LOW MULTILINEAR RANK TENSOR APPROXIMATION VIA SEMIDEFINITE PROGRAMMING

Carmeliza Navasca[♮] *and Lieven De Lathauwer*[♯♭]

[♮]Department of Mathematics, Clarkson University
Box 5815, Potsdam, New York 13699, USA
phone: + (1) 315 268 2369, email: cnavasca@clarkson.edu
web: people.clarkson.edu/∼cnavasca
and
[♯]K.U. Leuven, Research Group ESAT-SCD
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium and
[♭]K.U. Leuven Campus Kortrijk, Science, Engineering and Technology Group,
E. Sabbelaan 53, 8500 Kortrijk, Belgium
phone: +32 16 328651, e-mail: delathau@esat.kuleuven.be

ABSTRACT

We present a novel method for tensor dimensionality reduction. The tensor rank reduction has many applications in signal and image processing including various blind techniques. In this paper, we generalize the trace class norm to higher-order tensors. Recently, the matrix trace class has received much attention in the compressed sensing applications. It is known to provide bounds for the minimum rank of a matrix. In this paper, a new tensor trace class norm is used to formulate an optimization problem for finding the best low multilinear rank tensor approximation. Our new formulation leads to a set of semidefinite programming subproblems where the nth subproblem approximates a low multilinear rank factor in the nth modal direction. Our method is illustrated on a real-life data set.

1. INTRODUCTION

The optimization problem of approximating a tensor $\mathscr{T} \in \mathbb{R}^{I\times J\times K}$ by a low multilinear rank tensor $\widehat{\mathscr{T}}$,

$$\min_{\mathrm{rank}_n(\widehat{\mathscr{T}})=R_n} \|\mathscr{T} - \widehat{\mathscr{T}}\|_F$$

is of interest across many applications, for example, in signal and image processing. In independent component analysis, electro-encephalography, magneto-encephalography, nuclear magnetic resonance, etc., high-dimensional data with very few significant signal source contributions are ubiquitous. In image processing, tensor dimensionality reduction has been applied to image synthesis, analysis and recognition. Also tensor multilinear rank reduction has been particular useful in estimation of poles and complex amplitudes in harmonic retrieval. The tensor Frobenius norm $\|\cdot\|_F$ is defined as

$$\|\mathscr{T}\|_F = \left(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}|t_{ijk}|^2\right)^{\frac{1}{2}} \tag{1}$$

which is equivalent to the ℓ_2 norm of the singular values at each mode [4]. The desired solution, $\mathscr{T}^* = \arg\min\|\mathscr{T} - \widehat{\mathscr{T}}\|_F$ has a specific structure. If $\mathbf{T} \in \mathbb{R}^{I\times J}$ is a second order tensor (a matrix), then the best rank-R approximation is

$$\mathbf{T}^* = \arg\min\|\mathbf{T} - \widehat{\mathbf{T}}\|_F = \mathbf{U}_R\mathbf{\Sigma}_R\mathbf{V}_R^T$$

which can be computed by means of truncated SVD. The matrix $\mathbf{\Sigma}_R \in \mathbb{R}^{R\times R}$ is a diagonal matrix with the first R singular values of $\mathbf{T}$. The matrices $\mathbf{U}_R \in \mathbb{R}^{I\times R}$ and $\mathbf{V}_R \in \mathbb{R}^{J\times R}$ are the first R columns of the full column rank $\mathbf{U}$ and $\mathbf{V}$ through truncated SVD. Certainly, we would like to extend the similar concept to higher-order tensor. The higher-order SVD (HO-SVD) [6, 19, 20, 21] plays an important role in finding an approximation. Through HO-SVD, the tensor singular values and vectors are calculated mode by mode via successive SVD. Although the HO-SVD is extremely useful in finding singular values and vectors, it does not in general provide the best low multilinear rank approximation and certainly not the best low rank approximation.

There are several current well-known methods for tensor low multilinear rank approximation, namely, the truncated Higher-Order SVD (HO-SVD) [6, 21], Higher-Order Orthogonal Iteration (HOOI) [5, 12]. Although these methods are widely used in applications, these techniques have some shortcomings. The truncated HO-SVD, in general, does not give the best low multilinear rank approximation while HOOI does not guarantee global optimal solution and can be slow. There are also the current quasi-Newton schemes on Grassmannian manifolds [9, 10, 16, 7] which have been demonstrated to converge to local optima super-linearly and quadratically.

In this paper, we develop an alternative approach for reducing multilinear tensor rank by extending the matrix rank minimization problem to problems involving higher order tensors. The matrix rank minimization problems come from the subject areas of optimization and convex analysis. Lately, these rank minimization problems have received much attention in compressed sensing [3, 15, 2]. Thus it is worthwhile to extend these formulations to multilinear algebra to widen and improve the applicability of tensors.

1.1 Organization

Beginning with Section 2, we give some preliminaries which include basic definition, tensor decompositions and tensor unfolding techniques. In addition, HO-SVD is discussed in details. In Section 3, we describe the trace class norm and semidefinite programming for tensors. In the next section, we discuss our technique for computing the low multilinear rank factors through semidefinite programming. We also illustrate the attributes of the new method through some numerical results. Finally, we conclude in Section 5 by including some remarks on our future work.

2. PRELIMINARIES

We denote the scalars in $\mathbb{R}$ 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: $(\mathscr{A})_{ijk}=a_{ijk}$, $(\mathbf{A})_{ij}=a_{ij}$, $(\mathbf{a})_i=a_i$. 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. In addition, the lower-case superscripts on a matrix indicate the mode in which it has been matricized. For example, $\mathbf{R}^n$ is the mode-n matricization of the tensor $\mathscr{R}\in\mathbb{R}^{I\times J\times K}$ for $n=1,2,3$.

Definition 2.1 *The Kronecker product of matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *is defined as*

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

Definition 2.2 (Mode-n vector) *Given a tensor $\mathscr{T}\in\mathbb{R}^{I\times J\times K}$, 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 2.3 (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 cardinality of the index set. A matrix is a second-order tensor and a vector is a first-order tensor.

Definition 2.4 (rank-(L,M,N)) *A third-order tensor $\mathscr{T}\in\mathbb{R}^{I\times J\times K}$ 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. It is often denoted as $rank_1(\mathscr{T})=L$, $rank_2(\mathscr{T})=M$ and $rank_3(\mathscr{T})=N$.*

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

Definition 2.5 (Tucker mode-n product) *Given a tensor $\mathscr{T}\in\mathbb{R}^{I\times J\times K}$ and the matrices $\mathbf{A}\in\mathbb{R}^{\hat{I}\times I}$, $\mathbf{B}\in\mathbb{R}^{\hat{J}\times J}$ and $\mathbf{C}\in\mathbb{R}^{\hat{K}\times K}$, then the Tucker mode-n products are as follows:*

$$\begin{aligned}
(\mathscr{T}\bullet_1\mathbf{A})_{\hat{i},j,k} &= \sum_{i=1}^{I} t_{ijk}a_{\hat{i}i},\ \forall\hat{i},j,k \quad \textit{(mode-1 product)}\\
(\mathscr{T}\bullet_2\mathbf{B})_{\hat{j},i,k} &= \sum_{j=1}^{J} t_{ijk}b_{\hat{j}j},\ \forall\hat{j},i,k \quad \textit{(mode-2 product)}\\
(\mathscr{T}\bullet_3\mathbf{C})_{\hat{k},i,j} &= \sum_{k=1}^{K} t_{ijk}c_{\hat{k}k},\ \forall\hat{k},i,j \quad \textit{(mode-3 product)}
\end{aligned}$$

Definition 2.6 (Matrix Slice and Subtensor) *A third-order tensor $\mathscr{S}\in\mathbb{R}^{I\times J\times K}$ has three types of matrix slices obtained by fixing the index of one of the modes. The matrix slices of $\mathscr{S}\in\mathbb{R}^{I\times J\times K}$ are the following: $\mathbf{S}^1_{i=\alpha}\in\mathbb{R}^{J\times K}$ with fixed $i=\alpha$, $\mathbf{S}^2_{j=\alpha}\in\mathbb{R}^{I\times K}$ with fixed $j=\alpha$ and $\mathbf{S}^3_{k=\alpha}\in\mathbb{R}^{I\times J}$ with fixed $k=\alpha$. For an Nth-order tensor $\mathscr{S}\in\mathbb{R}^{I_1\times I_2\times I_3\times\ldots\times I_N}$, the subtensors are the $(N-1)$th-order tensors denoted by $\mathscr{S}^n_{i_n=\alpha}\in\mathbb{R}^{I_1\times I_2\times I_3\times\ldots\times I_{n-1}\times I_{n+1}\times\ldots\times I_N}$ which are obtained by fixing the index of the nth mode.*

2.1 Higher-Order SVD

The higher-order SVD is also referred to multilinear SVD. In the following theorem, we discuss the HO-SVD for third order tensors for the sake of simplicity and clarity. Of course the HO-SVD applies for Nth-order tensors.

Theorem 2.1 (Multilinear SVD [6]) *A third order tensor $\mathscr{T}\in\mathbb{R}^{I\times J\times K}$ can be represented as a product*

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

where

1. $\mathbf{A}\in\mathbb{R}^{I\times I}$ *is an orthogonal matrix and* $\mathbf{A}=[\mathbf{a_1}\ \ldots\mathbf{a_I}]$
2. $\mathbf{B}\in\mathbb{R}^{J\times J}$ *is an orthogonal matrix and* $\mathbf{B}=[\mathbf{b_1}\ \ldots\mathbf{b_J}]$
3. $\mathbf{C}\in\mathbb{R}^{K\times K}$ *is an orthogonal matrix and* $\mathbf{C}=[\mathbf{c_1}\ \ldots\mathbf{c_K}]$
4. $\mathscr{S}\in\mathbb{R}^{I\times J\times K}$ *is a third order tensor with subtensors (matrices) $\mathbf{S}^1_{i=\alpha}\in\mathbb{R}^{J\times K}$, $\mathbf{S}^2_{j=\alpha}\in\mathbb{R}^{I\times K}$ and $\mathbf{S}^3_{k=\alpha}\in\mathbb{R}^{I\times J}$ with the following properties:*
   - *all-orthogonality:*
     $$\langle\mathbf{S}^1_{i=\alpha},\mathbf{S}^1_{i=\beta}\rangle=(\sigma^{(1)}_\alpha)^2\delta_{\alpha,\beta},\ \ \alpha,\beta=1,\ldots,I,$$
     $$\langle\mathbf{S}^2_{j=\alpha},\mathbf{S}^2_{j=\beta}\rangle=(\sigma^{(2)}_\alpha)^2\delta_{\alpha,\beta},\ \ \alpha,\beta=1,\ldots,J,$$
     $$\langle\mathbf{S}^3_{k=\alpha},\mathbf{S}^3_{k=\beta}\rangle=(\sigma^{(3)}_\alpha)^2\delta_{\alpha,\beta},\ \ \alpha,\beta=1,\ldots,K$$
   - *ordering:*
     $\|\mathbf{S}^1_{i=1}\|_F\ge\|\mathbf{S}^1_{i=2}\|_F\ge\cdots\ge\|\mathbf{S}^1_{i=I}\|_F\ge0,$
     $\|\mathbf{S}^2_{j=1}\|_F\ge\|\mathbf{S}^2_{j=2}\|_F\ge\cdots\ge\|\mathbf{S}^2_{j=J}\|_F\ge0,$
     $\|\mathbf{S}^3_{k=1}\|_F\ge\|\mathbf{S}^3_{k=2}\|_F\ge\cdots\ge\|\mathbf{S}^3_{k=K}\|_F\ge0$

   where $\|\mathbf{S}^n_{i=\alpha}\|_F=\sigma^{(n)}_\alpha$ for $\alpha=1,\ldots,I_n$ ($I_1=I,I_2=J,I_3=K$).

The usual inner product of matrices, $\mathbf{A},\mathbf{B}\in\mathbb{R}^{I\times J}$ is denoted by $\langle\mathbf{A},\mathbf{B}\rangle=\sum_{ij}b_{ij}a_{ij}$. For a third order tensor, there are three sets of singular values: $\sigma^{(1)}_\alpha$'s are mode-1 singular values, $\sigma^{(2)}_\alpha$'s are the mode-2 singular values and $\sigma^{(3)}_\alpha$'s are the mode-3 singular values. The corresponding mode-1, mode-2 and mode-3 singular vectors are $\mathbf{a}_\alpha$, $\mathbf{b}_\alpha$ and $\mathbf{c}_\alpha$, respectively. The all-orthogonality property implies the simultaneous mutual orthogonality between different *horizontal* slices, *vertical* slices and *frontal* slices with respect to the scalar product of matrices.

For an Nth-order $\mathscr{T}\in\mathscr{R}^{I_1\times I_2\times I_3\times I_4\times\ldots\times I_N}$ tensor, the HOSVD is

$$\mathscr{T}=\mathscr{S}\bullet_1\mathbf{U^1}\bullet_2\mathbf{U^2}\bullet_3\mathbf{U^3}\bullet_4\mathbf{U^4}\bullet_5\ldots\bullet_N\mathbf{U^N} \qquad (2)$$

where $\mathbf{U^n}\in\mathbb{R}^{I_n\times I_n}$ are orthogonal matrices and $\mathscr{S}\in\mathbb{R}^{I_1\times I_2\times I_3\times I_4\times\ldots\times I_N}$ is a core tensor with subtensors $\mathscr{S}^n_n\in\mathbb{R}^{I_1\times I_2\times\ldots I_{n-1}\times I_{n+1}\times\ldots\times I_N}$ fixed at the i_nth index. The mode-n

singular values are $\|\mathscr{S}^n_{n=1}\|_F$, $\|\mathscr{S}^n_{n=2}\|_F$, ..., $\|\mathscr{S}^n_{n=I_n}\|_F$ corresponding to the mode-n singular vectors u^n_1, u^n_2, ..., $u^n_{I_n}$ of the orthogonal matrix $\mathbf{U^n}$. We denote the mode-n singular value as $\sigma^{(n)}_\alpha = \|\mathscr{S}^n_{n=\alpha}\|_F$.

The following matrix representations of the HO-SVD is obtained by unfolding the third-order $\mathscr{T}$ and $\mathscr{S}$ tensors in (2):

$$\mathbf{T}^1 = \mathbf{A}\mathbf{S}^1(\mathbf{B}\otimes\mathbf{C})^T, \mathbf{T}^2 = \mathbf{B}\mathbf{S}^2(\mathbf{C}\otimes\mathbf{A})^T, \mathbf{T}^3 = \mathbf{C}\mathbf{S}^3(\mathbf{A}\otimes\mathbf{B})^T$$

We denote $\mathbf{T}^1 = \mathbf{T}^{I\times JK}$, $\mathbf{T}^2 = \mathbf{T}^{J\times KI}$, $\mathbf{T}^3 = \mathbf{T}^{K\times IJ}$ and similarly for $\mathbf{S}^n$. In general, $\mathbf{T}^n$ and $\mathbf{S}^n$ are mode-n matrix representation of $\mathscr{T}$ and $\mathscr{S}$.

3. TRACE CLASS MINIMIZATION

We start with the trace class norm of the matrix (also referred to as the *Schatten-1 norm* [17] or more recently, *nuclear norm* [15], [2]). The matrix $\mathbf{T}\in\mathbb{R}^{I\times J}$ is a second order tensor which has an SVD of $\mathbf{T} = \mathbf{A\Sigma B}$ where $\mathbf{\Sigma} = \mathbf{diag}\{\sigma_1,\sigma_2,\ldots,\sigma_p\}$ where $p = \min\{I,J\}$. The trace class norm of $\mathbf{T}$ is the sum of its singular values; i.e.

$$\|\mathbf{T}\|_{tr} = \sigma_1+\sigma_2+\ldots+\sigma_p. \qquad (3)$$

In the paper [14], we generalize the trace class norm to higher-order tensors as follows

$$\|\mathscr{T}\|_{tr_{(n)}} = \|\mathbf{T}^n\|_{tr} = \sum_{\alpha=1}^{I_n}\sigma^{(n)}_\alpha \qquad (4)$$

where $\mathbf{T}^n$ is a matrix slice of $\mathscr{T}$ fixed at the nth mode. We refer to the norm $\|\cdot\|_{tr_{(n)}}$ as the mode-n tensor trace class norm. There are N trace class norms for an Nth order tensor which is consistent with the fact that there are also N mode-n ranks and N sets of mode-n singular values.

The nth mode tensor trace class norm (4) is the sum of the mode-n singular values. Recall that the mode-n singular values are $\sigma^{(n)}_\alpha = \sqrt{\langle\mathscr{S}^n_{n=\alpha},\mathscr{S}^n_{n=\alpha}\rangle} = \|\mathscr{S}^n_{n=\alpha}\|_F$ where the tensor $\mathscr{S}^n_{n=\alpha}\in\mathbb{R}^{I_1\times\ldots\times I_{n-1}\times I_{n+1}\times\ldots\times I_N}$ is the subtensor fixed at nth mode of the core tensor $\mathscr{S}\in\mathbb{R}^{I_1\times\ldots\times I_n\times\ldots\times I_N}$. The norm (4) is consistent with the matrix trace class norm for $\mathbf{T}\in\mathbb{R}^{I\times K}$ because the singular values of the two mode ranks are equivalent; i.e.

$$\|\mathbf{T}\|_{tr_{(1)}} = \sum_{i=1}^{I}\sqrt{\langle\Sigma^1_{i=1},\Sigma^1_{i=1}\rangle} \text{ and } \|\mathbf{T}\|_{tr_{(2)}} = \sum_{j=1}^{J}\sqrt{\langle\Sigma^2_{j=1},\Sigma^2_{j=2}\rangle}$$

where $\Sigma^1_{i=\alpha}$ $(\Sigma^2_{j=\alpha})$ is the αth row (column) vector of the diagonal core matrix $\mathbf{\Sigma}$. Due to the pseudodiagonality in the matrix SVD, the two sums are equivalent. For the higher order tensors, the sums of the singular values at each mode are not necessarily equivalent. The core tensor satisfies the all-orthogonality property which does not imply a tensor with only nonzero entries on its super-diagonal.

3.1 Semidefinite Programming for Tensors

The dual norm of the trace class norm (4) has been proven to be

$$\|\mathscr{T}\|_{tr^*_{(n)}} = \max_\alpha\left\{\sigma^{(n)}_\alpha\right\} = \sigma^{(n)}_1$$

in [14] where $\sigma^{(n)}_1 = \|\mathscr{S}^n_{i_n=1}\|_F$ is the maximum singular value in the set of the mode-n singular values. If we take the variational definition [11] of (4)

$$\|\mathscr{T}\|_{tr_{(n)}} = \max\ \langle\mathscr{T},\mathscr{R}\rangle, \quad \text{subject to } \|\mathscr{R}\|_{tr^*_{(n)}}\le 1 \qquad (5)$$

for all $n = 1,\ldots,N$ for a given tensor $\mathscr{T}$, then the nth mode tensor trace class norm provides an inherent optimization problem.

With the tensor unfolding techniques, the constraint in (5) becomes

$$\|\mathscr{R}\|_{tr^*_{(n)}}\le 1 \implies \|\mathbf{R}^n\|_{tr^*}\le 1$$

where $\mathbf{R}^n$ is the mode-n matricization of $\mathscr{R}$ and $\|\mathbf{R}^n\|_{tr^*}$ denotes the largest singular value of $\mathbf{R}^n$. Then from [1, 14],

$$\|\mathbf{R}^n\|_{tr*}\le 1 \implies \begin{bmatrix}\mathbf{I}^{I_n\times I_n} & \mathbf{R}^n\\ (\mathbf{R}^n)^T & \mathbf{I}^{I_m\times I_m}\end{bmatrix}\succeq 0 \qquad (6)$$

and $\mathbf{I}^{I_k\times I_k}$ is the identity matrix of dimension $I_k\times I_k$. The symbol $\succeq 0$ denotes a positive semidefinite matrix.

Let

$$\mathbf{M}^n = \begin{bmatrix}\mathbf{I}^{I_n\times I_n} & \mathbf{R}^n\\ (\mathbf{R^n})^T & \mathbf{I}^{I_m\times I_m}\end{bmatrix}, \qquad (7)$$

then we can formulate a set of N semidefinite programming problems (SDP),

$$\|\mathbf{T}^n\|_{tr_{(n)}} = \max\ \langle\mathbf{T}^n,\mathbf{R}^n\rangle, \quad \text{subject to } \mathbf{M}^n\succeq 0 \qquad (8)$$

for each mode n of an Nth-order tensor. Given the matrix representation of $\mathscr{T}$ by $\mathbf{T}^n$, we find an optimal matrix $\mathbf{R}^n$ which obtains the sum of the singular values of $\mathbf{T}^n$ constrained to the positive semidefinite $\mathbf{M}^n$. The SDP (8) is consistent with the matrix trace class norm defined in (3). It was in [22] that the matrix trace class norm was heuristically formulated as a SDP.

From [15, 2], it has been shown through convex analysis that the trace class norm of a matrix gives the optimal lower bound of the rank. In the algorithm described in the next section, we use the trace class norm to find low multilinear rank factors through N subproblems of matrix rank minimization.

4. ALGORITHM

The SDP framework (8) is used to find the low multilinear rank factors. For a given third order tensor $\mathscr{T}\in\mathbb{R}^{I\times J\times K}$, we initially implement three sub-SDP problems iteratively to find initial factors $\mathbf{A}^0$, $\mathbf{B}^0$ and $\mathbf{C}^0$:

$$\min_{rank_i(\widehat{\mathbf{T}})=L}\|\mathbf{T}^i-\widehat{\mathbf{T}}^i\|_{tr},\ \text{for } i=1,2,3, \text{respectively}, \qquad (9)$$

where $\mathbf{T}^n$ is a matricization of $\mathscr{T}$ and $\widehat{\mathbf{T}}^n$'s are the matricization of the the unknown $\widehat{\mathscr{T}}$. The matrices $\mathbf{A}^0$, $\mathbf{B}^0$ and $\mathbf{C}^0$) are the approximated dominant left singular vectors of $\mathbf{T}^1$, $\mathbf{T}^2$ and $\mathbf{T}^3$. For example, the first $I-R_1$ columns of the factor $\mathbf{A}^0$ is truncated. Then the truncated matrix is used to initialize the following algorithm below. Note that the factors are approximated by the solving (9) through an optimization solver in [8]. Now the column space of the iterate $\mathbf{A}^{it}$ ($\mathbf{B}^{it}$ and $\mathbf{C}^{it}$) is the dominant subspace of the column space of R_1

(R_2 and R_3). Then, to continue to approximate $\mathbf{A}^{it}$, $\mathbf{B}^{it}$ and $\mathbf{C}^{it}$, the following equations are updated and the SDPs are solved iteratively and alternatingly:

$$\begin{cases} (\widehat{\mathbf{S}}^1)^{it+1} = \mathbf{T}^1(\mathbf{B}^{it} \otimes \mathbf{C}^{it}) \\ \max \ \langle (\widehat{\mathbf{S}}^1)^{it+1}, \mathbf{R}^1 \rangle \\ \text{subject to } \mathbf{M}^1 \succeq 0 \end{cases}$$

$$\begin{cases} (\widehat{\mathbf{S}}^2)^{it+1} = \mathbf{T}^2(\mathbf{C}^{it} \otimes \mathbf{A}^{it+1}) \\ \max \ \langle (\widehat{\mathbf{S}}^2)^{it+1}, \mathbf{R}^2 \rangle \\ \text{subject to } \mathbf{M}^2 \succeq 0 \end{cases}$$

$$\begin{cases} (\widehat{\mathbf{S}}^3)^{it+1} = \mathbf{T}^3(\mathbf{A}^{it+1} \otimes \mathbf{B}^{it+1}) \\ \max \ \langle (\widehat{\mathbf{S}}^3)^{it+1}, \mathbf{R}^3 \rangle \\ \text{subject to } \mathbf{M}^3 \succeq 0 \end{cases}$$

where $\mathbf{T}^1$, $\mathbf{T}^2$ and $\mathbf{T}^3$ are matrix representations of the given tensor $\mathscr{T}$ and the matrices $\mathbf{R}^n$ and $\mathbf{M}^n$ are of the form (6) and (7). At each iteration above, we obtain the low multilinear rank factors: $\mathbf{A}^{it+1} \in \mathbb{R}^{I \times L}, \mathbf{B}^{it+1} \in \mathbb{R}^{J \times M}, \mathbf{C}^{it+1} \in \mathbb{R}^{K \times N}$ from each SDP problems. We assume $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are column-wise orthonormal.

4.1 Numerical Experiments

In these experiments, the original tensor data is approximated by a low multilinear rank tensor using both the HOOI and SDP algorithms. We have used the Matlab codes available at [8] and [18] for the SDP algorithm implementation.

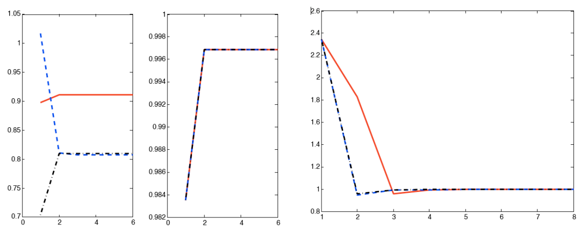

Figure 1: Compression Ratios of SDP (left,middle) and HOOI algorithm (right): mode-1(solid), mode-2 (dash), mode-3 (dash-dot)

A low multilinear rank core tensor of dimension $2 \times 2 \times 2$ (R=L=M=N=2) is computed from an original tensor data of dimension $5 \times 10 \times 14$ in the first experiment. The biomedical system dataset [13] is based on the tongue displacement shapes occurring in the pronunciation of English vowels by different English-speaking individuals which has a dominant rank-$(2,2,2)$. The dataset is a real-valued $(5 \times 10 \times 14)$-array obtained from high-quality audio recordings and cinefluorograms. In Figures 1, the initial factors are randomly generated. We plot the compression ratios using the HOOI and SDP algorithms versus the number of iterations where $\mathbf{T}$ is the original data and $\widehat{\mathbf{T}}$ is the approximated one. For the left and middle graphs, the convergence is achieved after three iterations using the SDP algorithm while the right graph took 4-7 iterations when the initial factors are generated randomly via the HOOI algorithm. In the left graph, the three curves correspond to $\frac{\|\widehat{\mathbf{T}}^1\|_{tr}}{\|\mathbf{T}^1\|_{tr}} = 0.9117$, $\frac{\|\widehat{\mathbf{T}}^2\|_{tr}}{\|\mathbf{T}^2\|_{tr}} = 0.8078$, and $\frac{\|\widehat{\mathbf{T}}^3\|_{tr}}{\|\mathbf{T}^3\|_{tr}} = 0.8093$, while the curves in the middle graph are the ratios $\frac{\|\widehat{\mathbf{T}}^n\|_F}{\|\mathbf{T}^n\|_F} = 0.9969$ for all modes and the right plot is the compression ratio of the HOOI algorithm with $\frac{\|\widehat{\mathbf{T}}^n\|_F}{\|\mathbf{T}^n\|_F} = 0.9968$. If the truncated SVD starters were used as an initial factors, the HOOI iterations are reduced while it has no effect in the SDP algorithm.

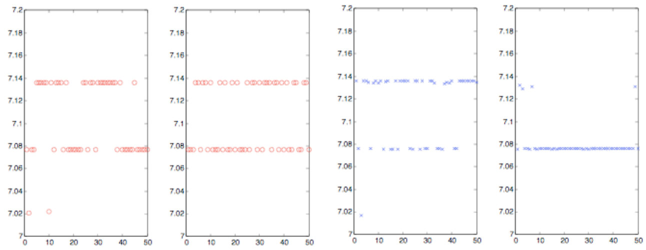

Figure 2: $\|\widehat{\mathbf{T}}^n\|_F$ versus 50 Monte Carlo Simulation Runs via HOOI (red, first two left) and SDP (blue, last two right). The stopping tolerance: $\varepsilon_{tol} = 10^{-5}$ (left,red) to $\varepsilon_{tol} = 10^{-9}$ (right,red) and $\varepsilon_{tol} = 10^{-4}$ (left,blue) to $\varepsilon_{tol} = 10^{-5}$ (right,blue).

In Figures 2-3, we implement the HOOI and SDP algorithms on the some particular tensors: tensors $\mathscr{T}$ of dimension $4 \times 7 \times 7$ and $7 \times 7 \times 7$ without a dominant low multilinear rank. We then find a low multilinear rank tensors $\widehat{\mathscr{T}}$ of rank-$(2,2,2)$ for the first tensor in Figure 2 and rank-$(3,3,3)$ in Figure 3. Now in the few test cases including the data in [13] where the original tensor has a dominant low multilinear rank, the approximation of both HOOI and SDP algorithms coincide. The plots in Figure 2 show that there are two local extrema, $\|\widehat{\mathbf{T}}^n\|_F = 7.1359$ and $\|\widehat{\mathbf{T}}^n\|_F = 7.0756$, generated by the HOOI algorithm with random initial factors. Several other extrema are visible on the plots which are artifacts caused by low stopping criteria settings. The stopping criteria is set through $\|\mathbf{A}^i - \mathbf{A}^{i+1}\|_F < \varepsilon_{tol}$ and a maximum iteration. As we decrease the stopping criteria from $\varepsilon_{tol} = 10^{-5}$ to $\varepsilon_{tol} = 10^{-9}$ and increase the maximum iteration tolerance, the artifacts diminish. The plots in Figure 3 show that there are two local extrema, $\|\widehat{\mathbf{T}}^n\|_F = 7.1359$ and $\|\widehat{\mathbf{T}}^n\|_F = 7.0756$, generated by the SDP algorithm with random initial factors. As we decrease the stopping criteria from $\varepsilon_{tol} = 10^{-4}$ to $\varepsilon_{tol} = 10^{-5}$ and increase the maximum iteration tolerance, we see that the SDP converges to the minimum of the two extrema, $\|\widehat{\mathbf{T}}^n\|_F = 7.0756$ ($\|\widehat{\mathbf{T}}^1\|_{tr} = 8.41$, $\|\widehat{\mathbf{T}}^2\|_{tr} = 8.41$, $\|\widehat{\mathbf{T}}^3\|_{tr} = 8.36$) which is the approximated lowest multilinear rank tensor. On the other hand, when both algorithms are started with the truncated HO-SVD initial factors, $\|\widehat{\mathbf{T}}^n\|_F = 7.1359$, was the only approximation for both algorithms.

In Figure 3, we also tested the both algorithms with tensor of size $7 \times 7 \times 7$ which has no dominant low multilinear rank structure. The plot shows in most cases that the SDP algorithm finds the extreme cost value of 9.1074 while the HOOI algorithm also converges to three values: 9.0815, 9.0790, 9.107 for approximation a low multilinear rank tensor of $(3,3,3)$.

In Figure 4, we ran a noisy test case with on data [13] where $\widetilde{\mathscr{T}} = \frac{\mathscr{T}}{\|\mathscr{T}\|_F} + \sigma \frac{\mathscr{N}}{\|\mathscr{N}\|_F}$ where $\mathscr{N}$ is a noise tensor and σ is the noise level. The noisy tensor $\widetilde{\mathscr{T}}$ is a tensor with

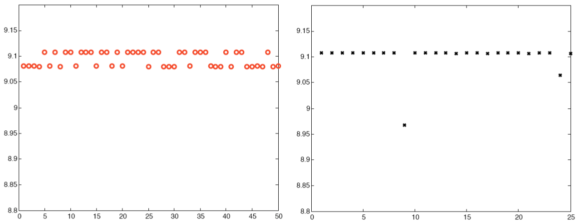

Figure 3: $\|\widehat{\mathbf{T}}^n\|_F$ versus Monte Carlo Simulation Runs via HOOI (red, left) and SDP (blue, right). The stopping tolerance: $\varepsilon_{tol} = 10^{-9}$(left) and $\varepsilon_{tol} = 10^{-4}$ (right) for nondominant low multilinear rank structure.

SNR Table

σ	$\|(\mathbf{T}^1)^{i+1} - (\mathbf{T}^1)^i\|_F$	$\|\mathbf{T}^1 - \widehat{\mathbf{T}}^1\|_F$
10^{-3}	1.0883×10^{-6}	0.0792
10^{-2}	6.604×10^{-6}	0.0863
10^{-1}	3.210×10^{-4}	0.3531

Figure 4: 30 Monte Carlo Simulation Runs with dominant low rank $(2,2,2)$

zero-mean Gaussian entries.

5. CONCLUSION

In this paper, we have presented a new method for computing low multilinear rank factors for higher-order tensors. Through the tensor trace class norm, we formulate a rank minimization problem for each mode. Thus, a set of semidefinite programming subproblems are solved. In general, this requires a high number of iterations. The results reported in this paper are only preliminary. In particular, we should examine whether the method always converges. Also the issue of local optima deserves further attention. We should determine the convergence rate, analyze noisy test cases and find efficient SDP algorithm for tensors.

Acknowledgments

L. De Lathauwer was supported in part by Research Council KUL: GOA-AMBioRICS, CoE EF/05/006 Optimization in Engineering, Campus Impuls Financiering (CIF1), STRT1/08/023; FWO: (a) project FWO-G.0321.06, (b) Research Communities ICCoS, ANMMM and MLDM; Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, "Dynamical systems, control and optimization", 2007–2011); EU: ERNSI.

REFERENCES

[1] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[2] E. Candès and B. Recht, "Exact Matrix Completion Via Convex Optimization," Preprint.

[3] E. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, **52** pp. 489-509, 2005.

[4] L. De Lathauwer, *Signal processing based on multilinear algebra.* Ph.D. Thesis, K.U. Leuven, Belgium, 1997.

[5] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-$(R_1;R_2;\ldots;R_N)$ approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications* **21**, pp. 1324-1342, 2000.

[6] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, **21**, pp. 1253-1278, 2000.

[7] L. Eldén and B. Savas, "A Newton–Grassmann method for computing the Best Multi-Linear Rank-(r_1,r_2,r_3) Approximation of a Tensor," *SIAM Journal on Matrix Analysis and Applications*, **31** 2, pp.248-271, 2009.

[8] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming*, `http://www.stanford.edu/~boyd/cvx/`.

[9] M. Ishteva, L. De Lathauwer, P.-A. Absil, S. Van Huffel, "Differential-geometric Newton method for the best rank-(R1,R2,R3) approximation of tensors," Approximation of Tensors", *Numerical Algorithms*, **51** 2, pp. 179–194, 2009.

[10] M. Ishteva, L. De Lathauwer, P.-A. Absil, S. Van Huffel, "Dimensionality reduction for higher-order tensors: algorithms and applications," *International Journal of Pure and Applied Mathematics*, **42** 3, pp. 337-343, 2008.

[11] E. Kreyszig, *Introductory Functional Analysis with Applications.* Wiley & Sons, New York, 1989.

[12] P. M. Kroonenberg and J. De Leeuw, "Principal component analysis of three-mode data by means of alternating least squares algorithms," *Psychometrika*, **45**, pp. 69-97, 1980.

[13] P. Ladefoged, R. Harshman, L. Goldstein, L. Rice, "Generating Vocal Tract Shapes from Formant Frequencies," *J. Acoust. Soc. Am.*, **64** 4, pp. 1027-1035, 1978.

[14] C. Navasca and L. De Lathauwer, "A new tensor norm for the best multilinear low rank approximation," Preprint.

[15] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed Minimum Rank Solutions to Linear Matrix Equations via Nuclear Norm Minimization," Preprint.

[16] B. Savas and L.-H. Lim, "Best multilinear rank approximation of tensors with quasi-Newton methods on Grassmannians," Preprint.

[17] B. Simon, *Trace Ideals and their Applications*. Cambridge University Press, 1979.

[18] J. Sturm. *SeDuMi*, `http://sedumi.ie.lehigh.edu/`.

[19] L. R. Tucker. "Implications of factor analysis of three-way matrices for measurement of change," in *Problems in Measuring Change*, C. W. Harris, ed., University of Wisconsin Press, pp. 122-137, 1963.

[20] L. R. Tucker, "The extension of factor analysis to three-dimensional matrices," in *Contributions to Mathematical Psychology*, H. Gulliksen and N. Frederiksen, eds., Holt, Rinehardt, & Winston, New York, 1963.

[21] L. R. Tucker. "Some mathematical notes on three-mode factor analysis," *Psychometrika*, **31**, pp. 279-311, 1966.

[22] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, **38**, pp. 49-95, 1996.