Boundedness of Modified Multiplicative Updates for Nonnegative Matrix Factorization

Jiro Katayama
Kyushu University
Fukuoka, 819–0395

Norikazu Takahashi
Okayama University
Okayama, 700–8530 Japan
Institute of Systems, Information
Technologies and Nanotechnologies
Fukuoka, 814–0001 Japan

Jun'ichi Takeuchi
Kyushu University
Fukuoka, 819–0395 Japan
Institute of Systems, Information
Technologies and Nanotechnologies
Fukuoka, 814–0001 Japan

***Abstract*—There have been proposed various types of multiplicative updates for nonnegative matrix factorization. However, these updates have a serious drawback in common: they are not defined for all pairs of nonnegative matrices. Furthermore, due to this drawback, their global convergence in the sense of Zangwill's theorem cannot be proved theoretically. In this paper, we consider slightly modified versions of various multiplicative update rules, that are defined for all pairs of matrices in the domain, and show that many of them have the boundedness property. This property is a necessary condition for update rules to be globally convergent in the sense of Zangwill's theorem.**

I. Introduction

Nonnegative matrix factorization (NMF), which is the problem of finding nonnegative matrices $\boldsymbol{W} \in \mathbb{R}_+^{m\times r}$ and $\boldsymbol{H} \in \mathbb{R}_+^{r\times n}$ such that the product $\boldsymbol{WH}$ is nearly equal to a given nonnegative matrix $\boldsymbol{X} \in \mathbb{R}_+^{m\times n}$ ($\mathbb{R}_+$ is the set of all nonnegative real numbers), has attracted great attention among researchers in many fields such as signal processing, machine learning and data analysis. NMF is usually formulated as an optimization problem in which an error function has to be minimized subject to the nonegativity constraints for $\boldsymbol{W}$ and $\boldsymbol{H}$. There have been proposed different types of algorithms for solving NMF optimization problems [1]–[3] . Among them, multiplicative updates proposed by Lee and Seung [1], [2] are the most widely used due to their simplicity and efficiency. Their original updates are restricted to the case where the error is measured by Euclidean distance or I-divergence, but their approach can be applied to a wide class of error functions such as Dual I-divergence, Itakura-Saito divergence and so on [4].

The multiplicative updates are practically useful, but theoretically have a serious drawback: the updates are not defined for all pairs of nonnegative matrices $\boldsymbol{W}$ and $\boldsymbol{H}$. Furthermore, due to this drawback, the global convergence is not theoretically guaranteed. Here, by the global convergence, we mean that for any initial matrices $\boldsymbol{W}^{(0)}$ and $\boldsymbol{H}^{(0)}$ the sequence $\{(\boldsymbol{W}^{(l)}, \boldsymbol{H}^{(l)})\}_{l=0}^{\infty}$ generated by the update contains at least one convergent subsequence and the limit of any convergent subsequence is a stationary point of the corresponding optimization problem. In order to overcome this drawback, Gillis and Glineur [5] proposed a slightly modified version of the multiplicative update for Euclidean distance minimization. The only difference between the original update and the modified one is that the latter returns a user-specified positive constant ϵ if the former returns a positive number smaller than ϵ. The global convergence of the modified update was later proved by Hibi and Takahashi [6], [7] with the help of Zangwill's global convergence theorem [8, p.91].

The objective of this study is to make clear whether or not the modification proposed by Gillis and Glineur can guarantee the global convergence for other multiplicative updates for NMF. However, in this paper, as the first step toward the goal, we focus our attention on the boundedness of the modified versions of multiplicative updates. In particular, we consider eleven multiplicative updates given in [4]. Because the boundedness is a necessary condition in Zangwill's global convergence theorem, the modified update rules with the boundedness property are expected to be globally convergent.

II. Modification of Multiplicative Updates

Let $\boldsymbol{X} = (X_{ij}) \in \mathbb{R}_+^{m\times n}$, $\boldsymbol{W} = (W_{ik}) \in \mathbb{R}_+^{m\times r}$ and $\boldsymbol{H} = (H_{kj}) \in \mathbb{R}_+^{r\times n}$. In this paper, we consider eleven multiplicative update given in [4], which are summarized in Table I. As pointed out in the previous section, they have a common problem: they are not defined for all pairs of nonnegative matrices $\boldsymbol{W}$ and $\boldsymbol{H}$. For example, if all elements of $\boldsymbol{H}$ are zero, the updates cannot be defined. It is certainly true that if both $\boldsymbol{W}$ and $\boldsymbol{H}$ are positive then $\boldsymbol{W}^{\text{new}}$ and $\boldsymbol{H}^{\text{new}}$ are also positive. Hence the update can be iterated infinitely many times. However, it is not guaranteed that both $\boldsymbol{W}$ and $\boldsymbol{H}$ are positive also at the limit point of the sequence generated by the update. If many elements of $\boldsymbol{W}$ and $\boldsymbol{H}$ converge to zero, some elements of $\boldsymbol{W}$ and $\boldsymbol{H}$ may go to infinity.

In order to avoid such a situation, Gillis and Glineur [5] proposed to modify updates as

$$W_{ik}^{\text{new}} = \max\left(\epsilon, f_{ik}(\boldsymbol{W}, \boldsymbol{H})\right), \tag{1}$$

where ϵ is a user-specified positive constant and $f_{ik}(\boldsymbol{W}, \boldsymbol{H})$ represents the right-hand side of each multiplicative update shown in Table I. For example, in the case of Euclidean distance minimization, $f_{ik}(\boldsymbol{W}, \boldsymbol{H})$ is given by

$$f_{ik}(\boldsymbol{W}, \boldsymbol{H}) = W_{ik}\frac{(\boldsymbol{XH}^T)_{ik}}{(\boldsymbol{WHH}^T)_{ik}}\,.$$

This work was partially supported by KAKENHI 24560076, 24500018 and 23310104, and by Proactive Response Against Cyber-attacks Through International Collaborative Exchange (PRACTICE), Ministry of Internal Affairs and Communications, Japan.

TABLE I. MULTIPLICATIVE UPDATE RULES FOR NMF SHOWN IN [4]. $\boldsymbol{Z} = (Z_{ij})$ IS AN $m \times n$ MATRIX DEFINED BY $Z_{ij} = X_{ij}/(\boldsymbol{WH})_{ij}$.

Objective	Update rule for W_{ik}
Euclidean distance	$W_{ik}^{\text{new}} = W_{ik} \frac{(\boldsymbol{XH}^T)_{ik}}{(\boldsymbol{WHH}^T)_{ik}}$
I-divergence	$W_{ik}^{\text{new}} = W_{ik} \frac{(\boldsymbol{ZH}^T)}{\sum_j H_{kj}}$
Dual I-divergence	$W_{ik}^{\text{new}} = W_{ik} \exp\left(\frac{\sum_j (\ln Z_{ij}) H_{kj}}{\sum_j H_{kj}}\right)$
Itakura-Saito divergence	$W_{ik}^{\text{new}} = W_{ik} \sqrt{\frac{\sum_j X_{ij} (\boldsymbol{WH})_{ij}^{-2} H_{kj}}{\sum_j (\boldsymbol{WH})_{ij}^{-1} H_{kj}}}$
α-divergence	$W_{ik}^{\text{new}} = \begin{cases} W_{ik} \left(\frac{\sum_j Z_{ij}^{\alpha} H_{kj}}{\sum_j H_{kj}}\right)^{\frac{1}{\alpha}}, & \alpha \neq 0 \\ W_{ik} \exp\left(\frac{\sum_j (\ln Z_{ij}) H_{kj}}{\sum_j H_{kj}}\right), & \alpha = 0 \end{cases}$
β-divergence	$W_{ik}^{\text{new}} = W_{ik} \left(\frac{\sum_j X_{ij} (\boldsymbol{WH})_{ij}^{\beta-1} H_{kj}}{\sum_j (\boldsymbol{WH})_{ij}^{\beta} H_{kj}}\right)^{\eta}, \eta = \begin{cases} \frac{1}{\beta}, & \beta > 1 \\ 1, & 0 < \beta \leq 1 \\ \frac{1}{1-\beta}, & \beta < 0 \end{cases}$
Log-Quad cost	$W_{ik}^{\text{new}} = W_{ik} \sqrt{\frac{(\boldsymbol{ZH}^T + 2\boldsymbol{XH}^T)_{ik}}{\sum_j H_{kj} + 2(\boldsymbol{WHH}^T)_{ik}}}$
$\alpha\beta$-Bregman divergence	$W_{ik}^{\text{new}} = W_{ik} \left(\frac{\alpha(\alpha-1)\sum_j X_{ij} (\boldsymbol{WH})_{ij}^{\alpha-2} H_{kj} + \beta(1-\beta)\sum_j X_{ij} (\boldsymbol{WH})_{ij}^{\beta-2} H_{kj}}{\alpha(\alpha-1)\sum_j (\boldsymbol{WH})_{ij}^{\alpha-1} H_{kj} + \beta(1-\beta)\sum_j (\boldsymbol{WH})_{ij}^{\beta-1} H_{kj}}\right)^{\frac{1}{\alpha-\beta+1}}$ $(\alpha \geq 1,\ 0 < \beta < 1)$
Kullback-Leibler divergence	$W_{ik}^{\text{new}} = W_{ik} \frac{(\boldsymbol{ZH}^T)_{ik}}{\sum_j H_{kj}} \sum_{ab} (\boldsymbol{WH})_{ab}$
γ-divergence	$W_{ik}^{\text{new}} = W_{ik} \left[\frac{\sum_j X_{ij} (\boldsymbol{WH})_{ij}^{\gamma-1} H_{kj}}{\sum_j (\boldsymbol{WH})_{ij}^{\gamma} H_{kj}} \cdot \frac{\sum_{ab} (\boldsymbol{WH})_{ab}^{1+\gamma}}{\sum_{ab} X_{ab} (\boldsymbol{WH})_{ab}^{\gamma}}\right]^{\eta}, \eta = \begin{cases} \frac{1}{1+\gamma}, & \gamma > 0 \\ \frac{1}{1-\gamma}, & \gamma < 0 \end{cases}$
Rényi divergence	$W_{ik}^{\text{new}} = W_{ik} \left[\frac{\sum_j Z_{ij}^r H_{kj}}{\sum_j H_{kj}} \cdot \frac{\sum_{ab} (\boldsymbol{WH})_{ab}}{\sum_{ab} X_{ab}^r (\boldsymbol{WH})_{ab}^{1-r}}\right]^{\eta}, \eta = \begin{cases} \frac{1}{r}, & r > 1 \\ 1, & 0 < r < 1 \end{cases}$

The basic idea behind the modified update is to prevent W_{ik} to become too small. The right-hand side of (1) returns ϵ if the original update makes W_{ik} smaller than ϵ.

The validity of the modified multiplicative update has been studied by some authors [5]–[7]. For example, Hibi and Takahashi [6], [7] showed that the global convergence of the modified update is guaranteed in the case of Euclidean distance minimization. Using Zangwill's global convergence theorem [8, p.91], they proved that the sequence $\{(\boldsymbol{W}^{(l)}, \boldsymbol{H}^{(l)})\}_{l=0}^{\infty}$ generated by the modified update has at least one convergent subsequence and the limit of any convergent subsequence is a stationary point of the modified optimization problem:

$$\begin{array}{ll} \text{Minimize} & \|\boldsymbol{X} - \boldsymbol{WH}\|^2 \\ \text{Subject to} & \boldsymbol{W} \geq \epsilon \boldsymbol{1}_{m \times r}, \boldsymbol{H} \geq \epsilon \boldsymbol{1}_{r \times n}\,. \end{array} \tag{2}$$

However, it is not clear for other objective functions in Table I whether the global convergence of the modified update is guaranteed or not.

III. BOUNDEDNESS OF MODIFIED UPDATES

When we try to prove the global convergence of an update by applying Zangwill's global convergence theorem, we need to show that the sequence $\{(\boldsymbol{W}^{(l)}, \boldsymbol{H}^{(l)})\}_{l=0}^{\infty}$ is contained in a closed bounded set for each initial value $(\boldsymbol{W}^{(0)}, \boldsymbol{H}^{(0)})$. In this section, we study the boundedness of the modified versions of the multiplicative updates shown in Table I. We will hereafter consider only the updates for $\boldsymbol{W}$. Each of the results given below can be easily applied to the corresponding update for $\boldsymbol{H}$ because it is symmetric to the update for $\boldsymbol{W}$.

We first give a fundamental lemma which will be needed in later discussions.

Lemma 1: Let ϵ be any positive constant. Let f be a map from $[\epsilon, \infty)$ to $\mathbb{R}$ such that

$$\forall x \geq \epsilon, \quad f(x) \leq cx^{\nu}\,, \tag{3}$$

where c is a positive constant and ν is a constant less than 1. Then, any sequence $\{x^{(l)}\}_{l=0}^{\infty}$ generated by the update rule:

$$x^{(l+1)} = \max(\epsilon, f(x^{(l)})), \quad l = 0, 1, 2, \ldots \tag{4}$$

with the initial value $x^{(0)} \geq \epsilon$ is contained in a bounded set.

Proof: We first consider the case where $0 \leq \nu < 1$. Note that a positive number x satisfies $cx^{\nu} \leq x$ if and only if $x \geq c^{\frac{1}{1-\nu}}$. Let l be any nonnegative integer. If $x^{(l)} \geq \max(\epsilon, c^{\frac{1}{1-\nu}})$, we have from (3), (4) and the inequality $c\left(x^{(l)}\right)^{\nu} \leq x^{(l)}$ that

$$\begin{aligned} x^{(l+1)} &= \max(\epsilon, f(x^{(l)})) \\ &\leq \max(\epsilon, c(x^{(l)})^{\nu}) \\ &\leq \max(\epsilon, x^{(l)}) \\ &= x^{(l)}\,. \end{aligned} \tag{5}$$

If $\epsilon \leq x^{(l)} \leq c^{\frac{1}{1-\nu}}$, we have from (3) and (4) that

$$\begin{aligned} x^{(l+1)} &= \max(\epsilon, f(x^{(l)})) \\ &\leq \max(\epsilon, c(x^{(l)})^{\nu}) \\ &\leq \max(\epsilon, c(c^{\frac{1}{1-\nu}})^{\nu}) \\ &= \max(\epsilon, c^{\frac{1}{1-\nu}}) \\ &= c^{\frac{1}{1-\nu}}\,. \end{aligned} \tag{6}$$

By (5) and (6), we have

$$x^{(l+1)} \le \max(x^{(l)}, c^{\frac{1}{1-\nu}}) .$$

Applying the same argument to $x^{(l)}$ in the right-hand side repeatedly, we have

$$\begin{aligned} \max(x^{(l)}, c^{\frac{1}{1-\nu}}) &\le \max(x^{(l-1)}, c^{\frac{1}{1-\nu}}) \\ &\vdots \\ &\le \max(x^{(0)}, c^{\frac{1}{1-\nu}}) . \end{aligned} \tag{7}$$

Therefore, the sequence $\{x^{(l)}\}_{l=0}^{\infty}$ is contained in the closed bounded set $[\epsilon, \max(x^{(0)}, c^{\frac{1}{1-\nu}})]$. We next consider the case where $\nu < 0$. In this case,

$$\begin{aligned} x^{(l+1)} &= \max(\epsilon, f(x^{(l)})) \\ &\le \max(\epsilon, c(x^{(l)})^{\nu}) \\ &\le \max(\epsilon, c\epsilon^{\nu}) \end{aligned}$$

holds for all $l \ge 0$. Therefore, the sequence $\{x^{(l)}\}_{l=0}^{\infty}$ is contained in the closed bounded set $[\epsilon, \max(x^{(0)}, c\epsilon^{\nu})]$. ■

The following theorem is the main result of this paper.

Theorem 1: Let ϵ be any positive constant. Any sequence $\{(\boldsymbol{W}^{(l)}, \boldsymbol{H}^{(l)})\}_{l=0}^{\infty}$ generated by the modified update (1) with $\boldsymbol{W}^{(0)} \ge \epsilon \mathbf{1}_{m\times r}$ and $\boldsymbol{H}^{(0)} \ge \epsilon \mathbf{1}_{r\times n}$ is contained in a bounded set, if the original update is one of the following: 1) Euclidean distance, 2) I-divergence, 3) Dual I-divergence, 4) Itakura-Saito divergence, 5) α-divergence, 6) β-divergence, 7) Log-Quad cost and 8) $\alpha\beta$-Bregman divergence.

Proof: By Lemma 1, it suffices for us to show that if $\boldsymbol{W} \ge \epsilon \mathbf{1}_{m\times r}$ and $\boldsymbol{H} \ge \epsilon \mathbf{1}_{r\times n}$ then there exist a positive constant c and a constant $\nu < 1$ such that

$$f_{ik}(\boldsymbol{W}, \boldsymbol{H}) \le c W_{ik}^{\nu}$$

for each of the eight update rules.

1) Euclidean distance:

$$\begin{aligned} f_{ik}(\boldsymbol{W}, \boldsymbol{H}) &= W_{ik} \sum_{j=1}^{n} \frac{X_{ij} H_{kj}}{\sum_{p=1}^{n} (\boldsymbol{WH})_{ip} H_{kp}} \\ &< W_{ik} \sum_{j=1}^{n} \frac{X_{ij} H_{kj}}{(\boldsymbol{WH})_{ij} H_{kj}} \\ &= W_{ik} \sum_{j=1}^{n} \frac{X_{ij}}{(\boldsymbol{WH})_{ij}} \\ &< W_{ik} \sum_{j=1}^{n} \frac{X_{ij}}{W_{ik} H_{kj}} \\ &= \sum_{j=1}^{n} \frac{X_{ij}}{H_{kj}} \\ &\le \frac{1}{\epsilon} \sum_{j=1}^{n} X_{ij} . \end{aligned}$$

2) I-divergence: By using the same techniques as in 1), we have

$$\begin{aligned} f_{ik}(\boldsymbol{W}, \boldsymbol{H}) &= W_{ik} \frac{(\boldsymbol{Z}\boldsymbol{H}^T)_{ik}}{\sum_{j=1}^{n} H_{kj}} \\ &< W_{ik} \sum_{j=1}^{n} \frac{Z_{ij} H_{kj}}{H_{kj}} \\ &= W_{ik} \sum_{j=1}^{n} \frac{X_{ij}}{(\boldsymbol{WH})_{ij}} \\ &\le \frac{1}{\epsilon} \sum_{j=1}^{n} X_{ij} . \end{aligned}$$

3) Dual I-divergence:

$$\begin{aligned} &f_{ik}(\boldsymbol{W}, \boldsymbol{H}) \\ &= W_{ik} \prod_{j=1}^{n} \exp\left(\frac{H_{kj}}{\sum_{p=1}^{n} H_{kp}} \cdot \ln \frac{X_{ij}}{(\boldsymbol{WH})_{ij}} \right) \\ &< W_{ik} \prod_{j=1}^{n} \exp\left(\frac{H_{kj}}{H_{kj}} \cdot \ln \frac{X_{ij}}{(\boldsymbol{WH})_{ij}} \right) \\ &= W_{ik} \prod_{j=1}^{n} \frac{X_{ij}}{(\boldsymbol{WH})_{ij}} \\ &< W_{ik} \prod_{j=1}^{n} \frac{X_{ij}}{W_{ik} H_{kj}} \\ &< \frac{1}{\epsilon} \prod_{j=1}^{n} X_{ij} . \end{aligned}$$

4) Itakura-Saito divergence:

$$\begin{aligned} f_{ik}(\boldsymbol{W}, \boldsymbol{H}) &= W_{ik} \sqrt{\sum_{j=1}^{n} \frac{X_{ij} (\boldsymbol{WH})_{ij}^{-2} H_{kj}}{\sum_{p=1}^{n} (\boldsymbol{WH})_{ip}^{-1} H_{kp}}} \\ &< W_{ik} \sqrt{\sum_{j=1}^{n} \frac{X_{ij} (\boldsymbol{WH})_{ij}^{-2} H_{kj}}{(\boldsymbol{WH})_{ij}^{-1} H_{kj}}} \\ &= W_{ik} \sqrt{\sum_{j=1}^{n} \frac{X_{ij}}{(\boldsymbol{WH})_{ij}}} \\ &< W_{ik} \sqrt{\sum_{j=1}^{n} \frac{X_{ij}}{W_{ik} H_{kj}}} \\ &< W_{ik}^{\frac{1}{2}} \left(\frac{1}{\epsilon} \sum_{j=1}^{n} X_{ij} \right)^{\frac{1}{2}} . \end{aligned}$$

5) α-divergence: We only consider the case where $\alpha \ne 0$ because the update is identical with Dual I-divergence if $\alpha \to 0$. If $\alpha > 0$ then we have

$$\begin{aligned} f_{ik}(\boldsymbol{W}, \boldsymbol{H}) &= W_{ik} \left(\sum_{j=1}^{n} \frac{Z_{ij}^{\alpha} H_{kj}}{\sum_{p=1}^{n} H_{kp}} \right)^{\frac{1}{\alpha}} \\ &< W_{ik} \left(\sum_{j=1}^{n} Z_{ij}^{\alpha} \right)^{\frac{1}{\alpha}} \\ &= W_{ik} \left(\sum_{j=1}^{n} \left(\frac{X_{ij}}{(\boldsymbol{WH})_{ij}} \right)^{\alpha} \right)^{\frac{1}{\alpha}} \\ &< W_{ik} \left(\sum_{j=1}^{n} \left(\frac{X_{ij}}{W_{ik} H_{kj}} \right)^{\alpha} \right)^{\frac{1}{\alpha}} \end{aligned}$$

$$= W_{ik} \left(\frac{1}{W_{ik}^{\alpha} \epsilon^{\alpha}} \sum_{j=1}^{n} X_{ij}^{\alpha} \right)^{\frac{1}{\alpha}}$$
$$= \frac{1}{\epsilon} \left(\sum_{j=1}^{n} X_{ij}^{\alpha} \right)^{\frac{1}{\alpha}}.$$

If $\alpha < 0$ then we have

$$f_{ik}(\boldsymbol{W}, \boldsymbol{H}) = W_{ik} \left(\sum_{j=1}^{n} \frac{H_{kj}}{\sum_{p=1}^{n} Z_{ip}^{\alpha} H_{kp}} \right)^{-\frac{1}{\alpha}}$$
$$< W_{ik} \left(\sum_{j=1}^{n} Z_{ij}^{-\alpha} \right)^{-\frac{1}{\alpha}}$$
$$< \frac{1}{\epsilon} \left(\sum_{j=1}^{n} X_{ij}^{-\alpha} \right)^{-\frac{1}{\alpha}}.$$

6) β-divergence:

$$f_{ik}(\boldsymbol{W}, \boldsymbol{H}) = W_{ik} \left(\sum_{j=1}^{n} \frac{X_{ij} (\boldsymbol{WH})_{ij}^{\beta-1} H_{kj}}{\sum_{p=1}^{n} (\boldsymbol{WH})_{ip}^{\beta} H_{kp}} \right)^{\eta}$$
$$< W_{ik} \left(\sum_{j=1}^{n} \frac{X_{ij} (\boldsymbol{WH})_{ij}^{\beta-1} H_{kj}}{(\boldsymbol{WH})_{ij}^{\beta} H_{kj}} \right)^{\eta}$$
$$< W_{ik} \left(\sum_{j=1}^{n} \frac{X_{ij}}{(\boldsymbol{WH})_{ij}} \right)^{\eta}$$
$$< W_{ik} \left(\sum_{j=1}^{n} \frac{X_{ij}}{W_{ik} H_{kj}} \right)^{\eta}$$
$$= W_{ik}^{1-\eta} \left(\frac{1}{\epsilon} \sum_{j=1}^{n} X_{ij} \right)^{\eta}.$$

Note that $1 - \eta$ is less than 1 because we easily see $0 < \eta \leq 1$ from the definition of η.

7) Log-Quad cost:

$$f_{ik}(\boldsymbol{W}, \boldsymbol{H})$$
$$< W_{ik} \sqrt{\frac{(\boldsymbol{ZH}^T)_{ik}}{\sum_{j=1}^{n} H_{kj}} + \frac{2(\boldsymbol{XH}^T)_{ik}}{2(\boldsymbol{WHH}^T)_{ik}}}$$
$$= W_{ik} \left(\sum_{j=1}^{n} \frac{Z_{ij} H_{kj}}{\sum_{p=1}^{n} H_{kp}} + \sum_{j=1}^{n} \frac{X_{ij} H_{kj}}{\sum_{p=1}^{n} (\boldsymbol{WH})_{ip} H_{kp}} \right)^{\frac{1}{2}}$$
$$< W_{ik} \left(\sum_{j=1}^{n} Z_{ij} + \sum_{j=1}^{n} \frac{X_{ij}}{(\boldsymbol{WH})_{ij}} \right)^{\frac{1}{2}}$$
$$= W_{ik} \left(2 \sum_{j=1}^{n} \frac{X_{ij}}{(\boldsymbol{WH})_{ij}} \right)^{\frac{1}{2}}$$
$$< W_{ik} \left(\frac{2}{W_{ik}} \sum_{j=1}^{n} \frac{X_{ij}}{H_{kj}} \right)^{\frac{1}{2}}$$
$$< W_{ik}^{\frac{1}{2}} \left(\frac{2}{\epsilon} \sum_{j=1}^{n} X_{ij} \right)^{\frac{1}{2}}.$$

8) $\alpha\beta$-Bregman divergence;

$$f_{ik}(\boldsymbol{W}, \boldsymbol{H})$$
$$< W_{ik} \left(\frac{\alpha(\alpha-1) \sum_{j=1}^{n} X_{ij} (\boldsymbol{WH})_{ij}^{\alpha-2} H_{kj}}{\alpha(\alpha-1) \sum_{j=1}^{n} (\boldsymbol{WH})_{ij}^{\alpha-1} H_{kj}} + \frac{\beta(1-\beta) \sum_{j=1}^{n} X_{ij} (\boldsymbol{WH})_{ij}^{\beta-2} H_{kj}}{\beta(1-\beta) \sum_{j=1}^{n} (\boldsymbol{WH})_{ij}^{\beta-1} H_{kj}} \right)^{\frac{1}{\alpha-\beta+1}}$$
$$< W_{ik} \left(\sum_{j=1}^{n} \frac{X_{ij}}{(\boldsymbol{WH})_{ip}} + \sum_{j=1}^{n} \frac{X_{ij}}{(\boldsymbol{WH})_{ip}} \right)^{\frac{1}{\alpha-\beta+1}}$$
$$< W_{ik}^{1-\frac{1}{\alpha-\beta+1}} \left(\frac{2}{\epsilon} \sum_{j=1}^{n} X_{ij} \right)^{\frac{1}{\alpha-\beta+1}}.$$

Note that $1 - 1/(\alpha - \beta + 1) < 1$ because $\alpha \geq 1$ and $0 < \beta < 1$.

■

In Theorem 1, the last three update rules in Table I are not considered. This is because the boundedness cannot be proved by the approach used in the proof of Theorem 1. However, this does not immediately mean that the modified versions of the last three update rules do not have the boundedness property. We may be able to prove it by using a different technique, or we may need to apply a different way of modification in order to guarantee their boundedness.

IV. Conclusions

We have shown that, for eight multiplicative update rules among eleven rules given in [4], the boundedness is guaranteed if the modification proposed by Gillis and Glineur [5] is applied. However, for the remaining three update rules, that is, rules obtained for Kullback-Leibler divergence, γ-divergence and Rényi divergence, the approach used in this paper cannot prove their boundedness. Further analysis is needed for a better understanding of their boundedness property.

References

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–792, 1999.

[2] ——, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds., vol. 13, 2001, pp. 556–562.

[3] C.-J. Lin, "Projected gradient methods for non-negative matrix factorization," *Neural Computation*, vol. 19, no. 10, pp. 2756–2779, 2007.

[4] Z. Yang and E. Oja, "Unified development of multiplicative algorithm for linear and quadratic nonnegative matrix factorization," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 1878–1891, December 2011.

[5] N. Gillis and F. Glineur, "Nonnegative factorization and the maximum edge biclique problem," *ArXiv e-prints*, October 2008.

[6] R. Hibi and N. Takahashi, "A modified multiplicative update algorithm for Euclidean distance-based nonnegative matrix factorization and its global convergence," in *Proceedings of 18th International Conference on Neural Information Processing, Part-II*, November 2011, pp. 655–662.

[7] N. Takahashi and R. Hibi, "Global convergence of modified multiplicative updates for nonnegative matrix factorization," *Computational Optimization and Applications*, 2013. [Online]. Available: http://dx.doi.org/10.1007/s10589-013-9593-0

[8] W. I. Zangwill, *Nonlinear programming: A unified approach*. Englewood Cliffs, NJ: Prentice-Hall, 1969.