New Concepts in Imaging: Optical and Statistical Models
D. Mary, C. Theys and C. Aime (eds)
EAS Publications Series, **59** *(2013) 357–379*

SGM TO SOLVE NMF – APPLICATION TO HYPERSPECTRAL DATA

C. Theys[1], H. Lantéri[1] and C. Richard[1]

Abstract. This article deals with the problem of minimization of a general cost function under non-negativity and flux conservation constraints. The proposed algorithm is founded on the Split Gradient Method (SGM) adapted here to solve the Non Negative Matrix Factorization (NMF). We show that SGM can be easily regularized, allowing to introduce some physical constraints. Finally, to validate the algorithm, we propose an example of application to hyperspectral data unmixing.

1 Introduction

In the field of image reconstruction or deconvolution, the minimization of a cost function between noisy measurements and a linear model is usually performed, subject to positivity and flux constraints. The well known, in astrophysical area, are the Iterative Space Reconstruction Algorithm (ISRA) (Daube-Witherspoon 1986), and the Expectation Minimization (EM) (Dempster *et al.* 1977) or Richardson Lucy (RL) (Lucy 1974; Richardson 1972) algorithm. In the last ten years, a general algorithmic method, called Split Gradient Method (SGM) (Lantéri *et al.* 2001, 2002), has been developed to derive multiplicative algorithms for minimizing any convex criterion under positivity constraints. It leads to ISRA and EM-RL algorithm as particular cases. SGM has recently been extended to take into account a flux conservation constraint (Lantéri *et al.* 2009).

During the last few years, many papers have been published in the field of Nonnegative Matrix Factorization (NMF) with multiplicative algorithms (Lee & Seung 2001; Cichoki *et al.* 2006; Févotte *et al.* 2009). This problem is closely related to the blind deconvolution one (Desidera *et al.* 2006; Lantéri *et al.* 1994) and consists in estimating $\mathbf{W}$ and $\mathbf{H}$, nonnegative, such that $\mathbf{V} \approx \mathbf{WH}$. The aim

[1] Laboratoire Lagrange, Université de Nice Sophia-Antipolis, Observatoire de la Côte d'Azur, CNRS, Nice, France

DOI: 10.1051/eas/1359016

of this paper is to propose a unified framework based on SGM, an interior-point algorithm, to derive algorithms for NMF, in a multiplicative form or not.

To illustrate the general interest of SGM for NMF, we also show how to regularize the problem by introducing smoothness or sparsity constraints on the columns of **W** and **H** respectively, Lantéri *et al.* (2011), Lantéri *et al.* (2011). The choice of these different regularization terms are motivated by the application on hyperspectral imagery, Theys *et al.* (2009). The paper is organized as follows. In Section 2, we describe the problem at hand and notations for non-negative matrix factorization. In Section 3, we describe the Split Gradient Method (SGM). In Section 4, we show how to add a sum-to-one constraint in the SGM algorithm. In Section 5, we briefly discuss the choice of the step size. Section 6 introduces the physical context and some simulation results are given in Section 7. The regularized SGM is developed in section 8 with a smoothness constraint on the columns of **W** and then a sparsity constraint on the columns of **H**, with typical numerical examples in Section 9. Section 10 concludes the paper.

2 Nonnegative matrix factorization

We consider here the problem of nonnegative matrix factorization (NMF), which is now a popular dimension reduction technique, employed for non-subtractive, part-based representation of nonnegative data. Given a nonnegative data matrix **V** of dimension $F \times N$, the NMF consists of seeking a factorization of the form

$$\mathbf{V} \approx \mathbf{W}\mathbf{H} \tag{2.1}$$

where **W** and **H** are nonnegative matrices of dimensions $F \times K$ and $K \times N$, respectively. Dimension K is usually chosen such that $FK + KN \ll FN$, that is, much more equations than unknowns. For example with $F = N = 3$ and $K = 1$:

$$\begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \mathbf{V}_{13} \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \mathbf{V}_{23} \\ \mathbf{V}_{31} & \mathbf{V}_{32} & \mathbf{V}_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{11} \\ \mathbf{W}_{21} \\ \mathbf{W}_{31} \end{bmatrix} \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \mathbf{H}_{13} \end{bmatrix}. \tag{2.2}$$

This problem is encountered at each time we want to find both the basis and the coefficients of projection. The factorization (2.1) is usually sought through the minimization problem

$$\min_{\mathbf{W},\mathbf{H}} \mathcal{D}(\mathbf{V}, \mathbf{W}\mathbf{H}) \quad \text{s.t.} \quad [\mathbf{W}]_{ij} \geq 0, [\mathbf{H}]_{ij} \geq 0 \tag{2.3}$$

with $[\mathbf{V}]_{ij}$ and $[\mathbf{W}\mathbf{H}]_{ij}$ the (i, j)-th entries of **V** and **WH**, respectively. In the above expression, $\mathcal{D}(\mathbf{V}, \mathbf{W}\mathbf{H})$ is a cost function defined by

$$\mathcal{D}(\mathbf{V}, \mathbf{W}\mathbf{H}) = \sum_{ij} d([\mathbf{V}]_{ij}, [\mathbf{W}\mathbf{H}]_{ij}) = \sum_{ij} d_{ij}. \tag{2.4}$$

In the general case, $d(u, v)$ is a positive convex function that is equal to zero if $u = v$.

2.1 Unicity

The solution of (2.3) is, obviously, not unique. One way to overcome this problem is to normalize the columns of $\mathbf{W}$ or $\mathbf{H}$. We propose, here, to normalize to *one* the columns of $\mathbf{W}$. As a direct consequence of (2.1), this implies a constraint-sum condition on the columns of $\mathbf{H}$.

The minimization problem (2.3) becomes:

$$\min_{\mathbf{W},\mathbf{H}} \mathcal{D}(\mathbf{V},\mathbf{WH}) \quad \text{s.t.} \quad [\mathbf{W}]_{ij} \geq 0, \quad [\mathbf{H}]_{ij} \geq 0,$$
$$\sum_i [\mathbf{W}]_{ij} = 1, \quad \sum_i [\mathbf{H}]_{ij} = \sum_i [\mathbf{V}]_{ij}. \qquad (2.5)$$

This constant-sum constraint is motivated by applications such as, for example, hyperspectral data unmixing. In this case, $\mathbf{W}$ is the matrix of basis spectra that are supposed to be normalized to *one*. Another source of indetermination is that the solutions are given up to a permutation on rows and columns of $\mathbf{W}$ and $\mathbf{H}$. The problem established by (2.3) is a convex optimization problem under inequality constraint and problem (2.5) is a convex optimization problem under both equality and inequality constraints. We propose to consider first the problem (2.3), the inequality constraint is treated by solving the Karush-Kuhn-Tucker conditions. Second, we consider the problem (2.5) and the equality constraint is added by introducing normalized variables. Once the conditions satisfying the constraints have been established, an iterative algorithm should be applied alternatively on $\mathbf{W}$ and $\mathbf{H}$. The proposed iterative algorithm founded on the Split Gradient Method (SGM), a scaled gradient descent algorithm. The way to obtain it is detailed in the following section.

3 Minimization under non-negativity constraints: The SGM

The SGM was initially formulated and developed to solve the minimization of a positive convex function under non-negativity constraint of the solution, problem (2.3).

3.1 The Lagrangian function

The non-negativity constraint is expressed by the Lagrangian function associated to (2.3), given by:

$$\mathcal{L}(\mathbf{V},\mathbf{WH};\boldsymbol{\Lambda},\boldsymbol{\Omega}) = \mathcal{D}(\mathbf{V},\mathbf{WH}) - \langle \boldsymbol{\Lambda},\mathbf{W}\rangle - \langle \boldsymbol{\Omega},\mathbf{H}\rangle \qquad (3.1)$$

where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ are the matrices of positive Lagrange multipliers, and $\langle \cdot,\cdot \rangle$ is the inner product defined by:

$$\langle \mathbf{U},\mathbf{V}\rangle = \sum_{ij} [\mathbf{U}]_{ij}[\mathbf{V}]_{ij}. \qquad (3.2)$$

The Lagrange multipliers method allows to find an optimum of a function under some constraints.

3.2 Minimization with respect to $\mathbf{W}$

Minimization of (3.1) with respect to $\mathbf{W}$ leads to the following Karush-Kuhn-Tucker conditions for all i, j at the solution $\mathbf{W}^*$, $\mathbf{\Lambda}^*$:

$$[\nabla_W \mathcal{L}(\mathbf{V}, \mathbf{W}^*\mathbf{H}; \mathbf{\Lambda}^*, \mathbf{\Omega})]_{ij} = 0, \tag{3.3}$$
$$[\mathbf{\Lambda}^*]_{ij} \geq 0, \tag{3.4}$$
$$[\mathbf{W}^*]_{ij} \geq 0, \tag{3.5}$$
$$\langle \Lambda^*, \mathbf{W}^* \rangle = 0 \Leftrightarrow [\Lambda^*]_{ij}[\mathbf{W}^*]_{ij} = 0. \tag{3.6}$$

Condition (3.3) immediately leads to

$$[\Lambda^*]_{ij} = [\nabla_W \mathcal{D}(\mathbf{V}, \mathbf{W}^*\mathbf{H})]_{ij}. \tag{3.7}$$

Condition (3.6) then becomes

$$\begin{aligned}&[\mathbf{W}^*]_{ij}[\nabla_W \mathcal{D}(\mathbf{V}, \mathbf{W}^*\mathbf{H})]_{ij} = 0\\ &\qquad \Leftrightarrow [\mathbf{W}^*]_{ij}[-\nabla_W \mathcal{D}(\mathbf{V}, \mathbf{W}^*\mathbf{H})]_{ij} = 0\end{aligned} \tag{3.8}$$

where the extra minus sign in the last expression is just used to make apparent the negative gradient descent direction of $\mathcal{D}(\mathbf{V}, \mathbf{WH})$.

The expression (3.6) gives the condition that must be satisfied for any optimization problem under non-negativity constraint. At the solution, the inner product between the gradient of the cost function and the variables must be equal to zero. The interpretation is the following: either our solution is the one that minimizes the cost function and the minimizer is positive, either the minimizer of the cost function is negative or zero and the constrained solution is zero.

This condition is non linear w.r.t. the unknowns, an analytical solution does not exist.

3.2.1 Gradient descent method

Since the gradient of the functional has an analytical form, a natural choice for the iterative algorithm is a gradient descent method.

If we consider first the minimization problem without non-negativity constraint:

$$\min_{\mathbf{W}, \mathbf{H}} \mathcal{D}(\mathbf{V}, \mathbf{WH}), \tag{3.9}$$

we use the negative gradient as a descent direction and we write:

$$[\mathbf{W}^{k+1}]_{ij} = [\mathbf{W}^k]_{ij} + \alpha_{ij}^k[-\nabla_W \mathcal{D}(\mathbf{V}, \mathbf{W}^*\mathbf{H})]_{ij} \tag{3.10}$$

with α_{ij}^k a positive step size that allows to control convergence of the algorithm.

If now, we consider the minimization problem with non-negativity constraint, Equation (2.3), the descent direction becomes $[\mathbf{W}^*]_{ij}[-\nabla_W \mathcal{D}(\mathbf{V}, \mathbf{W}^*\mathbf{H})]_{ij}$, Equation (3.8) and the descent algorithm is:

$$[\mathbf{W}^{k+1}]_{ij} = [\mathbf{W}^k]_{ij} + \alpha_{ij}^k[\mathbf{W}^*]_{ij}[-\nabla_W \mathcal{D}(\mathbf{V}, \mathbf{W}^*\mathbf{H})]_{ij}. \tag{3.11}$$

More generally:

$$\mathbf{M}\cdot\mathbf{W}\cdot[-\nabla_W\mathcal{D}(\mathbf{V},\mathbf{W}^*\mathbf{H})] \tag{3.12}$$

is a scaled gradient descent direction of $\mathcal{D}$ if $\mathbf{M}$ is a matrix with positive entries, where $\cdot$ denotes the Hadamard product. A particular choice for $\mathbf{M}$ with an adequate particular decomposition of $[-\nabla_W\mathcal{D}(\mathbf{V},\mathbf{W}^*\mathbf{H})]$ leads to the SGM algorithm.

3.2.2 Split Gradient Method (SGM)

The SGM algorithm is a descent algorithm whose direction is constructed in such a way that, for a step size equal to one, we obtain a multiplicative algorithm. To obtain it, an additional point is that $[-\nabla_W\mathcal{D}]_{ij}$ can always be decomposed as $[\mathbf{P}]_{ij}-[\mathbf{Q}]_{ij}$, where $[\mathbf{P}]_{ij}$ and $[\mathbf{Q}]_{ij}$ are positive entries, let us note that this decomposition is obviously not unique. If we take for $\mathbf{M}$, Equation (3.12):

$$[\mathbf{M}]_{ij}=\frac{1}{[\mathbf{Q}]_{ij}} \tag{3.13}$$

we obtain the following gradient-descent algorithm:

$$[\mathbf{W}^{k+1}]_{ij}=[\mathbf{W}^k]_{ij}+\alpha_{ij}^k\frac{[\mathbf{W}^k]_{ij}}{[\mathbf{Q}]_{ij}^k}[-\nabla_W\mathcal{D}(\mathbf{V},\mathbf{W}^k\mathbf{H})]_{ij} \tag{3.14}$$

with α_{ij}^k a positive step size that allows to control convergence of the algorithm. If we write explicitly the decomposition of the gradient, Equation (3.11) becomes:

$$[\mathbf{W}^{k+1}]_{ij}=[\mathbf{W}^k]_{ij}+\alpha_{ij}^k\frac{[\mathbf{W}^k]_{ij}}{[\mathbf{Q}^k]_{ij}}\left([\mathbf{P}^k]_{ij}-[\mathbf{Q}^k]_{ij}\right) \tag{3.15}$$

or

$$[\mathbf{W}^{k+1}]_{ij}=[\mathbf{W}^k]_{ij}+\alpha_{ij}^k[\mathbf{W}^k]_{ij}\left(\frac{[\mathbf{P}^k]_{ij}}{[\mathbf{Q}^k]_{ij}}-1\right). \tag{3.16}$$

Once we have the gradient type descent algorithm, we must determine the maximum value for the step size in order that $[\mathbf{W}^{k+1}]_{ij}\geq 0$, given $[\mathbf{W}^k]_{ij}\geq 0$. Note that, according to (3.15) or (3.16), a restriction must only apply if

$$[\mathbf{P}^k]_{ij}-[\mathbf{Q}^k]_{ij}<0 \tag{3.17}$$

since the other terms are positive. The maximum step size which ensures the positivity of $[\mathbf{W}^{k+1}]_{ij}$ is given by

$$(\alpha_{ij}^k)_{\max}=\frac{1}{1-\frac{[\mathbf{P}^k]_{ij}}{[\mathbf{Q}^k]_{ij}}} \tag{3.18}$$

which is strictly greater than 1. Finally, the maximum step size over all the components must satisfy

$$(\alpha^k)_{\max}\leq\min\{(\alpha_{ij}^k)_{\max}\}. \tag{3.19}$$

This choice ensures the non-negativity of all the components of $\mathbf{W}^k$ from iteration to iteration. Then, convergence of the algorithm is guaranteed by computing an appropriate step size, at each iteration, over the range $[0,(\alpha^k)_{\max}]$ by means of a simplified line search such as the Armijo rule for example. Finally, it is important to notice that the use of a step size equal to 1 leads to the very simple and well-known multiplicative form:

$$[\mathbf{W}^{k+1}]_{ij} = [\mathbf{W}^k]_{ij} \frac{[\mathbf{P}^k]_{ij}}{[\mathbf{Q}^k]_{ij}}. \tag{3.20}$$

This form is used because it is very easy to implement and it guarantees the non-negativity of successive iterates for an initial non-negative value $[\mathbf{W}^0]_{ij} \geq 0$. The main and important drawback is that the convergence of the algorithm is not assured in the general case, but only for specific cases of $[\mathbf{P}]$ and $[\mathbf{Q}]$.

3.3 Minimization with respect to $\mathbf{H}$

Minimization of (3.1) with respect to $\mathbf{H}$ leads to the following Karush-Kuhn-Tucker conditions for all i,j at the solution $\mathbf{W}^*$, $\mathbf{\Lambda}^*$:

$$[\nabla_H \mathcal{L}(\mathbf{V}, \mathbf{W}^*\mathbf{H}; \mathbf{\Lambda}, \mathbf{\Omega}^*)]_{ij} = 0, \tag{3.21}$$

$$[\mathbf{\Omega}^*]_{ij} \geq 0, \tag{3.22}$$

$$[\mathbf{H}^*]_{ij} \geq 0, \tag{3.23}$$

$$\langle \mathbf{\Omega}^*, \mathbf{H}^* \rangle = 0 \Leftrightarrow [\mathbf{\Omega}^*]_{ij}[\mathbf{H}^*]_{ij} = 0. \tag{3.24}$$

Condition (3.21) immediately leads to

$$[\mathbf{\Omega}^*]_{ij} = [\nabla_H \mathcal{D}(\mathbf{V}, \mathbf{W}\mathbf{H}^*)]_{ij}. \tag{3.25}$$

Condition (3.24) then becomes

$$\begin{aligned} &[\mathbf{H}^*]_{ij}[\nabla_H \mathcal{D}(\mathbf{V}, \mathbf{W}\mathbf{H}^*)]_{ij} = 0 \\ &\qquad \Leftrightarrow [\mathbf{H}^*]_{ij}[-\nabla_H \mathcal{D}(\mathbf{V}, \mathbf{W}\mathbf{H}^*)]_{ij} = 0. \end{aligned} \tag{3.26}$$

where the extra minus sign in the last expression is just used to make the negative gradient descent direction of $\mathcal{D}(\mathbf{V}, \mathbf{W}\mathbf{H})$ apparent.

The expression (3.24) gives the condition that must be satisfied for any optimization problem under non-negativity constraint. At the solution, the inner product between the gradient of the cost function and the variables must be equal to zero. The interpretation is the following: either our solution is the one that minimizes the cost function and the minimizer is positive, either the minimizer of the cost function is negative or zero and the constrained solution is zero.

This condition is non linear w.r.t. the unknowns, an analytical solution does not exist.

3.3.1 Gradient descent method

Since the gradient of the functional is computable, a natural choice for the iterative algorithm is a gradient descent method.

If we consider first the minimization problem without non-negativity constraint:

$$\min_{\mathbf{W},\mathbf{H}} \mathcal{D}(\mathbf{V},\mathbf{WH}), \tag{3.27}$$

we use the negative gradient as a descent direction and we write:

$$[\mathbf{H}^{k+1}]_{ij} = [\mathbf{H}^k]_{ij} + \beta_{ij}^k[-\nabla_H \mathcal{D}(\mathbf{V},\mathbf{WH}^*)]_{ij} \tag{3.28}$$

with β_{ij}^k a positive step size that allows to control convergence of the algorithm.

If now, we consider the minimization problem with non-negativity constraint, Equation (2.3), the descent direction becomes $[\mathbf{H}^*]_{ij}[-\nabla_H \mathcal{D}(\mathbf{V},\mathbf{WH}^*)]_{ij}$, Equation (3.26) and the descent algorithm is:

$$[\mathbf{H}^{k+1}]_{ij} = [\mathbf{H}^k]_{ij} + \beta_{ij}^k[\mathbf{H}^k]_{ij}[-\nabla_H \mathcal{D}(\mathbf{V},\mathbf{WH}^*)]_{ij}. \tag{3.29}$$

More generally:

$$\mathbf{N} \cdot \mathbf{H} \cdot [-\nabla_H \mathcal{D}(\mathbf{V},\mathbf{WH}^*)] \tag{3.30}$$

is a gradient descent direction of $\mathcal{D}$ if $\mathbf{N}$ is a matrix with positive entries, where $\cdot$ denotes the Hadamard product. A particular choice for $\mathbf{N}$ with a specific decomposition of $[-\nabla_H \mathcal{D}(\mathbf{V},\mathbf{WH}^*)]$ leads to the SGM algorithm.

3.3.2 Split Gradient Method (SGM)

The SGM algorithm is a descent algorithm whose direction is constructed in such a way that, for a step size equal to one, we obtain a multiplicative algorithm. To obtain it, an additional point is that $[-\nabla_H \mathcal{D}]_{ij}$ can always be decomposed as $[\mathbf{R}]_{ij} - [\mathbf{S}]_{ij}$, where $[\mathbf{R}]_{ij}$ and $[\mathbf{S}]_{ij}$ are positive entries, let us note that this decomposition is obviously not unique. If we take for $\mathbf{N}$, Equation (3.30):

$$[\mathbf{N}]_{ij} = \frac{1}{[\mathbf{S}]_{ij}}, \tag{3.31}$$

we obtain the following gradient-descent algorithm:

$$[\mathbf{H}^{k+1}]_{ij} = [\mathbf{H}^k]_{ij} + \beta_{ij}^k \frac{[\mathbf{R}^k]_{ij}}{[\mathbf{S}^k]_{ij}}[-\nabla_H \mathcal{D}(\mathbf{V},\mathbf{WH}^k)]_{ij} \tag{3.32}$$

with β_{ij}^k a positive step size that allows to control convergence of the algorithm. If we write explicitly the decomposition of the gradient, Equation (3.32) becomes:

$$[\mathbf{H}^{k+1}]_{ij} = [\mathbf{H}^k]_{ij} + \beta_{ij}^k \frac{[\mathbf{H}^k]_{ij}}{[\mathbf{R}^k]_{ij}}\left([\mathbf{R}^k]_{ij} - [\mathbf{S}^k]_{ij}\right) \tag{3.33}$$

or

$$[\mathbf{H}^{k+1}]_{ij} = [\mathbf{H}^k]_{ij} + \beta_{ij}^k [\mathbf{H}^k]_{ij} \left(\frac{[\mathbf{R}^k]_{ij}}{[\mathbf{S}^k]_{ij}} - 1 \right). \tag{3.34}$$

Once we have the gradient type descent algorithm, we must determine the maximum value for the step size in order that $[\mathbf{H}^{k+1}]_{ij} \geq 0$, given $[\mathbf{H}^k]_{ij} \geq 0$. Note that, according to (3.33), a restriction must only apply if

$$[\mathbf{R}^k]_{ij} - [\mathbf{S}^k]_{ij} < 0 \tag{3.35}$$

since the other terms are positive. The maximum step size which ensures the positivity of $[\mathbf{H}^{k+1}]_{ij}$ is given by

$$(\beta_{ij}^k)_{\max} = \frac{1}{1 - \frac{[\mathbf{R}^k]_{ij}}{[\mathbf{S}^k]_{ij}}} \tag{3.36}$$

which is strictly greater than 1. Finally, the maximum step size over all the components must satisfy

$$(\beta^k)_{\max} \leq \min\{(\beta_{ij}^k)_{\max}\}. \tag{3.37}$$

This choice ensures the non-negativity of all the components of $\mathbf{H}^k$ from iteration to iteration. Then, convergence of the algorithm is guaranteed by computing an appropriate step size, at each iteration, over the range $[0, (\beta^k)_{\max}]$ by means of a simplified line search such as the Armijo rule for example. Finally, it is important to notice that the use of a step size equal to 1 leads to the very simple and well-known multiplicative form:

$$[\mathbf{H}^{k+1}]_{ij} = [\mathbf{H}^k]_{ij} \frac{[\mathbf{R}^k]_{ij}}{[\mathbf{S}^k]_{ij}}. \tag{3.38}$$

This form is used because it is very easy to implement and it guarantees the non-negativity of successive iterates for an initial non-negative value $[\mathbf{H}^0]_{ij} \geq 0$. The main and important drawback is that the convergence of the algorithm is not assured.

3.4 Explicit expressions of the gradients

Before ending this section, let us compute $\nabla \mathcal{D}$ with respect to $\mathbf{H}$ and $\mathbf{W}$, using Equations (2.1) and (2.4). It can be expressed in matrix form as follows:

$$\nabla_H \mathcal{D} = \mathbf{W}^T \mathbf{A} \qquad \nabla_W \mathcal{D} = \mathbf{A} \mathbf{H}^T \tag{3.39}$$

where $\mathbf{A}$ is a matrix whose (i, j)-th entry is given by:

$$[\mathbf{A}]_{ij} = \frac{\partial d_{ij}}{\partial [\mathbf{WH}]_{ij}}. \tag{3.40}$$

Equations (3.20), (3.38) associated to (3.39), (3.40), lead to the multiplicative algorithms described in (Cichoki *et al.* 2006; Févotte *et al.* 2009; Lee & Seung 2001). These are particular cases of the relaxed algorithms (3.15) (3.33), when a unit step size is used.

4 Minimization under non-negativity constraints and flux conservation

Let us now consider problem (2.5), which differs from (2.3) by additional flux constraints.

4.1 Flux conservation constraints

We make the following variable changes:

$$[\mathbf{W}]_{ij} = \frac{[\mathbf{Z}]_{ij}}{\sum_m [\mathbf{Z}]_{mj}}; \tag{4.1}$$

$$[\mathbf{H}]_{ij} = \left(\sum_m [\mathbf{V}]_{mj}\right) \frac{[\mathbf{T}]_{ij}}{\sum_m [\mathbf{T}]_{mj}}. \tag{4.2}$$

The term $(\sum_m [\mathbf{V}]_{mj})$ comes from the fact that $[\mathbf{H}]_{ij}$ is normalized to the column j of $\mathbf{V}$. In so doing, the problem becomes unconstrained with respect to the flux but we must search the solution in a domain where the denominator is a constant to ensure that the problem remains convex w.r.t. the new variables. It is an important point performed by our method. The flux conservation being provided by the change of variables, we can proceed the SGM on the new variables to ensure both the non-negativity and the flux conservation.

To deal with the non-negativity constraints, let us consider again the SGM algorithm and compute the gradient with respect to new variables.

4.2 Explicit expressions of the gradients

Let us compute expression of the gradients w.r.t. the new variables:

$$\frac{\partial \mathcal{D}}{\partial [\mathbf{Z}]_{lj}} = \sum_i \frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{ij}} \times \frac{\partial [\mathbf{W}]_{ij}}{\partial [\mathbf{Z}]_{lj}}, \tag{4.3}$$

$$\frac{\partial \mathcal{D}}{\partial [\mathbf{T}]_{lj}} = \sum_i \frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{ij}} \times \frac{\partial [\mathbf{H}]_{ij}}{\partial [\mathbf{T}]_{lj}} \tag{4.4}$$

where, in a compact form,

$$\frac{\partial [\mathbf{W}]_{ij}}{\partial [\mathbf{Z}]_{lj}} = \frac{1}{\sum_m [\mathbf{Z}]_{mj}} \times (\delta_{li} - [\mathbf{W}]_{ij}), \tag{4.5}$$

$$\frac{\partial [\mathbf{H}]_{ij}}{\partial [\mathbf{T}]_{lj}} = \frac{\sum_m [\mathbf{V}]_{mj}}{\sum_m [\mathbf{T}]_{mj}} \times \left(\delta_{li} - \frac{[\mathbf{H}]_{ij}}{\sum_m [\mathbf{V}]_{mj}}\right) \tag{4.6}$$

with δ_{li} the Kronecker symbol. As a consequence, the components of the opposite of the gradient of $\mathcal{D}$ with respect to the new variables can now be written as

$$-\frac{\partial \mathcal{D}}{\partial [\mathbf{Z}]_{lj}} = \frac{1}{\sum_m [\mathbf{Z}]_{mj}} \left(\left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{lj}}\right) - \sum_i [\mathbf{W}]_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{ij}}\right)\right) \tag{4.7}$$

and
$$-\frac{\partial \mathcal{D}}{\partial [\mathbf{T}]_{lj}} = \frac{\sum_m [\mathbf{V}]_{mj}}{\sum_m [\mathbf{T}]_{mj}} \left(\left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{lj}} \right) - \frac{\sum_i [\mathbf{H}]_{ij}}{\sum_m [\mathbf{V}]_{mj}} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{ij}} \right) \right) \tag{4.8}$$

4.3 SGM with the normalized variables

We solve both the split of the gradient between two positive functions and the conservation of the convexity w.r.t. to the new variables by making the shift of the form:

$$\begin{aligned} (-\partial \mathcal{D} / \partial [\mathbf{W}]_{ij})_s &\longleftarrow (-\partial \mathcal{D} / \partial [\mathbf{W}]_{ij}) + \eta, \quad \forall (i,j), \\ (-\partial \mathcal{D} / \partial [\mathbf{H}]_{ij})_s &\longleftarrow (-\partial \mathcal{D} / \partial [\mathbf{H}]_{ij}) + \mu, \quad \forall (i,j). \end{aligned}$$

Let us notice that this shift leaves Equations (4.9) and (4.14) unchanged. Consequently, using

$$\eta = -\min_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{ij}} \right) + \epsilon, \quad \mu = -\min_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{ij}} \right) + \epsilon$$

does not modify the gradient of $\mathcal{D}$ with respect to the new variables $\mathbf{Z}$ and $\mathbf{T}$, but ensures the non-negativity of $(-\partial \mathcal{D} / \partial [\mathbf{W}]_{ij})_s$ and $(-\partial \mathcal{D} / \partial [\mathbf{H}]_{ij})_s$. A constant ϵ is added to avoid numerical instability, however, it must be chosen small enough not to slow down the minimization. Let us note that this particular decomposition allows to ensure that the denominator in (4.1) and (4.2) remains constant and then we are always in the convexity domain. We shall now apply the SGM method.

4.4 Minimization with respect to $\mathbf{W}$

Consider the following gradient (4.9) decomposition:

$$[-\nabla_Z \mathcal{D}]_{ij} = [\mathbf{P}]_{ij} - [\mathbf{Q}]_{ij} \tag{4.9}$$

that involves the non-negative entries defined as follows

$$[\mathbf{P}]_{ij} = \left(-\frac{\partial D}{\partial [\mathbf{W}]_{ij}} \right)_s, \tag{4.10}$$

$$[\mathbf{Q}]_{ij} = [\mathbf{Q}]_{.j} = \sum_i [\mathbf{W}]_{ij} \left(-\frac{\partial D}{\partial [\mathbf{W}]_{ij}} \right)_s. \tag{4.11}$$

The relaxed form of the minimization algorithm can be expressed as

$$[\mathbf{Z}^{k+1}]_{lj} = [\mathbf{Z}^k]_{lj} + \alpha^k [\mathbf{Z}^k]_{lj} \left(\frac{(-\partial \mathcal{D} / \partial [\mathbf{W}^k]_{lj})_s}{\sum_i [\mathbf{W}^k]_{ij} (-\partial \mathcal{D} / \partial [\mathbf{W}^k]_{ij})_s} - 1 \right).$$

We clearly have $\sum_l [\mathbf{Z}^{k+1}]_{lj} = \sum_l [\mathbf{Z}^k]_{lj}$, for all α^k. This allows us to express the algorithm with respect to the initial variable $\mathbf{W}$, that is,

$$[\mathbf{W}^{k+1}]_{lj} = [\mathbf{W}]^k_{lj} + \alpha^k [\mathbf{W}]^k_{lj} \left(\frac{(-\partial \mathcal{D} / \partial [\mathbf{W}]^k_{lj})_s}{\sum_i [\mathbf{W}]^k_{ij} (-\partial \mathcal{D} / \partial [\mathbf{W}]^k_{ij})_s} - 1 \right). \tag{4.12}$$

Again, with a constant step size equal to 1, the algorithm takes a simple multiplicative form:

$$[\mathbf{W}^{k+1}]_{lj} = [\mathbf{W}^k]_{lj} \frac{(-\partial\mathcal{D}/\partial[\mathbf{W}^k]_{lj})_s}{\sum_i [\mathbf{W}^k]_{ij}(-\partial\mathcal{D}/\partial[\mathbf{W}^k]_{ij})_s}. \tag{4.13}$$

4.5 Minimization with respect to H

In an analogous way, consider the following gradient (4.14) decomposition:

$$[-\nabla_T\mathcal{D}]_{ij} = [\mathbf{R}]_{ij} - [\mathbf{S}]_{ij} \tag{4.14}$$

that involves the non-negative entries given by

$$[\mathbf{R}]_{ij} = \frac{\sum_m [\mathbf{V}]_{mj}}{\sum_m [\mathbf{T}]_{mj}} \left(-\frac{\partial\mathcal{D}}{\partial[\mathbf{H}]_{ij}}\right)_s, \tag{4.15}$$

$$[\mathbf{S}]_{ij} = S_{.j} = \frac{\sum_m [\mathbf{V}]_{m,j}}{\sum_m [\mathbf{T}]_{mj}} \sum_i \frac{[\mathbf{H}]_{ij}}{\sum_m [\mathbf{V}]_{mj}} \left(-\frac{\partial\mathcal{D}}{\partial[\mathbf{H}]_{ij}}\right)_s. \tag{4.16}$$

This leads to the relaxed form of optimization algorithm with respect to variable **T**, that is,

$$[\mathbf{T}^{k+1}]_{lj} = [\mathbf{T}^k]_{lj} + \alpha^k [\mathbf{T}^k]_{lj} \left(\frac{(-\partial\mathcal{D}/\partial[\mathbf{H}^k]_{lj})_s}{\sum_i \frac{[\mathbf{H}^k]_{ij}}{\sum_m [\mathbf{V}]_{mj}} (-\partial\mathcal{D}/\partial[\mathbf{H}^k]_{ij})_s} - 1 \right).$$

It can be seen that $\sum_l [\mathbf{T}^{k+1}]_{lj} = \sum_l [\mathbf{T}^k]_{lj}$, for all α^k, which implies that

$$[\mathbf{H}^{k+1}]_{lj} = [\mathbf{H}^k]_{lj} + \alpha^k [\mathbf{H}^k]_{lj} \left(\frac{(-\partial\mathcal{D}/\partial[\mathbf{H}^k]_{lj})_s}{\sum_i \frac{[\mathbf{H}^k]_{ij}}{\sum_m [\mathbf{V}]_{mj}} (-\partial\mathcal{D}/\partial[\mathbf{H}^k]_{ij})_s} - 1 \right). \tag{4.17}$$

The multiplicative form is obtained with a constant step size equal to 1, namely,

$$[\mathbf{H}^{k+1}]_{lj} = [\mathbf{H}^k]_{lj} \frac{(-\partial\mathcal{D}/\partial[\mathbf{H}^k]_{lj})_s}{\sum_i [\mathbf{H}^k]_{ij}(-\partial\mathcal{D}/\partial[\mathbf{H}^k]_{ij})_s} \sum_m [\mathbf{V}]_{mj}. \tag{4.18}$$

In the next section, we propose to illustrate this algorithm within the field of hyperspectral imaging.

5 Choice of the descent step size and convergence speed

On one hand, if the descent step size is fixed to one, there is no way to modify the convergence speed and the iterations number can be high, moreover, the convergence is not ensured but the algorithm takes a simple form. On the other hand, if the descent step size is searched by a simple rule, Armijo for example, the iterations number decreases but the duration of one iteration increases, from our experience, when the step size is computed, the overall gain is about ten or twenty percents and in this case the convergence is ensured.

6 Physical context: Hyperspectral imagery

Hyperspectral imaging has received considerable attention in the last few years. See for instance (Chang 2003), (Landgrebe 2003) and references therein. It consists of data acquisition with high sensitivity and resolution in hundreds contiguous spectral bands, geo-referenced within the same coordinate system. With its ability to provide extremely detailed data regarding the spatial and spectral characteristics of a scene, this technology offers immense new possibilities in collecting and managing information for civilian and military application areas.

Each vector pixel of an hyperspectral image characterizes a local spectral signature. Usually, one consider that each vector pixel can be modeled accurately as a linear mixture of different pure spectral components, called endmembers. Referring to our notations, each column of $\mathbf{V}$ can thus be interpreted as a spectral signature obtained by linear mixing of the spectra of endmembers, *i.e.*, the columns of $\mathbf{W}$. The problem is then to estimate the endmember spectra $\mathbf{W}$ and the abundance coefficients $\mathbf{H}$ from the spectral signatures $\mathbf{V}$.

In all the simulations presented in this paper, the end members are extracted from the ENVI library (ENVI 2003).

7 Simulation results

Many simulations have been performed to validate the proposed algorithm, Equations (4.13) and (4.18). The experiment presented in this paper corresponds to 10 linear mixtures of 3 endmembers, the length of each spectrum being 826. The three endmembers used in this example correspond to the spectra of the construction concrete, green grass, and micaceous loam. The chosen cost function for $\mathcal{D}$ is the Frobenius norm:

$$\mathcal{D}(\mathbf{V}, \mathbf{WH}) = \sum_{ij} ([\mathbf{WH}]_{ij} - [\mathbf{V}]_{ij})^2 = (\mathbf{WH} - \mathbf{V})^T(\mathbf{WH} - \mathbf{V}). \tag{7.1}$$

The used procedure is the following:

1. Take the spectra from a library (ENVI here).
2. Generate randomly the KN abundance coefficients H_{ij} in a given interval.
3. Compute $\mathbf{V}$.
4. Generate randomly H^0 and W^0 in the space constraints.
5. Compute the chosen cost function, here the Frobenius norm:
6. Compute the decomposition of the gradient w.r.t. Z, *i.e.* (4.10) and (4.11).
7. Compute $\mathbf{W}^{k+1}$, (4.12).
8. Compute the decomposition of the gradient w.r.t. T, *i.e.* (4.15) and (4.16).

9. Compute $\mathbf{H}^{k+1}$, (4.17).

10. Until the stopping criterion:

$$\frac{\mathcal{D}(\mathbf{V}, \mathbf{W}^{k+1}\mathbf{H}^{k+1}) - \mathcal{D}(\mathbf{V}, \mathbf{W}^{k}\mathbf{H}^{k})}{\mathcal{D}(\mathbf{V}, \mathbf{W}^{k}\mathbf{H}^{k})} \leq 10^{-10}. \tag{7.2}$$

Figure 1 shows the behaviour of the criterion $\mathcal{D}$ as a function of the number of iterations, and the 10 reconstructed spectra in comparison with the true ones. Figure 2 shows the estimated endmembers (columns of $\mathbf{W}$), and their abundance coefficients (rows of $\mathbf{H}$) after 12 000 iterations, and compared with the true values. Note that the initial values for $\mathbf{W}$ and $\mathbf{H}$ were chosen to satisfy the constraints, *i.e.*, positivity, sum to one of the columns of $\mathbf{W}$. We clearly see that the curves coincide almost perfectly. The normalization of the columns of matrix $\mathbf{W}$, as well as the flux conservation between $\mathbf{V}$ and $\mathbf{H}$, are satisfied at each iteration. Let us note that $\mathbf{H}$ and $\mathbf{W}$ could be estimated up to a permutation of the columns of $\mathbf{W}$, and to an analogous permutation of the rows of $\mathbf{H}$.

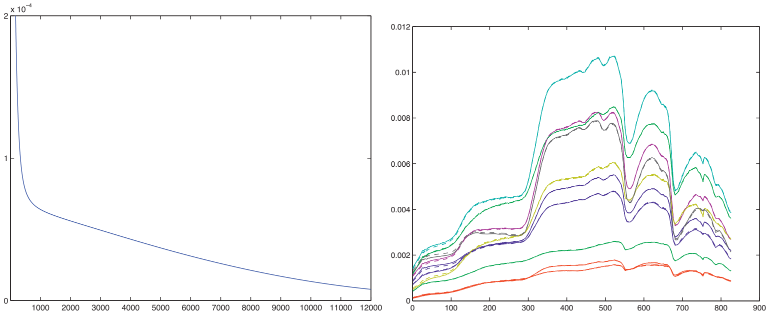

Fig. 1. Frobenius $\mathcal{D}(\mathbf{V}, \mathbf{WH})$ as a function of the number of iterations. Columns of $\mathbf{V}$ at the end of the iterations, solid line for true values, dashed line for estimated values.

8 Regularization

In full generality, we can add several regularization terms depending on one or two variables, the only constraint being that each regularization function must be convex w.r.t. the relevant variable. If the regularization term depends on the two variables, it must be convex w.r.t. one variable, the other being fixed. Here, we consider the case where the regularization penalty terms are incorporated separately on the columns of $\mathbf{W}$ and $\mathbf{H}$, and are added to the data consistency term $\mathcal{D}(\mathbf{V}, \mathbf{WH})$. Then the penalized objective function expresses as

$$\mathcal{D}_{\text{reg}}(\mathbf{V}, \mathbf{WH}) = \mathcal{D}(\mathbf{V}, \mathbf{WH}) + \gamma_1 \mathcal{F}_1(\mathbf{W}) + \gamma_2 \mathcal{F}_2(\mathbf{H}) \tag{8.1}$$

where $\mathcal{F}_1(\mathbf{W})$ and $\mathcal{F}_2(\mathbf{H})$ are penalty functions, and γ_1, γ_2 the respective regularization factors. The general rules given for SGM remain true for the regularized versions of the algorithms.

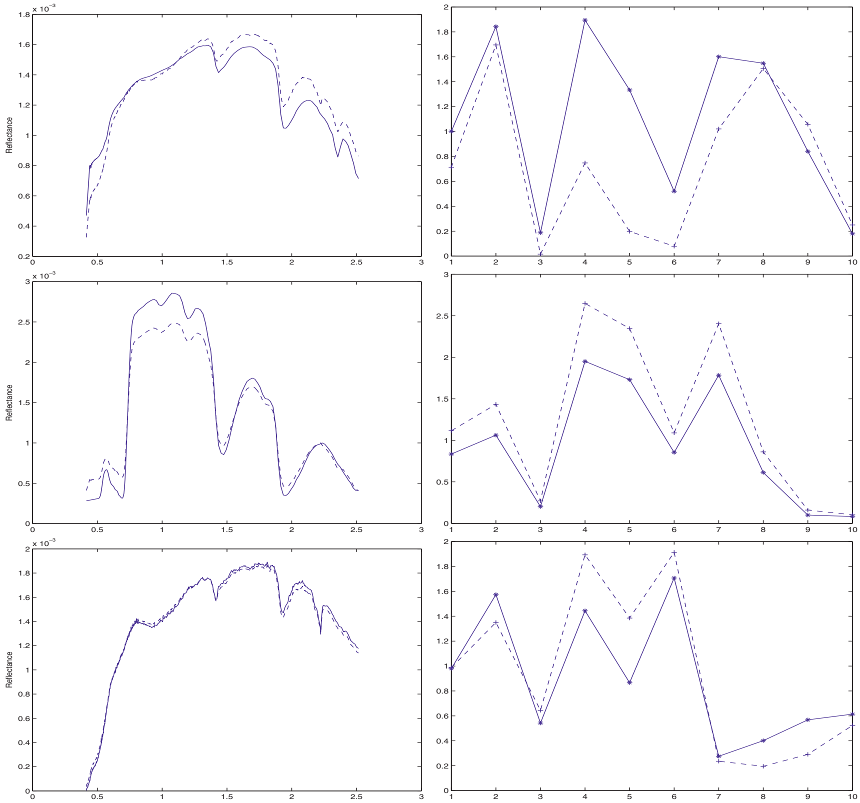

Fig. 2. On *the left*, columns of **W**. On *the right*, rows of **H**. On each plot: solid line for true values, dashed line for estimated values.

The minimization of $\mathcal{D}_{\text{reg}}$ w.r.t. the variable **Z** must take into account the regularization function $\mathcal{F}_1(\mathbf{W})$, Equation (4.1):

$$-\nabla_Z \mathcal{D}_{\text{reg}} = -\nabla_Z \mathcal{D} - \gamma_1 \nabla_Z \mathcal{F}_1, \tag{8.2}$$

and the minimization of $\mathcal{D}_{\text{reg}}$ w.r.t. the variable **T** must take into account the regularization function $\mathcal{F}_2(\mathbf{H})$, Equation (4.2).

$$-\nabla_T \mathcal{D}_{\text{reg}} = -\nabla_T \mathcal{D} - \gamma_2 \nabla_T \mathcal{F}_2. \tag{8.3}$$

In the following, we consider one regularization term at a time, that is, first on the spectra and then on the abundance coefficients.

8.1 Regularized SGM on the spectra $\mathbf{W}$

We develop in this section expressions of SGM for a regularization $\mathcal{F}_1$ on the normalized endmembers spectra $\mathbf{W}$, we have:

$$-\nabla_Z \mathcal{D}_{\text{reg}} = -\nabla_Z \mathcal{D} - \gamma_1 \nabla_Z \mathcal{F}_1, \tag{8.4}$$

$$-\nabla_T \mathcal{D}_{\text{reg}} = -\nabla_T \mathcal{D}. \tag{8.5}$$

The component of the opposite of the gradient of $\mathcal{D}_{\text{reg}}$ with respect to $\mathbf{Z}$ is:

$$\begin{aligned} -\frac{\partial \mathcal{D}_{\text{reg}}}{\partial [\mathbf{Z}]_{lj}} = \frac{1}{\sum_m [\mathbf{Z}]_{mj}} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{lj}} - \gamma_1 \frac{\partial \mathcal{F}_1}{\partial [\mathbf{W}]_{lj}} \right)_s - \\ \frac{1}{\sum_m [\mathbf{Z}]_{mj}} \sum_i [\mathbf{W}]_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{ij}} - \gamma_1 \frac{\partial \mathcal{F}_1}{\partial [\mathbf{W}]_{ij}} \right)_s \end{aligned} \tag{8.6}$$

In the same way that for the non regularized SGM, we solve both the split of the gradient between two positive functions and the conservation of the convexity w.r.t. the new variables by making the shift of the form:

$$\left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{ij}} - \gamma_1 \frac{\partial \mathcal{F}_1}{\partial [\mathbf{W}]_{ij}} \right)_s \quad \longleftarrow \quad \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{ij}} - \gamma_1 \frac{\partial \mathcal{F}_1}{\partial [\mathbf{W}]_{ij}} \right) + \eta + \epsilon \quad \forall (i,j)$$

with

$$\eta = -\min_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{ij}} - \gamma_1 \frac{\partial \mathcal{F}_1}{\partial [\mathbf{W}]_{ij}} \right).$$

The decomposition of the gradient of the regularized cost function w.r.t. $\mathbf{Z}$ is:

$$[-\nabla_Z \mathcal{D}_{\text{reg}}]_{ij} = [\mathcal{P}]_{ij} - [\mathcal{Q}]_{ij} \tag{8.7}$$

with

$$\begin{aligned} [\mathcal{P}]_{ij} &= \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{ij}} - \gamma_1 \frac{\partial \mathcal{F}_1}{\partial [\mathbf{W}]_{ij}} \right)_s, \\ [\mathcal{Q}]_{lj} = [\mathcal{Q}]_{.j} &= \sum_i [\mathbf{W}]_{.j} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{ij}} - \gamma_1 \frac{\partial \mathcal{F}_1}{\partial [\mathbf{W}]_{ij}} \right)_s \end{aligned} \tag{8.8}$$

and the iterate on $\mathbf{W}$ is:

$$[\mathbf{W}^{k+1}]_{lj} = [\mathbf{W}]^k_{lj} + \alpha^k [\mathbf{W}]^k_{lj} \left(\frac{\left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{lj}} - \gamma_1 \frac{\partial \mathcal{F}_1}{\partial [\mathbf{W}]_{lj}} \right)_s}{\sum_i [\mathbf{W}]_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{ij}} - \gamma_1 \frac{\partial \mathcal{F}_1}{\partial [\mathbf{W}]_{ij}} \right)_s} - 1 \right). \tag{8.9}$$

In the same way that for the non regularized SGM, with a constant step size equal to one, we get:

$$[\mathbf{W}^{k+1}]_{lj} = [\mathbf{W}]^k_{lj} \frac{\left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{lj}} - \gamma_1 \frac{\partial \mathcal{F}_1}{\partial [\mathbf{W}]_{lj}} \right)_s}{\sum_i [\mathbf{W}]_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{W}]_{ij}} - \gamma_1 \frac{\partial \mathcal{F}_1}{\partial [\mathbf{W}]_{ij}} \right)_s}. \tag{8.10}$$

The iterate on $\mathbf{H}$ is still given by Equation (4.17) or Equation (4.18) for a unit step size.

8.1.1 Tikhonov smoothness regularization

The well known Tikhonov regularization expresses some smoothness of the solution and is applied, here, on endmember spectra, *i.e.* on the columns of $\mathbf{W}$. This is justified by physical considerations, spectra varying slowly as a function of the wavelength. Consequently, the regularization function is:

$$\mathcal{F}_1(\mathbf{W}) = \frac{1}{2}\sum_{ij}([\mathbf{W}]_{ij} - c)^2 \tag{8.11}$$

with c a constant positive or zero, or more generally

$$\mathcal{F}_1(\mathbf{W}) = \frac{1}{2}\sum_{ij}[\partial_{1,2}\mathbf{W}]^2_{ij} \tag{8.12}$$

where $\partial_{1,2}$ is the first or second-order derivative operator. For simplicity, we approximate $\partial_{1,2}\mathbf{W}$ in a closed numerical form as

$$[\partial_{1,2}\mathbf{W}]_{ij} = [\mathbf{W}]_{ij} - [\mathbf{AW}]_{ij} \tag{8.13}$$

where $\mathbf{AW}$ stands for the convolution of each column of matrix $\mathbf{W}$ by a mask, *e.g.* $[1\ 0\ 0]$ and $[\frac{1}{2}\ 0\ \frac{1}{2}]$ for the first and second-order derivative operators, respectively. In this case, the opposite of the gradient can be expressed in matrix form as follows:

$$-[\nabla_{\mathbf{W}}\mathcal{F}_1]_{ij} = [(\mathbf{A} + \mathbf{A}^\top)\mathbf{W}]_{ij} - [(\mathbf{A}^\top\mathbf{A} + \mathbf{I})\mathbf{W}]_{ij}. \tag{8.14}$$

Note that Tikhonov regularization with the basic SGM algorithm was initially associated to the basic SGM algorithm in (Lantéri *et al.* 2011), *i.e.*, without flux constraint. The interested reader is invited to consult this reference for an overview of the results that have been obtained.

8.1.2 Simulations results

As for the non regularized SGM, many simulations have been performed to validate the proposed algorithm, Equations (8.10) and (4.18). Note that the different forms of the regularization term give approximatively the same practical results. The experiment corresponds to 10 linear mixtures of 3 endmembers, the length of each spectrum being 826. A noise vector distributed according to a Gaussian distribution with zero-mean and covariance matrix $\sigma^2\mathbf{I}_N$, where $\mathbf{I}_N$ is the $N \times N$ identity matrix has been added to each column of $\mathbf{V}$. Note that this statistical model assumes that the noise variances are the same in all bands. Results are given for a snr equal to $20dB$. Figure 3 shows the estimated endmembers (columns of $\mathbf{W}$) after 12 000 iterations, and compared with the true values with and without regularization. Figures 4 and 5 show the 10 reconstructed spectra in comparison with the true ones, respectively without and with regularization. We clearly see the interest of the regularization on the estimation.

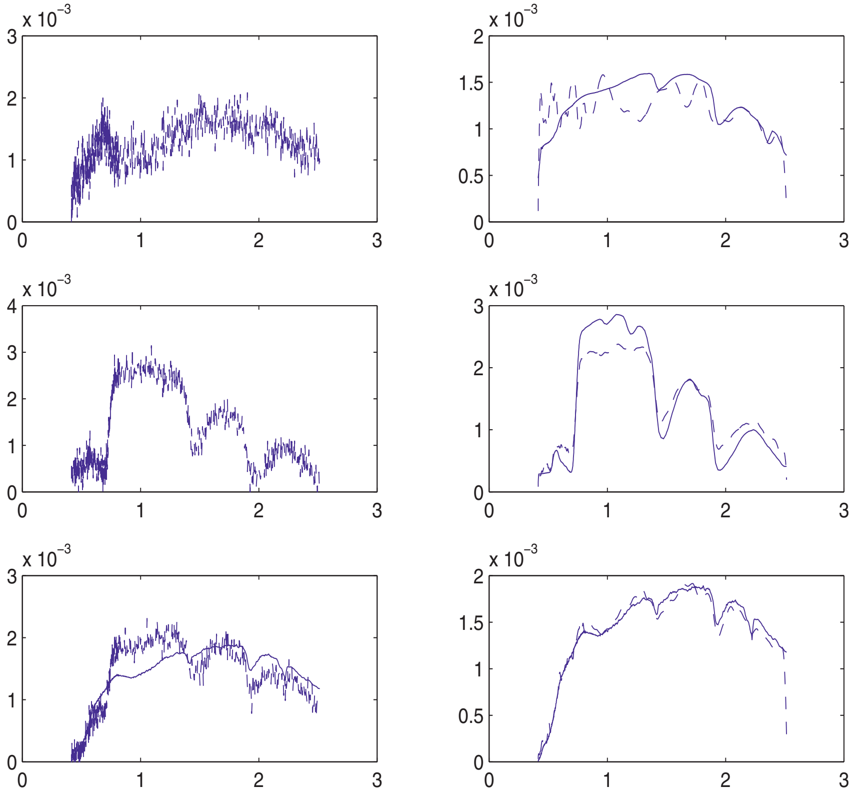

Fig. 3. Columns of $\mathbf{W}$. On each plot: solid line for true values, dashed line for estimated values. *Left column*: without regularization $\gamma = 0$. *Right column* with $\gamma = 0.1$.

8.2 Regularized SGM on the abundance coefficients $\mathbf{H}$

We develop in this section expressions of SGM for a regularization $\mathcal{F}_2$ on the normalized abundance coefficients $\mathbf{H}$, we have:

$$-\nabla_Z \mathcal{D}_{\text{reg}} = -\nabla_Z \mathcal{D} \tag{8.15}$$

$$-\nabla_T \mathcal{D}_{\text{reg}} = -\nabla_T \mathcal{D} - \gamma_2 \nabla_T \mathcal{F}_2. \tag{8.16}$$

In this case, the component of the opposite of the gradient of $\mathcal{D}_{\text{reg}}$ with respect to $\mathbf{T}$ is:

$$\begin{aligned} -\frac{\partial \mathcal{D}_{\text{reg}}}{\partial [\mathbf{T}]_{lj}} = \frac{1}{\sum_m [\mathbf{T}]_{mj}} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{lj}} - \gamma_2 \frac{\partial \mathcal{F}_2}{\partial [\mathbf{H}]_{lj}} \right)_s - \\ \frac{1}{\sum_m [\mathbf{T}]_{mj}} \sum_i [\mathbf{H}]_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{ij}} - \gamma_2 \frac{\partial \mathcal{F}_2}{\partial [\mathbf{H}]_{ij}} \right)_s . \end{aligned} \tag{8.17}$$

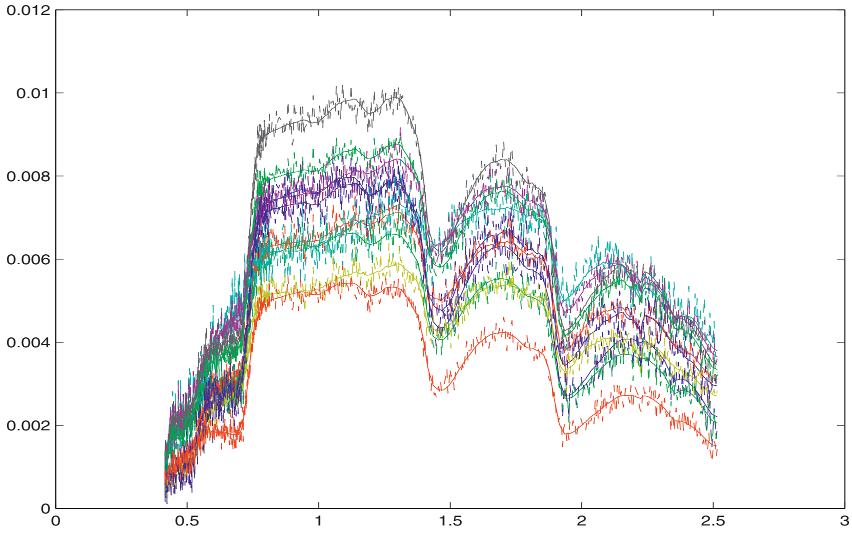

Fig. 4. Columns of $\mathbf{V}$, solid line for true values, dashed line for estimated values without regularization, $\gamma = 0$.

In the same way that for the non regularized SGM, we solve both the split of the gradient between two positive functions and the conservation of the convexity w.r.t. the new variables by making the shift of the form:

$$\left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{ij}} - \gamma_2 \frac{\partial \mathcal{F}_2}{\partial [\mathbf{H}]_{ij}}\right)_s \longleftarrow \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{ij}} - \gamma_2 \frac{\partial \mathcal{F}_2}{\partial [\mathbf{H}]_{ij}}\right) + \eta + \epsilon \quad \forall (i,j)$$

with

$$\eta = -\min_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{lj}} - \gamma_2 \frac{\partial \mathcal{F}_2}{\partial [\mathbf{H}]_{lj}}\right).$$

The decomposition of the gradient of the regularized cost function w.r.t. $\mathbf{T}$ is:

$$[-\nabla_T \mathcal{D}_{\text{reg}}]_{lj} = [\mathcal{R}]_{ij} - [\mathcal{S}]_{ij} \tag{8.18}$$

with

$$[\mathcal{R}]_{ij} = \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{ij}} - \gamma_2 \frac{\partial \mathcal{F}_2}{\partial [\mathbf{H}]_{lj}}\right)_s, \qquad [\mathcal{S}]_{ij} = \sum_i [\mathbf{H}]_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{ij}} - \gamma_2 \frac{\partial \mathcal{F}_2}{\partial [\mathbf{H}]_{ij}}\right)_s \tag{8.19}$$

and the iterate on $\mathbf{H}$ is:

$$[\mathbf{H}^{k+1}]_{lj} = [\mathbf{H}]^k_{lj} + \alpha^k [\mathbf{H}]^k_{lj} \left(\frac{\left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{lj}} - \gamma_2 \frac{\partial \mathcal{F}_2}{\partial [\mathbf{H}]_{lj}}\right)_s}{\sum_i [\mathbf{H}]_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{ij}} - \gamma_2 \frac{\partial \mathcal{F}_2}{\partial [\mathbf{H}]_{ij}}\right)_s} - 1 \right) \cdot \tag{8.20}$$

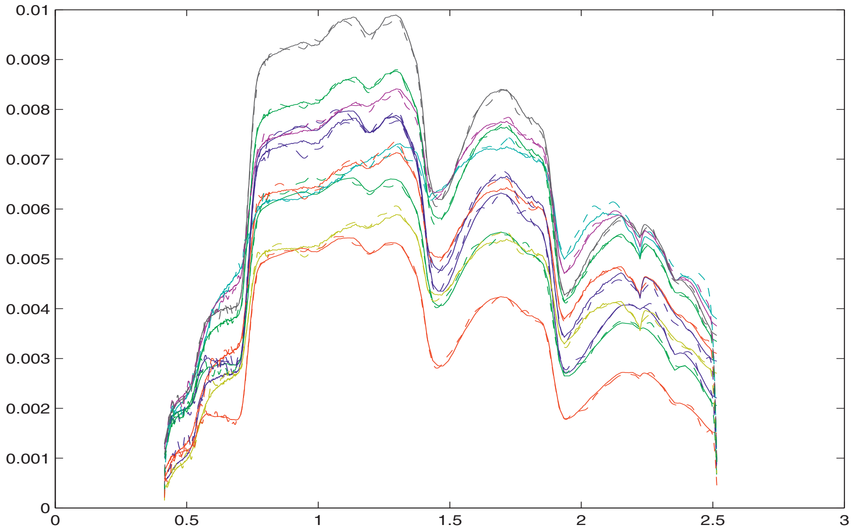

Fig. 5. Columns of **V**, solid line for true values, dashed line for estimated values with $\gamma = 0.1$.

In the same way that for the non regularized SGM, with a constant step size equal to one, we get:

$$[\mathbf{H}^{k+1}]_{lj} = [\mathbf{H}]^k_{lj} \frac{\left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{lj}} - \gamma_2 \frac{\partial \mathcal{F}_2}{\partial [\mathbf{H}]_{lj}}\right)_s}{\sum_i [\mathbf{H}]_{ij} \left(-\frac{\partial \mathcal{D}}{\partial [\mathbf{H}]_{ij}} - \gamma_2 \frac{\partial \mathcal{F}_2}{\partial [\mathbf{H}]_{ij}}\right)_s}. \tag{8.21}$$

The iterate on **W** is still given by Equation (4.12) or Equation (4.13) for a unit step size.

8.2.1 Sparsity-enforcing regularization

Such a penalty, which expresses that most of information may be concentrated in a few coefficients, mainly applies to the abundance coefficients, that is, to the columns of **H**. Keeping in mind that the algorithm satisfies flux conservation constraint, see (4.2), we are ready to consider the following sparsity measure σ introduced in (Hoyer (2004))

$$\sigma = \frac{\sqrt{K} - \frac{\|[\mathbf{H}]_{\bullet j}\|_1}{\|[\mathbf{H}]_{\bullet j}\|_2}}{\sqrt{K} - 1}, \qquad 0 \leq \sigma \leq 1 \tag{8.22}$$

with K the number of rows of $\mathbf{H}$, and $[\mathbf{H}]_{\bullet j}$ its j-th row. This clearly defines a relation between the ℓ_2-norm and the ℓ_1-norm of $[\mathbf{H}]_{\bullet j}$, the sum constraint on $\mathbf{H}$ associated with non negativity inducing a constant ℓ_1-norm.

$$\|[\mathbf{H}]_{\bullet j}\|_2^2 = \alpha^2 \|[\mathbf{H}]_{\bullet j}\|_1^2 \tag{8.23}$$

with

$$\alpha = \frac{1}{\sqrt{K} - \sigma(\sqrt{K}-1)}, \qquad \frac{1}{\sqrt{K}} \le \alpha \le 1. \tag{8.24}$$

Note that only two values for σ lead to unambiguous situations; if α is equal to one, only one entry of $[\mathbf{H}]_{\bullet j}$ is nonzero; if $\alpha = 1/\sqrt{K}$, all the entries of $[\mathbf{H}]_{\bullet j}$ are equal. Any other value for α can correspond to different sets of entries. As a consequence, we suggest to consider the following penalty function[2]:

$$\mathcal{F}_2(\mathbf{H}) = \frac{1}{2}\sum_j \left(\|[\mathbf{H}]_{\bullet j}\|_2^2 - \alpha^2 \|[\mathbf{H}]_{\bullet j}\|_1^2\right)^2 \tag{8.25}$$

with α equal to one, and use of the regularization factor γ_2 in (8.1) to push $[\mathbf{H}]_{\bullet j}$ toward a sparse solution. For convenience, let us provide the opposite of the gradient of $\mathcal{F}_2(\mathbf{H})$

$$\begin{aligned} -[\nabla_{\mathbf{H}}\mathcal{F}_2]_{ij} = (\alpha^2 \|[\mathbf{H}]_{\bullet j}\|_1^2 - \|[\mathbf{H}]_{\bullet j}\|_2^2) \\ ([\mathbf{H}]_{ij} - \alpha^2 \|[\mathbf{H}]_{\bullet j}\|_1) \end{aligned} \tag{8.26}$$

to be used in (8.21). In the next section, we shall test this algorithm for hyperspectral data unmixing.

8.2.2 Simulations results

To test interest of sparsity regularization on the abundance coefficients, we take 20 linear mixtures of 6 endmembers, the length of each spectrum being 826. The six endmembers correspond to the construction concrete, green grass, micaceous loam, olive green paint, bare red brick and galvanized steel metal.

In order to characterize the performance of our approach, and show that it tends to provide sparse solutions, we considered a matrix $\mathbf{H}$ with only one nonzero entry per column. This entry was selected randomly and set to one. See Figure 6. Each observed spectrum was corrupted by an additive white Gaussian noise at a signal-to-noise ratio equal to 20 dB. The matrices $\mathbf{H}$ obtained for $\gamma_2 = 0$ and $\gamma_2 = 10^{-3}$, respectively, are presented in Figures 7 and 8.

We clearly observe that the sparsity-enforcing function allowed us to recover, in most cases, the endmembers involved in each observed spectrum. On the contrary, when no sparsity penalty term was used, all the entries of the estimated matrix $\mathbf{H}$ were nonzero. Finally, we checked that normalization of the columns of the matrix $\mathbf{W}$, as well as the flux conservation between $\mathbf{V}$ and $\mathbf{H}$, were satisfied at each iteration in both cases. On Figure 9, the behaviour of s_j is plotted as a function of the number of iterations, one see clearly that s_j tends to 1, whatever j after a small number of iterations.

[2] Using (4.2), note that $[\mathbf{H}]_{\bullet j}\|_1^2$ remains constant along iterations.

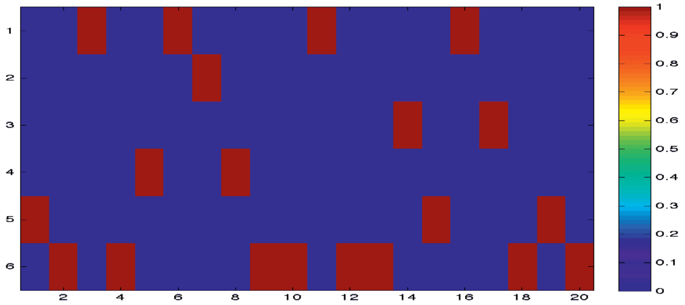

Fig. 6. True **H** with a sparsity $s_j = 1$, $\alpha = 1$.

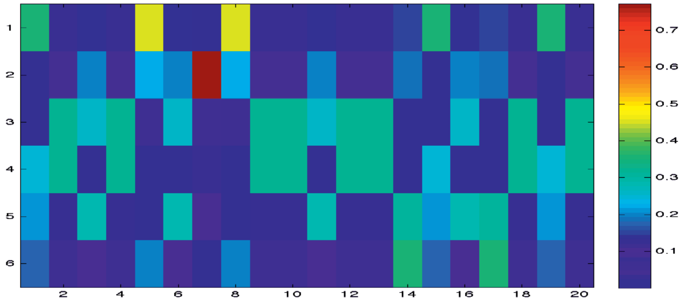

Fig. 7. Estimated **H** without sparsity constraint, $\mu = 0$.

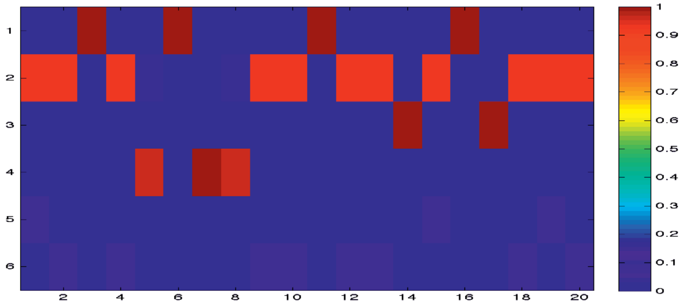

Fig. 8. Estimated **H** with a sparsity constraint, $\mu = 0.001$.

9 Conclusion

In this paper, we proposed a (split) gradient-descent method to solve the nonnegative matrix factorization problem subject to flux conservation constraints on each column of the estimated matrices. Tikhonov regularization and sparsity-enforcing

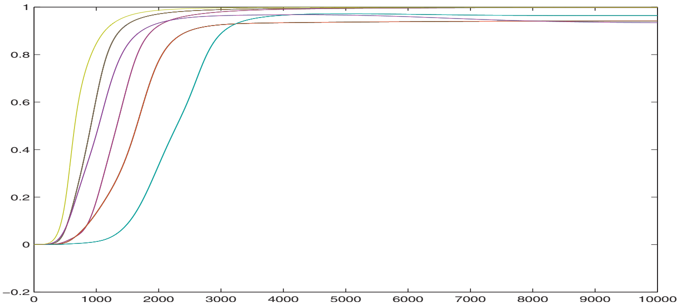

Fig. 9. s_j as a function of the number of iterations, $\mu = 0.001$.

regularization have been also considered. Application in the context of hyperspectral data unmixing shows the effectiveness and the interest of the proposed algorithms.

References

Chang, C.I., 2003, Hyperspectral Imaging: Techniques for Spectral Detection and Classification (New York: Plenum Publishing Co.)

Cichocki, A., Zdunek, R., & Amari, S., 2006, Csiszár's Divergences for Non-negative Matrix Factorization: Family of New Algorithms, Ser. Lectures Notes in Computer Science (Springer Berlin/ Heidelberg), Vol. 3889

Daube-Witherspoon, M.E., & Muehllehner, G., 1986, IEEE Trans. Medical Imaging, 5, 61

Dempster, A.D., Laird, N.M., & Rubin, D.B., 1977, J. R. Stat. Soc., B 39, 1

Desidera, G., Anconelli, B., Bertero, M., Boccacci, P., & Carbillet, M., 2006, "Application of iterative blind deconvolution to the reconstruction of lbt linc-nirvana images", A&A

Févotte, C., Bertin, N., & Durrieu, J.-L., 2009, "Nonnegative matrix factorization with the itakura-saito divergence, with application to music analysis", Neural Computation

Hoyer, P.O., 2004, J. Machine Learning, 5, 1457

Landgrebe, D.A., 2003, Signal Theory Methods in Multispectral Remote Sensing (New York: Wiley)

Lantéri, H., Roche, M., Cuevas, O., & Aime, C., 2001, Signal Processing, 54, 945

Lantéri, H., Roche, M., & Aime, C., 2002, Inverse Probl., 18, 1397

Lantéri, H., Theys, C., Benvenuto, F., & Mary, D., 2009, "Méthode algorithmique de minimisation de fonctions d'écart entre champs de données, application à la reconstruction d'images astrophysiques", in GRETSI

Lantéri, H., Aime, C., Beaumont, H., & Gaucherel, P., 1994, "Blind deconvolution using the richardson-lucy algorithm", in European Symposium on Satellite and Remote Sensing

Lantéri, H., Theys, C., & Richard, C., 2011, "Regularized split gradient method for non negative matrix factorization", in ICASSP, Prague

Lantéri, H., Theys, C., & Richard, C., 2011, "Nonnegative matrix factorization with regularization and sparsity-enforcing terms", in CAMSAP, Porto Rico

Lee, D.D., & Seung, H.S., 2001, Adv. NIPS, 13, 556

Lucy, L.B., 1974, AJ, 79, 745

Richardson, W.H., 1972, J. Opt. Soc. Am., 1, 55

RSI (Research Systems Inc.), 2003, ENVI User's guide Version 4.0, Boulder, CO 80301 USA, Sep.

Theys, C., Dobigeon, N., Tourneret, J.-Y., & Lantéri, H., 2009, "Linear unmixing of hyperspectral images using a scaled gradient method", in SSP, Cardiff