



HAL
open science

Informed split gradient non-negative matrix factorization using Huber cost function for source apportionment

Robert Chreiky, Gilles Delmaire, Matthieu Puigt, Gilles Roussel, Antoine Abche

► To cite this version:

Robert Chreiky, Gilles Delmaire, Matthieu Puigt, Gilles Roussel, Antoine Abche. Informed split gradient non-negative matrix factorization using Huber cost function for source apportionment. The 16th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2016), IEEE, Dec 2016, Limassol, Cyprus. pp.69-74, 10.1109/ISSPIT.2016.7886011 . hal-01393264

HAL Id: hal-01393264

<https://hal.science/hal-01393264v1>

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.

INFORMED SPLIT GRADIENT NON-NEGATIVE MATRIX FACTORIZATION USING HUBER COST FUNCTION FOR SOURCE APPORTIONMENT

Robert Chreiky^{*†}, Gilles Delmaire^{*}, Matthieu Puigt^{*}, Gilles Roussel^{*}, and Antoine Abche[†]

^{*} Univ. Littoral Côte d’Opale, LISIC – EA 4491, F-62228 Calais, France,

[†] University of Balamand, Balamand Al Kurah, Lebanon

ABSTRACT

Source apportionment is usually tackled with blind Positive/Non-negative Matrix factorization (PMF/NMF) methods. However, the obtained results may be poor due to the dependence between some rows of the second factor. We recently proposed to inform the estimation of this factor using some prior knowledge provided by chemists—some entries are set to some fixed values—and the sum-to-one property of each row. These constraints were recently taken into account by using a parameterization which gathers all of them. In this paper, a novel robust NMF approach able to cope with outliers is proposed. For that purpose, we consider the Huber loss function—a ℓ_2 - ℓ_1 cost function—which is robust to outliers, contrary to the Frobenius norm classically met in NMF. We thus propose new update rules for the informed Huber NMF in the framework of the split gradient techniques. The choice of the adaptive cutoff parameter—which links both single cost functions—is discussed along this paper. The proposed approach is shown to outperform state-of-the-art methods on several source apportionment simulations involving various input SNRs and outliers.

Index Terms— Informed source separation, Non-negative matrix factorization, Robustness, Split gradient, Source apportionment, Huber cost function.

1. INTRODUCTION

Source apportionment aims to find particulate matter sources and their emissions in the ambient air. In practice, a device is trapping particulate matter on filters, which are then analyzed by chemical experts. They actually estimate the masses of each chemical species present in the filters, which are converted into concentrations and provide a $n \times m$ non-negative data matrix X , where n is the number of samples and m the number of species. The data matrix may be approximated by a linear mixture model also called Receptor Model in source apportionment, i.e.,

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

where G is the $n \times p$ contribution matrix gathering the emissions of all the sources over time and F is the $p \times m$ profile

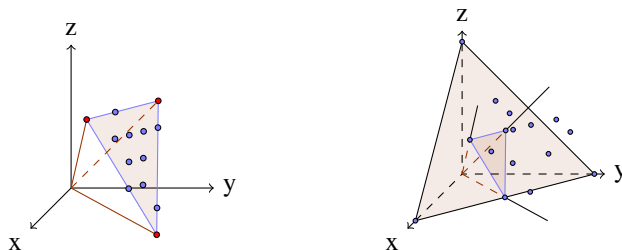


Fig. 1. Geometric point of view of the sum-to-one constraint for (left) hyperspectral data, and (right) chemical data.

matrix, where a profile is a source signature involving all the chemical species proportions (in ng/ng).

The non-negativity property of the factors usually leads to tackle the approximated factorization in Eq. (1) with NMF methods which generally carry out an alternating minimization between factors of the Frobenius norm of the residuals [1]. However, even a small number of deviating data points completely alter the quality of the approximation. Indeed, the Frobenius cost function is well-suited for Gaussian noise while degrading in the presence of atypical observations. As a result, robust NMF methods are devised to withstand a certain proportion of corrupted data points called outliers. In source apportionment, outliers may result from an additional contamination on the filter, or from a measurement error.

Two main strategies are used to handle outliers in NMF. A first class of methods [2–4] proposes to rewrite the matrix X as the sum of a low rank matrix L which can be factorized according to Eq. (1), and a sparse matrix S which should contain the outliers available in the collected data. The sparsity of S is reached by using special norms favoring this property. The second type of strategy suggests the use of a robust cost function between parts of Eq. (1), e.g., the ℓ_1 - ℓ_2 [5] or $\ell_{2,1}$ [6] norms, the Huber function [7], or a parametric divergence [8]. While the latter is asymmetric and implicitly assumes large residuals to be positive, the other cost functions generally behave as a Frobenius norm for small residuals while providing less penalization to large residuals than the Frobenius norm.

Moreover, in source apportionment, the definition of the profile matrix imposes that rows of F sum to one. It should

be noticed that sum-to-one constraints are also met in remote sensing [4, 9, 10], but they apply along the columns of the second matrix factor. As a result, each column of the data matrix may be viewed as a convex combination of the columns—called endmembers—of the first matrix factor. The geometric point of view (see the left plot of Fig. 1) shows that the column vectors of X belong to a non-negative polytope spanned by the endmembers. Searching for the mixing matrix thus consists of splitting the whole problem into independent sub-problems related to the search of one column together with the sum-to-one constraint of this column.

In source apportionment, the issue is different since in this case, the sum-to-one constraints apply to the rows of the profile matrix F . The right plot of Fig. 1 illustrates that each row of the data matrix is only a non-negative linear combination of the profile row vectors, so that it belongs to a simplicial cone in the non-negative orthant. The only noticeable property (shown on the right plot of Fig. 1) is that each row of the profile matrix lives in a non-negative polytope spanned by the usual orthonormal basis vectors. Therefore, it is impossible to split the NMF problem into independent sub-problems. As a result, source apportionment needs some specific methods.

Moreover, blind NMF methods for source apportionment may result in solutions without physical meaning [11], especially when some source profiles are geometrically close to each other. To overcome these drawbacks, we recently proposed some informed NMF approaches [12–15] which take into account some prior information—e.g., some set values in the profile matrix provided by chemical experts—together with the row sum-to-one property. In [12, 13], the constraints were alternately satisfied along iterations and only the limit matrix was fulfilling both of them. We recently proposed a new parameterization [14, 15] which fulfills both constraints at the same time, and we then derived informed NMF methods using the split gradient strategy [16]. While the approach in [14] is based on the Frobenius norm—and is then sensitive to outliers—the one in [15] is using a robust $\alpha\beta$ -divergence, which requires the parameters α and β to be properly tuned. In this paper, we propose a novel robust NMF approach—extending [14]—which is based on the Huber cost function [17] and which does not need to some parameter tuning.

This paper is organized as follows. Sections 2 and 3 briefly recall the parameterization proposed in [14] and the Huber function, respectively. We derive the update rules of the proposed informed NMF method in Section 4 and we investigate its performance in Section 5. The conclusion and the perspectives of this work are discussed in Section 6.

2. PARAMETERIZATION OF F

In source apportionment, prior information often concerns the profile matrix¹, for which some entries are provided by ex-

perts. Let Ω be a $p \times m$ binary matrix which shows the presence/absence of constraints on each entry of F , i.e.,

$$\Omega_{ij} = \begin{cases} 1 & \text{if } F_{ij} \text{ is known,} \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Conversely, we define $\bar{\Omega} \triangleq \mathbf{1}_{pm} - \Omega$, where $\mathbf{1}_{pm}$ is the $p \times m$ matrix of ones. Defining Φ and ΔF , the $p \times m$ matrices of set and free parts of F , respectively, we derive [13]

$$F = \Omega \circ \Phi + \bar{\Omega} \circ \Delta F, \quad (3)$$

where \circ stands for the Hadamard product. It should be noticed that updating F consists of updating its free part ΔF . However, at this stage, the sum-to-one constraint is not satisfied. For that purpose, we extended in [14] the idea proposed in [16] which consists in normalizing the constrained ΔF by replacing it by an unconstrained $p \times m$ matrix Z [14, 15], i.e.,

$$F = \Omega \circ \Phi + \frac{\bar{\Omega} \circ Z}{(\bar{\Omega} \circ Z) \cdot \mathbf{1}_{mm}} \cdot (\mathbf{1}_{pm} - \Phi \cdot \mathbf{1}_{mm}), \quad (4)$$

where $/$ stands for the elementwise division. As an example, if we assume that F is a 1×4 profile matrix and $\Phi \triangleq [0, 0.2, 0, 0.3]$, then the maximum free part value is $\mathbf{1}_{pm} - \Phi \cdot \mathbf{1}_{mm} = [0.5, 0.5, 0.5, 0.5]$ and Eq. (4) reads

$$F = \left[\frac{Z_{11} \cdot 0.5}{Z_{11} + Z_{13}}, 0.2, \frac{Z_{13} \cdot 0.5}{Z_{11} + Z_{13}}, 0.3 \right].$$

In the next sections, source apportionment is addressed with the above parameterization combined with the Huber cost function.

3. HUBER COST FUNCTION

Data are often corrupted by outliers so that quadratic cost functions—i.e., using the ℓ_2 norm—become inappropriate. Robust cost functions are designed to provide less penalization to large entries of E , the matrix of residuals defined as

$$E \triangleq X - G \cdot F. \quad (5)$$

On the other side, the ℓ_1 norm may be used to achieve robustness but—because of its sparsity promoting behaviour—it tends to overweight near-zero residuals. As a consequence, intermediate solutions may be found to reduce the weight of small residuals.

One of the widely used techniques is the M-estimators [19]—defined as a generalization of maximum likelihood estimators—of the form

$$\min_{G, F} \sum_{i,j} \rho(E_{ij}), \quad (6)$$

¹Please note that it is also possible to inform G [18].

where $\rho(\cdot)$ is a loss function, e.g., ℓ_1 - ℓ_2 [5] or $\ell_{2,1}$ [6] cost functions. Another popular choice of M-estimators is the Huber function which is based on the derivable connection between the ℓ_2 and ℓ_1 norms, i.e.,

$$\rho_{\text{huber}}(e) = \begin{cases} e^2 & \text{for } |e| < c, \\ 2c|e| - c^2 & \text{otherwise,} \end{cases} \quad (7)$$

where c is the cutoff parameter of the Huber cost function, which divides the domain between the quadratic and linear loss functions. Moreover, in source apportionment, individual standard deviations are available, thus yielding weighted residuals, i.e.,

$$E_{\sigma,ij} \triangleq \frac{E_{ij}}{\sigma_{ij}}. \quad (8)$$

The Weighted Huber-NMF problem is thus defined as:

$$\min_{G, F \succeq 0} \sum_{i=1}^n \sum_{j=1}^m \rho_{\text{huber}}(E_{\sigma,ij}). \quad (9)$$

It should be noticed that $\rho_{\text{huber}}(\cdot)$ behaves as an ℓ_2 norm for small weighted residuals while it is linear for large residuals. The function is convex and it has a bounded influence function—defined as the differentiation of the cost function—so that the influence of a single deviating point is minimized.

4. PROPOSED INFORMED NMF METHOD

In this section, we aim to solve Eq. (9) subject to the parameterization (4), i.e.,

$$\begin{aligned} \min_{G, F \succeq 0} \quad & \mathcal{J}(G, F) \\ \text{s.t.} \quad & F = \Omega \circ \Phi + \frac{\bar{\Omega} \circ Z}{(\bar{\Omega} \circ Z) \cdot 1_{mm}} \cdot (1_{pm} - \Phi \cdot 1_{mm}), \end{aligned} \quad (10)$$

where $\mathcal{J}(G, F) \triangleq \sum_{i=1}^n \sum_{j=1}^m \rho_{\text{huber}}(E_{\sigma,ij})$.

4.1. Update rules for F

We now aim to find the update rules of the profile matrix F which are consistent with the KKT conditions. This is done by differentiating \mathcal{J} in Eq. (10) with respect to F for both expressions of $\rho_{\text{huber}}(\cdot)$. For that purpose, we firstly define the following weight matrices W_1 and W_2 :

$$W_{1,ij} \triangleq \frac{1}{\sigma_{ij}^2}, \quad W_{2,ij} \triangleq \begin{cases} 1 & \text{if } \frac{(GF-X)_{ij}}{\sigma_{ij}} < c, \\ \frac{c \cdot \sigma_{ij}}{|GF-X|_{ij}} & \text{otherwise.} \end{cases} \quad (11)$$

W_2 should be seen as an adaptive re-weighting matrix, depending on small/large weighed residuals. In the case of a single data point labelled as an outlier, the weight in W_2 is proportional to the inverse of residual value and is below 1.

Noticing that $W_2 = 1_{nm}$ for the ℓ_2 part of the cost function, the differentiation of \mathcal{J} for a single entry F_{rj} reads

$$\frac{\partial \mathcal{J}}{\partial F_{rj}} = G_{ir} (GF - X)_{ij} W_{1,ij}. \quad (12)$$

The differentiation of \mathcal{J} for the ℓ_1 part yields

$$\frac{\partial \mathcal{J}}{\partial F_{rj}} = \sum_{i=1}^n \frac{c}{\sigma_{ij}} \cdot G_{ir} \cdot \text{sign}((GF - X)_{ij}). \quad (13)$$

Then, using the definition of the sign function, an equivalent expression is derived, i.e.,

$$\frac{\partial \mathcal{J}}{\partial F_{rj}} = \sum_{i=1}^n \frac{c \cdot \sigma_{ij}}{\sigma_{ij}^2} \cdot G_{ir} \cdot \frac{(GF - X)_{ij}}{|GF - X|_{ij}}. \quad (14)$$

We lastly derive from Eq. (11)

$$\frac{\partial \mathcal{J}}{\partial F_{rj}} = G_{ir} (GF - X)_{ij} W_{1,ij} W_{2,ij}. \quad (15)$$

Equations (12) and (15) yield a unified matrix expression, i.e.,

$$\frac{\partial \mathcal{J}}{\partial F} = G^T [W \circ (GF - X)], \quad (16)$$

where

$$W \triangleq W_1 \circ W_2. \quad (17)$$

Let us emphasize again that the differentiation of \mathcal{J} has a similar expression for both the quadratic and the linear parts of the Huber function, except that W is reweighted for large residuals, as explained when we introduced W_2 .

At this stage, we still need to express the differentiation of $\frac{\partial F_{ij}}{\partial Z_{ik}}$ based on the parameterization (4) where the (i, j) -th entry of F only depends on the i -th row of Z and $\forall k \in \{1 \dots m\}$. We thus obtain

$$\frac{\partial F_{ij}}{\partial Z_{ik}} = \left(1 - \sum_{l=1}^m \Phi_{il}\right) \frac{\bar{\Omega}_{ij}}{\sum_{l=1}^m \bar{\Omega}_{il} Z_{il}} \left[\delta_{jk} - \frac{\bar{\Omega}_{ik} Z_{ij}}{\sum_{l=1}^m \bar{\Omega}_{il} Z_{il}} \right], \quad (18)$$

where δ_{jk} is the Kronecker index which is equal to 1 when $j = k$ and 0 otherwise. For the sake of readability, we define B as a $p \times m$ matrix of scale factors, i.e.,

$$B_{ik} \triangleq \frac{1 - \sum_l \Phi_{il}}{\left(\sum_{l=1}^m Z_{il} \bar{\Omega}_{il}\right)}. \quad (19)$$

Using the chain rule and Eqs. (18) and (19), the matrix form of $\frac{\partial \mathcal{J}}{\partial Z}$ reads [14]

$$\frac{\partial \mathcal{J}}{\partial Z} = B \circ \left[\frac{\partial \mathcal{J}}{\partial F} \circ \bar{\Omega} - \bar{\Omega} \circ \left(\left(\frac{\partial \mathcal{J}}{\partial F} \circ F \circ \bar{\Omega} \right) \cdot 1_{mm} \right) \right]. \quad (20)$$

The first term (respectively, second) of the difference in the bracket of Eq. (20) corresponds to the first term (respectively, second) of the difference in Eq. (18).

The third KKT condition has to be fulfilled to get a stationary point with respect to Z , i.e., $Z \circ \frac{\partial \mathcal{J}}{\partial Z} = 0$. Equivalently and according to Eq. (4), it may be written as

$$F \circ \bar{\Omega} \circ \frac{\partial \mathcal{J}}{\partial Z} = 0. \quad (21)$$

A stationary point means that $F = F^k = F^{k+1}$, and using Eqs. (20) and (21), and letting $U \triangleq -\frac{\partial \mathcal{J}}{\partial F}$ (see Eq. (16)), the new update rules of F read

$$F^{k+1} = \Omega \circ \Phi + \bar{\Omega} \circ \frac{F^k \circ U}{[U \circ \bar{\Omega} \circ F^k] \cdot 1_{mm}} \circ (1_{pm} - \Phi \cdot 1_{mm}). \quad (22)$$

Moreover, we need to make sure that the updated rules of F in Eq. (22) remain positive which implies that U is positive at every iteration. In [15], we proved the following theorem:

Theorem 1. *Every U_s of the form $U_s = U + S \cdot 1_{mm}$ checks the third KKT condition (21).*

As a consequence, a shift equal to the sum of the negative part of $-\frac{\partial \mathcal{J}}{\partial F}$ was proposed in [14]. Applying the same idea to Eq. (22), we derive

$$U_s = G^T(W \circ X) - G^T(W \circ (GF)) + [G^T(W \circ (GF))]1_{mm}. \quad (23)$$

4.2. Update rules for G

Due to the application used in this paper, the contribution matrix G is unconstrained. Once again, the weight W should be taken into account in the updated rules. Due to symmetry in the expression of the weighted Huber cost function, the derivation with respect to G remains similar to Eq. (16) and reads

$$\frac{\partial \mathcal{J}}{\partial G} = [W \circ (GF - X)] \cdot F \quad (24)$$

Expression (24) is similar to the differentiation of the quadratic loss function in [20], except that we use the weight matrix (17). As a consequence, the update rules for G read [20]

$$G^{k+1} = G^k \circ \left[\frac{(W \circ (X)) F^T}{(W \circ (GF)) F^T} \right], \quad (25)$$

Please notice that the update of the weight matrix should be made prior to the computation of F and G , which implies that the cutoff parameter should be carefully chosen.

4.3. Choice of the cutoff parameter

The cutoff parameter is an adaptive parameter depending on the residuals. In [7], it is chosen as the median of the residuals. In our framework, it relies on weighted residuals. While it could also be chosen as the median—i.e., half of the data are seen as outliers—we consider that this choice prevents to modify the matrix factors in a sufficient way along iterations and as a consequence, we propose a different strategy. We also think that at the beginning of the NMF iterations, all the data should be processed in the quadratic part of the Huber function, resulting that c should go to the maximum value of the data. Conversely, at the end of the iterations, we feel

that c should tend to an intermediate value, enabling to point only the outlying points. As a consequence, we propose an heuristic choice for c based on the quantile function with a probability value linearly depending on the current iteration number k , i.e.,

$$c = \text{quantile}(E_\sigma, 1 - \frac{k}{\text{Itermax}} * 0.1) \quad (26)$$

where Itermax is the maximum number of iterations. This choice implicitly states that a maximum of 10 percent of the data may be labelled as outlying data, which was found to be a good trade-off in preliminary tests.

4.4. Algorithm

The algorithm may be outlined in the following form:

Algorithm 1 Weighted Huber Informed NMF description

Reading of X , Φ and Ω

Initialization of F and G

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

 Compute E and the weighted residuals E_σ

 Compute c according to Eq. (26)

 Update W according to Eqs. (11) and (17)

 Update F according to Eq. (22) with Eq. (23)

 Update G at fixed F according to Eq. (25)

end while

5. SIMULATIONS

In this section, we are investigating the performance of the newly proposed method—named Huber SG-CWNMF for Split-Gradient Constrained Weighted NMF using Huber cost function—in several tests with and without outliers. The collected data matrix X is made of 50 samples and 7 species. Each entry in X has a known variance σ_{ij} provided by chemical experts. The collected samples are possibly corrupted by outliers located at random locations and tuned by a gain which is always greater than one in order to simulate a pollutant contamination. Then a uniform noise is added, ranging in the interval $[-\min(\lambda\sigma_{ij}; x_{ij}); \lambda\sigma_{ij}]$ while taking into consideration that the data is always positive. As a result, λ is related to an input Signal-to-Noise Ratio (SNR) without any outliers. Moreover, the profile matrix F consists of three partially correlated industrial sources for which 9 entries are known and provided in Table 1. In the simulation, the initial F is consistent with the provided constraints and the initial G is computed with a quadratic program. The performance criterion is the Mixing-Error Ratio (MER) expressed in dB, computed over each column of G and averaged.

In addition to the proposed method, several methods are tested. They are divided into two categories, i.e., the blind

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	0.7	X	X	X	X	X	0
Source 2	X	0.4	5	0	X	0.075	X
Source 3	0.4	X	X	0	X	X	0

and the informed ones. In the first category, the tested methods which are all based on β -divergence, are the original unweighted β -NMF [21], the weighted β -NMF [20] and the robust NMF (rNMF) [4] in which the data matrix is split into a low rank matrix satisfying Eq. (1) and a sparse matrix containing the outliers. Two other informed approaches based on the β -divergence are experimented (tested with $\beta = 0.8$ in this paper), i.e., our previous β -CWNMF with Residuals (β -CWNMF-R) and without residuals (β -CWNMF) [13]. In addition, we test a weighted version of the Split Gradient approach developed by Lantéri *et al.* [16], which involves a sum-to-1 parameterization, and our previous proposed $\alpha\beta$ -SG-CWNMF with $\alpha = \beta = 0.6$ [15]. All these methods are compared for several input SNR with the newly proposed one, based on the weighted Huber cost function and for 200k iterations. Finally, the upper bound of the performance is computed with the Huber cost function by assuming F is set to its real value and using the update rules (25) together with the iterative re-weighting process (11).

In a first experiment, each method is tested over 400 simulations with respect to the input SNRs ranging from 15 to 60 dB without and with several outliers. For space considerations, we only plot the case of 5 outliers on Fig. 2. Moreover, in the sake of readability, we do not show the performance of all the tested methods. First of all, our proposed Huber method outperforms the blind and informed tested methods, for any input SNR interval and for both cases with and without outliers (except our previous $\alpha\beta$ -SG-CWNMF [15] which behaves very similarly and which is not shown as a consequence). Moreover, the achieved performance is almost equal to the Bound MERs. Interestingly, the rNMF method completely fails over all the tests. This is probably due to the fact that rNMF is not able to split the data matrix into appropriate sparse and low rank matrices. Indeed, in our experiments, outliers are relatively small for high input SNRs and medium-sized for low SNRs. Such a situation is challenging for rNMF which performs well in the presence of larger outliers.

In a second experiment, we investigate the influence of an increasing number of outliers on the behavior of the tested NMF methods, for an input SNR of 40 dB without outliers. Figure 3 shows the achieved MER with respect to the the numbers of outliers (associated with the input SNR with outliers). All the methods provide decreasing MERs when the number of outliers increase. However and similarly to the previous experiment, our proposed Huber SG-CWNMF method

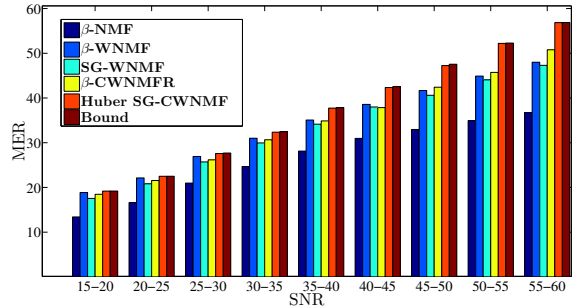


Fig. 2. MERs provided by the tested NMF methods vs. the input SNR for 5 outliers.

outperforms all the tested approaches and provides a performance very close to the Bound, which shows the relevance of this work.

Finally, we experiment different cutoff parameter strategies, either constant or adaptive with linear decrease along iterations. Figure 4 shows the obtained performance on the above simulations with each choice. In fact, the legend specifies the percentage of data points considered in the ℓ_1 part of the Huber function at the beginning and at the end of the iterations. The best choice is achieved with the cutoff proposed in Eq. (26) as indicated in Figure 4. Moreover, the usual choice of the median proposed in the literature is here shown to be inappropriate.

6. CONCLUSION

In this paper, a new robust informed NMF method is proposed for solving the source apportionment problem. It is based on a weighted symmetric Huber cost function which adaptively connects weighted ℓ_2 and ℓ_1 norms according to the size of the computed residuals. In addition, a specific parameterization and the split gradient framework allow to provide new update rules which take into account set entries and the sum-to-one constraint in one factor matrix. This method has been tested for a large range of SNRs and for several outliers. It overcomes all the tested state-of-the-art methods and besides, behaves very close to the bound computed with the Huber regression method and the exact profile matrix.

7. REFERENCES

- [1] 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.
- [2] B. Shen, L. Si, R. Ji, and B. Liu, “Robust nonnegative matrix factorization via ℓ_1 norm regularization,” *arXiv preprint arXiv:1204.2311*, 2012.

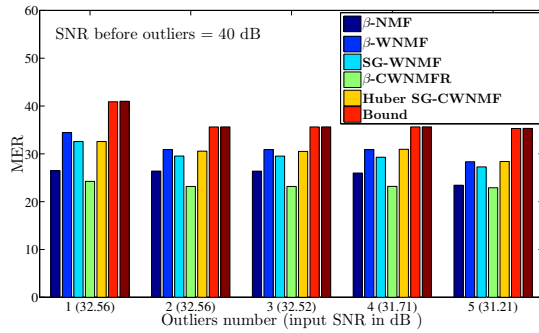


Fig. 3. NMF performance vs. the number of outliers.

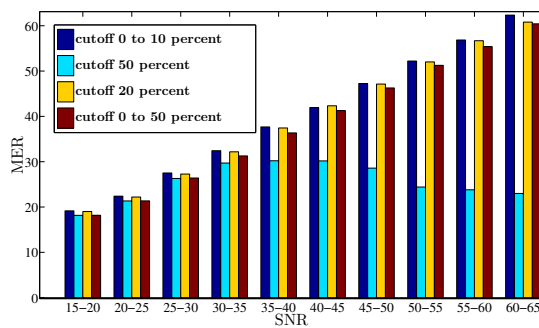


Fig. 4. MERs obtained with the Huber SG-CWNMF method vs. the input SNR for different cutoff strategies.

- [3] J. Rapin, J. Bobin, A. Larue, and J. L. Starck, “Sparse and non-negative BSS for noisy data,” *IEEE Trans. on Signal Processing*, vol. 61, no. 22, pp. 5620–5632, 2013.
- [4] C. Févotte and N. Dobigeon, “Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization,” *IEEE Trans. Image Processing*, vol. 24, no. 12, pp. 4810–4819, Dec. 2015.
- [5] A. Ben Hamza and D. J. Brady, “Reconstruction of reflectance spectra using robust nonnegative matrix factorization,” *IEEE Trans. on Signal Processing*, vol. 54, no. 9, pp. 3637–3642, 2006.
- [6] D. Kong, C. Ding, and H. Huang, “Robust non-negative matrix factorization using l_{21} -norm,” in *Proc. of CIKM*, 2011, pp. 673–682.
- [7] L. Du, X. Li, and Y.-D. Shen, “Robust non-negative matrix factorization via half-quadratic minimization,” in *Proc. of ICDM*, 2012, pp. 201–210, Springer.
- [8] 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.
- [9] E. Chouzenoux, S. Moussaoui, M. Legendre, and J. Idier, “Algorithme primal-dual de points intérieurs pour l’estimation pénalisée des cartes d’abondances en imagerie hyperspectrale,” *Traitement du Signal*, April 2013.
- [10] W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi, “A signal processing perspective on hyperspectral unmixing,” *IEEE Signal Proc. Mag.*, pp. 67–81, Jan. 2014.
- [11] M. Viana, A. Pandolfi, M. C. Minguillo, X. Querol, A. Alastuey, E. Monfort, and I. Celades, “Inter-comparison of receptor models for PM source apportionment: Case study in an industrial area,” *Atmospheric Environment*, vol. 42, pp. 3820–3832, 2008.
- [12] 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.
- [13] A. Limem, M. Puigt, G. Delmaire, G. Roussel, and D. Courcot, “Bound constrained weighted NMF for industrial source apportionment,” in *Proc. of MLSP*, 2014.
- [14] R. Chreiky, G. Delmaire, M. Puigt, G. Roussel, and A. Abche, “Split gradient method for informed non-negative matrix factorization,” in *Proc. of LVA-ICA*, 2015, pp. 376–383.
- [15] R. Chreiky, G. Delmaire, C. Dorffer, M. Puigt, G. Roussel, and A. Abche, “Robust informed split gradient NMF using Alpha Beta-divergence for source apportionment,” in *Proc. of MLSP*, 2016.
- [16] H. Lantéri, C. Theys, C. Richard, and C. Févotte, “Split gradient method for nonnegative matrix factorization,” in *Proc. of EUSIPCO*, 2010.
- [17] P. J. Huber, “Robust estimation of a location parameter,” *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, 03 1964.
- [18] 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.
- [19] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, Wiley, second edition, 2009.
- [20] N.-D. Ho, *Non negative matrix factorization algorithms and applications*, Phd thesis, Université Catholique de Louvain, 2008.
- [21] C. Févotte and J. Idier, “Algorithms for nonnegative matrix factorization with the beta-divergence,” *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011.