2014 International Symposium on Nonlinear Theory and its Applications
NOLTA2014, Luzern, Switzerland, September 14-18, 2014

New Update Rules based on Kullback-Leibler, Gamma, and Rényi Divergences for Nonnegative Matrix Factorization

Masato Seki† and Norikazu Takahashi†,‡

† Graduate School of Natural Science and Technology, Okayama University
3–1–1 Tsushima-naka, Kita-ku, Okayama 700–8530, Japan
‡ Institute of Systems, Information Technologies and Nanotechnologies
2–1–22 Momochihama, Sawara-ku, Fukuoka 814–0001 Japan
Email: seki@momo.cs.okayama-u.ac.jp, takahashi@cs.okayama-u.ac.jp

Abstract—Three new update rules based on Kullback-Leibler divergence, γ-divergence and Rényi divergence for nonnegative matrix factorization are presented in this paper. An important advantage of these update rules is that they are globally convergent in the sense that any sequence of solutions contains at least one convergent subsequence and the limit of any convergent subsequence is a stationary point of the corresponding optimization problem.

1. Introduction

Nonnegative matrix factorization (NMF) [1] is a technique to decompose a given nonnegative matrix $\boldsymbol{X} \in \mathbb{R}_+^{m\times n}$ into two nonnegative matrices $\boldsymbol{W} \in \mathbb{R}_+^{m\times r}$ and $\boldsymbol{H} \in \mathbb{R}_+^{r\times n}$ such that

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H} \tag{1}$$

where $\mathbb{R}_+$ denotes the set of nonnegative numbers and r is a positive integer less than $\min\{m, n\}$. NMF is usually formulated as an optimization problem to minimize an error function subject to the nonnegativity constraint on each variable. As a simple and efficient method for finding local optimal solutions of such problems, multiplicative updates developed by Lee and Seung [1, 2] are widely used.

Recently, Yang and Oja [3] proposed a unified method for constructing multiplicative update rules for NMF, and derived eleven update rules from eleven error functions. However, all of these rules have a common serious problem that the global convergence is not guaranteed. By the global convergence, we mean that any sequence of solutions has at least one convergent subsequence and the limit of any convergent subsequence is a stationary point of the corresponding optimization problem [4]. One of the reasons for this is that the update rules are not well-defined [5].

In order to avoid this problem, Takahashi *et al.* [6] applied a simple modification technique proposed by Gillis and Glineur [7] to the eleven multiplicative update rules and studied their global convergence. As a result, they proved that eight among the eleven update rules are globally convergent. However, they did not prove the global convergence of the remaining three update rules, that are based on Kullback-Leibler divergence, γ-divergence and Réyni divergence (see Table 1).

In this paper, we propose new update rules based on the above-mentioned three divergences, and prove their global convergence. The key idea is not to use each divergence directly as an error function but to add a penalty term to it. This is based on the observation that if each divergence is directly used as an error function then we can increase the values of variables as much as we want while keeping the value of the error function fixed.

2. Multiplicative Updates for NMF

Suppose that we are given a nonnegative matrix $\boldsymbol{X} \in \mathbb{R}_+^{m\times n}$ and a positive integer $r < \min\{m, n\}$. We will assume throughout this paper that every row and column of $\boldsymbol{X}$ has at least one nonzero entry. Then, the problem of finding $\boldsymbol{W}$ and $\boldsymbol{H}$ in (1) is formulated as

$$\begin{array}{ll} \text{maximize} & D(\boldsymbol{W}, \boldsymbol{H}) \\ \text{subject to} & \boldsymbol{W} \geq \boldsymbol{O}_{m\times r},\ \boldsymbol{H} \geq \boldsymbol{O}_{r\times n} \end{array} \tag{2}$$

where $D(\boldsymbol{W}, \boldsymbol{H})$ is an error function and $\boldsymbol{O}_{m\times r}$ ($\boldsymbol{O}_{r\times n}$, resp.) is the $m\times r$ ($r\times n$, resp.) zero matrix. Throughout this paper, matrix inequalities are understood componentwise.

Multiplicative updates are efficient iterative methods for finding local optimal solutions of the constrained optimization problem (2). Lee and Seung [1, 2] first derived two multiplicative updates based on Euclidean distance and I-divergence by making use of auxiliary functions. This approach was recently generalized by Yang and Oja [3] so that it can be applied to many error functions. In fact, they derived eleven multiplicative updates including the ones of Lee and Seung by using the method summarized below.

1. If the error function $D(\boldsymbol{W}, \boldsymbol{H})$ contains a logarithmic function, rewrite it in a generalized polynomial form by using the relationship:

$$\ln z = \lim_{\delta\to 0+} \frac{z^\delta - 1}{\delta}\,.$$

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

Table 1: Error functions based on Kullback-Leibler, gamma, and Rényi divergences [3].

Error function	$D(\boldsymbol{W},\boldsymbol{H})$
Kullback-Leibler divergence	$\sum_{ij} X_{ij} \ln \frac{X_{ij}}{(\boldsymbol{WH})_{ij}/\sum_{ab}(\boldsymbol{WH})_{ab}}$
γ-divergence	$\frac{1}{\gamma(1+\gamma)}\left(\ln\left(\sum_{ij} X_{ij}^{1+\gamma}\right) + \gamma \ln\left(\sum_{ij}(\boldsymbol{WH})_{ij}^{1+\gamma}\right) - (1+\gamma)\ln\left(\sum_{ij} X_{ij}(\boldsymbol{WH})_{ij}^{\gamma}\right)\right) \quad (\gamma \neq -1, 0)$
Rényi divergence	$\frac{1}{\rho-1}\ln\left(\sum_{ij}\left(\frac{X_{ij}}{\sum_{ab}X_{ab}}\right)^{\rho}\left(\frac{(\boldsymbol{WH})_{ij}}{\sum_{ab}(\boldsymbol{WH})_{ab}}\right)^{1-\rho}\right) \quad (\rho > 0, \rho \neq 1)$

2. Construct an auxiliary function $\bar{D}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}}) : \mathbb{R}_{++}^{m\times r}\times\mathbb{R}_{++}^{r\times n}\times\mathbb{R}_{++}^{m\times r}\times\mathbb{R}_{++}^{r\times n} \to \mathbb{R}$ of the error function $D(\boldsymbol{W},\boldsymbol{H})$, where $\mathbb{R}_{++}$ denotes the set of positive numbers, that satisfies

$$\forall \boldsymbol{W} > \boldsymbol{O}_{m\times r},\ \boldsymbol{H} > \boldsymbol{O}_{r\times n},\ \widetilde{\boldsymbol{W}} > \boldsymbol{O}_{m\times r},\ \widetilde{\boldsymbol{H}} > \boldsymbol{O}_{r\times n},$$
$$\bar{D}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}}) \geq D(\boldsymbol{W},\boldsymbol{H})$$

and

$$\forall \boldsymbol{W} > \boldsymbol{O}_{m\times r},\ \boldsymbol{H} > \boldsymbol{O}_{r\times n},$$
$$\bar{D}(\boldsymbol{W},\boldsymbol{H},\boldsymbol{W},\boldsymbol{H}) = D(\boldsymbol{W},\boldsymbol{H}).$$

3. Let $\widetilde{\boldsymbol{W}}$ and $\widetilde{\boldsymbol{H}}$ be any positive matrices and let $\boldsymbol{W}^*$ be the unique minimizer of the function $\phi(\boldsymbol{W}) = \bar{D}(\boldsymbol{W},\widetilde{\boldsymbol{H}},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})$. Then W_{ik}^* is expressed as a function of $\widetilde{\boldsymbol{W}}$ and $\widetilde{\boldsymbol{H}}$ which gives a multiplicative update rule for W_{ik}. Also, let $\widetilde{\boldsymbol{W}}$ and $\widetilde{\boldsymbol{H}}$ be any positive matrices and let $\boldsymbol{H}^*$ be the unique minimizer of the function $\psi(\boldsymbol{H}) = \bar{D}(\widetilde{\boldsymbol{W}},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})$. Then H_{kj}^* is expressed as a function of $\widetilde{\boldsymbol{W}}$ and $\widetilde{\boldsymbol{H}}$ which gives a multiplicative update rule for H_{kj}.

Table 2 shows three multiplicative update rules obtained by applying this method to the error functions in Table 1.

3. Modified Updates and Their Global Convergence

Every multiplicative update rules presented in [3] is not defined for all pairs of nonnegative matrices $\boldsymbol{W}$ and $\boldsymbol{H}$. This is because each update rule contains a rational function of $\boldsymbol{W}$ and $\boldsymbol{H}$ and the denominator can become zero (see Table 2). A simple way to avoid this problem is to use the modified update rule [7] which can be expressed as

$$W_{ik}^{(l+1)} = \max\left(\epsilon, f_{ik}(\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)})\right), \qquad (3)$$
$$H_{kj}^{(l+1)} = \max\left(\epsilon, g_{kj}(\boldsymbol{W}^{(l+1)},\boldsymbol{H}^{(l)})\right), \qquad (4)$$

where f_{ik} and g_{kj} represent the original updates for W_{ik} and H_{kj}, respectively, and ϵ is a small positive constant specified by the user. In the following, we will focus our attention on the modified update rule. Note that, in so doing, we should not consider the original optimization problem (2) but its modified version:

$$\begin{array}{ll}\text{maximize} & D(\boldsymbol{W},\boldsymbol{H})\\ \text{subject to} & \boldsymbol{W} \geq \epsilon\mathbf{1}_{m\times r},\ \boldsymbol{H} \geq \epsilon\mathbf{1}_{r\times n}\end{array} \qquad (5)$$

where $\mathbf{1}_{m\times r}$ ($\mathbf{1}_{r\times n}$, resp.) is the $m\times r$ ($r\times n$, resp.) matrix consisting of all ones. The feasible region of (5) will be denoted by F_ϵ.

Takahashi *et al.* [6] have recently presented a sufficient condition on the auxiliary function for the modified update rule to be globally convergent in the sense that any sequence $\{(\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)})\}_{l=0}^{\infty}$ generated by (3) and (4) with $(\boldsymbol{W}^{(0)},\boldsymbol{H}^{(0)}) \in F_\epsilon$ contains at least one convergent subsequence and the limit of any convergent subsequence is a stationary point of (5). Also, they have pointed out that the sufficient condition is not satisfied for the three multiplicative update rules shown in Table 2. The main reason for this is that the boundedness of the sequence of solutions generated by (3) and (4) is not guaranteed [5].

4. New Updates based on Three Divergences

In this section, we propose to modify error functions in Table 1 and then derive f_{ik} and g_{kj} by using the method of Yang and Oja [3] so that the modified update rule expressed by (3) and (4) is globally convergent.

By looking at the formulas in Table 1, we can easily see that the value of $D(\boldsymbol{W},\boldsymbol{H})$ does not change if $\boldsymbol{W}$ and $\boldsymbol{H}$ are multiplied by any positive scalars. In other words, we can increase the values of nonzero entries of $\boldsymbol{W}$ and $\boldsymbol{H}$ as much as we want, while keeping the value of $D(\boldsymbol{W},\boldsymbol{H})$ fixed. This may be the reason why boundedness of solutions is not guaranteed. As a simple way to keep $\boldsymbol{W}$ and $\boldsymbol{H}$ bounded, we propose to add a penalty term

$$\frac{C}{2}\left(\sum_{ij} X_{ij} - \sum_{ij}(\boldsymbol{WH})_{ij}\right)^2$$

to each error function, where C is a positive constant. For example, our new error function based on Kullback-Leibler divergence is given by

$$D(\boldsymbol{W},\boldsymbol{H}) = \sum_{ij} X_{ij}\ln\frac{X_{ij}}{(\boldsymbol{WH})_{ij}/\sum_{ab}(\boldsymbol{WH})_{ab}}$$

Table 2: Multiplicative update rules [3] for error functions in Table 1. $\boldsymbol{Z} = (Z_{ij})$ is defined by $Z_{ij} = X_{ij}/(\boldsymbol{WH})_{ij}$.

Error function	Multiplicative updates for W_{ik}
Kullback-Leibler divergence	$W_{ik}^{\text{new}} = W_{ik}\dfrac{(\boldsymbol{ZH}^T)_{ik}}{\sum_j H_{kj}}\sum_{ab}(\boldsymbol{WH})_{ab}$
γ-divergence	$W_{ik}^{\text{new}} = W_{ik}\left(\dfrac{\sum_j X_{ij}(\boldsymbol{WH})_{ij}^{\gamma-1}H_{kj}}{\sum_j(\boldsymbol{WH})_{ij}^{\gamma}H_{kj}}\cdot\dfrac{\sum_{ab}(\boldsymbol{WH})_{ab}^{1+\gamma}}{\sum_{ab}X_{ab}(\boldsymbol{WH})_{ab}^{\gamma}}\right)^{\eta},\quad \eta = \begin{cases}\frac{1}{1+\gamma}, & \text{if } \gamma > 0\\ \frac{1}{1-\gamma}, & \text{if } \gamma < 0 \text{ and } \gamma \neq -1\end{cases}$
Rényi divergence	$W_{ik}^{\text{new}} = W_{ik}\left(\dfrac{\sum_j Z_{ij}^{\rho}H_{kj}}{\sum_j H_{kj}}\cdot\dfrac{\sum_{ab}(\boldsymbol{WH})_{ab}}{\sum_{ab}X_{ab}^{\rho}(\boldsymbol{WH})_{ab}^{1-\rho}}\right)^{\eta},\quad \eta = \begin{cases}\frac{1}{\rho}, & \text{if } \rho > 1\\ 1, & \text{if } 0 < \rho < 1\end{cases}$

$$+\frac{C}{2}\left(\sum_{ij}X_{ij} - \sum_{ij}(\boldsymbol{WH})_{ij}\right)^2 .$$

Let us now apply the unified method of Yang and Oja to this error function. In the first step, we rewrite the error function in the form of a generalized polynomial as follows:

$$\begin{aligned}
&D(\boldsymbol{W},\boldsymbol{H})\\
&= \lim_{\delta\to 0+}\frac{1}{\delta}\Bigg[\sum_{ab}X_{ab}\left(\sum_{ij}(\boldsymbol{WH})_{ij}\right)^{\delta} - \sum_{ij}X_{ij}(\boldsymbol{WH})_{ij}^{\delta}\\
&\qquad + \frac{\delta C}{2}\left(\sum_{ij}X_{ij} - \sum_{ij}(\boldsymbol{WH})_{ij}\right)^2\Bigg] + \text{constant}
\end{aligned}$$

where "constant" represents terms independent of $\boldsymbol{W}$ and $\boldsymbol{H}$. In the second step, we obtain an auxiliary function:

$$\begin{aligned}
&\bar{D}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})\\
&= \lim_{\delta\to 0+}\frac{1}{\delta}\Bigg[\frac{\delta}{2}\sum_{ab}X_{ab}\left(\sum_{ab}(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ab}\right)^{\delta-1}\sum_{ijk}\frac{W_{ik}^2H_{kj}^2}{\widetilde{W}_{ik}\widetilde{H}_{kj}}\\
&\quad - \sum_{ijk}X_{ij}(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ij}^{\delta-1}(\widetilde{W}_{ik}\widetilde{H}_{kj})^{1-\delta}W_{ik}^{\delta}H_{kj}^{\delta}\\
&\quad + \frac{\delta C}{2}\sum_{ab}(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ab}\sum_{ijk}\frac{W_{ik}^2H_{kj}^2}{\widetilde{W}_{ij}\widetilde{H}_{kj}}\\
&\quad - C\sum_{ab}X_{ab}\sum_{ijk}(\widetilde{W}_{ij}\widetilde{H}_{kj})^{1-\delta}W_{ik}^{\delta}H_{kj}^{\delta}\Bigg] + \text{constant}.
\end{aligned}$$

In the last step, we find the unique minimizer of the function $\phi(\boldsymbol{W}) = \bar{D}(\boldsymbol{W},\boldsymbol{H}^{(l)},\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)})$, from which we obtain $f_{ik}(\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)})$. Similarly, we find the unique minimizer of the function $\psi(\boldsymbol{H}) = \bar{D}(\boldsymbol{W}^{(l+1)},\boldsymbol{H},\boldsymbol{W}^{(l+1)},\boldsymbol{H}^{(l)})$, from which we obtain $g_{kj}(\boldsymbol{W}^{(l+1)},\boldsymbol{H}^{(l)})$.

By using the same procedure as above, we can obtain f_{ik} in (3) and g_{kj} in (4) for each of the error functions based on γ-divergence and Rényi divergence. Explicit formulas of f_{ik} and g_{kj} are given in Table 3.

5. Global Convergence of New Updates

Let us consider the boundedness of the sequence of solutions generated by the three update rules obtained in the previous section.

Theorem 1 Let ϵ and C be any positive numbers. For any initial matrices $\boldsymbol{W}^{(0)} \geq \epsilon\boldsymbol{1}_{m\times r}$ and $\boldsymbol{H}^{(0)} \geq \epsilon\boldsymbol{1}_{r\times n}$, the sequence $\{(\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)})\}_{l=0}^{\infty}$ generated by any of the three update rules given in the previous section is bounded.

Proof We prove this by using Lemma 1 of [5]. To do so, it suffices for us to show that there exist constants c_{ik} (> 0) and ν_{ik} (< 1) such that

$$\forall\boldsymbol{W} \geq \epsilon\boldsymbol{1}_{m\times r},\ \forall\boldsymbol{H} \geq \epsilon\boldsymbol{1}_{r\times n},\quad f_{ik}(\boldsymbol{W},\boldsymbol{H}) \leq c_{ik}W_{ik}^{\nu_{ik}} \tag{6}$$

for all pairs of i and k and there exist constants d_{kj} (> 0) and μ_{kj} (< 1) such that

$$\forall\boldsymbol{W} \geq \epsilon\boldsymbol{1}_{m\times r},\ \forall\boldsymbol{H} \geq \epsilon\boldsymbol{1}_{r\times n},\quad g_{kj}(\boldsymbol{W},\boldsymbol{H}) \leq d_{kj}H_{kj}^{\mu_{kj}} \tag{7}$$

for all pairs of k and j. Let us consider the new update rule based on Kullback-Leibler divergence. Since

$$\begin{aligned}
&f_{ik}(\boldsymbol{W},\boldsymbol{H})\\
&< W_{ik}\left(\frac{\sum_j X_{ij}(\boldsymbol{WH})_{ij}^{-1}H_{kj} + C\sum_{ab}X_{ab}\sum_j H_{kj}}{C\sum_{ab}(\boldsymbol{WH})_{ab}\sum_j H_{kj}}\right)^{\frac{1}{2}}\\
&< W_{ik}\left(\frac{\frac{1}{r\epsilon^2}\sum_j X_{ij}H_{kj} + C\sum_{ab}X_{ab}\sum_j H_{kj}}{C\sum_{ab}(\boldsymbol{WH})_{ab}\sum_j H_{kj}}\right)^{\frac{1}{2}}\\
&= W_{ik}\left(\sum_j\frac{\frac{1}{r\epsilon^2}X_{ij}H_{kj} + C\sum_{ab}X_{ab}H_{kj}}{C\sum_{ab}(\boldsymbol{WH})_{ab}\sum_l H_{kl}}\right)^{\frac{1}{2}}\\
&< W_{ik}\left(\sum_j\frac{\frac{1}{r\epsilon^2}X_{ij}H_{kj} + C\sum_{ab}X_{ab}H_{kj}}{CW_{ik}H_{kj}n\epsilon}\right)^{\frac{1}{2}}\\
&= W_{ik}^{\frac{1}{2}}\left(\sum_j\frac{\frac{1}{r\epsilon^2}X_{ij} + C\sum_{ab}X_{ab}}{Cn\epsilon}\right)^{\frac{1}{2}},
\end{aligned}$$

the condition (6) is satisfied. In the same way, we can show that the condition (7) is also satisfied. As for the update rules based on γ-divergence and Rényi divergence, we can also prove that both (6) and (7) are satisfied. However, we omit the details due to lack of space. □

Table 3: f_{ik} and g_{kj} in new update rules.

Error function	$f_{ik}(\boldsymbol{W},\boldsymbol{H})$ and $g_{kj}(\boldsymbol{W},\boldsymbol{H})$
Kullback-Leibler divergence	$f_{ik}(\boldsymbol{W},\boldsymbol{H}) = W_{ik}\left(\frac{\sum_j X_{ij}(\boldsymbol{W}\boldsymbol{H})_{ij}^{-1}H_{kj} + C\sum_{ab}X_{ab}\sum_j H_{kj}}{\sum_{ab}X_{ab}(\sum_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab})^{-1}\sum_j H_{kj} + C\sum_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab}\sum_j H_{kj}}\right)^{\frac{1}{2}}$
	$g_{kj}(\boldsymbol{W},\boldsymbol{H}) = H_{kj}\left(\frac{\sum_i X_{ij}(\boldsymbol{W}\boldsymbol{H})_{ij}^{-1}W_{ik} + C\sum_{ab}X_{ab}\sum_i W_{ik}}{\sum_{ab}X_{ab}(\sum_{ab}(\boldsymbol{W}\boldsymbol{H}_{ab}))^{-1}\sum_i W_{ik} + C\sum_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab}\sum_i W_{ik}}\right)^{\frac{1}{2}}$
γ-divergence	$f_{ik}(\boldsymbol{W},\boldsymbol{H}) = W_{ik}\left(\frac{(\sum_{ab}X_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab}^{\gamma})^{-1}\sum_j X_{ij}(\boldsymbol{W}\boldsymbol{H})_{ij}^{\gamma-1}H_{kj} + C\sum_{ab}X_{ab}\sum_j H_{kj}}{(\sum_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab}^{1+\gamma})^{-1}\sum_j(\boldsymbol{W}\boldsymbol{H})_{ij}^{\gamma}H_{kj} + C\sum_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab}\sum_j H_{kj}}\right)^{\eta}$
	$g_{kj}(\boldsymbol{W},\boldsymbol{H}) = H_{kj}\left(\frac{(\sum_{ab}X_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab}^{\gamma})^{-1}\sum_i X_{ij}(\boldsymbol{W}\boldsymbol{H})_{ij}^{\gamma-1}W_{ik} + C\sum_{ab}X_{ab}\sum_i W_{ik}}{(\sum_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab}^{1+\gamma})^{-1}\sum_i(\boldsymbol{W}\boldsymbol{H})_{ij}^{\gamma}W_{ik} + C\sum_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab}\sum_i W_{ik}}\right)^{\eta}$
	where $\eta = \begin{cases} 1/(1+\gamma), & \text{if } \gamma > 1 \\ 1/2, & \text{if } 0 < \gamma < 1 \\ 1/(2-\gamma), & \text{if } \gamma < 0 \text{ and } \gamma \neq -1 \end{cases}$
Rényi divergence	$f_{ik}(\boldsymbol{W},\boldsymbol{H}) = W_{ik}\left(\frac{(\sum_{ab}X_{ab}^{\rho}(\boldsymbol{W}\boldsymbol{H})_{ab}^{1-\rho})^{-1}\sum_j X_{ij}^{\rho}(\boldsymbol{W}\boldsymbol{H})_{ij}^{-\rho}H_{kj} + C\sum_{ab}X_{ab}\sum_j H_{kj}}{(\sum_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab})^{-1}\sum_j H_{kj} + C\sum_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab}\sum_j H_{kj}}\right)^{\eta}$
	$g_{kj}(\boldsymbol{W},\boldsymbol{H}) = H_{kj}\left(\frac{(\sum_{ab}X_{ab}^{\rho}(\boldsymbol{W}\boldsymbol{H})_{ab}^{1-\rho})^{-1}\sum_i X_{ij}^{\rho}(\boldsymbol{W}\boldsymbol{H})_{ij}^{-\rho}W_{ik} + C\sum_{ab}X_{ab}\sum_i W_{ik}}{(\sum_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab})^{-1}\sum_i W_{ik} + C\sum_{ab}(\boldsymbol{W}\boldsymbol{H})_{ab}\sum_i W_{ik}}\right)^{\eta}$
	where $\eta = \begin{cases} 1/(1+\rho), & \text{if } \rho > 1 \\ 1/2, & \text{if } 0 < \rho < 1 \end{cases}$

We finally consider the global convergence of the new update rules. Because all auxiliary functions used for obtaining the three update rules satisfy the condition in Theorem 1 of [6], we have the following result.

Theorem 2 Let ϵ and C be any positive numbers. For any initial matrices $\boldsymbol{W}^{(0)} \geq \epsilon\boldsymbol{1}_{m\times r}$ and $\boldsymbol{H}^{(0)} \geq \epsilon\boldsymbol{1}_{r\times n}$, the sequence $\{(\boldsymbol{W}^{(l)}, \boldsymbol{H}^{(l)})\}_{l=0}^{\infty}$ generated by any of the three update rules given in the previous section has at least one convergent subsequence and the limit of any convergent subsequence is a stationary point of the problem (5).

6. Conclusion

We have proposed three new error functions based on Kullback-Leibler divergence, γ-divergence and Rényi divergence for NMF, and derived three new update rules. We have also shown that all of these update rules have a global convergence property. However, it is not clear how the solutions obtained by the proposed update rules depends on two parameters ϵ and C. We leave this as a future problem.

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] D. D. Lee and H. S. Seung, "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, pp. 556–562, 2001.

[3] Z. Yang and E. Oja, "Unified development of multiplicative algorithm for linear and quadratic nonnegative matrix factorization," *IEEE Transactions on Neural Networks*, vol. 22, pp. 1878–1891, December 2011.

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

[5] J. Katayama, N. Takahashi, and J. Takeuchi, "Boundedness of modified multiplicative updates for nonnegative matrix factorization," in *Proceedings of the Fifth International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, pp. 252–255, December 2013.

[6] N. Takahashi, J. Katayama, and J. Takeuchi, "A generalized sufficient condition for global convergence of modified multiplicative updates for NMF." to appear in the Proceedings of 2014 International Symposium on Nonlinear Theory and its Applications.

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