Supporting document for the paper "On the stability of multiplicative update algorithms. Application to non-negative matrix factorization"

Document de support pour l'article «Sur la stabilité des règles de mises à jour multiplicatives. Application à la factorisation en matrices positives»

Roland Badeau
Nancy Bertin
Emmanuel Vincent

2009D023

novembre 2009

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

Supporting document for the paper "On the stability of multiplicative update algorithms. Application to non-negative matrix factorization."

Document de support pour l'article "Sur la stabilité des règles de mises à jour multiplicatives. 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 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 reference [1], 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, and finally study the difficult case of non-negative matrix factorization.

In this supporting document, we present the proofs of some theoretical results presented in [1]. This document, written as a sequel of [1], is not intended to be read separately.

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 et ses nombreuses variantes. Néanmoins, malgré plusieurs années de recherche sur le sujet, la compréhension de leurs propriétés de convergence demeure imparfaite. Dans la référence [1], 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, et nous étudions finalement le cas difficile de la factorisation en matrices positives.

Dans ce document de support, nous présentons les preuves de certains résultats théoriques présentés dans [1]. Ce document, faisant suite à [1], n'est pas destiné à être lu séparément.

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.

INTRODUCTION

In the following developments, we use the notations introduced in sections II and IV of reference [1]. This supporting document is organized as follows: section VI is devoted to the proof of Lemma 12 in section IV-A of reference [1], and section VII is devoted to the proof of Proposition 13.

We begin with a key result in linear algebra:

Theorem 17. *Let $\boldsymbol{M}$ and $\boldsymbol{N}$ be two real square matrices of same dimension. Suppose that $\boldsymbol{M}$ is non-singular, $\boldsymbol{M}^T + \boldsymbol{N}$ is positive definite, and $\boldsymbol{M} - \boldsymbol{N}$ is positive semidefinite. Then any vector in the kernel of $\boldsymbol{M} - \boldsymbol{N}$ is an eigenvector of matrix $\boldsymbol{M}^{-1}\boldsymbol{N}$, whose eigenvalue is 1. Moreover, all other eigenvalues of $\boldsymbol{M}^{-1}\boldsymbol{N}$ have modulus lower than 1.*

Proof: First note that $\boldsymbol{M} - \boldsymbol{N} = (\boldsymbol{M} + \boldsymbol{M}^T) - (\boldsymbol{M}^T + \boldsymbol{N})$, which proves that this matrix is symmetric. Moreover, any vector in its kernel is obviously an eigenvector of matrix $\boldsymbol{M}^{-1}\boldsymbol{N}$, whose eigenvalue is 1. Then let $\boldsymbol{x}$ be a complex vector such that $(\boldsymbol{M} - \boldsymbol{N})\boldsymbol{x} \neq \boldsymbol{0}$, normalized so that $\boldsymbol{x}^H(\boldsymbol{M} - \boldsymbol{N})\boldsymbol{x} = 1$. Straightforward calculations show that

$$(\boldsymbol{M}^{-1}\boldsymbol{N}\boldsymbol{x})^H(\boldsymbol{M} - \boldsymbol{N})(\boldsymbol{M}^{-1}\boldsymbol{N}\boldsymbol{x}) = 1 - \boldsymbol{y}^H(\boldsymbol{M}^T + \boldsymbol{N})\boldsymbol{y} \tag{29}$$

where $\boldsymbol{y} = \boldsymbol{M}^{-1}(\boldsymbol{M} - \boldsymbol{N})\boldsymbol{x} \neq \boldsymbol{0}$. As $\boldsymbol{M}^T + \boldsymbol{N}$ is positive definite, the right member of equation (29) is lower than 1. If moreover $\boldsymbol{x}$ is an eigenvector of matrix $\boldsymbol{M}^{-1}\boldsymbol{N}$ associated to an eigenvalue $\lambda \neq 1$, then the left member of this equation is equal to $|\lambda|^2$, which proves that $|\lambda| < 1$. ■

VI. PRELIMINARY RESULTS

A. Gradient and Hessian matrix of the objective function

In this section, we provide closed form expressions of the gradient and the Hessian matrix of the objective function $J(\boldsymbol{x}) = D(\boldsymbol{V}|\boldsymbol{W}\boldsymbol{H})$, where D was defined in equation (1).

For all $f \in \{1 \dots F\}$ and $t \in \{1 \dots T\}$, we define the vectors $\boldsymbol{w}_f = [w_{f1}; \dots; w_{fK}]$ and $\boldsymbol{h}_t = [w_{1t}; \dots; w_{Kt}]$, so that we can write $\boldsymbol{w} = [\boldsymbol{w}_1; \dots; \boldsymbol{w}_F]$, $\boldsymbol{h} = [\boldsymbol{h}_1; \dots; \boldsymbol{h}_T]$, and $\boldsymbol{x} = [\boldsymbol{w}; \boldsymbol{h}]$.

1) First and Second order partial derivatives: For all $f \in \{1 \dots F\}$ and $t \in \{1 \dots T\}$, the first order partial derivatives of the objective function are

$$\left\{\begin{array}{rcl} \boldsymbol{\nabla}_{w_f} J(\boldsymbol{x}) &=& \sum\limits_{t=1}^{T} d'(v_{ft}|\hat{v}_{ft})\, \boldsymbol{h}_t \\ \boldsymbol{\nabla}_{h_t} J(\boldsymbol{x}) &=& \sum\limits_{f=1}^{F} d'(v_{ft}|\hat{v}_{ft})\, \boldsymbol{w}_f \end{array}\right.$$

where $\hat{v}_{ft}$ was defined in equation (5). The second order partial derivatives are obtained by differentiating these expressions. First note that if $f_1 \neq f_2$, $\boldsymbol{\nabla}^2_{w_{f_1} w_{f_2}} J(\boldsymbol{x}) = \boldsymbol{0}$, and if $t_1 \neq t_2$, $\boldsymbol{\nabla}^2_{h_{t_1} h_{t_2}} J(\boldsymbol{x}) = \boldsymbol{0}$, thus matrices $\boldsymbol{\nabla}^2_{ww} J(\boldsymbol{x})$ and $\boldsymbol{\nabla}^2_{hh} J(\boldsymbol{x})$ are block-diagonal. Moreover,

$$\left\{\begin{array}{rcl} \boldsymbol{\nabla}^2_{w_f w_f} J(\boldsymbol{x}) &=& \sum\limits_{t=1}^{T} d''(v_{ft}|\hat{v}_{ft})\, \boldsymbol{h}_t \boldsymbol{h}_t^T \\ \boldsymbol{\nabla}^2_{h_t h_t} J(\boldsymbol{x}) &=& \sum\limits_{f=1}^{F} d''(v_{ft}|\hat{v}_{ft})\, \boldsymbol{w}_f \boldsymbol{w}_f^T \\ \boldsymbol{\nabla}^2_{w_f, h_t} J(\boldsymbol{x}) &=& d''(v_{ft}|\hat{v}_{ft}) \boldsymbol{h}_t \boldsymbol{w}_f^T + d'(v_{ft}|\hat{v}_{ft}) \boldsymbol{I}_K \end{array}\right.$$

where $\boldsymbol{I}_K$ denotes the $K \times K$ identity matrix.

It follows that if the divergence d is convex (i.e. $d'' \geq 0$), then matrices $\boldsymbol{\nabla}^2_{ww} J(\boldsymbol{x})$ and $\boldsymbol{\nabla}^2_{hh} J(\boldsymbol{x})$ are both *positive semi-definite* and *non-negative* (they are called *doubly non-negative*). In particular, the positive semi-definite property confirms that the objective function J is convex w.r.t. $\boldsymbol{w}$ ($\boldsymbol{h}$ being fixed), and also convex w.r.t. $\boldsymbol{h}$ ($\boldsymbol{w}$ being fixed).

2) Case of the β-divergence: From now on, we focus on the β-divergence defined in equation (2). $\forall \beta \in \mathbb{R}$, the first and second order derivatives of $d_\beta(x|y)$ w.r.t. y are

$$\begin{array}{l} d'_\beta(x|y) = y^{\beta-1} \left(1 - \frac{x}{y}\right) \\ d''_\beta(x|y) = y^{\beta-2} \left((\beta - 1) + (2 - \beta)\frac{x}{y}\right) \end{array}$$

The last equality shows that the β-divergence is convex if and only if $1 \leq \beta \leq 2$. It follows that in this case, matrices $\boldsymbol{\nabla}^2_{ww} J(\boldsymbol{x})$ and $\boldsymbol{\nabla}^2_{hh} J(\boldsymbol{x})$ are doubly non-negative.

B. Notations

Using the results of section VI-A, the gradients of the objective function w.r.t. $\boldsymbol{w}$ and $\boldsymbol{h}$ can be written as the difference of two non-negative terms:

$$\left\{ \begin{array}{rcl} \boldsymbol{\nabla}_w J(\boldsymbol{x}) & = & \boldsymbol{p}^w(\boldsymbol{x}) - \boldsymbol{m}^w(\boldsymbol{x}) \\ \boldsymbol{\nabla}_h J(\boldsymbol{x}) & = & \boldsymbol{p}^h(\boldsymbol{x}) - \boldsymbol{m}^h(\boldsymbol{x}) \end{array} \right. \tag{30}$$

where $\boldsymbol{p}^w(\boldsymbol{x}) = [\boldsymbol{p}^{w_1}(\boldsymbol{x}); \ldots; \boldsymbol{p}^{w_F}(\boldsymbol{x})]$, vectors $\boldsymbol{m}^w(\boldsymbol{x})$, $\boldsymbol{p}^h(\boldsymbol{x})$, and $\boldsymbol{m}^h(\boldsymbol{x})$ being formed in a similar way, and

$$\left\{ \begin{array}{rcl} \boldsymbol{p}^{w_f}(\boldsymbol{x}) & = & \sum\limits_{t=1}^{T} \hat{v}_{ft}^{\beta-1}\, \boldsymbol{h}_t \\ \boldsymbol{m}^{w_f}(\boldsymbol{x}) & = & \sum\limits_{t=1}^{T} \hat{v}_{ft}^{\beta-1}\, \frac{v_{ft}}{\hat{v}_{ft}}\, \boldsymbol{h}_t \\ \boldsymbol{p}^{h_t}(\boldsymbol{x}) & = & \sum\limits_{f=1}^{F} \hat{v}_{ft}^{\beta-1}\, \boldsymbol{w}_f \\ \boldsymbol{m}^{h_t}(\boldsymbol{x}) & = & \sum\limits_{f=1}^{F} \hat{v}_{ft}^{\beta-1}\, \frac{v_{ft}}{\hat{v}_{ft}}\, \boldsymbol{w}_f \end{array} \right.$$

Note that these expressions correspond to those already provided in equations (4) and (20). In the next section, we will use the following result:

Lemma 18. $\forall \beta \in \mathbb{R}$, $\forall \boldsymbol{x} \in \mathbb{R}^{K(F+N)}$,

$$\boldsymbol{p}^w(\boldsymbol{x}) - \boldsymbol{\nabla}_{ww}^2 J(\boldsymbol{x})\ \boldsymbol{w} = (2-\beta)\boldsymbol{\nabla}_w J(\boldsymbol{x}), \tag{31}$$

$$\boldsymbol{p}^w(\boldsymbol{x}) - \boldsymbol{\nabla}_{wh}^2 J(\boldsymbol{x})\ \boldsymbol{h} = -(\beta-1)\boldsymbol{\nabla}_w J(\boldsymbol{x}). \tag{32}$$

Proof: Equation (31) is proved by noting that matrix $\boldsymbol{\nabla}_{ww}^2 J(\boldsymbol{x})$ is block-diagonal (as mentioned in section VI-A1), and that $\forall f \in \{1 \ldots F\}$,

$$\begin{array}{rl} & \boldsymbol{\nabla}_{w_f w_f}^2 J(\boldsymbol{x})\ \boldsymbol{w}_f \\ = & \left(\sum\limits_{t=1}^{T} \hat{v}_{ft}^{\beta-2} \left((\beta-1) + (2-\beta)\frac{v_{ft}}{\hat{v}_{ft}} \right) \boldsymbol{h}_t \boldsymbol{h}_t^T \right) \boldsymbol{w}_f \\ = & \sum\limits_{t=1}^{T} \hat{v}_{ft}^{\beta-1} \boldsymbol{h}_t - (2-\beta) \sum\limits_{t=1}^{T} \hat{v}_{ft}^{\beta-1} (1 - \frac{v_{ft}}{\hat{v}_{ft}}) \boldsymbol{h}_t \\ = & \boldsymbol{p}^{w_f}(\boldsymbol{x}) - (2-\beta)\boldsymbol{\nabla}_{w_f} J(\boldsymbol{x}) \end{array}$$

Equation (32) is proved by noting that $\forall f \in \{1 \ldots F\}$ and $\forall t \in \{1 \ldots T\}$,

$$\begin{array}{rl} & \boldsymbol{\nabla}_{w_f h_t}^2 J(\boldsymbol{x})\ \boldsymbol{h}_t \\ = & \hat{v}_{ft}^{\beta-1} \left((\beta-1) + (2-\beta)\frac{v_{ft}}{\hat{v}_{ft}} \right) \boldsymbol{h}_t + \hat{v}_{ft}^{\beta-1} (1 - \frac{v_{ft}}{\hat{v}_{ft}}) \boldsymbol{h}_t \end{array}$$

thus

$$\begin{array}{rl} & \sum\limits_{t=1}^{T} \boldsymbol{\nabla}_{w_f h_t}^2 J(\boldsymbol{x})\ \boldsymbol{h}_t \\ = & \sum\limits_{t=1}^{T} \hat{v}_{ft}^{\beta-1} \boldsymbol{h}_t + (\beta-1) \sum\limits_{t=1}^{T} \hat{v}_{ft}^{\beta-1} (1 - \frac{v_{ft}}{\hat{v}_{ft}}) \boldsymbol{h}_t \\ = & \boldsymbol{p}^{w_f}(\boldsymbol{x}) + (\beta-1)\boldsymbol{\nabla}_{w_f} J(\boldsymbol{x}) \end{array}$$

■

C. Upper bound for the exponent step size η

We can now prove Lemma 12:

Proof of Lemma 12: Let $\boldsymbol{x}$ be a local minimum of the objective function J. Lemma 19 below shows that both matrices $\boldsymbol{P}^w(\boldsymbol{x})$ and $\boldsymbol{P}^h(\boldsymbol{x})$ have an eigenvalue equal to 1 (see equations (34) and (36)). Thus $\|\boldsymbol{P}^w(\boldsymbol{x})\| \geq 1$ and $\|\boldsymbol{P}^h(\boldsymbol{x})\| \geq 1$. This proves that $\eta^\star \leq 2$. If moreover $\beta \in [1,2]$, lemma 20 additionally shows that 1 is the greatest eigenvalue of both matrices $\boldsymbol{P}^w(\boldsymbol{x})$ and $\boldsymbol{P}^h(\boldsymbol{x})$, thus $\|\boldsymbol{P}^w(\boldsymbol{x})\| = \|\boldsymbol{P}^h(\boldsymbol{x})\| = 1$. This finally proves that $\eta^\star = 2$. ∎

Lemma 19. *Let decompose the matrix $\boldsymbol{P}(\boldsymbol{x})$ defined in equation* (19) *into four sub-blocks:*

$$\boldsymbol{P}(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{P}^w(\boldsymbol{x}) & \boldsymbol{P}^{wh}(\boldsymbol{x}) \\ \boldsymbol{P}^{hw}(\boldsymbol{x}) & \boldsymbol{P}^h(\boldsymbol{x}) \end{bmatrix} \tag{33}$$

(where matrices $\boldsymbol{P}^w(\boldsymbol{x})$ and $\boldsymbol{P}^h(\boldsymbol{x})$ were already defined in equation (17)*). If $\boldsymbol{x}$ is a local minimum of the objective function J, then those sub-blocks satisfy the following equalities:*

$$\begin{aligned}
\boldsymbol{P}^w(\boldsymbol{x})\,\boldsymbol{D}^w(\boldsymbol{x})\,\boldsymbol{p}^w(\boldsymbol{x}) &= \boldsymbol{D}^w(\boldsymbol{x})\,\boldsymbol{p}^w(\boldsymbol{x}) && (34)\\
\boldsymbol{P}^{wh}(\boldsymbol{x})\,\boldsymbol{D}^h(\boldsymbol{x})\,\boldsymbol{p}^h(\boldsymbol{x}) &= \boldsymbol{D}^w(\boldsymbol{x})\,\boldsymbol{p}^w(\boldsymbol{x}) && (35)\\
\boldsymbol{P}^h(\boldsymbol{x})\,\boldsymbol{D}^h(\boldsymbol{x})\,\boldsymbol{p}^h(\boldsymbol{x}) &= \boldsymbol{D}^h(\boldsymbol{x})\,\boldsymbol{p}^h(\boldsymbol{x}) && (36)\\
\boldsymbol{P}^{hw}(\boldsymbol{x})\,\boldsymbol{D}^w(\boldsymbol{x})\,\boldsymbol{p}^w(\boldsymbol{x}) &= \boldsymbol{D}^h(\boldsymbol{x})\,\boldsymbol{p}^h(\boldsymbol{x}) && (37)
\end{aligned}$$

Proof: Let prove equation (34) (the same proof stands for (36)). Since $\boldsymbol{x}$ is a local minimum of the objective function J, all non-zero coefficients of vector $\boldsymbol{\nabla}_w J(\boldsymbol{x})$ correspond to zero coefficients of vector $\boldsymbol{w}$, thus $\boldsymbol{D}^w(\boldsymbol{x})\boldsymbol{\nabla}_w J(\boldsymbol{x}) = \boldsymbol{0}$. Consequently, equation (31) yields $\boldsymbol{D}^w(\boldsymbol{x})\boldsymbol{\nabla}^2_{ww} J(\boldsymbol{x})\,\boldsymbol{w} = \boldsymbol{D}^w(\boldsymbol{x})\boldsymbol{p}^w(\boldsymbol{x})$, which is equivalent to $\boldsymbol{P}^w(\boldsymbol{x})\,\boldsymbol{D}^w(\boldsymbol{x})\,\boldsymbol{p}^w(\boldsymbol{x}) = \boldsymbol{D}^w(\boldsymbol{x})\,\boldsymbol{p}^w(\boldsymbol{x})$.

Then let prove equation (35) (the same proof stands for (37)). Since $\boldsymbol{D}^w(\boldsymbol{x})\boldsymbol{\nabla}_w J(\boldsymbol{x}) = \boldsymbol{0}$, equation (32) yields $\boldsymbol{D}^w(\boldsymbol{x})\boldsymbol{\nabla}^2_{wh} J(\boldsymbol{x})\,\boldsymbol{h} = \boldsymbol{D}^w(\boldsymbol{x})\boldsymbol{p}^w(\boldsymbol{x})$, which is equivalent to $\boldsymbol{P}^{wh}(\boldsymbol{x})\,\boldsymbol{D}^h(\boldsymbol{x})\,\boldsymbol{p}^h(\boldsymbol{x}) = \boldsymbol{D}^w(\boldsymbol{x})\,\boldsymbol{p}^w(\boldsymbol{x})$. ∎

Lemma 20. *If $\beta \in [1,2]$ and if $\boldsymbol{x}$ is a local minimum of the objective function J, then all the eigenvalues of the positive semidefinite matrices $\boldsymbol{P}^w(\boldsymbol{x})$ and $\boldsymbol{P}^h(\boldsymbol{x})$ are lower than or equal to* 1.

Proof: Let prove that all the eigenvalues of $\boldsymbol{P}^w(\boldsymbol{x})$ are lower than or equal to 1 (the same proof stands for $\boldsymbol{P}^h(\boldsymbol{x})$). First, note that the eigenvalues of matrix $\boldsymbol{P}^w(\boldsymbol{x})$ are the same as those of matrix $\boldsymbol{M}^{-1}\boldsymbol{N}$, where $\boldsymbol{M} = \mathrm{diag}(\boldsymbol{p}^w(\boldsymbol{x}))$, and $\boldsymbol{N} = \mathrm{diag}(\sqrt{\boldsymbol{w}})\boldsymbol{\nabla}^2_{ww} J(\boldsymbol{x})\mathrm{diag}(\sqrt{\boldsymbol{w}})$. Obviously, $\boldsymbol{M}^T + \boldsymbol{N}$ is positive definite. Let us prove by contradiction that $\boldsymbol{M} - \boldsymbol{N}$ is positive semidefinite. Thus assume that $\boldsymbol{M} - \boldsymbol{N}$ has an eigenvalue $\lambda < 0$. Then matrix $\mathrm{diag}(\boldsymbol{p}^w(\boldsymbol{x})) - \boldsymbol{\nabla}^2_{ww} J(\boldsymbol{x})\,\mathrm{diag}(\boldsymbol{w}) - \lambda \boldsymbol{I}$ is singular. However, since $\boldsymbol{x}$ is a local minimum of the objective function J, all coefficients of vector $\boldsymbol{\nabla}_w J(\boldsymbol{x})$ are non-negative, thus equation (31) proves that all coefficients of vector $\boldsymbol{p}^w(\boldsymbol{x}) - \boldsymbol{\nabla}^2_{ww} J(\boldsymbol{x})\,\boldsymbol{w}$ are non-negative. Besides, since $\beta \in [1,2]$, d_β is convex, thus matrix $\boldsymbol{\nabla}^2_{ww} J(\boldsymbol{x})$ is non-negative. This proves that matrix $\mathrm{diag}(\boldsymbol{p}^w(\boldsymbol{x})) - \boldsymbol{\nabla}^2_{ww} J(\boldsymbol{x})\,\mathrm{diag}(\boldsymbol{w}) - \lambda \boldsymbol{I}$ is diagonally dominant (the difference between a diagonal coefficient and the sum of the absolute values of the other coefficients in the same row is greater than or equal to $-\lambda > 0$). Consequently, this matrix is non-singular, which contradicts the previous assertion. As a conclusion, all eigenvalues of $\boldsymbol{M} - \boldsymbol{N}$ are non-negative, thus this matrix is positive semidefinite. Lemma 20 is finally proved by applying theorem 17 to matrices $\boldsymbol{M}$ and $\boldsymbol{N}$. ∎

VII. LYAPUNOV'S FIRST METHOD

A. Expression of the Jacobian matrix

Straightforward calculations show that the Jacobian matrix of function $\boldsymbol{\phi}$ defined in equation (15) satisfies

$$\begin{aligned}
\boldsymbol{\nabla}\boldsymbol{\phi}^T &= \begin{bmatrix} \boldsymbol{\nabla}_w \boldsymbol{\phi}^{w^T} & \boldsymbol{0} \\ \boldsymbol{\nabla}_h \boldsymbol{\phi}^{w^T} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{\nabla}_w \boldsymbol{\phi}^{h^T} \\ \boldsymbol{0} & \boldsymbol{\nabla}_h \boldsymbol{\phi}^{h^T} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\nabla}_w \boldsymbol{\phi}^{w^T} & \boldsymbol{\nabla}_w \boldsymbol{\phi}^{w^T} \boldsymbol{\nabla}_w \boldsymbol{\phi}^{h^T} \\ \boldsymbol{\nabla}_h \boldsymbol{\phi}^{w^T} & \boldsymbol{\nabla}_h \boldsymbol{\phi}^{h^T} + \boldsymbol{\nabla}_h \boldsymbol{\phi}^{w^T} \boldsymbol{\nabla}_w \boldsymbol{\phi}^{h^T} \end{bmatrix}
\end{aligned} \tag{38}$$

where
$$\begin{cases} \nabla_w \boldsymbol{\phi}^{wT} & = & \boldsymbol{\Lambda}_w{}^{\eta} + \eta \nabla_w \boldsymbol{m}^{wT} \mathrm{diag}(\boldsymbol{\phi}^w / \boldsymbol{m}^w) \\ & & -\eta \nabla_w \boldsymbol{p}^{wT} \mathrm{diag}(\boldsymbol{\phi}^w / \boldsymbol{p}^w) \\ \nabla_h \boldsymbol{\phi}^{h^T} & = & \boldsymbol{\Lambda}_h{}^{\eta} + \eta \nabla_h \boldsymbol{m}^{h^T} \mathrm{diag}(\boldsymbol{\phi}^h / \boldsymbol{m}^h) \\ & & -\eta \nabla_h \boldsymbol{p}^{h^T} \mathrm{diag}(\boldsymbol{\phi}^h / \boldsymbol{p}^h) \\ \nabla_h \boldsymbol{\phi}^{wT} & = & \eta \nabla_h \boldsymbol{m}^{wT} \mathrm{diag}(\boldsymbol{\phi}^w / \boldsymbol{m}^w) \\ & & -\eta \nabla_h \boldsymbol{p}^{wT} \mathrm{diag}(\boldsymbol{\phi}^w / \boldsymbol{p}^w) \\ \nabla_w \boldsymbol{\phi}^{h^T} & = & \eta \nabla_w \boldsymbol{m}^{h^T} \mathrm{diag}(\boldsymbol{\phi}^h / \boldsymbol{m}^h) \\ & & -\eta \nabla_w \boldsymbol{p}^{h^T} \mathrm{diag}(\boldsymbol{\phi}^h / \boldsymbol{p}^h) \end{cases}$$

By differentiating equation (30), it can be noticed that if $\boldsymbol{x}$ is a fixed point of $\boldsymbol{\phi}$ (such that $\boldsymbol{\phi}^w(\boldsymbol{w}, \boldsymbol{h}) = \boldsymbol{w}$ and $\boldsymbol{\phi}^h(\boldsymbol{w}, \boldsymbol{h}) = \boldsymbol{h}$), these expressions can be written in a simpler form:

$$\begin{cases} \nabla_w \boldsymbol{\phi}^w(\boldsymbol{x})^T & = & \boldsymbol{\Lambda}_w(\boldsymbol{x})^{\eta} - \eta\, \nabla^2_{ww} J(\boldsymbol{x})\, \boldsymbol{D}^w(\boldsymbol{x})^2 \\ \nabla_h \boldsymbol{\phi}^h(\boldsymbol{x})^T & = & \boldsymbol{\Lambda}_h(\boldsymbol{x})^{\eta} - \eta\, \nabla^2_{hh} J(\boldsymbol{x})\, \boldsymbol{D}^h(\boldsymbol{x})^2 \\ \nabla_h \boldsymbol{\phi}^w(\boldsymbol{x})^T & = & -\eta\, \nabla^2_{hw} J(\boldsymbol{x})\, \boldsymbol{D}^w(\boldsymbol{x})^2 \\ \nabla_w \boldsymbol{\phi}^h(\boldsymbol{x})^T & = & -\eta\, \nabla^2_{wh} J(\boldsymbol{x})\, \boldsymbol{D}^h(\boldsymbol{x})^2 \end{cases}$$

Substituting these equations into (38), we obtain

$$\begin{array}{l} \nabla \boldsymbol{\phi}(\boldsymbol{x})^T = \\ \boldsymbol{\Lambda}(\boldsymbol{x})^{\eta} - \eta \begin{bmatrix} \nabla^2_{ww} J(\boldsymbol{x}) & \boldsymbol{\Lambda}_w(\boldsymbol{x})^{\eta}\, \nabla^2_{wh} J(\boldsymbol{x}) \\ \nabla^2_{hw} J(\boldsymbol{x}) & \nabla^2_{hh} J(\boldsymbol{x}) \end{bmatrix} \boldsymbol{D}(\boldsymbol{x})^2 \\ +\eta^2 \nabla^2_{xx} J(\boldsymbol{x}) \boldsymbol{D}(\boldsymbol{x})^2 \begin{bmatrix} \boldsymbol{0} & \nabla^2_{wh} J(\boldsymbol{x}) \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \boldsymbol{D}(\boldsymbol{x})^2 \end{array} \tag{39}$$

B. *Eigenvalues of the Jacobian matrix*

Using the expression of the Jacobian matrix $\nabla \boldsymbol{\phi}(\boldsymbol{x})^T$ provided in equation (39), we can now prove Proposition 13:

Proof of Proposition 13: It is easy to prove that any local minimum $\boldsymbol{x}$ of the objective function J is a fixed point of $\boldsymbol{\phi}$. Let now have a look at the eigenvalues of matrix $\nabla \boldsymbol{\phi}(\boldsymbol{x})^T$:

1) For all i such that $x_i = 0$, let $\boldsymbol{u}_i$ be the i^{th} column of the identity matrix. Then $\boldsymbol{u}_i$ is a right eigenvector of $\nabla \boldsymbol{\phi}(\boldsymbol{x})^T$, associated to the eigenvalue $\lambda_i = \left(\frac{m_i(\boldsymbol{x})}{p_i(\boldsymbol{x})}\right)^{\eta}$. 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 $\eta < 0$.
2) Let $\boldsymbol{u}$ be a right eigenvector of $\nabla \boldsymbol{\phi}(\boldsymbol{x})^T$ which does not belong to the subspace spanned by the previous ones, associated to an eigenvalue λ. Left multiplying equation (39) by $\boldsymbol{D}(\boldsymbol{x})$ and right multiplying it by $\boldsymbol{u}$ yields
$$\begin{array}{l} \lambda \boldsymbol{D}(\boldsymbol{x}) \boldsymbol{u} = \boldsymbol{D}(\boldsymbol{x}) \boldsymbol{u} - \eta \boldsymbol{D}(\boldsymbol{x}) \nabla^2_{xx} J(\boldsymbol{x}) \boldsymbol{D}(\boldsymbol{x})^2\, \boldsymbol{u} + \\ \eta^2 \boldsymbol{D}(\boldsymbol{x}) \nabla^2_{xx} J(\boldsymbol{x}) \boldsymbol{D}(\boldsymbol{x})^2 \begin{bmatrix} \boldsymbol{0} & \nabla^2_{wh} J(\boldsymbol{x}) \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \boldsymbol{D}(\boldsymbol{x})^2 \boldsymbol{u} \end{array} \tag{40}$$
Let $\boldsymbol{v} = \boldsymbol{D}(\boldsymbol{x}) \boldsymbol{u}$, and consider the decomposition of matrix $\boldsymbol{P}(\boldsymbol{x})$ in equation (33). Then equation (40) yields $\boldsymbol{J}(\eta)\, \boldsymbol{v} = \lambda\, \boldsymbol{v}$, where
$$\boldsymbol{J}(\eta) = \boldsymbol{I} - \eta \boldsymbol{P}(\boldsymbol{x}) + \eta^2 \boldsymbol{P}(\boldsymbol{x}) \begin{bmatrix} \boldsymbol{0} & \boldsymbol{P}^{wh}(\boldsymbol{x}) \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}. \tag{41}$$
However vector $\boldsymbol{v}$ is non-zero (otherwise $\boldsymbol{u}$ would belong to the space spanned by the previous set of eigenvectors). Therefore $\boldsymbol{v}$ is an eigenvector of matrix $\boldsymbol{J}(\eta)$, associated to the eigenvalue λ. Since for all i such that $x_i = 0$, $v_i = 0$, $[\boldsymbol{v}]^{\star}_{+}$ is also an eigenvector of matrix $[\boldsymbol{J}(\eta)]^{\star}_{+}$, associated to the same eigenvalue λ. As a conclusion, since they have the same cardinality, the set of eigenvalues of the Jacobian matrix $\nabla \boldsymbol{\phi}(\boldsymbol{x})^T$ which were not listed in 1) is equal to the set of eigenvalues of matrix $[\boldsymbol{J}(\eta)]^{\star}_{+}$.
Finally, applying Lemma 21 below to matrix $[\boldsymbol{J}(\eta)]^{\star}_{+}$, we prove that any vector in the kernel of $[\boldsymbol{P}(\boldsymbol{x})]^{\star}_{+}$ is a left eigenvector of matrix $[\boldsymbol{J}(\eta)]^{\star}_{+}$, whose eigenvalue is 1. If moreover $\eta \in]0, \eta^{\star}[$, where $\eta^{\star}$ was defined in equation (16), then all the other eigenvalues of $[\boldsymbol{J}(\eta)]^{\star}_{+}$ have modulus lower than 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 matrix $[\boldsymbol{P}(\boldsymbol{x})]_+$.

To finish this proof, we make the following remarks:

- Lemma 19 can be used to prove that vector
$$[(\eta-1)\boldsymbol{D}^w(\boldsymbol{x})\,\boldsymbol{p}^w(\boldsymbol{x}); \boldsymbol{D}^h(\boldsymbol{x})\,\boldsymbol{p}^h(\boldsymbol{x})]$$
is an eigenvector of $\boldsymbol{J}(\eta)$, whose eigenvalue is 1. Thus $\lambda = 1$ is always an eigenvalue of $\boldsymbol{\nabla\phi}(\boldsymbol{x})^T$.
- In the same way, lemma 19 can be used to prove that vector $\boldsymbol{D}(\boldsymbol{x})\,\boldsymbol{p}(\boldsymbol{x})$ is an eigenvector of $\boldsymbol{J}(\eta)$, whose eigenvalue is $\lambda = (1-\eta)^2$. Thus if $\eta \notin [0, 2]$, there is at least one eigenvalue greater than 1.

■

Lemma 21. *Let* $\boldsymbol{P}(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{P}^w(\boldsymbol{x}) & \boldsymbol{P}^{wh}(\boldsymbol{x}) \\ \boldsymbol{P}^{hw}(\boldsymbol{x}) & \boldsymbol{P}^h(\boldsymbol{x}) \end{bmatrix}$ *be a positive semidefinite matrix. For any* $\eta \in \mathbb{R}$, *define matrix*

$$\boldsymbol{J}(\eta) = \boldsymbol{I} - \eta \boldsymbol{P}(\boldsymbol{x}) + \eta^2 \boldsymbol{P}(\boldsymbol{x}) \begin{bmatrix} \boldsymbol{0} & \boldsymbol{P}^{wh}(\boldsymbol{x}) \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

Then any vector in the kernel of $\boldsymbol{P}(\boldsymbol{x})$ *is a left eigenvector of matrix* $\boldsymbol{J}(\eta)$, *whose eigenvalue is* 1. *If moreover* $\eta \in]0, \eta^\star[$, *where* $\eta^\star$ *was defined in equation* (16), *then all the other eigenvalues of* $\boldsymbol{J}(\eta)$ *have modulus lower than* 1.

Proof: Let

$$\left\{ \begin{array}{rcl} \boldsymbol{M} & = & \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \eta \boldsymbol{P}^{hw}(\boldsymbol{x}) & \boldsymbol{I} \end{bmatrix} \\ \boldsymbol{N} & = & \boldsymbol{M} - \eta \boldsymbol{P}(\boldsymbol{x}) \\ & = & \begin{bmatrix} \boldsymbol{I} - \eta \boldsymbol{P}^w(\boldsymbol{x}) & -\eta \boldsymbol{P}^{wh}(\boldsymbol{x}) \\ \boldsymbol{0} & \boldsymbol{I} - \eta \boldsymbol{P}^h(\boldsymbol{x}) \end{bmatrix} \end{array} \right.$$

Then $\boldsymbol{M}$ is non-singular, and its inverse matrix is $\boldsymbol{M}^{-1} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\eta \boldsymbol{P}^{hw}(\boldsymbol{x}) & \boldsymbol{I} \end{bmatrix}$. Consequently,

$$\begin{array}{rcl} \boldsymbol{M}^{-1}\boldsymbol{N} & = & \boldsymbol{M}^{-1}(\boldsymbol{M} - \eta \boldsymbol{P}(\boldsymbol{x})) \\ & = & \boldsymbol{I} - \eta \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\eta \boldsymbol{P}^{hw}(\boldsymbol{x}) & \boldsymbol{I} \end{bmatrix} \boldsymbol{P}(\boldsymbol{x}) \\ & = & \boldsymbol{J}(\eta)^T \end{array}$$

thus its eigenvalues are those of $\boldsymbol{J}(\eta)$. Besides,

$$\left\{ \begin{array}{rcl} \boldsymbol{M} - \boldsymbol{N} & = & \eta \boldsymbol{P}(\boldsymbol{x}) \\ \boldsymbol{M}^T + \boldsymbol{N} & = & \begin{bmatrix} 2\boldsymbol{I} - \eta \boldsymbol{P}^w(\boldsymbol{x}) & \boldsymbol{0} \\ \boldsymbol{0} & 2\boldsymbol{I} - \eta \boldsymbol{P}^h(\boldsymbol{x}) \end{bmatrix} \end{array} \right.$$

According to the definition of $\eta^\star$, $\boldsymbol{M}^T + \boldsymbol{N}$ is positive definite if and only if $\eta < \eta^\star$. If moreover $\eta > 0$, then $\boldsymbol{M} - \boldsymbol{N}$ is positive semi-definite, and its kernel is equal to that of $\boldsymbol{P}(\boldsymbol{x})$. Lemma 21 is thus proved by applying theorem 17 to matrices $\boldsymbol{M}$ and $\boldsymbol{N}$. ■

REFERENCES

[1] R. Badeau, N. Bertin, and E. Vincent, "On the stability of multiplicative update algorithms. Application to non-negative matrix factorization." *IEEE Trans. Neural Netw.*, 2009, submitted for publication.

Dépôt légal : 2009 – 4ème 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