Stability analysis of multiplicative update algorithms for non-negative matrix factorization

Roland Badeau, Nancy Bertin, Emmanuel Vincent

STABILITY ANALYSIS OF MULTIPLICATIVE UPDATE ALGORITHMS FOR NON-NEGATIVE MATRIX FACTORIZATION

*Roland Badeau**

Télécom ParisTech; CNRS LTCI
firstname.lastname@telecom-paristech.fr

Nancy Bertin, Emmanuel Vincent

INRIA
firstname.lastname@inria.fr

ABSTRACT

Multiplicative update algorithms have encountered a great success to solve optimization problems with non-negativity constraints, such as the famous non-negative matrix factorization (NMF) and its many variants. However, despite several years of research on the topic, the understanding of their convergence properties is still to be improved. In this paper, we show that Lyapunov's stability theory provides a very enlightening viewpoint on the problem. We prove the stability of supervised NMF and study the more difficult case of unsupervised NMF. Numerical simulations illustrate those theoretical results, and the convergence speed of NMF multiplicative updates is analyzed.

***Index Terms*—** Optimization methods, non-negative matrix factorization, multiplicative update algorithms, stability, Lyapunov methods.

1. INTRODUCTION

Non-negative matrix factorization (NMF) is a popular technique allowing the decomposition of two-dimensional non-negative data as a linear combination of meaningful elements in a dictionary [1]. Given an $F \times T$ data matrix $\boldsymbol{V}$ having non-negative entries, NMF consists in computing a rank-K truncated approximation $\widehat{\boldsymbol{V}}$ of matrix $\boldsymbol{V}$ (with $K < \min(F,T)$) as a product $\widehat{\boldsymbol{V}} = \boldsymbol{W}\boldsymbol{H}$, where both the $F \times K$ matrix $\boldsymbol{W}$ and the $K \times T$ matrix $\boldsymbol{H}$ have non-negative entries. The columns of matrix $\boldsymbol{W}$ form the elements of the dictionary, and the rows of $\boldsymbol{H}$ contain the coefficients of the decomposition. NMF can be considered either as a *supervised*, or as an *unsupervised* learning tool. In the case of supervised learning [2], the dictionary $\boldsymbol{W}$ is estimated from training data in a preprocessing stage, and matrix $\boldsymbol{H}$ only has to be computed given the data in matrix $\boldsymbol{V}$. In the case of unsupervised learning [1], both matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ have to be computed given $\boldsymbol{V}$. Several algorithms have been proposed in order to compute an NMF. The most popular are the multiplicative update algorithms initially proposed by Lee and Seung [3]. These algorithms can be applied both to supervised and to unsupervised NMF.

A curious point is that to the best of our knowledge, the convergence properties of multiplicative update algorithms for unsupervised NMF have not been clearly identified. Indeed, Lee and Seung proved that the objective function decreases at each iteration [3]. However, this proves neither that the limit value of the objective function is a local minimum, nor that the successive iterates converge to a limit point. In other respects, some numerical examples have been presented in [4], where the Karush-Kuhn-Tucker conditions are not fulfilled after a high (but finite) number of iterations, but this does not contradict possible asymptotic convergence to a local minimum. Finally, since multiplicative updates involve ratios, numerical problems can be encountered if the denominator becomes arbitrarily small. In order to circumvent this problem, it is proposed in [5] to add a small positive quantity to the denominator, and it is proved that any accumulation point of the sequence of the iterates computed in this way is a stationary point. However there is no guarantee that such a stationary point is a local minimum, nor that the algorithm converges to this accumulation point.

In this paper, we intend to analyze the convergence properties of NMF multiplicative update algorithms. We apply Lyapunov's first and second methods [6] to find some criteria which guarantee the exponential or asymptotic stability of the local minima of the objective function. This approach is applied to prove the stability of supervised NMF multiplicative updates, and we finally show how Lyapunov's first method provides some interesting insights into the convergence properties of unsupervised NMF multiplicative updates. The theoretical results presented in the paper are confirmed by numerical simulations involving both supervised and unsupervised NMF, and the convergence speed of NMF updates is investigated.

2. MULTIPLICATIVE UPDATE ALGORITHMS AND NMF

Given a matrix $\boldsymbol{V} \in \mathbb{R}_+^{F\times T}$ and an integer $K < \min(F,T)$, NMF consists in computing a reduced-rank approximation of $\boldsymbol{V}$ as a product $\widehat{\boldsymbol{V}} = \boldsymbol{W}\boldsymbol{H}$, where $\boldsymbol{W} \in \mathbb{R}_+^{F\times K}$ and $\boldsymbol{H} \in \mathbb{R}_+^{K\times T}$. This problem can be formalized as the minimization of an objective function

$$D(\boldsymbol{V}|\boldsymbol{W}\boldsymbol{H}) = \sum_{f=1}^{F}\sum_{t=1}^{T} d\left(v_{ft}\left|\sum_{k=1}^{K} w_{fk}h_{kt}\right.\right), \tag{1}$$

where d is a scalar divergence (*i.e.* a function such that $\forall x, y \in \mathbb{R}_+, d(x|y) \geq 0$, and $d(x|y) = 0$ if and only if $y = x$). β-divergences [7] are defined for all $\beta \in \mathbb{R}\backslash\{0,1\}$ as

$$d_\beta(x|y) = \frac{1}{\beta(\beta-1)}\left(x^\beta + (\beta-1)y^\beta - \beta x y^{\beta-1}\right). \tag{2}$$

The Euclidean distance corresponds to $\beta = 2$, and Kullback-Leibler and Itakura-Saito divergences are obtained when $\beta \to 1$ and $\beta \to 0$, respectively. The generalization of Lee and Seung's multiplicative updates to the β-divergence takes the following form [8]:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \otimes \left(\frac{(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})\boldsymbol{H}^T}{(\boldsymbol{W}\boldsymbol{H})^{\beta-1}\boldsymbol{H}^T}\right)^\eta \tag{3}$$

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \left(\frac{\boldsymbol{W}^T(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})}{\boldsymbol{W}^T(\boldsymbol{W}\boldsymbol{H})^{\beta-1}}\right)^\eta \tag{4}$$

*The research leading to this paper was supported by the French GIP ANR under contract ANR-06-JCJC-0027-01, *Décompositions en Éléments Sonores et Applications Musicales* - DESAM, and by the Quaero Programme, funded by OSEO, French State agency for innovation.

where $\eta = 1$, the symbol $\otimes$ and the fraction bar denote entrywise matrix product and division respectively, and the exponentiations must also be understood entrywise. In the case of unsupervised NMF, updates (3) and (4) are computed alternately, whereas in the case of supervised NMF, the update (4) only is computed at each iteration, matrix $\boldsymbol{W}$ being kept unchanged. We focus in this paper on the generalization of this approach to an exponent step size $\eta > 0$ (possibly different from 1) and we analyze the convergence properties of the multiplicative updates (3) and (4) in section 4. As will be shown in section 5, η actually permits to control the convergence rate, and in particular to outperform the standard case $\eta = 1$.

In [8], it was proved that if $\eta = 1$ and $\beta \in [1, 2]$, then the objective function is non-increasing at each iteration of (3) and (4). The following proposition proves that (3) and (4) actually satisfy the same decrease property for all $\eta \in]0, 1]$ (*i.e.* $0 < \eta \leq 1$).

Proposition 1. *Consider the objective function $D(\boldsymbol{V}|\boldsymbol{W}\boldsymbol{H})$ defined in equation* (1)*, involving the β-divergence* (2)*, with $\beta \in [1, 2]$. If $\eta \in]0, 1]$, if all entries in the numerator and denominator in recursion* (3) *are non-zero and if $(\boldsymbol{W}, \boldsymbol{H})$ is not a fixed point of* (3)*, then* (3) *makes the objective function strictly decrease. Similarly, if $\eta \in]0, 1]$, if all entries in the numerator and denominator in recursion* (4) *are non-zero and if $(\boldsymbol{W}, \boldsymbol{H})$ is not a fixed point of* (4)*, then* (4) *makes the objective function strictly decrease.*

Proposition 1 is proved in [9]. Note that this property does not guarantee that the limit value of the criterion is a local minimum, nor that the successive values of $\boldsymbol{W}$ and $\boldsymbol{H}$ converge to a limit point.

3. STABILITY DEFINITIONS

Let us recall a few definitions in Lyapunov's stability theory of discrete dynamical systems [6], which aim to characterize the convergence properties of the general recursion $\boldsymbol{x}^{(p+1)} = \boldsymbol{\phi}(\boldsymbol{x}^{(p)})$ (where p denotes the iteration index, $\boldsymbol{x}^{(p)}$ is the iterate at iteration p, and function $\boldsymbol{\phi}$ is called a *mapping*), in a neighborhood of a fixed point $\boldsymbol{x}$ (such that $\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{x}$). Notation $\|.\|$ denotes any vector norm.

Definition 1 (Lyapunov stability)**.** A fixed point $\boldsymbol{x} \in \mathbb{R}_+^n$ of the recursion $\boldsymbol{x}^{(p+1)} = \boldsymbol{\phi}(\boldsymbol{x}^{(p)})$, where mapping $\boldsymbol{\phi} : \mathbb{R}_+^n \to \mathbb{R}_+^n$ is continuous in a neighborhood of $\boldsymbol{x}$, is said to be *Lyapunov stable* if $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $\forall \boldsymbol{x}^{(0)} \in \mathbb{R}_+^n$, $\|\boldsymbol{x}^{(0)} - \boldsymbol{x}\| < \delta \Rightarrow \|\boldsymbol{x}^{(p)} - \boldsymbol{x}\| < \varepsilon \; \forall p \in \mathbb{N}$.

This property means that initializing the recursion close enough to $\boldsymbol{x}$ guarantees that the subsequent iterates remain in a bounded domain around $\boldsymbol{x}$. However, it does not guarantee local convergence. A fixed point which is not Lyapunov stable is called *unstable*.

Definition 2 (Asymptotic stability)**.** A fixed point $\boldsymbol{x} \in \mathbb{R}_+^n$ of the recursion $\boldsymbol{x}^{(p+1)} = \boldsymbol{\phi}(\boldsymbol{x}^{(p)})$, where mapping $\boldsymbol{\phi} : \mathbb{R}_+^n \to \mathbb{R}_+^n$ is continuous in a neighborhood of $\boldsymbol{x}$, is said to be *asymptotically stable* if it is Lyapunov stable and there exists $\delta > 0$ such that $\forall \boldsymbol{x}^{(0)} \in \mathbb{R}_+^n$, $\|\boldsymbol{x}^{(0)} - \boldsymbol{x}\| < \delta \Rightarrow \boldsymbol{x}^{(p)} \underset{p \to +\infty}{\longrightarrow} \boldsymbol{x}$.

This property means that initializing the recursion close enough to $\boldsymbol{x}$ guarantees the convergence to $\boldsymbol{x}$. A fixed point which is Lyapunov stable but not asymptotically stable is *marginally stable*.

Definition 3 (Exponential stability)**.** A fixed point $\boldsymbol{x} \in \mathbb{R}_+^n$ of the recursion $\boldsymbol{x}^{(p+1)} = \boldsymbol{\phi}(\boldsymbol{x}^{(p)})$, where mapping $\boldsymbol{\phi} : \mathbb{R}_+^n \to \mathbb{R}_+^n$ is continuous in a neighborhood of $\boldsymbol{x}$, is said to be *exponentially stable* if there exists $\delta, \alpha > 0$ and $\rho \in]0, 1[$ such that $\forall \boldsymbol{x}^{(0)} \in \mathbb{R}_+^n$,

$$\|\boldsymbol{x}^{(0)} - \boldsymbol{x}\| < \delta \Rightarrow \|\boldsymbol{x}^{(p)} - \boldsymbol{x}\| \leq \alpha \|\boldsymbol{x}^{(0)} - \boldsymbol{x}\| \, \rho^p \; \forall p \in \mathbb{N}. \quad (5)$$

In this case, the minimum value of ρ such that equation (5) stands is called the *rate of convergence* of the recursion.

This property ensures a *linear* speed of convergence; it also implies asymptotic stability. A fixed point which is asymptotically stable, but not exponentially stable, is generally characterized by a *sublinear* speed of convergence (depending on the initialization). Note that all the stability properties defined above are *local*, which means that those properties hold in a neighborhood of the fixed point $\boldsymbol{x}$.

4. STABILITY ANALYSIS OF NMF ALGORITHMS

In this section, we show how Lyapunov's stability theory applies to the particular problem of NMF introduced in section 2.

4.1. Supervised NMF

In supervised NMF, matrix $\boldsymbol{W}$ is kept unchanged and matrix $\boldsymbol{H}$ only is updated by the multiplicative update (4). Below, we remap the entries of $\boldsymbol{W}$ and $\boldsymbol{H}$ into vectors $\boldsymbol{w}$ and $\boldsymbol{h}$ of dimensions FK and KT, respectively. The following propositions are proved in [9, 10].

Proposition 2. *Let $\boldsymbol{w} \in \mathbb{R}_+^{FK}$. Let $\boldsymbol{h} \in \mathbb{R}_+^{KT}$ be a local minimum of the objective function $\boldsymbol{h} \mapsto D(\boldsymbol{w}, \boldsymbol{h})$ defined in equation* (1)*. Let $[\boldsymbol{\nabla}_h^2 D(\boldsymbol{w}, \boldsymbol{h})]_+$ be the matrix extracted from the Hessian matrix of D w.r.t. $\boldsymbol{h}$, by selecting the rows and columns whose index i is such that $\frac{\partial D}{\partial h_i}(\boldsymbol{w}, \boldsymbol{h}) = 0$. Finally, let $\boldsymbol{h} \mapsto \boldsymbol{\phi}^h(\boldsymbol{w}, \boldsymbol{h})$ denote the mapping defined by equation* (4)*, and $\nabla {\boldsymbol{\phi}^h}^T(\boldsymbol{w}, \boldsymbol{h})$ denote its Jacobian matrix*[1] *w.r.t. $\boldsymbol{h}$. Then $\forall \beta \in \mathbb{R}$, $\exists \eta^\star \in]0, 2]$ such that*

- *If $\eta \in]0, \eta^\star[$ and if matrix $[\boldsymbol{\nabla}_h^2 D(\boldsymbol{w}, \boldsymbol{h})]_+$ is positive definite, then $\boldsymbol{h}$ is an asymptotically stable fixed point of mapping $\boldsymbol{\phi}^h$.*
- *If $\eta = 0$, then $\boldsymbol{h}$ is a marginally stable fixed point.*
- *If $\eta \notin [0, \eta^\star]$, then $\boldsymbol{h}$ is an unstable fixed point.*
- *$\boldsymbol{h}$ is exponentially stable if and only if $\eta \in]0, \eta^\star[$, matrix $[\boldsymbol{\nabla}_h^2 D(\boldsymbol{w}, \boldsymbol{h})]_+$ is positive definite, and $\forall i$ such that $h_i = 0$, $\frac{\partial D}{\partial h_i}(\boldsymbol{w}, \boldsymbol{h}) > 0$. In this case, the rate of convergence is equal to the spectral radius $\rho^\star = \rho\left(\boldsymbol{\nabla} {\boldsymbol{\phi}^h}^T(\boldsymbol{w}, \boldsymbol{h})\right) < 1$.*

If moreover $\beta \in [1, 2]$, then $\eta^\star = 2$.

Proposition 2 proves the exponential or the asymptotic stability of the local minima of (1) with respect to $\boldsymbol{H}$ under mild conditions. The following proposition proves that conversely, exponentially stable fixed points of recursion (4) are local minima of (1).

Proposition 3. *Let $\boldsymbol{w} \in \mathbb{R}_+^{FK}$. For all $\beta \in \mathbb{R}$, if $\eta > 0$ and if $\boldsymbol{h}$ is an exponentially stable fixed point of mapping $\boldsymbol{\phi}^h$, then $\boldsymbol{h}$ is a local minimum of function $\boldsymbol{h} \mapsto D(\boldsymbol{w}, \boldsymbol{h})$.*

4.2. Unsupervised NMF

Analyzing the stability of the algorithm which alternates multiplicative updates (3) and (4) happens to be particularly difficult. Indeed, it is well known that unsupervised NMF admits several invariances: the product $\boldsymbol{W}\boldsymbol{H}$ is unchanged by replacing matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ by

[1] $\forall 1 \leq i, j \leq KT$, the $(i, j)^{\text{th}}$ coefficient of $\nabla {\boldsymbol{\phi}^h}^T(\boldsymbol{w}, \boldsymbol{h})$ is $\frac{\partial \phi_j^h}{\partial h_i}$.

the non-negative matrices $\boldsymbol{W}' = \boldsymbol{W}\boldsymbol{D}$ and $\boldsymbol{H}' = \boldsymbol{D}^{-1}\boldsymbol{H}$, where $\boldsymbol{D}$ is any diagonal matrix with positive entries. Consequently, the *local minima of the objective function are never isolated* (any local minimum is reached on a *continuum* of matrices $\boldsymbol{W}'$ and $\boldsymbol{H}'$ whose product is equal to $\boldsymbol{W}\boldsymbol{H}$). For this reason, the local minima of the objective function can never be exponentially stable. Below, we remap the entries of $\boldsymbol{w}$ and $\boldsymbol{h}$ into a vector $\boldsymbol{x}$, of dimension $FK + KT$.

Proposition 4. *Let $\boldsymbol{x}$ be a local minimum of the objective function D defined in equation* (1)*. Let $\boldsymbol{\phi}^x$ denote the mapping defined by the composition of* (3) *and* (4)*, and $\nabla\boldsymbol{\phi}^{xT}(\boldsymbol{x})$ denote its Jacobian matrix w.r.t. $\boldsymbol{x}$. Then $\forall\beta \in \mathbb{R}$, $\exists\eta^\star \in]0, 2]$ such that*

- *If $\eta \in]0, 2[$, $\rho\left(\nabla\boldsymbol{\phi}^{xT}(\boldsymbol{x})\right) \geq 1$. If moreover $\eta \in]0, \eta^\star[$, $\rho\left(\nabla\boldsymbol{\phi}^{xT}(\boldsymbol{x})\right) = 1$.*
- *If $\eta = 0$, then $\boldsymbol{x}$ is a marginally stable fixed point.*
- *If $\eta \notin [0, 2]$, then $\boldsymbol{x}$ is an unstable fixed point.*
- *$\forall\eta \in \mathbb{R}$, $\boldsymbol{x}$ is not an exponentially stable fixed point.*

If moreover $\beta \in [1, 2]$, then $\eta^\star = 2$.

This proposition is proved in [9, 10]. If $\eta \in]0, 2[$, Proposition 4 does not permit to conclude about the possible stability of the local minimum. This is because 1 is always an eigenvalue of the Jacobian matrix $\nabla\boldsymbol{\phi}^{xT}(\boldsymbol{x})$, which is generally multiple (we suppose that its multiplicity accounts for the invariances of the factorization).

5. SIMULATION RESULTS

In this section we propose some numerical simulations which illustrate the theoretical results presented in section 4.

5.1. Supervised NMF

First, we study the stability of the multiplicative update (4) applied to matrix $\boldsymbol{H}$, while keeping matrix $\boldsymbol{W}$ unchanged.

5.1.1. Example of sub-linear convergence speed

In this first experiment, the dimensions are $F = 3, T = 3$ and $K = 2$. The multiplicative update (4) is applied to the Kullback-Leibler divergence ($\beta = 1$) with a step size $\eta = 1$ (which corresponds to the standard multiplicative update). Let

$$\boldsymbol{V} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix} \text{ and } \boldsymbol{W} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}. \tag{6}$$

It can be noticed that $\boldsymbol{V}$ is singular, and that it can be exactly factorized as the product $\boldsymbol{V} = \boldsymbol{W}\boldsymbol{H}$, where

$$\boldsymbol{H} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}. \tag{7}$$

Thus we know that the lowest value of $D(\boldsymbol{V}|\boldsymbol{W}\boldsymbol{H})$ w.r.t. $\boldsymbol{H}$ is 0, and that this global minimum is reached for $\boldsymbol{H}$ in (7). This example was chosen so that matrix $[\boldsymbol{\nabla}_h^2 D(\boldsymbol{w}, \boldsymbol{h})]_+$ is positive definite, but $\exists i/h_i = 0$ and $\frac{\partial D}{\partial h_i}(\boldsymbol{w}, \boldsymbol{h}) = 0$. Thus according to Proposition 2, this global minimum is an asymptotically stable, but not exponentially stable fixed point. Therefore the speed of convergence of the multiplicative update (4) may be sub-linear. Fig. 1 shows the results obtained by initializing (4) with a matrix $\boldsymbol{H}$ having all coefficients equal to 2. As can be noticed in Fig. 1-(a), the objective function D monotonically converges to 0 (its global decrease was proven in Proposition 1). Besides, Fig. 1-(b) represents the sequence $\frac{1}{\|\boldsymbol{H}^{(p)}-\boldsymbol{H}\|_F} - \frac{1}{\|\boldsymbol{H}^{(p+1)}-\boldsymbol{H}\|_F}$ (where $\|.\|_F$ denotes the Frobenius norm, $\boldsymbol{H}^{(p)}$ is the matrix computed at iteration p and $\boldsymbol{H}$ is the matrix defined in equation (7)) as a solid blue line. This sequence converges to a finite negative value (represented by the dashed red line), which shows that $\|\boldsymbol{H}^{(p)} - \boldsymbol{H}\|_F = O(1/p)$. As predicted by the theoretical analysis, the convergence speed happens to be sub-linear (at least for the proposed initialization).

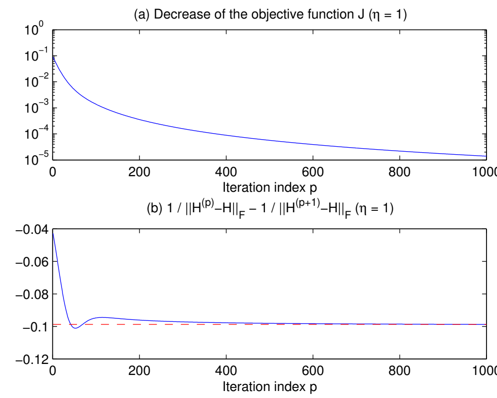

Fig. 1. Example of sub-linear convergence speed

5.1.2. Example of linear convergence speed

In this second experiment, all variables are defined as in section 5.1.1, except that the top left coefficient of $\boldsymbol{V}$ is replaced by 0.9. Consequently, this matrix is no longer singular, thus the global minimum of the objective function w.r.t. $\boldsymbol{H}$ cannot be zero. Instead, a local (possibly global) minimum w.r.t. $\boldsymbol{H}$ can be computed by means of multiplicative update (4), initialized as in section 5.1.1. Numerically, we observed that $[\boldsymbol{\nabla}_h^2 D(\boldsymbol{w}, \boldsymbol{h})]_+$ is positive definite and $\forall i/h_i = 0$, $\frac{\partial D}{\partial h_i}(\boldsymbol{w}, \boldsymbol{h}) > 0$, thus Proposition 2 now proves the exponential stability of this local minimum, with a convergence rate equal to the spectral radius $\rho^\star$. Fig. 2-(a) shows that the objective function D is monotonically decreasing. Besides, Fig. 2-(b) represents the sequence $\frac{\|\boldsymbol{H}^{(p+1)}-\boldsymbol{H}\|_F}{\|\boldsymbol{H}^{(p)}-\boldsymbol{H}\|_F}$ as a solid blue line, and the value $\rho^\star$ as a dashed red line. It can be noticed that this sequence converges to $\rho^\star$, which shows that $\|\boldsymbol{H}^{(p)} - \boldsymbol{H}\|_F = O\left(\rho^{\star p}\right)$. As predicted by the theoretical analysis, the convergence speed is linear, with a convergence rate equal to $\rho^\star$.

5.1.3. Optimal step size

In this third experiment, all variables are defined as in section 5.1.2, and we are looking for an optimal step size η. Since $\beta = 1$, Proposition 2 proves that the local minimum is exponentially stable if and only if $0 < \eta < 2$. In Fig. 2-(c), the solid red line presents the spectral radius $\rho^\star$ as a function of η, for all $\eta \in] - 0.1, 2.1[$. This simulation result confirms that $\rho^\star < 1$ if and only if $0 < \eta < 2$, and it shows that there is an optimal value of parameter η, for which the rate of convergence is optimal. In particular, the standard step size $\eta = 1$ is not optimal. Additional experiments (not displayed in the Figure) showed that the objective function D diverges for a value of eta outside the range $[0, 2]$.

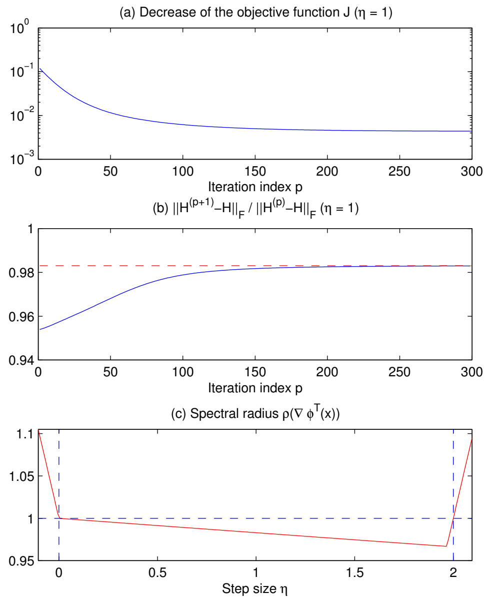

Fig. 2. Example of linear convergence speed

5.2. Unsupervised NMF

We now study the case of unsupervised NMF, which alternates multiplicative updates (3) and (4) for $\boldsymbol{W}$ and $\boldsymbol{H}$. In this fourth experiment, all variables are defined as in section 5.1.2, and the algorithm is initialized in the same way. Fig. 3-(a) shows that the objective function D is monotonically decreasing (to a non-zero value). As in section 5.1.3, the solid red line in Fig. 3-(b) represents the spectral radius $\rho\left(\boldsymbol{\nabla}\boldsymbol{\phi}^{x\,T}(\boldsymbol{x})\right)$ as a function of the step size η, for all $\eta \in]-0.1, 2.1[$. We note that $\rho\left(\boldsymbol{\nabla}\boldsymbol{\phi}^{x\,T}(\boldsymbol{x})\right) > 1$ if $\eta \notin [0, 2]$, and $\rho\left(\boldsymbol{\nabla}\boldsymbol{\phi}^{x\,T}(\boldsymbol{x})\right) = 1$ in the range $\eta =]0, 2[$, which confirms that the local minimum is not exponentially stable. Finally, the solid blue line represents the maximum among the magnitudes of the eigenvalues of matrix $\boldsymbol{\nabla}\boldsymbol{\phi}^{x\,T}(\boldsymbol{x})$ *which are different from 1* [2]. This suggests an optimal value $\eta \approx 1.875$, which is again different from the standard step size $\eta = 1$. Indeed it can be verified that the lowest value of the objective function D (after 100 iterations) is reached when the algorithm is run with this optimal value of η.

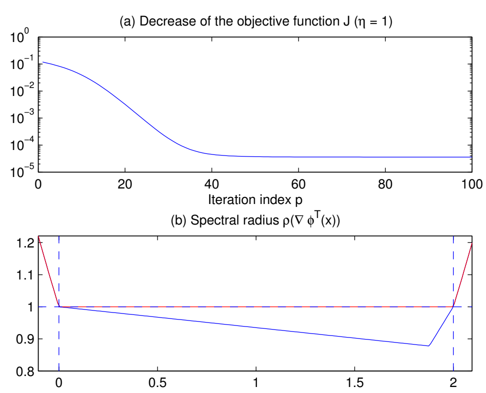

Fig. 3. Unsupervised NMF

[2]This maximum value discards the eigenvalues equal to 1, which are due to the invariances of the factorization.

6. CONCLUSIONS

In this paper, we analyzed the convergence properties of NMF multiplicative update algorithms based on β-divergences, where we introduced an exponent step size η. In the case of supervised NMF, we have presented some criteria which guarantee the exponential or asymptotic stability of the multiplicative updates for any $\eta \in]0, \eta^\star[$, where $\forall \beta \in \mathbb{R}$, $\eta^\star \in]0, 2]$, and if $\beta \in [1, 2]$, $\eta^\star = 2$. We then studied the more complex case of unsupervised NMF. In particular, we proved the unstability of the multiplicative updates if $\eta \notin [0, 2]$. Finally, the theoretical results presented in the paper were confirmed by numerical simulations involving both supervised and unsupervised NMF. Those simulations showed that the convergence rate depends on the value of η, and that there exists an optimal value of η which provides the fastest convergence rate.

An algorithmic outlook of this work would be the design of multiplicative update algorithms with an optimal or an adaptive exponent step size. The proposed stability analysis could also be extended to the hybrid case of constrained unsupervised NMF (as suggested in [10]), to non-negative tensor factorization, or to general multiplicative update algorithms, designed for any optimization problem with non-negativity constraints [9, 10]. Lyapunov's stability theory could also be used to study the stability of other alternating minimization procedures for NMF, such as that presented in [11].

7. REFERENCES

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.

[2] E. Vincent, "Musical source separation using time-frequency source priors," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 1, pp. 91–98, Jan. 2006.

[3] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. of the Conference on Advances in Neural Information Processing Systems*, vol. 13. Vancouver, British Columbia, Canada: MIT Press, Dec. 2001, pp. 556–562.

[4] E. F. Gonzalez and Y. Zhang, "Accelerating the Lee-Seung algorithm for nonnegative matrix factorization," TR-05-02, Rice University, Houston, Texas, USA, Tech. Rep., Mar. 2005.

[5] C.-J. Lin, "On the convergence of multiplicative update algorithms for nonnegative matrix factorization," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1589–1596, Nov. 2007.

[6] J. P. Lasalle, *The stability and control of discrete processes*. New York, NY, USA: Springer-Verlag, 1986.

[7] S. Eguchi and Y. Kano, "Robustifying maximum likelihood estimation," Tokyo Institute of Statistical Mathematics, Tokyo, Japan, Tech. Rep., 2001. [Online]. Available: http://www.ism.ac.jp/~eguchi/pdf/Robustify_MLE.pdf

[8] R. Kompass, "A generalized divergence measure for nonnegative matrix factorization," *Neural Computation*, vol. 19, no. 3, pp. 780–791, Mar. 2007.

[9] R. Badeau, N. Bertin, and E. Vincent, "Stability analysis of multiplicative update algorithms and application to non-negative matrix factorization," *IEEE Trans. Neural Netw.*, vol. 21, no. 12, pp. 1869–1881, Dec. 2010.

[10] ——, "Supporting document for the paper "stability analysis of multiplicative update algorithms and application to non-negative matrix factorization"," Télécom ParisTech, Paris, France, Tech. Rep. 2010D019, Sep. 2010. [Online]. Available: http://service.tsi.enst.fr/cgi-bin/valipub_download.cgi?dId=207

[11] L. Finesso and P. Spreij, "Approximate nonnegative matrix factorization via alternating minimization," in *Proc. of the 16th International Symposium on Mathematical Theory of Networks and Systems*, Leuven, Belgium, Jul. 2004.