



Non-negative matrix factorization using weighted beta divergence and equality constraints for industrial source apportionment

Abdelhakim Limem, Gilles Delmaire, Matthieu Puigt, Gilles Roussel,
Dominique Courcot

► To cite this version:

Abdelhakim Limem, Gilles Delmaire, Matthieu Puigt, Gilles Roussel, Dominique Courcot. Non-negative matrix factorization using weighted beta divergence and equality constraints for industrial source apportionment. 23rd IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2013), Sep 2013, Southampton, United Kingdom. 10.1109/MLSP.2013.6661903. hal-01367327

HAL Id: hal-01367327

<https://hal.science/hal-01367327>

Submitted on 19 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Non-negative matrix factorization using weighted beta divergence and equality constraints for industrial source apportionment

Abdelhakim Limem, Gilles Delmaire, Matthieu Puigt, Gilles Roussel,
Dominique Courcot

► To cite this version:

Abdelhakim Limem, Gilles Delmaire, Matthieu Puigt, Gilles Roussel, Dominique Courcot. Non-negative matrix factorization using weighted beta divergence and equality constraints for industrial source apportionment. 23rd IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2013), Sep 2013, Southampton, United Kingdom. 10.1109/MLSP.2013.6661903 . hal-01367327

HAL Id: hal-01367327

<https://hal.archives-ouvertes.fr/hal-01367327>

Submitted on 19 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NON-NEGATIVE MATRIX FACTORIZATION USING WEIGHTED BETA DIVERGENCE AND EQUALITY CONSTRAINTS FOR INDUSTRIAL SOURCE APPORTIONMENT

A. Limem*, G. Delmaire, M. Puigt, G. Roussel

LISIC, EA 4491, ULCO,
Université Lille Nord de France
Calais, France, FR-62228

D. Courcot

UCEIV, EA 4492, ULCO,
Université Lille Nord de France
Dunkerque, France, FR-59140

ABSTRACT

In this paper, we propose two weighted Non-negative Matrix Factorization (NMF) methods using a β -divergence cost function. This divergence is used as a dissimilarity measure which can be tuned by the parameter β . The weights allow to deal with the uncertainty associated to each data sample. Our first approach consists of generalizing weighted NMF methods proposed with specific divergences or norms to the β -divergence. In our second approach, we assume that some components of the factorization are known and we use them to inform our NMF algorithm. We thus consider a specific parameterization which involves these constraints. In particular, we propose specific multiplicative update rules for the minimization of this parameterization with a weighted divergence. Lastly, some experiments on simulated mixtures of particulate matter sources show the relevance of these approaches.

Index Terms— NMF, β -divergence, multiplicative updates, weighed factorization, equality constraints

1. INTRODUCTION

Non-negative Matrix Factorization (NMF) is often used in Blind Source Separation (BSS) as soon as positivity of the data is required. NMF aims to factorize a $n \times m$ data matrix X into two non-negative matrices G and F such that

$$X \approx G \cdot F. \quad (1)$$

In the BSS framework, G is a $n \times p$ mixing matrix while F is a $p \times m$ source matrix. Historically—apart from pioneering work [1]—many dissimilarity measures have been used to assess the discrepancy between the data matrix X and the estimated product $G \cdot F$: NMF was firstly using a Frobenius norm and a Kullback-Leibler divergence [2] but later, general parametric divergences, e.g., $\alpha\beta$ -divergences [3, 4], were investigated. In some problems, extra assumptions may be included in the algorithm, e.g., sparseness [5, 6], fixed row and column sums [7], or orthogonality constraints [8].

*This work was supported by the integrated steel and mining company ArcelorMittal.

In this paper, we propose two NMF approaches using the β -divergence, which is a special case of the $\alpha\beta$ -divergence. Our first method considers different uncertainties associated to the data points in X and adds Weights to a state-of-the-art β -NMF method, hence its name β -WNMF. We then assume to know some components of the matrix F . Such an assumption is realistic in the framework of the considered application, i.e., industrial source apportionment. This extra-information provides some equality Constraints in the NMF formulation, hence the name β -CWNMF of this latter. The relevance of both proposed approaches is shown by several simulations performed on a synthetic dataset of particulate matter source apportionment.

The remainder of the paper reads as follows. In Section 2, we present the concepts of the state-of-the-art β -NMF methods. Section 3 introduces our proposed methods while their performance is investigated in Section 4. We conclude and discuss of future directions of the incoming work in Section 5.

2. β -NMF METHODS

2.1. β -divergence

As distances, divergences are positive. However, they do not satisfy the symmetry nor the triangle inequality. They can be used to measure the discrepancy between two vectors or matrices. The β -divergence for NMF was first introduced by Basu *et al.* [9] for $\beta > 0$. It was then extended to $\beta \in \mathbb{R}$ by Cichocki *et al.* in [10]. The β -divergence between two arbitrary matrices P and Q may be expressed as¹

$$D^\beta(P||Q) = \begin{cases} \sum_{i,j} (\log \frac{q_{ij}}{p_{ij}} + (\frac{q_{ij}}{p_{ij}})^{-1} - 1) & \text{if } \beta = -1, \\ \sum_{i,j} (p_{ij} \log \frac{p_{ij}}{q_{ij}} - p_{ij} + q_{ij}) & \text{if } \beta = 0, \\ -\frac{1}{\beta} \sum_{i,j} (p_{ij} q_{ij}^\beta - \frac{1}{1+\beta} p_{ij}^{1+\beta} - \frac{\beta}{1+\beta} q_{ij}^{1+\beta}) & \text{otherwise,} \end{cases} \quad (2)$$

¹We used the definition adopted by Cichocki *et al.* in [4]. Please note that another formalism is also used in the literature, e.g., by F  votte and Idier in [11].

where p_{ij} and q_{ij} are elements of the matrices P and Q , respectively. The case $\beta = 0$ corresponds to Kullback-Leibler divergence whereas $\beta = 1$ accounts for the Frobenius norm. Another interesting property is the function convexity within $\beta \in [0; 1]$ if $p_{ij} > 0$ and $q_{ij} > 0$, for all i and j [4, 11]. In an NMF context, P and Q are respectively replaced by X and $G \cdot F$, i.e., β -NMF consists of estimating matrices G and F which minimize

$$\min_{G, F} D^\beta(X \| GF) \quad \text{such that} \quad G \geq 0, \quad F \geq 0. \quad (3)$$

Cichocki *et al.* [4] showed the influence of the parameter β on both the robustness to outliers and the efficiency of the factorization. Indeed, when β increases, robustness to outliers is observed whereas if β decreases towards 0, the factorization is more efficient. This implies that the Euclidean distance (i.e., $\beta = 1$) is more robust and less efficient than the Kullback-Leibler divergence (i.e., $\beta = 0$). As a result, the β parameter should be chosen as a trade-off between these values.

2.2. Existing NMF methods

Aside from pioneering methods [1], NMF is performed through an iterative procedure which alternatively minimizes Eq. (3) for fixed F (respectively G) under positivity of its components. First results were obtained Lee and Seung [2] who derived multiplicative update rules for the Frobenius norm and the Kullback-Leibler divergence. However, the initialization of the algorithm and the presence of a few outliers may drastically affect the performance of the algorithm. β -NMF reduces the effect of outliers, as shown in [10, 11] for example.

Using the convexity of the β -divergence within the range $\beta \in [0; 1]$, Févotte and Idier [11] proposed general update rules based on a Maximization Minimization (MM) strategy. These rules may be written in matrix form as follows

$$F = F \circ \frac{G^T (X \circ (GF)^{\beta-1})}{G^T (GF)^\beta}, \quad G = G \circ \frac{(X \circ (GF)^{\beta-1}) F^T}{(GF)^\beta F^T}, \quad (4)$$

where $X \circ Y$ and $\frac{X}{Y}$ respectively denote the componentwise product and the elementwise division between two matrices.

3. PROPOSED METHODS

3.1. Incorporating weights into β -NMF

Weights may be used to model (possibly drastically) different uncertainties associated to each data sample. Weighted NMF has been firstly investigated by Guillaumet *et al.* [12] and Ho [13] for Frobenius norm and Kullback-Leibler divergences, respectively. In this paper, we first extend their respective work to the β -NMF framework, hence providing a

β -Weighted NMF (β -WNMF). It leads to the minimization of the following β -divergence, for $\beta \in [0; 1]$ w.r.t. $G \cdot F$:

$$D_W^\beta(X \| GF) \triangleq \frac{1}{\beta} \sum_{ij} w_{ij} \left[\frac{1}{1+\beta} X^{1+\beta} + \frac{\beta}{1+\beta} [GF]^{1+\beta} - X \circ [GF]^\beta \right]_{ij}. \quad (5)$$

W is the weight matrix whose each element w_{ij} is set to $w_{ij} = \sigma_{ij}^{-(\beta+1)}$, where σ_{ij} is referred to the uncertainty associated to x_{ij} .

The update rules (the proof is omitted for space considerations) for β -WNMF are straightforward by combining Eq. (4) and the structure of the proof in [13]. It may be gathered with the case $\beta = 0$ [14] leading to a valid expression in the range $\beta \in [0; 1]$:

$$F = F \circ \frac{G^T (W \circ X \circ (GF)^{\beta-1})}{G^T (W \circ (GF)^\beta)}, \quad (6)$$

$$G = G \circ \frac{(W \circ X \circ (GF)^{\beta-1}) F^T}{(W \circ (GF)^\beta) F^T}.$$

3.2. Proposed constrained β -WNMF

3.2.1. Parameterization of linear equality constraints

In many cases, some known components on the source profile may be provided by experts. Here, we propose a parameterization that takes into account this knowledge. Let Ω be a $p \times m$ binary matrix which informs the presence or the absence of constraints on each element of the profile matrix F , i.e.,

$$\omega_{ij} = \begin{cases} 1 & \text{if } F_{ij} \text{ has to be set,} \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

We denote as Φ the $p \times m$ sparse matrix of set values, i.e.

$$\Phi \triangleq F \circ \Omega. \quad (8)$$

By construction, φ_{ij} —the (i, j) -th element of Φ —is equal to zero when $\omega_{ij} = 0$. We can easily prove that

$$\Phi \circ \Omega = \Phi, \quad \Phi \circ (1 - \Omega) = 0. \quad (9)$$

Let f_i and φ_i be the i -th column of F and Φ , respectively. Let M_i be the $l_i \times p$ constraint matrix issued from the i -th column of Ω (i.e., containing l_i constraints), and δ_i be the vector of the constraint values extracted from Φ . The constraints on F may be expressed for the i -th column vector of the profile matrix as

$$M_i f_i = M_i \varphi_i = \delta_i. \quad (10)$$

For the sake of clarity, let us introduce an example with, say $p = 5$ sources and $l_i = 2$ constraints. Let us assume that the constraints—respectively equal to 80 and 30—are set on the second and the fourth position. The above matrices then read

$$M_i = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad \varphi_i = \begin{bmatrix} 0 & 80 & 0 & 30 & 0 \end{bmatrix}^T, \quad \delta_i = \begin{bmatrix} 80 & 30 \end{bmatrix}^T. \quad (11)$$

Let Γ_i be the $p \times (p - l_i)$ matrix such that its columns span the supplementary space to the rows of M_i , i.e., such that

$$\text{rank} \begin{bmatrix} M_i^T & \Gamma_i \end{bmatrix} = p. \quad (12)$$

Γ_i satisfies the following normalization relationship

$$\begin{cases} M_i \Gamma_i &= 0_{l_i \times (p - l_i)} \\ \Gamma_i^T \Gamma_i &= I_{(p - l_i) \times (p - l_i)} \end{cases}. \quad (13)$$

The columns of F may be expressed as

$$f_i = \varphi_i + \Gamma_i \theta_i, \quad (14)$$

where θ_i is the $(p - l_i) \times 1$ vector of free parameters. From Eq. (14), we define Δf_i as

$$\Delta f_i \triangleq f_i - \varphi_i = \Gamma_i \theta_i, \quad (15)$$

and ΔF as the matrix gathering each column Δf_i , i.e.,

$$\Delta F \triangleq F - \Phi. \quad (16)$$

F can then be expressed with respect to ΔF and Φ as

$$F = \Delta F + \Phi, \quad (17)$$

and by noticing that $1 = \Omega + (1 - \Omega)$, we obtain

$$F = \Omega \circ \Phi + \Omega \circ \Delta F + (1 - \Omega) \circ \Delta F + (1 - \Omega) \circ \Phi. \quad (18)$$

Let us now simplify this expression. Eqs. (16), (8), and (9) yield

$$\Delta F \circ \Omega = 0. \quad (19)$$

Using the properties (9) and (19) into Eq. (18) lead to

$$F = \Phi + (1 - \Omega) \circ \Delta F. \quad (20)$$

Please note that this relationship² is the matrix form of the expression provided in Eq. (14). Both can be used in the sections below.

3.2.2. Update rules for β -CWNMF

Here we derive update rules by minimizing the weighted β -divergence cost function (5) according to the parameterization (20). In practice, we use a slightly different minimization, i.e.,

$$\min_{\Delta F \geq 0} D_W^\beta (X - G\Phi || G((1 - \Omega) \circ \Delta F)), \quad (21)$$

so that the right term of the above divergence contains the unknowns only. Minimizing such a divergence ensures that

$$X - G\Phi \approx G((1 - \Omega) \circ \Delta F). \quad (22)$$

²Actually, Eqs. (17) and (20) are theoretically equivalent. However, we consider the latter expression hereafter for algorithmic reasons.

We now focus on a column of the data since the divergence may be split into independent partial divergences. In the following, we propose to drop the index i for vectors Δf_i , φ_i , θ_i , and the Γ_i matrix. Let the superscript k be the current iteration number and let first introduce U by

$$U \triangleq G\Gamma. \quad (23)$$

The weighted β -divergence between two corresponding column vectors is written as

$$D_W^\beta (x - G\varphi || G\Delta f) = \sum_i w_i d^\beta ((x - G\varphi)_i || (G\Delta f)_i) \quad (24)$$

The proof below is performed by applying the MM theorem. By using Eqs. (15) and (23), Eq. (24) may be expressed as

$$D_W^\beta (x - G\varphi || G\Delta f) = \sum_i w_i (x - G\varphi)_i^{\beta+1} \cdot h^\beta \left(\frac{\sum_j u_{ij} \theta_j}{(x - G\varphi)_i} \right), \quad (25)$$

where

$$h^\beta(z) = -\frac{1}{\beta} \left(z^\beta - \frac{1}{\beta+1} - \frac{\beta}{\beta+1} z^{\beta+1} \right). \quad (26)$$

The second derivative of h^β reads

$$\frac{d''(h^\beta(z))}{dz} = \beta z^{\beta-1} - (\beta-1) z^{\beta-2}. \quad (27)$$

$h^\beta(z)$ turns out to be a convex function for $0 \leq \beta \leq 1$ and $z \geq 0$. Jensen's inequality may thus be applied, i.e.,

$$h^\beta \left(\frac{\sum_j u_{ij} \theta_j}{(x - G\varphi)_i} \right) \leq \sum_j \frac{u_{ij} \theta_j^k}{\sum_l u_{il} \theta_l^k} h^\beta \left(\frac{\theta_j \sum_l u_{il} \theta_l^k}{(x - G\varphi)_i \theta_j^k} \right), \quad (28)$$

where k is the current iteration number and θ_j is the j -th element of the free parameters vector. An MM algorithm is derived by minimizing an auxiliary function. The majorant function is derived from Eq. (28), i.e.,

$$J_W^\beta(\theta_j, \theta_j^k) = \sum_i w_i (x - G\varphi)_i^{\beta+1} \sum_j \frac{u_{ij} \theta_j^k}{\sum_l u_{il} \theta_l^k} h^\beta \left(\frac{\theta_j \sum_l u_{il} \theta_l^k}{(x - G\varphi)_i \theta_j^k} \right). \quad (29)$$

The differentiation of the auxiliary function $J_W^\beta(\cdot)$ leads to the following expression

$$\begin{aligned} \frac{\partial J_W^\beta}{\partial \theta_j} &= \left(\frac{\theta_j}{\theta_j^k} \right)^\beta \cdot \left(\sum_i w_i u_{ij} \cdot \left(\sum_l u_{il} \theta_l^k \right)^\beta \right) \\ &- \left(\frac{\theta_j}{\theta_j^k} \right)^{\beta-1} \cdot \left(\sum_i w_i u_{ij} \cdot (x - G\varphi)_i \cdot \left(\sum_l u_{il} \theta_l^k \right)^{\beta-1} \right). \end{aligned} \quad (30)$$

The optimum is obtained by cancelling the gradient $\frac{\partial J_w^\beta}{\partial \theta_j}$ in Eq. (30), thus providing the following relationship

$$\frac{\theta_j}{\theta_j^k} = \frac{\sum_i w_i u_{ij} \cdot (x - G\varphi)_i \cdot (\sum_l u_{il} \theta_l^k)^{\beta-1}}{\sum_i w_i u_{ij} \cdot (\sum_l u_{il} \theta_l^k)^\beta}, \quad (31)$$

which can be expressed in a vectorial form as

$$\frac{\theta}{\theta^k} = \frac{U^T [(x - G\varphi) \circ (U\theta^k)^{\beta-1} \circ w]}{U^T [w \circ (U\theta^k)^\beta]}. \quad (32)$$

From Eq. (15) (the index i is dropped in the proof), it yields the following update rule:

$$\Delta f^{k+1} = \Delta f^k \circ \frac{\Gamma U^T [(x - G\varphi) \circ (G\Delta f^k)^{\beta-1} \circ w]}{\Gamma U^T [w \circ (G\Delta f^k)^\beta]}. \quad (33)$$

By replacing U according to Eq. (23), Eq. (33) reads

$$\Delta f^{k+1} = \Delta f^k \circ \frac{\Gamma \Gamma^T G^T [(x - G\varphi) \circ (G\Delta f^k)^{\beta-1} \circ w]}{\Gamma \Gamma^T G^T [w \circ (G\Delta f^k)^\beta]}. \quad (34)$$

Defining $\bar{\omega} \triangleq 1_{p \times 1} - \omega$ and as $\Gamma \Gamma^T = \text{diag}(1_{p \times 1} - \omega)$, the previous equation may be written using the componentwise product for vectors

$$\Delta f^{k+1} = \Delta f^k \circ \frac{\bar{\omega} \circ (G^T [(x - G\varphi) \circ (G\Delta f^k)^{\beta-1} \circ w])}{\bar{\omega} \circ (G^T [w \circ (G\Delta f^k)^\beta])}. \quad (35)$$

This expression may be extended to matrices by noticing that the column vectors Δf , ω , φ are respectively part of the general matrices ΔF , Ω , Φ , thus yielding

$$\Delta F^{k+1} \leftarrow \Delta F^k \circ (1 - \Omega) \circ M_F, \quad (36)$$

where

$$M_F = \frac{G^T \cdot [W \circ (X - G\Phi) \circ [G(F - \Phi)]^{\beta-1}]}{G^T \cdot [W \circ [G(F - \Phi)]^\beta]}. \quad (37)$$

Finally, Eqs. (36) and (37), and the parameterization (20) are used together to update the profile matrix F .

The expression (36) is only valid within $\beta \in [0; 1]$ according to the convexity property of the β -divergence function. Extensions of this expression may be investigated by splitting the divergence into convex and concave parts (these ideas are particularly developed in [11]). Note also that $X - G\Phi$ can be viewed as the data matrix, but it changes with the update of G and it is required to be a positive matrix.

3.2.3. Algorithm

The proposed β -CWNMF approach, outlined in Algorithm 1, is based on successive estimations of F and G , followed by a normalization stage.

A projection step is firstly applied to ensure the positivity of $(X - G\Phi)$. A normalization step is also performed to the rows of the profile matrix F to keep the row sums unchanged. The contribution matrix G is then updated to keep the product $G \cdot F$ unchanged.

Algorithm 1 Update rules of the β -CWNMF method

Reading Φ and Ω .

Initialization of G and F

Normalize F and G

while the stopping rule is not checked **do**

 Check for the positivity of $(X - G\Phi)$ and project it if necessary

 Search for F at constant G according to Eq. (36)

 Search for G at constant F according to Eq. (6)

 Normalize F and G

end while

Table 1. Used matrix Ω for the β -CWNMF method.

Fe	Ca	SO ₄	Zn	Mg	Al	Cr
1	0	0	0	0	0	1
0	1	0	1	0	0	0
1	0	0	1	0	1	1

4. INDUSTRIAL SOURCE IDENTIFICATION

The proposed NMF algorithms have been tested in the framework of pollutant source separation. In this case, X describes concentrations of chemical species of all the available samples, G accounts for a contribution matrix, and F is a profile matrix. In this paper, synthetic data are generated according to industrial source profiles available in the literature [15]. Seven chemical species—listed in Table 1—are under consideration. Three industrial sources coming from steel industry (blast furnaces source, steel slag source, ores sintering plant) shown in this order on the top plot of Figure 1 are assumed to be active.

The data matrix consists of 50 samples and 7 species, with a known uncertainty measure σ_{ij} —provided by a chemical expert—associated to each data point x_{ij} . A uniform noise ranging in $[-\min\{\lambda \sigma_{ij}, x_{ij}\}; \lambda \sigma_{ij}]$ is added while keeping positivity of the data. Please note that λ is related to an input Signal-to-Noise Ratio (SNR). The actual profile matrix is given on Figure 1.

Three β -NMF methods are experimented in our tests for different values of the β parameter, i.e., the β -NMF [11], the β -WNMF, and the β -CWNMF. An uncertainty measure is used to compute the weight matrix W , according to $w_{ij} = \sigma_{ij}^{-(\beta+1)}$ in the last two methods. In the unweighted case, each element of the matrix W is set to one. An initialization of the profile matrix is provided by chemical experts, so that computations are directed toward the true solution. In the constrained method, we are able to select which components of the profile matrix are to be set, making it highly flexible. Here, 8 constraints are taken into account in Table 1, where the 1s indicate the location of the constraints in the profile matrix. To assess the quality of the fit, we use the Mixing-Error Ratio (MER) [16]. We could use instead a (normalized) divergence measure between $X - G\Phi$ and $G\Delta F$. While in

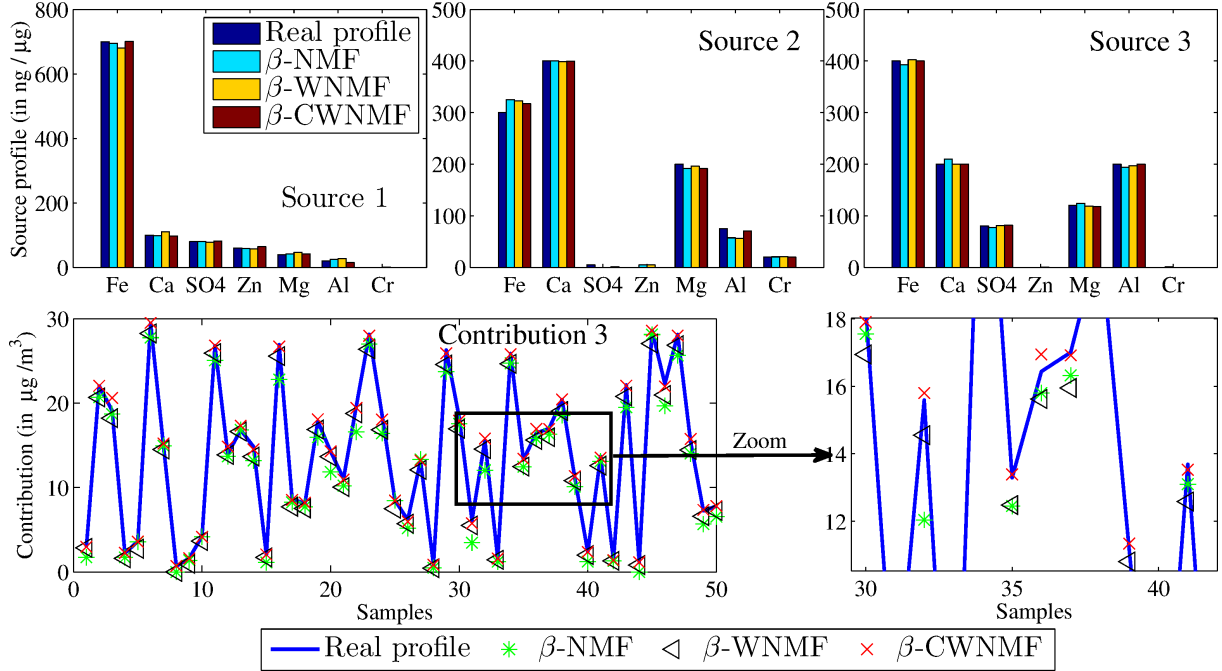


Fig. 1. Separation performance for the three tested β -NMF methods; with $\beta = 0.8$ and an input SNR=33 dB

some preliminary tests, this divergence enabled us to detect situations when the separation was not correctly performed (in these cases, the divergence was high), it was however not helpful when both G and F were correctly estimated. In that case, a lower divergence did not necessary mean a better separation. Moreover, some profiles in F are possibly correlated, then making the classical signal-to-interference ratio (computed over the estimated profiles f_i) non pertinent. As a consequence, we consider the MER which is a kind of SNR estimated on the contribution matrix G . We express each estimated column \hat{g}_j of the matrix G as

$$\hat{g}_j = g_j^{coll} + g_j^{orth}, \quad (38)$$

where g_j^{coll} and g_j^{orth} are respectively collinear and orthogonal to the true vector g_j . The MER (in dB) associated with this estimated column is then expressed as

$$\text{MER}_j = 10 \log_{10} \|g_j^{coll}\|^2 / \|g_j^{orth}\|^2. \quad (39)$$

We then derive a global MER for each estimated matrix G by adding all the MER_j .

Figure 2 provides the global MER versus the value of β , for the three tested β -NMF methods and with an SNR equal to 33 dB. Interestingly, the weighted NMF methods provide almost constant MERs for a wide range of values of β ($\beta \leq 0.7$ in our tests), hence showing the relevance of uncertainties in the NMF algorithm. Moreover, we notice that our β -CWNMF outperforms both unconstrained methods for all the β parameters in the range $[0; 1]$. This is consistent with the performance shown on Figure 1. This figure shows three industrial

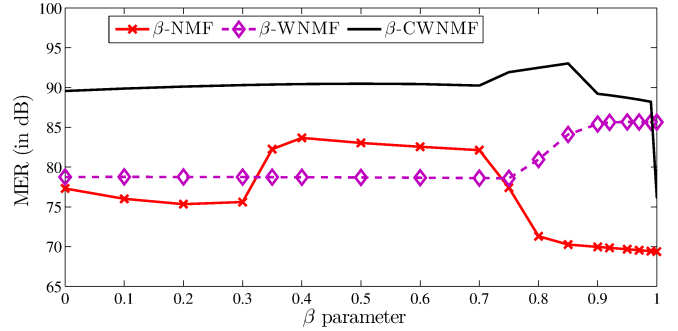


Fig. 2. Performance of the NMF methods with respect to β

source profiles and one contribution estimated by each tested NMF method (with $\beta = 0.8$) in an illustrative example. Here again, the β -CWNMF seems to better fit the true data than the other tested methods.

We then set the value of β to $\beta = 0.8$ in the following experiments. We generated several observed matrices X , with an SNR ranging from 20 to 55 dB. We then ran each tested NMF method and averaged their performance on small intervals of SNRs in Figure 3. The results are consistent with those presented on Figure 2: both proposed weighted NMF methods outperform the unweighted one and, especially for the highest tested input SNRs, the β -CWNMF clearly outperforms all the methods.

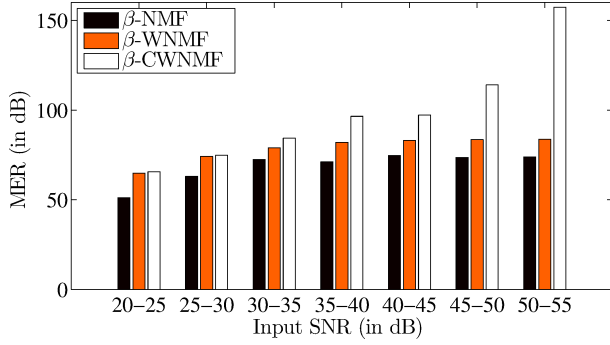


Fig. 3. Average performance of the tested NMF method, with respect to the input SNR. Performance criterion: MER

5. CONCLUSION

In this paper, we proposed two NMF methods using β -divergence and additional information. We first generalized the concept of weighted NMF—defined with the Kullback-Leibler divergence or the Frobenius norm in the literature—to the β -divergence. We then assumed to know some values of one of the factorized matrices and we introduced some equality constraints in the NMF formulation. To this end, we proposed a new parameterization which was able to involve this knowledge as an implicit constrained formulation. It resulted in multiplicative update rules which depend on two additional constraint matrices and on the weight matrix. These algorithms were tested with simulated mixtures of industrial particulate matter sources, with various signal-to-noise ratio conditions. These experiments showed the relevance of our latter approach which outperformed the tested unconstrained NMF methods. In future work, we will explore some alternatives to multiplicative update rules and, independently, other kinds of constraints to inform the NMF. We will also validate the enhancement provided by the above methods on real data.

6. REFERENCES

- [1] 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.
- [2] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Proc. of NIPS*, 2000, pp. 556–562.
- [3] A. Cichocki, R. Zdunek, S. Choi, R. Plemmons, and S. Amari, “Nonnegative tensor factorization using alpha and beta divergences,” in *Proc. of ICASSP*, 2007, pp. 1393–1396.
- [4] A. Cichocki, S. Cruces, and S. Amari, “Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization,” *Entropy*, vol. 13, pp. 134–170, 2011.
- [5] P. O. Hoyer, “Non-negative matrix factorization with sparseness constraints,” *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, November 2004.
- [6] W. Berry, M. Browne, N. Langville, V. Pauca, and J. Plemmons, “Algorithms and applications for approximate nonnegative matrix factorization,” *Computational Statistics and Data Analysis*, vol. 52, no. 1, pp. 155–173, September 2007.
- [7] N. Ho and P. Vandooren, “Non-negative matrix factorization with fixed row and column sums,” *Linear Algebra and its Applications*, vol. 429, no. 5-6, pp. 1020–1025, 2008.
- [8] J. Yoo and S. Choi, “Nonnegative matrix factorization with orthogonality constraints,” *Journal of computing science and engineering*, vol. 4, no. 2, pp. 97–109, 2010.
- [9] A. Basu, I. R. Harris, N. L. Hjort, and C. Jones, “Robust and efficient estimation by minimising a density power divergence,” *Biometrika*, vol. 3, no. 85, pp. 549–559, 1998.
- [10] A. Cichocki, R. Zdunek, and S. Amari, “Csiszár’s divergences for non-negative matrix factorization: Family of new algorithms,” in *Proc. of ICA*, 2006, pp. 32–39.
- [11] C. Févotte and J. Idier, “Algorithms for nonnegative matrix factorization with the beta-divergence,” *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011.
- [12] D. Guillaumet, J. Vitria, and B. Schiele, “Introducing a weighted non-negative matrix factorization for image classification,” *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2447–2454, 2003.
- [13] N.-D. Ho, *Non negative matrix factorizations algorithms and applications*, Phd thesis, Université Catholique de Louvain, 2008.
- [14] A. Limem, G. Delmaire, G. Roussel, and D. Courcot, “Kullback-Leibler NMF under linear equality constraints. Application to pollution source apportionment,” in *Proc. of ISSPA*, July 2012, pp. 752–757.
- [15] K. Oravissjarvi, K. L. Timonen, T. Wiikinkoski, A. R. Ruuskanen, K. Heinanen, and J. Ruuskanen, “Source contributions to pm2.5 particles in the urban air of a town situated close to a steel works,” *Atmospheric Environment*, vol. 37, no. 2, pp. 1013–1022, 2003.
- [16] E. Vincent, S. Araki, and P. Bofill, “The 2008 signal separation evaluation campaign: A community-based approach to large-scale evaluation,” in *Proc. of ICA*, 2009, pp. 734–741.