SHIFTED NON-NEGATIVE MATRIX FACTORIZATION

Morten Mørup, Kristoffer H. Madsen and Lars K. Hansen

Technical University of Denmark
Informatics and Mathematical Modelling
Richard Petersens Plads, Building 321
DK-2800 Kgs. Lyngby, Denmark
email: {mm,khm,lkh}@imm.dtu.dk

ABSTRACT

Non-negative matrix factorization (NMF) has become a widely used blind source separation technique due to its part based representation and ease of interpretability. We currently extend the NMF model to allow for delays between sources and sensors. This is a natural extension for spectrometry data where a shift in onset of frequency profile can be induced by the Doppler effect. However, the model is also relevant for biomedical data analysis where the sources are given by compound intensities over time and the onset of the profiles have different delays to the sensors. A simple algorithm based on multiplicative updates is derived and it is demonstrated how the algorithm correctly identifies the components of a synthetic data set. Matlab implementation of the algorithm and a demonstration data set is available from [1].

1. INTRODUCTION

Non-negative matrix factorization (NMF) is a blind source separation algorithm (BSS) given by the decomposition

$$\mathbf{V}_{n,m} = \sum_d \mathbf{W}_{n,d}\mathbf{H}_{d,m} + \mathbf{E}_{n,m}, \qquad (1)$$

where $\mathbf{V} \in \mathbb{R}_+^{N\times M}$, $\mathbf{W} \in \mathbb{R}_+^{N\times D}$ and $\mathbf{H} \in \mathbb{R}_+^{D\times M}$, i.e. such that the variables $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{H}$ are non-negative while $\mathbf{E}$ is noise. The decomposition is useful as it results in easy interpretable part based representations [2]. Non-negative decompositions is also named positive matrix factorization [3] but was popularized by Lee and Seung due to a simple algorithmic procedure based on multiplicative updates [4]. The decomposition has proven useful for a wide range of data where non-negativity is a natural constraint. These encompass data for text-mining based on counts, image data, biomedical data and spectral data. The algorithm can also be useful even when the data in itself is negative by considering the amplitude of a spectral representation [5]. Presently, we will advance the non-negative matrix factorization to incorporate delays between sources and sensors based on the following shifted non-negative matrix factorization (ShiftNMF) model

$$\mathbf{V}_{n,m} = \sum_d \mathbf{W}_{n,d}\mathbf{H}_{d,m-\boldsymbol{\tau}_{n,d}} + \mathbf{E}_{n,m}, \qquad (2)$$

where $\boldsymbol{\tau}_{n,d}$ denotes an arbitrary delay from the d^{th} source to the n^{th} sensor. Notice, the above model can be considered a special case of the Non-negative Matrix Factor Deconvolution (NMFD) introduced in [6] where only one delay is present between each sensor and source and $\tau_{n,d}$ takes integer values. In figure 1 is illustrated the impact of the delays for equally mixed sources into 9 sensors. In figure 2 is shown an example of how a given $\mathbf{W}$, $\mathbf{H}$ and $\boldsymbol{\tau}$ generates the data $\mathbf{V}$, see figure 3. Notice, the above model without non-negative constraints has previously been treated in the literature. In [7, 8] a procedure based on integer shifts was derived while Bell and Sejnowski [9] sketched how to handle time delays in networks based on a similar model. This was further explored in [10]. In [11] a model based on equally mixed sources, i.e. $\mathbf{W} = \mathbf{1}$ (a matrix of ones), formed by moving averages incorporated non-integer delays by signal interpolation. Yeredor [12] solved the model by joint diagonalization of the source cross spectra based on the AC-DC algorithm with non-integer shifts for the 2×2 system. This approach was extended to complex signals in [13]. In [14] we derived an algorithm for shifted sources based on the estimation of a shift-invariant subspace while rotating and shifting the sources found using maximum likelihood ICA to achieve unique solutions. However, despite the recent popularity of NMF, to our knowledge, delays have not previously been treated for models with non-negative constraints. Thus, the aim of the current paper: To form an algorithm for ShiftNMF

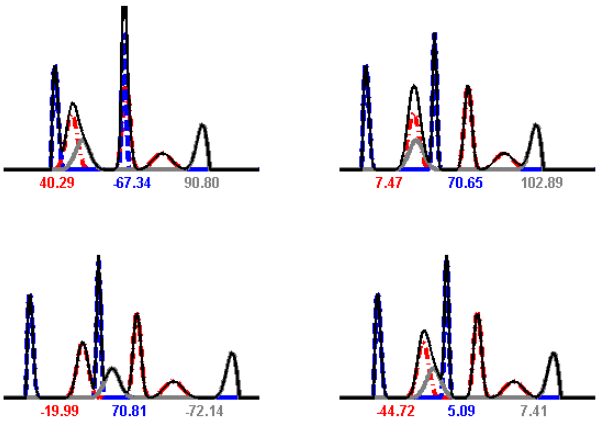

Fig. 1. Example of activities obtained (black graph) when summing three non-negative components (gray, blue dashed and red dash-dotted graphs) each shifted to various degrees (given in samples by the colored numbers). Clearly, the resulting activities are heavily impacted by the shifts such that a regular instantaneous NMF analysis would be inadequate.

based on multiplicative updates similar to the simple updates of regular NMF.

Recently, NMF has proven useful in the analysis of magnetic resonance spectra [15] and for analyzing data obtained using astronomical spectrometers for the identification and classification of space objects [16]. NMF has also been used in the analysis of fluorescence spectra [17]. For these types of data, ShiftNMF is potentially useful since shift in the spectral profiles often occurs for instance as a result of the Doppler effect. Furthermore, NMF has proven useful for extracting intensity profiles over time from biomedical data such as PET imaging [18, 19, 20]. Here, potential delays in the onset of the various profiles can be handled by the ShiftNMF model. Consequently, ShiftNMF might be useful for a wide range of data where NMF has previously been applied.

2. METHOD AND RESULTS

2.1. Notation

In the following $\mathbf{U}$ will denote a matrix in the time domain, while $\widetilde{\mathbf{U}}$ denotes the corresponding matrix in the frequency domain. $\mathcal{U}$ and $\widetilde{\mathcal{U}}$ denotes 3-way arrays in the time and frequency domains respectively. Furthermore, $\widetilde{\mathbf{U}}^H$ denotes the conjugate transpose of $\widetilde{\mathbf{U}}$ while $\mathbf{U} \bullet \mathbf{V}$ denotes the direct product, i.e. element-wise multiplication. Also, $\omega = 2\pi\frac{f-1}{M}$ and $\widetilde{\mathbf{U}}^{(f)} = \mathbf{U} \bullet e^{-i2\pi\frac{f-1}{M}\boldsymbol{\tau}}$ where $e^{-i2\pi\frac{f-1}{M}\boldsymbol{\tau}}$ denotes element wise raising the elements, i.e.$(e^{-i2\pi\frac{f-1}{M}\boldsymbol{\tau}})_{n,d} = e^{-i2\pi\frac{f-1}{M}\boldsymbol{\tau}_{n,d}}$. Finally, let $\mathbf{U}_d$ denote the d^{th} column, $\mathbf{U}_{n,:}$ the n^{th} row and $\mathbf{U}_{n,d}$ a given element of $\mathbf{U}$.

2.2. Multiplicative updates

Multiplicative updates were introduced in [2, 4] to solve the NMF model. Given a cost function $C(\mathbf{H})$ over the non-negative variables $\mathbf{H}$, define $\frac{\partial C(\mathbf{H})^+}{\partial \mathbf{H}_{d,m}}$ and $\frac{\partial C(\mathbf{H})^-}{\partial \mathbf{H}_{d,m}}$ as the positive and negative part of the derivative with respect to $\mathbf{H}_{d,m}$. Then the multiplicative update has the following form

$$\mathbf{H}_{d,m} \leftarrow \mathbf{H}_{d,m} \left(\frac{\frac{\partial C(\mathbf{H})^-}{\partial \mathbf{H}_{d,m}}}{\frac{\partial C(\mathbf{H})^+}{\partial \mathbf{H}_{d,m}}} \right)^{\alpha}. \tag{3}$$

A small constant $\varepsilon = 10^{-9}$ is added to the numerator and denominator to avoid division by zero or forcing $\mathbf{H}_{d,m}$ to zero. If the gradient is positive $\frac{\partial C(\mathbf{H})^+}{\partial \mathbf{H}_{d,m}} > \frac{\partial C(\mathbf{H})^-}{\partial \mathbf{H}_{d,m}}$, hence, $\mathbf{H}_{d,m}$ will decrease and vice versa if the gradient is negative. Thus, there is a one-to-one relation between fixed points and the gradient being zero. The attractive property of multiplicative updates is that they automatically ensure non-negativity as the updates is based on multiplication, division and raising to the power of purely non-negative variables. α is a "step size" parameter that potentially can be tuned. Notice, when $\alpha \to 0$ only very small steps in the negative gradient direction are taken. In [4] it was proven that for the least squares error and Kullback-Leibler divergence $\alpha = 1$ will keep decreasing the cost function.

2.3. Algorithm for ShiftNMF

Consider the ShiftNMF model as stated in both the time and frequency domain (using the DFT)

$$\mathbf{V}_{n,m} = \sum_d \mathbf{W}_{n,d}\mathbf{H}_{d,m-\boldsymbol{\tau}_{n,d}} + \mathbf{E}_{n,m}, \tag{4}$$

$$\widetilde{\mathbf{V}}_{n,f} = \sum_d \mathbf{W}_{n,d}\widetilde{\mathbf{H}}_{d,f}e^{-i2\pi\frac{f-1}{M}\boldsymbol{\tau}_{n,d}} + \widetilde{\mathbf{E}}_{n,f}. \tag{5}$$

In matrix notation the model is in the frequency domain stated as

$$\widetilde{\mathbf{V}}_f = \widetilde{\mathbf{W}}^{(f)}\widetilde{\mathbf{H}}_f + \widetilde{\mathbf{E}}_f. \tag{6}$$

We focus here on minimizing the least squares error

$$\begin{aligned} C_{LS}(W,H) &= \frac{1}{2}\sum_{n,m}(\mathbf{V}_{n,m} - \sum_d \mathbf{W}_{n,d}\mathbf{H}_{d,m-\boldsymbol{\tau}_{n,d}})^2 \\ &= \frac{1}{2M}\|\widetilde{\mathbf{V}}_f - \widetilde{\mathbf{W}}^{(f)}\widetilde{\mathbf{H}}_f\|_F^2. \end{aligned}$$

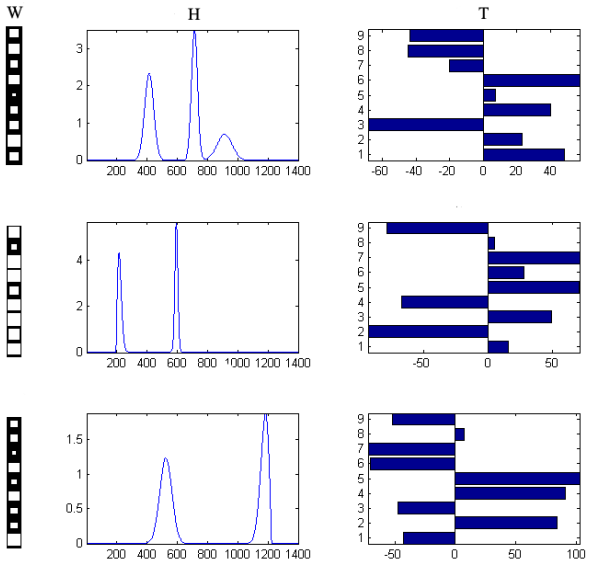

Fig. 2. The true factors forming a synthetic data set. To the left, the strength of the mixing $\mathbf{W}$ of each source is indicated in gray color scale. In the middle, the three sources are shown and to the right is given the time delays of each source to each channel.

The algorithm will be based on alternatingly solving for $\mathbf{W}$, $\mathbf{H}$ and $\boldsymbol{\tau}$.

*2.3.1. **W** update:*

Let $\widetilde{\mathbf{H}}_{d,f}^{(n)}$ denote the delayed version of the source signal $\widetilde{\mathbf{H}}_{d,f}$ to the n^{th} channel, i.e. $\widetilde{\mathbf{H}}_{d,f}^{(n)} = \widetilde{\mathbf{H}}_{d,f} e^{-i2\pi \frac{f-1}{M} \boldsymbol{\tau}_{n,d}}$. Then equation 4 can be restated as

$$\mathbf{V}_{n,:} = \mathbf{W}_{n,:}\mathbf{H}^{(n)} + \mathbf{E}_{n,:}, \tag{7}$$

Notice, since $\mathbf{H}^{(n)}$ corresponds to $\mathbf{H}$ where each source has been shifted a given amount $\mathbf{H}^{(n)}$ is still non-negative. Thus, this is the regular NMF problem which can be solved by the least squares NMF-update as given in [4]

$$\mathbf{W}_{n,d} = \mathbf{W}_{n,d} \frac{\mathbf{V}_{n,:}\mathbf{H}_{d,:}^{(n)^T}}{\mathbf{W}_{n,:}\mathbf{H}^{(n)}\mathbf{H}_{d,:}^{(n)^T}}. \tag{8}$$

Thus, the $\mathbf{W}$ update follows straight-forward from the update of regular NMF and will keep decreasing the cost function for $\alpha = 1$. The $\mathbf{H}$ update is however slightly more complicated.

*2.3.2. **H** update:*

Consider the model as stated in the frequency domain given in equation 6. Calculating the gradient of the least squares cost function in the frequency domain gives [21]

$$\mathbf{G}_f = \frac{\partial C_{LS}}{\partial \widetilde{\mathbf{H}}_f} = -\frac{1}{M} \widetilde{\mathbf{W}}^{(f)^H} (\widetilde{\mathbf{X}}_f - \widetilde{\mathbf{W}}^{(f)} \widetilde{\mathbf{H}}_f). \tag{9}$$

By taking the inverse DFT of the gradient in the frequency domain the corresponding gradient in the time domain is obtained. Splitting the gradient in the frequency domain into what constitutes the positive and negative part of the corresponding gradient in the time-domain gives

$$\widetilde{\mathbf{G}}_f^+ = \frac{1}{M} \widetilde{\mathbf{W}}^{(f)^H} \widetilde{\mathbf{W}}^{(f)} \widetilde{\mathbf{H}}_f, \tag{10}$$

$$\widetilde{\mathbf{G}}_f^- = \frac{1}{M} \widetilde{\mathbf{W}}^{(f)^H} \widetilde{\mathbf{X}}_f. \tag{11}$$

Consequently, by taking the inverse DFT of $\widetilde{\mathbf{G}}_f^+$ and $\widetilde{\mathbf{G}}_f^-$ the corresponding positive and negative part of the gradient in the time-domain can be found. As a result, $\mathbf{H}$ can be updated using multiplicative updates as

$$\mathbf{H}_{d,n} = \mathbf{H}_{d,n} \left(\frac{\mathbf{G}_{d,n}^-}{\mathbf{G}_{d,n}^+} \right)^{\alpha}. \tag{12}$$

Due to interpolation $\mathbf{G}_{d,n}^-$ and $\mathbf{G}_{d,n}^+$ can potentially take small negative values. In these rare cases these values are for stability reasons treated as zero. Since the gradient is interpolated through the DFT the step size α has to be tuned in order to guarantee that the cost function decreases.

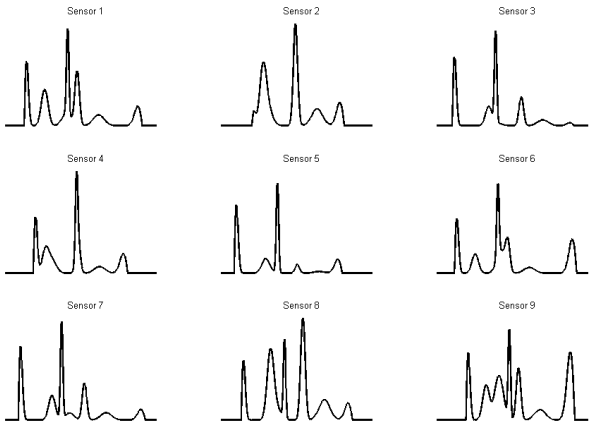

Fig. 3. The synthetic dataset generated from the factors given in figure 2 when mixed into each of the 9 sensors.

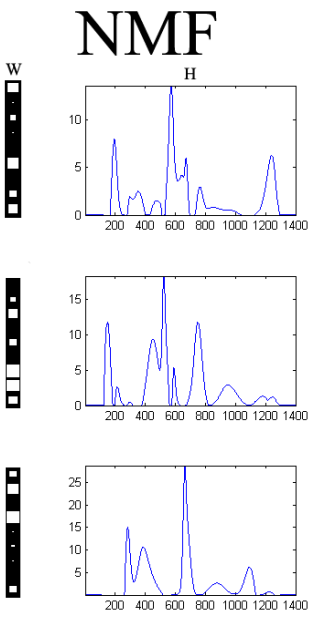

Fig. 4. Results obtained by regular instantaneous NMF for the synthetic data given in figure 3. To the left, the strength of the mixing $\mathbf{W}$ of each source is indicated in gray color scale. In the middle, the three sources are shown. Clearly, since the model can't account for the shifts in the data the sources estimated are mixtures of the 3 true sources. Notice, only 68 % of the variance of the data is accounted for.

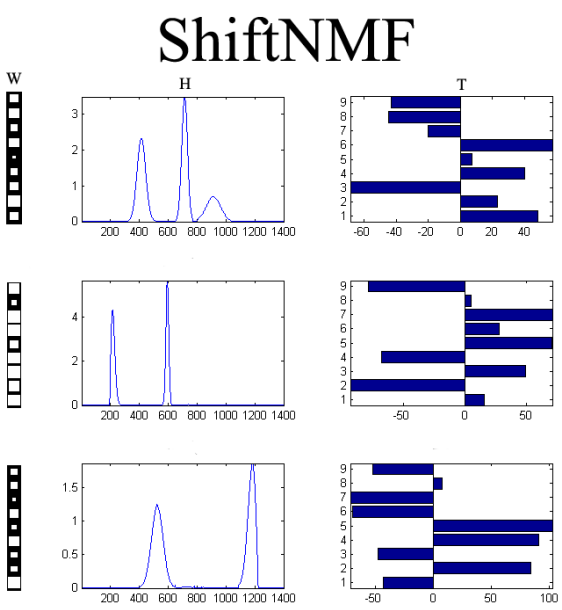

Fig. 5. The estimated factors obtained by a ShiftNMF analysis of the synthetic data given in figure 3. To the left, the strength of the mixing $\mathbf{W}$ of each source. In the middle, the three sources are shown and to the right is given the time delays of each source to each channel. Clearly, the model with shifts has correctly recovered the components of the synthetic data hence accounts for all the variance in the data.

2.3.3. $\boldsymbol{\tau}$ update:

The delays $\boldsymbol{\tau}$ are unconstrained. Consequently, we will estimate these by the Newton-Rhapson method as also proposed in [14]. The least square error for the Shift-NMF model as stated in equation 7, is given by

$$C_{LS} = \frac{1}{2M}\sum_f (\widetilde{\mathbf{V}}_f - \widetilde{\mathbf{W}}^{(f)}\widetilde{\mathbf{H}}_f)^H(\widetilde{\mathbf{X}}_f - \widetilde{\mathbf{W}}^{(f)}\widetilde{\mathbf{H}}_f).$$

Define $\mathbf{T}^{ND\times 1} = vec(\boldsymbol{\tau})$, i.e. the vectorized version of the matrix $\boldsymbol{\tau}$ such that $\mathbf{T}_{n+(d-1)N} = \boldsymbol{\tau}_{n,d}$. Let further

$$\widetilde{\mathcal{Q}}_{n,d,f} = \widetilde{\mathbf{W}}^{(f)}_{n,d}\widetilde{\mathbf{H}}_{d,f}, \quad \widetilde{\mathbf{E}}_f = \widetilde{\mathbf{V}}_f - \widetilde{\mathbf{W}}^{(f)}\widetilde{\mathbf{H}}_f. \tag{13}$$

Then the gradient of C_{LS} with respect to $\boldsymbol{\tau}_{n,d}$ is given as

$$\mathbf{g}_{n+(d-1)N} = \frac{\partial C_{LS}}{\partial \mathbf{T}_{n+(d-1)N}} = \frac{\partial C_{LS}}{\partial \boldsymbol{\tau}_{n,d}} \tag{14}$$

$$= \frac{-1}{M}\sum_f 2\omega\Im[\widetilde{\mathcal{Q}}_{n,d,f}\widetilde{\mathbf{E}}^*_{n,f}] \tag{15}$$

The Hessian has the following structure

$$\mathbf{B}_{t,t'} = \begin{cases} \frac{-2}{M}\sum_f \omega^2\Re[\widetilde{\mathcal{Q}}_{n,d,f}\widetilde{\mathcal{Q}}^*_{n',d',f}] & \text{if } n \neq n' \wedge d \neq d' \\ \frac{-2}{M}\sum_f \omega^2\Re[\widetilde{\mathcal{Q}}_{n,d,f}(\widetilde{\mathcal{Q}}^*_{n',d',f} + \widetilde{\mathbf{E}}^*_{n',f})] & \text{if } n = n' \wedge d = d' \end{cases}$$

where $t = n + (d-1)N$ and $t' = n' + (d'-1)N$. As a result, $\boldsymbol{\tau}$ can be estimated by the Newton-Raphson method as

$$\mathbf{T} \leftarrow \mathbf{T} - \eta\mathbf{B}^{-1}\mathbf{g}, \tag{16}$$

where η is a step size parameter that is tuned to keep decreasing the cost function.

The above iterative update is sensitive to local minima. However, we found that estimating the delays by the following cross-correlation procedure reduced the effect of local minima. Let

$$\widetilde{\mathbf{R}}_{n,f} = \widetilde{\mathbf{V}}_{n,f} - \sum_{d\neq d'}\widetilde{\mathbf{W}}^{(f)}_{n,d}\widetilde{\mathbf{H}}_{d,f}, \tag{17}$$

i.e. the signal at the n^{th} sensor at frequency f when projecting all but the d' source out of $\widetilde{\mathbf{V}}$. Then the cross-correlation between the d' source and n^{th} sensor is given as $\widetilde{\mathbf{c}}_f = \widetilde{\mathbf{R}}^*_{n,f}\widetilde{\mathbf{H}}_{d',f}$, such that $\boldsymbol{\tau}_{n,d}$ can be estimated as

$$t = \arg\max_m \mathbf{c}_m, \quad \boldsymbol{\tau}_{n,d'} = t - (M+1). \tag{18}$$

I.e. as the delay corresponding to maximum cross-correlation between the sensor and source. The value

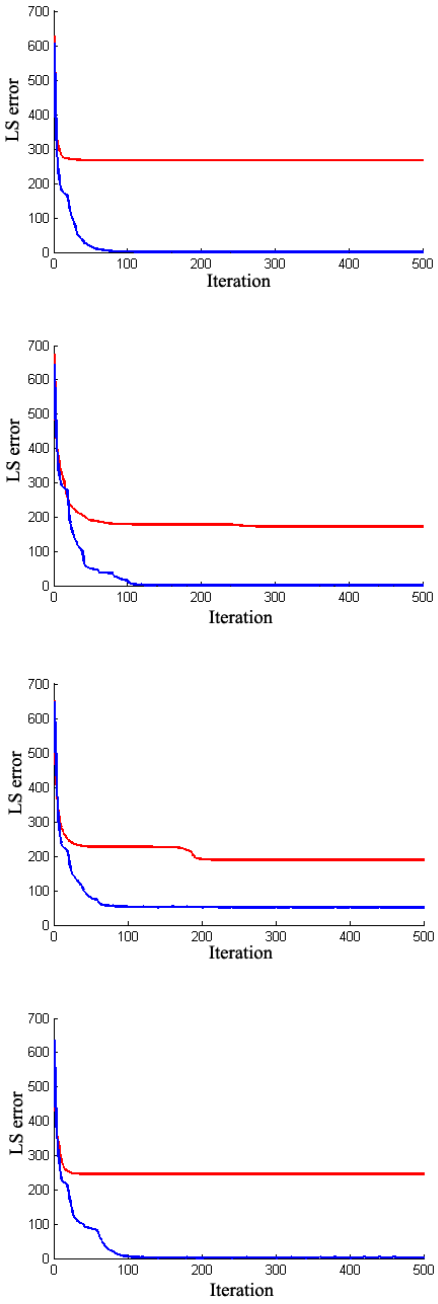

Fig. 6. Four runs illustrating how using the cross-correlation (CC) procedure every 20 iteration in combination with the Newton-Raphson (NR) update (blue curve) rather than relying solely on the NR update to estimate the delays (red curve) improves the algorithm from getting stuck in suboptimal solutions. Whereas the NR method alone does not identify the global minima, this is found in 3 out of the 4 trials when re-estimating the delays using the CC procedure. Notice, how the sudden jumps in the curves using the CC procedure are initiated at the iterations where the cross-correlation was used to re-estimate the delays.

of $\mathbf{W}_{n,d'}$ corresponding to this delay is given by

$$\mathbf{W}_{n,d'} = \frac{\mathbf{c}_t}{\mathbf{H}_{d',:}\mathbf{H}_{d',:}^T}. \tag{19}$$

Thus, to avoid being stuck in suboptimal solutions $\boldsymbol{\tau}$ was re-estimated by the cross-correlation procedure above every 20^{th} iteration. In figure 6 is demonstrated how indeed using the above cross-correlation approach improves the algorithm in finding the global optimum.

In figure 4 is shown the components found using a regular instantaneous NMF analysis. In figure 5 the results obtained by the ShiftNMF algorithm.

3. DISCUSSION

Clearly, the ShiftNMF algorithm correctly identified the components of the synthetic data set (see figure 5) whereas the regular instantaneous NMF failed to identify the correct component since the delays could not be accounted for (see figure 4). Furthermore, as shown in figure 6 using the cross-correlation approach as described in the section on the $\boldsymbol{\tau}$ update improved the component identifiability and avoided the algorithm to get stuck in many of the local minima.

Presently, the data had a unique decomposition since it was based on sparse sources, i.e. sources with many zero values. However, neither the NMF model [22] nor the ShiftNMF model is in general unique. In these situations additional constraints such as sparseness [23, 24] have proven useful. Furthermore, prior information such as smoothness has also been proposed to improve the component identification [16]. The present algorithm for ShiftNMF can straight forward be extended to incorporate these constraints in order to improve the identification where the model in general is not guaranteed to find a unique decomposition.

The DFT is based on the implicit assumption that the signals are periodic. In general, this is not the case. However, by zero padding the ends or introducing a window function periodicity can be enforced. Both windowing and zero padding will favor small delays and particularly windowing is also computationally expensive.

Presently, we derived an algorithm for blind (unsupervised) identification of sources and their respective mixing and delays. However, when analyzing for instance spectra of mixed source profiles the spectrum of the sources $\mathbf{H}$ are often known a priori. By keeping $\mathbf{H}$ fixed the algorithm presently derived can be used supervised. Hence, to estimate how much the various known sources are present in each sensor as well as to what extent the sources have been delayed to the

sensors. Future work should investigate the potential usefulness of this.

The present work investigated synthetic data. Current work focus on real data applications. It is our belief the algorithm derived will be useful for a range of data where NMF previously has been employed. Here accounting for delays can potentially improve component identification. The approach used to form the multiplicative updates by finding the positive and negative parts of the gradient using the spectral representation is valid due to Parseval's identity stating a correspondence between the cost function in the time and frequency domain. This correspondence does not exist for cost functions such as the Kullback-Leibler divergence. Thus, future work should investigate how the ShiftNMF model can be estimated for other cost functions than least squares. However, for least squares minimization the technique is applicable to other types of algorithms such as projected gradient [25] as well as the deconvolution extension (NMFD) [6]. These topics should be investigated in future work.

4. REFERENCES

[1] Morten Mørup and Kristoffer H. Madsen, "Shift nmf algorithm," http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/5232/zip/imm5232.zip.

[2] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–91, 1999.

[3] P Paatero and U Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.

[4] Daniel D Lee and H. Sebastian Seung, "Algorithms for non-negative matrix factorization," in *NIPS*, 2000, pp. 556–562.

[5] Paris Smaragdis and Judith C. Brown, "Non-negative matrix factorization for polyphonic music transcription," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 177–180, October 2003.

[6] Paris Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sourses from monophonic inputs," *International Symposium on Independent Component Analysis and Blind Source Separation (ICA)*, vol. 3195, pp. 494, sep 2004.

[7] R.A. Harshman, S. Hong, and M.E. Lundy, "Shifted factor analysisŮpart i: Models and properties," *Journal of Chemometrics*, vol. 17, pp. 363–378, 2003.

[8] R.A. Harshman, S. Hong, and M.E. Lundy, "Shifted factor analysisŮpart ii: Algorithms," *Journal of Chemometrics*, vol. 17, pp. 379–388, 2003.

[9] Anthony J. Bell and Terrence J. Sejnowski, "An information maximization approach to blind source separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[10] K Torkkola, "Blind separation of delayed sources based on informationmaximization," *Acoustics, Speech, and Signal Processing. ICASSP-96*, vol. 6, pp. 3509–3512, 1996.

[11] Bruno Emile and Pierre Comon, "Estimation of time delays between unknown colored signals," *Signal Processing*, vol. 68, no. 1, pp. 93–100, 1998.

[12] Arie Yeredor, "Time-delay estimation in mixtures," *ICASSP*, vol. 5, pp. 237–240, 2003.

[13] Arie Yeredor, "Blind source separation in the presence of doppler frequency shifts," *ICASSP*, vol. 5, pp. 277–280, 2005.

[14] Morten Mørup, Kristoffer H. Madsen, and Lars K. Hansen, "Shifted independent component analysis," *to appear, ICA2007*, 2007.

[15] Shuyan Du, Paul Sajda, Radka Stoyanova, and Truman. R. Brown, "Recovery of metabolomic spectral sources using non-negative matrix factorization," *Proceedings of the 2005 IEEE Engineering in Medicine and Biology*, pp. 4731– 4734, 2005.

[16] V. Paul Pauca Pauca, J. Piper, and Robert J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra and its Applications*, vol. 416 (2006), pp. 29–47, 2006.

[17] Cyril Gobinet, Eric Perrin, and Régis Huez, "Application of non-negative matrix factorization to fluorescence spectroscopy," *EUSIPCO*, pp. 1095–1098, 2004.

[18] J. S. Lee Lee, D. D. Lee, S. Choi, K. S. Park, and D. S. Lee, "Non-negative matrix factorization of dynamic images in nuclear medicine," *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 2027–2030, 2001.

[19] Kyeong Min Kim, H Watabe, M Shidahara, Ji Young Ahn, Seungjin Choi, N Kudomi, K Hayashida, Y Miyake, and H Iida, "Noninvasive estimation of cerebral blood flow using image-derived carotid input function in $h_2^{15}o$ dynamic pet," 2001, vol. 3, pp. 1282–1285.

[20] B. Bödvarsson, L. K. Hansen, C. Svarer, and G. M. Knudsen, "NMF on positron emission tomography," in *International conference on acoustics, speech and signal processing 2007, ICASSP.*, 2007.

[21] Kaare B. Petersen and Michael S. Pedersen, "The matrix cookbook," www.matrixcookbook.com, 2007.

[22] David Donoho and Victoria Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?," *NIPS*, 2003.

[23] P.O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, 2004.

[24] J. Eggert and E. Korner, "Sparse coding and nmf," in *Neural Networks*, 2004, vol. 4, pp. 2529–2533.

[25] C.-J. Lin, "Projected gradient methods for non-negative matrix factorization," *To appear in Neural Computation*, 2007.