Stability analysis of multiplicative update algorithms and application to non-negative matrix factorization

Analyse de la stabilité des règles de mises à jour multiplicatives et application à la factorisation en matrices positives

---

Roland Badeau
Nancy Bertin
Emmanuel Vincent

2010D018

Septembre 2010

Département Traitement du Signal et des Images
Groupe AAO : Audio, Acoustique et Ondes

Stability analysis of multiplicative update algorithms and application to non-negative matrix factorization

Analyse de la stabilité des règles de mises à jour multiplicatives et application à la factorisation en matrices positives

Roland Badeau [(1)], Nancy Bertin [(2)], and Emmanuel Vincent [(2)]

(1) Institut Télécom; Télécom ParisTech; CNRS LTCI, 75634 Paris Cedex 13, France

(2) INRIA, Centre Inria Rennes - Bretagne Atlantique, 35042 Rennes Cedex, France

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 exponential or asymptotic stability of the solutions to general optimization problems with non-negative constraints, including the particular case of supervised NMF, and finally study the more difficult case of unsupervised NMF. The theoretical results presented in the paper are confirmed by numerical simulations involving both supervised and unsupervised NMF, and the convergence speed of NMF multiplicative updates is investigated.

Index Terms

Optimization methods, non-negative matrix factorization, multiplicative update algorithms, convergence of numerical methods, stability, Lyapunov methods.

Résumé

Les règles de mises à jour multiplicatives ont connu un grand succès pour résoudre des problèmes d'optimisation avec contraintes de positivité, tels que la célèbre factorisation en matrices positives (NMF) et ses nombreuses variantes. Néanmoins, malgré plusieurs années de recherche sur le sujet, la compréhension de leurs propriétés de convergence demeure imparfaite. Dans cet article, nous prouvons que la théorie de la stabilité de Lyapunov fournit un point de vue très instructif sur le problème. Nous prouvons la stabilité exponentielle ou asymptotique des solutions de problèmes généraux d'optimisation avec contraintes de positivité, incluant le cas particulier de la NMF supervisée, et nous étudions finalement le cas difficile de la NMF non supervisée. Les résultats théoriques présentés dans cet article sont validés par des simulations numériques mettant en oeuvre les deux types de NMF, et la vitesse de convergence des mises à jour multiplicatives est examinée.

Mots clés

Méthodes d'optimisation, factorisation en matrices positives, algorithmes de mises à jour multiplicatives, convergence des méthodes numériques, stabilité, méthodes de Lyapunov.

The research leading to this document was supported by the French GIP ANR under contract ANR-06-JCJC-0027-01, Décompositions en Éléments Sonores et Applications Musicales - DESAM.

I. INTRODUCTION

NON-NEGATIVE matrix factorization (NMF) is a powerful decomposition technique allowing the decomposition of two-dimensional non-negative data as a linear combination of meaningful elements in a dictionary [24]. Given an $F \times T$ data matrix $\boldsymbol{V}$ having non-negative coefficients, 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 coefficients. The columns of matrix $\boldsymbol{W}$ form the elements of the dictionary, and the rows of $\boldsymbol{H}$ contain the coefficients of the decomposition. The computation of this factorization is generally formalized as a constrained optimization problem. The objective functions most often encountered in NMF literature rely on the Euclidean (EUC) distance, the Kullback-Leibler (KL) divergence [24], [25], or the Itakura-Saito (IS) divergence [19]. The three of them are enclosed in the general framework of β-divergences [10], [14].

NMF can be considered either as a *supervised*, or as an *unsupervised* learning tool. In the case of supervised learning [12], [18], [30], [32], 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 [24], 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 is the multiplicative update algorithm initially proposed by Lee and Seung [25] for the EUC and KL divergences, which has then been generalized to the β-divergence [10], [20]. This algorithm can be applied both to supervised and to unsupervised NMF. For the interested reader, other approaches have also been proposed, such as the projected gradient method [28], alternating least squares (ALS) algorithms [16], the quasi-Newton algorithm [11], a multilayer technique [9], or a space-alternating expectation-maximization (SAGE) algorithm derived in a statistical framework [15]. See for instance [11] for a recent survey on the topic. Theoretical and numerical comparisons between Lee and Sung's multiplicative updates and other algorithms are already available in the literature, see *e.g.* [2], [8], [9], [11], [17], [28].

After Lee and Seung's paper, the multiplicative update philosophy has been applied to various optimization problems involving non-negativity constraints, such as some variants of the NMF. These algorithms generally aim at enhancing (or enforcing) a particular property in the decomposition, depending on the application. In the context of image representation and recognition for instance, various properties have been investigated, such as orthogonality [7], spatial localization [26], or transformation-invariance [13]. In the context of multipitch and music transcription, some desired properties are spectral harmonicity [4], [5], [33] and temporal continuity [4], [5]. In source separation, classical constraints include sparseness and temporal continuity [34], decorrelation [36], and shift-invariance [31]. Note that multiplicative updates have also been applied to non-negative tensor factorization *via* unfolding [11].

A curious point is that to the best of our knowledge, despite many years of research and several papers on the topic, the convergence properties of multiplicative update algorithms for unsupervised NMF have not been clearly identified:

- Lee and Seung proved that the objective function based on EUC and KL decreases at each iteration [25] (and the proof was later generalized to β-divergences [20], for $\beta \in [1,2]$). 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 constrained optimization, all the local minima of the objective function are proved to be stationary points, defined as the solutions to Karush, Kuhn and Tucker's (KKT) optimality conditions. Stationary points of NMF were studied in [2], [8], [27]. Some numerical examples have been presented in [17], where the KKT conditions are not fulfilled after a high (but finite) number of iterations, but this does not contradict the possible asymptotic convergence to a local minimum.
- Since the multiplicative updates involve ratios, numerical problems could be encountered if the denominator becomes arbitrarily small. In order to circumvent this problem, it is proposed in [27] 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[1]. However there is no guarantee that such a stationary point is a local minimum, nor that the algorithm converges to this accumulation point.

[1]Note that this proof only stands for the Euclidean distance, and does not apply when the added quantity is zero.

Analyzing the convergence properties of unsupervised NMF multiplicative updates is difficult because at each iteration, these algorithms usually switch between two different updates: one for the left factor $\boldsymbol{W}$, and one for the right factor $\boldsymbol{H}$. Nevertheless, the convergence analysis happens to be simpler in the case of supervised NMF, where only one of the two factors is updated, the other one being kept unchanged throughout the iterations. In this paper, we intend to analyze the convergence of general multiplicative update algorithms, where all variables are updated at once (which includes the particular case of supervised NMF), before studying the case of unsupervised NMF. Two important aspects of these algorithms have to be taken into account:

- *Local convergence:* Since the objective function generally admits several local minima [2], [9], there is no guarantee that the algorithm converges to the global minimum. So the best result we can prove is the local convergence to a local minimum[2]. This means that if the algorithm is initialized in a given neighborhood of a local minimum called *basin of attraction*, then the algorithm will converge to this local minimum.
- *Stability:* Because of the multiplicative form of the algorithm, a zero entry remains zero in all subsequent iterations. Zeroing may happen because of a bad initialization or because of the finite machine precision for instance. This prevents the convergence to a local minimum whose corresponding coefficient would be non-zero. However, this problem can be very easily circumvented. Indeed, in this case, numerical simulations show that the limit point of the algorithm is generally unstable: replacing the zero entry by any arbitrarily small quantity will make the algorithm escape from this trap and finally converge to a stable limit point. Other well-known examples of unstable stationary points of the algorithm (with non-zero entries) are saddle points [2], [8], [16].

These remarks show that an appropriate notion for analyzing the convergence of multiplicative update algorithms is the *asymptotic stability* in the sense of Lyapunov's theory [22], which implies both local and stable convergence. In this paper, we analyze the convergence properties of general multiplicative update algorithms, designed to solve optimization problems with non-negativity constraints. We thus apply Lyapunov's first and second methods to find some criteria which guarantee the exponential or asymptotic stability of the local minima of the objective function. This analysis is then 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 numerical simulations illustrate those theoretical results, and the convergence speed of NMF multiplicative updates is investigated. The paper is organized as follows: in section II, we present some elementary results about NMF, general optimization problems with non-negativity constraints, and multiplicative update algorithms. The convergence of these algorithms is analyzed by means of Lyapunov's stability theory in section III, and the case of non-negative matrix factorization is studied in section IV. Some numerical simulations are presented in section V, and the main conclusions are summarized in section VI. Finally, the mathematical proofs of the main results presented in this paper are included in the Appendix (due to the lack of space, some proofs have been moved to a separate document [1]).

II. Theoretical background

A. Multiplicative update algorithms and NMF

Given a matrix $\boldsymbol{V} \in \mathbb{R}_+^{F\times T}$ and an integer $K < \min(F,T)$, unsupervised 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 [10], [14] are defined for all $\beta \in \mathbb{R}\backslash\{0,1\}$ as

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

[2]Avoiding to be trapped in a local minimum distinct from the global minimum is out of the scope of this paper (the interested reader can have a look at [3], [6]).

The Euclidean distance corresponds to $\beta = 2$, and KL and IS divergences are obtained when $\beta \to 1$ and $\beta \to 0$, respectively: $d_{KL}(x|y) = x \ln\left(\frac{x}{y}\right) - x + y$, and $d_{IS}(x|y) = \frac{x}{y} - \ln\left(\frac{x}{y}\right) - 1$.

The generalization of Lee and Seung's multiplicative updates to the β-divergence takes the following form [10], [20]:

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

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

where 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.

In [20], it is proved that if $\beta \in [1, 2]$, then the objective function is non-increasing at each iteration of (3) and (4). As in [25], this algorithm can be interpreted as a gradient descent with an adaptive step size for each entry, defined as a function of both matrices $\boldsymbol{W}$ and $\boldsymbol{H}$, chosen so that the successive iterates remain non-negative. Alternatively, we consider the approach used in [4], [7], [34]: the recursion for $\boldsymbol{W}$ in (3) can be written $\forall f, k,\ w_{fk} \leftarrow w_{fk} \frac{m^w_{fk}}{p^w_{fk}}$, where $p^w_{fk} \geq 0$ and $m^w_{fk} \geq 0$ are such that the partial derivative of the objective function w.r.t. w_{fk} is equal to $p^w_{fk} - m^w_{fk}$:

$$\frac{\partial D}{\partial w_{fk}} = \underbrace{\sum_{t=1}^{T} \hat{v}_{ft}^{\beta-1} h_{kt}}_{p^w_{fk}} - \underbrace{\sum_{t=1}^{T} v_{ft}\, \hat{v}_{ft}^{\beta-2} h_{kt}}_{m^w_{fk}}, \tag{5}$$

where

$$\hat{v}_{ft} = \sum_{k=1}^{K} w_{fk} h_{kt}. \tag{6}$$

Thus if $\frac{\partial D}{\partial w_{fk}} > 0$, $\frac{m^w_{fk}}{p^w_{fk}} < 1$ so that w_{fk} decreases, and conversely if $\frac{\partial D}{\partial w_{fk}} < 0$, $\frac{m^w_{fk}}{p^w_{fk}} > 1$ so that w_{fk} increases. The same remark stands for matrix $\boldsymbol{H}$. This confirms that the updates (3) and (4) form a descent method.

We focus in this paper on a generalization of this approach which involves an exponent step size $\eta > 0$:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \otimes \left(\frac{(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})\boldsymbol{H}^T}{(\boldsymbol{W}\boldsymbol{H})^{\beta-1}\boldsymbol{H}^T}\right)^{\eta} \tag{7}$$

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \left(\frac{\boldsymbol{W}^T(\boldsymbol{V} \otimes (\boldsymbol{W}\boldsymbol{H})^{\beta-2})}{\boldsymbol{W}^T(\boldsymbol{W}\boldsymbol{H})^{\beta-1}}\right)^{\eta} \tag{8}$$

Note that standard multiplicative updates (3) and (4) correspond to the particular case $\eta = 1$. As will be shown in section V, η actually permits to control the convergence rate, and in particular to outperform the standard case $\eta = 1$. This approach was first introduced in [4].

The convergence properties of these generalized updates will be analyzed in section IV. The following proposition proves that (7) and (8) satisfy the same decrease property as (3) and (4) for all $\eta \in]0, 1]$ (*i.e.* $0 < \eta \leq 1$)[3].

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 coefficients in the numerator and denominator in recursion* (7) *are non-zero and if $(\boldsymbol{W}, \boldsymbol{H})$ is not a fixed point of* (7) *, then* (7) *makes the objective function strictly decrease. Similarly, if $\eta \in]0, 1]$, if all coefficients in the numerator and denominator in recursion* (8) *are non-zero and if $(\boldsymbol{W}, \boldsymbol{H})$ is not a fixed point of* (8) *, then* (8) *makes the objective function strictly decrease.*

Proposition 1 is proved in Appendix A. Note that this decrease property does not guarantee that that the limit value of the objective function is a local minimum, nor that the successive values of $\boldsymbol{W}$ and $\boldsymbol{H}$ converge to a

[3]In this paper we use the ISO notation for intervals, which uses inwards pointing brackets to indicate inclusion of the endpoint, and outwards pointing brackets for exclusion.

limit point. From now on and until section IV, we will introduce a general framework for multiplicative update algorithms, where all variables are updated at once (which includes the particular case of supervised NMF).

B. General optimization problems with non-negativity constraints

We consider the minimization in the first orthant $\mathbb{R}^n_+$ of an objective function $J : \mathbb{R}^n_+ \to \mathbb{R}$, which is twice continuously differentiable in its domain. For any vector $\boldsymbol{x} = [x_1 \dots x_n]^T \in \mathbb{R}^n_+$, the constraint $x_i \geq 0$ is said to be *active* if $x_i = 0$, or *inactive* if $x_i > 0$. The following two propositions are classical results in constrained optimization theory [29].

Proposition 2 (First order KKT optimality conditions)**.** *Let $\boldsymbol{\nabla} J(\boldsymbol{x})$ denote the gradient vector of the objective function $J : \mathbb{R}^n_+ \to \mathbb{R}$, which is twice continuously differentiable in its domain. Then for any local minimum $\boldsymbol{x}$ of J in $\mathbb{R}^n_+$,*

- *$\forall i \in \{1 \dots n\}$ such that $x_i > 0$, $\nabla_i J = 0$;*
- *$\forall i \in \{1 \dots n\}$ such that $x_i = 0$, $\nabla_i J \geq 0$.*

If $x_i = 0$ and $\nabla_i J > 0$, the constraint is said to be *strictly active*. Following these considerations, we introduce the following notation for denoting the extraction of particular sub-vectors or sub-matrices:

Notation 1.

- $[.]_0$ is obtained by selecting the coefficients (of a vector) or the rows and columns (of a matrix) corresponding to strictly active constraints, *i.e.* whose index i is such that $\nabla_i J(\boldsymbol{x}) > 0$ (and $x_i = 0$).
- $[.]_+$ is obtained by selecting the coefficients, or the rows and columns, whose index i is such that $\nabla_i J(\boldsymbol{x}) = 0$ (and either $x_i > 0$ or $x_i = 0$).
- $[.]^\star_+$ is obtained by selecting the coefficients, or the rows and columns, corresponding to inactive constraints, *i.e.* whose index i is such that $x_i > 0$ (and $\nabla_i J(\boldsymbol{x}) = 0$).

We then have the following optimality condition at order 2:

Proposition 3 (Second order optimality condition)**.** *Let $\boldsymbol{\nabla}^2 J(\boldsymbol{x})$ denote the $n \times n$ Hessian matrix of the objective function $J : \mathbb{R}^n_+ \to \mathbb{R}$, which is twice continuously differentiable in its domain. Then for any local minimum $\boldsymbol{x}$ of J in $\mathbb{R}^n_+$, the sub-matrix $[\boldsymbol{\nabla}^2 J(\boldsymbol{x})]_+$ is positive semi-definite.*

C. Multiplicative update algorithms

In order to introduce general multiplicative update algorithms, we first have to assume that function J satisfies certain conditions.

Assumption 1 (Decomposability of the objective function)**.** Let $J : \mathbb{R}^n_+ \to \mathbb{R}$ be an objective function which is twice continuously differentiable in its domain. We assume that the gradient of J can be decomposed as the difference of two non-negative functions:

$$\boldsymbol{\nabla} J(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}) - \boldsymbol{m}(\boldsymbol{x}), \tag{9}$$

where both functions $\boldsymbol{p} : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ and $\boldsymbol{m} : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ are continuously differentiable in the domain of J.

Note that in Assumption 1, functions $\boldsymbol{p}$ and $\boldsymbol{m}$ are not unique. Indeed, any non-negative constant (or continuously differentiable function) can be added to both $\boldsymbol{p}$ and $\boldsymbol{m}$, without changing their difference.

Example. An example of function J satisfying Assumption 1 will be provided in equation (17), at the beginning of section IV. It corresponds to the NMF objective function introduced in equation (1), whose partial derivatives can be written as the difference of two continuously differentiable non-negative functions, defined in equations (5) and (32).

Assumption 2. Let $\boldsymbol{x} \in \mathbb{R}^n_+$. Given an objective function J satisfying Assumption 1, we assume that $\forall i \in \{1 \dots n\}$, $p_i(\boldsymbol{x}) > 0$ and $m_i(\boldsymbol{x}) > 0$.

Note that if Assumption 1 stands, there always exist functions $p_i(\boldsymbol{x})$ and $m_i(\boldsymbol{x})$ such that Assumption 2 also stands $\forall \boldsymbol{x} \in \mathbb{R}_+^n$. Indeed, any positive constant (or continuously differentiable function) can be added to both functions $p_i(\boldsymbol{x})$ and $m_i(\boldsymbol{x})$, without changing their difference.

Definition 1 (Multiplicative mapping)**.** Consider the minimization of an objective function J satisfying Assumption 1. For any step size $\eta \in \mathbb{R}$, the multiplicative mapping $\boldsymbol{\phi}$ is defined in the domain of all $\boldsymbol{x} \in \mathbb{R}_+^n$ satisfying Assumption 2, as

$$\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{\Lambda}(\boldsymbol{x})^\eta \, \boldsymbol{x} \tag{10}$$

where $\boldsymbol{\Lambda}(\boldsymbol{x}) = \operatorname{diag}\left(\frac{\boldsymbol{m}(\boldsymbol{x})}{\boldsymbol{p}(\boldsymbol{x})}\right)$, $\boldsymbol{m}(\boldsymbol{x})$ and $\boldsymbol{p}(\boldsymbol{x})$ have been defined in Assumption 1, and $\operatorname{diag}(.)$ denotes the diagonal matrix whose diagonal coefficients are those of the vector argument.

The proof of the following lemma is straightforward.

Lemma 4 (Regularity of the mapping)**.** *Let J be an objective function satisfying Assumption 1, and let $\boldsymbol{x} \in \mathbb{R}_+^n$ satisfying Assumption 2. Then $\forall \eta \in \mathbb{R}$, the mapping $\boldsymbol{\phi}$ introduced in Definition 1 is defined and continuously differentiable in a neighborhood of $\boldsymbol{x}$.*

The following lemma is a corollary of Proposition 2.

Lemma 5. *Let $\boldsymbol{x} \in \mathbb{R}_n^+$ be a local minimum of a function J satisfying Assumption 1. If Assumption 2 holds, then $\boldsymbol{x}$ is a fixed point of the mapping $\boldsymbol{\phi}$ introduced in Definition 1 (*i.e. $\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{x}$*).*

Proof: Obviously, if $x_i = 0$, $\phi_i(\boldsymbol{x}) = \boldsymbol{x}_i$. Otherwise Proposition 2 proves that $\forall i \in \{1 \dots n\}$ such that $x_i > 0$, $p_i(\boldsymbol{x}) = m_i(\boldsymbol{x})$, thus $\phi_i(\boldsymbol{x}) = \boldsymbol{x}_i$.

■

Definition 2 (Multiplicative update algorithm)**.** Consider the minimization of an objective function J satisfying Assumption 1. A multiplicative update algorithm is defined by a recursion of the form $\boldsymbol{x}^{(p+1)} = \boldsymbol{\phi}(\boldsymbol{x}^{(p)})$, where the mapping $\boldsymbol{\phi} : \mathbb{R}_+^n \to \mathbb{R}_+^n$ was defined in Definition 1.

Note that if there exists p such that $\boldsymbol{x}^{(p)}$ does not satisfy Assumption 2, then $\boldsymbol{x}^{(p+1)}$ may be undefined, and the algorithm must stop[4]. Also note that recursion $\boldsymbol{x}^{(p+1)} = \boldsymbol{\phi}(\boldsymbol{x}^{(p)})$ can be seen as a descent method. Indeed, if $\boldsymbol{x} \in \mathbb{R}_+^n$ satisfies Assumption 2, the first order expansion of function $\eta \mapsto J\left(\boldsymbol{\Lambda}(\boldsymbol{x})^\eta \, \boldsymbol{x}\right)$ in a neighborhood of $\eta = 0$ yields

$$\begin{aligned} &J(\boldsymbol{\phi}(\boldsymbol{x})) - J(\boldsymbol{x}) = \eta \boldsymbol{\nabla} J(\boldsymbol{x})^T \left(\ln\left(\boldsymbol{\Lambda}(\boldsymbol{x})\right) \boldsymbol{\phi}(\boldsymbol{x})\right) + O(\eta^2) \\ &= -\eta \sum_{i=1}^{n} \phi_i(\boldsymbol{x})(m_i(\boldsymbol{x}) - p_i(\boldsymbol{x})) \ln\left(\tfrac{m_i(\boldsymbol{x})}{p_i(\boldsymbol{x})}\right) + O(\eta^2) \end{aligned}$$

This equation shows that if $\boldsymbol{\phi}(\boldsymbol{x}) \neq \boldsymbol{x}$ and if $\eta > 0$ is small enough, then $J(\boldsymbol{\phi}(\boldsymbol{x})) - J(\boldsymbol{x}) < 0$, which means that the objective function decreases.

III. Stability analysis of multiplicative updates

We analyze the convergence of multiplicative update algorithms by means of Lyapunov's stability theory. In neural networks' litterature, this theory has been used for various problems, such as analyzing the global exponential stability of discrete recurrent neural networks with time-varying delays [35], [37] or analyzing the discrete-time dynamics of a class of self-stabilizing minor component analysis (MCA) extraction algorithms [21].

A. Stability definitions

Let us recall a few classical definitions in Lyapunov's stability theory of discrete dynamical systems [22]. Notation $\|.\|$ denotes any vector norm.

Definition 3 (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 $\forall \boldsymbol{x}^{(0)} \in \mathbb{R}_+^n$, $\|\boldsymbol{x}^{(0)} - \boldsymbol{x}\| < \delta \;\Rightarrow \|\boldsymbol{x}^{(p)} - \boldsymbol{x}\| < \varepsilon \;\forall p \in \mathbb{N}$.

[4]However this singular case is never observed in practical NMF problems.

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

Definition 4 (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 sometimes called *marginally stable*.

Definition 5 (Exponential stability and rate of convergence)**.** 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 $\forall \boldsymbol{x}^{(0)} \in \mathbb{R}_+^n$,

$$\|\boldsymbol{x}^{(0)} - \boldsymbol{x}\| < \delta \ \Rightarrow \|\boldsymbol{x}^{(p)} - \boldsymbol{x}\| \leq \alpha \|\boldsymbol{x}^{(0)} - \boldsymbol{x}\| \ \rho^p \ \forall p \in \mathbb{N}. \tag{11}$$

In this case, the minimum value of ρ such that equation (11) 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 *sub-linear* 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 $\boldsymbol{x}$.

B. Lyapunov's first (or indirect) method

Lyapunov's first (or indirect) method permits to characterize the exponential stability and the corresponding convergence rate of a dynamical system. Let us first recall its principle, that we apply in the domain $\mathbb{R}_+^n$.

Theorem 6 (Lyapunov's first stability theorem)**.** *Let $\boldsymbol{x} \in \mathbb{R}_+^n$ be a fixed point of the recursion $\boldsymbol{x}^{(p+1)} = \boldsymbol{\phi}(\boldsymbol{x}^{(p)})$, where mapping $\boldsymbol{\phi} : \mathbb{R}_+^n \to \mathbb{R}_+^n$ is continuously differentiable in a neighborhood of $\boldsymbol{x}$. Let $\nabla \boldsymbol{\phi}^T(\boldsymbol{x})$ be the Jacobian matrix[5] of mapping $\boldsymbol{\phi}$ at point $\boldsymbol{x}$. Then the exponential stability (or unstability) of $\boldsymbol{x}$ is characterized by the eigenvalues of $\nabla \boldsymbol{\phi}^T(\boldsymbol{x})$:*

- *$\boldsymbol{x}$ is an exponentially stable fixed point if and only if all the eigenvalues of $\nabla \boldsymbol{\phi}^T(\boldsymbol{x})$ have a magnitude lower than 1. In this case, the rate of convergence of the recursion is equal to the spectral radius[6] of matrix $\nabla \boldsymbol{\phi}^T(\boldsymbol{x})$, which is denoted $\rho\left(\nabla \boldsymbol{\phi}^T(\boldsymbol{x})\right) < 1$.*
- *If at least one eigenvalue of $\nabla \boldsymbol{\phi}^T(\boldsymbol{x})$ has a magnitude greater than 1, then $\boldsymbol{x}$ is unstable.*

Note that if all eigenvalues of $\nabla \boldsymbol{\phi}^T(\boldsymbol{x})$ have a magnitude lower than or equal to 1, and at least one of them has magnitude 1, then Theorem 6 does not permit to conclude: the fixed point can be Lyapunov stable or unstable. In order to apply Theorem 6 to the mapping $\boldsymbol{\phi}$ defined in equation (10), we characterize the eigenvalues of matrix $\nabla \boldsymbol{\phi}^T(\boldsymbol{x})$ in the following proposition.

Proposition 7. *Let $\boldsymbol{x} \in \mathbb{R}_+^n$ be a local minimum of an objective function J satisfying Assumption 1, and suppose that Assumption 2 holds. Let $\boldsymbol{\phi}$ be the mapping introduced in Definition 1, which is continuously differentiable in a neighborhood of $\boldsymbol{x}$. Moreover, let us define the positive semi-definite matrix*

$$\boldsymbol{P}(\boldsymbol{x}) = \boldsymbol{D}(\boldsymbol{x}) \boldsymbol{\nabla}^2 J(\boldsymbol{x}) \boldsymbol{D}(\boldsymbol{x}) \tag{12}$$

with $\boldsymbol{D}(\boldsymbol{x}) = \operatorname{diag}(\boldsymbol{x}/\boldsymbol{p}(\boldsymbol{x}))^{\frac{1}{2}}$ (where $\boldsymbol{p}(\boldsymbol{x})$, defined in Assumption 1, has no zero entry), and the positive scalar

$$\eta^\star = \frac{2}{\|\boldsymbol{P}(\boldsymbol{x})\|_2}, \tag{13}$$

[5]For $1 \leq i, j \leq n$, the $(i,j)^{\text{th}}$ coefficient of matrix $\nabla \boldsymbol{\phi}^T(\boldsymbol{x})$ is $\frac{\partial \phi_j}{\partial x_i}$.

[6]The spectral radius of a matrix is the maximum among the magnitudes of its eigenvalues.

where $\|.\|_2$ denotes the matrix 2-norm or spectral norm[7] *(if $\|\boldsymbol{P}(\boldsymbol{x})\|_2 = 0$, $\eta^\star$ is infinite).*

Then if $\eta = 0$, all the eigenvalues of the Jacobian matrix $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$ are equal to 1. Otherwise, the multiplicity of $\lambda = 1$ as an eigenvalue of $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$ is equal to the dimension of the kernel of matrix $[\boldsymbol{P}(\boldsymbol{x})]_+$ (with the use of Notation 1). Moreover, the other eigenvalues of $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$ are as follows:

- *If $\eta \in]0, \eta^\star[$, all the other eigenvalues have a magnitude lower than 1;*
- *If $\eta < 0$, all the other eigenvalues are greater than 1;*
- *If $\eta > \eta^\star$, at least one eigenvalue is lower than -1;*
- *If $\eta = \eta^\star$, at least one eigenvalue is equal to -1.*

Proposition 7 is proved in Appendix B. Note that in the case $\eta = 0$, $\boldsymbol{\phi}$ is the identity transform. In other respects, if $[\boldsymbol{P}(\boldsymbol{x})]_+$ is non-singular, 1 is *not* an eigenvalue of $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$. Now we can state the stability properties of mapping (10):

Proposition 8. *Let $\boldsymbol{x} \in \mathbb{R}^n_+$ be a local minimum of an objective function J satisfying Assumption 1, and suppose that Assumption 2 holds. Let $\boldsymbol{\phi}$ be the mapping introduced in Definition 1, which is continuously differentiable in a neighborhood of $\boldsymbol{x}$.*

Then, following the notation introduced in Proposition 7,

- *$\boldsymbol{x}$ is an exponentially stable fixed point if and only if $\eta \in]0, \eta^\star[$ and matrix $[\boldsymbol{P}(\boldsymbol{x})]_+$ is non-singular;*
- *if $\eta \notin [0, \eta^\star]$, $\boldsymbol{x}$ is an unstable fixed point;*
- *if $\eta = 0$, $\boldsymbol{x}$ is a marginally stable fixed point.*

Proof: The first and second assertion are a corollary of Theorem 6 and Proposition 7. The third assertion is trivial, since if $\eta = 0$, mapping $\boldsymbol{\phi}$ is the identity transform. ■

Note that the non-singularity of matrix $[\boldsymbol{P}(\boldsymbol{x})]_+$ is equivalent to the combination of the two following properties:

$$\forall i \text{ such that } x_i = 0,\ \nabla_i J(\boldsymbol{x}) > 0, \tag{14}$$

$$\text{matrix } [\boldsymbol{\nabla}^2 J(\boldsymbol{x})]_+ \text{ is positive definite.} \tag{15}$$

If $\eta = \eta^\star$, or if $\eta \in]0, \eta^\star[$ and matrix $[\boldsymbol{P}(\boldsymbol{x})]_+$ is singular, Lyapunov's first method does not permit to conclude, since there is at least one eigenvalue of magnitude 1.

Finally, the following proposition completes Proposition 8, and proves the equivalence between the exponentially stable fixed points of mapping $\boldsymbol{\phi}$, and the local minima of function J satisfying both properties (14) and (15).

Proposition 9. *Let J be an objective function such that Assumption 1 holds. Let $\boldsymbol{x} \in \mathbb{R}^n_+$ be a vector satisfying Assumption 2. Let $\boldsymbol{\phi}$ be the mapping introduced in Definition 1, which is continuously differentiable in a neighborhood of $\boldsymbol{x}$. Assume that $\eta > 0$ and that $\boldsymbol{x}$ is an exponentially stable fixed point of mapping $\boldsymbol{\phi}$. Then $\boldsymbol{x}$ is a local minimum of function J, which additionally satisfies properties* (14) *and* (15)*.*

Proposition 9 is proved in Appendix B.

C. Lyapunov's second (direct) method

For a fixed point which is not exponentially stable, Lyapunov's second method permits to further investigate its stability properties, and possibly prove its Lyapunov or asymptotic stability. In this section, we will prove the following result, which completes that of Proposition 8: if $\eta \in]0, \eta^\star[$, even if assumption (14) does not stand, assumption (15) alone is sufficient for guaranteeing the asymptotic stability of the dynamical system. Let us first recall the principle of Lyapunov's second method, that we apply in the domain $\mathbb{R}^n_+$.

Definition 6 (Lyapunov function)**.** For any $\boldsymbol{x} \in \mathbb{R}^n_+$, a Lyapunov function $\boldsymbol{y} \mapsto V(\boldsymbol{x}, \boldsymbol{y})$ is a continuous scalar function defined on a neighborhood of $\boldsymbol{x}$ included in $\mathbb{R}^n_+$, which is positive-definite (in the sense that $V(\boldsymbol{x}, \boldsymbol{x}) = 0$, and $V(\boldsymbol{x}, \boldsymbol{y}) > 0$ for all $\boldsymbol{y} \neq \boldsymbol{x}$).

Theorem 10 (Lyapunov's second stability theorem)**.** *Let $\boldsymbol{x} \in \mathbb{R}^n_+$ be a fixed point of a continuous mapping $\boldsymbol{\phi} : \mathbb{R}^n_+ \to \mathbb{R}^n_+$.*

[7]The spectral norm of a matrix is equal to its greatest singular value. In the particular case of Hermitian matrices, the spectral norm is equal to the spectral radius.

- *If there is a Lyapunov function V such that $V(\boldsymbol{x}, \boldsymbol{\phi}(\boldsymbol{y})) \leq V(\boldsymbol{x}, \boldsymbol{y})$ for all $\boldsymbol{y}$ in a neighborhood of $\boldsymbol{x}$, then $\boldsymbol{x}$ is Lyapunov stable.*
- *If there is a Lyapunov function V such that $V(\boldsymbol{x}, \boldsymbol{\phi}(\boldsymbol{y})) < V(\boldsymbol{x}, \boldsymbol{y})$ for all $\boldsymbol{y} \neq \boldsymbol{x}$ in a neighborhood of $\boldsymbol{x}$, then $\boldsymbol{x}$ is asymptotically stable.*

If $\boldsymbol{x}$ is a local minimum of a continuous function J, a natural candidate Lyapunov function for the mapping $\boldsymbol{\phi}$ defined in Definition 1 would be $V(\boldsymbol{x}, \boldsymbol{y}) = J(\boldsymbol{y}) - J(\boldsymbol{x})$. However, this choice raises two problems:

- In some cases, a fixed-point $\boldsymbol{x}$ is Lyapunov-stable, whereas the objective function J is not globally monotonically decreasing[8].
- The condition that V is positive-definite may not be satisfied in a neighborhood of $\boldsymbol{x}$ [9].

For these reasons, we propose an alternative Lyapunov function in the following lemma.

Lemma 11. *Consider an objective function J satisfying Assumption 1, and $\boldsymbol{x} \in \mathbb{R}_+^n$ such that Assumption 2 holds. Then the function*

$$V(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x})^T \operatorname{diag}\left(\frac{\boldsymbol{p}(\boldsymbol{x}) + \boldsymbol{p}(\boldsymbol{y})}{\boldsymbol{x} + \boldsymbol{y}}\right)(\boldsymbol{y} - \boldsymbol{x}) \tag{16}$$

defines a symmetric Lyapunov function on $\mathbb{R}_+^n \times \mathbb{R}_+^n$.

Proof: The definition and continuity of function V on the borders of the domain $\mathbb{R}_+^n$ follow from the inequality $\left|\frac{y_i - x_i}{x_i + y_i}\right| \leq 1 \ \forall x_i, y_i \in \mathbb{R}_+^\star$. ■

We can now state our main result:

Proposition 12. *Let $\boldsymbol{x} \in \mathbb{R}_+^n$ be a local minimum of a function J satisfying Assumption 1, and suppose that Assumption 2 and property* (15) *hold. If $\eta \in]0, \eta^\star[$ (where $\eta^\star$ was defined in equation* (13)*), then the mapping $\boldsymbol{\phi}$ introduced in Definition 1 makes the Lyapunov function $\boldsymbol{y} \mapsto V(\boldsymbol{x}, \boldsymbol{y})$ defined in equation* (16) *strictly decrease in a neighborhood of $\boldsymbol{x}$. As a consequence, $\boldsymbol{x}$ is an asymptotically stable fixed point of mapping $\boldsymbol{\phi}$.*

Proposition 12 is proved in Appendix C. Considering Proposition 8, it can be noticed that if all hypotheses in Proposition 12 stand, but property (14) is not satisfied, then $\boldsymbol{x}$ is an asymptotically stable, but not exponentially stable fixed point of mapping $\boldsymbol{\phi}$. This means that, depending on the initialization, the dynamical system generally has a sub-linear speed of convergence.

IV. Application to NMF

In this section, we show how Lyapunov's stability theory can be applied to the particular problem of NMF with an objective function based on the β-divergence, which was introduced in section II-A. We first focus on the simple case of supervised NMF, which is a direct application of the theory presented in section III, and then study the more complex case of unsupervised NMF. Due to the lack of space, the complete proofs have been moved to a separate document [1]. Nevertheless, the key ideas are provided whenever possible in the following discussion.

Notation 2. In the following, the entries of the $F \times K$ matrix $\boldsymbol{W}$ and the $K \times T$ matrix $\boldsymbol{H}$ are remapped into vectors $\boldsymbol{w}$ and $\boldsymbol{h}$ of dimensions KF and KT, respectively. Vector $\boldsymbol{x}$ is formed by concatenating $\boldsymbol{w}$ and $\boldsymbol{h}$. Then let us define the objective function

$$J(\boldsymbol{x}) = D(\boldsymbol{V} | \boldsymbol{W}\boldsymbol{H}) \tag{17}$$

where function D was defined in equation (1), and involves the β-divergence (2). The gradient of J w.r.t. $\boldsymbol{w}$ is decomposed as the difference of two non-negative functions $\boldsymbol{p}^w(\boldsymbol{x})$ and $\boldsymbol{m}^w(\boldsymbol{x})$ whose coefficients have been defined in equation (5), and similar notation is used for $\boldsymbol{h}$ (see equation (32)). Vector $\boldsymbol{p}(\boldsymbol{x})$ is formed by concatenating $\boldsymbol{p}^w(\boldsymbol{x})$ and $\boldsymbol{p}^h(\boldsymbol{x})$. The Hessian matrices of function J w.r.t. $\boldsymbol{x}$, $\boldsymbol{w}$ and $\boldsymbol{h}$ are denoted $\boldsymbol{\nabla}^2_{xx} J(\boldsymbol{x})$, $\boldsymbol{\nabla}^2_{ww} J(\boldsymbol{x})$ and $\boldsymbol{\nabla}^2_{hh} J(\boldsymbol{x})$, respectively. Let

$$\boldsymbol{\phi}^w(\boldsymbol{w}, \boldsymbol{h}) = \boldsymbol{\Lambda}_w(\boldsymbol{x})^\eta \, \boldsymbol{w} \tag{18}$$

$$\boldsymbol{\phi}^h(\boldsymbol{w}, \boldsymbol{h}) = \boldsymbol{\Lambda}_h(\boldsymbol{x})^\eta \, \boldsymbol{h} \tag{19}$$

[8] This is the case of NMF with $1 < \eta < 2$.

[9] In the case of unsupervised NMF, because of the invariances of the factorization, there is a continuous set of fixed points $\boldsymbol{y}$ satisfying $J(\boldsymbol{y}) = J(\boldsymbol{x})$ (cf. section IV).

where $\boldsymbol{\Lambda}_w(\boldsymbol{x}) = \operatorname{diag}\left(\frac{\boldsymbol{m}^w(\boldsymbol{x})}{\boldsymbol{p}^w(\boldsymbol{x})}\right)$ and $\boldsymbol{\Lambda}_h(\boldsymbol{x}) = \operatorname{diag}\left(\frac{\boldsymbol{m}^h(\boldsymbol{x})}{\boldsymbol{p}^h(\boldsymbol{x})}\right)$.

A. Application to supervised NMF

In the context of supervised NMF, vector $\boldsymbol{w}$ is kept unchanged, and vector $\boldsymbol{h}$ only is updated according to $\boldsymbol{h}^{(p+1)} = \boldsymbol{\phi}^h(\boldsymbol{w}, \boldsymbol{h}^{(p)})$, where mapping $\boldsymbol{\phi}^h$ was defined in equation (19), which is equivalent to the multiplicative update (8). The parameter $\eta_h^\star$ introduced in the following lemma will play the same role as $\eta^\star$ in sections III-B and III-C.

Lemma 13. *Given a constant vector $\boldsymbol{w}$, let $\boldsymbol{h}$ be a local minimum of the NMF objective function $\boldsymbol{h} \mapsto J(\boldsymbol{w}, \boldsymbol{h})$ defined in equation* (17)*. Function $\boldsymbol{h} \mapsto J(\boldsymbol{w}, \boldsymbol{h})$ satisfies Assumption 1, and we assume that $\boldsymbol{h}$ satisfies Assumption 2. Following Notation 2, let us define*

$$\eta_h^\star = \frac{2}{\|\boldsymbol{P}^h(\boldsymbol{x})\|_2}, \tag{20}$$

where $\boldsymbol{P}^h(\boldsymbol{x})$ is the positive semi-definite matrix

$$\boldsymbol{P}^h(\boldsymbol{x}) = \boldsymbol{D}^h(\boldsymbol{x}) \boldsymbol{\nabla}^2_{hh} J(\boldsymbol{x}) \boldsymbol{D}^h(\boldsymbol{x}). \tag{21}$$

with

$$\boldsymbol{D}^h(\boldsymbol{x}) = \operatorname{diag}(\boldsymbol{h}/\boldsymbol{p}^h(\boldsymbol{x}))^{\frac{1}{2}}. \tag{22}$$

Then $\forall \beta \in \mathbb{R}$, we have $0 < \eta_h^\star \leq 2$, and if $\beta \in [1, 2]$, $\eta_h^\star = 2$.

Proof: This lemma is proved by exhibiting an eigenvector of the positive semi-definite matrix $\boldsymbol{P}^h(\boldsymbol{x})$, whose eigenvalue is equal to 1. If, additionally, $\beta \in [1, 2]$, the convexity of function J w.r.t. $\boldsymbol{h}$ permits to prove that all the eigenvalues of $\boldsymbol{P}^h(\boldsymbol{x})$ are lower than or equal to 1. ■

Following Lemma 13, Propositions 8 and 12 directly prove the exponential or the asymptotic stability of the local minima of function $\boldsymbol{h} \mapsto J(\boldsymbol{w}, \boldsymbol{h})$ for all $\eta \in]0, \eta_h^\star[$, under mild conditions (numerical examples are presented in section V-A). Of course, the same result stands for the reciprocal algorithm, where $\boldsymbol{h}$ is kept unchanged, and vector $\boldsymbol{w}$ only is updated according to $\boldsymbol{w}^{(p+1)} = \boldsymbol{\phi}^w(\boldsymbol{w}^{(p)}, \boldsymbol{h})$, where mapping $\boldsymbol{\phi}^w$ was defined in equation (18), which is equivalent to the multiplicative update (7).

Lemma 14. *Given a constant vector $\boldsymbol{h}$, let $\boldsymbol{w}$ be a local minimum of the NMF objective function $\boldsymbol{w} \mapsto J(\boldsymbol{w}, \boldsymbol{h})$ defined in equation* (17)*. Function $\boldsymbol{w} \mapsto J(\boldsymbol{w}, \boldsymbol{h})$ satisfies Assumption 1, and we assume that $\boldsymbol{w}$ satisfies Assumption 2. Following Notation 2, let us define*

$$\eta_w^\star = \frac{2}{\|\boldsymbol{P}^w(\boldsymbol{x})\|_2}, \tag{23}$$

where $\boldsymbol{P}^w(\boldsymbol{x})$ is the positive semi-definite matrix

$$\boldsymbol{P}^w(\boldsymbol{x}) = \boldsymbol{D}^w(\boldsymbol{x}) \boldsymbol{\nabla}^2_{ww} J(\boldsymbol{x}) \boldsymbol{D}^w(\boldsymbol{x}); \tag{24}$$

with

$$\boldsymbol{D}^w(\boldsymbol{x}) = \operatorname{diag}(\boldsymbol{w}/\boldsymbol{p}^w(\boldsymbol{x}))^{\frac{1}{2}}; \tag{25}$$

Then $\forall \beta \in \mathbb{R}$, we have $0 < \eta_w^\star \leq 2$, and if $\beta \in [1, 2]$, $\eta_w^\star = 2$.

This lemma is proved in the same way as lemma 13.

B. Application to unsupervised NMF

Actually, analyzing the stability of the algorithm which alternates multiplicative updates (7) and (8) is particularly difficult for the following reasons:

- It is well known that unsupervised NMF admits several invariances (the problem of the uniqueness of unsupervised NMF has been addressed in [23] for instance). Indeed, the product $\boldsymbol{W}\boldsymbol{H}$ is unchanged by replacing matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ by 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 diagonal coefficients. For this simple reason, 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 $\boldsymbol{WH}$). The consequence is that assumption (15) never stands.
- Anyway, recursion (7)-(8) cannot be implemented with a mapping of the form (10), since it switches between updates for $\boldsymbol{W}$ and $\boldsymbol{H}$.

For these reasons, the results presented in section III cannot be straightforwardly applied to recursion (7)-(8), and the local minima of the objective function can never be exponentially stable (the Jacobian matrix always admits $\lambda = 1$ as an eigenvalue). Thus Lyapunov's first method will not permit to conclude on the stability of multiplicative updates. Nevertheless, this approach still provides an interesting insight into the stability properties of the algorithm. We summarize below the main results that we obtained by applying this approach.

Recursion (7)-(8) is rewritten in the form

$$\left\{ \begin{array}{rcl} \boldsymbol{w}^{(p+1)} & = & \boldsymbol{\phi}^w(\boldsymbol{w}^{(p)}, \boldsymbol{h}^{(p)}) \\ \boldsymbol{h}^{(p+1)} & = & \boldsymbol{\phi}^h(\boldsymbol{w}^{(p+1)}, \boldsymbol{h}^{(p)}) \end{array} \right. \tag{26}$$

where mappings $\boldsymbol{\phi}^w$ and $\boldsymbol{\phi}^h$ were defined in equations (18) and (19). Equivalently, we can write $\boldsymbol{x}^{(p+1)} = \boldsymbol{\phi}(\boldsymbol{x}^{(p)})$, with

$$\boldsymbol{\phi}(\boldsymbol{x}) = \left[\boldsymbol{\phi}^w(\boldsymbol{w}, \boldsymbol{h}); \boldsymbol{\phi}^h(\boldsymbol{\phi}^w(\boldsymbol{w}, \boldsymbol{h}), \boldsymbol{h})\right]. \tag{27}$$

The parameter $\eta_x^\star$ introduced in the following lemma will play the same role as $\eta^\star$ in sections III-B and III-C.

Lemma 15. *Let $\boldsymbol{x}$ be a local minimum of the NMF objective function J defined in equation* (17)*. Function J satisfies Assumption 1, and we assume that $\boldsymbol{x}$ satisfies Assumption 2. Let us define*

$$\eta_x^\star = \min\left(\eta_h^\star, \eta_w^\star\right), \tag{28}$$

where $\eta_h^\star$ and $\eta_w^\star$ were defined in equations (20) *and* (23)*. Then $\forall \beta \in \mathbb{R}$, we have $0 < \eta_x^\star \leq 2$, and if $\beta \in [1, 2]$, $\eta_x^\star = 2$.*

This lemma is a corollary of lemmas 13 and 14. Note that the definition of $\eta_x^\star$ does not follow the same scheme as the definitions of $\eta_h^\star$ and $\eta_w^\star$ in equations (20) and (23). Indeed in Proposition 16, which characterizes the eigenvalues of the Jacobian matrix $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$, the upper bound $\eta_x^\star$ is not equal to $\frac{2}{\|\boldsymbol{P}(\boldsymbol{x})\|_2}$ (with $\boldsymbol{P}(\boldsymbol{x})$ defined in equation (29)).

Proposition 16. *Let $\boldsymbol{x}$ be a local minimum of function J defined in equation* (17)*. Function J satisfies Assumption 1, and we assume that $\boldsymbol{x}$ satisfies Assumption 2. Consider the mapping $\boldsymbol{\phi}$ defined in equation* (27)*, which is continuously differentiable in a neighborhood of $\boldsymbol{x}$.*

Then if $\eta = 0$, all the eigenvalues of the Jacobian matrix $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$ are equal to 1. Otherwise, $\lambda = 1$ is always an eigenvalue of $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$. Following Notations 1 and 2, its multiplicity is greater than or equal to the dimension of the kernel of matrix $[\boldsymbol{P}(\boldsymbol{x})]_+$, where

$$\boldsymbol{P}(\boldsymbol{x}) = \boldsymbol{D}(\boldsymbol{x})\boldsymbol{\nabla}^2_{xx}J(\boldsymbol{x})\boldsymbol{D}(\boldsymbol{x}) \tag{29}$$

and $\boldsymbol{D}(\boldsymbol{x}) = \mathrm{diag}(\boldsymbol{x}/\boldsymbol{p}(\boldsymbol{x}))^{\frac{1}{2}}$. Moreover, the other eigenvalues of $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$ are as follows:

- *If $\eta \notin [0, 2]$, there is at least one eigenvalue greater than 1;*
- *If $\eta \in]0, \eta_x^\star[$ (where $\eta_x^\star$ was defined in equation* (28)*), the multiplicity of the eigenvalue 1 is equal to the dimension of the kernel of matrix $[\boldsymbol{P}(\boldsymbol{x})]_+$, and all other eigenvalues have a magnitude lower than 1.*

Proof: The proof of Proposition 16 follows the same outline as the proof of Proposition 7. It additionally relies on the observation that $\lambda = 1$ and $\lambda = (1-\eta)^2$ are always eigenvalues of the Jacobian matrix $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$ (which is proved by exhibiting the corresponding eigenvectors). ■

As mentioned above, Proposition 16 does not permit to conclude on the stability of multiplicative updates. Nevertheless, it enables a formal proof of the first and second assertions in the following proposition (the third one is trivial):

Proposition 17. *Let $\boldsymbol{x}$ be a local minimum of function J defined in equation* (17)*. Function J satisfies Assumption 1, and we assume that $\boldsymbol{x}$ satisfies Assumption 2. Consider the mapping $\boldsymbol{\phi}$ defined in equation* (27)*, which is continuously differentiable in a neighborhood of $\boldsymbol{x}$.*

Then

- $\forall \eta \in \mathbb{R}$, $\boldsymbol{x}$ *is not an exponentially stable fixed point;*
- *if* $\eta \notin [0, 2]$, $\boldsymbol{x}$ *is an unstable fixed point;*
- *if* $\eta = 0$, $\boldsymbol{x}$ *is a marginally stable fixed point.*

Proposition 17 does not tell us what happens if $\eta \in]0, 2]$. This is because 1 is always an eigenvalue of the Jacobian matrix $\nabla \boldsymbol{\phi}^T(\boldsymbol{x})$, as shown in Proposition 16. If $\eta \in]0, \eta_x^\star[$, its multiplicity is equal to the dimension of the kernel of matrix $[\boldsymbol{P}(\boldsymbol{x})]_+$, which we suppose accounts for the invariances of the factorization[10].

V. Simulation results

In this section we propose some numerical simulations which illustrate the theoretical results presented in sections III and IV. We first focus on supervised NMF, and then investigate the case of unsupervised NMF.

A. Supervised NMF

First, we study the stability of the multiplicative update (8) applied to matrix $\boldsymbol{H}$, while keeping matrix $\boldsymbol{W}$ unchanged.

1) Example of sub-linear convergence speed: In this first experiment, the dimensions are $F = 3$, $T = 3$ and $K = 2$. The multiplicative update (8) is applied to the Kullback-Leibler divergence ($\beta = 1$) with a step size $\eta = 1$ (which corresponds to the standard multiplicative update). The matrix $\boldsymbol{V}$ to be decomposed is defined as a square non-negative Hankel matrix: $\boldsymbol{V} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$, and the matrix $\boldsymbol{W}$ is defined as

$$\boldsymbol{W} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}. \tag{30}$$

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

$$\boldsymbol{H} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}. \tag{31}$$

Thus we readily 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 $\boldsymbol{H}$ defined in equation (31). This particular example was chosen so that $\widehat{\boldsymbol{V}} = \boldsymbol{V}$, thus $\boldsymbol{p}^h = \boldsymbol{m}^h$, as can be noticed in equation (32). Consequently, property (14) does not stand, since $h_{21} = 0$ and $\frac{\partial J}{\partial h_{21}} = 0$. Therefore Proposition 8, which would have proven the exponential stability of the global minimum, cannot be applied. Nevertheless all the hypotheses in Proposition 12 and Lemma 13 are satisfied, which proves the asymptotic stability. Thus the speed of convergence of the multiplicative update (8) may be sub-linear, which will be confirmed by the following simulation results.

Fig. 1 shows the results obtained by initializing (8) with a matrix $\boldsymbol{H}$ having all coefficients equal to 2. As can be noticed in Fig. 1-(a), the objective function J 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 (31)) as a solid blue line. It can be noticed that 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).

[10] Actually the invariances of the factorization imply some conditions on the first and second order derivatives of the objective function J. We managed to prove that if all the hypotheses in Proposition 17 and property (14) are satisfied, then the dimensionality of such conditions is equal to the dimension of the kernel of matrix $[\boldsymbol{P}(\boldsymbol{x})]_+$.

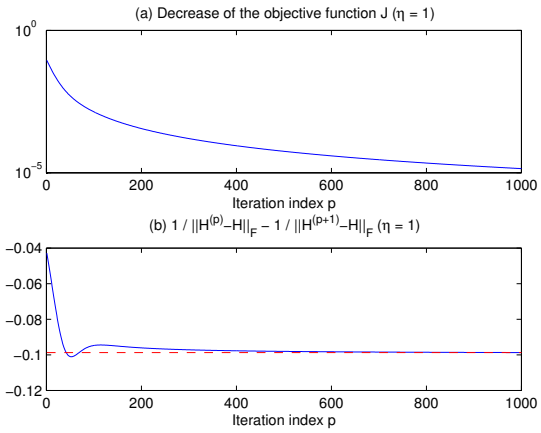

Fig. 1. Example of sub-linear convergence speed

2) Example of linear convergence speed: In this second experiment, all variables are defined as in section V-A1, 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 (8), initialized as in section V-A1 [11]. Numerically, we observed that the local minimum $\boldsymbol{x}$ is still such that $h_{21} = 0$, but $\frac{\partial J}{\partial h_{21}} > 0$, thus property (14) now stands, and Proposition 8 and Lemma 13 prove the exponential stability of this local minimum, with a convergence rate equal to the spectral radius $\rho\left(\boldsymbol{\nabla}_h \boldsymbol{\phi}^{h^T}(\boldsymbol{x})\right)$, which will be confirmed by the following simulation results.

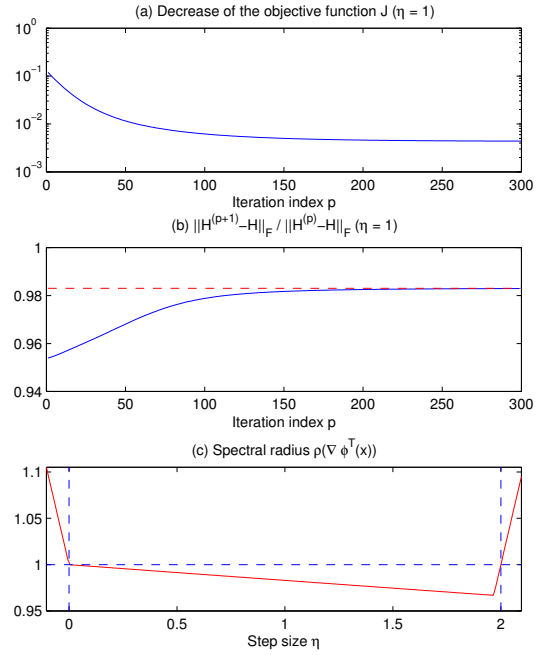

Fig. 2. Example of linear convergence speed

Fig. 2-(a) shows that the objective function J is monotonically decreasing. Besides, Fig. 2-(b) represents the

[11] Note that in this second experiment, the value of matrix $\boldsymbol{W}$ defined in equation (30), which remains unchanged throughout the iterations, does no longer correspond to a local minimum of the objective function w.r.t. $\boldsymbol{W}$.

sequence $\frac{\|\boldsymbol{H}^{(p+1)}-\boldsymbol{H}\|_F}{\|\boldsymbol{H}^{(p)}-\boldsymbol{H}\|_F}$ as a solid blue line, and the value $\rho\left(\boldsymbol{\nabla}_h\boldsymbol{\phi}^{h^T}(\boldsymbol{x})\right)$ as a dashed red line[12]. It can be noticed that this sequence converges to $\rho\left(\boldsymbol{\nabla}_h\boldsymbol{\phi}^{h^T}(\boldsymbol{x})\right)$, which shows that $\|\boldsymbol{H}^{(p)}-\boldsymbol{H}\|_F = O\left(\rho\left(\boldsymbol{\nabla}_h\boldsymbol{\phi}^{h^T}(\boldsymbol{x})\right)^p\right)$. As predicted by the theoretical analysis, the convergence speed is linear, with a convergence rate equal to $\rho\left(\boldsymbol{\nabla}_h\boldsymbol{\phi}^{h^T}(\boldsymbol{x})\right)$.

3) Optimal step size: In this third experiment, all variables are defined as in section V-A2, and we are looking for an optimal step size η. Since $\beta = 1$, Proposition 8 and Lemma 13 prove 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 $\rho\left(\boldsymbol{\nabla}_h\boldsymbol{\phi}^{h^T}(\boldsymbol{x})\right)$ as a function of η, for all $\eta \in]-0.1, 2.1[$. This simulation result confirms that $\rho\left(\boldsymbol{\nabla}_h\boldsymbol{\phi}^{h^T}(\boldsymbol{x})\right) < 1$ 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, we note that the standard step size $\eta = 1$ is not optimal. Besides, we observed that a value of η outside the range $[0, 2]$ results in a divergence of the objective function J (for $\eta = 0$ the objective function is constant, and for $\eta = 2$ it oscillates between two values).

B. Unsupervised NMF

We now study the case of unsupervised NMF, which alternates multiplicative updates (7) and (8) for $\boldsymbol{W}$ and $\boldsymbol{H}$. In this fourth experiment, all variables are defined as in section V-A2, and the algorithm is initialized in the same way.

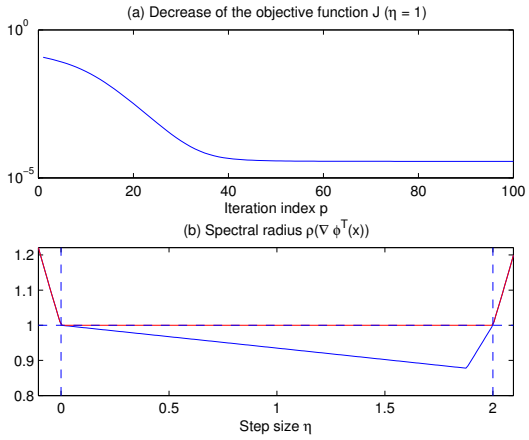

Fig. 3. Unsupervised NMF

Fig. 3-(a) shows that the objective function J is monotonically decreasing (to a non-zero value). As in section V-A3, the solid red line in Fig. 3-(b) represents the spectral radius $\rho\left(\boldsymbol{\nabla}\boldsymbol{\phi}^T(\boldsymbol{x})\right)$ as a function of the step size η, for all $\eta \in]-0.1, 2.1[$ [13]. As proved in Lemma 15 and Proposition 16, $\rho\left(\boldsymbol{\nabla}\boldsymbol{\phi}^T(\boldsymbol{x})\right) > 1$ if $\eta \notin [0, 2]$, and $\rho\left(\boldsymbol{\nabla}\boldsymbol{\phi}^T(\boldsymbol{x})\right) = 1$ in the range $\eta =]0, 2[$, which confirms that the local minimum is not exponentially stable, as stated in Proposition 17. Finally, the solid blue line represents the maximum among the magnitudes of the eigenvalues of matrix $\boldsymbol{\nabla}\boldsymbol{\phi}^T(\boldsymbol{x})$ *which are different from 1* [14]. 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 J (after 100 iterations) is reached when the algorithm is run with this optimal value of η.

VI. Conclusions

In this paper, we analyzed the convergence properties of general multiplicative update algorithms, designed to solve optimization problems with non-negativity constraints, where we introduced an exponent step size η. We have

[12]The spectral radius $\rho\left(\boldsymbol{\nabla}_h\boldsymbol{\phi}^{h^T}(\boldsymbol{x})\right)$ was computed from the closed form expression of matrix $\boldsymbol{\nabla}_h\boldsymbol{\phi}^{h^T}(\boldsymbol{x})$ presented in equation (52) in the supporting document [1].

[13]The spectral radius $\rho\left(\boldsymbol{\nabla}\boldsymbol{\phi}^T(\boldsymbol{x})\right)$ was computed from the closed form expression of matrix $\boldsymbol{\nabla}\boldsymbol{\phi}^T(\boldsymbol{x})$ presented in equation (55) in the supporting document [1].

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

applied Lyapunov's first and second methods to find some criteria which guarantee the exponential or asymptotic stability of the local minima of the objective function, either by analyzing the eigenvalues of the Jacobian matrix of the mapping, or by introducing a suitable Lyapunov function. We noticed that exponential stability ensures a linear convergence speed, whereas asymptotic stability generally leads to a sub-linear convergence speed (depending on the initialization). In both cases, we provided the closed-form expression of an upper bound $\eta^\star$ such that the exponential or asymptotic stability is guaranteed for $\eta \in]0, \eta^\star[$. This study was straightforwardly applied to supervised NMF algorithms based on β-divergences, which update all variables at once, instead of switching between two multiplicative updates. We have thus presented some criteria which guarantee the exponential or asymptotic stability of NMF multiplicative updates, and proved that the upper bound is such that $\forall \beta \in \mathbb{R}$, $\eta^\star \in]0, 2]$, and if $\beta \in [0, 2]$, $\eta^\star = 2$ (note that in this last case, Lee and Sung's multiplicative updates correspond to the particular case $\eta = 1$). We then applied the same methodology to study the more complex case of unsupervised NMF. Analyzing the stability properties of the multiplicative updates happened to be particularly difficult because of the invariances of the factorization, which make the local minima of the objective function non-isolated, thus non-asymptotically stable. Nevertheless Lyapunov's first method has provided some interesting insights into the convergence properties of those updates. In particular, we proved their unstability 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 (generally different from 1), which provides the fastest convergence rate (thus faster than that of Lee and Sung's multiplicative updates). Finally, this study can also be applied to non-negative tensor factorization *via* unfolding [11].

A possible extension of this stability analysis would focus on the hybrid case of constrained unsupervised NMF. If the constraint is expressed *via* a modification of the underlying model like in [13], [31], [33], the convergence analysis may be similar to that of unconstrained NMF. If the constraint is expressed *via* additional penaly terms in the objective function like in [7], [26], [34], [36], the convergence analysis may paradoxically be simpler than in unconstrained NMF: the algorithm still switches between two multiplicative updates, but the modified objective function may have isolated, thus asymptotically stable, local minima. In other respects, an algorithmic outlook of this work would be the design of multiplicative update algorithms with an optimal or an adaptive exponent step size. Similarly, since functions $\boldsymbol{p}$ and $\boldsymbol{m}$ are not unique, it would be interesting to investigate the impact of the choice of $\boldsymbol{p}$ and $\boldsymbol{m}$ onto the convergence rate.

APPENDIX

A. Multiplicative update algorithms for NMF

Proof of Proposition 1: We remark that $d_\beta(x|y)$ defined in (2) can be written in the form $d_\beta(x|y) = x^\beta \delta_\beta(\frac{y}{x})$, where

$$\begin{array}{rcl} \delta_\beta(u) & = & \frac{1}{\beta(\beta-1)}(1 + (\beta-1)u^\beta - \beta u^{\beta-1}), \\ \frac{d\delta_\beta}{du} & = & u^{\beta-1} - u^{\beta-2}, \\ \frac{d^2\delta_\beta}{du^2} & = & (\beta-1)u^{\beta-2} + (2-\beta)u^{\beta-3}. \end{array}$$

Function δ_β is strictly convex in $\mathbb{R}_+$ if and only if $\forall u > 0$, $\frac{d^2\delta_\beta}{du^2} > 0$, which is equivalent to $1 \leq \beta \leq 2$. Then we write

$$D(\boldsymbol{V}|\boldsymbol{WH}) = \sum_{f=1}^{F}\sum_{t=1}^{T} v_{ft}^\beta \delta_\beta\left(\frac{\hat{v}_{ft}}{v_{ft}}\right)$$

where $\hat{v}_{ft}$ was defined in equation (6). Because of the strict convexity of δ_β, we have $\forall \boldsymbol{h}'_t \in \mathbb{R}_+^K$

$$\begin{array}{rcl} \delta_\beta\left(\frac{\sum_{k=1}^{K} w_{fk}h'_{kt}}{v_{ft}}\right) & = & \delta_\beta\left(\sum_{k=1}^{K} \frac{w_{fk}h_{kt}}{\hat{v}_{ft}} \times \frac{\hat{v}_{ft}h'_{kt}}{v_{ft}h_{kt}}\right) \\ & \leq & \sum_{k=1}^{K} \frac{w_{fk}h_{kt}}{\hat{v}_{ft}} \delta_\beta\left(\frac{\hat{v}_{ft}h'_{kt}}{v_{ft}h_{kt}}\right) \end{array}$$

with equality if and only if $\forall k \in \{1 \dots K\}$, $h'_{kt} = h_{kt}$.

Thus $\forall \boldsymbol{H}' \in \mathbb{R}_+^{K\times T}$, $D(\boldsymbol{V}|\boldsymbol{W}\boldsymbol{H}') \leq G^{W,H}(\boldsymbol{H}')$, where $G^{W,H}(\boldsymbol{H}') = \sum_{k=1}^{K}\sum_{t=1}^{T} G_{kt}^{W,H}(h'_{kt})$ and $G_{kt}^{W,H}(h'_{kt}) = \sum_{f=1}^{F} \frac{w_{fk}h_{kt}}{\hat{v}_{ft}} v_{ft}^{\beta}\delta_\beta\left(\frac{\hat{v}_{ft}}{v_{ft}}\frac{h'_{kt}}{h_{kt}}\right)$ with equality if and only if $\boldsymbol{H}' = \boldsymbol{H}$. Then let $h'_{kt}(\eta) = h_{kt}\left(\frac{m^h_{kt}}{p^h_{kt}}\right)^{\eta}$ where p^h_{kt} and m^h_{kt} are defined similarly to p^w_{fk} and m^w_{fk} in equation (5):

$$\begin{aligned} p^h_{kt} &= \sum_{f=1}^{F} w_{fk}\hat{v}_{ft}^{\beta-1} \\ m^h_{kt} &= \sum_{f=1}^{F} w_{fk}v_{ft}\hat{v}_{ft}^{\beta-2} \end{aligned} \tag{32}$$

(here we assume $p^h_{kt} \neq 0$). Define the function $F_{kt}(\eta) = G_{kt}^{W,H}(h'_{kt}(\eta))$. Then
$\frac{dF_{kt}}{d\eta} = \frac{dG_{kt}^{W,H}}{dh'_{kt}}\frac{dh'_{kt}}{d\eta} = h_{kt}p^h_{kt}\ln\left(\frac{m^h_{kt}}{p^h_{kt}}\right)\left(\frac{m^h_{kt}}{p^h_{kt}}\right)^{\eta\beta}\left[1-\left(\frac{m^h_{kt}}{p^h_{kt}}\right)^{1-\eta}\right]$ (where we assume $m^h_{kt} \neq 0$).
If $h_{kt} \neq 0$ and $m^h_{kt} \neq p^h_{kt}$, then

- $\frac{dF_{kt}}{d\eta} < 0$ for all $\eta \in]-\infty, 1[$,
- $\frac{dF_{kt}}{d\eta} = 0$ for $\eta = 1$,
- $\frac{dF_{kt}}{d\eta} > 0$ for all $\eta \in]1, +\infty[$.

In particular, $\forall \eta \in]0,1]$, $F_{kt}(\eta) < F_{kt}(0)$. Finally, let $F(\eta) = \sum_{k=1}^{K}\sum_{t=1}^{T} F_{kt}(\eta) = G^{W,H}(\boldsymbol{H}'(\eta))$. If $\boldsymbol{H}'(\eta) \neq \boldsymbol{H}$, then $\forall \eta \in]0,1]$, $F(\eta) < F(0)$. Consequently, $\forall \eta \in]0,1]$, $D(\boldsymbol{V}|\boldsymbol{W}\boldsymbol{H}'(\eta)) \leq F(\eta) < F(0) = D(\boldsymbol{V}|\boldsymbol{W}\boldsymbol{H})$. Thus $\boldsymbol{H}'(\eta) \neq \boldsymbol{H} \;\Rightarrow\; D(\boldsymbol{V}|\boldsymbol{W}\boldsymbol{H}'(\eta)) < D(\boldsymbol{V}|\boldsymbol{W}\boldsymbol{H})$.

The same proof can be applied to the update of $\boldsymbol{W}$. ■

B. Lyapunov's first method

Proof of Proposition 7: By differentiating equation (10), we obtain the expression of the Jacobian matrix $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$:

$$\begin{aligned} &\nabla\boldsymbol{\phi}^T(\boldsymbol{x}) = \boldsymbol{\Lambda}(\boldsymbol{x})^{\eta} + \\ &\eta\left(\nabla\boldsymbol{m}^T(\boldsymbol{x})\mathrm{diag}(\tfrac{1}{\boldsymbol{m}(\boldsymbol{x})}) - \nabla\boldsymbol{p}^T(\boldsymbol{x})\mathrm{diag}(\tfrac{1}{\boldsymbol{p}(\boldsymbol{x})})\right)\mathrm{diag}(\boldsymbol{\phi}(\boldsymbol{x})). \end{aligned} \tag{33}$$

If $\boldsymbol{x}$ is a local minimum of the objective function J, Lemma 5 proves that it is a fixed point of $\boldsymbol{\phi}$, thus equation (33) yields

$$\nabla\boldsymbol{\phi}^T(\boldsymbol{x}) = \boldsymbol{\Lambda}(\boldsymbol{x})^{\eta} - \eta\boldsymbol{\nabla}^2 J(\boldsymbol{x})\,\mathrm{diag}(\boldsymbol{x}/\boldsymbol{p}(\boldsymbol{x})) \tag{34}$$

Now let us have a look at the eigenvalues of matrix $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$.

- For all i such that $x_i = 0$, it is easy to see that the i^{th} column of the identity matrix, that we denote by $\boldsymbol{u}_i$, is a right eigenvector of $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$, associated to the eigenvalue $\lambda_i = \left(\frac{m_i(\boldsymbol{x})}{p_i(\boldsymbol{x})}\right)^{\eta}$ (since the product of the last term in equation (34) and vector $\boldsymbol{u}_i$ is zero). We can conclude that if $\nabla_i J(\boldsymbol{x}) = 0$ or $\eta = 0$, then $\lambda_i = 1$; otherwise $\lambda_i \in [0,1[$ if $\eta > 0$, and $\lambda_i > 1$ if $\eta < 0$.
- Let $\boldsymbol{u}$ be a right eigenvector of $\nabla\boldsymbol{\phi}^T(\boldsymbol{x})$ which does not belong to the subspace spanned by the previous ones, associated to an eigenvalue λ. Then

$$\boldsymbol{\Lambda}(\boldsymbol{x})^{\eta}\boldsymbol{u} - \eta\boldsymbol{\nabla}^2 J(\boldsymbol{x})\mathrm{diag}(\boldsymbol{x}/\boldsymbol{p}(\boldsymbol{x}))\boldsymbol{u} = \lambda\boldsymbol{u}. \tag{35}$$

Let $\boldsymbol{v} = \boldsymbol{D}(\boldsymbol{x})\,\boldsymbol{u}$; this vector is non-zero, otherwise $\boldsymbol{u}$ would belong the subspace spanned by the previous set of eigenvectors $\boldsymbol{u}_i$. Left multiplying equation (35) by $\boldsymbol{D}(\boldsymbol{x})$ yields $\boldsymbol{\Lambda}(\boldsymbol{x})^{\eta}\boldsymbol{v} - \eta\boldsymbol{P}(\boldsymbol{x})\boldsymbol{v} = \lambda\boldsymbol{v}$, where the positive semi-definite matrix $\boldsymbol{P}(\boldsymbol{x})$ was defined in equation (12). Then noting that $\boldsymbol{\Lambda}(\boldsymbol{x})^{\eta}\boldsymbol{v} = \boldsymbol{v}$ (since for any index i, either $\Lambda_i(\boldsymbol{x}) = 1$ or $v_i = 0$), we obtain $(\boldsymbol{I}_n - \eta\,\boldsymbol{P}(\boldsymbol{x}))\,\boldsymbol{v} = \lambda\boldsymbol{v}$, where $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. This proves that λ is an eigenvalue of $\boldsymbol{I}_n - \eta\,\boldsymbol{P}(\boldsymbol{x})$. It is easy to see that the previous set of vectors $\boldsymbol{u}_i$ are also eigenvectors of $\boldsymbol{P}(\boldsymbol{x})$, associated to the eigenvalue 0, but they cannot be colinear to $\boldsymbol{v}$, since $v_i = 0$ for all i such that $x_i = 0$. Thus $\lambda = 1 - \eta\mu$, where μ is an eigenvalue of $[\boldsymbol{P}(\boldsymbol{x})]_+^{\star}$ (with the use of Notation 1). We can conclude that if $\mu = 0$, then $\lambda = 1$. Otherwise we note that $\eta^{\star}$ defined in equation (13) is equal to $\frac{2}{\|[\boldsymbol{P}(\boldsymbol{x})]_+^{\star}\|_2}$, and

- if $\eta = 0$, all the other eigenvalues are equal to 1;
- if $0 < \eta < \eta^\star$, all the other eigenvalues belong to $]-1, 1[$;
- if $\eta < 0$, all the other eigenvalues are greater than 1;
- if $\eta > \eta^\star$, there is at least one eigenvalue $\lambda < -1$;
- if $\eta = \eta^\star$, there is at least one eigenvalue $\lambda = -1$.

Finally, the total number of eigenvalues equal to 1 (if $\eta \neq 0$) is the number of coefficients i such that $x_i = 0$ and $\nabla_i J(\boldsymbol{x}) = 0$, plus the dimension of the kernel of matrix $[\boldsymbol{P}(\boldsymbol{x})]^\star_+$. In other words, it is equal to the dimension of the kernel of $[\boldsymbol{P}(\boldsymbol{x})]_+$. ■

Proof of Proposition 9: Since $\boldsymbol{x}$ is an exponentially stable fixed point of mapping $\boldsymbol{\phi}$, all the eigenvalues of $\nabla \boldsymbol{\phi}^T(\boldsymbol{x})$ have magnitude lower than 1. Moreover, $\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{x}$ and $\forall i$, either $x_i = 0$ or $m_i(\boldsymbol{x}) = p_i(\boldsymbol{x})$. Thus equation (33) still yields equation (34). Again, let us have a look at the eigenvalues of matrix $\nabla \boldsymbol{\phi}^T(\boldsymbol{x})$:

- For all i such that $x_i = 0$, the eigenvalue $\lambda_i = \left(\frac{m_i(\boldsymbol{x})}{p_i(\boldsymbol{x})}\right)^\eta$ associated to the eigenvector $\boldsymbol{u}_i$ is lower than 1. Thus $m_i(\boldsymbol{x}) < p_i(\boldsymbol{x})$ and $\nabla_i J(\boldsymbol{x}) > 0$.
- Previous developments in the Proof of Proposition 7 show that the others eigenvalues λ can be written in the form $\lambda = 1 - \eta\,\mu$, where μ is an eigenvalue of $[\boldsymbol{P}(\boldsymbol{x})]^\star_+$. Since $\lambda < 1$, we conclude that $\mu > 0$, thus $[\boldsymbol{P}(\boldsymbol{x})]^\star_+$ is a positive definite matrix, and so is $[\boldsymbol{\nabla}^2 J(\boldsymbol{x})]^\star_+ = [\boldsymbol{\nabla}^2 J(\boldsymbol{x})]_+$.

We have thus proved that properties (14) and (15) stand. This proves that $\boldsymbol{x}$ is a local minimum of function J. ■

C. Lyapunov's second method

Proof of Proposition 12: Function $V(\boldsymbol{x}, \boldsymbol{y})$ defined in equation (16) can be decomposed as follows:

$$\begin{aligned} V(\boldsymbol{x},\boldsymbol{y}) = & \sum_{i/x_i=0} y_i \tfrac{p_i(\boldsymbol{x})+p_i(\boldsymbol{y})}{2} \\ & + \sum_{i/x_i>0} \tfrac{1}{2}(y_i - x_i)^2 \tfrac{p_i(\boldsymbol{x})+p_i(\boldsymbol{y})}{x_i+y_i} \\ = & \sum_{i/x_i=0} y_i \left(p_i(\boldsymbol{x}) + O(\|\boldsymbol{y}-\boldsymbol{x}\|)\right) \\ & + \sum_{i/x_i>0} \tfrac{1}{2}(y_i - x_i)^2 \left(\tfrac{p_i(\boldsymbol{x})}{x_i} + O(\|\boldsymbol{y}-\boldsymbol{x}\|)\right). \end{aligned} \tag{36}$$

Note that, since $\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{x}$ and $\boldsymbol{\phi}$ is continuously differentiable at $\boldsymbol{x}$, $\|\boldsymbol{\phi}(\boldsymbol{y}) - \boldsymbol{x}\| = O(\|\boldsymbol{y} - \boldsymbol{x}\|)$. Therefore replacing $\boldsymbol{y}$ by $\boldsymbol{\phi}(\boldsymbol{y})$ in equation (36) yields

$$\begin{aligned} V(\boldsymbol{x},\boldsymbol{\phi}(\boldsymbol{y})) = & \sum_{i/x_i=0} \phi_i(\boldsymbol{y}) \left(p_i(\boldsymbol{x}) + O(\|\boldsymbol{y}-\boldsymbol{x}\|)\right) \\ & + \sum_{i/x_i>0} \tfrac{1}{2}(\phi_i(\boldsymbol{y}) - x_i)^2 \left(\tfrac{p_i(\boldsymbol{x})}{x_i} + O(\|\boldsymbol{y}-\boldsymbol{x}\|)\right) \end{aligned} \tag{37}$$

Then subtracting equation (36) to equation (37) yields

$$\begin{aligned} & V(\boldsymbol{x},\boldsymbol{\phi}(\boldsymbol{y})) - V(\boldsymbol{x},\boldsymbol{y}) \\ & = \sum_{i/x_i=0} (\phi_i(\boldsymbol{y}) - y_i) \left(p_i(\boldsymbol{x}) + O(\|\boldsymbol{y}-\boldsymbol{x}\|)\right) + \\ & \sum_{i/x_i>0} (\phi_i(\boldsymbol{y}) - y_i)(y_i - x_i + \tfrac{\phi_i(\boldsymbol{y})-y_i}{2})(\tfrac{p_i(\boldsymbol{x})}{x_i} + O(\|\boldsymbol{y}-\boldsymbol{x}\|)) \end{aligned} \tag{38}$$

However, equation (10) proves that $\phi_i(\boldsymbol{y}) - y_i = y_i \left(\left(\frac{m_i(\boldsymbol{y})}{p_i(\boldsymbol{y})}\right)^\eta - 1\right)$; in particular:

- if $x_i = 0$ and $\nabla_i J(\boldsymbol{x}) > 0$,

$$\phi_i(\boldsymbol{y}) - y_i = -y_i \left(1 - \left(\frac{m_i(\boldsymbol{x})}{p_i(\boldsymbol{x})}\right)^\eta + O(\|\boldsymbol{y}-\boldsymbol{x}\|)\right)$$

- if $x_i = 0$ and $\nabla_i J(\boldsymbol{x}) = 0$,

$$\phi_i(\boldsymbol{y}) - y_i = -\frac{\eta y_i}{p_i(\boldsymbol{x})} \left([\boldsymbol{\nabla}^2 J(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x})]_i + O(\|\boldsymbol{y}-\boldsymbol{x}\|^2)\right)$$

- if $x_i > 0$ (and $\nabla_i J(\boldsymbol{x}) = 0$),

$$\phi_i(\boldsymbol{y}) - y_i = -\frac{\eta\, x_i}{p_i(\boldsymbol{x})}[\boldsymbol{\nabla}^2 J(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x})]_i + O(\|\boldsymbol{y}-\boldsymbol{x}\|^2).$$

Substituting these three equalities into equation (38), plus a few manipulations show that (with the notation $[.]_0$ and $[.]_+$ introduced in Notation 1)

$$\begin{array}{l} V(\boldsymbol{x},\boldsymbol{\phi}(\boldsymbol{y})) - V(\boldsymbol{x},\boldsymbol{y}) = -[\boldsymbol{y}]_0^T\left([\boldsymbol{v}(\boldsymbol{x})]_0 + O(\|\boldsymbol{y}-\boldsymbol{x}\|)\right) \\ -\eta\,[\boldsymbol{y}-\boldsymbol{x}]_+^T[M(\boldsymbol{x})]_+[\boldsymbol{y}-\boldsymbol{x}]_+ + O(\|\boldsymbol{y}-\boldsymbol{x}\|^3) \end{array} \tag{39}$$

where $[\boldsymbol{v}(\boldsymbol{x})]_0 = [\boldsymbol{I}_n - \boldsymbol{\Lambda}(\boldsymbol{x})^\eta]_0\,[\boldsymbol{p}(\boldsymbol{x})]_0$ is a vector with (strictly) positive coefficients since $\eta > 0$, and

$$\begin{array}{rl} [M(\boldsymbol{x})]_+ & = [\boldsymbol{\nabla}^2 J(\boldsymbol{x})]_+ \\ & \quad -\frac{\eta}{2}[\boldsymbol{\nabla}^2 J(\boldsymbol{x})]_+\left[\mathrm{diag}\left(\frac{\boldsymbol{x}}{\boldsymbol{p}(\boldsymbol{x})}\right)\right]_+[\boldsymbol{\nabla}^2 J(\boldsymbol{x})]_+ \end{array} \tag{40}$$

is a positive definite matrix since $\eta < \eta^\star$ (cf. lemma 18 below). Equation (39) finally proves that there is a neighborhood of $\boldsymbol{x}$ such that $\forall \boldsymbol{y} \neq \boldsymbol{x}$, $V(\boldsymbol{x},\boldsymbol{\phi}(\boldsymbol{y})) - V(\boldsymbol{x},\boldsymbol{y}) < 0$.

■

Lemma 18. *Let $\boldsymbol{x} \in \mathbb{R}_+^n$ be a local minimum of a function J satisfying Assumption 1, and suppose that Assumption 2 and property* (15) *hold. Then the matrix $[\boldsymbol{M}(\boldsymbol{x})]_+$ defined in equation* (40) *is positive definite if and only if $\eta < \eta^\star$ (where $\eta^\star$ was defined in equation* (13)*).*

Proof of Lemma 18: Matrix $[\boldsymbol{M}(\boldsymbol{x})]_+$ is positive definite if and only if $[\boldsymbol{I}_n]_+ - \frac{\eta}{2}[\boldsymbol{P}'(\boldsymbol{x})]_+$ is positive definite, where

$$[\boldsymbol{P}'(\boldsymbol{x})]_+ = \left([\boldsymbol{\nabla}^2 J(\boldsymbol{x})]_+\right)^{\frac{1}{2}}\left[\mathrm{diag}\left(\frac{\boldsymbol{x}}{\boldsymbol{p}(\boldsymbol{x})}\right)\right]_+\left([\boldsymbol{\nabla}^2 J(\boldsymbol{x})]_+\right)^{\frac{1}{2}}$$

which is equivalent to $\eta < \frac{2}{\|[\boldsymbol{P}'(\boldsymbol{x})]_+\|_2}$. However, it is easy to prove that the eigenvalues of $[\boldsymbol{P}'(\boldsymbol{x})]_+$ are equal to those of

$$[\boldsymbol{P}(\boldsymbol{x})]_+ = [\boldsymbol{D}(\boldsymbol{x})]_+\,[\boldsymbol{\nabla}^2 J(\boldsymbol{x})]_+\,[\boldsymbol{D}(\boldsymbol{x})]_+\,.$$

Consequently, $\|[\boldsymbol{P}'(\boldsymbol{x})]_+\|_2 = \|[\boldsymbol{P}(\boldsymbol{x})]_+\|_2 = \|\boldsymbol{P}(\boldsymbol{x})\|_2$. ■

REFERENCES

[1] R. Badeau, N. Bertin, and E. Vincent, "Supporting document for the paper "Stability analysis of multiplicative update algorithms and application to non-negative matrix factorization"," Télécom ParisTech, Paris, France, Tech. Rep. 2009D023, Nov. 2009.

[2] M. Berry, M. Browne, A. Langville, V. Pauca, and R. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007.

[3] N. Bertin and R. Badeau, "Initialization, distances and local minima in audio applications of the non-negative matrix factorization," in *Proc. of Acoustics'08*. Paris, France: JASA, July 2008.

[4] N. Bertin, R. Badeau, and E. Vincent, "Fast Bayesian NMF algorithms enforcing harmonicity and temporal continuity in polyphonic music transcription," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, Oct. 2009, pp. 29–32.

[5] ——, "Enforcing harmonicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 538–549, Mar. 2010.

[6] N. Bertin, C. Févotte, and R. Badeau, "A tempering approach for Itakura-Saito non-negative matrix factorization. With application to music transcription," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, Apr. 2009, pp. 1545–1548.

[7] S. Choi, "Algorithms for orthogonal nonnegative matrix factorization," in *Proc. of IEEE International Joint Conference on Neural Networks*, Orlando, Florida, USA, June 2008, pp. 1828–1832.

[8] M. Chu, F. Diele, R. Plemmons, and S. Ragni, "Optimality, computation, and interpretation of nonnegative matrix factorizations," Wake Forest University, North Carolina, USA, Tech. Rep., Oct. 2004.

[9] A. Cichocki, S. Amari, R. Zdunek, R. Kompass, G. Hori, and Z. He, "Extended SMART algorithms for non-negative matrix factorization," *Springer LNAI*, vol. 4029, pp. 548–562, 2006.

[10] A. Cichocki, R. Zdunek, and S. Amari, "Csiszar's divergences for non-negative matrix factorization: Family of new algorithms," in *Proc. of the 6th International Conference on Independent Component Analysis and Blind Signal Separation*, Charleston, SC, USA, Mar. 2006, pp. 32–39.

[11] ——, "Nonnegative matrix and tensor factorization," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 142–145, Jan. 2008.

[12] A. Cont, "Realtime multiple pitch observation using sparse non-negative constraints," in *Proc. of International Symposium on Music Information Retrieval*, Victoria, Canada, Oct. 2006, pp. 206–212.

[13] J. Eggert, H. Wersing, and E. Korner, "Transformation-invariant representation and NMF," in *Proc. of IEEE International Joint Conference on Neural Networks*, Budapest, Hungary, July 2004, pp. 2535–2539.
[14] 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
[15] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, Mar. 2009.
[16] 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, July 2004.
[17] E. 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.
[18] D. Guillamet, J. Vitrià, and B. Schiele, "Introducing a weighted non-negative matrix factorization for image classification," *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2447–2454, 2003.
[19] F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," in *Proc. of the 6th International Congress on Acoustics*, Tokyo, Japan, Aug. 1968, pp. C17–C20.
[20] R. Kompass, "A generalized divergence measure for nonnegative matrix factorization," *Neural Computation*, vol. 19, no. 3, pp. 780–791, Mar. 2007.
[21] X. Kong, C. Hu, and C. Han, "On the discrete-time dynamics of a class of self-stabilizing MCA extraction algorithms," *IEEE Trans. Neural Netw.*, vol. 21, no. 1, pp. 175–181, Jan. 2010.
[22] J. Lasalle, *The stability and control of discrete processes*. New York, NY, USA: Springer-Verlag, 1986.
[23] H. Laurberg, "Uniqueness of non-negative matrix factorization," in *Proc. of IEEE Workshop on Statistical Signal Processing*, Madison, WI, USA, Aug. 2007, pp. 44–48.
[24] 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.
[25] ——, "Algorithms for non-negative matrix factorization," in *Proc. of Conference on Advances in Neural Information Processing Systems*, vol. 13. Vancouver, British Columbia, Canada: MIT Press, Dec. 2001, pp. 556–562.
[26] S. Li, X. Hou, H. Zhang, and Q. Cheng, "Learning spatially localized, parts-based representation," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Hawaii, USA, Dec. 2001, pp. 207–212.
[27] 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.
[28] ——, "Projected gradient methods for non-negative matrix factorization," *Neural Computation*, vol. 19, no. 10, pp. 2756–2779, 2007.
[29] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed., ser. Springer Series in Operations Research and Financial Engineering. New York Inc.: Springer-Verlag, Aug. 2006.
[30] G. Peyré, "Non-negative sparse modeling of textures," in *Proc. of Scale Space and Variational Methods in Computer Vision*, Ischia, Italy, May 2007, pp. 628–639.
[31] P. Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs," in *Proc. of International Conference on Independent Component Analysis and Blind Signal Separation*, Granada, Spain, Sept. 2004, pp. 494–499.
[32] 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.
[33] E. Vincent, N. Bertin, and R. Badeau, "Adaptive harmonic spectral decomposition for multiple pitch estimation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 528–537, Mar. 2010.
[34] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, no. 3, pp. 1066–1074, Mar. 2007.
[35] Z. Wu, H. Su, J. Chu, and W. Zhou, "Improved delay-dependent stability condition of discrete recurrent neural networks with time-varying delays," *IEEE Trans. Neural Netw.*, vol. 21, no. 4, pp. 692–697, Apr. 2010.
[36] Y. Zhang and Y. Fang, "A NMF algorithm for blind separation of uncorrelated signals," in *Proc. of IEEE International Conference on Wavelet Analysis and Pattern Recognition*, Beijing, China, Nov. 2007, pp. 999–1003.
[37] C.-D. Zheng, H. Zhang, and Z. Wang, "An augmented LKF approach involving derivative information of both state and delay," *IEEE Trans. Neural Netw.*, vol. 21, no. 7, pp. 1100–1109, July 2010.

Dépôt légal : 2010 – 3ème trimestre
Imprimé à Télécom ParisTech – Paris
ISSN 0751-1345 ENST D (Paris) (France 1983-9999)

Télécom ParisTech

Institut TELECOM - membre de ParisTech

46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.telecom-paristech.frfr

Département TSI