



HAL
open science

Estimation of the regularization parameter of an on-line NMF with minimum volume constraint

Ludivine Nus, Sebastian Miron, David Brie

► **To cite this version:**

Ludivine Nus, Sebastian Miron, David Brie. Estimation of the regularization parameter of an on-line NMF with minimum volume constraint. 10th IEEE Sensor Array and Multichannel Signal Processing Workshop, SAM 2018, Jul 2018, Sheffield, United Kingdom. hal-01938028

HAL Id: hal-01938028

<https://hal.science/hal-01938028>

Submitted on 28 Nov 2018

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.

Estimation of the regularization parameter of an on-line NMF with minimum volume constraint

Ludivine Nus, Sebastian Miron, David Brie
CRAN, Université de Lorraine, CNRS Vandœuvre-lès-Nancy, France
firstname.lastname@univ-lorraine.fr

Abstract—In this paper, the estimation of the regularization parameter of the on-line Non-negative Matrix Factorization (NMF) with minimum volume constraint on sources is addressed. Adding a volume constraint in the model is important to ensure uniqueness of the solution and good data representation. However, the effectiveness of this approach is hampered by the optimal determination of the strength of minimum volume term. To solve this problem, we formulate it as a bi-objective optimization problem and three Minimum Distance Criterion (MDC) strategies are proposed and evaluated. The three strategies yield similar results but one of them in particular yields an interesting tradeoff between accuracy and computation time.

Index Terms—on-line Non-negative Matrix Factorization, minimum volume simplex, bi-objective optimization, Minimum Distance Criterion

I. INTRODUCTION

Non-Negative Matrix Factorization (NMF) is a problem arising in applications such as hyperspectral imaging (which is the main concern of this work) where the observed non-negative data are non-negative linear mixtures of r non-negative *sources* (endmembers). For a non-negative matrix $\mathbf{X} \in \mathbb{R}_+^{n \times m}$, the NMF consists in finding two non-negative matrices $\mathbf{S} \in \mathbb{R}_+^{n \times r}$ and $\mathbf{A} \in \mathbb{R}_+^{r \times m}$ with $r \leq \min(n, m)$ such that [10]: $\mathbf{X} \approx \mathbf{S}\mathbf{A}$. In hyperspectral imagery, the m columns of \mathbf{X} represent the pixel spectra sampled at n wavelengths. \mathbf{S} is a matrix containing the r normalized endmembers and \mathbf{A} is a matrix containing on its column the mixing coefficients (the abundances) for the recorded samples. In general, the NMF suffers from non-uniqueness of the solution [14]. To be more precise, the uniqueness of the NMF relies on the sparsity of the underlying latent variables. In particular, if either \mathbf{S} and/or \mathbf{A} have only no zero entries, the NMF is not unique and additional constraints such as sparsity [8] or minimum volume constraint [12] need to be added to have an unique factorization. Over the last decade, a lot of attention has been paid to the development of on-line NMF, sequentially updating the parameters \mathbf{S} and \mathbf{A} as the size of the data matrix \mathbf{X} increases [2], [11], [18], [19]. In particular, we recently proposed [15] the on-line MVS-NMF (*Minimum Volume Simplex-NMF*) which includes a minimum volume constraint. This algorithm is well-adapted to real-time unmixing of hyperspectral images, which is required in industrial applications for controlling and sorting input materials. In MVS-NMF, the strength of the minimum volume constraint is

controlled by an hyperparameter; the main goal of this work is to provide a method to determine it automatically. Some classical approaches for regularization parameter estimation are the generalized cross validation (GCV) method [4] and the L-curve [6], [7]. As the NMF problem is bilinear, GCV is not applicable. The L-curve is a log-log plot of the data fitting versus regularization cost for varying regularization parameters; the optimal value of the hyperparameter corresponds to the maximum curvature of the L-curve. Its link to multi-objective optimization was first mentioned in [9]. However, the L-curve approach has some undesirable properties discussed in [17] and [5]: it is non-convex and the maximum curvature is not guaranteed to be unique. To overcome the non-uniqueness of the maximum curvature, [1] proposed the Minimum Distance Criterion (MDC) applied to the L-curve (bi-objective case) and L-hypersurface (multi-objective case). This leads to a simple fixed-point iterative algorithm for computing regularization parameters in both bi-objective and multi-objective cases but it requires to limit the range of the hyperparameter to ensure that the corresponding part of the L-curve remains convex. An approach using the L-curve for estimating the regularization parameters of a Tikhonov regularized NMF was proposed in [13]. Similarly to [1], this leads to a simple fixed point algorithm which also suffers from the same problems: hyperparameter range limitation, non-uniqueness of the maximal curvature. To overcome these two main drawbacks, it was proposed in [16] to estimate the hyperparameter by determining the MDC on response curve (bi-objective case) or response surface (multi-objective case) defined as the linear plot of the data fitting versus regularization cost. This paper aims at investigating the interest of such an approach for estimating the minimum volume hyperparameter of the on-line MVS-NMF. Unlike [16] which applies the MDC to a deconvolution case with a Tikhonov regularization, the problem considered in this paper is the on-line source separation with minimum volume constraint, and presents different implementation problems, as explained in the following sections. The remainder of the paper is organized as follows: section 2 is briefly presenting the on-line MVS-NMF algorithm. Section 3 aims at analyzing the response curves for the on-line MVS-NMF. Section 4 presents several MDC-based strategies for estimating the minimum volume constraint hyperparameter. Conclusions and perspectives are drawn in section 4.

II. ON-LINE MVS-NMF

The on-line MVS-NMF allows real-time blind unmixing of hyperspectral images acquired by a hyperspectral pushbroom imager installed on a production line. The main feature of pushbroom imaging system is that the hyperspectral data cube is acquired slice by slice, sequentially in time (which in fact is representing the so-called *along track* spatial dimension OY). Each slice is a $n_x \times n_\lambda$ image, where n_x represents the across track spatial dimension (one line of the scene) and n_λ , the spectral dimension (wavelengths). The hyperspectral image is created by moving the objects on the conveyor belt. To cast the data in matrix form, the hyperspectral image is unfolded as shown in figure 1. $\tilde{\mathbf{X}}^{(1)} = \mathbf{X}^{(1)}$ is the first slice of the hyperspectral image and $\tilde{\mathbf{X}}^{(k)}$, the k^{th} slice. The entire dataset at time instant $(k+1)$, $\mathbf{X}^{(k+1)}$, can be represented as the concatenation of the first k samples with the new incoming sample *i.e.*, $\mathbf{X}^{(k+1)} = [\mathbf{X}^{(k)} \tilde{\mathbf{X}}^{(k+1)}]$. Similarly, we define $\mathbf{A}^{(k+1)} = [\mathbf{A}^{(k)} \tilde{\mathbf{A}}^{(k+1)}]$.

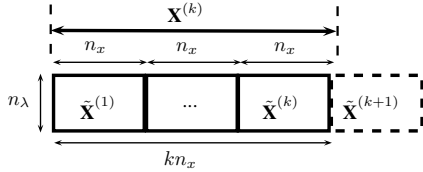


Fig. 1. Unfolded pushbroom hyperspectral image

To estimate $(\tilde{\mathbf{S}}^{(k+1)}, \tilde{\mathbf{A}}^{(k+1)})$, we consider the following cost function:

$$\begin{aligned} \mathcal{J}^{(k+1)}(\mathbf{S}^{(k+1)}, \mathbf{A}^{(k+1)}) &= \alpha \sum_{\ell=1}^k \left\| \tilde{\mathbf{X}}^{(\ell)} - \tilde{\mathbf{S}}^{(\ell)} \tilde{\mathbf{A}}^{(\ell)} \right\|_F^2 \\ &+ (1 - \alpha) \left\| \tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{S}}^{(k+1)} \tilde{\mathbf{A}}^{(k+1)} \right\|_F^2 \\ &+ \mu \ln \det \left(\tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{S}}^{(k+1)} \right), \end{aligned} \quad (1)$$

where the weighting coefficient α ($0 \leq \alpha \leq 1$) adds some tracking capacity to the algorithm. Following [2], we assume that the endmembers vary only slightly between consecutive samples *i.e.* $\tilde{\mathbf{S}}^{(k+1)} \approx \tilde{\mathbf{S}}^{(k)}$. The term $\ln \det \left(\tilde{\mathbf{S}}^{(k+1)T} \tilde{\mathbf{S}}^{(k+1)} \right)$ refers to the minimum volume constraint on the sources and $\mu > 0$ is the regularization parameter controlling the strength of the minimum volume constraint. The on-line MVS-NMF algorithm alternatively updates the matrices $\tilde{\mathbf{S}}^{(k+1)}$ and $\tilde{\mathbf{A}}^{(k+1)}$ by minimizing the cost function (1) using a gradient descent technique, where the step size is chosen in order to obtain multiplicative update rules (similarly to the classical NMF [10]). Let us mention that we introduce the logarithm in the minimum volume constraint because it results in simplifications of the update rule derivation. The different steps of MVS-NMF are given in Algorithm 1. The interested reader is referred to [15] for the complete derivation of the algorithm and the analysis of the influence of α and $Niter$ on its

convergence. While it was observed that the parameter μ strongly influences the quality of the results, its value was determined experimentally using a trial and error approach. The objective of this paper is to propose an unsupervised approach for estimating μ .

Algorithm 1 On-line MVS-NMF

Input : $\tilde{\mathbf{X}}^{(1)}$, r , α , μ , $Niter$ = number of iterations

Initialization:

$\mathbf{N} = \text{zeros}(n_\lambda, r)$, $\mathbf{M} = \text{zeros}(r, r)$,

$\tilde{\mathbf{A}} = \text{rand}(r, n_x)$, $\tilde{\mathbf{S}} = \text{rand}(n_\lambda, r)$

$\mathbf{A} = []$

Output : \mathbf{A} and $\tilde{\mathbf{S}}$

while New sample available **do**

$iter = 1$

while $iter < Niter$ **do**

$$\tilde{\mathbf{A}}_{rj} \leftarrow \tilde{\mathbf{A}}_{rj} \frac{(\tilde{\mathbf{S}}^T \tilde{\mathbf{X}})_{rj}}{(\tilde{\mathbf{S}}^T \tilde{\mathbf{S}} \tilde{\mathbf{A}})_{rj}}$$

$$\mathbf{N} \leftarrow \alpha \mathbf{N} + (1 - \alpha) \tilde{\mathbf{X}} \tilde{\mathbf{A}}^T$$

$$\mathbf{M} \leftarrow \alpha \mathbf{M} + (1 - \alpha) \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T$$

$$\tilde{\mathbf{S}}_{ir} \leftarrow \tilde{\mathbf{S}}_{ir} \frac{(\mathbf{N} \tilde{\mathbf{S}}^T \tilde{\mathbf{S}})_{ir}}{(\tilde{\mathbf{S}} \mathbf{M} \tilde{\mathbf{S}}^T \tilde{\mathbf{S}})_{ir} + \mu \tilde{\mathbf{S}}_{ir}}$$

$iter \leftarrow iter + 1$

end while

$\mathbf{A} \leftarrow [\mathbf{A}, \tilde{\mathbf{A}}]$

end while

III. RESPONSE CURVES FOR THE ON-LINE MVS-NMF

A. Response curve

Consider a hyperspectral image composed of K slices $\tilde{\mathbf{X}}^{(k)}$, $k = 1, \dots, K$ which is used to learn the value of μ . We assume that the values of α and $Niter$ are fixed. Let $\mathbf{S}_\mu^{(K)}$ and $\mathbf{A}_\mu^{(K)}$ denote the estimated endmembers and abundances for a given value of μ . Following [16] the response curve (bi-objective case) is the linear plot of the data fitting versus minimum volume constraint cost for $\mu \in [0, +\infty)$. The two objectives are respectively defined as:

$$\mathcal{J}_1(\mu) = \mathcal{J}_1(\mathbf{S}_\mu^{(K)}, \mathbf{A}_\mu^{(K)}) = \frac{1}{K} \sum_{k=1}^K \left\| \tilde{\mathbf{X}}^{(k)} - \tilde{\mathbf{S}}_\mu^{(k)} \tilde{\mathbf{A}}_\mu^{(k)} \right\|_F^2, \quad (2)$$

$$\mathcal{J}_2(\mu) = \mathcal{J}_2(\mathbf{S}_\mu^{(K)}) = \frac{1}{K} \sum_{k=1}^K \det \left(\tilde{\mathbf{S}}_\mu^{(k)T} \tilde{\mathbf{S}}_\mu^{(k)} \right). \quad (3)$$

Note that the minimum volume constraint cost (3) used here does not include the \ln function as in criterion (1). This is to avoid the non-convexity of the response curve induced by a log scale along the \mathcal{J}_2 -axis (we are currently working at proving the convexity of the response curve). It is well known that the NMF problems (including MVS-NMF) are bilinear and thus non-convex. This has important consequences on the convergence of the alternate algorithms such as the one proposed in Algorithm 1: only the convergence toward a local minimizer

can be guaranteed and the obtained solution depends on the initial values of the endmembers and abundances. To take into account this point, we defined a response curve for some given initial values ($\tilde{\mathbf{S}}^{(0)}, \tilde{\mathbf{A}}^{(0)}$). In the sequel, we consider the two following questions: (i) what is the shape of a response curve? (ii) what is the variability of the response curves for different initial values? These two points are addressed numerically, using a simulated hyperspectral image of size $119 \times 36 \times 36$ consisting in the non-negative mixture of three endmembers (with 119 wavelengths), not varying over time. None of the three endmembers has any zero value, which results in a non-unique NMF problem. In other words, without the minimum volume constraint the model is not identifiable. The three abundance maps are matrices of size 36×36 , randomly drawn from a continuous uniform distribution on the interval $[0, 1]$. By doing so, we ensure (with high probability) that the pure pixel condition is approximately fulfilled; thus, in the noise-free case, the minimum volume enclosing simplex yields the correct endmembers. For the noisy case, a noise drawn from a uniform distribution with positive support is added to reach a SNR of 26dB. In all simulations, the values of parameters for the on-line MVS-NMF were set to $r = 3$, $\alpha = 0.99$ and the number of iterations was fixed to 500 allowing to reach a good tradeoff between convergence rate and computational burden.

B. Shape of response curve

The response curves are evaluated for 44 values of $\mu \in [0.0001, 0.0028]$. Figure 2(a) corresponds to the noise-free situation while figure 2(b) is for the noisy one. In addition, the simplexes corresponding to the estimated endmembers for different values of μ are shown in figures 2(c) and 2(d).

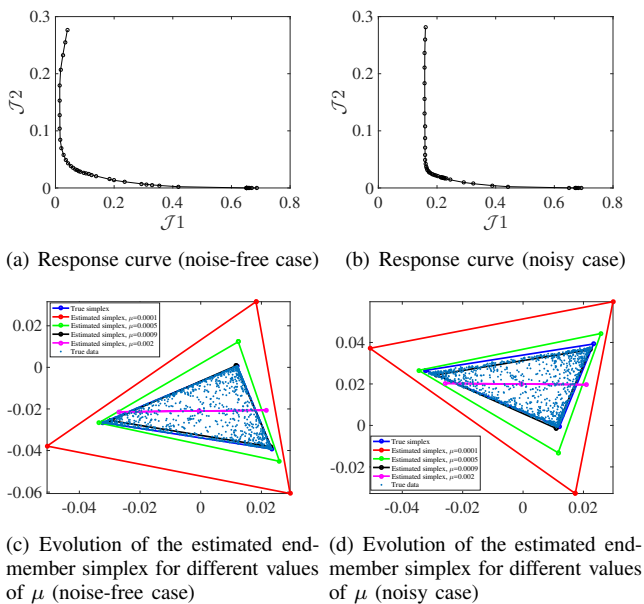


Fig. 2. Response curve and estimated endmember simplex

For both cases, the response curve includes three main parts: the first region corresponds to low values of μ . The

minimum volume simplex enclosing the data \mathcal{D} is included in the endmember simplex \mathcal{S} i.e. $\mathcal{D} \subset \mathcal{S}$. This situation is observed on figures 2(c) and 2(d) for $\mu = 0.0001$ and 0.0005 . The second region corresponds to large values of μ , and the endmember simplex volume tends to zero, indicating that the endmembers tend to be collinear. In that case, $\mathcal{D} \not\subset \mathcal{S}$. This situation is observed on figures 2(c) and 2(d) for $\mu = 0.002$. The third region is the corner of the response curve which corresponds to values of μ for which $\mathcal{D} \approx \mathcal{S}$. The optimal value of μ is in this part of the curve. This situation is observed on figures 2(c) and 2(d) for $\mu = 0.0009$. Comparing the noise-free and the noisy scenarios, we observe that the noise is right shifting the response curve by a value which increases with the noise level. We can observe that the response curve seems to be convex, on the considered domain; we are currently working at proving it formally. Also, it appears that for large value of μ , the data fitting tends to a maximum value which corresponds to the case of three collinear endmembers. In this limit case, the endmember is located in the gravity center of the data point cloud. Figure 3 illustrates the influence of μ on the estimated endmembers.

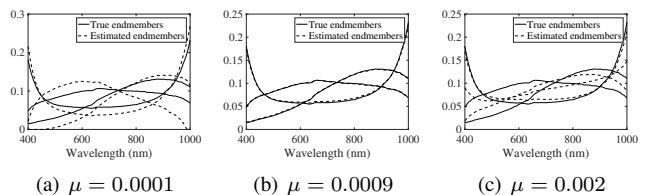


Fig. 3. Estimated endmember for different values of μ

C. Influence of the initial conditions

As mentioned before, the NMF is a non-convex problem and therefore it is very sensitive to initialization. This means that the response curves will depend on the initial values ($\tilde{\mathbf{S}}^{(0)}, \tilde{\mathbf{A}}^{(0)}$). To assess the influence of these initial values, we evaluated the response curves obtained for 31 different initial values drawn randomly from uniform distribution on the interval $[0, 1]$. It appears (figure 4) that most of the response curves are very similar, but some are deviating from the “mean response curve”. This is quite an important aspect since it opens the possibility to use a single realization of the response curve to estimate the optimal value of μ . We will address this issue in the next section.

IV. HYPERPARAMETER ESTIMATION

In order to estimate the optimal value of μ , we propose to use the MDC [16]. Three strategies are considered: the first one consists in applying the MDC to the Pareto Front estimated from the set of response curves, which corresponds in the best achievable performance. The second one applies the MDC to a single realization of the response curve. Finally, to assess how the variability induced by the different initialization is affecting the result we apply the MDC to the average response curve.

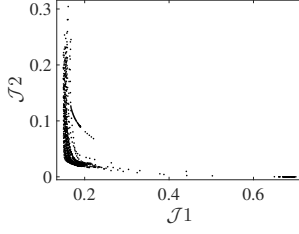


Fig. 4. Evolution of the response curve for different values of μ and different initializations

A. Pareto Front

The definition of the Pareto front relies on the notion of *domination* defined in [3]. A solution is said to dominate another solution, if the first solution is not worse than the second one in all objectives, and the first solution is strictly better than the second one in at least one objective. The solution is said to be non-dominated or *Pareto optimal* for a bi-objective problem if all other solution in the feasible set has a higher value in at least one of the objectives \mathcal{J}_i , with $i \in \{1, 2\}$. The image of all the non-dominated solutions is called *Pareto front*. The shape of the Pareto front reveals the nature of the trade-off between the different objective functions and represents the set of the best achievable trade-offs between conflicting objectives. It is worth noting that, even if each response curve is convex, the Pareto front cannot be guaranteed to be convex.

B. Minimum Distance Criterion

The Minimum Distance Criterion [16] consists in finding the point of the response curve or Pareto front which is at minimum distance to the *ideal point*. The ideal point I corresponds to the point whose coordinates are the minima of the two objective functions (see figure 5); it can be considered as a reference point and the *optimal point* M of the response surface will be the one having the minimum distance to this ideal point. Let $I = (I_1, I_2)$ denotes the coordinates of the ideal point. The function $D(\mu)$ is the squared distance from the ideal point I to the point $M(\mu) = (\mathcal{J}_1(\mu), \mathcal{J}_2(\mu))$ on the response surface:

$$D(\mu) = \sum_{i=1}^2 (\mathcal{J}_i(\mu) - I_i)^2 \quad (4)$$

The solution of the MDC is thus given by:

$$\mu^* = \arg \min_{\mu} D(\mu) \quad (5)$$

C. Results

The Pareto Front and the average response curve are evaluated from the 31 response curves. The MDC is applied to the Pareto front, to the average response curve and to a single realization of the response curve. The corresponding ideal points are I_p , I_a and I_s , respectively, while the optimal points determined by the MDC are M_p , M_a and M_s (figure 6). The red line segment represents the minimum distance

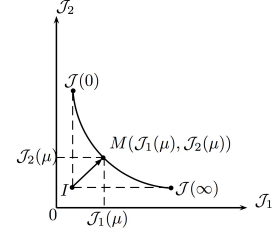


Fig. 5. Representation of the response curve. The ideal point is denoted by I and the optimal point is denoted by M

between points I and M . The M_p , M_a and M_s points on figure 6 may seem visually not optimal with respect to the MDC. This is because, for illustration purpose, we applied a much stronger zoom on the horizontal axis compared to the vertical axis. The estimated values of μ are: 0.001 (Pareto Front), 0.0011 (average response curve) and $\mu = 0.00095$ (single realization response curve). While these differences may seem important regarding the results of figure 3, all these values of μ yield endmembers which are very close to the true ones, when we are closed to the optimal value. Thus, the estimated μ is not very sensitive to the chosen strategy. Even for the very different response curve (see figure 4) the single realization response curve strategy yields an estimated $\mu = 0.0014$. This is a very important result, since it states that the hyperparameter can be estimated by the single realization strategy. This is the one with the lowest computational burden and opens the possibility to perform on-line estimation of μ .

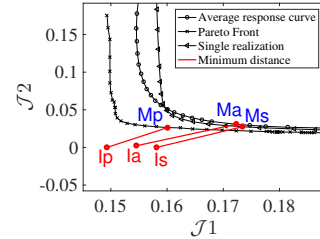


Fig. 6. Pareto curve and optimal point

A similar procedure could be used to estimate α . However, this would require the estimation of the Pareto Front each time a new sample arrives, which represents an unacceptable computational cost for an on-line algorithm.

V. CONCLUSION

This paper addresses the problem of regularization parameter estimation for the on-line MVS-NMF. Three different MDC-based strategies were proposed, all yielding similar results. Among the three strategies, the single realization response curve approach is the most attractive since it presents the lowest computational cost. The results presented in this paper open the possibility to develop an on-line learning of the regularization parameter, which is an aspect of utmost importance in industrial applications.

REFERENCES

- [1] M. Belge, M. E. Kilmer, and E. L. Miller. Efficient determination of multiple regularization parameters in a generalized l-curve framework. *Inverse Problems*, 18(4):1161, 2002.
- [2] S. S. Bucak and B. Günsel. Incremental subspace learning via non-negative matrix factorization. *Pattern recognition*, 42(5):788–797, 2009.
- [3] K. Deb. *Multiobjective optimization using evolutionary algorithms.*, volume 16. John Wiley & Sons, 2001.
- [4] G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.
- [5] M. Hanke. Limitations of the l-curve method in ill-posed problems. *BIT Numerical Mathematics*, 36(2):287–301, 1996.
- [6] P. C. Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM review*, 34(4):561–580, 1992.
- [7] P. C. Hansen. The l-curve and its use in the numerical treatment of inverse problems. 1999.
- [8] P. O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, 5(Nov):1457–1469, 2004.
- [9] L. Kaufman and A. Neumaier. Regularization of ill-posed problems by envelope guided conjugate gradients. *Journal of Computational and Graphical Statistics*, 6(4):451–463, 1997.
- [10] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.
- [11] A. Lefèvre, F. Bach, and C. Févotte. On-line algorithms for non-negative matrix factorization with the Itakura-Saito divergence. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 313–316, 2011.
- [12] L. Miao and H. Qi. Endmember extraction from highly mixed data using minimum volume constrained non-negative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3):765–777, 2007.
- [13] A. Mirzal. Nonparametric Tikhonov regularized NMF and its application in cancer clustering. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 11(6):1208–1217, 2014.
- [14] S. Moussaoui, D. Brie, and J. Idier. Non-negative source separation: range of admissible solutions and conditions for the uniqueness of the solution. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 5, pages v–289. IEEE, 2005.
- [15] L. Nus, S. Miron, and D. Brie. On-line blind unmixing for hyperspectral pushbroom imaging systems. *IEEE Statistical Signal Processing Workshop*, 2018.
- [16] Y. Song, D. Brie, E.-H. Djermoune, and S. Henrot. Regularization parameter estimation for non-negative hyperspectral image deconvolution. *IEEE Transactions on Image Processing*, 25(11):5316–5330, 2016.
- [17] C. R. Vogel. Non-convergence of the l-curve regularization parameter selection method. *Inverse problems*, 12(4):535, 1996.
- [18] Y. Wu, B. Shen, and H. Ling. Visual tracking via on-line non-negative matrix factorization. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(3):374–383, 2014.
- [19] G. Zhou, Z. Yang, S. Xie, and J.-M. Yang. On-line blind source separation using incremental non-negative matrix factorization with volume constraint. *IEEE transactions on neural networks*, 22(4):550–560, 2011.