CONSTRAINED NON-NEGATIVE MATRIX FACTORIZATION METHOD FOR EEG ANALYSIS IN EARLY DETECTION OF ALZHEIMER DISEASE

Zhe Chen, Andrzej Cichocki and Tomasz M. Rutkowski

Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute
Wako-shi, Saitama 351-0198, Japan
{zhechen,cia,tomek}@brain.riken.jp

ABSTRACT

Approximate non-negative matrix factorization (NMF) is an emerging technique with a wide spectrum of potential applications in biomedical data analysis. In this paper, we proposed a new NMF algorithm with temporal smoothness constraint that aims to extract non-negative components that have meaningful physical or physiological interpretations. We propose two constraints and derive new multiplicative learning rules. Specifically, we apply the proposed algorithm, combined with advanced time-frequency analysis and machine learning techniques, to early detection of Alzheimer disease using clinical EEG recordings. Empirical results show promising performance.

1. INTRODUCTION

Many real-life data, such as the image pixels, stock indices, gene expressions, and power spectra, are known to be non-negative. Positive or nonnegative matrix factorization (NMF) is a powerful technique for analyzing such positive or nonnegative data [5-8]. Specifically, Lee and Seung's algorithms are simple yet elegant in that the multiplicative learning rules follow an expectation-maximization (EM) like procedure and monotonically decrease the cost function. While being viewed as a generative model, NMF is intrinsically linked to another popular generative models such as the independent component analysis (ICA). However, NMF differs from ICA in that the decomposed components from NMF are not necessarily independent and subject to nonnegative constraints; in addition, their different objective functions also lead to different optimization procedures. Recently, there are research efforts that attempt to enhance the NMF by exploiting further constraints apart from non-negativeness (e.g., [4,6,10]). In this paper, we study the NMF problem with temporal smoothness (with unit variance) and spatial decorrelation constraints. We propose several new regularized cost functions and derive corresponding learning rules; similar to the original NMF procedure, the learning rules are EM-like and guarantee monotonic convergence to local minima. Another contribution of this paper is to apply the proposed NMF algorithm for feature extraction in electroencephalographic (EEG) data for early detection of Alzheimer disease; we combine the NMF features with time-frequency analysis and machine learning techniques and achieve satisfactory performance in a two-class pattern classification problem.

2. NON-NEGATIVE MATRIX FACTORIZATION AND CONSTRAINED NMF ALGORITHMS

Consider a temporal linear mixing model as follows

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad \text{or} \quad \mathbf{x}_t = \mathbf{A}\mathbf{s}_t \quad (t = 1, 2, \dots, T) \tag{1}$$

where all entries of vectors $\mathbf{s}_t$ and a mixing matrix $\mathbf{A}$ are assumed to be non-negative. In the matrix form, (1) can be rewritten as

$$\mathbf{X} = \mathbf{A}\mathbf{S} \tag{2}$$

where $\mathbf{X} = [x_{it}] \in \mathbb{R}_+^{m\times T}$ is a non-negative data matrix, whose columns $\mathbf{x}_t$ are $m \times 1$ vectors with entries $x_{it} = x_i(t)$ and T denotes the total number of observations; $\mathbf{A} = [a_{ij}] \in \mathbb{R}_+^{m\times r}$ and $\mathbf{S} = [s_{it}] \in \mathbb{R}_+^{r\times T}$ are two factorized, non-negative matrices, with individual elements $a_{ij} \geq 0, s_{it} \geq 0$. In general, low-rank factorization ($r \leq m$) is used. Here, $\mathbf{A}$ can be viewed as a set of r basis vectors (of size $m \times 1$) associated with T observations in $\mathbf{S}$, for which a generative model $\mathbf{x}_t = \mathbf{A}\mathbf{s}_t$ is valid. The goal of the NMF is to find non-negative matrices $\mathbf{A}$ and $\mathbf{S}$ such that a predefined cost function is minimized. Obviously, the solution to this problem is generally non-unique; in practice, we wish to specify further constraints or impose prior knowledge to derive desired solutions.

2.1. Temporal smoothness constraint

Suppose each "factorized" temporal source (the row vector of the matrix $\mathbf{S}$) represents a temporal signal with length T, we then write the ith source as $s_i(t) \equiv s_{it}$ and the ith row vector of $\mathbf{S}$ as $\mathbf{s}_i$. When $s_i(t)$ is temporally locally smooth, its short-term variance is relatively small compared to a larger long-term variance [9]. Let $m_i = \frac{1}{T}\sum_t s_i(t)$ denote the mean value of $\{s_i(t)\}$ given the total T observations, and let $\overline{s}_i(t)$ denote the short-term exponentially weighted average of temporal signal $s_i(t)$, namely

$$\overline{s}_i(t) = \alpha\overline{s}_i(t-1) + (1-\alpha)s_i(t) \equiv \alpha\overline{s}_i(t-1) + \beta s_i(t) \tag{3}$$

where $0 < \alpha < 1$ is a forgetting factor that determines the local smoothness range, and $\beta = 1 - \alpha$. In particular, the temporal smoothness under consideration is measured by the ratio of short-term variance against long-term (i.e., with the complete data) variance in the temporal domain[1]

$$R = \log\frac{\sum_t (s_i(t) - \overline{s}_i(t))^2}{\sum_t (s_i(t) - m_i)^2} = \log\frac{\sum_t \alpha^2 (s_i(t) - \overline{s}_i(t-1))^2}{\sum_t (s_i(t) - m_i)^2} \tag{4}$$

[1] We do not simply minimize the "regular" variance of the temporal signal since the optimization will favor the output as a constant signal.

Hence, the smaller the ratio value R, the smoother is the temporal signal $s_i(t)$. In vector notation, let $\bar{\mathbf{s}}_i$ denote the short-term average vector corresponding to the row vector $\mathbf{s}_i = [s_{i1}, \ldots, s_{iT}]$; if we further constrain the variance of the row vector $\mathbf{s}_i$ as 1, then the minimization problem (4) may be equivalently rewritten as

$$R = \frac{1}{T}\|\mathbf{s}_i - \bar{\mathbf{s}}_i\|^2, \quad s.t. \;\; \mathrm{Var}[\mathbf{s}_i] = 1 \tag{5}$$

The unit variance constraint is imposed in order to simplify the minimization of (4); this trick is used in our algorithms described below. Now, we wish to represent $\bar{\mathbf{s}}_i$ in terms of $\mathbf{s}_i$. Note that the exponentially average $\bar{s}_i(t)$ is indeed the *convolution product* between $s_i(t)$ and a template operator. Suppose $\bar{s}_i(0) = 0$ and the template vector has an exponentially decreasing property with a length L (e.g., when $L = 5$, $template = [\beta, \alpha\beta, \alpha^2\beta, \alpha^3\beta, \alpha^4\beta]$; notably if $\alpha = 0.5$ then `sum`(*template*)=0.9688). The convolution operation can also be conveniently expressed as a matrix product operation: $\bar{\mathbf{s}}_i^T = \mathbf{T}\mathbf{s}_i^T$, where

$$\mathbf{T} = \begin{bmatrix} \beta & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \alpha\beta & \beta & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \alpha^2\beta & \alpha\beta & \beta & 0 & 0 & 0 & 0 & \cdots & 0 \\ \alpha^3\beta & \alpha^2\beta & \alpha\beta & \beta & 0 & 0 & 0 & \cdots & 0 \\ \alpha^4\beta & \alpha^3\beta & \alpha^2\beta & \alpha\beta & \beta & 0 & 0 & \cdots & 0 \\ 0 & \alpha^4\beta & \alpha^3\beta & \alpha^2\beta & \alpha\beta & \beta & 0 & \cdots & 0 \\ & & & & & & & \ddots & \vdots \\ 0 & 0 & \cdots & & & & & & \\ 0 & \cdots & & 0 & \alpha^4\beta & \alpha^3\beta & \alpha^2\beta & \alpha\beta & \beta \end{bmatrix},$$

that is, $\mathbf{T}$ is a $T \times T$ Toeplitz matrix with the right-shifted template appearing at each row. Then (5) can be rewritten as

$$R_i = \frac{1}{T}\|\mathbf{s}_i^T - \mathbf{T}\mathbf{s}_i^T\|^2 = \frac{1}{T}\|(\mathbf{I}-\mathbf{T})\mathbf{s}_i^T\|^2. \tag{6}$$

To achieve constrained NMF, two regularized cost functions are proposed here

$$J_1 = \sum_{i=1}^{m}\sum_{t=1}^{T}[x_{it} - (\mathbf{AS})_{it}]^2 + \lambda\sum_{i=1}^{r} R_i \tag{7}$$

$$J_2 = \sum_{i=1}^{m}\sum_{t=1}^{T}\Big(x_{it}\log\frac{x_{it}}{(\mathbf{AS})_{it}} - x_{it} + (\mathbf{AS})_{it}\Big) + \lambda\sum_{i=1}^{r} R_i \tag{8}$$

where λ is a small regularization coefficient that balances the trade-off between the reconstruction error and temporal smoothness constraint. In the below we state several theorems for optimizing the above-defined regularized cost functions.

Theorem 1 *Under the condition that λ is sufficiently small such that $\mathbf{A}^T\mathbf{AS}+\lambda\mathbf{SQ}$ is non-negative, where $\mathbf{Q} = \frac{1}{T}(\mathbf{I}-\mathbf{T})^T(\mathbf{I}-\mathbf{T})$ is a symmetric square matrix, the regularized cost function (7) is monotonically non-increasing under the learning rules*

$$s_{jt} \leftarrow s_{jt}\frac{(\mathbf{A}^T\mathbf{X})_{jt}}{(\mathbf{A}^T\mathbf{AS}+\lambda\mathbf{SQ})_{jt}}, \quad a_{ij} \leftarrow a_{ij}\frac{(\mathbf{XS}^T)_{ij}}{(\mathbf{ASS}^T)_{ij}} \tag{9}$$

where the alternating update between $\mathbf{S}$ and $\mathbf{A}$ is inserted by proper scaling of row vectors $\mathbf{s}_i$ ($i = 1, \ldots, r$) and $\mathbf{A}$ in order to assure each row vector of $\mathbf{S}$ to have a unit variance.[2] *The cost function is invariant under these updates if and only if $\mathbf{A}$ and $\mathbf{S}$ are at a stationary point of the distance metric.*

[2] Suppose for each row vector $\mathbf{s}_i$ is scaled by a coefficient γ_i to assure unit variance, then we need to rescale $\mathbf{A} \leftarrow \mathbf{A}\mathbf{\Gamma}^{-1}$ ($\mathbf{\Gamma} = \mathrm{diag}\{\boldsymbol{\gamma}_i\}$) before updating a_{ij}. See [2] for details.

Proof: For $0 < \alpha < 1$, it is obvious that $(\mathbf{I}-\mathbf{T})$ has positive diagonals and non-positive entries elsewhere. It is easy to verify that in the matrix $\mathbf{Q} = \frac{1}{T}(\mathbf{I}-\mathbf{T})^T(\mathbf{I}-\mathbf{T})$ only the diagonal values are positive (e.g., $0.333/T$ if $\alpha = 0.5$), whereas the rest of entries are non-positive. Since $\lambda > 0$, and $\mathbf{A}^T\mathbf{AS}$ and $\mathbf{S}$ are both non-negative, then $\mathbf{A}^T\mathbf{AS} + \lambda\mathbf{SQ}$ is also non-negative if the scaled off-diagonal elements of λq_{tk} is sufficiently small; this is indeed the case that was confirmed in our implementation. Similar to [6], we may verify that the update rules in (9) guarantee monotonic non-increasing convergence [2]. □

Theorem 2 *Given the above-defined Toeplitz matrix $\mathbf{T}$ and a small positive λ, the regularized cost function (8) is monotonically non-increasing under the learning rules*

$$s_{jt} \leftarrow \frac{-b+\sqrt{b^2-4ac}}{2a}, \tag{10}$$

$$where \;\; a = \lambda\sum_{k=1}^{T} q_{tk}, \; b = \sum_i a_{ij}, \; c = -\sum_i x_{it}\frac{a_{ij}s_{jt}}{\sum_j a_{ij}s_{jt}}$$

$$a_{ij} \leftarrow a_{ij}\frac{\sum_t s_{jt}x_{it}/(\mathbf{AS})_{it}}{\sum_t s_{jt}}, \tag{11}$$

where the alternating update between $\mathbf{S}$ and $\mathbf{A}$ is inserted by proper scalings of row vectors $\mathbf{s}_i$ ($i = 1, \ldots, r$) and $\mathbf{A}$ in order to assure each row vector of $\mathbf{S}$ to have a unit variance. The cost function is invariant under these updates if and only if $\mathbf{A}$ and $\mathbf{S}$ are at a stationary point of the distance metric.

Proof: This is done in the same line of analysis for Theorem 1, see [2] for detailed proof. Note that the non-negative constraint remains valid in (10): it is easy to verify that the denominator $a = \lambda\sum_k q_{tk} > 0$; namely, the column sum of each row of the matrix $\mathbf{Q}$ is positive; in addition, since c is non-positive, then $-4ac$ is non-negative and so is the nominator of (10). □

2.2. Spatial decorrelation constraint

The spatial decorrelation constraint implies the column vectors $\mathbf{s}_t$ are uncorrelated, and the sample correlation matrix $\mathbf{V} = \mathbf{SS}^T/T$ is *diagonal dominant* [6]. For this purpose, we propose to minimize the following regularized cost function (similar to [6,10])

$$\begin{aligned} J_3 \;\; = \;\; & \sum_{i=1}^{m}\sum_{t=1}^{T}\Big(x_{it}\log\frac{x_{it}}{(\mathbf{AS})_{it}} - x_{it} + (\mathbf{AS})_{it}\Big) \\ & + \frac{\lambda}{T}\sum_{i,j\neq i}^{m}(\mathbf{SS}^T)_{ij} - \frac{\lambda}{2T}\sum_{i=1}^{m}(\mathbf{SS}^T)_{ii}. \end{aligned} \tag{12}$$

Theorem 3 *Given a small positive λ, the regularized cost function (12) is monotonically non-increasing under the learning rule*

$$s_{jt} \;\; \leftarrow \;\; \frac{b-\sqrt{b^2-4ac}}{-2a} \tag{13}$$

$$where \qquad a = -\frac{\lambda}{T}, \; b = \sum_{i=1}^{m} a_{ij} + \frac{\lambda}{T}\sum_{i=1, i\neq j}^{r} s_{it},$$

$$c = -\sum_i x_{it}\frac{a_{ij}s_{jt}}{\sum_j a_{ij}s_{jt}}$$

followed by the same update equation (11). The cost function is invariant under these updates if and only if $\mathbf{A}$ and $\mathbf{S}$ are at a stationary point of the distance metric.

Proof: The proof is similar to that of Theorem 2, except the definition of the auxiliary function; we omit the proof [2] due to lack of space. Note that $b^2 \geq 4ac \geq 0$ is necessary to assure the square root to be real value; this assumption is often valid when λ is small and T is relatively large. $\square$

Remarks: When $b^2 \gg 4ac$, (13) may be approximated by the following learning rule

$$s_{jt} \approx \mu\sqrt{\frac{c}{a}} \triangleq \mu\sqrt{\frac{T}{\lambda}\sum_i x_{it}\frac{a_{ij}s_{jt}}{\sum_j a_{ij}s_{jt}}} \tag{14}$$

where $\mu > 0$ is a scaling parameter, whose effect will be wiped out by the normalization of a_{ij}. It is also possible to integrate the temporal smoothness and spatial decorrelation constraints together into one cost function, for example

$$\begin{aligned} J_4 \quad = \quad & \sum_{i=1}^{m}\sum_{t=1}^{T}[x_{it} - (\mathbf{AS})_{it}]^2 + \frac{\lambda_1}{T}\sum_{i=1}^{r}\|(\mathbf{I}-\mathbf{T})\mathbf{s}_i^T\|^2 \\ & + \frac{\lambda_2}{2T}\Big(2\sum_{i,j\neq i}^{m}(\mathbf{SS}^T)_{ij} - \sum_{i=1}^{m}(\mathbf{SS}^T)_{ii}\Big), \end{aligned}$$

the minimization of which will lead us to a new update rule for entries of $\mathbf{S}$:

$$s_{jt} \leftarrow s_{jt}\frac{(\mathbf{A}^T\mathbf{X})_{jt}}{(\mathbf{A}^T\mathbf{AS} + \lambda_1\mathbf{SQ})_{jt} + \frac{\lambda_2}{T}(\sum_{i=1,i\neq j}^{r} s_{it} - s_{jt})} \tag{15}$$

where the choices of λ_1 and λ_2 depend on the practical need regarding specific constraints. Practically, the monotonic convergence of (15) can also be satisfied if λ_1 and λ_2 are sufficiently small. A final note is that in order to prevent numerical problems in practice, whenever a negative or zero value appears in the update equations, we replace it with a small (say 10^{-9}) positive value.

3. SIMULATION RESULTS

Early detection of Alzheimer disease (AD) is a practically important yet challenging problem for evaluating human brain degeneration processing. In the following experiment, the EEG data [1] were collected from Japanese hospitals with three distinct groups: normal control subjects, AD patients, and MCI (mild cognitive impairment) patients with high risks for developing AD in 1 to 1.5 year. During recordings all subjects were in relaxed state. The goal here is to extract features that help discriminate between healthy age-matched control and MCI patients. The EEG recordings (200Hz for 20s, 21 channels in international 10/20 system) were initially preprocessed to remove the artifacts. Two strategies were employed for feature extraction. The first strategy works in the frequency domain. We start with an AMUSE-BSS algorithm [1] to extract 5 most significant spatiotemporal components for each subject's recording, which yields 6×4000 data points in temporal domain. Then the *Welch's method* is used to calculate the *non-negative* power spectra of the (bipolar) temporal signals, which are further normalized (i.e., the integration of power spectra equals 1). For illustration purpose, the normalized power spectra of two typical categories of data are shown in Fig. 1. Given the power spectra, we further apply the constrained NMF learning rule (15) (parameters $\alpha = 0.8, \lambda_1 = 0.1, \lambda_2 = 0.05$ were selected by many empirical experiments) to extract 5 smooth spectral components from the power spectra of each group. The results showed

Table 1. The categorized channel sets according to their spatial locations and relative importance given some prior knowledge (using 21-channel electrode cap of the standard 10/20 system[1]).

class I	left frontal	FP1, F3, F7
class II	right frontal	FP2, F4, F8
class III	occipital	O1, O2
class IV	parietal	P3, P4

that for MCI patients the relative peaks are higher in *Theta band* (4-7Hz) than the control subjects, whereas MCI also has more significant components in *Alpha band* (8-12Hz) and *Beta band* (13-30Hz, esp. $Beta_1$, 13-20Hz) compared to the normal case. It was found that the constrained NMF produced similar yet smoother components compared to the standard NMF.

The second strategy works in the time-frequency domain. It is known that Fourier-based estimate of power spectra is limited by the stationarity assumption of the EEG data. To overcome this drawback, we apply time-frequency analysis using Wigner-Ville distribution (WVD) to the same data. First, 21 time-frequency maps were calculated, each image (the magnitude of the map) is normalized in the time-frequency plane and therefore can be treated as a two-dimensional probability density function $P(t, \omega)$. Then the marginal spectrum $h(\omega) = \int_t P(t,\omega)dt$ is calculated, which yields a measure of energy distribution from each frequency band representing the cumulated amplitude over the entire data span in a probabilistic sense. Among the resulted $h(\omega)$ obtained from the selected channel sets (Table 1), several averaged energy (power) statistics are calculated across different frequency bands

$$\begin{aligned} \rho_1 \quad &= \quad h_\theta(\omega), \quad \text{(class I, II, III, IV)} \\ \rho_2 \quad &= \quad h_\alpha(\omega)/h_\theta(\omega), \quad \text{(class III, IV)} \end{aligned}$$

The choices of above statistics are motivated by the various reported clinical and psychiatric observations (e.g., [4]). The plots of some selected features ρ_1 vs ρ_2 are also shown in Fig. 1.

Here we investigate the early detection of AD and focus on the two-class (MCI vs. control) pattern classification problem. We first use the features extracted from the proposed constrained NMF algorithm alone. For each subject, 100 Monte Carlo runs (each with different random initialization) were conducted, the mean and variance of two features are calculated: the powers of Alpha and Beta bands (i.e., the area summed over specific bandwidths and then averaged over three power spectrum curves). The extracted features for all 60 subjects (38 for control and 22 for MCI) are shown in Fig. 2. At the first sight (by assuming the class labels are known for each subject), these two classes have quite distinct features, despite the obvious overlapping in some cases. Given the limited data at hand, we used a support vector machine (SVM) and leave-one-out cross-validation procedure (www.kyb.tuebingen.mpg.de/bs/people/spider/) to classify the two classes based on the extracted two-dimensional features (using the mean values out of 100 Monte Carlo runs). A Gaussian kernel is used for training the SVM with standard parameters setup $\sigma = 1.1, C = 100$. We obtained 13.3% misclassification rate (7 false negatives, 1 false positive), achieving a similar performance as the early results reported in [1] using the same data.

Second, we combined the above two features with the other two obtained from the time-frequency analysis, and fed the total four inputs into the SVM classifier. Using the same parameter setup, a slightly improved classification result was obtained (10%

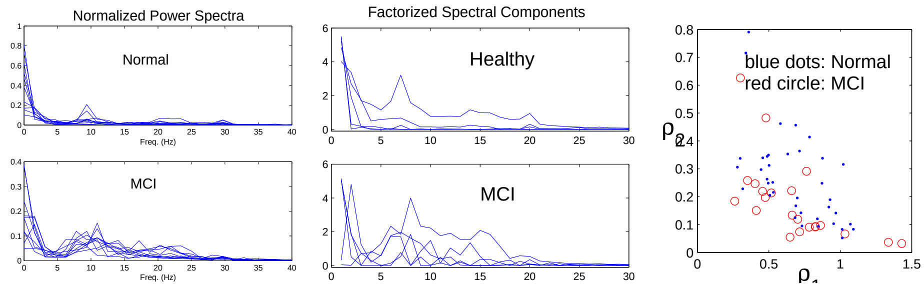

Fig. 1. *From left to right:* a) EEG normalized power spectra of two randomly selected normal control and MCI subject. b) Factorized spectral components. c) Feature statistics plots from time-frequency maps of WVD.

misclassification rate, 5 false negatives, 1 false positive). The classification boundary, however, cannot be visualized because of the high-dimensionality of the features. As observed, the classifier performance can be improved by incorporating more relevant features.

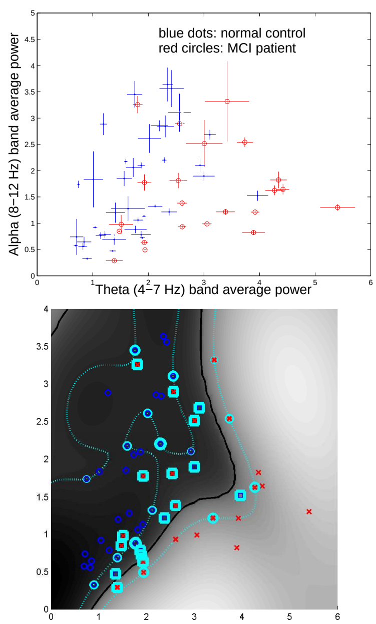

Fig. 2. *Top*: The scatter plot of control vs. MCI subjects' features projected on two-dimensional space. Two extracted features correspond to the averaged power in Theta band (abscissa) and Alpha band (ordinate). The central dots or circles correspond to mean values over 100 Monte Carlo runs, whereas the crossings represent the respective variances. *Bottom*: The decision boundary found by SVM (circled points represent the support vectors; squared points represent the misclassified points).

4. SUMMARY

Non-negative matrix factorization has become an increasingly popular tool for feature extraction. By enforcing temporal smoothness, unit variance, and/or spatial decorrelation constraints on the matrix $\mathbf{S}$, we derived constrained NMF rules for extracting physically meaningful temporal/contextural components, which may be potentially useful for a wide range of engineering and biomedical problems. We applied the proposed NMF algorithm, integrated with prior knowledge, time-frequency analysis, and machine learning techniques to the EEG signal analysis for early detection of Alzheimer disease. Given limited data at hand, our preliminary simulations show a promising result that deserves further investigation.

5. REFERENCES

[1] Cichocki, A., Shishkin, S. L., Musha, T., Leonowicz, Z., Asada, T., and Kurachi, T. (2005). EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease. *Clinic Neurophysiology, 116*(3), 729–737.

[2] Chen, Z. and Cichocki, A. (2005). Nonnegative matrix factorization with temporal smoothness and/or spatial decorrelation constraints. Tech. Rep. (Available at http://www.bsp.brain.riken.jp/~zhechen/download/nmf_TR.ps)

[3] Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research, 5*, 1457–1469.

[4] Knott, V., Mohr, E., Mahoney, C., and Ilivitsky, V. (2001). Quantitative electroencephalography in Alzheimer's disease: comparison with a control group, population norms and mental status. *J. Psychiatry Neuroscience, 26*, 2. 106–116.

[5] Lee, D. D. and Seung, H. S. (2000). Algorithms for nonnegative matrix factorization. in *Proc. NIPS2000*.

[6] Li, S. Z., Hou, X., Zhang, H., and Cheng, Q. (2001). Learning spatially localized, parts-based representation. *Proc. IEEE CVPR2001*, pp. 207–210.

[7] Sajda, P., Du, S., Brown, T., et al. (2004). Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain. *IEEE Trans. Medical Imaging, 23*(12), 1453–1465.

[8] Sha, F. and Saul, L. (2004). Real-time pitch determination of one or more voices by nonnegative matrix factorization. *Proc. NIPS2004*.

[9] Stone, J. V. (2001). Blind source separation using temporal predictability. *Neural Computation, 13*, 1559–1574.

[10] Wang, Y., Jia, Y., Hu, C. and Turk, M. (2005). Non-negative matrix factorization framework for face recognition. *Int. J. Pattern Recognition and Artificial Intellligence, 19*(4), 495–511.