s}_i a_{i,t}$, where \mathbf{s}_i is the spectrum of i -th component, $a_{i,t}$ is the gain of i -th component in frame t , and n is the component number. Individual musical instrument sounds extraction using NMF was exploited by Benetos et al. in [14]. A number of 300 audio files are used, corresponding to 6 different instrument classes (piano, violin, cello, flute, bassoon, and soprano saxophone) [15]. Two sorts of features are used to form the input matrix. The first feature set is composed of 9 audio specific features and MPEG-7 specifications (such as zero-crossing rate, delta spectrum, mel-frequency cepstral coefficients, etc). The second feature set is given by the rhythm pattern described by several other audio characteristics (such as power spectrum, critical bands, modulation amplitude, etc).

One particular NMF application is on spectral data analysis. Source spectra separation from magnetic resonance (MR) chemical shift imaging (CSI) of human brain using constrained NMF analysis was investigated by Sajda et al. [16]. In CSI, each tissue is characterized by a spectral profile or a set of profiles corresponding to the chemical composition of the tissue. In tumors, for instance, metabolites are heterogeneously distributed and, in a given voxel, multiple metabolites and tissue types may be present, so that the observed spectra are a combination of different constituent spectra. Consequently, the spectral amplitudes of the different coherent resonators are additive, making the application of NMF reasonable. The overall gain with which a tissue type contributes to this addition is proportional to its concentration in each voxel, such that \mathbf{X} is the observed spectra, the columns of \mathbf{W} represents concentration of the constituent materials, and the rows of \mathbf{H} comprises their corresponding spectra.

The spectral data analysis was also investigated in [17] by proposing a constraint NMF algorithm to extract spectral reflectance data from a mixture for space object identification. In this application, each observation of an object is stored as a column of a spectral trace matrix \mathbf{X} , while its rows correspond to different wavelengths. Each column of \mathbf{W} , called endmember, is a vector containing nonnegative spectral measurements along p different spectral bands, where each row of \mathbf{H} comprises the fractional concentration.

Multichannel EEG signals have been analyzed via NMF concept by Rutkowski et al. in [18]. The signals are firstly decomposed into intrinsic modes in order to represent them as a superposition of components with well defined instantaneous frequencies called IMF. The resulting trace IMF components form the input matrix \mathbf{X} , while \mathbf{W} is the mixing matrix containing the true sub-spectra.

3.2 Image applications

One of the first NMF applications in images is on face recognition tasks. Li et al [3] explored this issue for both NMF and LNMF techniques, when a simple Euclidean distance is used as classifier. Their experiments revealed the superiority of LNMF over the standard NMF for the ORL face database [19], especially for occluded faces. Guillaumet and Vitrià [20] also applied NMF to a face recognition task. A third framework dealing with the face recognition task is described in [21], where the DNMF is employed along NMF and LNMF for comparison. Also, besides the Euclidean distance, two other classifiers (cosine similarity measure and SVMs) are utilized. Two databases, namely ORL and YALE [22] are utilized here. The experiments showed that the NMF seems to be more robust to illumination changes than LNMF and DNMF, since the variation of illumination conditions for the faces pertaining to Yale database is much more intense than for images from the ORL database. Contrary to the

results obtained for ORL, where LNMF gave the highest recognition rate, when face recognition is performed on the YALE database, the best results are obtained by the NMF algorithm. Although the ORL database, generally, contains frontal faces or slightly rotated facial poses. This can contribute to the superior performance of LNMF, since this algorithm is rotation invariant (up to some degree), because it generates local features in contrast to NMF which yields more distributed features.

Buciu and Pitas applied NMF and LNMF for facial expression recognition in [23] and compared them with the DNMF algorithm in [4] for the same task. It was found that, for the facial expression recognition task, the DNMF method outperforms the other two techniques for the Cohn-Kanade AU-coded facial expression database [24].

The NMF for object recognition is investigated by Spratling [25], where an empirical investigation of the NMF performance with respect to its sparseness issue for occluded images is reported. The experiments were conducted for the standard bars problem, where the training data consists of 8×8 pixel images in which each of the 16 possible (one-pixel wide) horizontal and vertical bars can be present with a probability of $1/8$. The occlusion was simulated by overlapping between horizontal and vertical bars. Several NMF variants (i.e., NMF, LNMF, and NNSC) have been tested. It was found that no NMF method was able to identify all the components in the unequal bars (occlusion) problem for any value of the sparseness parameter. To overcome this situation, the author proposed a non-negative dendritic inhibition neural network, where the neural activations identified in the rows of \mathbf{H} reconstruct the input patterns \mathbf{X} via a set of feedforward (recognition) weights \mathbf{W} . When applied to face images, the proposed NMF neural network learns representations of elementary image features more accurately than other NMF variants. Guillaumet et al. experimentally compared NMF to the Principal Component Analysis (PCA) for image patch classification in [20]. In these experiments, 932 color images from the Corel Image database were used. Each of these images belongs to one of 10 different classes of image patches (clouds, grass, ice, leaves, rocky mountains, sand, sky, snow mountains, trees and water). NMF outperformed PCA. Finally, a face detection approach based on LNMF was proposed by Chen et al. in [26].

Genomic signal processing is another task where NMF recently done a good job. The approach has been used to discover metagenes and molecular patterns in [27]. NMF recovered meaningful biological information from cancer-related microarray data. NMF appears to have advantages over other methods such as hierarchical clustering or self-organizing maps. NMF was found less sensitive to a priori selection of genes or initial conditions and able to detect alternative or context-dependent patterns of gene expression in complex biological systems. This ability, similar to semantic polysemy in text, provides a general method for robust molecular pattern discovery. More recent, a robust NMF variant was proposed in [28]. A least-square NMF which incorporates uncertainty estimates is developed in [29], while factoring gene expressions using NMF was also utilized in [30] where a statistical sparse NMF was developed.

3.3 Other applications

The application of NMF for text classification was undertaken in [31]. This application is characterized by a large number of classes and a small training data size. In their formulation, the elements $w_{ik} \geq 0$ represent the confidence score of assigning the k -th class label to the i -th example. Furthermore, $\mathbf{H} = \mathbf{B}\mathbf{W}^T$, where the non-negative matrix \mathbf{B} captures the correlation (similarity) among different classes.

The extraction and detection of concepts or topics from electronic mail messages is a NMF application proposed by Berry and Browne in [32]. The input matrix \mathbf{X} contains n messages indexed by m terms (or keywords). Each matrix element $x_{i,j}$ defines a weighted frequency at which the term i occurs in message j . Furthermore, $x_{i,j}$ is decomposed as $x_{i,j} = l_{i,j}g_id_j$, where $l_{i,j}$ is the local weight for the term i occurring in message j , g_i is the global weight for i -th term in the subset, and d_j is a document normalization factor, which specifies whether or not the columns of \mathbf{X} (i.e., the documents) are normalized. Next, a normalized term $p_{i,j} = f_{i,j}/\sum_j f_{i,j}$ is defined, where $f_{i,j}$ denotes frequency that term i appears in the message j . Then, two possible definitions exist for $x_{i,j}$. The first one sets $l_{i,j} = f_{i,j}$, $g_{i,j} = 1$, while the second interpretation sets $l_{i,j} = \log(1 + f_{i,j})$ and $g_{i,j} = 1 + (\sum_j p_{i,j}\log(p_{i,j})/\log n)$, respectively. After NMF decomposition, the semantic feature represented by a given basis vector \mathbf{w}_k (k -th column of the matrix) by simply sorting (in descending order) its i elements and generating a list of the corresponding dominant terms (or keywords) for that feature. In turn, a given row of \mathbf{H} having n elements can be used to reveal messages sharing common basis vectors \mathbf{w}_k , i.e., similar semantic features or meaning. The columns of \mathbf{H} are the projections of the columns (messages) of \mathbf{X} onto the basis spanned by the columns of \mathbf{W} .

A chemometric application of the NMF method is proposed by Li et al. [33] where several NMF variants are used to detect chemical compounds from a chemical substances represented through Raman spectroscopy. The input matrix contains the observed total mixture chemical spectra, the basis vectors denote the contribution of chemical spectra, while the spectra is encoded into \mathbf{H} .

4 NMF Shortcomings

As far as the NMF open problems are concerned, several challenges exist, as follows:

- The optimization problem. All the described NMF variants suffer from the same drawback: no global minimum is guaranteed; they only lead to a local minimum, thus several algorithm runs may be necessary to avoid getting stuck in an undesired local minimum. Having an approach which conducts to a global minimum will greatly improve the numerical NMF stability.
- Initialization of \mathbf{H} and \mathbf{W} . Basically, the factors are initialized with random nonnegative values. A few efforts were undertaken in order to speed up the convergence of the standard NMF. Wild [34] proposed a spherical k -means clustering to initialize \mathbf{W} . More recently, Boutsidis and E. Gallopoulos [35] employed an SVD-based initialization, while Buciu et al. [36] constructed initial basis vectors, whose values are not randomly chosen but contain information taken from the original database. However, this issue is an open problem and needs further improvements for the standard NMF approach and its variants.
- Subspace selection. To date, there is no approach suggesting, a priori, the optimal choice of p for the best performances. The issue is difficult and data-dependent. Typically, the algorithms run for several values of p and the subspace dimension corresponding to the highest recognition rate is reported. Also, before data projection, the resulting basis vectors may be re-ordered according to some criteria (descending order of sparseness degree, discriminative capabilities, etc).
- Nonlinear nonnegative features. Standard NMF linearly decomposes the data. The kernel-based NMF approach proposed in [12] tends to retrieve nonlinear negative features.

References

- [1] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.
- [2] A. P. Dempster and N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data using the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, pp. 1–38, 1977.
- [3] S. Z. Li, X. W. Hou and H. J. Zhang, "Learning spatially localized, parts-based representation," *Int. Conf. Computer Vision and Pattern Recognition*, pp. 207–212, 2001.
- [4] I. Buciu and I. Pitas, "A new sparse image representation algorithm applied to facial expression recognition," in *Proc. IEEE Workshop on Machine Learning for Signal Processing*, pp. 539–548, 2004.
- [5] P. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.
- [6] A. Pascual-Montano, J. M. Carazo, K. Kochi, D. Lehmann, and R. D. Pascual-Marqui, "Nonsmooth nonnegative matrix factorization (nsNMF)," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, nr. 3, pp. 403–415, 2006.
- [7] Z. Yuan and E. Oja, "Projective nonnegative matrix factorization for image compression and feature extraction," *14th Scandinavian Conference on Image Analysis*, pp. 333–342, 2005.
- [8] Z. Chen and A. Cichocki, "Nonnegative matrix factorization with temporal smoothness and/or spatial decorrelation constraints," Preprint, 2005.
- [9] M. Mørup, K. H. Madsen, and L. K. Hansen, "Shifted Non-negative Matrix Factorization," *IEEE Workshop on Machine Learning for Signal Processing*, 2007.
- [10] S. S. Bucak, B. Gunsel, and O. Gursoy, "Incremental non-negative matrix factorization for dynamic background modelling," *ICEIS 8th International Workshop on Pattern Recognition in Information Systems*, 2007.
- [11] M. Mørup, L. K. Hansen, and S. M. Arnfred, "Algorithms for sparse higher order non-negative matrix factorization (HONMF)," *Technical Report*, 2006.
- [12] I. Buciu, N. Nikolaidis, and I. Pitas, "Non-negative matrix factorization in polynomial feature space," *IEEE Trans. on Neural Networks*, in Press, 2008.

- [13] M. Helén and T. Virtanen, "Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine," *13th European Signal Processing Conference*, 2005.
- [14] E. Benetos, C. Kotropoulos, T. Lidy, A. Rauber, "Testing supervised classifiers based on non-negative matrix factorization to musical instrument classification," in *Proc. of the 14th European Signal Processing Conference*, 2006.
- [15] Univ. of Iowa Musical Instrument Sample Database, <http://theremin.music.uiowa.edu/index.html>.
- [16] P. Sajda, S. Du, T. Brown, R. Stoyanova, D. Shungu, X. Mao, and L. Parra, "Non-negative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain," *IEEE Trans. on Medical Imaging*, vol. 23, no. 12, pp. 1453–1465, 2004.
- [17] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative Matrix Factorization for Spectral Data Analysis," *Linear Algebra and its Applications*, vol. 416, pp. 29–47, 2006.
- [18] T. M. Rutkowski, R. Zdunek, and A. Cichocki, "Multichannel EEG brain activity pattern analysis in time-frequency domain with nonnegative matrix factorization support," *International Congress Series*, vol. 1301, pp. 266–269, 2007.
- [19] <http://www.uk.research.att.com/>
- [20] D. Guillaumet and Jordi Vitrià, "Non-negative matrix factorization for face recognition," *Topics in Artificial Intelligence*, Springer Verlag Series: Lecture Notes in Artificial Intelligence, vol. 2504, pp. 336–344, 2002.
- [21] I. Buciu, N. Nikolaidis, and I. Pitas, "A comparative study of NMF, DNMF, and LNMF algorithms applied for face recognition," *2006 Second IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing*, 2006.
- [22] <http://cvc.yale.edu>
- [23] I. Buciu and I. Pitas, "Application of non-negative and local non negative matrix factorization to facial expression recognition," *International Conference on Pattern Recognition*, pp. 288–291, 2004.
- [24] T. Kanade, J. Cohn and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. IEEE Inter. Conf. on Face and Gesture Recognition*, pp. 46–53, 2000.
- [25] M. W. Spratling, "Learning image components for object recognition," *Journal of Machine Learning Research*, vol. 7, pp. 793–815, 2006.
- [26] X. Chen, L. Gu, S. Z. Li, and H.-J. Zhang, "Learning representative local features for face detection," *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1126–1131, 2001.
- [27] J. P. Brunet, P. Tamayo, T.R. Golub, and J. P. Mesirov, "Metagenes and molecular pattern discovery using matrix factorization", *Proc Natl Acad Sci U S A.*, vol. 101, no. 12, pp. 4164–4169, 2004.
- [28] P. Fogel, S. S. Young, D. M. Hawkins, and N. Lédirac, "Inferential, robust non-negative matrix factorization analysis of microarray data", *Bioinformatics*, vol. 23, no. 1, pp.44 – 49, 2007.
- [29] G. Wang, A. V. Kossenkova, and M. F. Ochs, "LS-NMF: a modified non-negative matrix factorization algorithm utilizing uncertainty estimates", *BMC Bioinformatics*, vol. 7, no. 1, pp. 175, 2006.
- [30] D. Dueck, Q. D. Morris and B. J. Frey, "Multi-way clustering of microarray data using probabilistic sparse matrix factorization", *Bioinformatics*, vol. 21, no. 1, pp. 144–151, 2005.
- [31] Y. Liu, R. Jin, and L. Yang, "Semi-supervised Multi-label Learning by Constrained Non-negative Matrix Factorization," in *Proc. of The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference*, vol. 21, pp. 421–426, 2006.
- [32] M. W. Berry and M. Browne, "Email Surveillance Using Nonnegative Matrix Factorization," *Computational & Mathematical Organization Theory*, vol. 11, pp. 249–264, 2005.
- [33] H. Li, T. Adali, W. Wang, D. Emge, A. Cichocki, "Non-negative matrix factorization with orthogonality constraints and its application to Raman spectroscopy," *Journal of VLSI Signal Processing*, vol. 48, no. 1-2, pp. 83–97, 2007.
- [34] S. Wild, "Seeding non-negative matrix factorization with the spherical k-means clustering," Master thesis, University of Colorado, 2002.

- [35] C. Boutsidis and E. Gallopoulos, "On SVD-based initialization for nonnegative matrix factorization," Tech. Rep. HPCLAB-SCG-6/08-05, University of Patras, Patras, Greece., 2005.
- [36] I. Buciu, N. Nikolaidis, and I. Pitas, "On the initialization of the DNMF algorithm," *IEEE International Symposium on Circuits and Systems*, pp. 4671–4674, 2006.

Ioan Buciu
University of Oradea
Faculty of Electrical Engineering and Information Technology
Department of Electronics
Address: Universitatii 1, 410087, Oradea, Romania
<http://webhost.uoradea.ro/ibuciu/>
E-mail: ibuciu@uoradea.ro