Weighted nonnegative tensor factorization: on monotonicity of multiplicative update rules and application to user-guided audio source separation

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

Weighted nonnegative tensor factorization: on monotonicity of multiplicative update rules and application to user-guided audio source separation

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

Abstract

This report focuses on so-called *weighted* variants of nonnnegative matrix factorization (NMF) and more generally nonnnegative tensor factorization (NTF) approximations. First, we consider multiplicative update (MU) rules to optimize these approximations, and we prove that some results on monotonicity of MU rules for NMF generalize without restrictions to both the NTF and the WNTF cases. Second, we propose new weighting strategies for an existing NTF-based user-guided audio source separation method. Experimental evaluation shows that these weightings lead to better source separation than the same model without using the weights. The best configuration of the proposed method was entered into the fourth community-based Signal Separation Evaluation Campaign (SiSEC 2013).

Index Terms

Weighted nonnegative matrix factorization, weighted nonnegative tensor factorization, multiplicative updates, monotonicity, user-guided audio source separation

I. Introduction

Nonnegative matrix factorization (NMF) [1] or more generally nonnegative tensor factorization (NTF) [2] approaches consist in approximating nonnegative matrices (tensors) by lower rank structured matrices (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]–[9] communities. As such, they were applied for non-supervised image classification [5], image inpainting [6], polyphonic music transcription [7], audio source separation [10]–[15], audio coding [4], etc.

In this report 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 [16], ratings prediction [17], [18], mass spectrometry analysis [19], as well as for audio source separation with perceptual modeling [10], [11]. A particular WNTF modeling, namely a weighted three-way "PARAFAC" factor analysis, was considered in [20]. The NTF modeling for tensors with possibly missing entries [21] 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 [17], [19]. 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], [22], [23], less work (e.g., [16], [18]) just reports the results on monotonicity of MU rules for WNMF in particular cases of the Euclidean (EU) distance and the Kullback-Leibler

The authors are with Technicolor, 1, avenue de la Belle Fontaine F-35576 Cesson Sévigné, France, e-mail: first-name.lastname@technicolor.com.

(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], [24].

The contribution of this work is two-fold. First, we provide some theoretical results on monotonicity of MU rules for WNMF and WNTF. Second, we propose new WNTF strategies applied to a user-guided audio source separation framework inspired by [13] and using a particular NTF model (multichannel NMF model from [12]). The user-guided source separation approach from [13] is based on a temporal user-specified annotation of source activities (or *time codes*) that are further used to constrain NTF model learning. Our original WNTF framework targets to (i) re-weight the data so as to enhance the importance of segments with less sources (or less NTF components), and to (ii) re-weight the data so as to re-equilibrate the impact of different data type segments that may have different lengths.

The remaining of this report is organized as follows. WNMF and WNTF together with a general formulation of the MU rules are presented in section II. New results on monotonicity of these rules for both WNMF and WNTF are given in section III. Section IV is devoted to the description of the proposed WNTF-based user-guided audio source separation framework, which is then evaluated in section V on the development data of the fourth community-based Signal Separation Evaluation Campaign (SiSEC 2013) "professionally produced music recordings" (PPMRs) task. The best configuration according to this evaluation was then entered into the evaluation campaign. Some conclusions are drawn in section VI.

II. Weighted NMF and weighted NTF

A. 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 $\mathbf{V} \approx \hat{\mathbf{V}} = \mathbf{WH}$, which can be rewritten in a scalar form as

$$v_{fn} \approx \hat{v}_{fn} = \sum\nolimits_k w_{fk} h_{kn}, \tag{1}$$

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:

$$C_{\mathrm{NMF}}(\mathbf{Z}) = \mathbf{D}(\mathbf{V}|\hat{\mathbf{V}}) = \sum\nolimits_{f,n} d(v_{fn}|\hat{v}_{fn}), \tag{2}$$

where $\hat{v}_{fn}$ is given by (1) and $d(x|y)$ is some divergence (e.g., α-divergence [24] 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 [23], [25]. 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 (2), except that all the entries in summation are weighted by b_{fn}:

$$C_{\mathrm{WNMF}}(\mathbf{Z}) = \sum\nolimits_{f,n} b_{fn}\, d(v_{fn}|\hat{v}_{fn}). \tag{3}$$

The MU rules [1]–[3], [23] consist in updating in turn each scalar parameter z as follows:

$$z \leftarrow z \left([\nabla_z C(\mathbf{Z})]_- / [\nabla_z C(\mathbf{Z})]_+ \right)^\eta, \tag{4}$$

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

$$\nabla_z C(\mathbf{Z}) = [\nabla_z C(\mathbf{Z})]_+ - [\nabla_z C(\mathbf{Z})]_-, \tag{5}$$

and the summands are both nonnegative. Note that the decomposition (5) is not unique and this algorithm is rather a heuristic one. Thus, neither its convergence, no its monotonicity is guaranteed and should be studied case by case [1], [22], [23].

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

$$d'(x|y) = d'_+(x|y) - d'_-(x|y), \tag{6}$$

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

$$w_{fk} \leftarrow w_{fk} \left(\frac{\sum_n b_{fn}\, d'_-(v_{fn}|\hat{v}_{fn}) h_{kn}}{\sum_n b_{fn}\, d'_+(v_{fn}|\hat{v}_{fn}) h_{kn}} \right)^{\eta}, \tag{7}$$

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

Note that this is not the only way to write the MU rules, since the decomposition (5) could be probably obtained differently. However, this is the way the MU rules are derived in many cases, e.g., for the β-divergence as in [3], [22] and for all separable divergences (including α-divergence and $\alpha\beta$-Bregman divergence) considered in [23].

B. Weighted NTF

We build our presentation following a general formulation of tensor decompositions originally called probabilistic latent tensor factorization (PLTF) [9], [21]. However, we rather call it here NTF, since we do not push its probabilistic aspect. Our presentation follows very closely [9], [21], 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 [9], [21] 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 (1) can be extended to

$$v_{i_0} \approx \hat{v}_{i_0} = \sum\nolimits_{\bar{i}_0 \in \bar{I}_0} \prod\nolimits_{\alpha=1}^{T} z^\alpha_{i_\alpha}. \tag{9}$$

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 (1). The TUCKER3 decomposition [26] (this example is from [21])

$$v_{jkl} \approx \hat{v}_{jkl} = \sum\nolimits_{p,q,r} z^1_{jp} z^2_{kq} z^3_{lr} z^4_{pqr} \tag{10}$$

can be represented as (9) 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 $I_4 = \{p, q, r\}$. These models can be also visualized by graphical representations from [9], [21], where all indices are represented by graph nodes and the latent factors by corresponding non-oriented graph vertices. NMF (1) and TUCKER3 (10) models are represented by such graphs on Fig. 1 (A) and (B), where the visible indices are shaded and the non-visible ones are white.

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:

$$C_{\mathrm{WNTF}}(\mathbf{Z}) = \sum\nolimits_{i_0} b_{i_0}\, d(v_{i_0}|\hat{v}_{i_0}). \tag{11}$$

Finally, relying on the decomposition (6), as in the WNMF case, one can derive the following MU rules for WNTF:

$$z^\alpha_{i_\alpha} \leftarrow z^\alpha_{i_\alpha} \left(\frac{\sum_{\bar{i}_\alpha} b_{i_0}\, d'_-(v_{i_0}|\hat{v}_{i_0}) \prod_{\alpha' \neq \alpha} z^{\alpha'}_{i'_\alpha}}{\sum_{\bar{i}_\alpha} b_{i_0}\, d'_+(v_{i_0}|\hat{v}_{i_0}) \prod_{\alpha' \neq \alpha} z^{\alpha'}_{i'_\alpha}} \right)^{\eta}. \tag{12}$$

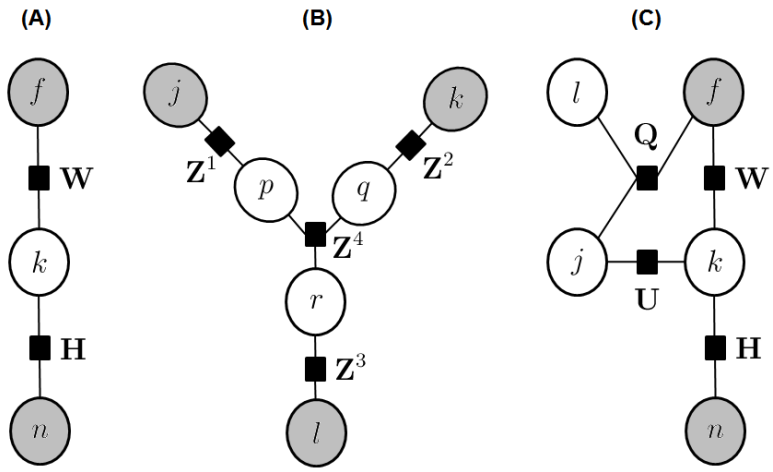

Fig. 1. Graphical representations of different NTF models. **(A)**: NMF model (1); **(B)**: TUCKER3 model (10); **(C)**: Multichannel NMF model (18).

III. 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 [22] and NMF with other divergences (e.g., α-divergence or $\alpha\beta$-Bregman divergence) [23]. 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 [16], [18], 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* (3) *is non-increasing under WNMF MU rules* (12) *if and only if for each $i_0 \cap i_\alpha$ criterion*

$$C_{\mathrm{WNTF}}(\mathbf{Z}^{\alpha}_{i_0\cap i_\alpha}) = \sum\nolimits_{i_0\cap\bar{i}_\alpha} b_{i_0\cap\bar{i}_\alpha}\, d(v_{i_0\cap\bar{i}_\alpha}|\hat{v}_{i_0\cap\bar{i}_\alpha}), \tag{13}$$

(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 (13) (corresponding to two different indices $i_0 \cap i_\alpha \neq i'_0 \cap i'_\alpha$) do not overlap. ■

Proposition 1 (WNMF monotonicity $\Leftrightarrow$ WNTF monotonicity)**.** *Assume WNMF MU rules* (7), (8) *and WNTF MU rules* (12) *are derived for the same η, for the same divergence $d(x|y)$ and under the same decomposition* (6)*. WNMF criterion* (3) *is non-increasing under the WNMF MU rules if and only if WNTF criterion* (11) *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 sub-factor $\mathbf{Z}^{\alpha}_{i_0\cap i_\alpha}$, 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 (9), (13) and (12) can be recast into the form of expressions (1), (3) and (8) for WNMF.

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 multi-way index sets $I_0 \cap \bar{I}_\alpha$ and $\bar{I}_0 \cap I_\alpha$ onto 1-way index sets $\{f\} = \{1, \ldots, F\}$ and $\{k\} = \{1, \ldots, K\}$ 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 (9), (13) and (12) rewrite as expressions (1), (3) and (8) 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 $\Leftrightarrow$ 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 $\Leftrightarrow$ WNMF monotonicity)**.** *Assume that WNMF MU rules* (7)*,* (8) *are derived for some* η*, some divergence* $d(x|y)$ *and under some decomposition* (6)*. WNMF criterion* (3) *is non-increasing under the WNMF MU rules for a trivial weighting* $\mathbf{B}_0 = [1]_{f,n}$ *(making WNMF* (3) *equivalent standard NMF* (2)*) if and only if WNMF criterion* (3) *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 (8), 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 [22]):

$$C_{\mathrm{WNMF}}(\mathbf{h}_{\tilde{n}}) = \sum\nolimits_f b_{f\tilde{n}}\, d(v_{f\tilde{n}}|\hat{v}_{f\tilde{n}}). \tag{14}$$

Thus, in the following we fix $n = \tilde{n}$ and consider (14) as target criterion.

We first assume $b_{f\tilde{n}} \in \mathbb{N}$ $(f = 1, \ldots, F)$. 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:

$$\mathbf{A}(\tilde{n}) = [\mathbf{A}_1(\tilde{n})^T, \ldots, \mathbf{A}_F(\tilde{n})^T]^T. \tag{15}$$

Using $\mathbf{A}(\tilde{n})$ we rewrite approximation (1) as follows:

$$v'_{l\tilde{n}} \approx \hat{v}'_{l\tilde{n}} = \sum\nolimits_{f,k} a_{lf}(\tilde{n}) w_{fk} h_{k\tilde{n}}, \tag{16}$$

where $v'_{l\tilde{n}} = \sum_f a_{lf}(\tilde{n}) v_{f\tilde{n}}$. Let us first remark that (16) is an NTF approximation (w.r.t. $v'_{l\tilde{n}}$ and $\hat{v}'_{l\tilde{n}}$) according to our general formulation (9). Non-weighted NTF criterion (11) (i.e., (11) with trivial weighting $b'_{l\tilde{n}} = 1$) for approximation (16) writes

$$C_{\mathrm{NTF}}(\mathbf{h}_{\tilde{n}}) = \sum\nolimits_l d(v'_{l\tilde{n}}|\hat{v}'_{l\tilde{n}}). \tag{17}$$

It can be easily shown that criterion (17) is strictly equivalent to criterion (14) 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 (17), and thus that of criterion (14).

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 $\Leftrightarrow$ WNTF monotonicity)**.** *Assume WNMF MU rules* (7)*,* (8) *and WNTF MU rules* (12) *are derived for the same* η*, for the same divergence* $d(x|y)$ *and under the same decomposition* (6)*. WNMF criterion* (3) *is non-increasing under the WNMF MU rules for a trivial weighting* $\mathbf{B}_0 = [1]_{f,n}$ *(making WNMF* (3) *equivalent standard NMF* (2)*) if and only if WNTF criterion* (11) *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 II-A (in particular, those from [22] for β-divergence and those from [23] for separable divergences, e.g., for α and $\alpha\beta$-Bregman divergences) generalize to WNMF, NTF and WNTF cases.

IV. Application to user-guided audio source separation

The considered user-guided audio source separation framework is similar to the one from [13], except that here we assume all model parameters, including mixing coefficients, nonnegative. Thus, we rather adopt the multichannel NMF modeling from [12] (the one referred as "MU rules" in [12]). We then introduce two complementary weighting strategies.

A. WNTF model for source separation

Let $\mathbf{X} = [x_{lfn}]_{l,f,n} \in \mathbb{C}^{L\times F\times N}$ the short-time Fourier transform (STFT) of the multichannel mixture, where l, f and n denote, respectively, the channel, the frequency and the time indices. We consider its power-spectrogram $\mathbf{V} = [v_{lfn}]_{l,f,n} \in \mathbb{R}_+^{L\times F\times N}$ ($[v_{lfn}]_{l,f,n} = |x_{lfn}|^2_{l,f,n}$) as data tensor. Assuming there are J spatial sources altogether represented by K latent additive components, the model approximation writes

$$v_{lfn} \approx \hat{v}_{lfn} = \sum\nolimits_{j,k} q_{ljf} u_{jk} w_{fk} h_{kn}, \tag{18}$$

where tensor $\mathbf{Q} = [q_{ljf}]_{l,j,f} \in \mathbb{R}_+^{L\times J\times F}$ models the time-invariant spatial filters of J sources, matrix $\mathbf{U} = [u_{jk}]_{j,k} \in \mathbb{R}_+^{J\times K}$ represent the correspondence between J sources and K latent additive components, and matrices $\mathbf{W} = [w_{fk}]_{f,k} \in \mathbb{R}_+^{F\times K}$ and $\mathbf{H} = [h_{kn}]_{k,n} \in \mathbb{R}_+^{K\times N}$, model, respectively, the spectral patterns and the time activations of latent additive components. A graphical representation of this model is given on Fig. 1 (C).

Let $\mathbf{B} = [b_{lfn}]_{l,f,n} \in \mathbb{R}_+^{L\times F\times N}$ a tensor of nonnegative weights and $d(x|y)$ β-divergence (see, e.g., [3] for a definition). According to [3] $d'(x|y) = y^{\beta-2}(y-x)$, thus it can be represented as (6) with $d'_+(x|y) = y^{\beta-1}$ and $d'_-(x|y) = y^{\beta-2}x$, and general WNTF MU rules (12) become:

$$q_{ljf} \leftarrow q_{ljf}\left(\frac{\sum_{n,k} b_{lfn} v_{lfn} \hat{v}_{lfn}^{\beta-2} u_{jk} w_{fk} h_{kn}}{\sum_{n,k} b_{lfn} \hat{v}_{lfn}^{\beta-1} u_{jk} w_{fk} h_{kn}}\right)^{\eta}, \tag{19}$$

$$u_{jk} \leftarrow u_{jk}\left(\frac{\sum_{l,f,n} b_{lfn} v_{lfn} \hat{v}_{lfn}^{\beta-2} q_{ljf} w_{fk} h_{kn}}{\sum_{l,f,n} b_{lfn} \hat{v}_{lfn}^{\beta-1} q_{ljf} w_{fk} h_{kn}}\right)^{\eta}, \tag{20}$$

$$w_{fk} \leftarrow w_{fk}\left(\frac{\sum_{l,j,n} b_{lfn} v_{lfn} \hat{v}_{lfn}^{\beta-2} q_{ljf} u_{jk} h_{kn}}{\sum_{l,j,n} b_{lfn} \hat{v}_{lfn}^{\beta-1} q_{ljf} u_{jk} h_{kn}}\right)^{\eta}, \tag{21}$$

$$h_{kn} \leftarrow h_{kn}\left(\frac{\sum_{l,j,f} b_{lfn} v_{lfn} \hat{v}_{lfn}^{\beta-2} q_{ljf} u_{jk} w_{fk}}{\sum_{l,j,f} b_{lfn} \hat{v}_{lfn}^{\beta-1} q_{ljf} u_{jk} w_{fk}}\right)^{\eta}. \tag{22}$$

B. User-guided approach

User-guided source separation approach we consider here is the one from [13] developed for separation of PPMRs. The difference lies in the fact that a weighting is used within this modeling. This approach consists in introducing some structure into approximation (18) by setting to zero some factor entries, which then remain zero under multiplicative updates (19) - (22). The main steps of this approach are summarized below, and more details can be found in [13].

1) The user (manually) performs a temporal segmentation of the sources to separate (see Fig. 2 (A) for an example) and decides on the number of latent components (indexed by k in section IV-A) per source (e.g., 2 for bass, 7 for vocals, etc.).
2) The temporal segmentation and the number of components per source are reflected in matrix $\mathbf{H}$ in the forms of zeros. E.g., if source 1 is assigned two first components and is silent between frames 100 and 200, then we set $h_{1n} = h_{2n} = 0$ for $n = 100, \ldots, 200$. The other coefficients are randomly initialized to positive values.
3) The remaining parameters are initialized using and ad-hoc procedure described in Section IV.H of [12], based on WNMF of the stacked channel spectrograms (initialization of $\mathbf{W}$) and clustering of the spatial cues (initialization of $\mathbf{Q}$ and $\mathbf{U}$). This WNMF is performed with a similar temporal weighting as the one described in section IV-C and used in the next step.
4) Iterate WNMF MU rules (19) - (22) with weights $\mathbf{B}$ specified by one of strategies from section IV-C below.
5) Compute K latent components estimates in the STFT domain by Wiener filtering as follows:

$$\hat{c}_{klfn} = \left(\hat{v}_{lfn}^{-1} \sum\nolimits_j q_{ljf} u_{jk} w_{fk} h_{kn}\right) x_{lfn}, \tag{23}$$

   where $\hat{v}_{lfn}$ is defined in (18).
6) As discussed in [13], for PPMR separation, neither temporal segmentation, nor spatial cues can completely disambiguate component - source assignments. Thus, if needed, the user can listen to the components and manually fine-tune their grouping.

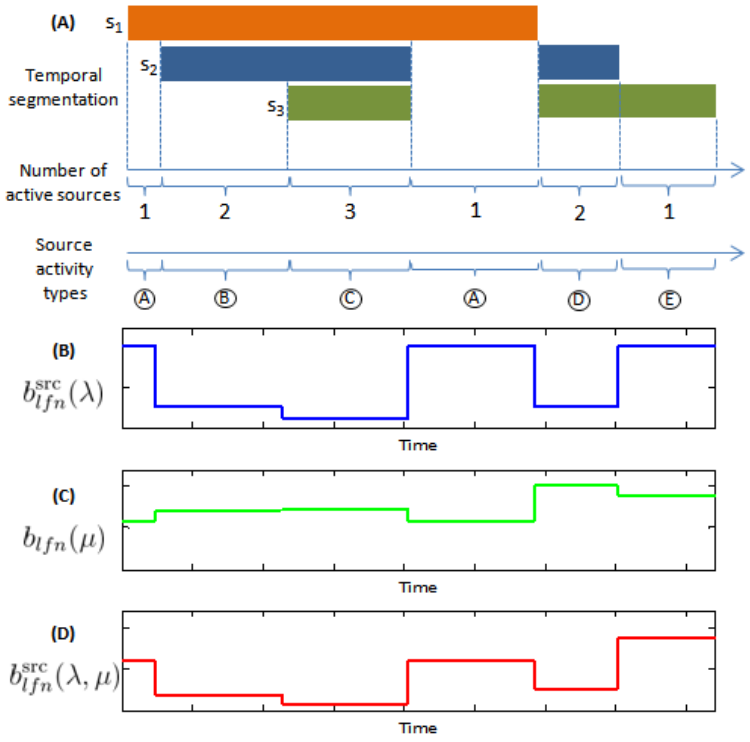

Fig. 2. Examples of proposed weighting strategies. **(A)**: temporal segmentation; **(B)**: weighting enhancing data purity $b^{\mathrm{src}}_{lfn}(\lambda)$, $(\lambda = 2)$; **(C)**: weighting equilibrating segment types $b_{lfn}(\mu)$, $(\mu = 0.8)$; **(D)**: combined weighting $b^{\mathrm{src}}_{lfn}(\lambda, \mu)$ $(\lambda = 2, \mu = 0.8)$.

C. Proposed temporal weighting strategies

We introduce two new temporal weighting strategies for our model, where the weights b_{lfn} vary only over the temporal dimension indexed by n and are constant over l and f. We also explain how these strategies, being complementary, can be combined.

1) Weighting to enhance data purity: Some source separation approaches, e.g., [27], first learn source spectral patterns $\mathbf{W}$ from some training data consisting of clean source examples, and then perform NMF decomposition on a test mixture, while keeping pre-learned spectral patterns $\mathbf{W}$ fixed. Other approaches, e.g., [13] summarized in section IV-B, introduce information about source activities via zeros in $\mathbf{H}$, and, assuming trivial weights ($b_{lfn} = 1$) as in [13], spectral patterns $\mathbf{W}$ are learned from all data without making any distinction between segments with clean sources and those with mixed sources.

We propose a weighting providing a continuum of strategies between these two, thus possibly allowing to choose a better intermediate strategy. Given a temporal segmentation, as explained above, we define the weights for estimating parameters of (18) as one of the following two options:

$$b^{\mathrm{src}}_{lfn}(\lambda) = [1/\#(\mathrm{act\_src}_n)]^{\lambda}, \tag{24}$$

$$b^{\mathrm{cmp}}_{lfn}(\lambda) = [1/\#(\mathrm{act\_cmp}_n)]^{\lambda}, \tag{25}$$

where $\lambda \in [0, +\infty)$ is a fixed parameter, and $\#(\mathrm{act\_src}_n)$ and $\#(\mathrm{act\_cmp}_n)$ are, respectively, the number of active sources and the number of latent NTF components in n-th frame. An example of such weighting is represented on Fig. 2 (B). One can see that, in case of two sources in the mixture and for $\lambda \to +\infty$, this approach is equivalent to the first state-of-the-art strategy [27] (for more than two sources, it generalizes [27] to a slightly more tricky learning). For $\lambda = 0$ this approach becomes equivalent to the second strategy [13]. The proposed weighting is also related to [14], [15], where clean and mixed sources are re-weighted with one weight in case of two sources, but extends these approaches to multi-source scenario with segmentation and multiple weights.

2) Weighting to equilibrate segment types: Another potential problem of learning strategy presented in section IV-B is that the types of segment activities for an arbitrary segmentation can have different sizes (e.g., "bass"-only segment is too long as compared to "piano"-only segment or "bass +vocals"-only segment is too long as

compared to "piano+vocals"-only segment). Such a "dis-equilibrated" structure could lead to over-fitting of the NTF model on a particular source, while modeling poorly other sources. To overcome this issue we propose the following weighting

$$b_{lfn}(\mu) = [1/\mathrm{segm\_len}_n]^{\mu}, \tag{26}$$

where $\mu \in [0, 1]$ is a fixed parameter, and $\mathrm{segm\_len}_n$ denotes the number of frames of the same segment type as the n-th frame (e.g., "piano+vocals" or "bass+vocals"). An example is shown on Fig. 2 (C).

3) Combined weightings: Finally, the two above weighting strategies being complementary, they can be combined as

$$b^{\mathrm{src}}_{lfn}(\lambda, \mu) = b^{\mathrm{src}}_{lfn}(\lambda)\, b_{lfn}(\mu), \tag{27}$$

$$b^{\mathrm{cmp}}_{lfn}(\lambda, \mu) = b^{\mathrm{cmp}}_{lfn}(\lambda)\, b_{lfn}(\mu). \tag{28}$$

V. Experiments

A. Data

We evaluate our approach on dev2 subset of SiSEC 2013 [1] PPMRs task development dataset including three 18 - 25 second 44100 Hz sampled stereo mixtures to be separated together with the corresponding full 2 - 4 minute recordings. The dev1 subset is not considered since it does not include full recordings, and thus the segmental information is very poor (almost all sources are active at the same time).

B. Parameter setting

The STFT is computed using a half-overlapping 93 ms (4096 samples) length sine window. Parameters of MU rules (19) - (22) are set to $\eta = 1$ and $\beta = 0$, which correspond to the IS divergence [3]. Thanks to the results reported in [22] on NMF and to our generalization of these results to WNTF (Sec. III), criterion (11) is non-increasing under these updates for any weighting $\mathbf{B}$.

C. Oracle components grouping

Our goal is to assess the separation performance for different considered weighting strategies we proposed. Within the user-guides source separation approach described in section IV-B, a manual user intervention is needed in the first step (to create temporal segmentation) and in the last step (to fine-tune component grouping). While the segmentation step is independent on the weighting, which is itself built on segmentation, the grouping step depends on the weighting. Given a considerable number of weightings we would like to assess, performing manual grouping for each weighting is too time consuming. Thus, inspired by [8], and given that for development data the corresponding clean sources are available, the grouping is performed in an *oracle greedy* manner, as follows. First, source estimates are initialized by zero and reconstructed components (23) are ranged in order of decreasing energy. Then, while iterating over $k = 1, \ldots, K$, source estimates are updated by adding each time component $\hat{c}_{klfn}$ to only one source so as the total mean squared error between the current source estimates and the clean sources is minimum.

D. Simulation results

The approach was run for both combined weighting types (27) and (28), for different parameters $(\lambda, \mu) \in \{0, 1, 3, 9, 27\} \times \{0, 0.33, 0.66, 0.83, 1\}$, and assessed in terms of signal-to-distortion ratio (SDR) [28] and overall perceptual score (OPS) [29] source separation measures. The results are given in table I. The best average results in terms of SDR are achieved with $b^{\mathrm{cmp}}_{lfn}(\lambda = 3, \mu = 0.66)$, and those in terms of OPS with $b^{\mathrm{cmp}}_{lfn}(\lambda = 9, \mu = 0)$. The corresponding configurations outperform the baseline non-weighted approach ($\lambda = \mu = 0$) by 0.61 dB for SDR and by 2.96 for OPS. For our SiSEC 2013 PPMRs task submission, we have chosen $b^{\mathrm{cmp}}_{lfn}(\lambda = 3, \mu = 0.66)$ weighting. With this weighting we have also assessed the results on the same dev2 subset, but with manual component grouping instead of the oracle one. Average SDR and OPS are, respectively, 3.28 dB and 27.98, which is not very different compared to the oracle results (SDR = 3.10 dB and OPS = 28.02, see table I).

[1] http://sisec2013.wiki.irisa.fr

	Measure	Average SDR (dB)					Average OPS (0 - 100)				
Weighting type	$\lambda \setminus \mu$	0	0.33	0.66	0.83	1	0	0.33	0.66	0.83	1
$b_{lfn}^{\mathrm{src}}(\lambda,\mu)$	0	*2.49*	2.20	1.96	2.36	2.42	*25.87*	22.60	26.09	24.88	24.76
	1	1.81	2.37	2.45	2.51	2.42	27.73	24.39	24.86	24.08	23.81
	3	2.42	2.74	2.59	2.09	3.09	24.62	25.93	25.16	23.86	24.06
	9	2.15	2.66	2.44	2.49	2.26	25.53	23.26	22.65	22.12	20.75
	27	1.99	2.22	1.85	1.97	2.24	21.17	20.08	19.84	19.75	20.12
$b_{lfn}^{\mathrm{cmp}}(\lambda,\mu)$	0	*2.49*	2.20	1.96	2.36	2.42	*25.87*	22.60	26.09	24.88	24.76
	1	2.15	2.36	2.54	2.38	2.59	26.86	24.80	26.78	25.68	24.30
	3	2.61	2.43	**3.10**	***3.00***	1.83	24.55	25.31	***28.02***	24.49	21.80
	9	2.81	1.79	2.13	2.57	2.34	**28.83**	25.55	21.27	24.36	22.69
	27	2.24	2.10	1.71	2.00	1.98	21.50	23.15	23.15	23.22	22.42

TABLE I
AVERAGE SDRS / OPSS FOR DIFFERENT WEIGHTING STRATEGIES ON THREE MIXTURES OF DEV2 DATASET OF SISEC 2013 PPMRS TASK.

VI. CONCLUSION

We have proven that certain results on monotonicity of MU rules for NMF generalize to WNMF, NTF and WNTF cases. Moreover, we have proposed new parametric weighting strategies for NTF-based user-guided separation of PPMRs. Some of the parametrizations of these weighting have shown improving source separation performance over the baseline non-weighted case. We have chosen the best weighting setup of the proposed approach to enter the fourth community-based Signal Separation Evaluation Campaign (SiSEC 2013). Future work will include the development and investigation of other weighting strategies for audio source separation, as well as applying WNTF for other applications.

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*, 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*, 2013, to appear.

[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] T. Virtanen, "Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1066–1074, 2007.

[9] 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.

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

[11] 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.

[12] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 3, pp. 550–563, March 2010.

[13] 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.

[14] M. Kim, J. Yoo, K. Kang, and S. Choi, "Nonnegative matrix partial co-factorization for spectral and temporal drum source separation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1192–1204, 2011.

[15] A. Lefèvre, F. Bach, and C. Févotte, "Semi-supervised NMF with time-frequency annotations for single-channel source separation," in *Proc. of the International Conference on Music Information Retrieval (ISMIR)*, Porto, Portugal, Oct. 2012, pp. 115–120.

[16] 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.
[17] 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.
[18] 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.
[19] 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.
[20] 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.
[21] 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.
[22] 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.
[23] 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.
[24] 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.
[25] 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.
[26] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *Journal of Chemometrics*, vol. 14, pp. 105–122, 2000.
[27] B. Wang and M. D. Plumbley, "Investigating single-channel audio source separation methods based on non-negative matrix factorization," in *Proc. of ICA Research Network International Workshop*, Sep. 2006, pp. 17 – 20.
[28] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, July 2006.
[29] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann, "Subjective and objective quality assessment of audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 7, pp. 2046–2057, 2011.