

Non-negative Tensor Factorization Based on Alternating Large-scale Non-negativity-constrained Least Squares

Hyunsoo Kim, Haesun Park
 College of Computing
 Georgia Institute of Technology
 266 Ferst Drive, Atlanta, GA 30332, USA
 E-mail: {hskim,hpark}@cc.gatech.edu

Lars Eldén
 Department of Mathematics
 Linköping University
 SE-581, 83 Linköping, Sweden
 E-mail: laeld@math.liu.se

Abstract—Non-negative matrix factorization (NMF) and non-negative tensor factorization (NTF) have attracted much attention and have been successfully applied to numerous data analysis problems where the components of the data are necessarily non-negative such as chemical concentrations in experimental results or pixels in digital images. Especially, Andersson and Bro’s PARAFAC algorithm with non-negativity constraints (AB-PARAFAC-NC) provided the state-of-the-art NTF algorithm, which uses Bro and de Jong’s non-negativity-constrained least squares with single right hand side (NLS/S-RHS). However, solving an NLS with multiple right hand sides (NLS/M-RHS) problem by multiple NLS/S-RHS problems is not recommended due to hidden redundant computation. In this paper, we propose an NTF algorithm based on alternating large-scale non-negativity-constrained least squares (NTF/ANLS) using NLS/M-RHS. In addition, we introduce an algorithm for the regularized NTF based on ANLS (RNTF/ANLS). Our experiments illustrate that our NTF algorithms outperform AB-PARAFAC-NC in terms of computing speed on several data sets we tested.

I. INTRODUCTION

Non-negativity constraints play an important role in analyzing non-negative data such as chemical concentrations or spectrometry signal intensities. Thus, there have been much efforts of developing efficient non-negative matrix factorization (NMF) algorithms [1], [2]. They have been applied to many practical problems including image processing [2], text data mining [3], subsystem identification [4], cancer class discovery [5], [6], [7], etc. Recently, an NMF algorithm based on alternating non-negativity constrained least squares (NMF/ANLS) [8] has been developed, which is theoretically sound and practically efficient. It is much faster than NMFs based on multiplicative update rules [9]. The framework of NMF/ANLS that has a good convergence property has been utilized to develop several variations of NMF such as sparse NMFs [7] for imposing additional sparse constraints on one of factors, the regular-

ized NMF [8] for increasing numerical stabilities, and one-sided NMFs [10] for imposing non-negativity constraints on only one of factors. In this paper, we propose a fast algorithm for non-negative tensor factorization (NTF) that is a multilinear extension of NMF.

A given a non-negative N -order tensor $\underline{\mathbf{T}} \in \mathbb{R}_+^{m_1 \times m_2 \times m_3 \times \dots \times m_N}$ is decomposed into a set of loading matrices (factors) $\{A_1, A_2, A_3, \dots, A_N\}$, where $A_r \in \mathbb{R}_+^{m_r \times k}$ for $1 \leq r \leq N$, k is a positive integer such that $k \leq \min(m_1, m_2, m_3, \dots, m_N)$, and \mathbb{R}_+ denotes the non-negative orthant with appropriate dimensions. For the sake of simplicity, we will present a three-way NTF model. The three-way NTF model of a non-negative tensor $\underline{\mathbf{T}} \in \mathbb{R}_+^{m \times n \times p}$ can be described as

$$t_{ijz} = \sum_{q=1}^k w_{iq} b_{jq} c_{zq} + e_{ijz}, \quad (1)$$

where t_{ijz} is the element of $\underline{\mathbf{T}}$ in the i -th row, j -th column, and z -th tube, w_{iq} is the element of i -th row and q -th column of the loading matrix $W \in \mathbb{R}_+^{m \times k}$ of the first mode, b_{jq} is the element of j -th row and q -th column of the loading matrix $B \in \mathbb{R}_+^{n \times k}$ of the second mode, c_{zq} is the element of z -th row and q -th column of the loading matrix $C \in \mathbb{R}_+^{p \times k}$ of the third mode, and e_{ijz} is the residual element for t_{ijz} . The subscripts m , n and p are used for indicating the dimension of the first, second, and third mode of a three-way array and i , j , and z are used as indices for each of these modes.

The rest of this paper is organized as follows. We present an algorithm for non-negative tensor factorization in Section II. In Section III, we describe the regularized non-negative tensor factorization. Our NTF algorithms are compared with Andersson and Bro’s algorithm [21] on several test data sets in Section IV. Finally, summary is given in Section V.

II. NON-NEGATIVE TENSOR FACTORIZATION (NTF)

The PARALLEL FACtor (PARAFAC) analysis model [11] is a multilinear model for a tensor decomposition. For simplicity, we present a fast NTF algorithm for a three-way PARAFAC model with non-negativity constraints, although it can be extended to a general algorithm for higher order non-negative PARAFAC models. We deal with a non-negative tensor $\underline{\mathbf{T}} \geq 0$ and want to identify three non-negative factors $W, B, C \geq 0$ in Eqn. (1). $\underline{\mathbf{T}}, W, B, C \geq 0$ means that all elements of a tensor and matrices are non-negative. The loading matrices (W, B , and C) can be iteratively estimated by non-negativity constrained least squares (NLS).

For example, given non-negative matrices $B \in \mathbb{R}_+^{n \times k}$, $C \in \mathbb{R}_+^{p \times k}$, and a tensor $\underline{\mathbf{T}} \in \mathbb{R}_+^{m \times n \times p}$, we can estimate non-negative matrix $W \in \mathbb{R}_+^{m \times k}$ by

$$\min_{W \geq 0} \|Y_W^T - X\|_F^2, \quad (2)$$

where X is the $(np) \times m$ unfolded matrix of $\underline{\mathbf{T}}$, Y is $(np) \times k$ matrix obtained from $Y(:, q) = B(:, q) \otimes C(:, q)$ for $1 \leq q \leq k$, where \otimes stands for the Kronecker product ($Y = B \odot C$, where \odot stands for the Khatri-Rao product). The i -th column of W^T (i.e. $\mathbf{w}^i \in \mathbb{R}_+^{k \times 1}$) can be computed from the following non-negativity-constrained least squares for single right hand side (NLS/S-RHS) [12]:

$$\min_{\mathbf{w}^i \geq 0} \|Y \mathbf{w}^i - \mathbf{x}_i\|_F^2, \quad (3)$$

where \mathbf{x}_i is the i -th column vector of X . This NLS problem can be solve by the active set method of Lawson and Hanson [13], which is implemented in MATLAB [14] as function *lsqnonneg*. The terms $Y^T Y \in \mathbb{R}_+^{k \times k}$ and $X^T Y \in \mathbb{R}_+^{m \times k}$ can be computed from

$$\begin{aligned} Y^T Y &= (C^T C) \bullet (B^T B) \\ X^T Y &= T_1 B D_1 + T_2 B D_2 + \dots + T_p B D_p, \end{aligned} \quad (4)$$

where \bullet is the Hadamard product (i.e. entrywise product), T_z is the z -th frontal slice ($m \times n$ layer) of $\underline{\mathbf{T}}$, and D_z is the $k \times k$ diagonal matrix containing the z -th row of C in its diagonal [12].

When $(np) \gg k$, the size of cross-product terms are small due to the small k . In addition, a submatrix of $Y^T Y$ containing only the rows and columns corresponding to the passive set and a part of $Y^T \mathbf{x}_i$ containing only the rows corresponding to the passive set are used for solving the smaller unconstrained least squares problems for passive set in the main loop and inner loop of the Lawson and Hanson's NLS algorithm. Thus, it is possible to reduce overall computing time by precomputing the cross-product terms. This term is also repetitively used for computing all columns of W^T in a series of NLS/S-RHS problems. The matrices B and C can also be estimated via similar ways.

Consequently, an NTF can be obtained from iteratively solving a series of NLS/S-RHS problems [12].

However, we can still develop a faster NTF algorithm that solves an NLS for multiple right hand sides (NLS/M-RHS) problem instead of a series of NLS/S-RHS problems for estimating each factor. Given a non-negative tensor $\underline{\mathbf{T}} \in \mathbb{R}_+^{m \times n \times p}$, two of the factors, say $B \in \mathbb{R}_+^{n \times k}$ and $C \in \mathbb{R}_+^{p \times k}$, are initialized with non-negative values. Then, we iterate the following alternating non-negativity-constrained least squares (ANLS) for multiple right hand sides until a convergence criterion is satisfied:

$$\min_{W \geq 0} \|Y_{BC} W^T - X_{(1)}\|_F^2, \quad (5)$$

where $Y_{BC} = B \odot C$ and $X_{(1)}$ is the $(np) \times m$ unfolded matrix of $\underline{\mathbf{T}}$, and

$$\min_{B \geq 0} \|Y_{WC} B^T - X_{(2)}\|_F^2, \quad (6)$$

where $Y_{WC} = W \odot C$ and $X_{(2)}$ is the $(mp) \times n$ unfolded matrix of $\underline{\mathbf{T}}$, and

$$\min_{C \geq 0} \|Y_{WB} C^T - X_{(3)}\|_F^2, \quad (7)$$

where $Y_{WB} = W \odot B$ and $X_{(3)}$ is the $(mn) \times p$ unfolded matrix of $\underline{\mathbf{T}}$ across the third mode. Alternatively, one can choose to begin with initializing another pair of factors, change orders of NLS/M-RHS, and/or change the order of unfolding of $\underline{\mathbf{T}}$ so as to design alternative NTF/ANLS. Each subproblem shown in Eqns. (5)-(7) can be solved by projected quasi-Newton optimization [15], [16], projected gradient descent optimization [17], non-negativity-constrained least squares [18], [19], and so forth. We implemented our algorithm by NLS/M-RHS [20] since it is based on the active set method that is guaranteed to terminate in a *finite number of steps*, unlike other NLS algorithms that are based on nonlinear optimization techniques. One can impose non-negativity constraints on only a subset of factors. Higher-order NTF can be done via the simple extension of above example for a three-order NTF.

III. REGULARIZED NTF (RNTF)

We also introduce an algorithm for the regularized NTF based on ANLS (RNTF/ANLS). Although our RNTF algorithm is applicable to N -way tensor, we describe a three-way RNTF algorithm for simple presentation. Given a non-negative tensor $\underline{\mathbf{T}} \in \mathbb{R}_+^{m \times n \times p}$, two of the factors, say $B \in \mathbb{R}_+^{n \times k}$ and $C \in \mathbb{R}_+^{p \times k}$, are initialized with non-negative values. Then, we iterate the following alternating non-negativity-constrained least squares (ANLS) for multiple right hand sides with regularization parameters $\alpha_{(r)}$ for $1 \leq r \leq N$ until a convergence criterion is satisfied:

$$\min_{W \geq 0} \left\| \left(\begin{array}{c} Y_{BC} \\ \sqrt{\alpha_{(1)}} I_k \end{array} \right) W^T - \left(\begin{array}{c} X_{(1)} \\ 0_{k \times m} \end{array} \right) \right\|_F^2, \quad (8)$$

where $Y_{BC} = B \odot C$, I_k is a $k \times k$ identity matrix, $0_{k \times m}$ is a zero matrix of size $k \times m$, and $X_{(1)}$ is the $(np) \times m$ unfolded matrix of $\underline{\mathbf{T}}$, and

$$\min_{B \geq 0} \left\| \left(\frac{Y_{WC}}{\sqrt{\alpha(2)} I_k} \right) B^T - \begin{pmatrix} X_{(2)} \\ 0_{k \times n} \end{pmatrix} \right\|_F^2, \quad (9)$$

where $Y_{WC} = W \odot C$, $0_{k \times n}$ is a zero matrix of size $k \times n$, and $X_{(2)}$ is the $(mp) \times n$ unfolded matrix of $\underline{\mathbf{T}}$, and

$$\min_{C \geq 0} \left\| \left(\frac{Y_{WB}}{\sqrt{\alpha(3)} I_k} \right) C^T - \begin{pmatrix} X_{(3)} \\ 0_{k \times p} \end{pmatrix} \right\|_F^2, \quad (10)$$

where $Y_{WB} = W \odot B$, $0_{k \times p}$ is a zero matrix of size $k \times p$, and $X_{(3)}$ is the $(mn) \times p$ unfolded matrix of $\underline{\mathbf{T}}$ across the third mode. Alternatively, one can choose to begin with initializing another pair of factors, change orders of NLS/M-RHS, and/or change the order of unfolding of $\underline{\mathbf{T}}$ so as to design alternative RNTF/ANLS. Each NLS/M-RHS can be solved by one of optimization techniques. One can impose non-negativity constraints on only a subset of factors.

IV. EXPERIMENTS AND ANALYSIS

A. Experimental Results

We used the N -way toolbox 2.11 [21] for MATLAB 6.5 [14] for testing Andersson and Bro's PARAFAC algorithm with non-negativity constraints and implementing our NTF/ANLS algorithm for N -way NTF. We executed all algorithms in MATLAB 6.5 on a P3 600MHz machine with 512MB memory. The N -way toolbox contains a tensor $\underline{\mathbf{T}} \in \mathbb{R}^{5 \times 201 \times 61}$, i.e. a fluorescence data set (AMINO) of five samples with different amount of tryptophan, phenylalanine and tyrosine. Even though there are some small negative values in the data set, they came from the intrinsic uncertainty in real experimental measurements with noise. Thus, having such small negative values is not contradictory to non-negativity issues, in other words, true parameters are still non-negative. We built the following PARAFAC model with non-negativity constraints:

$$\underline{\mathbf{T}} = \sum_{q=1}^3 \mathbf{w}_q \circ \mathbf{b}_q \circ \mathbf{c}_q + \underline{\mathbf{E}}, \quad (11)$$

where \mathbf{w}_q , \mathbf{b}_q , and \mathbf{c}_q are the q -th columns of non-negative matrices W , B , and C , respectively, \circ means outer product of vectors, and $\underline{\mathbf{E}} \in \mathbb{R}^{5 \times 201 \times 61}$ represents approximation errors. Each sample was excited at 61 wavelengths (240 - 300 nm in 1 nm interval) and fluorescence emission intensities are measured at 201 wavelengths (250 - 450 nm in 1 nm interval). Each element of $\underline{\mathbf{T}}$ represents fluorescence emission signal intensity. A three-component PARAFAC model ($k = 3$) was chosen for this data set, and this is undoubtedly correct since we already knew that each signal intensity comes from three components

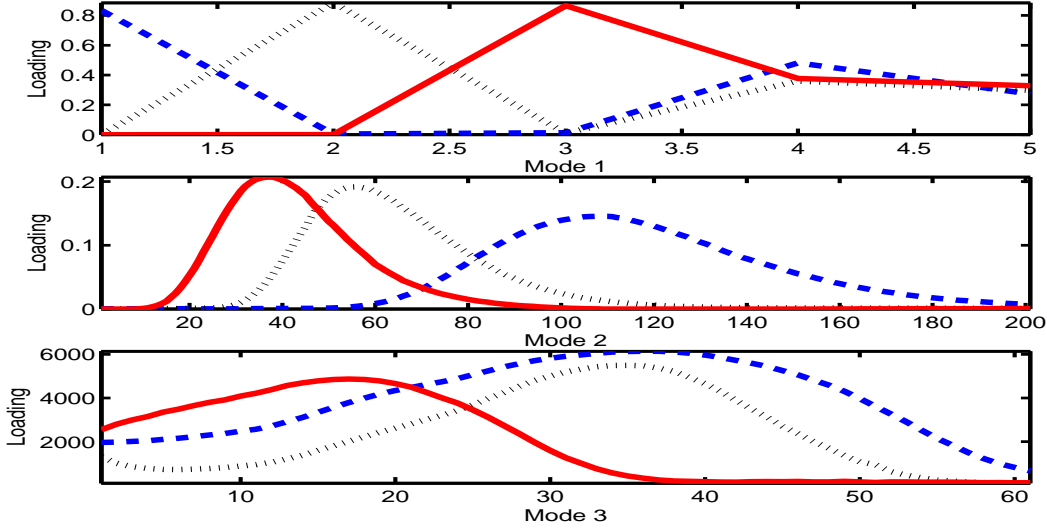
TABLE I
COMPARISON BETWEEN ANDERSSON AND BRO'S PARAFAC ALGORITHM WITH NON-NEGATIVITY CONSTRAINTS (AB-PARAFAC-NC) [21] AND NTF/ANLS ON THE AMINO DATA SET $\underline{\mathbf{T}} \in \mathbb{R}^{5 \times 201 \times 61}$. WE PRESENT THE SUM-OF-SQUARES OF RESIDUALS ($SSR = \sum_{i,j,z} \underline{\mathbf{E}}(i, j, z)^2$), THE NUMBER OF ITERATIONS, AND COMPUTING TIMES.

Algorithm	AB-PARAFAC-NC [21]	NTF/ANLS
SSR	1455820.7088609075	1455817.9774463808
Iteration	37	26
Time	18.206 sec.	3.124 sec.

(analytes). The scores/loadings in W are the sample mode loadings, the loadings in B are the emission mode loadings and the loadings in C are the excitation mode loadings. $W \in \mathbb{R}_+^{5 \times 3}$ has information on the effect of three analytes on five samples. $B \in \mathbb{R}_+^{201 \times 3}$ has information on the fluorescence emission of three analytes at 201 wavelengths. $C \in \mathbb{R}_+^{61 \times 3}$ has information on the response of three analytes at excitation 240 - 300 nm in 1 nm interval. This three-way data analysis cannot be done by multiple bilinear data analysis.

Figure 1 illustrates loadings of three components of three modes (W , B , and C) obtained from NTF/ANLS. We also obtained loadings from Andersson and Bro's PARAFAC algorithm with non-negativity constraints [21], which is referred to as AB-PARAFAC-NC. We compared our NTF/ANLS algorithm with AB-PARAFAC-NC algorithm. The convergence was decided when the relative difference of the sum-of-squares of residuals between two successive iterations was lower than 10^{-6} . The initialization was performed by direct trilinear decomposition via generalized rank annihilation method (DTLD/GRAM). The PARAFAC model is a least-squares model whereas the DTLD model has no well-defined optimization criterion, and many authors found that GRAM was inferior to PARAFAC but suggested that GRAM could be used for the initialization of the PARAFAC algorithm [22]. Line search acceleration scheme [21] was not used for both methods. We did not normalize factors during iterations, but normalized them after convergence. Table I shows the performance comparison between AB-PARAFAC-NC and our NTF/ANLS on the AMINO data set. The sum-of-squares of residuals ($SSR = \sum_{i,j,z} \underline{\mathbf{E}}(i, j, z)^2$), the number of iterations, and computing times were compared. Our NTF/ANLS produced more accurate decomposition within a shorter computing time. When we used the line search acceleration scheme and normalization during iterations,

Fig. 1. Loadings of three modes obtained from NTF/ANLS on the AMINO data set $\underline{\mathbf{T}} \in \mathbb{R}_+^{5 \times 201 \times 61}$: (mode-1: $W \in \mathbb{R}_+^{5 \times 3}$, mode-2: $B \in \mathbb{R}_+^{201 \times 3}$, and mode-3: $C \in \mathbb{R}_+^{61 \times 3}$). Three components are drawn in different color and line types: (1st component: dashed blue line, 2nd component: dotted black line, and 3rd component: solid red line). Components (i.e. column vectors) were normalized to column vectors of unit L_2 -norm in all modes but the last mode. Components were sorted according to contribution.



which may affect the convergence property of algorithms, we observed that our algorithm was still superior to AB-PARAFAC-NC.

An artificial non-negative tensor $\underline{\mathbf{A}} \in \mathbb{R}_+^{50 \times 70 \times 90}$ was built from $A_z = W_a D_z B_a^T$ for $1 \leq z \leq 90$, where $W_a \in \mathbb{R}_+^{50 \times 6}$ and $B_a \in \mathbb{R}_+^{70 \times 6}$ were artificial non-negative matrices, and $D_z \in \mathbb{R}_+^{6 \times 6}$ was a diagonal matrix holding the z -th row of the artificial matrix $C_a \in \mathbb{R}_+^{90 \times 6}$ in its diagonal. Components (i.e. k column vectors) in W_a and B_a were normalized to column vectors of unit L_2 -norm and components in C_a were correspondingly scaled so that $\underline{\mathbf{A}}$ was not changed. The maximum value in $\underline{\mathbf{A}}$ was 255. After adding a positive noise tensor, we obtained $\underline{\mathbf{A}}_N = \underline{\mathbf{A}} + \underline{\mathbf{N}}$, where each element of $\underline{\mathbf{N}}$ was a random positive real number in the range of $(0, 255)$. The sum-of-squares of elements in $(\underline{\mathbf{N}} = \underline{\mathbf{A}}_N - \underline{\mathbf{A}})$ was about 6.825×10^9 . We built the following PARAFAC model with non-negativity constraints:

$$\underline{\mathbf{A}}_N = \sum_{q=1}^6 \mathbf{w}_q \circ \mathbf{b}_q \circ \mathbf{c}_q + \underline{\mathbf{E}}_N, \quad (12)$$

where \mathbf{w}_q , \mathbf{b}_q , and \mathbf{c}_q are the q -th columns of non-negative matrices W , B , and C , respectively, \circ means outer product of vectors, and $\underline{\mathbf{E}}_N \in \mathbb{R}_+^{50 \times 70 \times 90}$ represents approximation errors for $\underline{\mathbf{A}}_N$. We also define approximation errors $\underline{\mathbf{E}}_A$ for $\underline{\mathbf{A}}$:

$$\underline{\mathbf{E}}_A = \underline{\mathbf{A}} - \sum_{q=1}^6 \mathbf{w}_q \circ \mathbf{b}_q \circ \mathbf{c}_q, \quad (13)$$

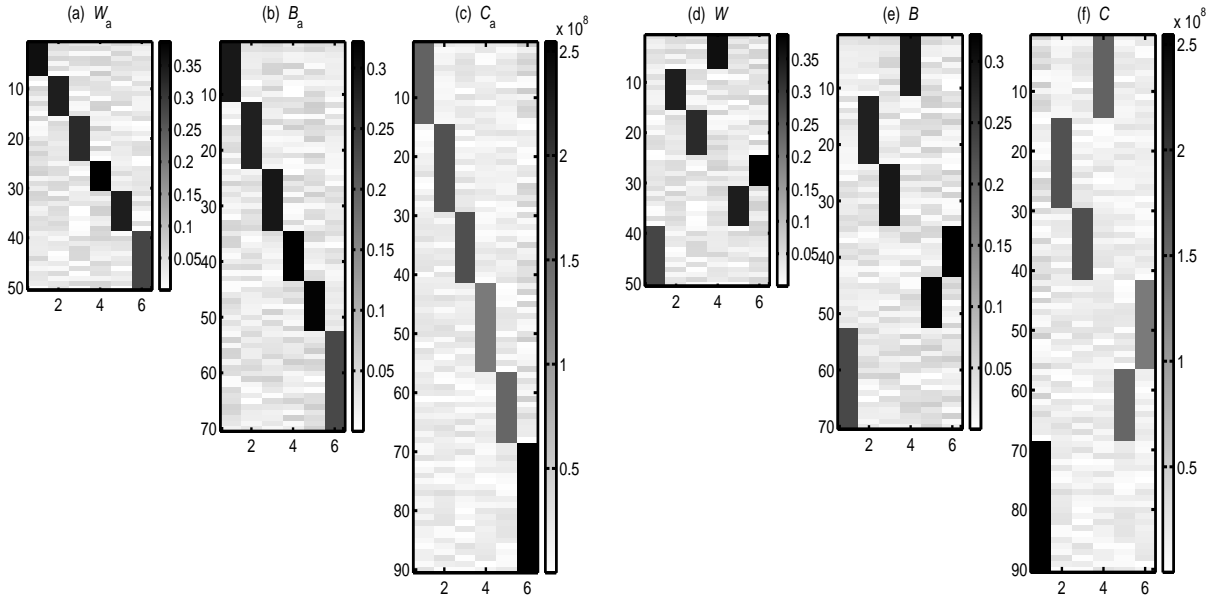
which can measure how close the approximation is to the original noise-free $\underline{\mathbf{A}}$.

TABLE II
COMPARISON BETWEEN ANDERSSON AND BRO'S PARAFAC ALGORITHM WITH NON-NEGATIVITY CONSTRAINTS (AB-PARAFAC-NC) [21] AND NTF/ANLS ON THE ARTIFICIAL NOISE-ADDED DATA SET $\underline{\mathbf{A}}_N \in \mathbb{R}_+^{50 \times 70 \times 90}$. WE PRESENT THE SUM-OF-SQUARES OF RESIDUALS ($\text{SSR} = \sum_{i,j,z} \underline{\mathbf{E}}_N(i,j,z)^2$) FOR $\underline{\mathbf{A}}_N$, THE SUM-OF-SQUARES OF APPROXIMATION ERROR ($\text{SSE} = \sum_{i,j,z} \underline{\mathbf{E}}_A(i,j,z)^2$) FOR THE ARTIFICIAL NOISE-FREE TENSOR $\underline{\mathbf{A}} \in \mathbb{R}_+^{50 \times 70 \times 90}$, THE NUMBER OF ITERATIONS, AND COMPUTING TIMES.

Algorithm	AB-PARAFAC-NC [21]	NTF/ANLS
SSR	2231896540.2836580	2231896540.2836003
SSE	4592875129.9958649	4592875129.9836903
Iteration	10	10
Time	153.631 sec.	5.228 sec.

Table II shows the performance comparison between AB-PARAFAC-NC and our NTF/ANLS on the artificial noise-added data set $\underline{\mathbf{A}}_N$. We used the same initialization method and convergence criterion as the first comparison in Table I. The sum-of-squares of residuals ($\text{SSR} = \sum_{i,j,z} \underline{\mathbf{E}}_N(i,j,z)^2$) for $\underline{\mathbf{A}}_N$, the sum-of-squares of approximation error ($\text{SSE} = \sum_{i,j,z} \underline{\mathbf{E}}_A(i,j,z)^2$) for the artificial noise-free tensor $\underline{\mathbf{A}}$, the number of iterations, and computing times were compared. Our NTF/ANLS produced

Fig. 2. (a)-(c) Loadings of three modes of the artificial noise-free data set $\underline{\mathbf{A}} \in \mathbb{R}_+^{50 \times 70 \times 90}$ (mode-1: $W_a \in \mathbb{R}_+^{50 \times 6}$, mode-2: $B_a \in \mathbb{R}_+^{70 \times 6}$, and mode-3: $C_a \in \mathbb{R}_+^{90 \times 6}$). (d)-(f) Loadings of three modes (W , B , and C) obtained from NTF/ANLS on the artificial noise-added data set $\underline{\mathbf{A}}_N$ ($\underline{\mathbf{A}}_N = \underline{\mathbf{A}} + \underline{\mathbf{N}}$ where $\underline{\mathbf{N}}$ is a positive noise tensor). Components were normalized to column vectors of unit L_2 -norm in all modes but the last mode. Components were sorted according to contribution.



more accurate decomposition of $\underline{\mathbf{A}}_N$ within much shorter computing time. More than 96% of computing time was reduced in this case. The approximated tensor obtained from NTF/ANLS was closer to $\underline{\mathbf{A}}$ than that obtained from AB-PARAFAC-NC.

Figure 2 illustrates that factors obtained from NTF/ANLS on the artificial noise-added data set $\underline{\mathbf{A}}_N = \underline{\mathbf{A}} + \underline{\mathbf{N}}$. The sub-figures (a), (b), and (c) present loadings of three modes of the artificial noise-free data set $\underline{\mathbf{A}}$, i.e. W_a , B_a , and C_a , respectively. The sub-figures (d), (e), and (f) describe loadings of three modes obtained from NTF/ANLS. Even though we used $\underline{\mathbf{A}}_N$ that has considerable noise, NTF/ANLS identified the components in three modes that were very similar to the original components in W_a , B_a , and C_a . Components were normalized to column vectors of unit L_2 -norm in all modes but the last mode. Components were ordered according to contribution. The descending order was determined by diagonal elements in $C^T C \in \mathbb{R}_+^{6 \times 6}$. The order of k values was used to change the order of components in all modes at the same time so that changing the order of components does not affect the approximation error $\underline{\mathbf{E}}_N$. For example, $\sum_{q=1}^6 \mathbf{w}_q \circ \mathbf{b}_q \circ \mathbf{c}_q$ is equal to $(\mathbf{w}_6 \circ \mathbf{b}_6 \circ \mathbf{c}_6 + \mathbf{w}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_2 + \mathbf{w}_3 \circ \mathbf{b}_3 \circ \mathbf{c}_3 + \mathbf{w}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \mathbf{w}_5 \circ \mathbf{b}_5 \circ \mathbf{c}_5 + \mathbf{w}_4 \circ \mathbf{b}_4 \circ \mathbf{c}_4)$ since the order of summation is only changed.

Using the regularized NTF with $\alpha_{(1)} = \alpha_{(2)} = \alpha_{(3)} = 10^{-8}$, we obtained $\text{SSR}=2231896599.7608781$

and $\text{SSE}=4592541808.8592987$ after 10 iterations in 6.31 seconds on the artificial data set $\underline{\mathbf{A}}_N$. We used the same initialization method and convergence criterion as the first comparison in Table I. We did not normalize factors after convergence in this case. Some elements in B factor were larger than 10^8 and all elements in W and C were less than 1.0. The SSR is slightly larger than those of other algorithms, while the SSE is much smaller than those of other NTF algorithms. This result suggests that RNTF would sometimes be helpful in handling noise. The RNTF/ANLS can also be used to control the size of factors or increase numerical stabilities of NTF/ANLS when it is required.

B. Analysis

The computing time of NTF was significantly reduced by using alternating large-scale non-negativity-constrained least squares for multiple right hand sides. AB-PARAFAC-NC uses Bro and de Jong's fast NLS algorithm (*fastnls.m*) [12] for NLS/S-RHS as described in Eqn. (3). However, solving an NLS problem with multiple right hand sides by multiple NLS/S-RHS problems is not recommended due to hidden redundant computation [8]. Thus, we utilized a faster NLS algorithm for multiple right hand sides [20] that improved the performance by initializing with an unconstrained least squares solution and reorganizing the calculations of unconstrained least squares for each unique passive set in each step of the active set iteration.

Let us suppose that we are dealing with the following NLS/M-RHS:

$$\min_{K \geq 0} \|YK - X\|_F^2 \quad (14)$$

Then, the active set method for this NLS solves unconstrained least squares problems by the normal equations $Y^T Y K_S = Y^T X_S$ under certain set of columns S for passive sets $\mathcal{P}_S = \{\mathcal{P}_1, \dots, \mathcal{P}_{|S|}\}$. Each passive set contains a set of row indices of passive variables in the corresponding column of K . \mathcal{P}_S can be represented by a passive logical matrix $P_S \in \mathbb{R}^{k \times |S|}$ of which element is true or false when it is passive or active variable, respectively, for each column. In the sub-routine for solving the unconstrained least squares problems, u unique passive sets $\{\mathcal{U}_1, \dots, \mathcal{U}_u\}$ are found from \mathcal{P}_S . This grouping strategy is an essential part that contributes to the computational efficiency of this algorithm. For each unique passive set \mathcal{U}_j , ($1 \leq j \leq u$), the system of normal equations $\Gamma(\mathcal{U}_j, \mathcal{U}_j)K(\mathcal{U}_j, \mathcal{E}_j) = \Theta(\mathcal{U}_j, \mathcal{E}_j)$ is solved inside the sub-routine, where $\Gamma = Y^T Y$ and $\Theta = Y^T X$ are precomputed matrices, and \mathcal{E}_j is a set of column indices sharing the same passive set of \mathcal{U}_j .

C. Remarks on NTF algorithms

The singular value decomposition (SVD) has been extended to the multilinear SVD or the higher-order singular value decomposition (HOSVD) [23]. It does not have non-negativity constraints but has orthogonality constraints. Therefore, it is not appropriate when we search for an approximation related with non-subtractive combinations of non-negative basis vectors.

Recently, a critical theoretical problem on PARAFAC models has been claimed: the problem of computing the PARAFAC decomposition is ill-posed, except in the case when there are non-negativity constraints [24]. This is one of the most important reasons why we focus on non-negative PARAFAC. NTF can also be considered as the PARAFAC with non-negativity constraints or the higher-order non-negative matrix factorization (HONMF). It can be applied to signal processing especially in neuroscience (EEG, fMRI) and analytical chemistry, image analysis, text data mining for a term-document-author tensor, and gene expression data analysis for a gene-experiment-time tensor. The three-dimensional gene expression data analysis is useful for understanding biological systems since we can simultaneously analyze temporal variations and sample-dependent variations of gene expression levels.

V. SUMMARY

We introduce a fast algorithm for NTF based on alternating large-scale non-negativity-constrained least squares. It is computationally more efficient than the current state-of-the-art Andersson and Bro's PARAFAC algorithm with

non-negativity constraints [21] since it solves an NLS/M-RHS problem instead of a series of NLS/S-RHS problems for estimating each factor. We show that our NTF/ANLS algorithm can be used for multi-way blind source separation with non-negativity constraints. Due to its efficiency, it can also be applied to large-scale multi-dimensional data analysis in neuroscience, psychometrics, chemometrics, computational biology, and bioinformatics when we deal with chemical concentrations or signal intensities.

ACKNOWLEDGMENT

We thank Prof. Rasmus Bro for helpful comments on non-negativity on the fluorescence data set. This work is supported by the National Science Foundation Grants ACI-0305543 and CCF-0621889. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

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, pp. 111–126, 1994.
- [2] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
- [3] V. P. P. Pauca, F. Shahnaz, M. W. Berry, and R. J. Plemmons, "Text mining using non-negative matrix factorizations," in *Proc. SIAM Int'l Conf. Data Mining (SDM'04)*, April 2004.
- [4] P. M. Kim and B. Tidor, "Subsystem identification through dimensionality reduction of large-scale gene expression data," *Genome Research*, vol. 13, pp. 1706–1718, 2003.
- [5] J. P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov, "Metagenes and molecular pattern discovery using matrix factorization," *Proc. Natl Acad. Sci. USA*, vol. 101, no. 12, pp. 4164–4169, 2004.
- [6] Y. Gao and G. Church, "Improving molecular cancer class discovery through sparse non-negative matrix factorization," *Bioinformatics*, vol. 21, no. 21, pp. 3970–3975, 2005.
- [7] H. Kim and H. Park, "Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis," 2007, *Bioinformatics*, to appear.
- [8] —, "Non-negative matrix factorization via alternating non-negativity-constrained least squares and active set method," 2007, submitted.
- [9] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proceedings of Neural Information Processing Systems*, 2000, pp. 556–562. [Online]. Available: <http://citeseer.ist.psu.edu/lee01algorithms.html>
- [10] H. Park and H. Kim, "One-sided non-negative matrix factorization and non-negative centroid dimension reduction for text classification," in *Proceedings of the Workshop on Text Mining at the 6th SIAM International Conference on Data Mining (SDM06)*, M. Castellanos and M. W. Berry, Eds., 2006.
- [11] R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multimodal factor analysis," *UCLA working papers in phonetics*, vol. 16, no. 16, pp. 1–84, 1970.
- [12] R. Bro and S. de Jong, "A fast non-negativity-constrained least squares algorithm," *J. Chemometrics*, vol. 11, pp. 393–401, 1997.
- [13] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
- [14] MATLAB, *User's Guide*. Natick, MA 01760: The MathWorks, Inc., 1992.

- [15] R. Zdunek and A. Cichocki, "Non-negative matrix factorization with quasi-Newton optimization," in *The Eighth International Conference on Artificial Intelligence and Soft Computing (ICAISC)*, 2006, pp. 870–879.
- [16] D. Kim, S. Sra, and I. S. Dhillon, "Fast Newton-type methods for the least squares nonnegative matrix approximation problem," in *Proceedings of the 2007 SIAM International Conference on Data Mining (SDM07)*, 2007, to appear.
- [17] C. J. Lin, "Projected gradient methods for non-negative matrix factorization," Department of Computer Science, National Taiwan University, Tech. Rep. Information and Support Service ISSTECH-95-013, 2005.
- [18] H. Kim and H. Park, "Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares," in *Proceedings of the IASTED International Conference on Computational and Systems Biology (CASB2006)*, D.-Z. Du, Ed., Nov. 2006, pp. 95–100.
- [19] —, "Cancer class discovery using non-negative matrix factorization based on alternating non-negativity-constrained least squares," in *Springer Verlag Lecture Notes in Bioinformatics (LNBI)*, vol. 4463, May 2007, pp. 477–487.
- [20] M. H. van Benthem and M. R. Keenan, "Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems," *J. Chemometrics*, vol. 18, pp. 441–450, 2004.
- [21] C. A. Andersson and R. Bro, "The N-way Toolbox for MATLAB," *Chemometrics and Intelligent Laboratory Systems*, vol. 52, no. 1, pp. 1–4, 2000.
- [22] R. Bro, "PARAFAC: Tutorial and applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 38, pp. 149–171, 1997.
- [23] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, pp. 1253–1278, 2000.
- [24] V. de Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM J. Matrix Anal. Appl.*, to appear, 2007.