2017 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2017, Cancun, Mexico, December 4-7, 2017

Derivation of Multiplicative Update Rules for Nonnegative Matrix Factorization with Regularization Terms

Akihiro Koso and Norikazu Takahashi

Graduate School of Natural Science and Technology
3–1–1 Tsushima-naka, Kita-ku, Okayama 700–8530, Japan
Email: kouso@momo.cs.okayama-u.ac.jp, takahashi@cs.okayama-u.ac.jp

Abstract—Nonnegative Matrix Factorization (NMF) is an operation that decomposes a given nonnegative matrix into two nonnegative factor matrices. NMF is usually formulated as a constrained optimization problem in which an objective function has to be minimized under the constraint that all variables are nonnegative, and the multiplicative update rules are widely used for solving this problem. In this paper, we give a unified method for deriving a multiplicative update rule from a given objective function including regularization terms. We then apply it to 22 objective functions obtained by adding two types of regularization terms to 11 error functions.

1. Introduction

Nonnegative Matrix Factorization (NMF) [1, 2, 3] is an operation that decomposes a given nonnegative matrix $\boldsymbol{X} \in \mathbb{R}_+^{M\times N}$ ($\mathbb{R}_+$ denotes the set of nonnegative real numbers) into two nonnegative matrices $\boldsymbol{W} \in \mathbb{R}_+^{M\times K}$ and $\boldsymbol{H} \in \mathbb{R}_+^{K\times N}$ such that $\boldsymbol{WH}$ is approximately equal to $\boldsymbol{X}$ (see Fig. 1). NMF is usually formulated as optimization problems of the form:

$$\begin{array}{ll} \text{minimize} & D(\boldsymbol{W},\boldsymbol{H}) \\ \text{subject to} & \boldsymbol{W} \geq \boldsymbol{0},\ \boldsymbol{H} \geq \boldsymbol{0}, \end{array} \tag{1}$$

where $D(\boldsymbol{W},\boldsymbol{H})$ represents an error between $\boldsymbol{X}$ and $\boldsymbol{WH}$. The Euclidean distance $\|\boldsymbol{X} - \boldsymbol{WH}\|_F^2$ is often used for $D(\boldsymbol{W},\boldsymbol{H})$, but many other error functions can also be used (see [4] for example).

Since $D(\boldsymbol{W},\boldsymbol{H})$ is not convex in general, it is difficult to find an optimal solution of (1). The most well-known approach to finding a local optimal solution is the multiplicative update rule [2, 3]. In this approach, $\boldsymbol{W}$ and $\boldsymbol{H}$ are updated alternatively. When updating $\boldsymbol{W}$ ($\boldsymbol{H}$, resp.), $\boldsymbol{H}$ ($\boldsymbol{W}$, resp.) is fixed and $\boldsymbol{W}$ ($\boldsymbol{H}$, resp.) is updated to the minimum point of an auxiliary function of $D(\boldsymbol{W},\boldsymbol{H})$. For example, the multiplicative update rule for the Euclidean distance is given by

$$W_{ik}^{\text{new}} = W_{ik}\frac{(\boldsymbol{XH}^{\mathrm{T}})_{ik}}{(\boldsymbol{WHH}^{\mathrm{T}})_{ik}}, \quad H_{kj}^{\text{new}} = H_{kj}\frac{(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X})_{kj}}{(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{WH})_{kj}}.$$

The idea behind the multiplicative update rule can be applied to a wide class of error functions. In fact, Yang and Oja [4] proposed a unified method for deriving multiplicative update rules and obtained 11 update rules from 11 error functions shown in Table 1.

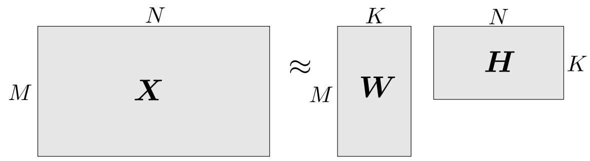

Figure 1: Nonnegative matrix factorization.

In some applications of NMF, it is desired that $\boldsymbol{W}$ and $\boldsymbol{H}$ are smooth or sparse. In order to control the smoothness or sparseness of $\boldsymbol{W}$ and $\boldsymbol{H}$, Cichocki *et al.* [5] proposed to add regularization terms $C_1J_1(\boldsymbol{W})$ and $C_2J_2(\boldsymbol{H})$ to the Euclidean distance based error function $\frac{1}{2}\|\boldsymbol{X}-\boldsymbol{WH}\|_F^2$ and derived a multiplicative update rule. However, the validity of this update rule was not clearly shown.

In this paper, we consider the objective functions of the form:

$$D(\boldsymbol{W},\boldsymbol{H}) + C_1\sum_{ik} W_{ik}^b + C_2\sum_{kj} H_{kj}^b, \tag{2}$$

where $b \in \{1,2\}$. We apply the unified method of Yang and Oja to 22 objective functions obtained from the 11 error functions shown in Table 1, and show that the multiplicative update rule can be obtained for all cases except one.

2. Unified Method for Deriving Multiplicative Update Rules

In this section, we review the unified method proposed by Yang and Oja [4] for deriving multiplicative update rules from various error functions. Let $\mathcal{F} = \{(\boldsymbol{W},\boldsymbol{H}) \mid \boldsymbol{W} \geq \boldsymbol{0},\ \boldsymbol{H} \geq \boldsymbol{0}\} = \mathbb{R}_+^{M\times K}\times\mathbb{R}_+^{K\times N}$ and let $\operatorname{int}\mathcal{F}$ denote the interior of $\mathcal{F}$. That is, $\operatorname{int}\mathcal{F} = \mathbb{R}_{++}^{M\times K}\times\mathbb{R}_{++}^{K\times N}$ where $\mathbb{R}_{++}$ is the set of positive real numbers.

We first give the definition of the auxiliary function.

Definition 1 (Auxiliary Function) For a given function $D(\boldsymbol{W},\boldsymbol{H}) : \operatorname{int}\mathcal{F} \to \mathbb{R}$, any function $\bar{D}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}}) : \operatorname{int}\mathcal{F}\times\operatorname{int}\mathcal{F} \to \mathbb{R}$ that satisfies the following two conditions is called an auxiliary function of $D(\boldsymbol{W},\boldsymbol{H})$.

1. $\bar{D}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}}) \geq D(\boldsymbol{W},\boldsymbol{H})$ for all $(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}}) \in \operatorname{int}\mathcal{F}\times\operatorname{int}\mathcal{F}$.

Table 1: Error functions considered by Yang and Oja [4]. As for the last three error functions, we have applied the modification proposed by Seki and Takahashi [6].

Name	$D(\boldsymbol{W},\boldsymbol{H})$
Euclidean distance	$\sum_{ij}(X_{ij}-(\boldsymbol{WH})_{ij})^2$
I-divergence	$\sum_{ij}\left(X_{ij}\ln\left(\frac{X_{ij}}{(\boldsymbol{WH})_{ij}}\right)-X_{ij}+(\boldsymbol{WH})_{ij}\right)$
Dual I-divergence	$\sum_{ij}\left((\boldsymbol{WH})_{ij}\ln\left(\frac{(\boldsymbol{WH})_{ij}}{X_{ij}}\right)-(\boldsymbol{WH})_{ij}+X_{ij}\right)$
Itakura-Saito divergence	$\sum_{ij}\left(-\ln\left(\frac{X_{ij}}{(\boldsymbol{WH})_{ij}}\right)+\frac{X_{ij}}{(\boldsymbol{WH})_{ij}}-1\right)$
α-divergence	$\frac{1}{\alpha(1-\alpha)}\sum_{ij}\left(\alpha X_{ij}+(1-\alpha)(\boldsymbol{WH})_{ij}-X_{ij}^{\alpha}(\boldsymbol{WH})_{ij}^{1-\alpha}\right)$ $(\alpha\neq 0,1)$
β-divergence	$\sum_{ij}\left(X_{ij}\frac{X_{ij}^{\beta}-(\boldsymbol{WH})_{ij}^{\beta}}{\beta}-\frac{X_{ij}^{\beta+1}-(\boldsymbol{WH})_{ij}^{\beta+1}}{\beta+1}\right)$ $(\beta\neq 0,-1)$
Log-Quad cost	$\sum_{ij}\left((X_{ij}-(\boldsymbol{WH})_{ij})^2+X_{ij}\ln\left(\frac{X_{ij}}{(\boldsymbol{WH})_{ij}}\right)-X_{ij}+(\boldsymbol{WH})_{ij}\right)$
$\alpha\beta$-Bregman divergence	$\sum_{ij}\Big[X_{ij}^{\alpha}-X_{ij}^{\beta}-(\boldsymbol{WH})_{ij}^{\alpha}+(\boldsymbol{WH})_{ij}^{\beta}$ $-\left(\alpha(\boldsymbol{WH})_{ij}^{\alpha-1}-\beta(\boldsymbol{WH})_{ij}^{\beta-1}\right)\left(X_{ij}-(\boldsymbol{WH})_{ij}\right)\Big]$ $(\alpha\geq 1, 0<\beta<1)$
Kullback-Leibler divergence	$\sum_{ij}\frac{X_{ij}}{\sum_{pq}X_{pq}}\ln\left(\frac{X_{ij}/\sum_{pq}X_{pq}}{(\boldsymbol{WH})_{ij}/\sum_{pq}(\boldsymbol{WH})_{pq}}\right)+\frac{C_0}{2}\left(\sum_{ij}(\boldsymbol{WH})_{ij}-\sum_{ij}X_{ij}\right)^2$
γ-divergence	$\frac{1}{\gamma(1+\gamma)}\left(\ln\left(\sum_{ij}X_{ij}^{1+\gamma}\right)+\gamma\ln\left(\sum_{ij}(\boldsymbol{WH})_{ij}^{1+\gamma}\right)-(1+\gamma)\ln\left(\sum_{ij}X_{ij}(\boldsymbol{WH})_{ij}^{\gamma}\right)\right)$ $+\frac{C_0}{2}\left(\sum_{ij}(\boldsymbol{WH})_{ij}-\sum_{ij}X_{ij}\right)^2$ $(\gamma\neq 0,-1)$
Renyi divergence	$\frac{1}{\rho-1}\ln\left(\sum_{ij}\left(\frac{X_{ij}}{\sum_{pq}X_{pq}}\right)^{\rho}\left(\frac{(\boldsymbol{WH})_{ij}}{\sum_{pq}(\boldsymbol{WH})_{pq}}\right)^{1-\rho}\right)+\frac{C_0}{2}\left(\sum_{ij}(\boldsymbol{WH})_{ij}-\sum_{ij}X_{ij}\right)^2$ $(\rho>0,\rho\neq 1)$

2. $\bar{D}(\boldsymbol{W},\boldsymbol{H},\boldsymbol{W},\boldsymbol{H}) = D(\boldsymbol{W},\boldsymbol{H})$ for all $(\boldsymbol{W},\boldsymbol{H}) \in \operatorname{int}\mathcal{F}$.

Let $\bar{D}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})$ be an auxiliary function of $D(\boldsymbol{W},\boldsymbol{H})$. Let $\{(\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)})\}_{l=0}^{\infty}$ be a sequence satisfying the following three conditions.

1. $(\boldsymbol{W}^{(0)},\boldsymbol{H}^{(0)})$ belongs to $\operatorname{int}\mathcal{F}$.
2. For each $l \geq 0$, $\boldsymbol{W}^{(l+1)}$ is an optimal solution of the optimization problem:

$$\begin{array}{ll}\text{minimize} & \bar{D}(\boldsymbol{W},\boldsymbol{H}^{(l)},\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)})\\ \text{subject to} & \boldsymbol{W} > \boldsymbol{O}_{m\times r}.\end{array} \tag{3}$$

3. For each $l \geq 0$, $\boldsymbol{H}^{(l+1)}$ is an optimal solution of the optimization problem:

$$\begin{array}{ll}\text{minimize} & \bar{D}(\boldsymbol{W}^{(l+1)},\boldsymbol{H},\boldsymbol{W}^{(l+1)},\boldsymbol{H}^{(l)})\\ \text{subject to} & \boldsymbol{H} > \boldsymbol{O}_{r\times n}.\end{array} \tag{4}$$

Then $\{D(\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)})\}_{l=0}^{\infty}$ is a nonincreasing sequence because we have $D(\boldsymbol{W}^{(l+1)},\boldsymbol{H}^{(l+1)}) \leq \bar{D}(\boldsymbol{W}^{(l+1)},\boldsymbol{H}^{(l+1)},\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)}) \leq \bar{D}(\boldsymbol{W}^{(l+1)},\boldsymbol{H}^{(l)},\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)}) \leq \bar{D}(\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)},\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)}) = D(\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)})$. Furthermore, if $D(\boldsymbol{W},\boldsymbol{H})$ is bounded below on $\operatorname{int}\mathcal{F}$, the sequence $D(\boldsymbol{W}^{(l)},\boldsymbol{H}^{(l)})$ converges to some constant as l goes to infinity.

The unified method of Yang and Oja [4] is described as follows.

1. If the objective function contains a logarithm, replace it with a generalized polynomial by using

$$\ln x = \lim_{\mu\to 0^+}\frac{x^{\mu}-1}{\mu}.$$

2. Obtain an auxiliary function of the objective function by applying Theorems 1–3 [1] given below. Then take the limit $\mu \to 0^+$ if necessary. One may need to apply L'Hôpital's rule when taking the limit.
3. Find the optimal solutions of (3) and (4).

Theorem 1 (Yang and Oja [4]) Let

$$D(\boldsymbol{W},\boldsymbol{H}) = a\left(\sum_{ij} b_{ij}(\boldsymbol{WH})_{ij}^{c}\right)^{d}$$

where $a \neq 0$, $b_{ij} > 0$, $c \neq 0$ and $d \neq 1$. Then $\bar{D}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})$ defined as follows is an auxiliary function of $D(\boldsymbol{W},\boldsymbol{H})$.

1. If $g(x) = ax^d$ is convex for $x > 0$, let

$$\bar{D}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}}) = a\left(\sum_{ij} b_{ij}(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ij}^{c}\right)^{d-1} \times \sum_{ij} b_{ij}(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ij}^{c}\left(\frac{(\boldsymbol{WH})_{ij}}{(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ij}}\right)^{cd}.$$

2. If $g(x) = ax^d$ is concave for $x > 0$, let

$$\begin{aligned}&\bar{D}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})\\ &= a\left(\sum_{ij} b_{ij}(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ij}^{c}\right)^{d} + ad\left(\sum_{ij} b_{ij}(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ij}^{c}\right)^{d-1}\\ &\times\left(b_{ij}\sum_{ij}(\boldsymbol{WH})_{ij}^{c} - b_{ij}\sum_{ij}(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ij}^{c}\right).\end{aligned}$$

[1] Proof of these theorems are given by Takahashi *et al.* [7]

Theorem 2 (Yang and Oja [4]) Let

$$D(\boldsymbol{W},\boldsymbol{H}) = \sum_{ij} a_{ij}(\boldsymbol{W}\boldsymbol{H})_{ij}^{b}$$

where $a_{ij} \neq 0$. We define $\bar{D}_{ij}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})$ as follows.

1. If $g_{ij}(x) = a_{ij}x^b$ is convex for $x > 0$, let

$$\begin{aligned}&\bar{D}_{ij}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})\\&\quad = a_{ij}(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ij}^{b-1}\sum_{k}(\widetilde{W}_{ik}\widetilde{H}_{kj})^{1-b}(W_{ik}H_{kj})^{b}.\end{aligned}$$

2. If $g_{ij}(x) = a_{ij}x^b$ is concave for $x > 0$, let

$$\begin{aligned}\bar{D}_{ij}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}}) &= a_{ij}(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ij}^{b}\\&+ a_{ij}b(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ij}^{b-1}\left((\boldsymbol{W}\boldsymbol{H})_{ij} - (\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{H}})_{ij}\right).\end{aligned}$$

Then $\sum_{ij}\bar{D}_{ij}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})$ is an auxiliary function of $D(\boldsymbol{W},\boldsymbol{H})$.

Theorem 3 (Yang and Oja [4]) Let

$$D(\boldsymbol{W},\boldsymbol{H}) = \sum_{t}\sum_{ijk} a_{tijk}(W_{ik}H_{kj})^{b_t}$$

where $a_{tijk} \neq 0$. We also assume that $D_{tijk}(\boldsymbol{W},\boldsymbol{H}) = a_{tijk}(W_{ik}H_{kj})^{b_t}$ is convex on $\operatorname{int}\mathcal{F}$ for all i, j, k, t and that $\{b_t\}$ contains at least two distinct nonzero numbers. Let $b_{\max} = \max\{b_t \mid b_t \neq 0\}$ and $b_{\min} = \min\{b_t \mid b_t \neq 0\}$. We define $\bar{D}_{tijk}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})$ as follows.

1. If $b_t \in \{b_{\min}, b_{\max}, 0\}$, let

$$\bar{D}_{tijk}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}}) = a_{tijk}(W_{ik}H_{kj})^{b_t}.$$

2. If $b_t \notin \{b_{\min}, b_{\max}, 0\}$ and

   (a) if $(b_t > 1) \vee ((b_t = 1) \wedge (a_{tijk} > 0))$, let

$$\begin{aligned}&\bar{D}_{tijk}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})\\&\quad = \frac{a_{tijk}b_t}{b_{\max}}(\widetilde{W}_{ik}\widetilde{H}_{kj})^{b_t-b_{\max}}(W_{ik}H_{kj})^{b_{\max}}\\&\quad\quad + a_{tijk}(\widetilde{W}_{ik}\widetilde{H}_{kj})^{b_t}\left(1 - \frac{b_t}{b_{\max}}\right),\end{aligned}$$

   (b) if $(b_t < 1) \vee ((b_t = 1) \wedge (a_{tijk} < 0))$, let

$$\begin{aligned}&\bar{D}_{tijk}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})\\&\quad = \frac{a_{tijk}b_t}{b_{\min}}(\widetilde{W}_{ik}\widetilde{H}_{kj})^{b_t-b_{\min}}(W_{ik}H_{kj})^{b_{\min}}\\&\quad\quad + a_{tijk}(\widetilde{W}_{ik}\widetilde{H}_{kj})^{b_t}\left(1 - \frac{b_t}{b_{\min}}\right).\end{aligned}$$

Then $\sum_{tijk}\bar{D}_{tijk}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})$ is an auxiliary function of $D(\boldsymbol{W},\boldsymbol{H})$ and strictly convex in $\operatorname{int}\mathcal{F}$.

3. Derivation of Multiplicative Update Rules for NMF with Regularization

We now consider objective functions given by (2), and derive multiplicative update rules by using the unified method of Yang and Oja. We assume that C_1 and C_2 are positive constants and b is either 1 or 2. Because each of the regularization terms contains only $\boldsymbol{W}$ or $\boldsymbol{H}$, we cannot apply Theorems 1–3 directly. We thus provide a new tool for deriving auxiliary functions.

Corollary 1 Let

$$D(\boldsymbol{W},\boldsymbol{H}) = \sum_{t}\sum_{ijk} a_{tijk}(W_{ik}H_{kj})^{b_t} + C_1\sum_{ik}W_{ik}^{b} + C_2\sum_{kj}H_{kj}^{b}$$

where $a_{tijk} \neq 0$, $C_1 > 0$, $C_2 > 0$ and $b \in \{1, 2\}$. We also assume that $D_{tijk}(\boldsymbol{W},\boldsymbol{H}) = a_{tijk}(W_{ik}H_{kj})^{b_t}$ is convex on $\operatorname{int}\mathcal{F}$ for all i, j, k, t and that $\{b_t\}$ contains at least two distinct nonzero numbers. Let $b_{\max} = \max\{b, \max\{b_t \mid b_t \neq 0\}\}$ and $b_{\min} = \min\{b, \min\{b_t \mid b_t \neq 0\}\}$. We define $\bar{D}_{tijk}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})$ as described in Theorem 3. We also define $E(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})$ as follows.

1. If $b \in \{b_{\min}, b_{\max}, 0\}$, let

$$E(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}}) = C_1\sum_{ik}W_{ik}^{b} + C_2\sum_{kj}H_{kj}^{b}.$$

2. If $b \notin \{b_{\min}, b_{\max}, 0\}$, let

$$\begin{aligned}&E(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})\\&= \frac{C_1 b}{b_{\max}}\sum_{ik}\widetilde{W}_{ik}^{b-b_{\max}}W_{ik}^{b_{\max}} + C_1\left(1 - \frac{b}{b_{\max}}\right)\sum_{ik}\widetilde{W}_{ik}^{b}\\&+ \frac{C_2 b}{b_{\max}}\sum_{kj}\widetilde{H}_{kj}^{b-b_{\max}}H_{kj}^{b_{\max}} + C_2\left(1 - \frac{b}{b_{\max}}\right)\sum_{kj}\widetilde{H}_{kj}^{b}.\end{aligned}$$

Then $\sum_{tijk}\bar{D}_{tijk}(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}}) + E(\boldsymbol{W},\boldsymbol{H},\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{H}})$ is an auxiliary function of $D(\boldsymbol{W},\boldsymbol{H})$ and strictly convex in $\operatorname{int}\mathcal{F}$.

For each of the 22 objective functions obtained from the 11 error functions shown in Table 1, we apply Theorems 1 and 2 to the first term of (2) and then apply Corollary 1 to all terms to obtain a multiplicative update rule. Obtained update rules for some popular error functions are shown in Tables 2 and 3. The only objective function we could not obtain the multiplicative update rule is the dual I-divergence combined with the regularization terms with $b = 2$. In this case, the auxiliary function could not be obtained because the limit is not well-defined.

4. Conclusion

We have derived various multiplicative update rules for NMF with regularization by applying the unified method of Yang and Oja. Future problems are to study global convergence of these update rules and to derive a multiplicative update rule for the objective function based on the dual I-divergence and the regularization terms with $b = 2$.

Table 2: Some examples of multiplicative update rules for $b = 1$.

Error	Update rule for W_{ik}
Euclidean distance	$W_{ik}^{\mathrm{new}} = W_{ik} \frac{\sum_j X_{ij} H_{kj} - C_1/2}{\sum_j (\boldsymbol{W}\boldsymbol{H})_{ij} H_{kj}}$
I-divergence	$W_{ik}^{\mathrm{new}} = W_{ik} \frac{\sum_j X_{ij} (\boldsymbol{W}\boldsymbol{H})_{ij}^{-1} H_{kj}}{\sum_j H_{kj} + C_1}$
Itakura-Saito divergence	$W_{ik}^{\mathrm{new}} = W_{ik} \left(\frac{\sum_j X_{ij} (\boldsymbol{W}\boldsymbol{H})_{ij}^{-2} H_{kj}}{\sum_j (\boldsymbol{W}\boldsymbol{H})_{ij}^{-1} H_{kj} + C_1} \right)^{\frac{1}{2}}$
α-divergence	$W_{ik}^{\mathrm{new}} = W_{ik} \left(\frac{\sum_j X_{ij}^{\alpha} (\boldsymbol{W}\boldsymbol{H})_{ij}^{-\alpha} H_{kj}}{\sum_j H_{kj} + \alpha C_1} \right)^{\frac{1}{\alpha}}$
Kullback-Leibler divergence	$W_{ik}^{\mathrm{new}} = W_{ik} \left(\frac{(\sum_{pq} X_{pq})^{-1} \sum_j X_{ij} (\boldsymbol{W}\boldsymbol{H})_{ij}^{-1} H_{kj} + C_0 (\sum_{pq} X_{pq}) \sum_j H_{kj}}{(\sum_{pq} (\boldsymbol{W}\boldsymbol{H})_{pq})^{-1} \sum_j H_{kj} + C_0 (\sum_{pq} (\boldsymbol{W}\boldsymbol{H})_{pq}) \sum_j H_{kj} + C_1} \right)^{\frac{1}{2}}$

Table 3: Some examples of multiplicative update rules for $b = 2$.

Error	Update rule for W_{ik}
Euclidean distance	$W_{ik}^{\mathrm{new}} = W_{ik} \frac{\sum_j X_{ij} H_{kj}}{\sum_j (\boldsymbol{W}\boldsymbol{H})_{ij} H_{kj} + C_1 W_{ik}}$
I-divergence	$W_{ik} \frac{\sum_j X_{ij} (\boldsymbol{W}\boldsymbol{H})_{ij}^{-1} H_{kj}}{\sum_j H_{kj} + 2C_1 W_{ik}}$
Itakura-Saito divergence	$W_{ik}^{\mathrm{new}} = W_{ik} \left(\frac{\sum_j X_{ij} (\boldsymbol{W}\boldsymbol{H})_{ij}^{-2} H_{kj}}{\sum_j (\boldsymbol{W}\boldsymbol{H})_{ij}^{-1} H_{kj} + 2C_1 W_{ik}} \right)^{\frac{1}{3}}$
α-divergence	
1) $\alpha > 0, \alpha \neq 1$	$W_{ik}^{\mathrm{new}} = W_{ik} \left(\frac{\sum_j X_{ij}^{\alpha} (\boldsymbol{W}\boldsymbol{H})_{ij}^{-\alpha} H_{kj}}{\sum_j H_{kj} + 2\alpha C_1 W_{ik}} \right)^{\frac{1}{\alpha+1}}$
2) $-1 \leq \alpha < 0$	$W_{ik}^{\mathrm{new}} = W_{ik} \left(\frac{\sum_j H_{kj}}{\sum_j X_{ij}^{\alpha} (\boldsymbol{W}\boldsymbol{H})_{ij}^{-\alpha} H_{kj} - 2\alpha C_1 W_{ik}} \right)$
3) $\alpha < -1$	$W_{ik}^{\mathrm{new}} = W_{ik} \left(\frac{\sum_j X_{ij}^{\alpha} (\boldsymbol{W}\boldsymbol{H})_{ij}^{-\alpha} H_{kj} - 2\alpha C_1 W_{ik}}{\sum_j H_{kj}} \right)^{\frac{1}{\alpha}}$
Kullback-Leibler divergence	$W_{ik}^{\mathrm{new}} = W_{ik} \left(\frac{(\sum_{pq} X_{pq})^{-1} \sum_j X_{ij} (\boldsymbol{W}\boldsymbol{H})_{ij}^{-1} H_{kj} + C_0 (\sum_{pq} X_{pq}) \sum_j H_{kj}}{(\sum_{pq} (\boldsymbol{W}\boldsymbol{H})_{pq})^{-1} \sum_j H_{kj} + C_0 (\sum_{pq} (\boldsymbol{W}\boldsymbol{H})_{pq}) \sum_j H_{kj} + 2C_1 W_{ik}} \right)^{\frac{1}{2}}$

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Number JP15K00035.

References

[1] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol.5, no.2, pp.111–126, June 1994.

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

[3] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in T. K. Leen, T. G. Dietterich and V. Tresp (Eds), *Advances in Neural Information Processing Systems*, vol.13, pp.556–562, 2001.

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

[5] A. Cichocki, S. Amari, R. Zdunek, R. Kompass, G. Hori and Z. He, "Extended SMART algorithms for non-negative matrix factorization," *Proceedings of the 8th International Conference on Artificial Intelligence and Soft Computing*, pp.548–562, 2006.

[6] M. Seki and N. Takahashi, "New update rules based on Kullback-Leibler, gamma, and Renyi divergences for nonnegative matrix factorization," *Proceedings of 2014 International Symposium on Nonlinear Theory and its Applications*, pp.48–51, September 2014.

[7] N. Takahashi, J. Katayama, M. Seki and J. Takeuchi, "A unified global convergence analysis of multiplicative update rules for nonnegative matrix factorization," submitted.