2014 International Symposium on Nonlinear Theory and its Applications
NOLTA2014, Luzern, Switzerland, September 14-18, 2014

On monotonicity of multiplicative update rules for weighted nonnegative tensor factorization

Alexey Ozerov, Ngoc Q. K. Duong and Louis Chevallier

technicolor
975 avenue des Champs Blancs, CS 17616, 35576 Cesson Sévigné, France
Email: {alexey.ozerov, quang-khanh-ngoc.duong, louis.chevallier}@technicolor.com

Abstract—This paper focuses on so-called *weighted* variants of nonnnegative matrix factorization (NMF) and more generally nonnnegative tensor factorization (NTF) approximations. We consider multiplicative update (MU) rules to optimize these approximations, and we prove that under certain conditions the results on monotonicity of MU rules for NMF generalize to both the NTF and the weighted NTF (WNTF) cases.

1. Introduction

Nonnegative matrix factorization (NMF) [1] or more generally nonnegative tensor factorization (NTF) [2] approaches consist in approximating nonnegative matrices (or tensors) by lower rank structured matrices (or tensors) composed by nonnegative latent factors. These approximations can be useful for revealing some latent data structure [3] or for compressing the data [4]. Thus they have recently gained a great popularity in both machine learning [5, 6] and signal processing [7, 8] communities. As such, they were applied for non-supervised image classification [5], image inpainting [6], polyphonic music transcription [7], audio source separation [9–11], where such approaches have recently became the de facto state of the art, audio coding [4], etc.

In this paper we focus on a particular kind of NMF (NTF) methods called *weighted NMF (WNMF)* (*weighted NTF (WNTF)*), where the contribution of each data point to the approximation is weighted by a nonnegative weight. Weighted NMF was already used for face feature extraction [12], ratings prediction [13, 14], mass spectrometry analysis [15], as well as for audio source separation with perceptual modeling [9, 10]. A particular WNTF modeling, namely a weighted three-way "PARAFAC" factor analysis, was considered in [16]. The NTF modeling for tensors with possibly missing entries [17] could be considered as a partial case of WNTF, where the weights can be either ones (observed) or zeros (missing).

Concerning the algorithms to compute NMF (NTF) decompositions, one of the most popular choice among the others [2] are the multiplicative update (MU) rules [1, 3]. While in terms of convergence speed MU is not the fastest approach [2], its popularity can be explained by the simplicity of the derivation and implementation, as well as by the fact that the nonnegativity constraints are inherently taken into account. Deriving MU rules for WNMF (WNTF) is quite straightforward, and it was already done for WNMF, e.g., in [13, 15]. However, few works analyse their convergence properties in terms of *monotonicity* of the optimized criterion, i.e., by theoretically studying whether the criterion to be minimized remains non-increasing at each update. While several results on monotonicity of MU rules for NMF exist [1, 18, 19], less work (e.g., [12, 14]) just reports the results on monotonicity of MU rules for WNMF in particular cases of the Euclidean (EU) distance and the Kullback-Leibler (KL) divergence. To our best knowledge there are no such results either for WNTF, or for other divergences, such as α or β-divergences [3, 20].

This is a purely theoretical work and its contribution is two-fold. We consider a quite general NTF formulation inspired by the probabilistic latent tensor factorization (PLTF) [17] (we however do not push forward the probabilistic aspect) and covering many existing divergences and MU rules (e.g., those considered in [1, 18, 19]). First, within this formulation we show that the results on the MU rules monotonicity for the NMF generalize to the general NTF case. Second, we show that within the same formulation all the results on the MU monotonicity for the NMF and the NTF generalize to the corresponding weighted cases. While such results are quite natural and expected, no formal proves were provided so far. A more detailed presentation of these results together with a practical application of WNTF to audio source separation can be found in our longer report [21].

The remaining of this paper is organized as follows. WNMF and WNTF together with a general formulation of the MU rules are presented in section 2. New results on monotonicity of these rules are given in section 3. Some conclusions are drawn in section 4.

2. Weighted NMF and weighted NTF

2.1. Weighted NMF

Let $\mathbf{V} \in \mathbb{R}_+^{F \times N}$ a nonnegative matrix of data that is approximated by a nonnegative matrix $\hat{\mathbf{V}} \in \mathbb{R}_+^{F \times N}$ being a product of two nonnegative latent matrices $\mathbf{W} \in \mathbb{R}_+^{F \times K}$ and $\mathbf{H} \in \mathbb{R}_+^{K \times N}$ as

$$\mathbf{V} \approx \hat{\mathbf{V}} = \mathbf{W}\mathbf{H}. \quad (1)$$

This approximation can be rewritten in a scalar form as

$$v_{fn} \approx \hat{v}_{fn} = \sum\nolimits_k w_{fk} h_{kn}, \quad (2)$$

where v_{fn}, $\hat{v}_{fn}$, w_{fk} and h_{kn} denote, respectively, the entries of $\mathbf{V}$, $\hat{\mathbf{V}}$, $\mathbf{W}$ and $\mathbf{H}$. The goal of NMF consists in finding the latent parameters $\mathbf{Z} \triangleq \{\mathbf{W}, \mathbf{H}\}$ minimizing the following criterion:

$$C_{\text{NMF}}(\mathbf{Z}) = \mathbf{D}(\mathbf{V}|\hat{\mathbf{V}}) = \sum\nolimits_{f,n} d(v_{fn}|\hat{v}_{fn}), \quad (3)$$

where $\hat{v}_{fn}$ is given by (2) and $d(x|y)$ is some divergence (e.g., α-divergence [20] or β-divergence [3]). As such, we here consider only the case of *separable matrix divergences* $\mathbf{D}(\mathbf{V}|\hat{\mathbf{V}})$, i.e., those computed by element-wise summing of scalar divergences [19, 22]. However, we believe that our results below can be easily generalized to some nonseparable divergences.

Let $\mathbf{B} = [b_{fn}]_{f,n} \in \mathbb{R}_+^{F\times N}$ a matrix of nonnegative weights, the goal of WNMF is to optimize the same criterion as (3), except that all the entries in summation are weighted by b_{fn}:

$$C_{\text{WNMF}}(\mathbf{Z}) = \sum\nolimits_{f,n} b_{fn}\, d(v_{fn}|\hat{v}_{fn}). \quad (4)$$

The MU rules [1–3, 19] consist in updating in turn each scalar parameter z as follows:

$$z \leftarrow z \left([\nabla_z C(\mathbf{Z})]_- / [\nabla_z C(\mathbf{Z})]_+\right)^\eta, \quad (5)$$

where $\eta > 0$, $C(\mathbf{Z})$ is the cost function to be minimized, its derivative with respect to (w.r.t.) the parameter writes

$$\nabla_z C(\mathbf{Z}) = [\nabla_z C(\mathbf{Z})]_+ - [\nabla_z C(\mathbf{Z})]_-, \quad (6)$$

and the summands are both nonnegative. Note that the decomposition (6) is not unique and this algorithm is rather a heuristic one. Thus, neither its convergence, nor its monotonicity is guaranteed and should be studied case by case [1, 18, 19].

Assuming the derivative over the second argument of our divergence $d(x|y)$ can be written

$$d'(x|y) = d'_+(x|y) - d'_-(x|y), \quad (7)$$

where $d'_+(x|y)$ and $d'_-(x|y)$ are both nonnegative, one can write the following MU rules for WNMF:

$$w_{fk} \leftarrow w_{fk} \left(\frac{\sum_n b_{fn}\, d'_-(v_{fn}|\hat{v}_{fn}) h_{kn}}{\sum_n b_{fn}\, d'_+(v_{fn}|\hat{v}_{fn}) h_{kn}} \right)^\eta, \quad (8)$$

$$h_{kn} \leftarrow h_{kn} \left(\frac{\sum_f b_{fn}\, d'_-(v_{fn}|\hat{v}_{fn}) w_{fk}}{\sum_f b_{fn}\, d'_+(v_{fn}|\hat{v}_{fn}) w_{fk}} \right)^\eta. \quad (9)$$

Note that this is not the only way to write the MU rules, since the decomposition (6) could be obtained differently [1].

[1] In other words, $[\nabla_z C(\mathbf{Z})]_+$ and $[\nabla_z C(\mathbf{Z})]_-$ from (5) are not obliged to be representable in a from as in (8). Indeed, for example it can be noted that $[\nabla_z C(\mathbf{Z})]_+$ and $[\nabla_z C(\mathbf{Z})]_-$ from (8) are sums over n of some terms with each term depending on its own n. It is obvious that in general not any decomposition as in (6) can be represented as such a sum.

However, this is the way the MU rules are derived in many cases, e.g., for the β-divergence as in [3, 18] and for all separable divergences (including α-divergence and $\alpha\beta$-Bregman divergence) considered in [19].

2.2. Weighted NTF

We build our presentation following a general formulation of tensor decompositions originally called probabilistic latent tensor factorization (PLTF) [8, 17]. However, we rather call it here NTF, since we do not push its probabilistic aspect. Our presentation follows very closely the one from [8, 17], except that we are using slightly different notations and we consider the weighted case.

Instead of matrices (so-called 2-way arrays) we now consider tensors (so-called multi-way arrays) that are all assumed nonnegative. For example $\mathbf{E} = [e_{fnk}]_{f,n,k} \in \mathbb{R}_+^{F,N,K}$ is a 3-way array. However, for the sake of conciseness and following [8, 17] we use single-letter notations for both tensor indices and their domains of definition, e.g., $j = fnk$ and $J = \{f, k, n\}_{f,n,k}$ in the example above.

Let us introduce the following notations:

- I is the set of all indices,
- $\mathbf{V} = [v_{i_0}]_{i_0 \in I_0}$ is the data tensor and $I_0 \subset I$ is the set of visible indices,
- $\mathbf{Z}^\alpha = [z^\alpha_{i_\alpha}]_{i_\alpha \in I_\alpha}$ ($\alpha = 1, \ldots, T$) are T latent factors (tensors), $I_\alpha \subset I$, and we also require $I = I_0 \cup I_1 \cup \ldots \cup I_T$.
- $\bar{I}_\alpha = I \backslash I_\alpha$ denotes the set of indices that are not in I_α.

With these conventions the matrix approximation (2) can be extended to

$$v_{i_0} \approx \hat{v}_{i_0} = \sum\nolimits_{\bar{i}_0 \in \bar{I}_0} \prod\nolimits_{\alpha=1}^{T} z^\alpha_{i_\alpha}. \quad (10)$$

This formulation generalizes in fact many existing models. Let us give some examples for a better understanding. Assuming $\mathbf{Z}^1 = \mathbf{W}$, $\mathbf{Z}^2 = \mathbf{H}$, $I = \{f, n, k\}$, $I_0 = \{f, n\}$, $I_1 = \{f, k\}$ and $I_2 = \{n, k\}$, we get back to the NMF decomposition (2). The TUCKER3 decomposition [23] (this example is from [17])

$$v_{jkl} \approx \hat{v}_{jkl} = \sum\nolimits_{p,q,r} z^1_{jp} z^2_{kq} z^3_{lr} z^4_{pqr} \quad (11)$$

can be represented as (10) by defining $I = \{j, k, l, p, q, r\}$, $I_0 = \{j, k, l\}$, $I_1 = \{j, p\}$, $I_2 = \{k, q\}$, $I_3 = \{l, r\}$ and $I_4 = \{p, q, r\}$.

Let $\mathbf{Z} = \{\mathbf{Z}^\alpha\}_{\alpha=1,\ldots,T}$ set of all latent factors and $\mathbf{B} = [b_{i_0}]_{i_0 \in I_0}$ a tensor of nonnegative weights. WNTF criterion to be minimized writes:

$$C_{\text{WNTF}}(\mathbf{Z}) = \sum\nolimits_{i_0} b_{i_0}\, d(v_{i_0}|\hat{v}_{i_0}). \quad (12)$$

Finally, relying on the decomposition (7), as in the WNMF case, one can derive the following MU rules for WNTF:

$$z^\alpha_{i_\alpha} \leftarrow z^\alpha_{i_\alpha} \left(\frac{\sum_{\bar{i}_\alpha} b_{i_0}\, d'_-(v_{i_0}|\hat{v}_{i_0}) \prod_{\alpha' \neq \alpha} z^{\alpha'}_{i'_\alpha}}{\sum_{\bar{i}_\alpha} b_{i_0}\, d'_+(v_{i_0}|\hat{v}_{i_0}) \prod_{\alpha' \neq \alpha} z^{\alpha'}_{i'_\alpha}} \right)^\eta. \quad (13)$$

3. New results on monotonicity of MU rules for WNMF and WNTF

There exist several results on monotonicity of MU rules for NMF with β-divergence [18] and NMF with other divergences (e.g., α-divergence or $\alpha\beta$-Bregman divergence) [19]. However, these results are not really extended neither to a general NTF case, nor to the WNMF or WNTF cases. Some results for WNMF exist [12, 14], but only in particular cases of the EU distance and the KL divergence. In order to fill in these gaps in the state of the art we here provide the insite on the monotonicity of the NMF MU rules for NTF, WNMF and WNTF cases. We start with the following lemma.

Lemma 1. *When updating one latent factor $\mathbf{Z}^\alpha$, given all other factors $\{\mathbf{Z}^{\alpha'}\}_{\alpha'\neq\alpha}$ fixed, criterion* (4) *is non-increasing under WNMF MU rules* (13) *if and only if for each $i_0 \cap i_\alpha$ criterion*

$$C_{\mathrm{WNTF}}(\mathbf{Z}^\alpha_{i_0\cap i_\alpha}) = \sum\nolimits_{i_0\cap\bar{i}_\alpha} b_{i_0\cap\bar{i}_\alpha}\, d(v_{i_0\cap\bar{i}_\alpha}|\hat{v}_{i_0\cap\bar{i}_\alpha}), \tag{14}$$

(where $\mathbf{Z}^\alpha_{i_0\cap i_\alpha} = [z^\alpha_{i_0\cap i_\alpha, i_0\cap\bar{i}_\alpha}]_{i_0\cap\bar{i}_\alpha}$) is non-increasing under these rules.

Proof: The sufficiency is evident, the necessity follows from the fact that two sets of entries of $\mathbf{Z}^\alpha$ involved in two different criteria (14) (corresponding to two different indices $i_0 \cap i_\alpha \neq i'_0 \cap i'_\alpha$) do not overlap. □

Proposition 1 (WNMF monotonicity ⇔ WNTF monotonicity)**.** *Assume WNMF MU rules* (8), (9) *and WNTF MU rules* (13) *are derived for the same η, for the same divergence $d(x|y)$ and under the same decomposition* (7). *WNMF criterion* (4) *is non-increasing under the WNMF MU rules if and only if WNTF criterion* (12) *is non-increasing under the WNTF MU rules.*

Proof: The sufficiency follows from the fact that WNTF generalizes WNMF. To prove the necessity, it is enough to show, thanks to lemma 1, that for updating one latent subfactor $\mathbf{Z}^\alpha_{i_0\cap i_\alpha}$, given $i_0 \cap i_\alpha \in I_0 \cap I_\alpha$ and given all other factors $\{\mathbf{Z}^{\alpha'}\}_{\alpha'\neq\alpha}$ fixed, expressions (10), (14) and (13) can be recast into the form of expressions (2), (4) and (9) for WNMF. We rely on an NTF to NMF reduction trick that is somehow similar to the one used in [24].

Let us define $F = |I_0 \cap \bar{I}_\alpha|$, $N = 1$ and $K = |\bar{I}_0 \cap I_\alpha|$, where $|A|$ denotes cardinality of a set A. We can now unfold multiway index sets $I_0 \cap \bar{I}_\alpha$ and $\bar{I}_0 \cap I_\alpha$ onto 1-way index sets $\{f\} = \{1, \dots, F\}$ and $\{k\} = \{1, \dots, K\}$ using some bijections $f \to [i_0 \cap \bar{i}_\alpha](f)$ and $k \to [\bar{i}_0 \cap i_\alpha](k)$. We then define elements of matrices/vectors $\mathbf{V} \in \mathbb{R}_+^{F\times 1}$, $\mathbf{B} \in \mathbb{R}_+^{F\times 1}$, $\mathbf{W} \in \mathbb{R}_+^{F\times K}$ and $\mathbf{H} \in \mathbb{R}_+^{K\times 1}$ as $v_{f1} = v_{i_0\cap i_\alpha,[i_0\cap\bar{i}_\alpha](f)}$, $b_{f1} = b_{i_0\cap i_\alpha,[i_0\cap\bar{i}_\alpha](f)}$, $h_{k1} = z^\alpha_{i_0\cap i_\alpha,[\bar{i}_0\cap i_\alpha](k)}$ and $w_{fk} = \tilde{w}_{[i_0\cap\bar{i}_\alpha](f)[\bar{i}_0\cap i_\alpha](k)}$, where $\tilde{w}_{i_0\cap\bar{i}_\alpha,\bar{i}_0\cap i_\alpha} = \sum_{\bar{i}_0\cap\bar{i}_\alpha}\prod_{\alpha'\neq\alpha} z^{\alpha'}_{i'_\alpha}$. It can be easily checked that with these notations expressions (10), (14) and (13) rewrite as expressions (2), (4) and (9) for WNMF. □

Even if according to proposition 1 the monotonicity of the WNMF MU rules implies that of the WNTF MU rules, the monotonicity of the WNMF MU rules has not been guaranteed yet. To obtain such results, let us first formulate the following lemma, which strictly speaking is not a direct consequence of proposition 1, but it is very similar to it.

Lemma 2 (NMF monotonicity ⇔ NTF monotonicity)**.** *This lemma formulates exactly as proposition 1, but without weighting, i.e., with trivial weighting: $b_{i_0} = 1$ ($i_0 \in I_0$).*

Proof: The proof is exactly as that of proposition 1, except with trivial weighting. □

Proposition 2 (NMF monotonicity ⇔ WNMF monotonicity)**.** *Assume that WNMF MU rules* (8), (9) *are derived for some η, some divergence $d(x|y)$ and under some decomposition* (7). *WNMF criterion* (4) *is non-increasing under the WNMF MU rules for a trivial weighting $\mathbf{B}_0 = [1]_{f,n}$ (making WNMF* (4) *equivalent standard NMF* (3)*) if and only if WNMF criterion* (4) *is non-increasing under the WNMF MU rules for any weighting* $\mathbf{B}$.

Proof: The sufficiency being evident, let us prove the necessity. We carry the proof for $\mathbf{H}$ update (9), given $\mathbf{W}$ fixed. According to lemma 1, it is enough to show that for each $n = \tilde{n}$ the following criterion is non-increasing under the WNMF MU updates of $\mathbf{h}_{\tilde{n}} = [h_{\tilde{n}k}]_k$ (the same trick is used in proofs in [18]):

$$C_{\mathrm{WNMF}}(\mathbf{h}_{\tilde{n}}) = \sum\nolimits_f b_{f\tilde{n}}\, d(v_{f\tilde{n}}|\hat{v}_{f\tilde{n}}). \tag{15}$$

Thus, in the following we fix $n = \tilde{n}$ and consider (15) as target criterion.

We first assume $b_{f\tilde{n}} \in \mathbb{N}$ ($f = 1, \dots, F$). Let us introduce binary matrices $\mathbf{A}_f(\tilde{n}) \in \{0,1\}^{b_{f\tilde{n}}\times F}$. Each matrix $\mathbf{A}_f(\tilde{n})$ is zero everywhere except the f-th column that contains ones. We then define a binary matrix $\mathbf{A}(\tilde{n}) = [a_{lf}(\tilde{n})]_{l,f} \in \{0,1\}^{L\times F}$, where $L = \sum_f b_{f\tilde{n}}$, that stacks vertically matrices $\mathbf{A}_f(\tilde{n})$ as follows:

$$\mathbf{A}(\tilde{n}) = [\mathbf{A}_1(\tilde{n})^T, \dots, \mathbf{A}_F(\tilde{n})^T]^T. \tag{16}$$

Using $\mathbf{A}(\tilde{n})$ we rewrite approximation (2) as follows:

$$v'_{l\tilde{n}} \approx \hat{v}'_{l\tilde{n}} = \sum\nolimits_{f,k} a_{lf}(\tilde{n}) w_{fk} h_{k\tilde{n}}, \tag{17}$$

where $v'_{l\tilde{n}} = \sum_f a_{lf}(\tilde{n}) v_{f\tilde{n}}$. Let us first remark that (17) is an NTF approximation (w.r.t. $v'_{l\tilde{n}}$ and $\hat{v}'_{l\tilde{n}}$) according to our general formulation (10). Non-weighted NTF criterion (12) (i.e., (12) with trivial weighting $b'_{l\tilde{n}} = 1$) for approximation (17) writes

$$C_{\mathrm{NTF}}(\mathbf{h}_{\tilde{n}}) = \sum\nolimits_l d(v'_{l\tilde{n}}|\hat{v}'_{l\tilde{n}}). \tag{18}$$

It can be easily shown that criterion (18) is strictly equivalent to criterion (15) and that the corresponding MU updates w.r.t. $\mathbf{h}_{\tilde{n}}$ are the same. Moreover, since the monotonicity of non-weighted NMF is assumed, it implies, according to lemma 2, the monotonicity of non-weighted NTF criterion (18), and thus that of criterion (15).

We have proven the result for $b_{f\tilde{n}} \in \mathbb{N}$. Since multiplying all the weights by a positive constant factor does not affect monotonicity of MU updates, the result is proven for $b_{f\tilde{n}} \in \mathbb{Q}_+$. Finally, since $\mathbb{Q}_+$ is dense in $\mathbb{R}_+$ and since both MU updates and the corresponding criteria are all continuous w.r.t. weights and parameters, the result is proven for $b_{f\tilde{n}} \in \mathbb{R}_+$. □

Propositions 2 and 1 can be summarized by the following theorem.

Theorem 1 (NMF monotonicity ⇔ WNTF monotonicity)**.** *Assume WNMF MU rules* (8)*,* (9) *and WNTF MU rules* (13) *are derived for the same* η*, for the same divergence* $d(x|y)$ *and under the same decomposition* (7)*. WNMF criterion* (4) *is non-increasing under the WNMF MU rules for a trivial weighting* $\mathbf{B}_0 = [1]_{f,n}$ *(making WNMF* (4) *equivalent standard NMF* (3)*) if and only if WNTF criterion* (12) *is non-increasing under the WNTF MU rules.*

We have shown that the results on the monotonicity of NMF MU rules derived as in section 2.1 (in particular, those from [18] for β-divergence and those from [19] for separable divergences, e.g., for α and $\alpha\beta$-Bregman divergences) generalize to WNMF, NTF and WNTF cases.

4. Conclusion

We have proven that certain results on monotonicity of MU rules for NMF generalize to WNMF, NTF and WNTF cases. The underlined meaning is that that in most cases considered in the literature the conditions on MU rules monotonicity depend mostly on the divergence and not on the NTF structure or weighing. Such results are quite natural and were expected. However, to the best of our knowledge, no formal proves were provided so far. Future work will consist in trying to relax the current result conditions including a restriction to the case separable divergences and a specific decomposition (7), which does not lead to the most general form of MU rules.

References

[1] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural and Information Processing Systems 13*, 2001, pp. 556–562.

[2] A. Cichocki, R. Zdunek, A.H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations*, John Wiley & Sons Ltd, Chichester, UK, 2009.

[3] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, Mar. 2009.

[4] A. Ozerov, A. Liutkus, R. Badeau, and G. Richard, "Coding-based informed source separation: Nonnegative tensor factorization approach," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 8, pp. 1699 – 1712, Aug. 2013.

[5] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman, "Discovering objects and their location in images," in *Proc. International Conference on Computer Vision (ICCV)*, 2005.

[6] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, no. 1, pp. 19–60, 2010.

[7] 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 (WASPAA'03)*, 19-22 Oct. 2003, pp. 177–180.

[8] A. T. Cemgil, U. Simsekli, and Y. C. Subakan, "Probabilistic latent tensor factorization framework for audio modeling," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA Š11)*, 2011.

[9] T. O. Virtanen, "Monaural sound source separation by perceptually weighted non-negative matrix factorization," Tech. Rep., Tampere University of Technology, 2007.

[10] S. Kirbiz and B. Gunsel, "Perceptually weighted non-negative matrix factorization for blind single-channel music source separation," in *21st International Conference on Pattern Recognition (ICPR)*, 2012, pp. 226–229.

[11] A. Ozerov, C. Févotte, R. Blouet, and J.-L. Durrieu, "Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation," in *Proc. IEEE Int. Conf. on Acoustics, speech, and signal processing*, Prague, Czech Republic, May 2011, pp. 257 – 260.

[12] V. D. Blondel, N.-D. Ho, and P. V. Dooren, "Weighted non-negative matrix factorization and face feature extraction," in *Image and Vision Computing*, 2008.

[13] Y.-D. Kim and S. Choi, "Weighted nonnegative matrix factorization," in *Proc. of International Conference on Acoustics, Speech, and Signal Processing*, 2009, pp. 1541–1544.

[14] Q. Gu, J. Zhou, and C.H.Q. Ding, "Collaborative filtering weighted nonnegative matrix factorization incorporating user and item graphs," in *SIAM Conference on Data Mining (SDM)*, 2010, pp. 199–210.

[15] R. Dubroca, C. Junor, and A. Souloumiac, "Weighted NMF for high-resolution mass spectrometry analysis," in *Proc. of the 20th European Signal Processing Conference (EUSIPCO'12)*, 2012.

[16] P. Paatero, "A weighted non-negative least squares algorithm for three-way "PARAFAC" factor analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 38, pp. 223–242, 1997.

[17] K. Yilmaz and A. T. Cemgil, "Probabilistic latent tensor factorisation," in *Proc. of International Conference on Latent Variable analysis and Signal Separation*, 2010, pp. 346–353.

[18] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the beta-divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, Sep. 2011.

[19] Z. Yang and E. Oja, "Unified development of multiplicative algorithms for linear and quadratic nonnegative matrix factorization," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 1878–1891, 2011.

[20] A. Cichocki, H. Lee, Y.D. Kim, and S. Choi, "Nonnegative matrix factorization with alpha-divergence," *Pattern Recogn. Lett.*, vol. 29, pp. 1433–1440, 2008.

[21] A. Ozerov, N. Q. K. Duong, and L. Chevallier, "Weighted nonnegative tensor factorization: on monotonicity of multiplicative update rules and application to user-guided audio source separation," Tech. Rep., Technicolor, Oct. 2013, Available online: `http://hal.inria.fr/hal-00878685`.

[22] I. Dhillon and S. Sra, "Generalized nonnegative matrix approximations with Bregman divergences," in *Proc. of the Neural Information Processing Systems (NIPS) Conference*, Vancouver, BC, 2005.

[23] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *Journal of Chemometrics*, vol. 14, pp. 105–122, 2000.

[24] Y. K. Yilmaz, A. T. Cemgil, and U. Simsekli, "Generalised coupled tensor factorisation," in *25th Annual Conference on Neural Information Processing Systems (NIPS)*, 2011.