A Sufficient Condition for the Unique Solution of Non-Negative Tensor Factorization

Toshio Sumi and Toshio Sakata*

Faculty of Design, Kyushu University, Japan
sumi@design.kyushu-u.ac.jp, sakata@design.kyushu-u.ac.jp

Abstract. The applications of Non-Negative Tensor Factorization (NNTF) is an important tool for brain wave (EEG) analysis. For it to work efficiently, it is essential for NNTF to have a unique solution. In this paper we give a sufficient condition for NNTF to have a unique global optimal solution. For a third-order tensor T we define a matrix by some rearrangement of T and it is shown that the rank of the matrix is less than or equal to the rank of T. It is also shown that if both ranks are equal to r, the decomposition into a sum of r tensors of rank 1 is unique under some assumption.

1 Introduction

In the past few years, Non-Negative Tensor Factorization (NNTF) is becoming an important tool for brain wave (EEG) analysis through Morlet wavelet analysis (for example, see Miwakeichi [MMV] and Morup [MHH]). The NNTF algorithm is based on Non-Negative Matrix Factorization (NNMF) algorithms, amongst the most well-known algorithms contributed by Lee-Seung [LS]. Recently, Chichoki et al. [CZA] deals with a new NNTF algorithm using Csiszar's divergence. Furthermore, Wang et al. [WZZ] also worked on NNMF algorithms and its interesting application in preserving privacy in datamining fields. These algorithms converged to some stationary points and do not converge to a global minimization point. In fact, it is easily shown that the problem has no unique minimization points in general (see [CSS]). In applications of NNTF for EEG analysis, it is important for NNTF to have a unique solution. However this uniqueness problem has not been addressed sufficiently as far as the authors are aware of. Similarly as in Non-Negative Matrix Factorization (NNMF), it seems that the uniqueness problem has not been solved. However we managed to obtain the uniqueness and proved it. (see Proposition 1). In this paper we give a sufficient condition for NNTF to have a unique solution and for the usual NNTF algorithm to find its minimization point in the case when NNTF exists strictly, not approximation (see Theorem 3).

* The research is supported partially by User Science Institute in Kyushu University (Special Coordination Funds for Promoting Science and Technology).

M.E. Davies et al. (Eds.): ICA 2007, LNCS 4666, pp. 113–120, 2007.

2 Quadratic Form

As the NNMF problem is a minimization of a quadratic function, we shall first review quadratic functions generally. Let us consider the quadratic form defined by $f(x) = x^T Ax - 2b^T x$ where A is a $n \times n$ symmetric matrix and b is a n vector. The symmetric matrix A is a diagonalized by an orthogonal matrix P as

$$PAP^T = \mathrm{diag}(e_1, \dots, e_n).$$

Then by assigning $y = (y_1, \dots, y_n)^T = Px$ and $c = (c_1, \dots, c_n)^T = Pb$, we obtain the equality

$$f(x) = y^T(PAP^T)y - 2c^T y = \sum_i (e_i y_i^2 - 2c_i y_i) = \sum_i \left(e_i (y_i - \frac{c_i}{e_i})^2 - \frac{c_i^2}{e_i} \right).$$

We assume that the matrix A is positive definite. Then,when $f(x)$ reaches its minimum at $y = (PAP^T)^{-1}c = (PAP^T)^{-1}Pb = PA^{-1}b$ in $\mathbb{R}^n$, with the value $f(A^{-1}b) = -b^T A^{-1}b$ at $x = P^T y = A^{-1}b \in \mathbb{R}^n$. The minimal value is under the condition $x \geq 0$. Here, some basic facts will be explained. Let $a \in \mathbb{R}^n$ and let $h\colon \mathbb{R}^n \to \mathbb{R}$ be a function defined as $h(x) = \| x - a \|^2$, where $\| \cdot \|$ stands for the common Euclidean norm.

Lemma 1. *On the arbitrary closed set S of $\mathbb{R}^n$, $h(t)$, $t \in S$ takes a global minimal value in S.*

Proof. Choose an arbitrary $t_0 \in S$, and set $s = h(t_0)$ and $U = h^{-1}([0, s]) \cap S$. The set U is a closed subset of $\mathbb{R}^n$. By triangular inequality, we know that $h(t) \geq \| t \| - \| a \|$. Since $s \geq h(t)$ for $t \in U$, it holds that $\| t \| \leq s + \| a \|$ which shows that U is bounded. Hence, since U is bounded and closed, it is compact. Thus $h(t)$, $t \in U$ becomes a closed map, and $h(U)$ is also compact. That is, $h(t)$, $t \in U$ takes a global minimum value, say s_0. Thus, it holds that for $t \in S$, $h(t) > s$ if $t \notin U$, and $h(t) \geq s_0$ if t in U. This means that s_0 is the global minimum of h on S. □

Lemma 2. *Let S be a closed convex subset of $\mathbb{R}^n$. Then $h(t)$, $t \in S$ reaches a global minimal value at a unique point in S.*

Proof. The existence of a global minimal value follows from Lemma 1. Let x and y be points in $\mathbb{R}^n$ which attain a global minimal value $r := \min_{z \in S} f(z)$. Note that $x, y \in S \cap \partial B_r(z_0)$, where $B_r(a) := \{x \mid \| x - a \| \leq r\}$ and $\partial B_r(a) := \{x \mid \| x - a \| = r\}$. Since $S \cap B_r(a)$ is also convex, $tx + (1-t)y \in S \cap B_r(a)$ for each $0 \leq t \leq 1$. If $x \neq y$, then $\| a - (x+y)/2 \| < r$, which is contradiction. Therefore $x = y$. □

Let $D = \mathrm{diag}(\sqrt{e_1}, \dots, \sqrt{e_n})$, $z = DPx$ and $S = \{ z \in \mathbb{R}^n \mid x \geq 0 \}$. Note that S is a convex set of $\mathbb{R}^n$ and $f(x) = h(z)$ for $a = D^{-1}Pb$. Therefore $f(x)$ reaches a global minimal value at a unique point under the condition $x \geq 0$.

The following are some basic facts about matrix decompositions. Let A, W and H be $m \times n$, $m \times r$ and $r \times n$ matrix respectively. Then $\| A - WH \|$ for any H with $H \geq O$ and any W with $W \geq O$ reaches a global minimal value at a unique $m \times n$ matrix WH but W and H are not unique. We state this precisely below.

Proposition 1. *The following properties hold:*

1. *If* $r = \text{rank}(A)$*, there exist* W *and* H *such that* $A = WH$*.*
2. *If* $r > \text{rank}(A)$*, there exists an infinite number of pairs of* W *and* H *such that* $A = WH$*.*
3. *Let* $r = \text{rank}(A)$*. Then if* $A = WH = W'H'$ *there exists a non-singular matrix* X *such that* $W' = WX$*,* $H' = X^{-1}H$ *([CSS, Full-Rank Decomposition Theorem]).*
4. *If* $r < \text{rank}(A)$*, there exists no pair of* W *and* H *such that* $A = WH$*.*

Proof. (1) In this case, let W be a matrix whose columns are linearly independent vectors of length m. From the assumption it is clear that the columns of A are expressed as linear combination of columns of W hence $A = WH$.

(2) In this case, put $s = \text{rank}(A)$, $(r > s)$. By property (1), we know there exists a $m \times s$ matrix W_1 and a $s \times m$ matrix H_1 such that $A = W_1 H_1$. Place $W = \begin{pmatrix} W_1 & W_2 \end{pmatrix}$ and $H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}$ where W_2 and H_2 are $m \times (r - s)$ matrix and $(r - s) \times n$ matrix respectively and satisfy $W_2 H_2 = 0$. There are infinitely many such pairs of (W_2, H_2), and for all of those it clearly holds that $A = WH$.

(3) From $r = \text{rank}(A)$, in the expression of $A = WH = W'H'$, the columns of W and W' are linearly independent respectively. Hence we have $W' = WX$ for some regular $r \times r$ matrix X. From this the rest of (3) is derived trivially.

(4) Since $\text{rank}(WH) \leq r$, it is impossible to have $A = WH$. □

3 Non-Negative Matrix Factorization

It is well known NNMF (Non-Negative Matrix Factorization) is not unique ([CSS]). Let V, W and H be a $m \times n$, $m \times r$ and $r \times n$ matrix respectively. For a matrix A, we denote by A_{ij} the (i, j)-component of A and its Frobenius norm is defined by

$$\| A \|_F := \sqrt{tr(A^T A)} = \sqrt{\sum_{i,j} A_{ij}^2},$$

where tr takes the sum of all diagonal entries.

Lemma 3. *Fixing* H*,* $f(W) = \| V - WH \|_F$ *attains the minimum at the solution* W *of the equation* $W(HH^T) = VH^T$*. Especially, if* HH^T *is non-singular, the minimum is attained at the unique point* $W = VH^T(HH^T)^{-1}$*.*

Proof. It holds that

$$\begin{aligned} f(W) &= \sum_{i,j}(V_{ij} - \sum_{p} W_{ip}H_{pj})^2 \\ &= \sum_{i,j}\left(\sum_{p,q} W_{ip}H_{pj}W_{iq}H_{qj} - 2\sum_{p} V_{ij}W_{ip}H_{pj} + V_{ij}^2\right) \\ &= \sum_{i,p,q}(HH^T)_{pq}W_{ip}W_{iq} - 2\sum_{i,p}(VH^T)_{ip}W_{ip} + \sum_{i,j} V_{ij}^2. \end{aligned}$$

Therefore $f(W)$ is a quadratic function of W_{ij} $(i = 1, 2, \cdots, m,\ j = 1, 2, \cdots, r)$. Put

$$\begin{aligned} x &= (W_{11}, \ldots, W_{1r}, \ldots, W_{m1}, \ldots, W_{mr})^T \in \mathbb{R}^{mr}, \\ a &= ((VH^T)_{11}, \ldots, (VH^T)_{1r}, \ldots, (VH^T)_{m1}, \ldots, (VH^T)_{mr})^T \in \mathbb{R}^{mr} \end{aligned}$$

and define a $mr \times mr$ matrix M by $\mathrm{diag}(HH^T, \ldots, HH^T)$. Then, M is positive semidefinite and $f(W)$ is expressed as

$$f(W) = x^T Mx - 2a^T x + \sum_{i,j} V_{ij}^2.$$

Assume that HH^T is non-singular. Then M is positive definite and thus the minimum of $f(W)$ is attained at the unique point $x = M^{-1}a$, that is, $W^T = (HH^T)^{-1}(VH^T)^T$, equivalent to, $W = VH^T(HH^T)^{-1}$. The minimum value is

$$f(W) = \| V \|_F^2 - \| WH \|_F^2 \qquad (1)$$

and we also have $\| WH \|_F^2 = tr(W^T VH^T) = tr((HH^T)^{-1}(VH^T)^T(VH^T))$. □

Since $\| V - WH \|_F = \| V^T - H^T W^T \|_F$, fixing W, $\| V - WH \|_F$ attains the minimum at the unique point $V = (W^T W)^{-1} W^T V$ if $W^T W$ is non-singular.

We recall the Lee-Seung NNMF Algorithm for the Frobenius norm property.

Theorem 1 ([LS]). *The Frobenius norm* $\| V - WH \|_F$ *is non-increasing under the update rules:*

$$H_{ij} \leftarrow H_{ij}\frac{(W^T V)_{ij}}{(W^T WH)_{ij}} \quad W_{ij} \leftarrow W_{ij}\frac{(VH^T)_{ij}}{(WHH^T)_{ij}}$$

Now we propose the following improvement of the Lee-Seung NNMF Algorithm for the Frobenius norm property. For matrices X with $X \geq 0$ and Y, let $t_{max}(X, Y) = \max\{t \mid (1-t)X + tY \geq O,\ 0 \leq t \leq 1\}$.

Theorem 2. *The Frobenius norm* $\| V - WH \|_F$ *is non-increasing under the update rules:*

$$H \leftarrow \begin{cases} (1-h_0)(W^TW)^{-1}W^TV + h_0H, & \text{if } W^TW \text{ is non-singular and } h_0 > 0 \\ H_{ij}\dfrac{(W^TV)_{ij}}{(W^TWH)_{ij}}, & \text{otherwise} \end{cases}$$

$$W \leftarrow \begin{cases} (1-w_0)VH^T(HH^T)^{-1} + w_0W, & \text{if } HH^T \text{ is non-singular and } w_0 > 0 \\ W_{ij}\dfrac{(VH^T)_{ij}}{(WHH^T)_{ij}}, & \text{otherwise} \end{cases}$$

where $h_0 = t_{max}((W^TW)^{-1}W^TV, H)$ *and* $w_0 = t_{max}(VH^T(HH^T)^{-1}, W)$.

Proof. If either W^TW is singular or $h_0 = 0$, the claim follows from Theorem 1. Suppose both W^TW is non-singular and $h_0 > 0$. By Lemma 3, fixing W, $\| V - WH \|_F$ takes minimum at $(W^TW)^{-1}W^TV$ without the assumption $x \geq 0$. Let us denote $H' = (1-h_0)(W^TW)^{-1}W^TV + h_0H$ for clarity. On the line from H to H', the Frobenius norm decreases and thus $\| V - WH \|_F \geq \| V - WH' \|_F$. Clearly $H' \geq 0$ which follows from the definition of h_0. □

4 Non-Negative Tensor Factorization

4.1 Existence of a Global Optimal Solution

Let $\mathbb{R}_{\geq 0}$ be the set of all non-negative real numbers. Let T be a third-order tensor in $\mathbb{R}_{\geq 0}^{a \times b \times c}$. Let $X = \begin{pmatrix} x_1 \dots x_r \end{pmatrix}$, $Y = \begin{pmatrix} y_1 \dots y_r \end{pmatrix}$ and $Z = \begin{pmatrix} z_1 \dots z_r \end{pmatrix}$ be $a \times r$, $b \times r$ and $c \times r$ matrices, respectively. We define a function f over $\mathbb{R}_{\geq 0}^{(a+b+c)r}$ as

$$f(X, Y, Z) = \sum_{ijk} \left(t_{ijk} - \sum_{\ell} X_{i\ell} Y_{j\ell} Z_{k\ell} \right)^2 .$$

Let $S_b = \{x \in \mathbb{R}_{\geq 0}^b \mid \| x \| = 1\}$ be an intersection of an unit sphere in $\mathbb{R}^b$ with $\mathbb{R}_{\geq 0}^b$. Put $S = (\mathbb{R}^a)^{\times r} \times (S_b)^{\times r} \times (S_c)^{\times r}$ for short, where $M^{\times r} = M \times \cdots \times M$ (r times). Then S is a closed subspace of $\mathbb{R}_{\geq 0}^{(a+b+c)r}$ and the image $f(S)$ coincides with the full image $f(\mathbb{R}_{\geq 0}^{(a+b+c)r})$. Let $(X, Y, Z) \in S$. Then $X \geq O$, $Y \geq O$, $Z \geq O$ and $\| y_j \| = \| z_j \| = 1$ for all j. Noting that

$$\sum_{i,j,k} \left(\sum_{\ell} X_{i\ell} Y_{j\ell} Z_{k\ell} \right)^2 \geq \sum_{i,j,k} (X_{i\ell} Y_{j\ell} Z_{k\ell})^2 = \| x_\ell \|^2$$

if $\sum_{i,j,k} \left(\sum_{\ell} X_{i\ell} Y_{j\ell} Z_{k\ell} \right)^2$ is bounded, $\| X \|_F$ is also bounded and thus so is S. Hence, we can apply the proof of Lemma 1 for the function f on S instead of h and we obtain an existence of a global minimal value.

4.2 Uniqueness

We show the uniqueness under some assumption. First,several facts are presented. For convenience, we define

$$X_1 \circ \cdots \circ X_k = (x_1^{(1)} \otimes \cdots \otimes x_1^{(k)}, \ldots, x_r^{(1)} \otimes \cdots \otimes x_r^{(k)})$$

for matrices $X_1 = (x_1^{(1)}, \ldots, x_r^{(1)})$, ..., $X_k = (x_1^{(k)}, \ldots, x_r^{(k)})$ with r-columns. For $u = (1, \ldots, 1)^T \in \mathbb{R}^r$, we have $f(X, Y, Z) = \| T - (X \circ Y \circ Z)u \|_F^2$. For a transformation $M_\sigma = (m_{ij})$ among $\{1, \ldots, r\}$ σ, a permutation matrix M_σ is defined by $m_{ij} = \delta_{i\sigma(j)}$. For a permutation matrix M_σ it does hold that

$$M_\sigma^T = M_{\sigma^{-1}} = M_\sigma^{-1}.$$

Proposition 2. *In a general P, the following equation does not hold*

$$(X_1 P) \circ \cdots \circ (X_r P) = (X_1 \circ \cdots \circ X_r)P.$$

However, if P is a permutation matrix, and $P_1, \ldots, P_r$ are diagonal matrices,

$$(X_1 P) \circ \cdots \circ (X_r P) = (X_1 \circ \cdots \circ X_r)P$$
$$(X_1 P_1) \circ \cdots \circ (X_r P_r) = (X_1 \circ \cdots \circ X_r)P_1 \cdots P_r$$

does hold.

Lemma 4. *Let A and C be $m \times r$ matrices and B and D be $n \times r$ matrices, and Q be $r \times r$ non-singular matrix. Assume that $A \circ B = (C \circ D)Q$ and* $\mathrm{rank}(C) = \mathrm{rank}(C \circ D) = r$ *Then there exists a permutation matrix $P = M_\sigma$ such that both of PQ and QP^{-1} become diagonal matrices and $A = CQX$ and $B = DP^{-1}X^{-1}$ hold for some diagonal matrix X. Further suppose that $A, B, C, D \geq 0$. Let $Q_{1/2}$ be a $r \times r$ matrix whose (i, j)-component is the square root of the (i, j)-component of Q. Then $Q_{1/2}$ is a real matrix, and both $A = CQ_{1/2}X$ and $B = DQ_{1/2}X^{-1}$ hold for some diagonal matrix X.*

Proof. We use the notations $A = (a_1, \ldots a_r)$, $B = (b_1, \ldots b_r) = (b_{ij})$, $C = (c_1, \ldots c_r)$, $D = (d_1, \ldots d_r) = (d_{ij})$, $Q = (q_{ij})$. Since $\otimes$ is a bilinear operation and $\mathrm{rank}(C \circ D) = r$, it holds that $d_k \neq 0$ ($^\forall k$) and that $d_k \,/\!/\, d_\ell$ implies $k = \ell$. Since $A \circ B = (C \circ D)Q$, we have

$$a_k \otimes b_k = \sum_\ell q_{\ell k} c_\ell \otimes d_\ell, \quad {}^\forall k, \quad \text{and} \quad b_{ik} a_k = \sum_\ell q_{\ell k} d_{i\ell} c_\ell, \quad {}^\forall i, k.$$

Since Q is non-singular, for each k there exists a permutation $\sigma(k)$ such that $q_{\sigma(k)k} \neq 0$. Now we will show that for each ℓ there exists an i such that $b_{i\ell} \neq 0$. Assume that $b_s = 0$ for some s. Then, it holds that $\sum_\ell q_{\ell s} c_\ell \otimes d_\ell = 0$, and since $\mathrm{rank}(C \circ D) = r$, it holds that $q_{\ell s} = 0$ ($^\forall \ell$). This contradicts to the fact that Q is non-singular. Therefore, for each ℓ, there exists a $\tau(\ell)$ such that $b_{\tau(\ell)\ell} \neq 0$. Then, it follows

$$q_{\ell k} b_{ik} d_{\tau(k)\ell} = q_{\ell k} b_{\tau(k)k} d_{i\ell}, \quad {}^\forall i, k, \ell$$

from the equality

$$b_{\tau(k)k}b_{ik}a_k = \sum_\ell b_{ik}q_{\ell k}d_{\tau(k)\ell}c_\ell = \sum_\ell b_{\tau(k)k}q_{\ell k}d_{i\ell}c_\ell, \quad {}^\forall i,k$$

and $\mathrm{rank}(C) = r$. On the assumption of $q_{\ell k} \neq 0$, since $d_{i\ell} = \dfrac{d_{\tau(k)\ell}}{b_{\tau(k)k}} \cdot b_{ik}$ for all i it holds that $d_\ell = \dfrac{d_{\tau(k)\ell}}{b_{\tau(k)k}} b_k$. Especially it holds that $d_{\tau(k)\ell} \neq 0$. That is, it holds that $d_\ell \,//\, b_k$. Hence, by $\mathrm{rank}(C \circ D) = r$, if $q_{\ell k} \neq 0$, then $\ell = \sigma(k)$. This implies that there exists a permutation matrix $P = M_\sigma$ such that both of PQ and QP^{-1} are diagonal. Then, if we choose $X = \mathrm{diag}\left(\dfrac{d_{\tau(k)\sigma(k)}}{b_{\tau(k)k}}\right)$, it holds that

$$a_k = q_{\sigma(k)k} \cdot \frac{d_{\tau(k)\sigma(k)}}{b_{\tau(k)k}} c_{\sigma(k)},\ b_k = \frac{b_{\tau(k)k}}{d_{\tau(k)\sigma(k)}} d_{\sigma(k)}, \quad {}^\forall k$$

that is, it holds that $A = CQX,\ B = DP^{-1}X^{-1}$. Further, on the assumption of $Q \geq 0$, if we choose $Y = \mathrm{diag}\left(\dfrac{\sqrt{q_{\sigma(k)k}}\, d_{\tau(k)\sigma(k)}}{b_{\tau(k)k}}\right)$, it holds that $A = CQ_{1/2}Y,\ B = DQ_{1/2}Y^{-1}$. These completes the proof of Lemma 4. □

We should note that the factorization $(X \circ Y \circ Z)$ has the scalar uncertainty such that for scalars a, b, c, it holds

$$(a'X) \circ (b'Y) \circ (c'Z) = (abc)(X \circ Y \circ Z).$$

where (a', b', c') denotes any permutation of (a, b, c). Now we give a sufficient condition that NNTF has the unique global solution. From now set $u = (1, \ldots, 1)^T \in \mathbb{R}^r$ and let $fl_1(T)$ be a $a \times bc$ matrix whose $(i, j + b(k-1))$-component is t_{ijk}. Then the following theorem holds.

Theorem 3. *For* $f(X, Y, Z) = \| T - (X \circ Y \circ Z)u \|_F^2$, *we assume* $\mathrm{rank}(fl_1(T)) = r$ *and* $\min f(X, Y, Z) = 0$. *Then, under the condition* $\mathrm{rank}(Y) = \mathrm{rank}(Y \circ Z) = r$, *the optimal global point is unique up to permutations and scalar uncertainty.*

Proof. By triangular inequality we have

$$\| (X_1 \circ Y_1 \circ Z_1)u - (X_0 \circ Y_0 \circ Z_0)u \|_F \leq f(X_0, Y_0, Z_0) + f(X_1, Y_1, Z_1) = 0,$$

and thus $(X_0 \circ Y_0 \circ Z_0)u = (X_1 \circ Y_1 \circ Z_1)u$ which is equivalent to the following equation $X_0(Y_0 \circ Z_0)^T = X_1(Y_1 \circ Z_1)^T$. By Proposition 1 (3), there exists a non-singular matrix Q such that $X = X_0(Q^T)^{-1}$, $Y \circ Z = (Y_0 \circ Z_0)Q$. From Lemma 4, for some permutation matrix P and diagonal matrix D_1, it holds that $D_2 := PQ$ is a diagonal matrix and $Y = Y_0 Q D_1$ and $Z = Z_0 P^{-1} D_1^{-1}$. Hence, noting $P^{-1} = P^T$, it holds that $X = X_0 P^{-1} D_2^{-1}$, $Y = Y_0 P^{-1} D_2 D_1$, $Z = Z_0 P^{-1} D_1^{-1}$. Up to scalar uncertainty, (X, Y, Z) is equal to $(X_0 P^{-1}, Y_0 P^{-1}, Z_0 P^{-1}) = (X_0, Y_0, Z_0)P^{-1}$, and also it is, up to permutation, equal to (X_0, Y_0, Z_0). □

In general, it does not hold $(X_0 \circ Y_0 \circ Z_0)u = (X_1 \circ Y_1 \circ Z_1)u$, but we can show the following property.

Proposition 3. *For the function* $f(X, Y, Z) = \| T - (X \circ Y \circ Z)u \|_F^2$, *assume that* (X_0, Y_0, Z_0), (X_1, Y_1, Z_1) *are two stationary points which attain the minimal value such that* $f(X_0, Y_0, Z_0) = f(X_1, Y_1, Z_1)$. *Then it holds that*

$$\| (X_0 \circ Y_0 \circ Z_0)u \|_F = \| (X_1 \circ Y_1 \circ Z_1)u \|_F \ .$$

Proof. Since $f(X, Y, Z) = \| fl_1(T) - X(Y \circ Z)^T \|_F^2$, from the equation (1), we have $\| fl_1(T) \|_F^2 - \| X_0(Y_0 \circ Z_0)^T \|_F^2 = \| fl_1(T) \|_F^2 - \| X_1(Y_1 \circ Z_1)^T \|_F^2$. That is, it holds that $\| X_0(Y_0 \circ Z_0)^T \|_F = \| X_1(Y_1 \circ Z_1)^T \|_F$. □

Finally we remark that the equality

$$\| X_0(Y_0 \circ Z_0)^T \|_F = \| Y_0(Z_0 \circ X_0)^T \|_F = \| Z_0(X_0 \circ Y_0)^T \|_F \ .$$

5 Conclusion

For a third-order tensor T and each r, there exists a sum of r tensors of rank 1 which is the closest to T in the sense of Frobenius norm (Existence property). Generally, a global optimal solution is not unique for NNTF. For this problem we proved that if T is of rank r the rank of the matrix made by an arrangement of T is less than or eaual to r, and that if the equality of both ranks holds the decomposition of T into a sum of r tensors of rank 1 is unique under some condition (Uniqueness property).

References

[CSS] Cao, B., Shen, D., Sun, J.-T., Wang, X., Yang, Q., Chen, Z.: Detect and Track Latent Factors with Online Nonnegative Matrix Factorization, IJCAI 2007, pp. 2689–2694 (2007)

[CZA] Chichoki, A., Zdunek, R., Amari, S.-i.: Non-Negative Tensor Factorization Using Csiszar's Divergence. In: Rosca, J., Erdogmus, D., Príncipe, J.C., Haykin, S. (eds.) ICA 2006. LNCS, vol. 3889, pp. 32–39. Springer, Heidelberg (2006)

[LS] Lee, D.D., Seung, H.S.: Algorithms for Non-negative Matrix Factorization. In: Leen, T.K., Dietterich, T.G., Tresp, V. (eds.) Advances in Neural Information Processing Systems, vol. 13, pp. 556–562. MIT Press, Cambridge (2001)

[MMV] Miwakeichi, F., Martínez-Montes, E., Valdés-Sosa, P.A., Nishiyama, N., Mizuhara, H., Yamaguchi, Y.: Decomposing EEG Data into Space–Time–Frequency Components Using Parallel Factor Analysis. NeuroImage 22, 1035–1045 (2004)

[MHH] Morup, M., Hansen, L.K., Herman, C.S., Parnas, J., Arnfred, S.M.: Parallel Factor Analysis as an Exploratory Tool for Wavelet Transformed Event-related EEG. NeuroImage 29, 938–947 (2006)

[WZZ] Wang, J., Zhong, W., Zhang, J.: NNMF-Based Factorization techniques for High-Accuracy Privacy Protection on Non-negative-valued Datasets. In: Perner, P. (ed.) ICDM 2006. LNCS (LNAI), vol. 4065, pp. 513–517. Springer, Heidelberg (2006)