SWAMP REDUCING TECHNIQUE FOR TENSOR DECOMPOSITION

Carmeliza Navasca[†], *Lieven De Lathauwer*[‡¶] *and Stefan Kindermann*[§]

[†]Department of Mathematics, Clarkson University
PO Box 5815, Potsdam, NY 13699, USA
email: cnavasca@gmail.com
[‡]Subfaculty Science and Technology, Katholieke Universiteit Leuven Campus Kortrijk,
E. Sabbelaan 53, 8500 Kortrijk, Belgium
e-mail: Lieven.DeLathauwer@kuleuven-kortrijk.be
[¶]Electrical Engineering Department (ESAT-SCD), Katholieke Universiteit Leuven
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium
e-mail: delathau@esat.kuleuven.be
[§]Industrial Mathematics Institute, Johannes Kepler University Linz
Altenbergerstrasse 69,A-4040 Linz, Austria
email: kindermann@indmath.uni-linz.ac.at

ABSTRACT

There are numerous applications of tensor analysis in signal processing, such as, blind multiuser separation-equalization-detection and blind identification. As the applicability of tensor analysis widens, the numerical techniques must improve to accommodate new data. We present a new numerical method for tensor analysis. The method is based on the iterated Tikhonov regularization and a parameter choice rule. Together these elements dramatically accelerate the well-known Alternating Least-Squares method.

1. INTRODUCTION

The purpose of this paper is to present a novel technique for tensor decomposition, namely the parallel factor (PARAFAC) analysis and block term decomposition (BTD). Tensor decomposition or multi-way analysis has been playing a major role in advancing signal processing and communications. For example, it has appeared in problems involving higher order statistics, with source independence, in blind multichannel system identification including the case where the number of sources exceeds the number of sensors, in source separation and in blind multiuser detection. In particular, Sidiropoulos et al. [28] have pioneered the application of tensor analysis to the blind multiuser separation-equalization-detection for direct-sequence code-division multiple access (DS-CDMA) systems where no space/time statistical independence is assumed. There have been numerous contributions thereafter; e.g. see [32] for PARAFAC and [25, 3] for BTD. Much of the success of applying tensor analysis is attributed to the Kruskal essential uniqueness property of PARAFAC [21] and BTD [11].

Moreover tensor decomposition has also been a vital tool in scientific computing [8, 16, 18], data analysis [2, 20], biomedical engineering [1, 14, 23] and chemometrics [29]. For more applications of tensor decomposition, see the survey paper of Kolda and Bader [19]. These tensor analysis techniques rely on numerical methods in optimization and numerical linear algebra. Current methods like the Alternating Least-Squares become inefficient and at times inadequate for these problems. Thus it is important to improve numerical methods for tensor analysis.

In this paper, we introduce our regularized method for PARAFAC and BTD. We provide a new formulation for increasing the convergence rate of the well-known Alternating-Least-Squares method by applying the iterated Tikhonov regularization. Regularization methods have also appeared in the context of tensor decomposition; e.g. see [6], [7] for techniques in approximating nonnegative matrix. Our new formulation results in a new objective functional which requires a selection of the regularization parameter α. We will describe our strategies for an easy and quick computation of the parameter α. The implementation only requires only a few extra lines of coding so they are no more complicated than the original algorithms. This is another motivating factor for using regularization methods. They are extremely easy to implement unlike the purely algorithmic and complicated line search schemes [26, 24, 33] for accelerating the ALS algorithm.

1.1 Preliminaries

We denote the scalars in $\mathbb{R}$ and $\mathbb{C}$ with lower-case letters $(a,b,\ldots)$ and the vectors with bold lower-case letters $(\mathbf{a},\mathbf{b},\ldots)$. The matrices are written as bold upper-case letters $(\mathbf{A},\mathbf{B},\ldots)$ and the symbol for tensors are calligraphic letters $(\mathscr{A},\mathscr{B},\ldots)$. The subscripts represent the following scalars: $(\mathscr{A})_{ijk} = a_{ijk}$, $(\mathbf{A})_{ij} = a_{ij}$, $(\mathbf{a})_i = a_i$. If $\mathbf{A}$ is a partitioned matrix, then $\mathbf{A} = [\mathbf{A_1}\ \mathbf{A_2}\ \ldots \mathbf{A_R}]$ where $\mathbf{A_r} \in \mathbb{C}^{I\times L_r}$ is a submatrix for $r = 1,\ldots,R$. Note that when $L_r = 1$ for all r then the submatrices collapse into column vectors; i.e. $\mathbf{A_r} = \mathbf{a_r}$, and $\mathbf{A} = [\mathbf{a_1}\ \mathbf{a_2}\ \ldots\ \mathbf{a_R}]$. The superscripts indicate the length of the vector or the size of the matrices. For example, $\mathbf{b}^K$ is a vector with length K and $\mathbf{B}^{N\times K}$ is a $N\times K$ matrix.

Definition 1.1 *The Kronecker product of matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *is defined as*

$$\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \ldots \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \ldots \\ \vdots & \vdots & \end{bmatrix}.$$

Definition 1.2 *The Khatri-Rao product of* $\mathbf{A}$ *and* $\mathbf{B}$ *is de-*

fined in [27] as

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{A_1} \otimes \mathbf{B_1} \;\; \mathbf{A_2} \otimes \mathbf{B_2} \;\; \ldots].$$

while the column-wise Khatri-Rao product is the special case,

$$\mathbf{A} \odot_c \mathbf{B} = [\mathbf{a_1} \otimes \mathbf{b_1} \; \mathbf{a_2} \otimes \mathbf{b_2} \; \ldots]$$

when $\mathbf{A} = [\mathbf{a_1} \; \mathbf{a_2} \; \ldots \; \mathbf{a_R}]$ *and* $\mathbf{B} = [\mathbf{b_1} \; \mathbf{b_2} \; \ldots \; \mathbf{b_R}]$.

Definition 1.3 (Mode-*n* vector) *Given a tensor* $\mathscr{T} \in \mathbb{C}^{I\times J\times K}$*, there are three types of mode vectors, namely, mode-*1*, mode-*2*, and mode-*3*. There are* $J \cdot K$ *mode-*1 *vectors that are of length* I *which are obtained by fixing the indices* (j,k) *while varying* i*. Similarly, the mode-*2 *vector (mode-*3 *vector) is of length* J (K) *obtained from the tensor by varying* j (k) *with fixed* (k,i) (i,j).

Definition 1.4 (Mode-*n* rank) *The mode-n rank of a tensor* $\mathscr{T}$ *is the dimension of the subspace spanned by the mode-n vectors.*

The order of a tensor refers to the dimension of the index set. A matrix is a second-order tensor and a vector is a first-order tensor.

Definition 1.5 (rank-(L,M,N)) *A third-order tensor is rank-*(L,M,N) *if the mode-*1 *rank is* L*, the mode-*2 *rank is* M *and the mode-*3 *rank is* N.

In the case when a third-order tensor has rank-$(1,1,1)$, it is simply called a *rank-1* tensor.

Definition 1.6 (Tucker mode-*n* product) *Given a tensor* $\mathscr{T} \in \mathbb{C}^{I_1\times I_2\times I_3}$ *and the matrices* $\mathbf{A} \in \mathbb{C}^{J_1\times I_1}$, $\mathbf{B} \in \mathbb{C}^{J_2\times I_2}$ *and* $\mathbf{C} \in \mathbb{C}^{J_3\times I_3}$*, then the Tucker mode-n products are the following:*

$$\begin{aligned}
(\mathscr{T} \bullet_1 \mathbf{A}) &= \sum_{i_1=1}^{I_1} t_{i_1 i_2 i_3} a_{j_1 i_1},\ \forall j_1, i_2, i_3 \quad \textit{(mode-1 product)} \\
(\mathscr{T} \bullet_2 \mathbf{B}) &= \sum_{i_2=1}^{I_2} t_{i_1 i_2 i_3} b_{j_2 i_2},\ \forall j_2, i_1, i_3 \quad \textit{(mode-2 product)} \\
(\mathscr{T} \bullet_3 \mathbf{C}) &= \sum_{i_3=1}^{I_3} t_{i_1 i_2 i_3} c_{j_3 i_3},\ \forall j_3, i_1, i_2 \quad \textit{(mode-3 product)}
\end{aligned}$$

2. TENSOR ANALYSIS

The Parallel Factor Decomposition (PARAFAC) first appeared in [17] in the context of psychometrics. Independently, [9] introduced this decomposition as the Canonical Decomposition (CANDECOMP) in phonetics. Given a tensor $\mathscr{T} \in \mathbb{C}^{I\times J\times K}$, the PARAFAC decomposition is

$$\mathscr{T} = \sum_{r=1}^{R} \mathbf{a_r} \circ \mathbf{b_r} \circ \mathbf{c_r} \tag{1}$$

where $\mathbf{a_r}$, $\mathbf{b_r}$, $\mathbf{c_r}$ are the r-th columns of the matrices $\mathbf{A} \in \mathbb{C}^{I\times R}$, $\mathbf{B} \in \mathbb{C}^{J\times R}$, and $\mathbf{C} \in \mathbb{C}^{K\times R}$. The symbol '$\circ$' denotes the usual outer product; i.e. if $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, then $\mathbf{v} \circ \mathbf{w} = \mathbf{v} \otimes \mathbf{w}^T$. PARAFAC is a sum of rank-1 tensors. Another decomposition is the so-called Tucker decomposition [12, 30, 31]:

$$\mathscr{T} = \sum_{l=1}^{L}\sum_{m=1}^{M}\sum_{n=1}^{N} (\mathscr{D})_{lmn}\, \mathbf{a_l} \circ \mathbf{b_m} \circ \mathbf{c_n} \tag{2}$$

where $(\mathscr{D})_{lmn}$ is a component of the tensor $\mathscr{D} \in \mathbb{C}^{L\times M\times N}$ with full mode-1, mode-2 and mode-3 rank, $\mathbf{a_l}$ is l-th column of $\mathbf{A} \in \mathbb{C}^{I\times L}$, $\mathbf{b_m}$ is the m-th column of $\mathbf{B} \in \mathbb{C}^{J\times M}$, and $\mathbf{c_n}$ is the n-th column of $\mathbf{C} \in \mathbb{C}^{K\times N}$. The decomposition (2) is written more succinctly in terms of the Tucker mode-n product,

$$\mathscr{T} = \mathscr{D} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} \bullet_3 \mathbf{C}.$$

A normalized version of this formulation is referred to as the Higher-Order Singular Value Decomposition (HOSVD). In [12], the HOSVD has been shown as a generalization of the matrix SVD.

In the papers of [10, 11, 13], BTD is introduced. The following is a BTD in rank-$(L_r, L_r, 1)$:

$$\mathscr{T} = \sum_{r=1}^{R} \mathbf{E_r} \circ \mathbf{c_r} = \sum_{r=1}^{R} (\mathbf{A_r} \cdot \mathbf{B_r}^T) \circ \mathbf{c_r} \tag{3}$$

where $\mathbf{E_r} \in \mathbb{C}^{I\times J}$ with rank($\mathbf{E_r}$)$= L_r$ and $\mathbf{E_r} = \mathbf{A_r} \cdot \mathbf{B_r^T}$ and $\mathbf{c_r}$ is a column vector for $r = 1, \ldots, R$. Note that $\mathbf{A_r} \in \mathbb{C}^{I\times L_r}$ with rank$(\mathbf{A_r}) = L_r$ and $\mathbf{B_r} \in \mathbb{C}^{J\times L_r}$ with rank$(\mathbf{B_r}) = L_r$. When $L_r = L$ for all $r = 1, \ldots, R$, it is called the BTD in rank-$(L,L,1)$. Furthermore, the BTD in rank-$(L_r, L_r, 1)$ can be recast as

$$\mathscr{T} = \sum_{r=1}^{R} \mathbf{D_r} \bullet_1 \mathbf{A_r} \bullet_2 \mathbf{B_r} \bullet_3 \mathbf{c_r} \tag{4}$$

where $\mathbf{D_r} \in \mathbb{C}^{L_r\times L_r}$ is full rank $L_r \times L_r$ matrix. Lastly, the BTD in rank-(L,M,N) is

$$\mathscr{T} = \sum_{r=1}^{R} \mathscr{D}_r \bullet_1 \mathbf{A_r} \bullet_2 \mathbf{B_r} \bullet_3 \mathbf{C_r} \tag{5}$$

where $\mathscr{D}_r \in \mathbb{C}^{L\times M\times N}$ has full mode-1, mode-2 and mode-3 ranks, $\mathbf{A_r} \in \mathbb{C}^{I\times L}$, $\mathbf{B_r} \in \mathbb{C}^{J\times M}$ and $\mathbf{C_r} \in \mathbb{C}^{K\times N}$ for $r = 1, \ldots, R$.

2.1 Matrix And Vector Representation

To apply the numerous well-developed tools in numerical linear algebra and optimization, matrix and vector representations of tensor are necessary. In the PARAFAC framework, the standard matricization of the tensors are in the directions of the left-right, front-back, and top-bottom to generate slices of matrices. Concatenating these sliced matrices allows us to build these long matrices:

$$\begin{aligned}
&\mathbf{T}^{JK\times I} = (\mathbf{B} \odot \mathbf{C})\mathbf{A}^T, \;\; \mathbf{T}^{KI\times J} = (\mathbf{C} \odot \mathbf{A})\mathbf{B}^T \\
&\text{and } \mathbf{T}^{IJ\times K} = (\mathbf{A} \odot \mathbf{B})\mathbf{C}^T
\end{aligned} \tag{6}$$

where the superscripts of $\mathbf{T}$ reflect the matrix size. For the BTD in rank-$(L_r, L_r, 1)$, the matricization results into these matrices

$$\mathbf{T}^{JK\times I} = [\mathbf{B_1} \otimes \mathbf{c_1} \ldots \mathbf{B_R} \otimes \mathbf{c_R}]\mathbf{A}^T, \tag{7}$$

$$\mathbf{T}^{KI\times J} = [\mathbf{c_1} \otimes \mathbf{A_1} \ldots \mathbf{c_R} \otimes \mathbf{A_R}]\mathbf{B}^T \tag{8}$$

and

$$\mathbf{T}^{IJ\times K} = [(\mathbf{A_1} \odot \mathbf{B_1})1_{L_1} \ldots (\mathbf{A_R} \odot \mathbf{B_R})1_{L_R}]\mathbf{C}^T \tag{9}$$

where 1_{L_r} is a column vector of 1's of length L_r.

3. ITERATED TIKHONOV REGULARIZATION

3.1 Alternating Least Squares

The problem we are solving is the following: we would like to recover the *best* tensor $\mathscr{T} \in \mathbb{C}^{I\times J\times K}$ of rank-$(L_r, L_r, 1)$ or rank-(L_r, M_r, N_r) if it exists from the given a noisy tensor $\widehat{\mathscr{T}} \in \mathbb{C}^{I\times J\times K}$. Define the residual tensor as $\mathscr{R} = \widehat{\mathscr{T}} - \mathscr{T}$. The standard approach is then to minimize the norm of the residual tensor in the least-square sense:

$$\min \|\mathscr{R}\|_F^2 = \min_{\mathscr{T}} \|\widehat{\mathscr{T}} - \mathscr{T}\|_F^2 \tag{10}$$

where $\|\cdot\|_F$ is the Frobenius norm. Here we discuss the recovery of the tensor decomposition in rank-$(L_r, L_r, 1)$, but the techniques are easily extended to the BTD in rank-(L_r, M_r, N_r). Recall the Frobenius norm of a tensor is defined as

$$\|\mathscr{A}\|_F^2 = \sum_{i=1}^{I}\sum_{j=}^{J}\sum_{k=1}^{K} |a_{ijk}|^2$$

which is a direct extension of the Frobenius norm of a matrix. Here $\mathscr{T} = \sum_{r=1}^{R} \mathbf{A_r}\mathbf{B_r}^T \circ \mathbf{c_r}$ is tensor of rank-$(L_r, L_r, 1)$ where $\mathbf{A_r} \in \mathbb{R}^{I\times L_r}$, $\mathbf{B_r} \in \mathbb{R}^{J\times L_r}$, $\mathbf{c_r}$ is column vector of $\mathbf{C} \in \mathbb{R}^{K\times R}$. Then it follows that the optimization problem (10) is

$$\min_{A,B,C} \left\| \widehat{\mathscr{T}} - \sum_{r=1}^{R} \mathbf{A_r}\mathbf{B_r}^T \circ \mathbf{c_r} \right\|_F^2 \tag{11}$$

where the minimization is over the partitioned matrices $\mathbf{A}$ and $\mathbf{B}$ and the matrix $\mathbf{C}$. Recall that $\mathbf{A} = [\mathbf{A_1} \ldots \mathbf{A_R}]$, $\mathbf{B} = [\mathbf{B_1} \ldots \mathbf{B_R}]$ and $\mathbf{C} = [\mathbf{c_1} \ldots \mathbf{c_R}]$. Although there exists methods which directly approximate the factors $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ through the least-squares formulation (10), it is more tractable and natural to approximate using these three sub-least squares problems:

$$\begin{cases} \min_A \|\widehat{\mathbf{T}}^{JK\times I} - \mathbf{T}^{JK\times I}\|_F^2 = \min_A \|\widehat{\mathbf{T}}^{JK\times I} - \mathbf{Q}\mathbf{A}^T\|_F^2 \\ \min_B \|\widehat{\mathbf{T}}^{KI\times J} - \mathbf{T}^{KI\times J}\|_F^2 = \min_B \|\widehat{\mathbf{T}}^{KI\times J} - \mathbf{R}\mathbf{B}^T\|_F^2 \\ \min_C \|\widehat{\mathbf{T}}^{IJ\times K} - \mathbf{T}^{IJ\times K}\|_F^2 = \min_C \|\widehat{\mathbf{T}}^{IJ\times K} - \mathbf{S}\mathbf{C}^T\|_F^2 \end{cases} \tag{12}$$

where $\widehat{\mathbf{T}}^{JK\times I}$, $\widehat{\mathbf{T}}^{KI\times J}$ and $\widehat{\mathbf{T}}^{IJ\times K}$ are the matricization of tensor $\widehat{\mathscr{T}}$ through the top-bottom, left-right, front-back slices as in Section 2 and the matrices are

$$\begin{aligned} \mathbf{Q} &= [\mathbf{B_1} \otimes \mathbf{c_1} \ldots \mathbf{B_R} \otimes \mathbf{c_R}], \\ \mathbf{R} &= [\mathbf{c_1} \otimes \mathbf{A_1} \ldots \mathbf{c_R} \otimes \mathbf{A_R}] \\ &\quad \textit{and} \\ \mathbf{S} &= [(\mathbf{A_1} \odot \mathbf{B_1})1_{L_1} \ldots (\mathbf{A_R} \odot \mathbf{B_R})1_{L_R}]. \end{aligned}$$

The three sub-least squares problems is implemented consecutively and iteratively. Each sub-least squares problem is updated immediately by the newly estimated matrix from the preceding sub-least squares problem much like the Gauss-Seidel scheme [5]. If instead

$$\mathbf{Q} = \mathbf{B} \odot \mathbf{C}, \;\; \mathbf{R} = \mathbf{C} \odot \mathbf{A}, \;\; \text{and} \;\; \mathbf{S} = \mathbf{A} \odot \mathbf{B}$$

and the noisy version of the matrices (6) are in (12), then this is a formulation for PARAFAC. These formulations are known as the Alternating Least-Squares (ALS) method which can be found in [29] for PARAFAC and [13] for BTD.

There are drawbacks of the ALS method. An artifact of the ALS algorithm is the so-called *swamp*. Swamp behavior occurs when there are exceedingly high number of iterations causing the convergence rate to slow down dramatically. Swamps have been observed when the component matrices are ill-conditioned or when collinearity occurs in the columns of these matrices.

3.2 Regularization Method

We reformulate the least squares sub-problems into three iterated Tikhonov sub-problems [15]:

$$\mathbf{A_{i+1}}^T = \text{argmin}_A \|\widehat{\mathbf{T}}^{JK\times I} - \mathbf{Q}\mathbf{A}^T\|_F^2 + \alpha_i \|\mathbf{A}^T - \mathbf{A_i}^T\|_F^2 \tag{13}$$

$$\mathbf{B_{i+1}}^T = \text{argmin}_B \|\widehat{\mathbf{T}}^{KI\times J} - \mathbf{R}\mathbf{B}^T\|_F^2 + \alpha_i \|\mathbf{B}^T - \mathbf{B_i}^T\|_F^2 \tag{14}$$

$$\mathbf{C_{i+1}}^T = \text{argmin}_C \|\widehat{\mathbf{T}}^{IJ\times K} - \mathbf{S}\mathbf{C}^T\|_F^2 + \alpha_i \|\mathbf{C}^T - \mathbf{C_i}^T\|_F^2 \tag{15}$$

where α_i is the regularization parameter and $\mathbf{A_i}$, $\mathbf{B_i}$ and $\mathbf{C_i}$ are the approximation from the ith iteration. The regularization terms $\alpha_i\|\mathbf{A} - \mathbf{A_i}\|$, $\alpha_i\|\mathbf{B} - \mathbf{B_i}\|$ and $\alpha_i\|\mathbf{C} - \mathbf{C_i}\|$ are the fitting terms for the factors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. It also forces the matrices $\mathbf{A_i}$, $\mathbf{B_i}$ and $\mathbf{C_i}$ to have the same scalings and permutation of the previous iterates. Both PARAFAC and BTD are *essentially unique*. For PARAFAC, this means that the factors are unique up to scalings and permutation. For BTD, essential uniqueness includes conjugation. As a consequence, the objective functional is *flat* which causes a sluggish rate of convergence. Moreover, the regularization term allows a sub-optimal *regularized* solution to an otherwise ill-posed problem.

3.2.1 Parameter Choice

Iterated Tikhonov regularization can be understood in two ways: either α is fixed and the iteration number is the main steering parameter or the iteration number is set *a priori* and α is tuned. Both strategies give a convergent regularization if the tuning parameter is based on the parameter choice rule. The parameter choice must be related to the noise level in the data according to a well-known result by Bakushinskii [4]. Otherwise a choice independent of the noise level cannot be expected to give convergent results in the worst case. In addition, the two schemes above can be combined by choosing a different regularization parameter at each iteration, e.g. a geometrically decaying sequence $\alpha_i = q^i$ with $0 < q < 1$. In fact, this is the strategy that we employed in the numerical examples in Section 4.

The parameter choice rule is based on Morozov's discrepancy rule [22]: the *a posteriori* choice of parameter α requires no prior knowledge of the parameter but the knowledge of the noise level which is defined as the distance between given noisy tensor and unknown true tensor, i.e.

$$\|\widehat{\mathscr{T}} - \mathscr{T}\|_F^2 \le \tau \cdot \sigma \tag{16}$$

where $\tau > 1$ is a factor to ensure that the residual is attained and σ is the noise level. The noise level is typically unknown but it must be assumed *a priori*. Thus, the discrepancy rule defines α or the iteration number as the optimal parameter when the residual has the same order as of the noise level. If α is fixed and the iteration number is the parameter to

Table 1: Alternating Least-Squares with iterated Tikhonov regularization

Given I_{max}, $\widehat{\mathscr{T}}$, c, ε
$\mathbf{A_0} = \mathtt{randn(I,R)}$, $\mathbf{B_0} = \mathtt{randn(J,R)}$ and $\mathbf{C_0} = \mathtt{randn(K,R)}$
$\alpha_0 = 1$
for $i = 1,2,3,\ldots,I_{max}$
  $\mathbf{A_i} \longleftarrow \min \|\widehat{\mathbf{T}}^{JK\times I} - \mathbf{Q_{i-1}}\mathbf{A}^T\|_F^2 + \alpha_i\|\mathbf{A}^T - \mathbf{A_{i-1}}^T\|_F^2$
    $\mathbf{R_i} = [(\mathbf{c_{i-1}})_\mathbf{1} \otimes (\mathbf{A_i})_\mathbf{1} \ldots (\mathbf{c_{i-1}})_\mathbf{R} \otimes (\mathbf{A_i})_\mathbf{R}]$
  $\mathbf{B_i} \longleftarrow \min \|\widehat{\mathbf{T}}^{KI\times J} - \mathbf{R_i}\mathbf{B}^T\|_F^2 + \alpha_i\|\mathbf{B}^T - \mathbf{B_{i-1}}^T\|_F^2$
    $\mathbf{S} = [(\mathbf{A_i})_\mathbf{1} \odot (\mathbf{B_i})_\mathbf{1}) 1_{L_1} \ldots (\mathbf{A_i})_\mathbf{R} \odot (\mathbf{B_i})_\mathbf{R}) 1_{L_R}$
  $\mathbf{C_i} \longleftarrow \min \|\widehat{\mathbf{T}}^{IJ\times K} - \mathbf{S_i}\mathbf{C}^T\|_F^2 + \alpha_i\|\mathbf{C}^T - \mathbf{C_{i-1}}^T\|_F^2$
    $\mathbf{Q_i} = [(\mathbf{B_i})_\mathbf{1} \otimes (\mathbf{c_i})_\mathbf{1} \ldots (\mathbf{B_i})_\mathbf{R} \otimes (\mathbf{c_i})_\mathbf{R}]$
  $\mathscr{T}_i = \mathtt{create\_tensor}(\mathbf{A_i}, \mathbf{B_i}, \mathbf{C_i})$
  if $\|\widehat{\mathscr{T}} - \mathscr{T}_i\| < \varepsilon$
    $i = I_{max}$
  end
  $\alpha_i = c\alpha_{i-1}$
end

be tuned, then the discrepancy rule dictates the termination of the loop. It stops at the first iteration index where the inequality (16) is satisfied. If the iteration number is set and the iterated Tikhonov regularization is implemented to obtain a sequence of decreasing parameters α_i, then the discrepancy principle essentially terminates the procedure at the largest α_i for which (16) holds. Similarly for the mixed method of geometrically decaying α_i, we stop the iteration when (16) is satisfied. Typically for linear problems, there exists such a parameter α_i or maximum number of iterations such that (16) holds and the algorithm usually terminates. Moreover, it has been shown in [15] that the regularization together with this parameter choice rule result in a convergent method; i.e. as the noise level $\sigma \to 0$, the regularized solution will tend to the true one.

Also there is the possibility of an *a priori* parameter choice rule. For instance, one can choose

$$\alpha^* \approx \tau \cdot \sigma. \tag{17}$$

However, there are cases when this choice does not give the optimal order of convergence in contrast to the discrepancy rule. See the pseudocode in Table 1. Notice only a few additional lines of coding is required.

4. NUMERICAL RESULTS

We consider two models, PARAFAC and BTD, to demonstrate the swamp shortening property of the iterated Tikhonov regularization. We start with simple PARAFAC model [26] with

$$\mathbf{A} = \begin{bmatrix} 1 & \cos\theta & 0 \\ 0 & \sin\theta & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 3 & \sqrt{2}\cos\theta & 0 \\ 0 & \sin\theta & 1 \\ 0 & \sin\theta & 0 \end{bmatrix},$$

and $\mathbf{C} = \mathbf{I}_{3\times 3}$. Notice that as $\theta \to 0$, the columns of $\mathbf{A}$ and $\mathbf{B}$ become collinear. In Figure 1, the plot shows the error $\|\widehat{\mathscr{T}} - \mathscr{T}_{est}\|_F^2$ versus the number of iterations it takes to obtain an error of 1×10^{-5} for $\theta = \frac{\pi}{60}$. At $\theta = \frac{\pi}{60}$ the iterated Tikhonov method takes 41 iterations to reach an error within 10^{-5} while the least-squares method need 683 iterations. As θ is decreased both methods need more iterations. For the

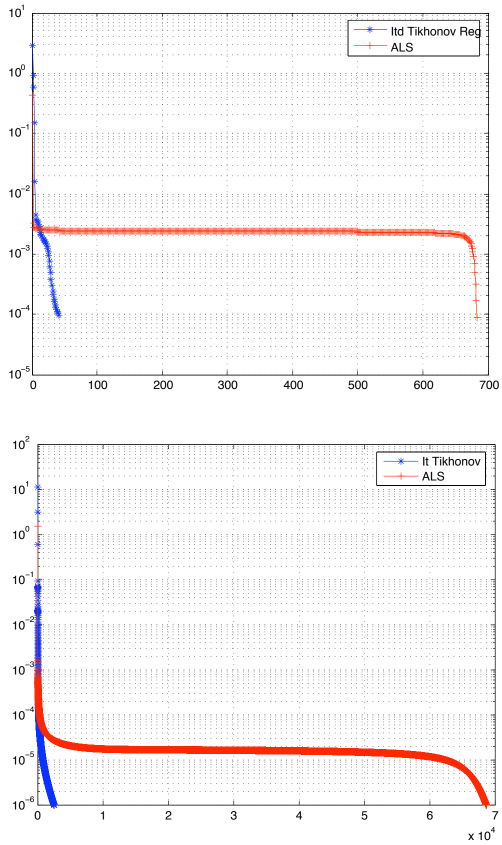

Figure 1: Error vs Iteration using iterated Tikhonov regularization (blue) and least-squares method (red) with $\theta = \frac{\pi}{60}$. PARAFAC (top) and BTD (bottom)

iterated Tikhonov regularization, 69 and 311 iterations are needed to attain an error of 1×10^{-5} for $\theta = \frac{\pi}{90}$ and $\theta = \frac{\pi}{120}$, respectively. However, ALS requires 12007 ($\theta = \frac{\pi}{90}$) and 12026 ($\theta = \frac{\pi}{120}$) iterations for the error to be within 1×10^{-3}.

In the second example, we find the BTD of the tensor $\mathscr{T} \in \mathbb{R}^{4\times 4\times 6}$ of rank-$(2,2,1)$ with the component matrices $\mathbf{A_i}, \mathbf{B_i} \in \mathbb{R}^{4\times 2}$ for $i = 1,2,3$ and $\mathbf{C} \in \mathbb{R}^{6\times 3}$. Here $\mathbf{A} = [\mathbf{A_1}\mathbf{A_2}\mathbf{A_3}]$ where

$$\mathbf{A_1} = \begin{bmatrix} 1 & \cos\theta \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{A_2} = \begin{bmatrix} 0 & 0 \\ 1 & \cos\theta \\ 0 & \sin\theta \\ 1 & \cos\theta \end{bmatrix}, \quad \mathbf{A_3} = \begin{bmatrix} 0 & \sin\theta \\ 1 & \cos\theta \\ 0 & 0 \\ 1 & 1 \end{bmatrix},$$

$\mathbf{B} = \mathbf{A}$ and $\mathbf{C}$ is randomly generated. The bottom graph in Figure 1 displays 2500 iterations for our method versus 68630 ALS iterations for the residual to be within 1×10^{-7}. The *swamp* regime lengthens as the columns become more collinear in the ALS case. In general, more iterations are required for both methods when the dimensions I, J, K are increased.

Acknowledgments

L. De Lathauwer was supported in part by (1) Research Council K.U.Leuven: GOA-Ambiorics, CoE EF/05/006 Optimization in Engineering (OPTEC), CIF1, (2) F.W.O.: (a) project G.0321.06, (b) Research Communities ICCoS, AN-

MMM and MLDM, (3) the Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, “Dynamical systems, control and optimization”, 2007–2011), (4) EU: ERNSI. A large part of this research was carried out when C. Navasca and L. De Lathauwer were with the French Centre National de la Recherche Scientifique (C.N.R.S.).

REFERENCES

[1] E. Acar, C. A. Bingol, H. Bingol, R. Bro and B.Yener, “Multiway Analysis of Epilepsy Tensors,” *ISMB 2007 Conference Proceedings, Bioinformatics*, 23(13): i10-i18, 2007.

[2] E. Acar and B. Yener, “Unsupervised Multiway Data Analysis: a Literature Survey”, submitted for publication.

[3] A. L. F . de Almeida, G. Favier and J. C. M. Mota, ”PARAFAC-Based Unified Tensor Modeling for Wireless Communication Systems with Application to Blind Multiuser Equalization,” *Journal of Signal Processing*, vol 87, pp. 337-351, 2007.

[4] A. Bakushinsky, “Remarks on Choosing a Regularization Parameter Using the Quasioptimality and Ratio Criterion,” *USSR Comput. Maths. Math. Phys.*, vol 24, pp. 181–182, 1984.

[5] R. Barrett et al., *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, SIAM, 1994.

[6] B. W. Bader and R. B. Schnabel, “On the Performance of Tensor Methods for Solving Ill-conditioned Problems,” *SIAM J. Scientific Computing, 2007.*

[7] M. Berry, M. Browne, A. Langville, P. Pauca, and R. Plemmons, “Algorithms and Applications for Approximate Nonnegative Matrix Factorization,” *Computational Statistics & Data Analysis*, vol 52 no 1, pp. 155-173, 2007.

[8] G. Beylkin and M.J. Mohlenkamp, “Algorithms for Numerical Analysis in High Dimensions,” *SIAM J. Sci. Comput.*, vol. 26, pp. 2133–2159, 2005.

[9] J. Carrol and J. Chang, “Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of ”Eckart-Young” Decomposition” *Psychometrika*, 9, 267-283, 1970.

[10] L. De Lathauwer, “Decompositions of a Higher-Order Tensor in Block Terms — Part I: Lemmas for Partitioned Matrices,” *SIAM J. Matrix Anal. Appl.*, to appear.

[11] L. De Lathauwer, “Decompositions of a Higher-Order Tensor in Block Terms — Part II: Definitions and Uniqueness,” *SIAM J. Matrix Anal. Appl.*, to appear.

[12] L. De Lathauwer, B. De Moor, and J. Vandewalle, “A Multilinear Singular Value Decomposition,” *SIAM J. Matrix Anal. Appl.*, 21, pp. 1253–1278, 2000.

[13] L. De Lathauwer and D. Nion, “Decompositions of a Higher-Order Tensor in Block Terms — Part III: Alternating Least Squares Algorithms,” *SIAM J. Matrix Anal. Appl.*, to appear.

[14] M. De Vos, A. Vergult, L. De Lathauwer, W. De Clercq, S. Van Huffel, P. Dupont, A. Palmini, W. Van Paesschen, “Canonical Decomposition of Ictal EEG Reliably Detects the Seizure Onset Zone”, *NeuroImage*, vol. 37, no. 3, pp. 844–854, Sep. 2007.

[15] H. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*. Kluwer Dordrecht, 1996.

[16] W. Hackbush, B.N. Khoromskij and E. Tyrtyshnikov, “Hierarchical Kronecker Tensor-Product Approximations,” *J. Numer. Math.*, vol. 13, no. 2, pp. 119–156, 2005.

[17] R. A. Harshman, “Foundations of the PARAFAC procedure: Model and Conditions for an ”Explanatory” Mutli-code Factor Analysis,” *UCLA Working Papers in Phonetics*, 16, pp. 1-84, 1970.

[18] B. N. Khoromskij and V. Khoromskaia, “Low Rank Tucker-Type Tensor Approximation to Classical Potentials,” *Central European Journal of Mathematics*, vol. 5, no. 3, pp. 523–550, 2007.

[19] T. G. Kolda and B. W. Bader, “Tensor Decompositions and Applications,” Technical Report Number SAND2007-6702, Sandia National Laboratories, Albuquerque, NM and Livermore, CA, Novermber 2007.

[20] P.M. Kroonenberg, *Applied Multiway Data Analysis*, Wiley, 2008.

[21] J. B. Kruskal, “Three-way Arrays: Rank and Uniqueness of Trilinear Decompositions with Application to Arithmetic Complexity and Statistics,” *Linear Algebra Applicat.,* vol. 18 pp. 95-138, 1977.

[22] V. A. Morozov, *Regularization Methods for Ill-Posed Problems*. CRC Press, Boca Raton, FL, 1993.

[23] M. Mørup, L.K. Hansen, C.S. Herrmann, J. Parnas and S.M. Arnfred, “Parallel Factor Analysis as an Exploratory Tool for Wavelet Transformed Event-Related EEG,” *NeuroImage*, vol. 29, pp. 938–947, 2006.

[24] D. Nion and L. De Lathauwer, “An Enhanced Line Search Scheme for Complex-Valued Tensor Decompositions. Application in DS-CDMA,” *Signal Processing*, vol. 88, no. 3, pp. 749–755, 2008.

[25] D. Nion and L. De Lathauwer, “Block Component Model Based Blind DS-CDMA Receivers,” *IEEE Trans. Signal Processing*, to appear.

[26] M. Rajih and P. Comon, ”Enhanced Line Search: A Novel Method to Accelerate PARAFAC,” *Proc. EUSIPCO 2005* Antalya, Turkey, September 4-8, 2005.

[27] C.R. Rao and S.K. Mitra, *Generalized Inverse of Matrices and Its Applications*. Wiley, New York, 1971.

[28] N.D. Sidiropoulos, G.B. Giannakis and R. Bro, “Blind PARAFAC Feceivers for DS-CDMA Systems,” *IEEE Trans. on Signal Processing*, vol. 48, no. 3, pp. 810–823, 2000.

[29] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis. Applications in the Chemical Sciences*. John Wiley and Sons, Chichester, U.K., 2004.

[30] L.R. Tucker, “The Extension of Factor Analysis to Three-dimensional Matrices,“in *Contributions to mathematical psychology*, H. Gulliksen and N. Frederiksen, eds., Holt, Rinehart & Winston, NY, 1964, pp. 109-127.

[31] L.R. Tucker, “Some Mathematical Notes on Three-mode Factor Analysis,” *Psychometrika*, 31, pp. 279-311, 1966.

[32] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos and A.B. Gershman, “Robust Iterative Fitting of Multilinear Models”, *IEEE Trans. on Signal Processing*, 53, pp. 2678–2689, 2005.

[33] H.-G. Zhao, “A Heuristic Method for Computing the Best Rank-r Approximation to Higher-Order Tensors,” *Int. J. Contemp. Math. Sciences*, vol 3, no. 10, pp. 471 - 476, 2008.