SIAM J. OPTIM.
Vol. 26, No. 1, pp. 171–196
© 2016 Society for Industrial and Applied Mathematics

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

RANK-1 TENSOR PROPERTIES WITH APPLICATIONS TO A CLASS OF TENSOR OPTIMIZATION PROBLEMS*

YUNING YANG†, YUNLONG FENG†, XIAOLIN HUANG†, AND JOHAN A. K. SUYKENS†

Abstract. This paper studies models and algorithms for a class of tensor optimization problems, based on a rank-1 equivalence property between a tensor and certain unfoldings. It is first shown that in dth order tensor space, the set of rank-1 tensors is the same as the intersection of $\lceil \log_2(d) \rceil$ tensor sets, of which tensors have a specific rank-1 balanced unfolding matrix. Moreover, the number $\lceil \log_2(d) \rceil$ is proved to be optimal in some sense. Based on the above equivalence property, three relaxation approaches for solving the best rank-1 tensor approximation problems are proposed, including two convex relaxations and a nonconvex one. The two convex relaxations utilize the matrix nuclear norm regularization/constraints. They have the advantage of identifying whether the solution is a global optimizer of the original problem, by computing the nuclear norm or the Frobenius norm of a certain matrix. Under certain assumptions, the optimal solution of the original problem is characterized by the solution to the dual of the nuclear norm constrained problem. The nonconvex relaxation can be solved via the conventional alternating minimization scheme, with the output being always a rank-1 tensor. Numerical experiments demonstrate the effectiveness of the proposed methods.

Key words. higher order tensor, balanced unfolding, rank-1 approximation, rank-1 equivalence property, convex relaxation, nuclear norm

AMS subject classifications. 90C26, 15A18, 15A69, 41A50

DOI. 10.1137/140983689

1. Introduction.

Tensors are higher order generalizations of vectors and matrices. The structure of tensors makes it possible to explore the information behind data that have intrinsically many dimensions from a higher order perspective. In recent years, a variety of real-world problems have been modeled as optimization problems where the variable is a low rank or even rank-1 tensor. For example, in video completion, one needs to recover a video from its partial observations, which can be modeled as a tensor completion problem [35, 21, 44]; in [32], the web links are represented as a third order tensor, where web link analysis is implemented by solving a tensor decomposition problem; in higher order graph matching, higher order affinities are represented by tensors, while finding the matching has been modeled as a rank-1 tensor approximation problem [13]; in nonlinear elastic materials analysis, determining whether the strong or ordinary ellipticity condition holds or not amounts to solving an optimization problem where the variable is again a rank-1 tensor [41, 47].

Mathematically, applications mentioned above involve the following optimization problem (or subproblem)

$$\max \langle \mathcal{A}, x_1 \circ \cdots \circ x_d \rangle \quad \text{s.t. } x_i \in \mathbb{R}^{N_i},\ \|x_i\| = 1,\ \ i = 1, \ldots, d. \tag{1.1}$$

That is, one wants to find a rank-1 tensor $\mathcal{X} := x_1 \circ \cdots \circ x_d$ to maximize the inner product between the dth order tensors $\mathcal{A}$ and $\mathcal{X}$, where $\circ$ denotes outer product.

*Received by the editors August 25, 2014; accepted for publication (in revised form) November 16, 2015; published electronically January 20, 2016. This work was supported by ERC AdG A-DATADRIVE-B (290923), GOA/10/09 MaNet, CoE PFV/10/002 (OPTEC), BIL12/11T, FWO G.0377.12, G.088114N, iMinds SBO 2014; IUAP P7/19 (DYSCO).
http://www.siam.org/journals/siopt/26-1/98368.html
†Department of Electrical Engineering, ESAT-STADIUS, KU Leuven, Kasteelpark Arenberg 10, Leuven, B-3001, Belgium (yuning.yang@esat.kuleuven.be, yunlong.feng@esat.kuleuven.be, xiaolin.huang@esat.kuleuven.be, johan.suykens@esat.kuleuven.be).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The maximum value of (1.1) is defined as the spectral norm of $\mathcal{A}$ [34], and the KKT system of (1.1) is known as the tensor singular value/eigenvalue problem [34, 40]. On the other hand, (1.1) is closely related to the best rank-1 approximation problem [17], i.e., to find a projection of $\mathcal{A}$ onto the set of rank-1 tensors. Theoretically, (1.1) and its related tensor singular value/eigenvalue problem has been studied using tools from algebraic geometry; see, e.g., [40, 10, 19]. However, it is proved in [28] that solving (1.1) is NP-hard; more precisely, determining the spectral norm is NP-hard when $d > 2$, and for any fixed nonzero rational number σ, deciding whether σ is the spectral norm of $\mathcal{A}$ is also NP-hard. Therefore, several algorithms were proposed to find local or approximation solutions to (1.1). To find a local maximum or a KKT point of (1.1), the higher order power method (and its symmetric version) was proposed [17, 30, 33]; a Newton-type method was introduced in [51] for (1.1) when $d = 3$; Chen et al. [11] developed a maximum block improvement approach that can find a KKT point. Another category of algorithms is focused on finding an approximation solution to (1.1) with provable lower bound in polynomial time; see, e.g., [36, 27, 45, 52, 26, 53], just to name a few.

Recently, Jiang, Ma, and Zhang [29] proposed convex relaxations for solving the following symmetric version of (1.1),

$$\max \langle \mathcal{A}, x \circ \cdots \circ x \rangle \quad \text{s.t. } x \in \mathbb{R}^N, \ \|x\| = 1, \tag{1.2}$$

where $\mathcal{A}$ is a symmetric tensor, i.e., every entry $\mathcal{A}_{i_1 \cdots i_d}$ is invariant under any permutation of $\{i_1, \ldots, i_d\}$. In particular, [29] proved an equivalence property between a rank-1 symmetric tensor and its balanced-unfolding matrix (the concept of balanced unfolding will be introduced later), based on which the above tensor optimization problem can be casted into a matrix optimization problem. Specifically, it is proved that the set of even order symmetric rank-1 tensors is equivalent to the set of even order symmetric tensors whose balanced-unfolding matrix is rank-1. Results reported in [29] are promising in that in most cases, the solution to the convex relaxation is found to be also an optimal solution to (1.2) numerically. Extensions of reformulating third order and even order cases of (1.1) into (1.2) (or partially symmetric forms) have been discussed in [29, section 8].

In this work, motivated by [29], without the symmetric property and the even order assumptions, a rank-1 equivalence property between a dth order tensor and $\lceil \log_2(d) \rceil$ unfolding matrices is studied. More discussions on this property will be detailed later. The equivalence property provides a simple way to determine whether a tensor is rank-1 by only examining the ranks of $\lceil \log_2(d) \rceil$ unfolding matrices. Moreover, thanks to this equivalence property, it is possible to cast (1.1) into a matrix optimization problem with $\lceil \log_2(d) \rceil$ rank-1 matrix constraints. To solve the resulting matrix optimization problem, three relaxations are introduced, two of which are convex relaxations, based on matrix nuclear norm regularization/constraints, while the other one is nonconvex. By using the equivalence property, it is proved that the convex relaxations can identify whether the solution is a global optimizer of the original problem (1.1), by only examining the magnitude of the nuclear norm or the Frobenius norm of the solution. The nuclear norm constrained problem is then studied from the dual. In particular, if a critical point of the dual satisfies certain assumptions, then it is shown to relate to an optimal solution of (1.1). The nonconvex relaxation is tailored for third or fourth order tensor cases. The relaxation can be solved via the simple alternating minimization scheme, with the output being always a rank-1 tensor. The nonconvex model can be extended to rank-R tensor approximation problems with $R > 1$ as well.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

To be more specific, the rank-1 equivalence property consists of the following two results. First, it is shown that the set of dth order rank-1 tensors is the same as the intersection of $\lceil \log_2(d) \rceil$ tensor sets, of which tensors have a specific rank-1 balanced-unfolding matrix. In other words, if $\lceil \log_2(d) \rceil$ specific balanced unfoldings of tensor $\mathcal{X}$ are known to be rank-1, then $\mathcal{X}$ is also rank-1. How to choose these balanced unfoldings is important to the equivalence property; to achieve this, a procedure is provided. Then, it is proved that the number $\lceil \log_2(d) \rceil$ is optimal in the sense specified below. For any given $\lceil \log_2(d) \rceil - 1$ tensor sets, of which tensors have a rank-1 unfolding matrix, the intersection of these $\lceil \log_2(d) \rceil - 1$ sets properly includes the set of rank-1 tensors. That is to say, if $\lceil \log_2(d) \rceil - 1$ unfolding matrices of $\mathcal{X}$ are known to be rank-1, then it is not sufficient to deduce that $\mathcal{X}$ is rank-1.

In fact, the first part of the equivalence property introduced above can be deduced from [8, Theorem 8], which characterizes the upper bound of the rank of a certain unfolding matrix via the product of ranks of some unfoldings (see Remark 3.2 for details). Nonetheless, the equivalence property introduced in this work differs from that of [8] in the following ways. First, the ideas behind the equivalence property, e.g., how to choose $\lceil \log_2(d) \rceil$ balanced unfoldings, are proposed in this paper; second, although the rank-1 equivalence property can be deduced from [8], for better understanding the principle behind the equivalence property, an alternative proof is in turn provided; third, the rank-1 equivalence property can be more practical, as introduced above.

This paper is organized as follows. Section 2 introduces basic multilinear algebra. Properties and results of the rank-1 equivalence property are studied in section 3. Relaxations based on the equivalence property for tensor approximation problems are proposed and analyzed in section 4. Numerical experiments are provided in section 5. This paper is ended with conclusions in section 6.

2. Notations and multilinear algebra. $\mathbb{R}$ denotes the real field. Vectors are written as lowercase letters $(x, y, \ldots)$, matrices correspond to italic capitals $(A, B, \ldots)$, and tensors are written as calligraphic capitals $(\mathcal{A}, \mathcal{B}, \ldots)$.

Tensors, tensor inner product, and Frobenius norm. A tensor is a multiway array. A dth order tensor $\mathcal{A} \in \mathbb{R}^{N_1 \times \cdots \times N_d}$ is defined as $\mathcal{A} = (A_{i_1 \cdots i_d})$, $1 \le i_j \le N_j$, $j = 1, \ldots, d$. For two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{N_1 \times \cdots \times N_d}$, their inner product is given by $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{N_1} \cdots \sum_{i_d=1}^{N_d} \mathcal{A}_{i_1 \cdots i_d} \mathcal{B}_{i_1 \cdots i_d}$. The Frobenius norm of $\mathcal{A}$ is defined by $\|\mathcal{A}\|_F = \langle \mathcal{A}, \mathcal{A} \rangle^{1/2}$.

Kronecker product and outer product. The Kronecker product $A \otimes B$ of matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$ is the $mp \times nq$ block matrix given by

$$A \otimes B := \begin{bmatrix} a_{11}B & \cdots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{m,1}B & \cdots & a_{m,n}B \end{bmatrix} = [a_1 b_1 \ a_1 b_2 \ \cdots \ a_1 b_q \ a_2 b_1 \ \cdots \ a_n b_q].$$

The outer product $a \circ b$ of vectors a and b is the rank-1 matrix given by $a \circ b := ab^T$. Similarly, the outer product $\mathcal{A} := a_1 \circ \cdots \circ a_d$ of d vectors $a_i \in \mathbb{R}^{N_i}$, $i = 1, \ldots, d$, is a rank-1 tensor $\mathcal{A}$, whose entries are the product of the corresponding vector entries $\mathcal{A}_{i_1 \cdots i_d} = a_{1,i_1} \cdots a_{d,i_d}$ for all $1 \le i_j \le N_j$.

Tensor-tensor unfolding and tensor-matrix balanced unfolding. Let the dth order tensor be $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_d}$, and denote μ_k, $1 \le k \le m$ as m $(m \le d)$ mode sets, satisfying $\cap_{k=1}^{m} \mu_k = \emptyset$ and $\cup_{k=1}^{m} \mu_k = \{1, \ldots, d\}$. The tensor-tensor unfolding is to map $\mathcal{X}$ to an mth order tensor $\mathcal{X}_{[\mu_1;\ldots;\mu_m]}$ in the space $\mathbb{R}^{\prod_{i_1 \in \mu_1} N_{i_1} \times \cdots \times \prod_{i_m \in \mu_m} N_{i_m}}$, where the jth mode of $\mathcal{X}_{[\mu_1;\ldots;\mu_m]}$ is constructed by merg-

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

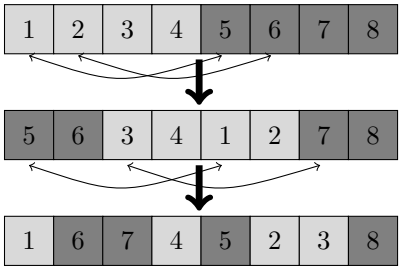

FIG. 1. *An illustration of choosing the balanced unfoldings. Nodes in light and dark gray represent the modes corresponding to the row and column of the first matrix; curves with double arrows represent the exchanging procedure.*

ing the modes in μ_j together. Here a semicolon indicates a new mode. When $m = 2$, it reduces to tensor-matrix unfolding.

The tensor-matrix balanced unfolding is to unfold $\mathcal{X}$ to a matrix $X_{[\mu;\nu]}$ with $\text{card}(\mu) = \lfloor \frac{d}{2} \rfloor$ and $\text{card}(\nu) = d - \lfloor \frac{d}{2} \rfloor$, where $\text{card}(\cdot)$ denotes the cardinality of a set. Consider a 4th order tensor $\mathcal{X}$; $X_{[1,2;3,4]}$ is given by $(X_{[1,2;3,4]})_{(i_1-1)N_2+i_2,(i_3-1)N_4+i_4} = \mathcal{X}_{i_1,i_2,i_3,i_4}$, for all $1 \leq i_j \leq N_j$, $j = 1, 2, 3, 4$. The idea of the balanced-unfolding technique has been adopted in the literature, e.g., [22, 29, 38].

Tensor CP-rank. The CP-rank of a tensor $\mathcal{X}$, denoted by $\text{rank}_{\text{CP}}(\mathcal{X})$, is defined as the smallest number R such that $\mathcal{X}$ can be factorized as a sum of R rank-1 tensors. In this paper, tensor rank is referred to CP-rank.

3. The rank-1 equivalence property. In this section, it is first shown that the set of dth order rank-1 tensors is equivalent to the intersection of $\lceil \log_2(d) \rceil$ tensor sets with tensors having a specific rank-1 balanced-unfolding matrix. Moreover, it is proved that the number $\lceil \log_2(d) \rceil$ is optimal: for any given $\lceil \log_2(d) \rceil - 1$ tensor sets with tensors having a certain rank-1 unfolding matrix, the intersection of these $\lceil \log_2(d) \rceil - 1$ sets properly includes the set of rank-1 tensors.

3.1. Choosing the balanced-unfolding matrices. To clarify the analysis, in this part, we restrict ourselves to tensors of order $d = 2^n$, $n \geq 2$, and consider the general cases in what follows. For notation simplification, we may specify a matrix or a tensor by its subscript, e.g., $X_{[1,2;3,4]}$ is represented by $[1, 2; 3, 4]$.

In 4th order tensor cases, the two matrices $[1, 2; 3, 4]$ and $[3, 2; 1, 4]$ are chosen. The proof of Lemma 3.1 explains why the above two matrices are selected. In 8th order tensor cases, the following 3 specific balanced unfoldings are chosen. The first is $[1, 2, 3, 4; 5, 6, 7, 8]$; then the second one $[5, 6, 3, 4; 1, 2, 7, 8]$ is chosen by exchanging $(1, 2)$ and $(5, 6)$; finally, $[1, 6, 7, 4; 5, 2, 3, 8]$ is chosen by carrying out $(5) \leftrightarrow (1)$ and $(3) \leftrightarrow (7)$. The procedure of choosing the three matrices is visualized in Figure 1. The principle behind this procedure is that, $[1, 2, 3, 4; 5, 6, 7, 8]$ and $[5, 6, 3, 4; 1, 2, 7, 8]$ being rank-1 implies the rank-1 property of tensor $[1, 2; 3, 4; 5, 6; 7, 8]$; $[1, 2; 3, 4; 5, 6; 7, 8]$ being rank-1 together with the last matrix $[1, 6, 7, 4; 5, 2, 3, 8]$ being rank-1 lead to the rank-1 property of $[1; 2; 3; 4; 5; 6; 7; 8]$. This will be shown in Lemma 3.2 and Theorem 3.3.

For a 2^nth order tensor, the pseudocode of the procedure for choosing balanced unfoldings is provided in Algorithm 1. The principle of the procedure is based on Lemmas 3.1 and 3.2 and Theorem 3.3, which discusses the equivalence between a 2^nth order rank-1 tensor and the generated rank-1 balanced-unfolding matrices.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ALGORITHM 1. PROCEDURE TO CHOOSE n UNFOLDING MATRICES FOR A 2^nTH ORDER TENSOR.

Input a tuple $\{1,\dots,2^n\}$. Denote $\mu_1 := \{1,\dots,2^{n-1}\}$ and $\mu_2 := \{2^{n-1}+1,\dots,2^n\}$.
Denote $[\overline{\mu}_1;\overline{\nu}_1] := [\mu_1;\mu_2]$.
for $i = 1 : n-1$ **do** % each i refers to a tuple to be generated
- Let $\mu_k = \mu_k^l \cup \mu_k^r$ with $\mu_k^l \cap \mu_k^r = \emptyset$ and $\text{card}(\mu_k^l) = \text{card}(\mu_k^r) = \text{card}(\mu_k)/2$, $1 \le k \le 2^i$.
- Denote $[\overline{\mu}_{i+1};\overline{\nu}_{i+1}] := [\mu_{2^{i-1}+1}^l, \mu_1^r, \mu_{2^{i-1}+2}^l, \mu_2^r, \dots, \mu_{2^i}^l, \mu_{2^{i-1}}^r; \mu_1^l, \mu_{2^{i-1}+1}^r, \dots, \mu_{2^{i-1}}^l, \mu_{2^i}^r]$.
- Rewrite notations: $\mu_{2^{i-1}}^l \leftarrow \mu_1,\ \mu_1^r \leftarrow \mu_2, \dots, \mu_{2^i}^l \leftarrow \mu_{2^i-1},\ \mu_{2^{i-1}}^r \leftarrow \mu_{2^i}, \dots, \mu_{2^{i-1}}^l \leftarrow \mu_{2^{i+1}-1},\ \mu_{2^i}^r \leftarrow \mu_{2^{i+1}}$.

end for
Output $[\overline{\mu}_1;\overline{\nu}_1],\ [\overline{\mu}_2;\overline{\nu}_2], \dots, [\overline{\mu}_n;\overline{\nu}_n]$.

3.2. The equivalence property. To introduce the equivalence property between a rank-1 tensor and its rank-1 balanced unfoldings, some notations and lemmas are introduced below. Denote

$$\boldsymbol{\mathcal{X}}_d := \{\mathcal{X} \in \mathbb{R}^{N_1 \times \cdots \times N_d} \mid \text{rank}_{\text{CP}}(\mathcal{X}) = 1\}$$

as the set of dth order rank-1 tensors, and

$$\boldsymbol{\mathcal{X}}_{d,[\mu;\nu]} := \{\mathcal{X} \in \mathbb{R}^{N_1 \times \cdots \times N_d} \mid \text{rank}(X_{[\mu;\nu]}) = 1\}$$

as the set of tensors whose unfolding matrix $X_{[\mu;\nu]}$ is rank-1, where μ and ν are certain mode sets satisfying $\mu \cap \nu = \emptyset$, $\mu \cup \nu = \{1,\dots,d\}$.

LEMMA 3.1 (the equivalence property for 4th order tensors). *There holds*

$$\boldsymbol{\mathcal{X}}_4 = \boldsymbol{\mathcal{X}}_{4,[1,2;3,4]} \cap \boldsymbol{\mathcal{X}}_{4,[3,2;1,4]}. \tag{3.1}$$

In other words, if $X_{[1,2;3,4]}$ and $X_{[3,2;1,4]}$ of $\mathcal{X}$ are rank-1, then $\mathcal{X}$ is also rank-1.

Remark 3.1. The proof of Lemma 3.1 is based on matrix SVD, which is similar to that of [8, Theorem 8]; nevertheless, to see more clearly why $X_{[1,2;3,4]}$ and $X_{[3,2;1,4]}$ are chosen, a complete proof is still provided here.

Proof of Lemma 3.1. $\boldsymbol{\mathcal{X}}_4 \subseteq \boldsymbol{\mathcal{X}}_{4,[1,2;3,4]} \cap \boldsymbol{\mathcal{X}}_{4,[3,2;1,4]}$ is clear. Suppose $\mathcal{X} \in \boldsymbol{\mathcal{X}}_{4,[1,2;3,4]} \cap \boldsymbol{\mathcal{X}}_{4,[3,2;1,4]}$. Then $X_{[1,2;3,4]}$ may be expressed as the outer product of two vectors, i.e.,

$$X_{[1,2;3,4]} = x_{[1,2]} \circ x_{[3,4]}, \quad \text{where } x_{[1,2]} \in \mathbb{R}^{N_1 N_2}, x_{[3,4]} \in \mathbb{R}^{N_3 N_4}.$$

Let $X_{[1;2]}$ and $X_{[3;4]}$ be the folded matrices of $x_{[1,2]}$ and $x_{[3,4]}$, respectively. Let the SVDs of the two matrices be given by

$$X_{[1;2]} = \sum_{s=1}^{R_1} a_s \circ b_s \ \text{ and } \ X_{[3;4]} = \sum_{t=1}^{R_2} c_t \circ d_t, \tag{3.2}$$

where $R_1 = \text{rank}(X_{[1;2]})$ and $R_2 = \text{rank}(X_{[3;4]})$, and the singular values are absorbed in the corresponding singular vectors. Then

$$X_{[1,2;3,4]} = \left(\sum_{s=1}^{R_1} a_s \otimes b_s\right) \circ \left(\sum_{t=1}^{R_2} c_t \otimes d_t\right) = \sum_{s=1}^{R_1}\sum_{t=1}^{R_2} (a_s \otimes b_s) \circ (c_t \otimes d_t).$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

As a result, $\mathcal{X}$ can be decomposed as $\mathcal{X} = \sum_{s=1}^{R_1} \sum_{t=1}^{R_2} a_s \circ b_s \circ c_t \circ d_t$. Therefore, the decomposition of $X_{[3,2;1,4]}$ is given by

$$X_{[3,2;1,4]} = \sum_{s=1}^{R_1} \sum_{t=1}^{R_2} (c_t \otimes b_s) \circ (a_s \otimes d_t)\,.$$

Since $c_t \otimes b_s, s = 1, \ldots, R_1, t = 1, \ldots, R_2$, are linearly independent, and $a_s \otimes d_t, s = 1, \ldots, R_1, t = 1, \ldots, R_2$, are also linearly independent, $X_{[3,2;1,4]}$ is rank-1 if and only if $R_1 = R_2 = 1$. This implies that (3.1) holds. ∎

To extend Lemma 3.1 to higher order cases, we need the following lemma, where u_i^l, μ_i^r are some disjoint mode sets, $i = 1, \ldots, 2^n$, $n \geq 1$, with $\text{card}(\mu_i) \geq 1$.

LEMMA 3.2. *Let the tensor* $[\mu_1^l, \mu_1^r; \mu_2^l, \mu_2^r; \ldots; \mu_{2^n}^l, \mu_{2^n}^r]$ *be rank-*1. *Furthermore, assume that its unfolding matrix*

$$[\mu_{2^{n-1}+1}^l,\ \mu_1^r,\ \mu_{2^{n-1}+2}^l,\ \mu_2^r, \ldots, \mu_{2^n}^l,\ \mu_{2^{n-1}}^r;\ \mu_1^l,\ \mu_{2^{n-1}+1}^r, \ldots, \mu_{2^{n-1}}^l,\ \mu_{2^n}^r]$$

*is rank-*1. *Then the tensor* $[\mu_1^l; \mu_1^r; \mu_2^l; \mu_2^r; \ldots; \mu_{2^n}^r]$ *is also rank-*1.

Proof. This can also be proved based on SVD, following the argument of Lemma 3.1. ∎

The following theorem establishes the equivalence property for 2^nth order tensors.

THEOREM 3.3 (the equivalence property for 2^nth order tensors). *There holds*

$$\boldsymbol{\mathcal{X}}_{2^n} = \bigcap_{i=1}^{n} \boldsymbol{\mathcal{X}}_{2^n,[\overline{\mu}_i;\overline{\nu}_i]}, \tag{3.3}$$

where $[\overline{\mu}_1; \overline{\nu}_1], \ldots, [\overline{\mu}_n; \overline{\nu}_n]$ *are generated by Algorithm* 1. *In other words, if the matrices* $X_{[\overline{\mu}_1;\overline{\nu}_1]}, \ldots, X_{[\overline{\mu}_n;\overline{\nu}_n]}$ *are known rank-*1, *then* $\mathcal{X}$ *is also rank-*1.

Proof. Suppose $\mathcal{X} \in \cap_{i=1}^{n} \boldsymbol{\mathcal{X}}_{2^n,[\overline{\mu}_i;\overline{\nu}_i]}$. Consider the induction method on n. Lemma 3.1 gives the result when $n = 2$. Suppose (3.3) holds when $n = k$. For $n = k + 1$, let

$$\mu_1 = \{1, 2\},\ \mu_2 = \{3, 4\}, \ldots,\ \mu_{2^k} = \{2^{k+1} - 1, 2^{k+1}\}.$$

By induction, the unfolding tensor $\mathcal{X}_{[\mu_1;\mu_2;\cdots;\mu_{2^k}]}$ is rank-1 in the 2^kth order tensor space $\mathbb{R}^{N_1 N_2 \times N_3 N_4 \times \cdots \times N_{2^{k+1}-1} N_{2^{k+1}}}$. Let

$$\mu_i = \mu_i^l \cup \mu_i^r,\ i = 1, \ldots, 2^k,\ \text{with } \text{card}(\mu_i^l) = \text{card}(\mu_i^r) = 1.$$

Then the rank-1 property of the last unfolding matrix $[\overline{\mu}_{k+1}; \overline{\nu}_{k+1}]$ together with Lemma 3.2 gives the rank-1 property of $\mathcal{X}_{[1;2;\cdots;2^{k+1}]}$ in the 2^{k+1}th order tensor space $\mathbb{R}^{N_1 \times \cdots \times N_{2^{k+1}}}$. ∎

Denote $f(d) := \lceil \log_2(d) \rceil$ in the following. To generalize Algorithm 1 as well as Theorem 3.3 to any dth order tensor, we introduce the following virtual mode, which can be used to lift a dth order tensor to a $2^{f(d)}$th order tensor.

DEFINITION 3.4 (virtual mode). *For a* d*th order tensor* $\mathcal{X} \in \mathbb{R}^{N_1 \times \cdots \times N_d}$, *we call a mode* $k \in \{1, \ldots, d\}$ *virtual if* $N_k = 1$.

In general, a dth order tensor can always be lifted to a $2^{f(d)}$th order tensor by adding $2^{f(d)} - d$ virtual modes. A balanced lifting way is illustrated in section 3.4.

In the following, the equivalence property for dth order tensors is provided.

COROLLARY 3.5 (the equivalence property for dth order tensors). *There holds*

$$\boldsymbol{\mathcal{X}}_d = \bigcap_{i=1}^{f(d)} \boldsymbol{\mathcal{X}}_{d,[\overline{\mu}_i;\overline{\nu}_i]},$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where the matrices $[\overline{\mu}_1;\overline{\nu}_1],\ldots,[\overline{\mu}_{f(d)};\overline{\nu}_{f(d)}]$ *are generated by first applying Algorithm* 1 *to the lifted* $2^{f(d)}$*th order tensor, and then eliminating the virtual modes.*

In the introduction, we have mentioned that Theorem 3.3 can be deduced from [8, Theorem 8]. We recall the theorem in the following. For a set μ, denote $\mu^C := \{1,\ldots,d\}\setminus\mu$ the complement of μ.

THEOREM 3.6 (cf. Theorem 8 of [8]). *For any* $\mu_1,\ldots,\mu_m \subseteq \{1,\ldots,d\}$*, let* $\nu = \cap_{i=1}^m \mu_i$ *or* $\nu = \cup_{i=1}^m \mu_i$*. Then for a dth order tensor* $\mathcal{X}$

$$\operatorname{rank}(X_{[\nu;\nu^C]}) \le \prod_{i=1}^{m} \operatorname{rank}(X_{[\mu_i;\mu_i^C]}). \tag{3.4}$$

Remark 3.2. To use the above theorem to prove Theorem 3.3, it suffices to prove that

$$\operatorname{rank}(X_{(i)}) \le \prod_{i=1}^{n} \operatorname{rank}(X_{[\overline{\mu}_i;\overline{\nu}_i]}) \tag{3.5}$$

for $i = 1,\ldots,2^n$, which in turn implies that $\operatorname{rank}_{\mathrm{CP}}(\mathcal{X}) = 1$, where $X_{(i)}$ is the mode-i unfolding of $\mathcal{X}$. Nevertheless, to verify (3.5), it is also necessary to apply the induction method on n, as that of Theorem 3.3. On the other hand, the analysis of Theorem 3.3 might be helpful to understand the principle behind the equivalence property. Therefore, we provide an alternative proof for Theorem 3.3, instead of applying [8, Theorem 8].

Remark 3.3. The rank-1 equivalence property of [29] can be deduced from Corollary 3.5, because for an even order symmetric tensor, all the balanced unfoldings are the same. Therefore $\boldsymbol{\mathcal{X}}_d = \boldsymbol{\mathcal{X}}_{d,[\mu;\nu]}$, where $\operatorname{card}(\mu) = \operatorname{card}(\nu) = d/2$.

3.3. The number $\lceil \log_2(d) \rceil$ is optimal. In this subsection, it is shown that the number $f(d) = \lceil \log_2(d) \rceil$ is optimal.

THEOREM 3.7 ($\lceil \log_2(d) \rceil$ is optimal). *Let tensors be of order* d*,* $d \ge 3$*, with* $N_i > 1$*,* $1 \le i \le d$*. Denote* μ_i, ν_i*,* $i = 1,\ldots,f(d)-1$*, as some mode sets, with*

$$\mu_i \cup \nu_i = \{1,\ldots,d\},\ \ \mu_i \cap \nu_i = \emptyset; \tag{3.6}$$

then for any μ_i, ν_i *satisfying* (3.6)*, there holds*

$$\boldsymbol{\mathcal{X}}_d \subsetneq \bigcap_{i=1}^{f(d)-1} \boldsymbol{\mathcal{X}}_{d,[\mu_i;\nu_i]}. \tag{3.7}$$

The above theorem indicates that the set of dth order rank-1 tensors is properly included in the intersection of less than $f(d)$ sets of tensors whose certain unfolding matrix is rank-1. In other words, there exists at least a tensor that, even it has $f(d)-1$ unfoldings which are known to be rank-1, the tensor itself may not be rank-1. To prove the above theorem, we need the following technical lemma.

LEMMA 3.8. *Let tensors be of order* d*,* $d \ge 3$*, with* $N_i > 1$*,* $1 \le i \le d$*. Let mode sets* μ_i, ν_i*,* $i = 1,\ldots,f(d)-1$*, satisfy* (3.6)*. Then there exist at least two modes* $\{s,t\} \subsetneq \{1,\ldots,d\}$*, such that for* $i = 1,\ldots,f(d)-1$*,*

$$\text{either} \quad \{s,t\} \subsetneq \mu_i \quad \text{or} \quad \{s,t\} \subsetneq \nu_i. \tag{3.8}$$

Proof. We start with μ_1 and ν_1. Either μ_1 or ν_1 contains at least $\lceil \frac{d}{2} \rceil$ modes. We may use α_1 to denote the set of these $\lceil \frac{d}{2} \rceil$ modes and, without loss of generality,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

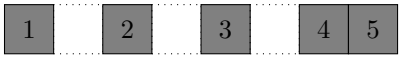

FIG. 2. *An illustration of lifting a 5th order tensor to an 8th order tensor. Nodes in gray represent the original modes. Nodes with a dotted line represent the added virtue modes.*

TABLE 1
Matrices generated by Algorithm 1 for tensors from order 3 to order 10.

order	3	4	5	6
matrices	[1;2,3] [2;1,3]	[1,2;3,4] [3,2;1,4]	[1,2;3,4,5] [3,2;1,4,5] [1,4;3,2,5]	[1,2,3;4,5,6] [4,2,3;1,5,6] [1,5,3;4,2,6]
order	7	8	9	10
matrices	[1,2,3;4,5,6,7] [1,6,7;4,5,2,3] [1,5,6,3;4,2,7]	[1,2,3,4;5,6,7,8] [5,6,3,4;1,2,7,8] [1,6,7,4;5,2,3,8]	[1,2,3,4;5,6,7,8,9] [5,6,3,4;1,2,7,8,9] [1,6,7,4;5,2,3,8,9] [5,2,3,8;1,6,7,4,9]	[1,2,3,4,5;6,7,8,9,10] [6,7,3,4,5;1,2,8,9,10] [1,7,8,4,5;6,2,3,9,10] [6,2,3,9,5;1,7,8,4,10]

assume that $\alpha_1 \subseteq \mu_1$. Consider μ_2 and ν_2. By (3.6), either μ_2 or ν_2 contains at least $\lceil \frac{d}{4} \rceil$ of the modes in α_1. Without loss of generality, we suppose that these modes are contained in μ_2, which are denoted as α_2. Continue this argument. At last, either $\mu_{f(d)-1}$ or $\nu_{f(d)-1}$ contains at least $\lceil d/2^{f(d)-1} \rceil$ modes from $\alpha_{f(d)-2}$, say, $\alpha_{f(d)-1}$. Without loss of generality assume that $\alpha_{f(d)-1} \subseteq \mu_{f(d)-1}$. As a result, $\alpha_{f(d)-1} \subseteq \alpha_i \subseteq \mu_i, i = 1, \ldots, f(d)-1$. Since $d \geq 2^{f(d)-1}+1$, there holds $\mathrm{card}(\alpha_{f(d)-1}) \geq 2$, and the results follow. ∎

Proof of Theorem 3.7. According to Lemma 3.8, there are two modes $\{s,t\}$, such that for $i = 1, \ldots, f(d)-1$, either $\{s,t\} \subsetneq \mu_i$ or $\{s,t\} \subsetneq \nu_i$. Denote the set

$$\boldsymbol{\mathcal{X}}_{d,\{s,t\}} := \{\mathcal{X} \in \mathbb{R}^{N_1 \times \cdots \times N_d} \mid \mathrm{rank}_{\mathrm{CP}}(\mathcal{X}_{[\{s,t\};1;2;\cdots;s-1;s+1;\cdots;t-1;t+1;\cdots;d]}) = 1\},$$

i.e., $\boldsymbol{\mathcal{X}}_{d,\{s,t\}}$ stands for the set of tensors which are rank-1 in a $(d-1)$th order tensor space. One then notices that for any $\mathcal{X} \in \boldsymbol{\mathcal{X}}_{d,\{s,t\}}$, it follows from (3.8) that $\mathcal{X} \in \boldsymbol{\mathcal{X}}_{d,[\mu_i;\nu_i]}$, $i = 1, \ldots, f(d)-1$, and so $\boldsymbol{\mathcal{X}}_{d,\{s,t\}} \subseteq \cap_{i=1}^{f(d)-1} \boldsymbol{\mathcal{X}}_{d,[\mu_i;\nu_i]}$. On the other hand, since $N_i > 1$, $1 \leq i \leq d$, one also notices that $\boldsymbol{\mathcal{X}}_d \subsetneq \boldsymbol{\mathcal{X}}_{d,\{s,t\}}$, and finally (3.7) holds, as desired. ∎

3.4. Balanced unfolding of a general tensor. We describe a way to add virtual modes to proper positions of the modes of the tensor, such that a general tensor can be unfolded into a balanced matrix.

Assume that $M = 2^{f(d)} - d \neq 0$. At the right-hand side of each mode beginning from mode-1, mode-2, until mode-$\lceil \frac{M}{2} \rceil$, a virtual mode is added; at the right-hand side of each mode beginning from mode-$(\lfloor \frac{d}{2} \rfloor + 1)$ until mode-$(\lfloor \frac{d}{2} \rfloor + \lfloor \frac{M}{2} \rfloor)$, a virtual mode is also added. As a result, we get a $2^{f(d)}$th order tensor. Consider an example. Suppose $\mathcal{X} \in \mathbb{R}^{N_1 \times \cdots \times N_5}$. In this case, $M = 8 - 5 = 3$. Then the lifted tensor $\mathcal{X}$ belongs to $\mathbb{R}^{N_1 \times 1 \times N_2 \times 1 \times N_3 \times 1 \times N_4 \times N_5}$; see Figure 2.

Table 1 illustrates the matrices generated by Algorithm 1 for tensors from order 3 to order 10. For convenience, the matrix is denoted by its subscript.

4. Applications to low rank tensor approximation problems. This section concentrates on low rank tensor approximation problems, particularly the best

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

rank-1 approximation problems

(4.1)
$$\min \ \frac{1}{2}\|\mathcal{A} - \lambda x_1 \circ x_2 \circ \cdots \circ x_d\|_F^2 \quad \text{s.t. } \lambda \in \mathbb{R},\ x_i \in \mathbb{R}^{N_i},\ \|x_i\| = 1,\ \ i = 1, \ldots, d,$$

or its equivalent formulation

$$\max \ \langle \mathcal{A}, x_1 \circ x_2 \circ \cdots \circ x_d \rangle \quad \text{s.t. } x_i \in \mathbb{R}^{N_i},\ \|x_i\| = 1,\ \ i = 1, \ldots, d, \tag{4.2}$$

that has been introduced in the introduction. The key idea of the relaxations proposed in this section is to first reformulate (4.1) or (4.2) to a tensor optimization problem with rank-1 constraint. Then by using Corollary 3.5, the rank-1 tensor constrained problem can be further reformulated to a matrix rank constrained problem.

To solve the matrix rank constrained problem, three approaches (relaxations) are proposed, two of which are convex relaxations, based on matrix nuclear norm regularization/constraints, while the other one is nonconvex. By examining the magnitude of the nuclear norm or the Frobenius norm of the solution, the convex relaxations are able to identify whether a solution is still a global optimizer of the original problem (4.2) or not. The solution to the nuclear norm constrained problem is then studied from the dual. The nonconvex relaxation can be solved by the alternating minimization scheme, with the output being always a rank-1 tensor, serving as an approximation to (4.1) or (4.2). Finally, the nonconvex relaxation is extended to rank-R approximation problems.

For ease of notation, for a dth order tensor $\mathcal{X}$, in this section we denote the $f(d)$ unfolding matrices chosen by Algorithm 1 as $X_{[1]}, \ldots, X_{[f(d)]}$.

4.1. Convex relaxations of (4.2). In (4.2), by letting $\mathcal{X} := x_1 \circ x_2 \circ \cdots \circ x_d$, we get the following reformulation

$$\min \ \langle -\mathcal{A}, \mathcal{X} \rangle \quad \text{s.t. } \mathcal{X} \in \mathbb{R}^{N_1 \times \cdots \times N_d},\ \|\mathcal{X}\|_F = 1,\ \operatorname{rank}_{\mathrm{CP}}(\mathcal{X}) = 1. \tag{4.3}$$

By using Corollary 3.5, the rank-1 tensor constraint $\operatorname{rank}_{\mathrm{CP}}(\mathcal{X}) = 1$ can be replaced by a set of rank-1 matrix constraints, i.e., (4.3) can be equivalently reformulated into the following optimization problem

(4.4)
$$\min \ \langle -\mathcal{A}, \mathcal{X} \rangle \ \text{ s.t. } \mathcal{X} \in \mathbb{R}^{N_1 \times \cdots \times N_d}, \|\mathcal{X}\|_F = 1,\ \operatorname{rank}(X_{[i]}) = 1,\ \ i = 1, \ldots, f(d).$$

In recent years, it is popular to replace the rank function by the matrix nuclear norm. By doing this, one expects to obtain a low rank solution via solving the nuclear norm based problems. The nuclear norm $\|\cdot\|_*$ is defined as the sum of singular values of a matrix and serves as a surrogate function to the rank function [43]. Following this line, (4.4) can be relaxed to the following nuclear norm regularized convex optimization problem

$$\min \ \langle -\mathcal{A}, \mathcal{X} \rangle + \rho \, |||\, \mathcal{X} \,|||_* \qquad \text{s.t.} \qquad \mathcal{X} \in \mathbb{R}^{N_1 \times \cdots \times N_d},\ \|\mathcal{X}\|_F \le 1, \tag{4.5}$$

where $\rho > 0$ is a regularization parameter, and we denote

$$|||\mathcal{X}|||_* := f(d)^{-1} \sum_{i=1}^{f(d)} \|X_{[i]}\|_*$$

as a regularizer to control the rank of the tensor.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

It is also possible to employ the nuclear norms as constraints, instead of using them as regularizers. This results in the following problem

$$\min \langle -\mathcal{A}, \mathcal{X} \rangle \text{ s.t. } \mathcal{X} \in \mathbb{R}^{N_1 \times \cdots \times N_d},\ \|X_{[i]}\|_* \leq 1,\quad i = 1, \ldots, f(d), \tag{4.6}$$

where the spherical constraint $\|\mathcal{X}\|_F = 1$ is dropped.

As convex relaxations, one is concerned with the relationship between their solutions and the solution of the original problem. This will be studied in sections 4.1.1 and 4.1.2. On the other hand, as well-formulated convex optimization problems, (4.5) and (4.6) can be solved by many state-of-the-art algorithms, e.g., the alternating direction method of multipliers (ADMM). The implementation of ADMM for solving (4.5) and (4.6) is given in section 4.1.3.

4.1.1. Determining whether a solution to (4.5) or (4.6) is a global minimizer of (4.3). Before stating the main results, we present a simple observation.

PROPOSITION 4.1. *Let* $X \in \mathbb{R}^{N_1 \times N_2}$. *If* $\|X\|_F = 1$ *and* $\|X\|_* = 1$, *then* $\operatorname{rank}(X) = 1$.

Proof. Let $\|X\|_* = \sum_{k=1} \sigma_k$ be the sum of its singular values. From the assumptions we have $\sum_{k=1} \sigma_k^2 = 1$ and $\sum_{k=1} \sigma_k = 1$, which imply $\sum_{k_1 \neq k_2} \sigma_{k_1} \sigma_{k_2} = 0$. Since $\sigma_k \geq 0$, the equality holds if and only if there is only one $\sigma_{\overline{k}} = 1$ while other singular values are zero. Therefore the result follows. ∎

In the following, we study properties of (4.5). Denote $\hat{p}$ as the optimal value of (4.5). The following observations, which characterize some properties of the global minimizers of (4.5), are presented first.

PROPOSITION 4.2. *If* $\hat{\mathcal{X}}$ *is an optimal solution to* (4.5), $\hat{\mathcal{X}} \neq 0$, $\hat{p} \neq 0$, *and* $\mathcal{X}^*$ *is a global optimal solution to* (4.3), *then*

1. $\|\hat{\mathcal{X}}\|_F = 1$,
2. $|||\hat{\mathcal{X}}|||_* \geq 1$,
3. *if* $|||\hat{\mathcal{X}}|||_* = 1$, *then* $\operatorname{rank}_{\mathrm{CP}}(\hat{\mathcal{X}}) = 1$,
4. $\lambda^* \leq \hat{\lambda}$ *with* $\lambda^* = \langle \mathcal{A}, \mathcal{X}^* \rangle$ *and* $\hat{\lambda} = \langle \mathcal{A}, \hat{\mathcal{X}} \rangle$.

Proof. 1. Since the zero tensor is a feasible solution to (4.5) with the associated objective value being zero, it holds that $\hat{p} < 0$. Suppose $\|\hat{\mathcal{X}}\|_F < 1$. Then the objective value of (4.5) evaluated at $\hat{\mathcal{X}}/\|\hat{\mathcal{X}}\|_F$ is $\hat{p}/\|\hat{\mathcal{X}}\|_F < \hat{p}$, which gives a contradiction.

2. Suppose that $|||\hat{\mathcal{X}}|||_* < 1$. This means that there exists at least an $\hat{X}_{[\overline{i}]}$ such that $\|\hat{X}_{[\overline{i}]}\|_* < 1$. Let $\|\hat{X}_{[\overline{i}]}\|_* = \sum_{k=1} \sigma_{\overline{i},k}$ be the sum of its singular values. Then

$$\|\hat{\mathcal{X}}\|_F^2 = \|\hat{X}_{[\overline{i}]}\|_F^2 = \sum_{k=1} \sigma_{\overline{i},k}^2 \leq \left(\sum_{k=1} \sigma_{\overline{i},k} \right)^2 < 1.$$

This contradicts item 1. Therefore, $|||\hat{\mathcal{X}}|||_* \geq 1$.

3. From $|||\hat{\mathcal{X}}|||_* = 1$, we have $\|\hat{X}_{[i]}\|_* = 1$ for $i = 1, \ldots, f(d)$, otherwise the same contradiction as for item 2 would happen. Then it follows from item 1 and Proposition 4.1 that $\operatorname{rank}(\hat{X}_{[i]}) = 1$, $1 \leq i \leq f(d)$, which together with Corollary 3.5 implies that $\operatorname{rank}_{\mathrm{CP}}(\hat{\mathcal{X}}) = 1$.

4. Since $\hat{\mathcal{X}}$ is optimal to (4.5), we have

$$\langle -\mathcal{A}, \hat{\mathcal{X}} \rangle - \langle -\mathcal{A}, \mathcal{X}^* \rangle \leq \rho \,|||\, \mathcal{X}^* \,|||_* - \rho \,|||\, \hat{\mathcal{X}} \,|||_* .$$

From item 2, $|||\mathcal{X}^*|||_* = 1 \leq |||\hat{\mathcal{X}}|||_*$, and so $\lambda^* \leq \hat{\lambda}$. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proposition 4.2 implies that it is possible to identify whether an optimizer of (4.5) is an optimizer of (4.3) by computing the sum of balanced nuclear norms. The following theorem summarizes this result.

THEOREM 4.3. *Assume that $\mathcal{A} \neq 0$. Assume that $\hat{\mathcal{X}}$ is a global minimizer of the convex problem* (4.5)*. Then $\hat{\mathcal{X}}$ is a global minimizer of the original problem* (4.3) *if and only if $\hat{\mathcal{X}} \neq 0$, $|||\hat{\mathcal{X}}|||_* = 1$, and $\hat{p} \neq 0$.*

Proof. The "only if" part is easy to verify by recalling the definition of problem (4.3). For the "if" part, by Proposition 4.2 and the assumption, there holds $\|\hat{\mathcal{X}}\|_F = 1, \mathrm{rank}_{\mathrm{CP}}(\hat{\mathcal{X}}) = 1$, and $\hat{\lambda} = \langle \mathcal{A}, \hat{\mathcal{X}} \rangle \geq \langle \mathcal{A}, \mathcal{X}^* \rangle \geq \lambda^*$. Therefore, $\hat{\mathcal{X}}$ is a global minimizer of (4.3). ∎

Next we study (4.6). The following observations also characterize some properties of the global minimizers of (4.6).

PROPOSITION 4.4. *Assume that $\mathcal{A} \neq 0$. If $\hat{\mathcal{X}}$ is an optimal solution to* (4.6)*, then*

1. *at least one constraint of* (4.6) *is active, i.e., there exists an $\overline{i}$ such that $\|\hat{X}_{[\overline{i}]}\|_* = 1$,*
2. *$\|\hat{\mathcal{X}}\|_F \leq 1$,*
3. *if $\|\hat{\mathcal{X}}\|_F = 1$, then all the constraints are active. Moreover, $\mathrm{rank}_{\mathrm{CP}}(\hat{\mathcal{X}}) = 1$.*

Proof. Items 1 and 2 are easy to verify. We consider 3. Consider the ith unfolding $\hat{X}_{[i]}$, whose nuclear norm is given by $\|\hat{X}_{[i]}\|_* = \sum_{k=1} \sigma_{i,k}$, where $\sigma_{i,k}$ denotes the kth singular value of $\hat{X}_{[i]}$. It then follows from $\|\hat{\mathcal{X}}\|_F = 1$ that

$$\|\hat{X}_{[i]}\|_*^2 = \left(\sum_{k=1} \sigma_{i,k} \right)^2 \geq \sum_{k=1} \sigma_{i,k}^2 = \|\hat{\mathcal{X}}\|_F^2 = 1.$$

This together with the constraint $\|\hat{X}_{[i]}\|_* \leq 1$ implies that $\|X_{[i]}\|_* = 1, i = 1, \ldots, f(d)$. Finally, applying Proposition 4.1 to $X_{[i]}$ and using Corollary 3.5, we have $\mathrm{rank}_{\mathrm{CP}}(\hat{\mathcal{X}}) = 1$. The proof is completed. ∎

Based on the above observations, it is also possible to determine whether an optimizer of (4.6) is still an optimizer of (4.3) by computing the Frobenius norm of the optimizer. The results are summarized in the following theorem.

THEOREM 4.5. *Assume that $\mathcal{A} \neq 0$ and $\hat{\mathcal{X}}$ is a global minimizer of the convex relaxation* (4.6)*. Then $\hat{\mathcal{X}}$ is a global minimizer of the original problem* (4.3) *if and only if $\|\hat{\mathcal{X}}\|_F = 1$.*

Proof. Similar to Theorem 4.3 we only verify the if part, which follows from Proposition 4.4 and the fact that (4.6) is a relaxation of (4.3). ∎

4.1.2. Solution properties of (4.6) from the dual. By introducing auxiliary variables $\mathcal{Y}_i$ and imposing the constraint $\|\mathcal{X}\|_F \leq 1$, (4.6) can be rewritten as

$$\min_{\mathcal{X}, \mathcal{Y}_i} \langle -\mathcal{A}, \mathcal{X} \rangle \ \text{ s.t. } \ \mathcal{Y}_i = \mathcal{X}, \ \|\mathcal{X}\|_F \leq 1, \ \|Y_{i,[i]}\|_* \leq 1, \ i = 1, \ldots, f(d). \tag{4.7}$$

According to item 2 of Proposition 4.4, $\|\mathcal{X}\|_F \leq 1$ is redundant in (4.7); in other words, (4.7) is in fact equivalent to (4.6). Nonetheless, imposing this constraint is helpful in deriving its dual. Denote the Lagrangian function of (4.7) as

$$L(\mathcal{X}, \mathcal{Y}_i, \Lambda_i) = \langle -\mathcal{A}, \mathcal{X} \rangle + \sum_{i=1}^{f(d)} \langle \Lambda_i, \mathcal{X} - \mathcal{Y}_i \rangle.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Its dual problem is given by

$$\sup_{\Lambda_i} d(\Lambda_1, \ldots, \Lambda_{f(d)}) \text{ with } d(\Lambda_i) := \min_{\|\mathcal{X}\|_F \le 1, \|Y_{i,[i]}\|_* \le 1} L(\mathcal{X}, \mathcal{Y}_i, \Lambda_i). \tag{4.8}$$

It can be shown that the dual function has the following expression

$$\begin{aligned} d(\Lambda_1, \ldots, \Lambda_{f(d)}) &= \min_{\|\mathcal{X}\|_F \le 1} \left\langle -A + \sum_{i=1}^{f(d)} \Lambda_i, \mathcal{X} \right\rangle + \sum_{i=1}^{f(d)} \min_{\|Y_{i,[i]}\|_* \le 1} \langle \Lambda_i, -\mathcal{Y}_i \rangle \\ &= -\left\| \mathcal{A} - \sum_{i=1}^{f(d)} \Lambda_i \right\|_F - \sum_{i=1}^{f(d)} \|\Lambda_{i,[i]}\|_2, \end{aligned}$$

where $\Lambda_{i,[i]}$ is the ith balanced unfolding of Λ_i. For any given $(\Lambda_1, \ldots, \Lambda_{f(d)})$, denote the set

$$\begin{aligned} &\boldsymbol{\mathcal{X}}(\Lambda_1, \ldots, \Lambda_{f(d)}) \\ &\quad := \left\{ (\mathcal{X}, \mathcal{Y}_1, \ldots, \mathcal{Y}_{f(d)}) \mid d(\Lambda_i) = L(\mathcal{X}, \mathcal{Y}_i, \Lambda_i), \|\mathcal{X}\|_F \le 1, \|Y_{i,[i]}\|_* \le 1 \right\}. \end{aligned}$$

Since the sets associated with $\mathcal{X}$ and $\mathcal{Y}_i$ are compact, $\boldsymbol{\mathcal{X}}(\Lambda_i)$ is nonempty. According to the expression of $d(\Lambda_i)$, $\boldsymbol{\mathcal{X}}(\Lambda_i)$ can also be written as

$$\begin{aligned} \boldsymbol{\mathcal{X}}(\Lambda_1, \ldots, \Lambda_{f(d)}) = \Big\{ (\mathcal{X}, \mathcal{Y}_1, \ldots, \mathcal{Y}_{f(d)}) \mid \mathcal{X} \in \arg\min_{\|\mathcal{X}\|_F \le 1} \left\langle -\mathcal{A} + \sum \Lambda_i, \mathcal{X} \right\rangle, \\ \mathcal{Y}_i \in \arg\min_{\|Y_{i,[i]}\|_* \le 1} \langle -\Lambda_i, \mathcal{Y}_i \rangle \Big\}. \end{aligned} \tag{4.9}$$

Note that the $\mathcal{Y}_i$-related subproblem of (4.9) amounts to computing the leading singular value of $\Lambda_{i,[i]}$, and hence the solution $Y_{i,[i]}$ to the subproblem can be chosen as a rank-1 matrix. With the above notation, the first order optimality condition of (4.8) can be written as

$$0 \in \partial d(\Lambda_1, \ldots, \Lambda_{f(d)}) \Longleftrightarrow 0 \in \operatorname{conv} \left\{ (\mathcal{X} - \mathcal{Y}_1, \ldots, \mathcal{X} - \mathcal{Y}_{f(d)}) \mid (\mathcal{X}, \mathcal{Y}_i) \in \boldsymbol{\mathcal{X}}(\Lambda_i) \right\}, \tag{4.10}$$

where $\partial d(\cdot)$ is the subgradient of $d(\cdot)$ at Λ_i. The following proposition shows that, under certain assumptions, a critical point of (4.8) can indicate a global minimizer of the original problem (4.3).

PROPOSITION 4.6. *Let Λ_i^* be a critical point of* (4.8). *Assume that $\sum \Lambda_i^* - \mathcal{A} \neq 0$, and the leading singular value of Λ_i^* is unique, $1 \le i \le f(d)$. Assume that $(\mathcal{X}^*, \mathcal{Y}_i^*) \in \boldsymbol{\mathcal{X}}(\Lambda_i^*)$. Then $\mathcal{X}^*$ is a global minimizer of the original problem* (4.3).

Proof. We first show that $\mathcal{X}^*$ is a rank-1 tensor. According to the assumption, given Λ_i^*, the $\mathcal{Y}_i$-related subproblem in (4.9) must have a unique solution. Moreover, it holds that $\mathcal{Y}_{i,[i]}^*$ must be a rank-1 matrix, $1 \le i \le f(d)$. On the other hand, since $\sum \Lambda_i^* - \mathcal{A} \neq 0$, $\mathcal{X}^*$ is unique, and $\mathcal{X}^* = \frac{\mathcal{A} - \sum \Lambda_i^*}{\|\sum \Lambda_i^* - \mathcal{A}\|_F}$. These two facts show that $\partial d(\Lambda_i)$ at Λ_i^* is singleton, i.e., $d(\cdot)$ is differentiable at Λ_i^*, which together with (4.10) implies that $\mathcal{X}^* = \mathcal{Y}_i^*$. In particular, $\operatorname{rank}(X_{[i]}^*) = 1$, which by Corollary 3.5 in turn gives that $\operatorname{rank}_{\mathrm{CP}}(\mathcal{X}^*) = 1$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

To show that $\mathcal{X}^*$ is a global minimizer of (4.3), it remains to show that $\mathcal{X}^*$ is a global minimizer of (4.7). By the definitions of $\mathcal{X}^*$ and $\mathcal{Y}_i^*$, and that $\mathcal{X}^* = \mathcal{Y}_i^*$,

$$\begin{aligned}\langle -\mathcal{A}, \mathcal{X}^* \rangle &= \left\langle -\mathcal{A} + \sum_{i=1} \Lambda_i^*, \mathcal{X}^* \right\rangle - \left\langle \sum_{i=1} \Lambda_i^*, \mathcal{X}^* \right\rangle \\ &= \left\langle -\mathcal{A} + \sum_{i=1} \Lambda_i^*, \frac{\mathcal{A} - \sum_{i=1} \Lambda_i^*}{\|\mathcal{A} - \sum_{i=1} \Lambda_i^*\|_F} \right\rangle - \sum_{i=1} \langle \Lambda_i^*, \mathcal{Y}_i^* \rangle \\ &= -\left\| \mathcal{A} - \sum_{i=1} \Lambda_i^* \right\|_F - \sum_{i=1} \|\Lambda_i^*\|_2 \\ &= d(\Lambda_1^*, \ldots, \Lambda_{f(d)}^*).\end{aligned}$$

The above equality shows that there is no duality gap between (4.7) and (4.8). As a result, $\mathcal{X}^*$ is a minimizer of (4.7). This together with the rank-1 property of $\mathcal{X}^*$ and the fact that (4.7) is a convex relaxation of the original problem (4.3) shows that $\mathcal{X}^*$ is also a global minimizer of (4.3). ∎

4.1.3. Implementation of (4.5) and (4.6) via ADMM. By introducing $f(d)$ variables $\mathcal{Y}_i, i = 1, \ldots, f(d)$, problem (4.5) can be formulated as

$$\begin{aligned}\min_{\mathcal{X}, \mathcal{Y}_i} \ & \langle -\mathcal{A}, \mathcal{X} \rangle + \frac{\rho}{f(d)} \sum_{i=1}^{f(d)} \|Y_{i,[i]}\|_* \\ \text{s.t.} \quad & \mathcal{Y}_i = \mathcal{X}, \ i = 1, \ldots, f(d), \ \|\mathcal{X}\|_F \le 1,\end{aligned} \tag{4.11}$$

where $Y_{i,[i]}$ is the ith balanced unfolding of $\mathcal{Y}_i$. The ADMM for (4.11) is given in Algorithm 2, where $\mathcal{D}_{\rho/(\tau \cdot f(d))}$ is the matrix shrinkage operator [7].

ALGORITHM 2. ADMM FOR (4.5).

Input: tensor $\mathcal{A}$, parameters $\rho > 0, \tau > 0$, starting guess $\mathcal{X}^0, \mathcal{Y}_i^0, \Lambda_i^0$

for $i = 1, 2, \ldots$ **do**

$\mathcal{X}^{k+1} = \mathcal{P}_{\|\cdot\|_F}\left[\frac{1}{\tau \cdot f(d)}(\mathcal{A} + \sum_{i=1}^{f(d)} \Lambda_i^k + \tau \sum_{i=1}^{f(d)} \mathcal{Y}_i^k)\right]$.

$Y_{i,[i]}^{k+1} = \mathcal{D}_{\rho/(\tau \cdot f(d))}\left[X_{[i]}^{k+1} - \frac{1}{\tau}\Lambda_{i,[i]}^k\right], \qquad i = 1, \ldots, f(d)$.

$\Lambda_i^{k+1} = \Lambda_i^k - \tau(\mathcal{X}^{k+1} - \mathcal{Y}_i^{k+1}), \qquad i = 1, \ldots, f(d)$.

end for

Similarly, (4.6) can be formulated as

$$\min_{\mathcal{X}, \mathcal{Y}_i} \langle -\mathcal{A}, \mathcal{X} \rangle \ \text{ s.t. } \mathcal{Y}_i = \mathcal{X}, \ \|Y_{i,[i]}\|_* \le 1, \ i = 1, \ldots, f(d). \tag{4.12}$$

The ADMM for solving (4.12) is given in Algorithm 3, where $\mathcal{P}_{\|\cdot\|_*}$ denotes the projection onto the nuclear norm ball $\{X | \|X\|_* \le 1\}$. To compute $\mathcal{P}_{\|\cdot\|_*}$, one can first compute the SVD, and then project the vector of singular values onto the L_1 norm ball.

4.2. A nonconvex relaxation of (4.1). We consider a nonconvex relaxation of (4.1) in this subsection, and employ an alternating minimization scheme to solve it. To ensure that the solution generated by the alternating minimization scheme is rank-1, the tensor involved in the relaxation is of third or fourth order. First, we

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ALGORITHM 3. ADMM FOR (4.6).

Input: tensor $\mathcal{A}$, parameters $\tau > 0$, starting guess $\mathcal{X}^0, \mathcal{Y}_i^0, \Lambda_i^0$
for $i = 1, 2, \ldots$ **do**
$\mathcal{X}^{k+1} = \frac{1}{\tau \cdot f(d)}(\mathcal{A} + \sum_{i=1}^{f(d)} \Lambda_i^k + \tau \sum_{i=1}^{f(d)} \mathcal{Y}_i^k)$.
$Y_{i,[i]}^{k+1} = \mathcal{P}_{\|\cdot\|_*}\left[X_{[i]}^{k+1} - \frac{1}{\tau}\Lambda_{i,[i]}^k\right], \qquad i = 1, \ldots, f(d)$.
$\Lambda_i^{k+1} = \Lambda_i^k - \tau(\mathcal{X}^{k+1} - \mathcal{Y}_i^{k+1}), \qquad i = 1, \ldots, f(d)$.
end for

reformulate (4.1) as

$$\min \ \frac{1}{2}\|\mathcal{A} - \mathcal{X}\|_F^2 \qquad \text{s.t. } \mathcal{X} \in \mathbb{R}^{N_1 \times \cdots \times N_d}, \ \operatorname{rank}_{\mathrm{CP}}(\mathcal{X}) = 1, \tag{4.13}$$

where $d = 3$ or 4. By using the rank-1 equivalence property and introducing an auxiliary variable $\mathcal{Y}$, (4.13) can be reformulated as

$$\min \ \frac{1}{2}\|\mathcal{A} - \mathcal{X}\|_F^2 \qquad \text{s.t. } \mathcal{X} = \mathcal{Y}, \ \operatorname{rank}(Y_{[1]}) = 1, \ \operatorname{rank}(X_{[2]}) = 1.$$

By relaxing the constraint $\mathcal{X} - \mathcal{Y}$ and imposing it on the objective function, we obtain the following nonconvex relaxation of (4.13)

$$\min \ F(\mathcal{X}, \mathcal{Y}) := \frac{1}{2}\|\mathcal{A} - \mathcal{Y}\|_F^2 + \frac{\rho}{2}\|\mathcal{X} - \mathcal{Y}\|_F^2 \ \text{ s.t. } \operatorname{rank}(Y_{[1]}) \le 1, \ \operatorname{rank}(X_{[2]}) \le 1, \tag{4.14}$$

where $\rho > 0$ is a regularization parameter. It is possible to employ the alternating minimization scheme to solve (4.14), namely, for $k = 0, 1, 2, \ldots$, compute

$$\mathcal{Y}^{k+1} \in \arg\min\nolimits_{\operatorname{rank}(Y_{[1]}) \le 1} F(\mathcal{X}^k, \mathcal{Y}), \qquad \mathcal{X}^{k+1} \in \arg\min\nolimits_{\operatorname{rank}(X_{[2]}) \le 1} F(\mathcal{X}, \mathcal{Y}^{k+1}). \tag{4.15}$$

If the above scheme terminates at step K, then $\overline{\mathcal{X}} := \mathcal{X}^K / \|\mathcal{X}^K\|_F$ is returned, which serves as an approximation solution to (4.13).

For any $\mathcal{X}^k$, finding $\mathcal{Y}^{k+1}$ reduces to solving

$$\min_{\operatorname{rank}(Y_{[1]}) \le 1} \|\mathcal{Y} - (1+\rho)^{-1}(\mathcal{A} + \rho\mathcal{X}^k)\|_F,$$

while computing $\mathcal{X}^{k+1}$ amounts to solving $\min_{\operatorname{rank}(X_{[2]}) \le 1} \|\mathcal{X} - \mathcal{Y}^k\|_F$. Therefore, the alternating minimization scheme (4.15) is well-defined.

Since $F(\cdot, \cdot)$ is coercive and $\{F(\mathcal{X}^k, \mathcal{Y}^k)\}$ is nonincreasing, the sequence $\{(\mathcal{X}^k, \mathcal{Y}^k)\}$ generated by scheme (4.15) is bounded. The following proposition holds.

PROPOSITION 4.7. *Let $\{(\mathcal{X}^k, \mathcal{Y}^k)\}$ be a sequence generated by the alternating minimization scheme* (4.15). *Then every limit point $(\mathcal{X}^*, \mathcal{Y}^*)$ of $\{(\mathcal{X}^k, \mathcal{Y}^k)\}$ is a stationary point of* (4.14)*, i.e., there holds*

$$F(\mathcal{X}^*, \mathcal{Y}^*) \le F(\mathcal{X}, \mathcal{Y}^*), \qquad F(\mathcal{X}^*, \mathcal{Y}^*) \le F(\mathcal{X}^*, \mathcal{Y}) \qquad \forall \mathcal{X}, \mathcal{Y} \text{ feasible for (4.14)}.$$

A brief proof of the above proposition is provided in Appendix A.

It will be shown in the following that $\mathcal{X}^k$ is always a rank-1 tensor. To this end, denote $\mathcal{T}_{[i]}$ the folding operator such that $\mathcal{X} = \mathcal{T}_{[i]}(X_{[i]})$, and $P_{[i]}$ such that $P_{[i]}(\mathcal{X}) := \arg\min_{\operatorname{rank}(Z) \le 1} \|X_{[i]} - Z\|_F$. Let $\tau_{[i]} = \mathcal{T}_{[i]} \circ P_{[i]}$, $1 \le i \le f(d)$. The following result holds for a dth order tensor.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

PROPOSITION 4.8. *For any dth order tensor $\mathcal{X}$, $d \geq 3$, $\mathcal{X} \neq 0$, let $\mathcal{X}_0 = \tau_{[f(d)]} \cdots \tau_{[1]}(\mathcal{X})$. Then* $\operatorname{rank}_{\mathrm{CP}}(\mathcal{X}_0) = 1$.

Proof. Since any dth order tensor can be lifted to a $2^{f(d)}$th order tensor, without loss of generality, we consider $d = 2^n$. When $n = 2$, let $\mathcal{Y} = \tau_{[1]}(\mathcal{X})$. Then $Y_{[1]}$ can be expressed as $Y_{[1]} = \lambda y_{[1,2]} \circ y_{[3,4]}$, where $\lambda \in \mathbb{R}$ is a scalar, $y_{[1,2]}$ and $y_{[3,4]}$ are two normalized vectors. We further denote $y_{[1,2]} = \sum_{i=1} \sigma_i y_{1,i} \otimes y_{2,i}$ and $y_{[3,4]} = \sum_{j=1} \rho_j y_{3,j} \otimes y_{4,j}$ as the SVDs of $y_{[1,2]}$ and $y_{[3,4]}$ (in terms of their folded matrix forms), where σ_i and ρ_j are in decreasing order. Then

$$Y_{[2]} = \sum_{i=1} \sum_{j=1} \lambda \sigma_i \rho_j \big(y_{3,j} \otimes y_{2,i}\big) \circ \big(y_{1,i} \otimes y_{4,j}\big). \tag{4.16}$$

Since $y_{3,j} \otimes y_{2,i}$, $y_{1,i} \otimes y_{4,j}$ are normalized and mutually orthogonal, they form the left and right singular vectors for $Y_{[2]}$, respectively, with $\sigma_i \rho_j$ being the singular values. As a result, $\mathcal{X}_0 = \lambda \sigma_1 \rho_1 y_{1,1} \circ y_{2,1} \circ y_{3,1} \circ y_{4,1}$. Suppose that the result holds when $n = k$; when $n = k + 1$, let $\mu_1 = \{1, 2\}, \mu_2 = \{3, 4\}, \ldots, \mu_{2^k} = \{2^{k+1} - 1, 2^{k+1}\}$. By induction, the unfolding tensor $\mathcal{Y}_{[\mu_1;\mu_2;\ldots;\mu_{2^k}]}$ of $\mathcal{Y} = \tau_{[k]} \cdots \tau_{[1]}(\mathcal{X})$ is rank-1. As a result, $Y_{[k+1]}$ can be expressed in a similar manner to (4.16)

$$\begin{aligned} &Y_{[k+1]} \\ &= \sum\nolimits_{i_1=1} \cdots \sum\nolimits_{i_{2^k}=1} \hat{\lambda} \sigma_{1,i_1} \cdots \sigma_{2^k,i_{2^k}} (y_{\mu_1^l,i_1} \otimes \cdots \otimes y_{\mu_{2^k}^l,i_{2^k}}) \circ (y_{\mu_1^r,i_1} \otimes \cdots \otimes y_{\mu_{2^k}^r,i_{2^k}}), \end{aligned}$$

where $\mu_i^l \cup \mu_i^r = \mu_i$, with $\operatorname{card}(\mu_i^l) = \operatorname{card}(\mu_i^r) = 1$, and so $\mathcal{X}_0 = \hat{\lambda} \sigma_{1,1} \cdots \sigma_{2^k,1} y_{1,1} \circ \cdots y_{2^{k+1},1}$, showing that $\mathcal{X}_0$ is rank-1. ∎

COROLLARY 4.9. *If $\mathcal{A} \neq 0$, $\mathcal{A}$ is not rank-1, and $\mathcal{X}^0 \neq -\rho^{-1}\mathcal{A}$, then $\mathcal{X}^k$ generated by* (4.15) *is always a rank-1 tensor.*

Proof. From the discussion after (4.15) and using the notation $\tau_{[i]}$, the alternating minimization scheme (4.15) can be written as

$$\mathcal{X}^{k+1} = \tau_{[2]} \tau_{[1]}((1 + \rho)^{-1}(\mathcal{A} + \rho \mathcal{X}^k)). \tag{4.17}$$

If $\mathcal{X}^0 \neq -\rho^{-1}\mathcal{A}$, then $\mathcal{X}^1$ is rank-1. This together with $\mathcal{A}$ being not rank-1 implies that $\mathcal{X}^2, \mathcal{X}^3, \ldots$ are rank-1. ∎

Remark 4.1. Proposition 4.8 also tells us that to compute $\mathcal{X}^{k+1}$, one can first compute the best rank-1 approximation to $Y_{[1]} = (A + \rho X^k)_{[1]}$, and then, respectively, compute the best rank-1 approximations to $y_{[1,2]}$ and $y_{[3,4]}$.

Remark 4.2. We explain why the nonconvex relaxation (4.14) are not suitable for tensors of order higher than 4. Consider the following nonconvex relaxation

$$\min_{\operatorname{rank}(\mathrm{X}_{\mathrm{i},[\mathrm{i}]})=1, i=1,\ldots,f(d)} F(\mathcal{X}_1, \ldots, \mathcal{X}_{f(d)}) := \frac{1}{2}\|\mathcal{A} - \mathcal{X}_1\|_F^2 + \frac{\rho}{2} \sum_{i=2}^{f(d)} \|\mathcal{X}_1 - \mathcal{X}_i\|_F^2.$$

When $f(d) > 2$, with simple computation, one can verify that none of the resulting $\mathcal{X}_i^k$ are rank-1 tensors, if the alternating minimization scheme is applied to solve the above problem. In this sense, the relaxation (4.14) and the scheme (4.15) are not suitable for tensors of order higher than 4. However, it is possible to extend (4.17) to the following iterative scheme

$$\mathcal{X}^{k+1} = \tau_{[f(d)]} \cdots \tau_{[1]}((1 + \rho)^{-1}(\mathcal{A} + \rho \mathcal{X}^k))$$

as a heuristic algorithm for rank-1 approximation to a general dth order tensor.

Remark 4.3. At first glance, there exists a similarity between the formula $\mathcal{X}_0 = \tau_{[f(d)]} \cdots \tau_{[1]}(\mathcal{X})$ in Proposition 4.8 and higher order SVDs (HOSVDs) [16]. However,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

the noncommutativity of $\tau_{[i]}$ and $\tau_{[j]}$, $i \neq j$ makes the above formula quite different from HOSVDs. In fact, due to the noncommutativity, it is not easy to estimate an upper bound for $\|\mathcal{X} - \mathcal{X}_0\|_F$. However, it has been proved that for HOSVDs, one has $\|\mathcal{X} - \mathcal{X}^{HOSVD}\|_F \leq \sqrt{d}\|\mathcal{X} - \mathcal{X}^*\|_F$, where $\mathcal{X}^{HOSVD}$ is generated by an HOSVD, and $\mathcal{X}^*$ denotes a best rank-1 approximation to $\mathcal{X}$; see, e.g., [16, Property 10] and [22, Lemma 2.6]. Nonetheless, with additional assumptions on the leading singular value of $X_{[1]}$, it is still possible to give a reasonable upper bound for $\|\mathcal{X} - \mathcal{X}_0\|_F$, as shown in the following proposition.

PROPOSITION 4.10. *For any dth order tensor $\mathcal{X}$, $d \geq 3$, $\mathcal{X} \neq 0$, let $\mathcal{X}_0$ be defined as in Proposition* 4.8. *Assuming* $\operatorname{rank}(X_{[1]}) = r$, *denoting the singular values of $X_{[1]}$ as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and satisfying $\sigma_1 \leq \alpha\sqrt{\sum_{i=2}^r \sigma_i^2}$ with $\alpha \geq 1/\sqrt{r-1}$, we have*

$$\|\mathcal{X} - \mathcal{X}_0\|_F \leq (1 + 2\alpha)\|\mathcal{X} - \mathcal{X}^*\|_F, \tag{4.18}$$

*where $\mathcal{X}^*$ is a best rank-*1 *approximation to $\mathcal{X}$. If $d = 3$ or* 4, *then the factor $1 + 2\alpha$ can be improved to $1 + \alpha$.*

Proof. The proof goes as follows:

$$\begin{aligned}
\|\mathcal{X} - \mathcal{X}_0\|_F &\leq \|\mathcal{X} - \tau_{[1]}(\mathcal{X})\|_F + \|\tau_{[1]}(\mathcal{X}) - \mathcal{X}_0\|_F \\
&\leq \|\mathcal{X} - \tau_{[1]}(\mathcal{X})\|_F + \|\tau_{[1]}(\mathcal{X})\|_F + \|\mathcal{X}_0\|_F \\
&\leq \|\mathcal{X} - \tau_{[1]}(\mathcal{X})\|_F + 2\|\tau_{[1]}(\mathcal{X})\|_F \qquad (4.19a) \\
&\leq (1 + 2\alpha)\|\mathcal{X} - \tau_{[1]}(\mathcal{X})\|_F \qquad (4.19b) \\
&\leq (1 + 2\alpha)\|\mathcal{X} - \mathcal{X}^*\|_F, \qquad (4.19c)
\end{aligned}$$

where (4.19a) comes from the property of $\tau_{[i]}$ that

$$\|\mathcal{X}_0\|_F = \|\tau_{[f(d)]} \cdots \tau_{[1]}(\mathcal{X})\|_F \leq \|\tau_{[f(d)-1]} \cdots \tau_{[1]}(\mathcal{X})\|_F \leq \cdots \leq \|\tau_{[1]}(\mathcal{X})\|_F;$$

(4.19b) holds because

$$\|\tau_{[1]}(\mathcal{X})\|_F = \|P_{[1]}(\mathcal{X})\|_F = \sigma_1 \leq \alpha\sqrt{\sum_{i=2}^r \sigma_i^2} = \alpha\|\mathcal{X} - \tau_{[1]}(\mathcal{X})\|_F,$$

and (4.19c) is due to that $P_{[1]}(\mathcal{X})$ is a best rank-1 matrix approximation to $X_{[1]}$.

If $d = 3$ or 4, then we have $\|\tau_{[1]}(\mathcal{X}) - \tau_{[2]}\tau_{[1]}(\mathcal{X})\|_F \leq \|\tau_{[1]}(\mathcal{X})\|_F$. The result follows. ∎

Remark 4.4. In the above proposition, if in particular $\sigma_1 \leq \sqrt{\sum_{i=2}^r \sigma_i^2}$, then the factor in (4.18) is 3. The proposition also indicates that if all the singular values of $X_{[1]}$ lie in a small interval, and $\operatorname{rank}(X_{[1]})$ is large enough, then $\|\mathcal{X} - \mathcal{X}_0\|_F$ may get close to $\|\mathcal{X} - \mathcal{X}^*\|_F$: Consider an extreme case that all the modes of $\mathcal{X}$ take the same dimension n, $\operatorname{rank}(X_{[1]}) = n^{\lfloor d/2 \rfloor}$, and $\sigma_1 = \cdots = \sigma_{\operatorname{rank}(X_{[1]})}$. Then the factor in (4.18) is $1 + 2/\sqrt{n^{\lfloor d/2 \rfloor} - 1}$, showing that $\mathcal{X}_0$ is nearly optimal as n and d go large. On the other hand, numerically we observe that without additional assumptions, the relation $\|\mathcal{X} - \mathcal{X}_0\|_F \leq \lceil \log_2(d) \rceil \|\mathcal{X} - \tau_{[1]}(\mathcal{X})\|_F \leq \lceil \log_2(d) \rceil \|\mathcal{X} - \mathcal{X}^*\|_F$ always holds. This needs further research.

4.3. Extension of relaxation (4.14) to rank-R approximation when $R > 1$. Given a dth order tensor $\mathcal{A} \in \mathbb{R}^{N_1 \times \cdots \times N_d}$ and an integer R, the low rank tensor

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

approximation can be formulated as [17, 31]

$$\begin{aligned} &\inf \ \left\| \mathcal{A} - \sum_{i=1}^{R} \lambda_i x_{1,i} \circ \cdots \circ x_{d,i} \right\|_F^2 \\ &\text{s.t. } \lambda_i \in \mathbb{R}, x_{k,i} \in \mathbb{R}^{N_k}, \|x_{k,i}\| = 1, \ i = 1, \ldots, R, \ k = 1, \ldots, d. \end{aligned} \tag{4.20}$$

The problem has succeeded in a wide range of applications [15, 20, 3, 2], and a variety of methods have been proposed to address this problem; see, e.g., [9, 24, 42, 14, 39]. In particular, the alternating least squares (ALS) method [31] has been used most widely due to its simple implementation. The best rank-R approximation problem is degenerate when $R > 1$, due to the fact that the set of tensors of at most rank R is not closed if $R > 1$; see, e.g., [18, 31].

With slight modifications, (4.14) can be extended as a relaxation for (4.20) when $d = 3$ or 4 by denoting $\mathcal{X} := \sum_{i=1}^{R} \lambda_i x_{1,i} \circ \cdots \circ x_{d,i}$,

$$\min \ F(\mathcal{X}, \mathcal{Y}) := \frac{1}{2}\|\mathcal{A} - \mathcal{Y}\|_F^2 + \frac{\rho}{2}\|\mathcal{X} - \mathcal{Y}\|_F^2 \ \text{ s.t. } \ \mathrm{rank}(Y_{[1]}) \le R, \ \mathrm{rank}(X_{[2]}) \le R, \tag{4.21}$$

and the following alternating minimization scheme can be applied

$$\mathcal{Y}^{k+1} \in \arg \min_{\mathrm{rank}(Y_{[1]}) \le R} F(\mathcal{X}^k, \mathcal{Y}), \qquad \mathcal{X}^{k+1} \in \arg \min_{\mathrm{rank}(X_{[2]}) \le R} F(\mathcal{X}, \mathcal{Y}^{k+1}). \tag{4.22}$$

When $R > 1$, the tensor $\overline{\mathcal{X}}$ generated by (4.22) might not be rank-R. In this case, one can apply the ALS method [31] to $\overline{\mathcal{X}}$ to get a rank-R tensor.

5. Numerical experiments. Numerical experiments on rank-1 and rank-R tensor approximations are presented in this section. In particular, in rank-1 tensor approximation problem, we show that the proposed methods can be used to solve the maximum-clique problem (MCP) and tensor completion. All the computations are conducted on an Intel i7-3770 CPU desktop computer with 16 GB of RAM. The supporting software is MATLAB R2013a.

5.1. Rank-1 approximation. The first experiment is focused on solving (4.2). The tensors $\mathcal{A}$ in this experiment are generated as $\mathcal{A} = \sum_{i=1}^{5} a_{1,i} \circ \cdots \circ a_{d,i}$, where every vector $a_{i,k}$ is randomly generated by the standard normal distribution, and $d = 3$ or 4, with the same dimension along each mode. Specifically, the dimensions vary from 20 to 50. For each specified order and dimension, we generate 50 instances.

The two convex relaxations (4.5), (4.6), and the nonconvex one (4.14) are tested. (4.5) and (4.6) are solved by ADMM (Algorithms 2 and 3, respectively), while (4.14) is solved by scheme (4.15). In the following, these three methods are respectively denoted as Model (4.5), Model (4.6), and Scheme (4.15). For Models (4.5) and (4.6), we examine their ability to find the global solution to (4.2); for Scheme (4.15), the value $\langle \mathcal{A}, \overline{\mathcal{X}} \rangle$ will be evaluated, where $\overline{\mathcal{X}}$ is the tensor generated by Scheme (4.15) and is normalized. Besides, because Proposition 4.10 indicates that $\tau_{[2]}\tau_{[1]}(\mathcal{A})$ is an approximation solution, we thus examine its quality as an initial guess for other local search methods.

The ρ in Algorithm 2 for solving Model (4.6) is chosen as follows. Suppose that $\mathcal{X}^*$ is a global minimizer of (4.5) with $\mathrm{rank}_{\mathrm{CP}}(\mathcal{X}^*) = 1$. Then

$$\langle -\mathcal{A}, \mathcal{X}^* \rangle + \rho \, |||\, \mathcal{X} |||_* = \langle -\mathcal{A}, \mathcal{X}^* \rangle + \rho \le 0 \quad \Rightarrow \quad \rho \le \langle \mathcal{A}, \mathcal{X}^* \rangle = v^*.$$

Therefore, it is reasonable to choose $\rho \in (0, v^*)$ such that the global solution is not

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 2
Reports of Model (4.5)*, Model* (4.6)*, Scheme* (4.15)*, HOPM and MBI on different cases.*

	Model (4.5)		Model (4.6)		Scheme (4.15)			HOPM			MBI		
$[d, n]$	#succ	time	#succ	time	value	time	iter	value	time	iter	value	time	iter
[3,20]	42	0.36	45	0.79	111.2	0.10	17.7	106.4	0.02	8.5	108.4	0.13	26.7
[3,25]	46	0.31	46	0.60	154.3	0.11	18.7	148.1	0.02	7.4	150.9	0.12	22.9
[3,30]	45	0.36	46	0.68	195.2	0.11	18.7	182.6	0.02	7.8	183.7	0.13	25.9
[3,35]	45	0.54	47	0.85	250.1	0.12	17.6	225.7	0.02	6.9	230.2	0.11	20.2
[3,40]	42	0.61	45	1.15	299.3	0.14	18.9	267.1	0.03	7.6	280.7	0.14	25.6
[3,45]	41	0.83	42	1.57	337.1	0.16	21.1	319.4	0.02	6.7	321.3	0.11	21.3
[3,50]	41	1.31	44	2.23	412.9	0.18	20.5	390.0	0.02	6.2	389.1	0.09	16.5
[4,20]	43	4.20	48	5.56	513.5	0.20	12.7	405.9	0.03	5.2	412.3	0.13	15.4
[4,25]	47	18.13	46	21.61	810.9	0.30	12.7	670.3	0.04	4.8	663.4	0.17	14.3
[4,30]	37	44.35	48	51.44	1123.7	0.53	12.4	916.7	0.07	4.8	880.7	0.26	13.6
[4,35]	33	87.26	46	98.51	1455.5	0.96	12.9	1226.0	0.12	4.7	1239.3	0.47	13.0
[4,40]	35	369.76	45	396.41	2061.9	1.50	13.0	1733.7	0.17	4.6	1772.5	0.62	12.8
[4,45]	26	451.92	43	478.15	2536.0	2.30	12.7	2106.2	0.28	4.5	2116.1	1.02	12.3
[4,50]	26	1408.08	40	1405.89	3102.0	3.40	12.7	2619.7	0.42	4.6	2658.3	1.32	12.0

TABLE 3
Reports of Scheme (4.15)*, Model* (4.5)+*Scheme* (4.15)*, and Model* (4.6)+*Scheme* (4.15) *on different cases in terms of objective values.*

$[d, n]$	[3,20]	[3,30]	[3,40]	[3,50]	[4,20]	[4,30]	[4,40]	[4,50]
Scheme (4.15)	125.2	253.2	378.9	541.8	611.3	1556.8	2368.8	3539.6
Model (4.5)+Scheme (4.15)	127.9	258.2	384.3	548.6	611.5	1555.9	2369.8	3539.9
Model (4.6)+Scheme (4.15)	127.9	258.2	384.3	548.6	612.6	1553.8	2370.1	3540.8

zero. On the other hand, a large ρ can potentially reduce the tensor rank. We therefore apply Scheme (4.15) to $\mathcal{A}$ within 5 iterations, and let $\rho = \langle \mathcal{A}, \overline{\mathcal{X}} \rangle$, where $\overline{\mathcal{X}}$ is generated by Scheme (4.15). The τ in Algorithm 2 is 50. The τ in Algorithm 3 is 100. The ρ in Scheme (4.15) is 0.25.

The stopping criterion for Algorithms 2 and 3 is $\max_i \|\Lambda_{[i]}^{k+1} - \Lambda_{[i]}^{k}\| \le 10^{-4}$; for Scheme (4.15) it is $\|\mathcal{X}^{k+1} - \mathcal{X}^{k}\|_F \le 10^{-4}$; for higher order power method (HOPM) [17] (the code in Tensorlab [48] is used) and maximum block improvement (MBI) [11] (the code is implemented from scratch, where the necessary tensor-vector multiplications are supported also by Tensorlab) that will be used later, it is $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^{k}\| \le 10^{-4}$, where $\boldsymbol{x}$ stands for the tuple $(x_1, \ldots, x_d)$. The starting guess for Algorithms 2 and 3 is the zero tensor, while that for Scheme (4.15), HOPM, and MBI will be specified later.

Results of Models (4.5) and (4.6) are reported in Table 2, from columns 2 to 5, where "#succ" denotes the number of times that Models (4.5) and (4.6) achieve the global optimum of (4.2); the unit of computational time is seconds. From the table, it seems that (4.6) performs better than (4.5). This is probably because it is not easy to determine a good parameter ρ for Model (4.5). On the other hand, (4.5) is more efficient than (4.6). In case that the two convex models cannot find a global solution to the original problem, we compute an approximation solution as follows: Let $\hat{\mathcal{X}}$ be generated by Model (4.5) or Model (4.6); then Scheme (4.15) is applied to find a rank-1 approximation $\overline{\mathcal{X}}$ to $\hat{\mathcal{X}}$, with starting guess $\mathcal{X}^0$ being the zero tensor. The methods are, respectively, termed Model (4.5)+Scheme (4.15) and Model (4.6)+Scheme (4.15). Their results in terms of values $\langle \mathcal{A}, \overline{\mathcal{X}} \rangle$ are compared with Scheme (4.15) and shown in Table 3, from which we can observe that the combined methods may have better performance than only using Scheme (4.15).

Then, Scheme (4.15) is compared with HOPM and MBI. The starting guess takes the same randomly generated and normalized $\boldsymbol{x}^0$ for HOPM, MBI, and Scheme (4.15) (for Scheme (4.15), $\mathcal{X}^0$ is the outer product of vectors in $\boldsymbol{x}^0$). Results are reported in Table 2, from columns 6 to 14, where "value" denotes $\langle \mathcal{A}, \overline{\mathcal{X}} \rangle$ for Scheme (4.15) with $\overline{\mathcal{X}}$ normalized, and $\langle \mathcal{A}, \overline{x}_1 \circ \cdots \overline{x}_d \rangle$ for HOPM or MBI; "iter" specifies the iterations. From the table, we observe that in terms of the objective value, Scheme (4.15) may

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

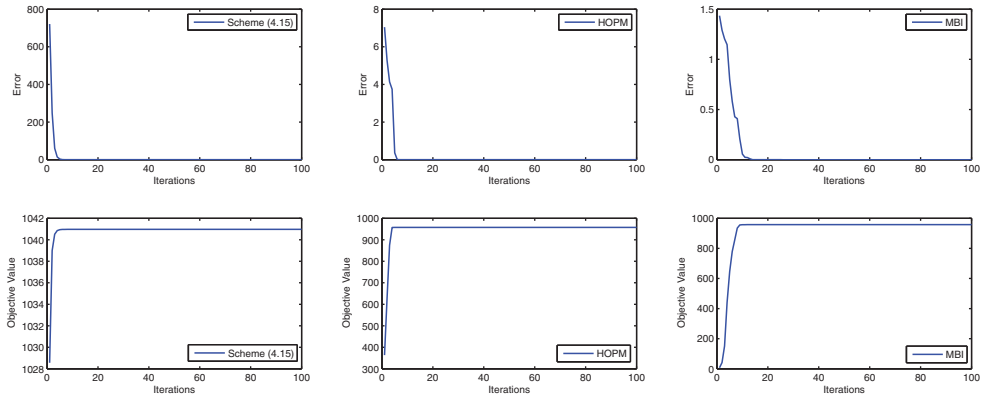

FIG. 3. *Plots of error versus iteration, and objective value versus iteration.*

TABLE 4
Reports of initial values from different initialization schemes.

	Random		HOSVD		seq-HOSVD		SP	
$[d, n]$	value	time	value	time	value	time	value	time
[3,30]	1.8	0.00	162.4	0.03	189.7	0.03	189.7	0.02
[3,40]	1.6	0.00	222.7	0.03	278.3	0.03	278.3	0.02
[3,50]	1.3	0.00	328.9	0.03	394.2	0.04	394.2	0.02
[4,30]	1.4	0.00	765.0	0.20	1048.7	0.22	1152.7	0.06
[4,40]	1.2	0.00	1207.2	0.51	1756.1	0.59	1970.6	0.13
[4,50]	1.2	0.00	1927.1	1.21	2669.7	1.38	3017.7	0.28

have a better performance, followed by MBI; however, since Scheme (4.15) requires us to solve matrix singular value problems, it is not as efficient as the other two methods.

To show that the three methods have become stable after reaching the stopping criterion, we run the methods to 10^4 iterations on several instances, and illustrate the results of one instance ($d = 4, n = 30$) in Figure 3. The results are shown by means of plots of error versus iteration, and objective value versus iteration, where "error" means $\|\mathcal{X}^{k+1} - \mathcal{X}^k\|_F$ for Scheme (4.15), and $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\|$ for HOPM and MBI. To give a clearer view, only the results of the first 100 iterations are shown. In this instance, the errors of Scheme (4.15), HOPM, and MBI reach the tolerance 10^{-4} within 13, 7, and 18 iterations, while their associated objective values become stable after 9, 5, and 13 iterations, respectively. In fact, we find that when the errors reach the tolerance, the $\overline{\boldsymbol{x}}$ generated by HOPM and MBI are already stationary points of (4.2).

In this experiment, we use $\mathcal{X}^0 = \tau_{[2]}\tau_{[1]}(\mathcal{A})$ (we denote it as SP, namely sequential projection) as an initialization scheme, and compare it with random initializations, HOSVD [16] and sequential HOSVD [23]. Their corresponding initial value $\langle \mathcal{A}, \mathcal{X}^0 \rangle$ or $\langle \mathcal{A}, x_1^0 \cdots x_d^0 \rangle$ (with all the $\mathcal{X}^0$ or $\boldsymbol{x}^0$ being normalized) and computational time are recorded. Results are reported in Table 4. Empirically, in terms of the value, SP has comparable or better results, while it is also more efficient. The efficiency is because by Proposition 4.8, it only needs to compute two matrix singular value problems, where according to Remark 4.1, the latter one can be further reduced to two much smaller matrix singular value problems. Then, HOPM and MBI are initialized by HOSVD, sequential HOSVD, and SP, respectively. Besides, HOPM with a randomly generated guess served as a baseline. The results are reported in Table 5, which shows that when $d = 3$, MBI initialized by HOSVD is slightly better, while HOPM and MBI initialized by SP slightly outperform the others when $d = 4$. Overall, at least in $d = 3$

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 5
Reports of HOPM and MBI initialized by different guesses.

	Random+ HOPM		HOSVD+ HOPM		seq-HOSVD+ HOPM		SP+ HOPM		HOSVD+ MBI		seq-HOSVD+ MBI		SP+ MBI	
$[d, n]$	value	time	value	time	value	time	value	time	value	time	value	time	value	time
[3,30]	189.6	0.03	208.2	0.02	208.0	0.04	208.0	0.04	208.3	0.17	208.0	0.14	208.0	0.13
[3,40]	273.2	0.03	310.8	0.03	310.4	0.05	310.4	0.05	311.7	0.17	310.4	0.15	310.4	0.13
[3,50]	388.7	0.04	431.2	0.04	431.8	0.07	431.8	0.05	432.0	0.19	431.8	0.16	431.8	0.15
[4,30]	998.5	0.08	1163.3	0.28	1159.7	0.29	1164.0	0.11	1163.5	0.49	1159.7	0.45	1164.0	0.21
[4,40]	1603.0	0.17	1986.1	0.68	1988.2	0.73	1991.7	0.26	1991.1	1.06	1988.2	0.97	1991.7	0.51
[4,50]	2557.0	0.41	3026.1	1.60	3033.0	1.71	3036.4	0.57	3035.5	2.44	3033.0	2.20	3036.4	1.04

TABLE 6
Reports of counts that different methods need to find a global optimizer to problem (4.2). *The order of the tensors is* 4. *For each specified dimension,* 50 *instances with known optimal value have been generated.*

n	Random+ HOPM	HOSVD+ HOPM	seq-HOSVD+ HOPM	SP+ HOPM	Random+ MBI	HOSVD+ MBI	seq-HOSVD+ MBI	SP+ MBI
20	19	36	36	38	19	37	36	38
30	15	36	35	39	13	37	35	39
40	11	36	37	38	12	37	37	38
50	9	33	30	37	10	33	30	37

or $d = 4$ cases, SP can be an alternative way of generating a starting guess. Of course, this procedure becomes expensive when the size goes larger, as it needs to solve larger matrix singular value problems.

Remark 5.1. It is also interesting to examine whether HOPM and MBI with different initializations can find a global optimizer to problem (4.2). In our experiment, this is achieved by running the methods on instances where global optimizers have been found by Model (4.6) solved by Algorithm 3. For each specified dimension, in 50 instances, the counts of different methods achieving a global optimum are reported in Table 6.[1] From the table, we observe that (1) at least on these instances, HOPM and MBI can find a global optimizer in many cases; (2) a good starting guess can indeed improve the chance of finding a global optimizer; (3) empirically, HOPM and MBI initialized by SP, namely, $\tau_{[2]}\tau_{[1]}(\mathcal{A})$ seem to have a slightly better chance of finding a global optimizer.

5.2. The MCP. In this experiment, we try to solve the MCP by relaxing it to Model (4.6). Given an undirected unweighted graph $G = (V, E)$ with vertex set V and edge set $E \in V \times V$, the MCP consists in finding a complete subgraph of G of maximum cardinality $\omega(G)$. The MCP is known as one of the most famous NP-complete problem [25]. It has been proved in [37] that

$$\omega(G) = (1 - 2L_G(x^*))^{-1}, \quad \text{where} \quad L_G(x) := \sum\nolimits_{\{i,j\}\in E} x_i x_j$$

is the Lagrangian of G, and x^* is a global maximizer of $L_G(\cdot)$ over the standard simplex $\Delta = \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 1, x \geq 0\}$. Moreover, if S is a maximum clique of G, then the characteristic vector x^S of S given by

$$x_i^S = \begin{cases} \text{card}(S)^{-1}, & i \in S, \\ 0, & i \notin S, \end{cases} \tag{5.1}$$

[1] Because Scheme (4.15) is not sure to find a stationary point to (4.2), we do not compare it with HOPM and MBI. Nevertheless, we also run the whole Scheme (4.15) with randomly generated initializations on the same instances. We observe that although Scheme (4.15) cannot achieve the global optimum, it can give a good approximation in that the difference between the objective value generated by Scheme (4.15) and the global optimum is less than 10 in up to 70% of the instances.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

TABLE 7
Number of successes on computing the MCP by Model (4.6) *on* 50 *instances.*

[#clique, #edge]	[10,200]	[10,250]	[10,300]	[15,225]	[15,275]	[15,325]	[20,500]	[20,550]	[20,600]
#succ	46	50	43	50	49	46	41	47	46

is a global maximizer of $\max_{x\in\Delta} L_G(x)$. This problem has been formulated as a quartic problem over the unit sphere [6] as follows:

$$\max \sum\nolimits_{\{i,j\}\in E} x_i^2 x_j^2 \ \text{ s.t. } \ \|x\| = 1.$$

The above problem can be relaxed to

$$\max \sum\nolimits_{\{i,j\}\in E} u_i v_i x_j y_j \ \text{ s.t. } \ \|u\| = \|v\| = \|x\| = \|y\| = 1. \tag{5.2}$$

Since (5.2) is of the form (4.2), it can be further relaxed to Model (4.6).

In the experiment, we randomly construct graphs G with 40 vertices containing a maximum clique S, where $\text{card}(S)$ varies in $\{10, 15, 20\}$, and the vertices corresponding to S are known at first; the edges of G varies from 200 to 600. When using Algorithm 3 to solve Model (4.6), $\tau = 0.1$, and we terminate the algorithm within 10 iterations. According to (5.1), during the experiment, we use the following way to identify the maximum clique: Assume that $\hat{\mathcal{X}}$ is returned by Model (4.6); we compute the left singular vector x_1 associated with the leading singular value of $\hat{X}_{[1;2,3,4]}$, and record the largest $\text{card}(S)$ entries of x_1 as the maximum clique. The results are reported in Table 7, showing the number of success over 50 instances, from which we see that Model (4.6) can successfully find the maximum clique in most cases.

5.3. Tensor completion. In this experiment, we incorporate Scheme (4.15) as a subroutine into the higher order relaxed matching pursuit (HoRMP) method [49] for tensor completion (TC). The goal of TC is to recover a possibly low rank tensor from its partial observations. TC finds applications in image/video processing, pattern recognition, and spectral data recovery; see, e.g., [1, 35, 21, 46, 44]. Specifically, we model TC as

$$\min J(\mathcal{X}) = \frac{1}{2}\|\mathcal{X}_\Omega - \mathcal{B}_\Omega\|_F^2 \ \text{ s.t. } \ \text{rank}_{\text{CP}}(\mathcal{X}) \le R, \tag{5.3}$$

where $\mathcal{B}_\Omega$ denotes the partially observed tensor and Ω denotes the index set of observed entries, and $R > 0$ is a parameter. The scheme of HoRMP for solving TC is given by

$$\mathcal{X}^{k+1} = \overline{\alpha}_1\mathcal{X}^k + \overline{\alpha}_2\mathcal{S}^k, \ (\overline{\alpha}_1, \overline{\alpha}_2) = \arg\min\nolimits_{\alpha_1\in\mathbb{R},\alpha_2\in\mathbb{R}} J(\alpha_1\mathcal{X}^k + \alpha_2\mathcal{S}^k),$$

where $\mathcal{S}^k$ is a normalized rank-1 tensor that approximately solves the following problem

$$\max\nolimits_{\|\mathcal{S}\|_F=1,\text{rank}_{\text{CP}}(\mathcal{S})=1} \langle \nabla J(\mathcal{X}^k), \mathcal{S}\rangle, \tag{5.4}$$

which may be approximately solved by Scheme (4.15).

In the following, we denote our method as HoRMP+(4.15) ($\rho = 0.25$ in (4.15)), and compare it with some state-of-the-art TC algorithms: GCG [50], HaLRTC [35], and STDC [12]. To examine the performance of Scheme (4.15), we also compare HoRMP+(4.15) with HoRMP, where the subproblem (5.4) is solved by the subroutine

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 8

Comparisons of different methods on tensor completion on 10 *instances. Due to efficiency, on the dataset Tomato, the max iterations of STDC and HaLRTC are, respectively, set to* 50 *and* 100.

		HoRMP + (4.15)		HoRMP_ori [49]		GCG [50]		HaLRTC [35]		STDC [12]	
Datasets	MR (%)	relerr (%)	time	relerr (%)	time	relerr(%)	time	relerr (%)	time	relerr (%)	time
MRI (181×217×181)	70	1.52	63.80	1.55	51.70	6.02	65.29	0.003	133.55	0.13	85.96
	80	2.09	49.58	2.13	39.97	6.61	46.93	0.29	148.43	0.20	115.82
	90	3.83	31.17	3.94	24.96	7.33	27.70	9.06	160.64	0.42	151.10
	95	8.17	10.42	8.29	7.83	8.67	17.23	34.48	188.61	27.29	124.78
Hyperspectral (205×246×96)	70	1.53	39.06	1.18	52.24	3.36	48.37	2.99	35.53	3.66	443.82
	80	1.77	30.88	1.46	40.03	3.72	34.46	4.23	34.30	6.62	445.66
	90	2.40	21.57	2.25	26.76	3.90	21.97	6.46	56.09	6.73	209.95
	95	3.81	10.38	3.83	7.54	4.39	14.78	10.50	131.89	17.62	91.20
Tomato (242×320×3×167)	70	7.18	286.72	7.19	211.72	10.01	362.53	7.41	327.57	7.57	1195.18
	80	7.83	212.30	7.78	152.94	10.70	264.34	9.63	437.16	8.00	1203.36
	90	8.58	127.36	8.54	87.66	11.40	158.03	13.75	645.45	9.51	1214.15
	95	9.54	75.27	9.56	50.41	11.80	98.16	24.22	632.10	43.20	822.46

used in [49] (to distinguish it, it is denoted as HoRMP_ori). At each iteration, we stop Scheme (4.15) within 5 steps. The stopping criterion for our method is when $|J(\mathcal{X}^{k+1}) - J(\mathcal{X}^k)|$ is less than a threshold. For all the methods except HaLRTC, the threshold of the stopping criteria is $\epsilon = 10^{-5}$. For HaLRTC, 10^{-6} is chosen since 10^{-5} is too loose. The max iteration for all the methods is 500. The parameter R in (5.3) is selected from $\{200, 300, 400, 500\}$ via cross validation. Datasets including brain MRI, Hyperspectral images, and Tomato[2] are used to evaluate the performance of the methods. In the datasets, some entries are randomly missing, with the missing ratio (MR) varying in $\{0.7, 0.8, 0.9, 0.95\}$. All the results are averaged over ten instances.

The relative error $\|\overline{\mathcal{X}} - \mathcal{B}\|_F / \|\mathcal{B}\|_F$ and computational time are reported in Table 8, where the best two results of each setting are marked by underlines. From the table, we can first observe that in most cases, the recovery results of our method are comparable to or better than other methods; even comparing with the original HoRMP, our method has slight improvements, e.g., on the MRI dataset. Considering the efficiency, the two HoRMP methods and GCG perform better, which is more evident on the Tomato dataset; on the Hyperspectral dataset, our method is a bit faster than the original HoRMP. Totally speaking, HoRMP+(4.15) may be a practical method for tensor completion.

5.4. Rank-R approximation. The tensors $\mathcal{A}$ in this experiment are generated as follows: First, we randomly generate a dth order tensor $\mathcal{X}^* = \sum_{i=1}^{R} x_{1,i} \circ \cdots \circ x_{d,i}$. Then we impose a noisy tensor $\mathcal{N}$ to $\mathcal{X}^*$, whose nonzero entries obey the standard normal distribution, with $\frac{\sharp\text{nonzeros entries of } \mathcal{N}}{\sharp\text{entries of } \mathcal{N}}$ = ratio, where ratio is a prescribed number. Finally $\mathcal{A} := \mathcal{X}^* + \mathcal{N}$. The rank R is chosen from $\{5, 7, 9, 12, 15\}$, while the ratio varies among $\{0.2, 0.5, 0.9, 1\}$.

We test Scheme (4.22) and Scheme (4.22)+ALS in this experiment, where the latter one means that Scheme (4.22) is first applied to $\mathcal{A}$, and then ALS is applied to the result generated by Scheme (4.22). There are available MATLAB toolboxes that implement the ALS method [4, 48]. Here we employ the latter one. Our methods are compared with ALS+ELS and ALS+ELS+GEVD, where ALS+ELS stands for the ALS incorporated with the enhanced line search rule [42]. GEVD is an initialization strategy for ALS implemented in [48]. This initialization requires that the first two factor matrices of the tensors have full column rank, otherwise it will fall back to a random initialization. The results are reported in Table 9, where relerr= $\|\overline{\mathcal{X}} - \mathcal{X}^*\|_F / \|\mathcal{X}^*\|_F$. All the results presented are averaged over 50 instances. From

[2]The datasets are available from https://code.google.com/p/tensor-related-code/source/browse/trunk/Model/Tensor+Completion/LRTC_Package_Ji/?r=6#LRTC_Package_Ji%2Fadditional%20results.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 9
Comparisons of different methods on rank-R approximation on 50 *instances.*

		Scheme (4.22)			Scheme (4.22)+ALS			ALS+ELS			ALS+ELS+GEVD		
R	ratio	relerr (%)	iter	time	relerr (%)	iter	time	relerr (%)	iter	time	relerr (%)	iter	time
3rd order tensors of size 50 × 50 × 50													
7	0.5	4.16	10.0	0.26	2.42	10.0	0.25	18.57	6.9	0.14	2.69	2.5	0.17
	0.9	5.66	10.0	0.26	3.31	9.7	0.24	14.40	6.4	0.13	3.75	2.5	0.16
	1	5.89	10.0	0.29	4.00	9.8	0.28	14.60	6.8	0.15	4.10	2.4	0.18
9	0.5	4.68	10.0	0.28	3.02	9.9	0.29	12.95	7.0	0.17	3.53	2.8	0.20
	0.9	6.18	10.0	0.26	3.26	10.1	0.28	13.65	7.3	0.16	3.25	3.2	0.19
	1	6.51	10.0	0.29	3.61	10.3	0.31	16.39	7.3	0.18	3.44	3.4	0.21
12	0.5	5.20	10.0	0.29	2.44	10.0	0.34	12.2	8.0	0.18	3.69	3.3	0.21
	0.9	7.05	10.0	0.32	3.30	10.3	0.39	11.8	8.2	0.22	5.31	3.8	0.25
	1	7.45	10.0	0.29	3.49	11.3	0.36	13.3	7.9	0.18	3.72	4.1	0.23
15	0.5	5.72	10.0	0.32	2.44	11.6	0.44	12.7	8.4	0.2	3.77	3.9	0.23
	0.9	7.72	10.0	0.3	3.29	11.0	0.41	15.02	8.7	0.21	5.72	4.7	0.24
	1	8.16	10.0	0.32	3.46	12.1	0.48	13.55	8.7	0.22	5.18	4.8	0.27
4th order tensors of size 30 × 30 × 30 × 30													
7	0.5	0.93	10.4	1.99	0.88	5.1	0.47	17.72	5.7	1.11	9.81	2.5	1.71
	0.9	1.23	10.3	2.02	1.17	4.9	0.47	18.57	5.8	1.15	8.30	2.5	1.75
	1	1.31	10.2	1.99	1.24	5.0	0.47	17.95	6.4	1.26	8.52	3.1	1.84
9	0.5	0.92	10.9	2.36	0.85	5.5	0.55	17.4	6.6	1.40	8.83	4.4	2.19
	0.9	1.26	10.9	2.56	1.17	6.3	0.66	21.43	6.5	1.51	13.85	4.2	2.35
	1	1.31	10.9	2.35	1.22	6.6	0.61	19.64	6.6	1.40	15.93	4.5	2.19
12	0.5	0.97	11.0	2.88	0.88	6.7	0.7	14.56	7.9	1.77	15.97	6.4	2.76
	0.9	1.28	11.0	2.84	1.16	7.5	0.77	14.98	7.3	1.69	13.40	6.2	2.72
	1	1.35	11.0	2.94	1.22	6.9	0.76	22.19	7.0	1.68	15.14	6.9	3.01
15	0.5	0.98	11.0	3.33	0.87	8.8	0.97	17.67	7.9	2.02	18.82	9.5	3.74
	0.9	1.29	11.0	3.62	1.14	8.5	1.05	16.45	8.0	2.25	17.04	9.2	3.91
	1	1.36	11.0	3.18	1.20	8.1	0.88	18.76	7.8	1.91	18.56	9.0	3.48

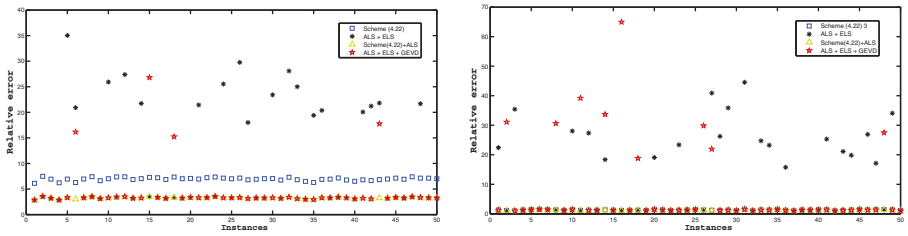

FIG. 4. *Each figure shows* 50 *instances on different* $\mathcal{A}$*'s. The left one presents* $50 \times 50 \times 50$ *tensors with* $R = 12$ *and ratio* $= 0.9$*, while the right one presents* $30 \times 30 \times 30 \times 30$ *tensors, with* $R = 7$ *and ratio* $= 1$*. The y-axis stands for the relative error, while the x-axis denotes runs.*

the table, we first observe that all the algorithms perform efficiently, while ALS-type methods need less time. Considering the relative error, Scheme (4.22) outperforms ALS+ELS. Based upon the results generated by Scheme (4.15), Scheme (4.22)+ALS also performs well, particularly on 4th order tensors. To emphasize the effectiveness of the proposed method, Figure 4 illustrates 50 instances, which shows that the proposed methods may be more stable than other algorithms.

6. Conclusions and remarks. In this paper, the rank-1 equivalence property as well as its applications to tensor optimization problems was studied. It was shown that in the dth order tensor space, the set of rank-1 tensors is the same as the intersection of the $\lceil \log_2(d) \rceil$ set of tensors for which a certain balanced unfolding is rank-1. The number $\lceil \log_2(d) \rceil$ was shown to be optimal as well. Although the equivalence property can be deduced from [8, Theorem 8] (Theorem 3.6), differences should be clarified: The ideas behind the property are proposed, a new proof is presented, and the property can be directly used in tensor optimization problems.

Based on the rank-1 equivalence property, three relaxation approaches were proposed for solving the best rank-1 approximation problem, where the nonconvex relaxation was generalized to solving the rank-R approximation problem when $R > 1$. Numerical experiments demonstrated the effectiveness of the proposed methods.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

From the discussions in section 4.1.2, it is possible to use the dual subgradient ascent [5] to solve the nuclear norm constrained problem (4.6) with the scheme

$$\Lambda_i^{k+1} = \Lambda_i^k + \alpha_i^k(\mathcal{X}^{k+1} - \mathcal{Y}_i^{k+1})$$

with suitably chosen α_i^k, where

$$\mathcal{X}^{k+1} \in \arg\min_{\|\mathcal{X}\|_F \le 1} \left\langle \sum_{i=1}^{f(d)} \Lambda_i^k - \mathcal{A}, \mathcal{X} \right\rangle,\ \mathcal{Y}_i^{k+1} \in \arg\min_{\|Y_{i,[i]}\|_* \le 1} \langle -\Lambda_i^k, \mathcal{Y}_i \rangle.$$

Solving the $\mathcal{Y}_i$-related subproblem can be more efficient than computing a full SVD, however, it is commonly known that the subgradient method is not efficient. Therefore, how to solve (4.8) by taking advantage of the efficiency on solving the subproblems while improving the performance of the subgradient method is a possible direction for future research.

Appendix A. Convergence of the alternating minimization scheme (4.15).

Proof of Proposition 4.7. We have demonstrated that $\{(\mathcal{X}^k, \mathcal{Y}^k)\}_{k=1}^{\infty}$ is bounded. Suppose $\{\mathcal{X}^*, \mathcal{Y}^*\}$ is a limit point and suppose the subsequence $\{\mathcal{X}^{k_l}, \mathcal{Y}^{k_l}\}_{l=1}^{\infty} \to \{\mathcal{X}^*, \mathcal{Y}^*\}$. Denote the best response functions by

$$R(\mathcal{X}^*) \in \arg\min_{\text{rank}(Y_{[1]})=1} F(\mathcal{X}^*, \mathcal{Y}), \quad R(\mathcal{Y}^*) \in \arg\min_{\text{rank}(X_{[2]})=1} F(\mathcal{X}, \mathcal{Y}^*).$$

Then the following two inequalities hold:

$$F(R(\mathcal{Y}^*), \mathcal{Y}^{k_l}) \ge F(\mathcal{X}^{k_l}, \mathcal{Y}^{k_l}),$$
$$F(\mathcal{X}^{k_l}, R(\mathcal{X}^*)) \ge F(\mathcal{X}^{k_l}, \mathcal{Y}^{k_l+1}) \ge F(\mathcal{X}^{k_l+1}, \mathcal{Y}^{k_l+1}) \ge \cdots \ge F(\mathcal{X}^{k_{l+1}}, \mathcal{Y}^{k_{l+1}}).$$

Letting $l \to \infty$ in the above two inequalities gives us that

$$F(R(\mathcal{Y}^*), \mathcal{Y}^*) \ge F(\mathcal{X}^*, \mathcal{Y}^*), \quad F(\mathcal{X}^*, R(\mathcal{X}^*)) \ge F(\mathcal{X}^*, \mathcal{Y}^*).$$

These in connection with the definition of the best response function give the result. ∎

Acknowledgments. We would like to thank the associates editor and two anonymous referees for their insightful comments and suggestions that helped improve this manuscript.

REFERENCES

[1] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, *Scalable tensor factorizations for incomplete data*, Chemometrics Intell. Lab. Syst., 106 (2011), pp. 41–56.
[2] C. M. Andersen and R. Bro, *Practical aspects of PARAFAC modeling of fluorescence excitation-emission data*, J. Chemometrics, 17 (2003), pp. 200–215.
[3] B. W. Bader, M. W. Berry, and M. Browne, *Discussion tracking in enron email using PARAFAC*, in Survey of Text Mining II, Springer, New York, 2008, pp. 147–163.
[4] B. W. Bader, et al., *MATLAB Tensor Toolbox Version* 2.5, http://www.sandia.gov/∼tgkolda/TensorToolbox (2012).
[5] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999.
[6] I. M. Bomze and L. Palagi, *Quartic formulation of standard quadratic optimization problems*, J. Global Optim., 32 (2005), pp. 181–205.
[7] J.-F. Cai, E. J. Candès, and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM J. Optim., 20 (2010), pp. 1956–1982.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[8] E. Carlini and J. Kleppe, *Ranks derived from multilinear maps*, J. Pure Appl. Algebra, 215 (2011), pp. 1999–2004.
[9] J. D. Carroll and J. J. Chang, *Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition*, Psychometrika, 35 (1970), pp. 283–319.
[10] D. Cartwright and B. Sturmfels, *The number of eigenvalues of a tensor*, Linear Algebra Appl., 438 (2011), pp. 942–952.
[11] B. Chen, S. He, Z. Li, and S. Zhang, *Maximum block improvement and polynomial optimization*, SIAM J. Optim., 22 (2012), pp. 87–107.
[12] Y.-L. Chen, C.-T. Hsu, and H.-Y. Liao, *Simultaneous tensor decomposition and completion using factor priors*, IEEE Trans. Pattern Anal. Mach. Intell., 36 (2014), pp. 577–591.
[13] M. Chertok and Y. Keller, *Efficient high order matching*, IEEE Trans. Pattern Anal. Mach. Intell., 32 (2010), pp. 2205–2215.
[14] L. De Lathauwer, *A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 642–666.
[15] L. De Lathauwer and J. Castaing, *Tensor-based techniques for the blind separation of DS–CDMA signals*, Signal Process., 87 (2007), pp. 322–336.
[16] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.
[17] 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 higer-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.
[18] 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., 30 (2008), pp. 1084–1127.
[19] S. Friedland and G. Ottaviani, *The number of singular vector tuples and uniqueness of best rank-one approximation of tensors*, Found. Comput. Math., 14 (2014), pp. 1209–1242.
[20] M. Fumikazu, M.-M. Eduardo, A. V.-S. Pedro, N. Nobuaki, M. Hiroaki, and Y. Yoko, *Decomposing EEG data into space-time-frequency components using parallel factor analysis*, NeuroImage, 22 (2004), pp. 1035–1045.
[21] S. Gandy, B. Recht, and I. Yamada, *Tensor completion and low-n-rank tensor recovery via convex optimization*, Inverse Problems, 27 (2011), 025010.
[22] L. Grasedyck, *Hierarchical singular value decomposition of tensors*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2029–2054.
[23] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, Springer Ser. Comput. Math. 42, Springer, Heidelberg, 2012.
[24] R. A. Harshman, *Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multimodal Factor Analysis*, UCLA Working Papers in Phonetics, 16 (1970), pp. 1–84.
[25] J. Hastad, *Clique is hard to approximate within $n^{1-\epsilon}$*, in Proceedings of the 37th Annual Symposium on Foundations of Computer Science, 1996, IEEE Computer Society, Los Alamitos, CA, 1996, pp. 627–636.
[26] S. He, B. Jiang, Z. Li, and S. Zhang, *Probability bounds for polynomial functions in random variables*, Math. Oper. Res., 39 (2014), pp. 889–907.
[27] S. He, Z. Li, and S. Zhang, *Approximation algorithms for homogeneous polynomial optimization with quadratic constraints*, Math. Program., 125 (2010), pp. 353–383.
[28] C. J. Hillar and L.-H. Lim, *Most tensor problems are NP-hard*, J. ACM, 60 (2013), 45.
[29] B. Jiang, S. Ma, and S. Zhang, *Tensor principal component analysis via convex optimization*, Math. Program., 150 (2014), pp. 423–457.
[30] E. Kofidis and P. A. Regalia, *On the best rank-1 approximation of higher-order supersymmetric tensors*, SIAM J. Matrix Anal. Appl., 23 (2002), pp. 863–884.
[31] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.
[32] T. G. Kolda, B. W. Bader, and J. P. Kenny, *Higher-order web link analysis using multilinear algebra*, in Proceedings of the Fifth IEEE International Conference on Data Mining, IEEE Computer Society, Los Alamitos, CA, 2005, pp. 242–249.
[33] T. G. Kolda and J. R. Mayo, *Shifted power method for computing tensor eigenpairs*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1095–1124.
[34] L.-H. Lim, *Singular values and eigenvalues of tensors: A variational approach*, in Proceedings of the First IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Vol. 1, IEEE, Piscataway, NJ, 2005, pp. 129–132.
[35] J. Liu, P. Musialski, P. Wonka, and J. Ye, *Tensor completion for estimating missing values in visual data*, IEEE Trans. Pattern Anal. Mach. Intell., 35 (2013), pp. 208–220.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[36] Z.-Q. Luo and S. Zhang, *A semidefinite relaxation scheme for multivariate quartic polynomial optimization with quadratic constraints*, SIAM J. Optim., 20 (2010), pp. 1716–1736.
[37] T. S. Motzkin and E. G. Straus, *Maxima for graphs and a new proof of a theorem of Turán*, Canad. J. Math., 17 (1965), pp. 533–540.
[38] C. Mu, B. Huang, J. Wright, and D. Goldfarb, *Square deal: Lower bounds and improved relaxations for tensor recovery*, in Proceedings of the 31st International Conference on Machine Learning (ICML-14), 2014, pp. 73–81.
[39] A.-H. Phan, P. Tichavský, and A. Cichocki, *Low complexity damped Gauss–Newton algorithms for CANDECOMP/PARAFAC*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 126–147.
[40] L. Qi, *Eigenvalues of a real supersymmetric tensor*, J. Symbolic Comput., 40 (2005), pp. 1302–1324.
[41] L. Qi, H.-H. Dai, and D. Han, *Conditions for strong ellipticity and M-eigenvalues*, Front. Math. China, 4 (2009), pp. 349–364.
[42] M. Rajih, P. Comon, and R. A. Harshman, *Enhanced line search: A novel method to accelerate PARAFAC*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1128–1147.
[43] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.
[44] M. Signoretto, Q. T. Dinh, L. De Lathauwer, and J. A. K. Suykens, *Learning with tensors: A framework based on convex optimization and spectral regularization*, Mach. Learn., 94 (2014), pp. 303–351.
[45] A. M.-C. So, *Deterministic approximation algorithms for sphere constrained homogeneous polynomial optimization problems*, Math. Program., 129 (2011), pp. 357–382.
[46] M. Signoretto, R. Van de Plas, B. De Moor, and J. A. K. Suykens, *Tensor versus matrix completion: A comparison with application to spectral data*, IEEE Signal Process. Lett., 18 (2011), pp. 403–406.
[47] S. Soare, J. W. Yoon, and O. Cazacu, *On the use of homogeneous polynomials to develop anisotropic yield functions with applications to sheet forming*, Int. J. Plasticity, 24 (2008), pp. 915–944.
[48] L. Sorber, M. Van Barel, and L. De Lathauwer, *Tensorlab: A MATLAB Toolbox for Tensor Computations*, http://www.tensorlab.net/ (2014).
[49] Y. Yang, S. Mehrkanoon, and J. A. K. Suykens, *Higher Order Matching Pursuit for Low Rank Tensor Learning*, preprint, arXiv:1503.02216, 2015.
[50] Y. Yu, H. Cheng, and X. Zhang, *Approximate low-rank tensor learning*, 32nd NIPS Workshop on Optimization for Machine Learning, JMLR Workshop Conf. Proc. 37, 2014.
[51] T. Zhang and G. H. Golub, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 534–550.
[52] X. Zhang, L. Qi, and Y. Ye, *The cubic spherical optimization problems*, Math. Comp., 81 (2012), pp. 1513–1525.
[53] G. Zhou, L. Caccetta, K. L. Teo, and S.-Y. Wu, *Nonnegative polynomial optimization over unit spheres and convex programming relaxations*, SIAM J. Optim., 22 (2012), pp. 987–1008.

Downloaded 01/28/16 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.