Bound constrained weighted NMF for industrial source apportionment

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

Bound constrained weighted NMF for industrial source apportionment

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

BOUND CONSTRAINED WEIGHTED NMF FOR INDUSTRIAL SOURCE APPORTIONMENT

A. Limem, M. Puigt, G. Delmaire, 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 our recent work, we introduced a constrained weighted Non-negative Matrix Factorization (NMF) method using a β-divergence cost function. We assumed that some components of the factorization were known and were used to inform our NMF algorithm. In this paper, we are provided some intervals of possible values for some factorization components. We thus introduce an extended version of our previous work combining an improved divergence expression and some matrix normalizations while using the known / bounded information. Some experiments on simulated mixtures of particulate matter sources show the relevance of these approaches.

***Index Terms*—** Blind source separation; Non-negative matrix factorization; Beta divergence; Normalization;

1. INTRODUCTION

Source apportionment consists of estimating which particulate matter sources are present in the ambiant air, with their relative concentrations. A source is fully characterized by a *profile* which gathers the m chemical species proportions (expressed in ng/ng) that constitute it. Usually, several, say n, data samples are collected from a chemical sampler. Each of them can be written as a mixture of p profiles, with different concentrations (expressed in ng/m^3). Mathematically, if we respectively denote by X, G, and F the non-negative $n \times m$ data matrix, $n \times p$ contribution matrix, and $p \times m$ profile matrix, the collected data read

$$X \approx G \cdot F. \tag{1}$$

G and F are usually unknown and estimating them from X is a Blind Source Separation (BSS) problem [1], that Non-negative Matrix Factorization (NMF) [1, Ch. 13] can solve.

The general idea behind NMF is to minimize a dissimilarity measure between X and the estimated product $G \cdot F$. Historically—apart from pioneering work [2]—many dissimilarity measures have been explored. First ones were based on the Frobenius norm and the Kullback Leibler (KL) divergence [3]. Then, some elegant parametric divergences were considered and gave rise to flexible and robust algorithms, e.g., α-NMF [4], β-NMF [5], and $\alpha\beta$-NMF [6]. Moreover, additional assumptions such as sparseness [7], fixed row and column sums [8] or orthogonality constraints [9] were investigated. However, the authors in [10] showed that, for the considered application, the state-of-the-art methods did not provide a consistent performance. As a consequence, we investigated in our recent work [11–13] the enhancement provided by *informed* NMF which takes into consideration the known values of some terms of F [11, 12] or G [13] in order to improve the separation. In [11], we introduced a specific parameterization for NMF methods using a Frobenius norm while we extended this approach to a Constrained Weighted NMF method using a β-divergence (β-CWNMF) in [12]. These approaches should be considered as a flexible NMF counterpart of [14] in between BSS—where no information on F is provided—and regression, where F is fully known. In [13], we used a pollution dispersion model using wind directions to derive the (possibly) sparse structure of G.

In this paper, we extend our previous work [12] by (i) investigating and discussing several β-divergence expressions, (ii) exploring different data normalization procedures (as profiles are chemical species proportions, the rows of F are normalized), and (iii) adding maximum and minimum bounds to some of the unknown values of F. The relevance of the proposed approaches is shown on simulations of particulate matter source apportionment.

2. UNCONSTRAINED β-NMF METHODS

Aside from pioneering work [2], NMF is classically performed through an iterative procedure which alternatively minimizes—for a fixed F (respectively G)—a discrepancy between X and $G \cdot F$. Multiplicative update rules for the Frobenius norm and the KL divergence were firstly proposed in [3]. However, the algorithm is sensitive to the presence of outliers but using the β-divergence reduces this effect [5].

The β-divergence between two arbitrary non-negative matrices P and Q is convex [5, 6] if $\beta \in [0, 1]$, and may be ex-

*This work was supported partly by the integrated steel and mining company ArcelorMittal, and partly by the DREAL Nord – Pas de Calais Agency.

pressed [6] on $]0,1]$ w.r.t. the terms p_{ij} and q_{ij} of P and Q:

$$D^{\beta}(P\|Q) = -\frac{1}{\beta}\sum_{i,j}(p_{ij}q_{ij}^{\beta} - \frac{1}{1+\beta}p_{ij}^{1+\beta} - \frac{\beta}{1+\beta}q_{ij}^{1+\beta}). \quad (2)$$

In an NMF context, P and Q are respectively replaced by X and $G \cdot F$, i.e., the β-NMF estimates G and F according to

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

where the matrix operator $\geq$ means that each entry f_{ij} or g_{ij} is positive. The work in [6] showed the influence of the parameter β within the interval $[0,1]$—the higher (respectively the lower) the value of β, the more robust (respectively the more efficient) the β-NMF method—and concluded that the β parameter should be set as a trade-off in the interval $[0,1]$.

Using a Majorization-Minimization (MM) strategy, some general update rules for β-NMF were proposed in [5]. Extending the work in [15, 16] from one hand and in [5] from the other hand, we introduced in [12] a Weighted β-NMF (β-WNMF) defined for $\beta \in [0;1]$ and using

$$F = F \circ \mathcal{N}_F(G), \qquad G = G \circ \mathcal{N}_G(F), \quad (4)$$

$$\mathcal{N}_F(G) \triangleq [G^T(W\circ X\circ(GF)^{\beta-1})]/[G^T(W\circ(GF)^{\beta})], \quad (5)$$

$$\mathcal{N}_G(F) \triangleq [(W\circ X\circ(GF)^{\beta-1})F^T]/[(W\circ(GF)^{\beta})F^T], \quad (6)$$

where $X \circ Y$ and $\frac{X}{Y}$ respectively denote the componentwise product and division between two matrices. W is a weight matrix used to model the uncertainties σ_{ij} associated to the data samples x_{ij}, and whose each element w_{ij} is set to $w_{ij} \triangleq \sigma_{ij}^{-(\beta+1)}$. This approach is equivalent to β-NMF [5] if $W = 1_{nm}$, i.e., if for any i and j, $w_{ij} = 1$.

3. PARAMETERIZATION OF CONSTRAINTS

In this paper, we assume the values of some components of the profile matrix F to be provided or bounded by experts. We thus propose a parameterization which takes into account this knowledge. It extends our previous parameterization [12] which only considered equality constraints.

Let Ω^E and Ω^I be two $p \times m$ binary matrices which inform the presence / absence of equality and inequality constraints on each element f_{ij} of the matrix F, respectively, i.e.,

$$\omega_{ij}^E = \begin{cases} 1 & \text{if } f_{ij} \text{ is known,} \\ 0 & \text{otherwise,} \end{cases} \quad \omega_{ij}^I = \begin{cases} 1 & \text{if } f_{ij} \text{ is bounded,} \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

We then define the $p\times m$ binary matrices $\overline{\Omega}^E$ and $\overline{\Omega}^I$ as $\overline{\Omega}^E \triangleq 1_{pm} - \Omega^E$ and $\overline{\Omega}^I \triangleq 1_{pm} - \Omega^I$. By construction, we obtain

$$\Omega^E \circ \Omega^I = 0, \qquad \Omega^I \leq \overline{\Omega^E}. \quad (8)$$

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

$$\Phi^E \triangleq F \circ \Omega^E. \quad (9)$$

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

$$\Phi^E \circ \Omega^E = \Phi^E, \qquad \Phi^E \circ \overline{\Omega}^E = 0. \quad (10)$$

Similarly, we define Φ^{I+} and Φ^{I-} the $p \times m$ sparse matrices of upper and lower bounds[1], respectively, i.e.,

$$\Phi^{I-} \leq F \circ \Omega^I \leq \Phi^{I+}. \quad (11)$$

Let f_i and φ_i^E be the i-th column of F and Φ^E, respectively. A column f_i may be expressed as

$$f_i = \varphi_i^E + \Gamma_i\theta_i, \quad (12)$$

where θ_i and Γ_i are respectively the $(p - l_i) \times 1$ vector of free parameters and the $p \times (p - l_i)$ orthonormal basis of free parameters [12]. From Eq. (12), we define Δf_i as

$$\Delta f_i \triangleq f_i - \varphi_i^E = \Gamma_i\theta_i, \quad (13)$$

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

$$\Delta F \triangleq F - \Phi^E. \quad (14)$$

Following the stages in [12]—which combines Eqs. (14), (9), and (10)—we obtain the matrix form of Eq. (12):

$$F = \Omega^E \circ \Phi^E + \overline{\Omega}^E \circ \Delta F. \quad (15)$$

The proposed informed NMF methods consist of estimating the matrices G and F satisfying Eq. (3) under the above constraints, i.e.,

$$\min_{G\geq 0, F\geq 0} D_W^{\beta}(X||G\cdot F) \text{ s.t. } \begin{cases} F \circ \Omega^E = \Phi^E, \\ \Phi^{I-} \leq F \circ \Omega^I \leq \Phi^{I+}, \\ F \cdot 1_{mm} = 1_{pm}. \end{cases} \quad (16)$$

The last condition forces each row of F to be normalized[2], i.e., $\sum_{j=1}^{m} f_{ij} = 1, \forall i = 1 \ldots p$. Using the parameterization (15), Eq. (16) can be expressed as

$$\min_{G\geq 0, \Delta F\geq 0} D_W^{\beta}\left(X||G\cdot(\Omega^E\circ\Phi^E) + G\cdot(\overline{\Omega}^E\circ\Delta F)\right) \text{ s.t. } \begin{cases} \Phi^{I-} \leq \Delta F \circ \Omega^I \leq \Phi^{I+}, \\ \Delta F \cdot 1_{mm} = 1_{pm} - \Phi^E \cdot 1_{mm}. \end{cases} \quad (17)$$

The last condition is derived from the last one in Eq. (16) combined with the definition (14).

[1] Equality constraints could be considered as inequalities, with the same upper and lower bounds. However, in some preliminary tests, we found our proposed approaches to outperform those using bound constraints only.

[2] Please note that the normalization met in remote sensing is not similar, as the sum of each row of F is only *approximately* known in that case.

In [12], we used a different minimization. We recall it hereafter while adding in this paper the bound constraints:

$$\min_{G\geq 0, \Delta F\geq 0} D_W^\beta\left(X - G\Phi^E || G\cdot(\overline{\Omega}^E \circ \Delta F)\right) \quad \text{s.t.} \begin{cases} \Phi^{I-} \leq \Delta F \circ \Omega^I \leq \Phi^{I+}, \\ \Delta F \cdot 1_{mm} = 1_{pm} - \Phi^E \cdot 1_{mm}. \end{cases} \tag{18}$$

We show in Section 4 how the update rules derived from Eq. (17) extends those obtained from Eq. (18).

Instead of looking for the solution of Eq. (17) directly, we propose to sequentially consider each additionnal information, i.e., we first estimate ΔF and F that we then normalize and project in the set of admissible solutions (or that we project and then normalize, respectively).

4. NEW β-CWNMF UPDATE RULES

As explained above, we first aim to find an estimate of F which satisfies the equality constraints. In [12], we considered the minimization problem (18)—using non-negativity constraints only—and derived the following update rules:

$$\triangle F^{k+1} \leftarrow \triangle F^k \circ \overline{\Omega}^E \circ M_{F^k}, \tag{19}$$

where

$$M_{F^k} = \frac{G^T \cdot \left[W \circ (X - G\Phi^E) \circ [G(F^k - \Phi^E)]^{\beta-1}\right]}{G^T \cdot \left[W \circ [G(F^k - \Phi^E)]^{\beta}\right]}. \tag{20}$$

In this paper, we introduce alternative update rules that we derive from Eq. (17). We first focus on a column of the data since the divergence may be split into independent partial divergences. We drop hereafter the index i for the vectors $\triangle f_i$, φ_i^E, θ_i, and the matrix Γ_i. Let k be the current iteration index and let us define

$$U \triangleq G\Gamma. \tag{21}$$

The weighted β-divergence between two corresponding column vectors reads

$$D_W^\beta(x||G\varphi^E + G\Delta f) = \sum_i w_i D^\beta\left(x_i||(G\varphi^E)_i + (G\Delta f)_i\right), \tag{22}$$

and may be rewritten using Eq. (2) as

$$D_W^\beta(x||G\varphi^E + G\Delta f) = \sum_i w_i x_i^{\beta+1} h^\beta\left(\frac{(G\varphi^E)_i + \sum_j u_{ij}\theta_j}{x_i}\right), \tag{23}$$

where

$$h^\beta(z) = -\frac{1}{\beta}\left(z^\beta - \frac{1}{\beta+1} - \frac{\beta}{\beta+1} z^{\beta+1}\right). \tag{24}$$

Noticing that $h^\beta(z)$ is convex for $z \geq 0$ and $0 \leq \beta \leq 1$ [12], and that $h^\beta(1) = 0$, Jensen's inequality may be applied twice:

$$h^\beta\left(\frac{(G\varphi^E)_i + \sum_j u_{ij}\theta_j}{x_i}\right) \leq \frac{(x - G\varphi^E)_i}{x_i} h^\beta\left(\frac{\sum_j u_{ij}\theta_j}{(x-G\varphi^E)_i}\right) \tag{25}$$

and

$$h^\beta\left(\frac{\sum_j u_{ij}\theta_j}{(x-G\varphi^E)_i}\right) \leq \sum_j \frac{u_{ij}\theta_j^k}{\sum_l u_{il}\theta_l^k} h^\beta\left(\frac{\theta_j \sum_l u_{il}\theta_l^k}{(x-G\varphi^E)_i \theta_j^k}\right), \tag{26}$$

where the superscript 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 a majorant function derived from Eqs. (25) and (26), i.e.,

$$J_W^\beta(\theta_j, \theta_j^k) = \sum_i w_i x_i^\beta (x - G\varphi^E)_i \sum_j \frac{u_{ij}\theta_j^k}{\sum_l u_{il}\theta_l^k} h^\beta\left(\frac{\theta_j \sum_l u_{il}\theta_l^k}{(x - G\varphi^E)_i \theta_j^k}\right). \tag{27}$$

Cancelling its gradient $\frac{\partial J_W^\beta}{\partial \theta_j}$ leads to its optimum, i.e.,

$$\frac{\theta}{\theta^k} = \frac{U^T\left[w \circ x^\beta \circ [(x - G\varphi^E)^+]^{1-\beta} \circ (U\theta^k)^{\beta-1}\right]}{U^T\left[w \circ x^\beta \circ [(x - G\varphi^E)^+]^{-\beta} \circ (U\theta^k)^{\beta}\right]}. \tag{28}$$

The superscript $^+$ here denotes the function defined as

$$(z)^+ \triangleq \max\{\epsilon, z\}, \tag{29}$$

where ϵ is a user-defined threshold chosen to avoid the division by zero in Eq. (28). By combining the definition (13) with the above relationship, by replacing U according to Eq. (21), and by noticing that $\Gamma\Gamma^T = \text{diag}(\overline{\omega})$, we derive the following update rule: $\triangle f^{k+1} \leftarrow \triangle f^k \circ N_{f^k}$ where

$$N_{f^k} \triangleq \frac{\overline{\omega}^E \circ \left(G^T\left[w \circ x^\beta \circ [(x - G\varphi^E)^+]^{1-\beta} \circ (U\theta^k)^{\beta-1}\right]\right)}{\overline{\omega}^E \circ (G^T[w \circ x^\beta \circ [(x-G\varphi^E)^+]^{-\beta} \circ (U\theta^k)^\beta])}. \tag{30}$$

Similarly to [12], it yields to the update rule by combining Eq. (14) and the matrix form of Eq. (30), i.e.,

$$F^{k+1} \leftarrow \Phi^E + \Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}, \tag{31}$$

where[3]

$$N_{F^k} = \frac{G^T\left[W \circ X^\beta \circ (X - G\Phi^E)^{1-\beta} \circ \{G[(F^k - \Phi^E) \circ \overline{\Omega}^E]\}^{\beta-1}\right]}{G^T\left[W \circ X^\beta \circ (X - G\Phi^E)^{-\beta} \circ \{G[(F^k - \Phi^E) \circ \overline{\Omega}^E]\}^{\beta}\right]}. \tag{32}$$

It should be noticed that N_{F^k} can be written under the same form as M_{F^k} in Eq. (20) by defining

$$W' \triangleq W \circ X^\beta \circ (X - G\Phi^E)^{-\beta}. \tag{33}$$

This means that our newly proposed update rule extends the previous one by iteratively updating the weights.

[3] It should be noticed that the operator (29) is also applied in practice to the terms of N_{F^k}. However, we dropped it in Eq. (32) for the sake of readability.

5. NORMALIZATION PROCEDURES

In the considered application, the rows of the profile matrix are summed to 1. As a consequence, in our previous work [11, 12], we used to normalize the matrices G and F in each iteration, after estimating them. In this paper, we investigate an alternative normalization procedure. They are introduced in the framework of the above approach, but may be extended to our previous ones [11, 12] as well.

Here we express the first normalization, used in our previous papers, as a special matrix product. Eq. (31) stands for the update rules for the unnormalized version of F, denoted $\widetilde{F}$ hereafter. The normalization is obtained by dividing by each line of $\widetilde{F}^{k+1}$ by its sum, which provides the following normalized update rule:

$$F^{k+1} \leftarrow \frac{\Phi^E + \Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}}{[\Phi^E + \Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}] \cdot 1_{mm}}. \tag{34}$$

We then derive G^{k+1} as

$$G^{k+1} \leftarrow G^k \circ N_{G^k}(\widetilde{F}^{k+1}) \circ \left[1_{nm} \cdot (\Phi^E + \Delta F^k \circ \overline{\Omega}^E \circ N_{F^k})^T\right], \tag{35}$$

where $N_{G^k}(\widetilde{F}^{k+1})$ is set according to Eq. (6). The right term in Eq. (35) is chosen so that $G^{k+1}F^{k+1} = \widetilde{G}^{k+1}\widetilde{F}^{k+1}$ while ensuring the decrease of the divergence (17) at each iteration. Although the constraints are lost within iterations, we noticed in preliminary tests that they were recovered asymptotically.

As an alternative, here we propose to keep the constraints on F and to normalize the free parameters only (second normalization). Using the above notations, the normalization then reads

$$F^{k+1} \leftarrow \Phi^E + \frac{\Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}}{(\Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}) \cdot 1_{mm}} \circ (1_{pm} - \Phi^E \cdot 1_{mm}). \tag{36}$$

This normalization keeps the constraints verified within the iterations but may prevent from moving along the steepest descent direction. Once this matrix is normalized, we estimate G^{k+1} using the rule in Eq. (4).

6. EXTENDED NMF WITH BOUND CONSTRAINTS

We now introduce our proposed Bound Constrained Weighted β-NMF (β-BCWNMF). It is well known that classical NMF approaches—i.e., [3] and its extensions—do not necessarily converge to a global optimum. In the considered application, we are provided some intervals of possible values for some entries of F. This might prevent the NMF methods to move along wrong directions.

Using the properties in Eq. (8), Eq. (31) may be split into free and bounded terms, i.e.,

$$F^{k+1} \leftarrow \Phi^E + \Delta F^k \circ \overline{\Omega}^E \circ \overline{\Omega}^I \circ N_{F^k} + \Delta F^k \circ \overline{\Omega}^E \circ \Omega^I \circ N_{F^k}. \tag{37}$$

We propose to project the bounded terms—which would be outside their interval—to the closest bound. This might be considered as an extension of both the work in [17]—which is based on projected gradients—and Eq. (31). After this projection, the update rule for an unormalized estimate $\widetilde{F}$ of F reads

$$\begin{aligned}\widetilde{F}^{k+1} \leftarrow \Phi^E + \Delta F^k \circ \overline{\Omega}^E \circ \overline{\Omega}^I \circ N_{F^k} \\ + P_{\Omega^I}\left(\Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}\right),\end{aligned} \tag{38}$$

where $P_{\Omega^I}(.)$ is the projection operator defined as

$$P_{\Omega^I}(U) \triangleq \begin{cases} \Omega^I \circ \Phi^{I-} & \text{if } \Omega^I \circ U \leq \Phi^{I-}, \\ \Omega^I \circ \Phi^{I+} & \text{if } \Omega^I \circ U \geq \Phi^{I+}, \\ \Omega^I \circ U & \text{otherwise.} \end{cases} \tag{39}$$

We lastly normalize the above profile matrix $\widetilde{F}^{k+1}$, using one of both normalization procedures discussed in Section 5. We thus derive two possible update rules, i.e.,

$$F^{k+1} \leftarrow \frac{\widetilde{F}^{k+1}}{\widetilde{F}^{k+1} \cdot 1_{mm}} \tag{40}$$

and

$$F^{k+1} \leftarrow \Phi^E + \frac{\widetilde{F}^{k+1} - \Phi^E}{(\widetilde{F}^{k+1} - \Phi^E) \cdot 1_{mm}} \circ (1_{pm} - \Phi^E \cdot 1_{mm}). \tag{41}$$

As an alternative, we can also apply the projection operator P_{Ω^I} on the normalized matrices obtained in Eqs. (34) and (36). The resulting update rules respectively read

$$\begin{aligned}F^{k+1} \leftarrow \frac{\Phi^E + \Delta F^k \circ \overline{\Omega}^E \circ \overline{\Omega}^I \circ N_{F^k}}{[\Phi^E + \Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}] \cdot 1_{mm}} \\ + P_{\Omega^I}\left(\frac{\Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}}{[\Phi^E + \Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}] \cdot 1_{mm}}\right)\end{aligned} \tag{42}$$

and

$$\begin{aligned}F^{k+1} \leftarrow \Phi^E + \frac{\Delta F^k \circ \overline{\Omega}^E \circ \overline{\Omega}^I \circ N_{F^k}}{(\Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}) \cdot 1_{mm}} \circ (1_{pm} - \Phi^E \cdot 1_{mm}) \\ + P_{\Omega^I}\left(\frac{\Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}}{(\Delta F^k \circ \overline{\Omega}^E \circ N_{F^k}) \cdot 1_{mm}} \circ (1_{pm} - \Phi^E \cdot 1_{mm})\right).\end{aligned} \tag{43}$$

Let us re-emphasize that these projections are only a guarantee for the NMF not to be far from a "realistic" solution. We thus expect the inequality constraints to be useless after a finite number of iterations. Our previous β-CWNMF [12] can be extended as well using the same approach.

7. EXPERIMENTAL VALIDATION

In this section, we investigate the enhancement provided by our methods and compare it to the one provided by the

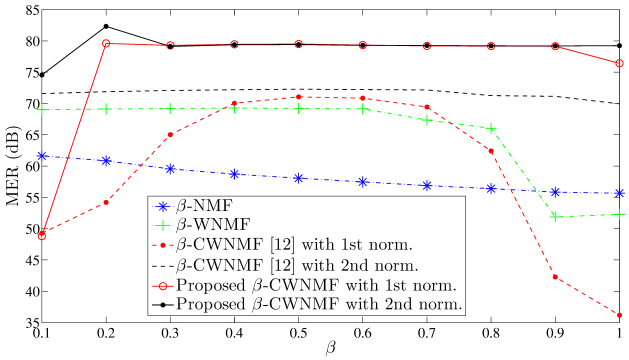

Fig. 1. NMF performance vs β. Perf. criterion: MER (in dB).

β-NMF [5], the β-WNMF [12], and the projected gradient-based NMF [17]. We use a similar configuration as in [12], i.e., 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 additive noise [12] is added and lets us vary the input Signal-to-Noise Ratio (SNR) in the experiments. We vary β from 0.1 to 1 with a step of 0.1 and for each value of β, we run 227 tests with various input SNRs ranging between 17 dB and 85 dB. We also run Lin's approach on the same data. 3 equality and 3 bound constraints are considered in informed NMF and are shown with their values in Table 1, where the crosses mean that no constraint is applied (bound constraints should be seen as crosses for the β-CWNMF variants). All the algorithms are stopped after 5e4 iterations and their performance is evaluated [12] using the Mixing-Error Ratio (MER) [18].

Table 1. Positions and values of the constraints used in the informed NMF methods. X means no constraint.

	Fe	Ca	SO_4	Zn	Mg	Al	Cr
Source 1	X	X	X	40/80	X	X	0
Source 2	280/320	X	5	X	X	X	20
Source 3	X	180/220	X	X	X	X	X

Firstly, we only compare the performance of the β methods which do *not* use the bound constraints. Figure 1 shows their average MERs (computed over the input SNRs) with respect to the value of β. Most informed NMF methods always outperform both blind approaches. In particular, the methods using the divergence (17) provide MERs around 80 dB on a wide range of β values while the approaches using the previous parameterization [12] provide MERs lower than 73 dB. We also notice that the second normalization does not seem to be sensitive to the value of β while the methods using the first one provide a limited BSS enhancement for low and high values of β. This is particularly visible with our previous β-CWNMF [12] which slightly outperforms the blind β-NMF methods when $0.4 \leq \beta \leq 0.7$ only. These tests emphasize the relevance of both the new parameterization and the second normalization we introduced in this paper.

Figure 2 shows the mean MERs (averaged over small groups of input SNRs [12] and over the value of β for β methods) and the associated standard deviations, obtained with both β-CWNMF—providing the best performance in Figure 1—and with the five bound constrained methods (the mean MERs provided by [17] are plotted in dash line on the same plots as our proposed methods). All the plots provide the same shape: the MER grows with the input SNR. However, the mean values and the spreads may differ. Let us first focus on the β-CWNMF approaches. Again, the second normalization makes the NMF less sensitive to the choice of β than the first one. Indeed, the mean MERs and their spreads are respectively slightly higher and much lower than with the first one. These approaches are outperformed by the β-BCWNMF methods, except for the lowest input SNRs. The latter may be explained as follows: the estimates of F are close to the theoretical values and all the effects of the noise in X are reported in the estimation of G, from which we estimate the MERs. Adding information on G helps improving the MERs in this case, as shown in [13]. Except for the highest SNRs—where the approach using Eq. (40) slightly outperforms the others—the β-BCWNMF methods almost provide the same performance. This might be due to the fact that the projections help the NMF to keep close to a relevant solution, which might limit the effects due to the normalizations noticed above. Moreover, we found in some preliminary tests that the final estimates of F did not get any value of Φ^{I-} or Φ^{I+}, which means that no projection was applied during at least the last iteration. This implies that the update rules (40) and (42) from one hand, and (41) and (43) from the other hand, are equal to Eq. (34) and Eq. (36), respectively, thus providing the same performance. Finally, all the β-BCWNMF variants outperform Lin's approach [17], thus showing the relevance of our approaches.

8. CONCLUSION

In this paper, we extended our previous informed β-NMF [12]—assuming some values of one of the factorized matrices to be known—to alternative update rules and normalizations, while adding some bound constraints. The relevance of these extensions were shown on simulated mixtures of industrial particulate matter sources, with various input SNR conditions. In future work, we will investigate other constraints to inform the NMF, and alternatives to multiplicative updates. As in [11], we will also validate the enhancement provided by our methods on real data.

9. REFERENCES

[1] P. Comon and C. Jutten, *Handbook of blind source separation. Independent component analysis and applications*, Academic press, 2010.

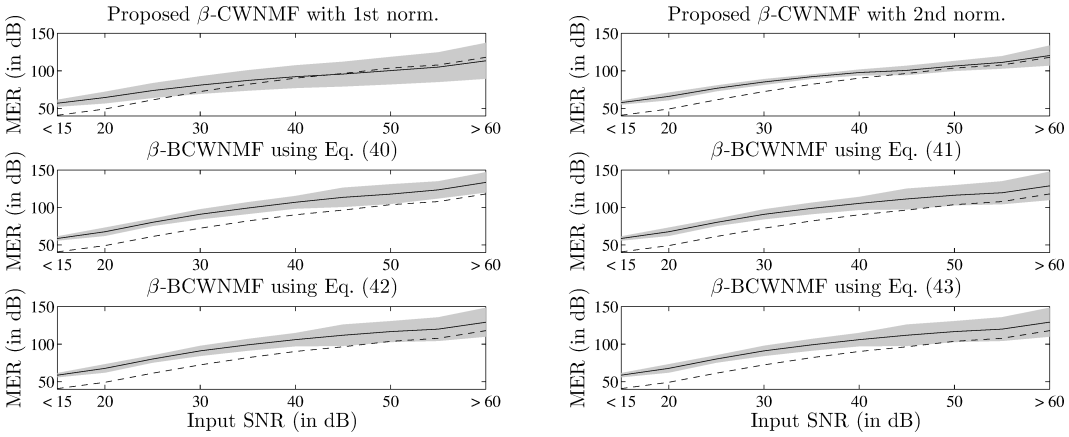

Fig. 2. NMF performance vs input SNR. Perf. criterion: MER (in dB). The plain lines and the grey areas denote the mean values and the spread of the estimated MERs. The dash lines show the mean MERs provided by Lin's approach [17].

[2] 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.

[3] D.D. Lee and H.S. Seung, "Learning the parts of objects by non negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[4] A. Cichocki, H. Lee, Y. Kim, and S. Choi, "Nonnegative matrix factorization with alpha-divergence," *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1433–1440, 2008.

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

[6] 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.

[7] P.O. Hoyer, "Non-negative matrix factorization with sparseness constraint," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, November 2004.

[8] H. Lantéri, C. Theys, C. Richard, and C. Févotte, "Split gradient method for nonnegative matrix factorization," in *Proc. of EUSIPCO*, 2010.

[9] 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.

[10] M. Viana, T. A. J. Kuhlbusch, X. Querol, and A. Alastuey, "The effects of particle-size distribution and chloride depletion of sea-salt aerosols on estimating atmospheric deposition at a coastal site," *Journal of Aerosol Science*, vol. 39, no. 10, pp. 827–849, 2008.

[11] A. Limem, G. Delmaire, M. Puigt, G. Roussel, and D. Courcot, "Non-negative matrix factorization under equality constraints—a study of industrial source identification," *Applied Numerical Mathematics*, vol. 85, pp. 1–15, November 2014.

[12] A. Limem, G. Delmaire, M. Puigt, G. Roussel, and D. Courcot, "Non-negative matrix factorization using weighted beta divergence and equality constraints for industrial source apportionment," in *Proc. of MLSP*, 2013.

[13] M. Plouvin, A. Limem, M. Puigt, G. Delmaire, G. Roussel, and D. Courcot, "Enhanced NMF initialization using a physical model for pollution source apportionment," in *Proc. of ESANN*, 2014, pp. 261–266.

[14] M. De Vos, S. Van Huffel, and L. De Lathauwer, "Spatially constrained ICA algorithm with an application in EEG processing," *Signal Processing*, vol. 91, no. 8, pp. 1963–1972, August 2011.

[15] D. Guillamet, 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.

[16] N.-D. Ho, *Non negative matrix factorization algorithms and applications*, Phd thesis, Université Catholique de Louvain, 2008.

[17] C.-J. Lin, "Projected gradients methods for non-negative matrix factorization," *Neural Computation*, vol. 19, pp. 2756–2779, 2007.

[18] 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.