Noname manuscript No.
(will be inserted by the editor)

Rank-1 Approximation for Entangled Multipartite Real Systems

Matthew M. Lin · Moody T. Chu

Received: date / Accepted: date

Abstract The interaction of multiple parts with each other within a system according to certain intrinsic rules is a crucial natural phenomenon. The notion of entanglement and its decomposition of high-dimensional arrays is particularly intriguing since it opens a new way of thinking in data processing and communication, of which the applications will be broad and significant. Depending on how the internal parts engage with each other, there are different types of entanglements with distinct characteristics. This paper concerns the approximation over a multipartite system whose subsystems consist of symmetric rank-1 matrices that are entangled via the Kronecker tensor product. Such a structure resembles that arising in quantum mechanics where a mixed state is to be approximated by its nearest separable state, except that the discussion in this paper is limited to real-valued matrices. Unlike the conventional low-rank tensor approximations, the added twist due to the involvement of the Kronecker product destroys the multilinearity, which makes the problem harder. As a first step, this paper explores the rank-1 multipartite approximation only. Reformulated as a nonlinear eigenvalue problem and a nonlinear singular value problem, respectively, the problem can be tackled numerically by power-like iterative methods and SVD-like iterative methods. The iteration in both classes of methods can be implemented cyclically or acyclically. Motivations, schemes, and convergence theory are discussed in this paper. Preliminary numerical experiments suggest these methods are effective and efficient when compared with some general-purpose optimization packages.

Matthew M. Lin's research was supported in part by the National Center for Theoretical Sciences of Taiwan and by the Ministry of Science and Technology of Taiwan under grant 109-2636-M-006-013.

Moody T. Chu's research was supported in part by the National Science Foundation under grant DMS-1912816.

Matthew M. Lin
Department of Mathematics, National Cheng Kung University, Tainan 701, Taiwan.
E-mail: mhlin@mail.ncku.edu.tw

Moody T. Chu
Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205.
E-mail: chu@math.ncsu.edu

Keywords entanglement · separability · multipartite system · low-rank approximation · nonlinear eigenvalue problem · nonlinear singular value problem

1 Introduction

The problem of matrix approximation via Kronecker product, that is, given a matrix $T \in \mathbb{R}^{m_1 m_2 \times n_1 n_2}$, find $B \in \mathbb{R}^{m_1 \times n_1}$ and $C \in \mathbb{R}^{m_2 \times n_2}$ so that

$$\|T - B \otimes C\|_F \tag{1}$$

is minimized, was probably first discussed by Van Loan and Pitsianis in 1992 at a NATO ASI conference [40]. This problem can be solved effectively by computing the largest singular value and the associated singular vector of a properly permuted version of T, referred to as the $\mathscr{R}$-folding of T or the realignment of T [5]. In [39], Van Loan suggested that the operation by the Kronecker product would have an increasingly greater role to play in the future and demonstrated under the same framework a collection of matrix nearness problems arising from matrix equations, multidimensional quadrature rules, fast transforms, least squares, and semi-definite programming. In the following years, applications to stochastic automata networks [29], micro-array analysis [18,22], image restoration [32,35,45], and computational physics [16] are just another short list added to the Kronecker product approximation.

Given a positive integer k, denote the set $[\![k]\!] := \{1, 2, \ldots, k\}$. The problem (1) can be generalized to the problem of finding $B_r \in \mathbb{R}^{m_1 \times n_1}, C_r \in \mathbb{R}^{m_2 \times n_2}$, $r \in [\![R]\!]$, such that

$$\|T - \sum_{r=1}^{R} B_r \otimes C_r\|_F \tag{2}$$

is minimized, where R is a fixed positive integer no greater than $\min\{m_1 n_1, m_2 n_2\}$. The same singular value decomposition (SVD) technique can be applied and the global minimum is always guaranteed.

The so-called Kronecker SVD technique proposed in [39,40] cannot be applied when the factors B_k and C_k in (2) are required to be structured. Of particular interest is the structure of the form

$$\begin{cases} B_r = \mathbf{x}_r \mathbf{x}_r^\top, & \mathbf{x}_r \in \mathbb{R}^m, \\ C_r = \mathbf{y}_r \mathbf{y}_r^\top, & \mathbf{y}_r \in \mathbb{R}^n, \end{cases} \quad r \in [\![R]\!], \tag{3}$$

that is, the factors are all rank-1 symmetric matrices. Upon normalizing the vectors $\mathbf{x}_r$ and $\mathbf{y}_r$, the problem

$$\min_{\substack{\lambda_r \in \mathbb{R}_+; \mathbf{x}_r \in S^{m-1}; \mathbf{y}_r \in S^{n-1} \\ r \in [\![R]\!]}} \|T - \sum_{r=1}^{R} \lambda_r (\mathbf{x}_r \mathbf{x}_r^\top) \otimes (\mathbf{y}_r \mathbf{y}_r^\top)\|_F^2, \tag{4}$$

where S^{m-1} denotes the unit sphere in $\mathbb{R}^m$, is referred to as a rank-R entangled bipartite approximation to T.

It might be worth mentioning why the structure (3) and the problem (4) are of interest. The motivation can be traced back to the theory of quantum mechanics

whose scope, however, is so broad that it is not possible, nor proper, to provide a general overview in a single discourse. We shall briefly sketch only some elementary ideas here, relating the basic notion of quantum entanglement to the linear algebra setting used in this paper [7]. For readers interested in more formal, in-depth and mathematically oriented discussions, we suggest [1, 23, 33] and the classic book [34]. A quantum mechanical system is generally cast as a complex Hilbert space [1]. Any unit vector in the space is referred to as a pure state which typically is denoted by the Dirac's ket notation $|\mathbf{w}\rangle$. For practical reasons, it is more convenient to represent a pure state $|\mathbf{w}\rangle$ in the projector form $|\mathbf{w}\rangle\langle\mathbf{w}|$, referred to as a density matrix. With respect to a specified basis, a finitely dimensional quantum system can be identified as $\mathbb{C}^m$, where we can write the state $|\mathbf{w}\rangle$ as a unit vector $\mathbf{w} \in \mathbb{C}^m$, and the corresponding density matrix as the rank-1 matrix $\mathbf{w}\mathbf{w}^*$ with * standing for the conjugate transpose. A mixed state, typically denoted by ρ in the literature, is a probabilistic ensemble of density matrices of pure states in that system, which therefore is a positive semi-definite matrix with unit trace.

A k-partite quantum system consists of k interacting quantum subsystems. With respect to properly specified bases among the subsystems, the intersection can be described mathematically via the Kronecker product. We say that a mixed state ρ in the k-partite system is separable if ρ can be expressed as a finite sum in the form

$$\rho = \sum_{r=1}^{R} \mu_r (|\mathbf{w}_1^{(r)}\rangle\langle\mathbf{w}_1^{(r)}|) \otimes \ldots \otimes (|\mathbf{w}_k^{(r)}\rangle\langle\mathbf{w}_k^{(r)}|), \tag{5}$$

where $|\mathbf{w}_j^{(r)}\rangle$, $r \in [\![R]\!]$, are pure states in the j-th subsystems, $j \in [\![k]\!]$, and

$$\mu_r \geq 0, \quad \sum_{r=1}^{R} \mu_r = 1. \tag{6}$$

There is no restriction on the value of R. When ρ is not separable, we say that it is entangled. Quantum entanglement plays an increasingly more important role in modern quantum technologies. Quantum informatics and quantum communication, for example, exploit the entanglement for faster and more secure delivery of information than classical algorithms.

Determining whether a given mixed state in a k-partite system is entangled or not is NP hard [15, 21]. On the other hand, note that separable states form a convex compact subset in the ambient space. Approximating a given mixed state ρ with the nearest separable state is a problem of interest in its own right. Depending on what statistical properties are to be quantified, the nearness can be measured under different metrics [6].

The above notion can be expressed in terms of the classical linear algebra notations. Observe that the tensor product of complex vectors

$$(\mathbf{u} + \imath\mathbf{v}) \otimes (\mathbf{p} + \imath\mathbf{q}) = (\mathbf{u} \otimes \mathbf{p} - \mathbf{v} \otimes \mathbf{q}) + \imath(\mathbf{v} \otimes \mathbf{p} + \mathbf{u} \otimes \mathbf{q}),$$

involves a nontrivial intertwinement between the real and the imaginary parts of the variables. Suppose that, as a first step, we limit ourselves to the real values. Then the structure of the rank-R k-partite problem

$$\min_{\substack{\lambda_r \in \mathbb{R}_+, \mathbf{x}_i^{(r)} \in S^{I_i - 1}, \\ i \in [\![k]\!], r \in [\![R]\!]}} \|T - \sum_{r=1}^{R} \lambda_r (\mathbf{x}_1^{(r)} {\mathbf{x}_1^{(r)}}^\top) \otimes \cdots \otimes (\mathbf{x}_k^{(r)} {\mathbf{x}_k^{(r)}}^\top)\|_F^2 \tag{7}$$

for a given matrix T resembles the fabric (5) underlying the quantum entanglement, where the Frobenius norm is used.

In an earlier paper [7], we have already studied thoroughly the rank-1 approximation of an entangled bipartite system. In particular, by casting the approximation as a nonlinear eigenvalue problem and a nonlinear singular value problem, we have developed numerical algorithms and accomplished the convergence analysis. The goal of this paper is to take into account another aspect of complexity when more than two interacting subsystems are involved. That is, we shall consider the rank-1 approximation of a k-partite system

$$\min_{\lambda\in\mathbb{R}_+,\, \mathbf{x}_i\in S^{I_i-1},\, i\in[\![k]\!]} \|T-\lambda(\mathbf{x}_1\mathbf{x}_1^\top)\otimes\cdots\otimes(\mathbf{x}_k\mathbf{x}_k^\top)\|_F^2, \tag{8}$$

where $k\geq 2$, the dimensions I_i, $i\in[\![k]\!]$, are preselected, and $T\in\mathbb{R}^{\prod_{i=1}^k I_i\times\prod_{i=1}^k I_i}$ is a given symmetric and positive definite (SPD) matrix. A generalization of both the numerical methods and the convergence analysis from the rank-1 approximation of a bipartite system to the rank-1 approximation of a k-partite problem (4) is not trivial because of the many more factors involved in the tensor product. Our main contribution in this paper is to fill that gap.

To demonstrate the difficulty of this generalization, observe that with a proper folding of T into an order-4 tensor $\mathfrak{T}$ [39,40], the bipartite problem can be treated as a specially structured rank-1 tensor approximation, known as the canonical polyadic decomposition with symmetry:

$$\min_{\substack{\lambda\in\mathbb{R}_+,\mathbf{x}\in\mathbb{R}^m,\mathbf{y}\in\mathbb{R}^n\\ \|\mathbf{x}\|=1,\|\mathbf{y}\|=1}} \|\mathfrak{T}-\lambda\,\mathbf{x}\circ\mathbf{x}\circ\mathbf{y}\circ\mathbf{y}\|_F^2, \tag{9}$$

where $\circ$ denotes the outer product. Many techniques, e.g., those in the Tensorlab toolbox [41], are readily available to handle (9), albeit some disadvantages of such a formulation when compared with our methods [7]. The sticking point when generalizing this idea to the case $k>2$ is that, to our knowledge, there is no explicit strategy for folding the given T into an order-$2k$ tensor. This difficulty is in line with the fact that thus far the Kronecker SVD technique for (2) has not been successfully generalized when the summation involves terms with more than two factors in the tensor product.

We must stress also that the optimization problem (8) should not be confused with the general low-rank tensor approximation that had been discussed extensively in recent years [3,8,10,24,26–28,38,43,44]. The latter involves tensor products of vectors and is multilinear in its factors, but our problem involves Kronecker products of rank-1 matrices and is not multilinear at all. Existing techniques for conventional tensor decomposition are inadequate to handle this structured problem. We thus propose in this paper two new methods for this rank-1 multipartite approximation (8). Our main focus is on addressing the multi-indices effectively, proposing the numerical algorithms and proving the global convergence.

Finally, to prepare our presentation, we point out that the minimization of (8) is equivalent to the maximization of the orthogonal component of T in the direction of the unit "vector" $(\mathbf{x}_1\mathbf{x}_1^\top)\otimes(\mathbf{x}_2\mathbf{x}_2^\top)\otimes\cdots\otimes(\mathbf{x}_k\mathbf{x}_k^\top)$, i.e., we may consider the problem

$$\max_{\substack{\mathbf{x}_i\in S^{I_i-1}\\ i\in[\![k]\!]}} \lambda(\mathbf{x}_1,\ldots,\mathbf{x}_k), \tag{10}$$

where

$$\lambda = \lambda(\mathbf{x}_1, \ldots, \mathbf{x}_k) := \langle T, (\mathbf{x}_1\mathbf{x}_1^\top) \otimes (\mathbf{x}_2\mathbf{x}_2^\top) \otimes \cdots \otimes (\mathbf{x}_k\mathbf{x}_k^\top)\rangle \tag{11}$$

and $\langle \cdot, \cdot \rangle$ stands for the Frobenius inner product of real matrices.

This paper is organized as follows. We begin in Section 2 by introducing a convenient notation system which will help circumvent the otherwise tedious multi-indexed descriptions. In Section 3, we reformulate the problem (10) as a nonlinear eigenvalue problem and propose a power-like iterative scheme. Despite the simplicity in its appearance, the iteration is inherently nonlinear. Our first contribution in this regard is the convergence analysis. The implication of multi-indices also allows us to formulate the problem as a nonlinear singular value problem in Section 4. The proposed SVD-like iterative scheme updates two vectors at a time. Our second contribution is the proof that, regardless of the order of updating, convergence can be achieved. A comparison of our two methods with some existing routines in Matlab Optimization Toolbox is given in Section 5.

2 Basic notation

To enumerate the data in a multi-array consistently, we adopt the practice of counting the multi-indexed entry $\tau_{i_1,\ldots,i_k}$ of an order-k tensor $\mathfrak{T} \in \mathbb{R}^{J_1\times J_2\times\ldots\times J_k}$ as the I-the entry in the corresponding linear array, where

$$I := (i_k - 1)J_{k-1}J_{k-2}\ldots J_1 + (i_{k-1} - 1)J_{k-2}\ldots J_1 + \ldots + (i_2 - 1)J_1 + i_1. \tag{12}$$

In this way, we say that the tensor $\mathfrak{T}$ is vectorized by the operation vec. Therefore, the classical Kronecker product $\otimes$ of column vectors, resulting in a long vector, is related to the tensor product $\circ$ which produces multi-indexed tensor in a reversed order, i.e.,

$$\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \ldots \otimes \mathbf{x}_k = \mathsf{vec}(\mathbf{x}_k \circ \ldots \circ \mathbf{x}_2 \circ \mathbf{x}_1).$$

So as to discuss our algorithms for the k-partite system in a more concise way, we adopt the following notations. First, we introduce the abbreviations

$$\begin{cases} \bigotimes_{i=1}^{k} \mathbf{x}_i := \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \cdots \otimes \mathbf{x}_k, \\ \bigcirc_{i=k}^{1} \mathbf{x}_i := \mathbf{x}_k \circ \ldots \circ \mathbf{x}_1, \end{cases}$$

and define the order-k tensor

$$\mathscr{D}(\mathbf{x}_1, \ldots, \mathbf{x}_k) := \mathsf{reshape}(T \bigotimes_{i=1}^{k} \mathbf{x}_i, [I_k, \ldots, I_1]) \in \mathbb{R}^{I_k\times\ldots\times I_1}, \tag{13}$$

where the operation reshape imitates the same command in Matlab that folds data into a multi-dimensional array according to the enumeration rule (12). Second, given a fixed partition $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$ with $\boldsymbol{\alpha} := \{\alpha_1, \ldots, \alpha_\ell\}$ and $\boldsymbol{\beta} := \{\beta_1, \ldots, \beta_{k-\ell}\}$, let $\mathcal{I} = (i_1, \ldots, i_\ell)$ and $\mathcal{J} = (j_1, \ldots, j_{k-\ell})$ denote the multi-indices at locations $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively.

We shall regard an order-k tensor $\mathfrak{T} = [\tau_{s_1\ldots s_k}] \in \mathbb{R}^{J_1\times J_2\times\ldots\times J_k}$ as the matrix representation of the linear operator $\mathscr{T}_{\boldsymbol{\alpha}}$

$$\mathscr{T}_{\boldsymbol{\alpha}} : \mathbb{R}^{J_{\beta_1}\times\ldots\times J_{\beta_{k-\ell}}} \to \mathbb{R}^{J_{\alpha_1}\times\ldots\times J_{\alpha_\ell}} \tag{14}$$

in the sense that, given any $U = [u_{j_1,\ldots,j_{k-\ell}}] \in \mathbb{R}^{J_{\beta_1}\times J_{\beta_2}\times\ldots\times J_{\beta_{k-\ell}}}$, the $\mathcal{I}$-th entry of its image $\mathscr{T}_{\boldsymbol{\alpha}}(U)$ is given by

$$\begin{aligned}\big(\mathscr{T}_{\boldsymbol{\alpha}}(U)\big)_{\mathcal{I}} &= (\mathfrak{T}\circledast_{\boldsymbol{\alpha}} U)_{\mathcal{I}} \\ &:= \sum_{\mathcal{J}} \tau^{(\boldsymbol{\alpha},\boldsymbol{\beta})}_{[\mathcal{I}|\mathcal{J}]} u_{\mathcal{J}} = \sum_{j_1=1}^{J_{\beta_1}} \cdots \sum_{j_{k-\ell}=1}^{J_{\beta_{k-\ell}}} \tau^{(\boldsymbol{\alpha},\boldsymbol{\beta})}_{[\mathcal{I}|j_1\ldots j_{k-\ell}]} u_{j_1,\ldots,j_{k-\ell}} \in \mathbb{R}^{J_{\alpha_1}\times\ldots\times J_{\alpha_\ell}},\end{aligned}$$

where the symbol $\tau^{(\boldsymbol{\alpha},\boldsymbol{\beta})}_{[\mathcal{I}|\mathcal{J}]}$ represents the entry $\tau_{s_1\ldots s_k}$ of $\mathfrak{T}$ with $s_{\alpha_\mu} = i_\mu$ and $s_{\beta_\nu} = j_\nu$, $\mu \in [\![\ell]\!]$, $\nu \in [\![k-\ell]\!]$. In this way, the subsets $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ generalize the notion of rows and columns, respectively.

Using the associative law of multiplication, it can be seen that

$$\langle \mathfrak{T}, \bigcirc_{i=1}^{k} \mathbf{u}^{(i)} \rangle = \langle \mathfrak{T}\circledast_{\boldsymbol{\alpha}} (\bigcirc_{s=1}^{k-\ell} \mathbf{u}^{(\beta_s)}), \bigcirc_{t=1}^{\ell} \mathbf{u}^{(\alpha_t)} \rangle \tag{15}$$

for any tensors $\bigcirc_{t=1}^{\ell} \mathbf{u}^{(\alpha_t)} \in \mathbb{R}^{J_{\alpha_1}\times\ldots\times J_{\alpha_\ell}}$ and $\bigcirc_{s=1}^{k-\ell} \mathbf{u}^{(\beta_s)} \in \mathbb{R}^{J_{\beta_1}\times J_{\beta_2}\times\ldots\times J_{\beta_{k-\ell}}}$. The relationship (15) can be interpreted as a generalization of the adjoint equation. Such a notation system offers the convenience that by merely specifying the elements in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we have a clear indication of which part of the tensor is to be multiplied with another tensor.

3 Nonlinear eigenvalue formulation

The following first order optimality condition for $\lambda(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ is easy to derive.

Lemma 1 *The first order optimality condition for maximizing $\lambda(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_\ell)$ subject to $\mathbf{x}_j \in S^{I_j-1}$, $j \in [\![k]\!]$, is that*

$$\mathscr{D}(\mathbf{x}_1,\ldots,\mathbf{x}_k)\circledast_j \left(\bigcirc_{i=k,i\neq j}^{1} \mathbf{x}_i \right) = \lambda(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k)\mathbf{x}_j, \quad j \in [\![k]\!], \tag{16}$$

where the multiplication $\circledast_j$ is to sum over the location indicators $\boldsymbol{\beta} = [\![k]\!]\backslash\{j\}$.

The system of equations (16) can be regarded as a nonlinear eigenvalue problem which thus motivates a power-like iterative scheme:

$$\mathbf{x}_j^{[p+1]} := \frac{\mathscr{D}(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})\circledast_j \bigcirc_{i=k,i\neq j}^{1} \mathbf{x}_i^{[p]}}{\|\mathscr{D}(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})\circledast_j \bigcirc_{i=k,i\neq j}^{1} \mathbf{x}_i^{[p]}\|_2}, \quad j \in [\![k]\!], \quad p = 0,1,\ldots. \tag{17}$$

If $\mathscr{D}(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})$ were invariant in p, the scheme would be a Jacobi version [24] of the conventional ALS method for rank-1 tensor approximation [3,8,26–28,44] whose convergence theory is well established [38,43]. In our case, $\mathscr{D}(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})$ does vary in p. Therefore, the analysis of its dynamical behavior will be more involved.

3.1 Convergence of λ values

Define a functional $G : \mathbb{R}^{I_1} \times \mathbb{R}^{I_1} \times \cdots \times \mathbb{R}^{I_k} \times \mathbb{R}^{I_k} \to \mathbb{R}$ by

$$G(\mathbf{x}_1, \widetilde{\mathbf{x}}_1; \ldots; \mathbf{x}_k, \widetilde{\mathbf{x}}_k) := \langle T \bigotimes_{i=1}^{k} \mathbf{x}_i, \bigotimes_{i=1}^{k} \widetilde{\mathbf{x}}_i \rangle. \tag{18}$$

Since T is a real and symmetric, it is clear that

$$G(\mathbf{x}_1, \widetilde{\mathbf{x}}_1; \ldots; \mathbf{x}_k, \widetilde{\mathbf{x}}_k) = G(\widetilde{\mathbf{x}}_1, \mathbf{x}_1; \ldots; \widetilde{\mathbf{x}}_k, \mathbf{x}_k). \tag{19}$$

Note also that

$$G(\mathbf{x}_1, \mathbf{x}_1; \ldots; \mathbf{x}_k, \mathbf{x}_k) = \lambda(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k). \tag{20}$$

We now argue that the sequence $\{\lambda(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})\}$ generated by the scheme (17) is monotone.

Theorem 1 *Assume that T is SPD. Let $\{(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})\}$ be the sequence generated by the scheme* (17)*. Then the functional G satisfies the interlacing property:*

$$\begin{aligned} &G(\mathbf{x}_1^{[p]}, \mathbf{x}_1^{[p]}; \mathbf{x}_2^{[p]}, \mathbf{x}_2^{[p]}; \ldots; \mathbf{x}_k^{[p]}, \mathbf{x}_k^{[p]}) \le G(\mathbf{x}_1^{[p]}, \mathbf{x}_1^{[p+1]}; \mathbf{x}_2^{[p]}, \mathbf{x}_2^{[p]}; \ldots; \mathbf{x}_k^{[p]}, \mathbf{x}_k^{[p]}) \\ &\le G(\mathbf{x}_1^{[p+1]}, \mathbf{x}_1^{[p+1]}; \mathbf{x}_2^{[p]}, \mathbf{x}_2^{[p]}; \ldots; \mathbf{x}_k^{[p]}, \mathbf{x}_k^{[p]}) \le G(\mathbf{x}_1^{[p+1]}, \mathbf{x}_1^{[p+1]}; \mathbf{x}_2^{[p]}, \mathbf{x}_2^{[p+1]}; \ldots; \mathbf{x}_k^{[p]}, \mathbf{x}_k^{[p]}) \\ &\le \cdots \le G(\mathbf{x}_1^{[p+1]}, \mathbf{x}_1^{[p+1]}; \ldots; \mathbf{x}_k^{[p+1]}, \mathbf{x}_k^{[p+1]}). \end{aligned} \tag{21}$$

Therefore, the sequence $\{\lambda(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})\}$ converges.

Proof To prove the first inequality, observe that by (20) and the adjoint equation (15), we can write

$$G(\mathbf{x}_1^{[p]}, \mathbf{x}_1^{[p]}; \mathbf{x}_2^{[p]}, \mathbf{x}_2^{[p]}; \ldots; \mathbf{x}_k^{[p]}, \mathbf{x}_k^{[p]}) = \langle \mathscr{D}(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]}) \circledast_1 (\bigcirc_{i=k}^{2} \mathbf{x}_i^{[p]}), \mathbf{x}_1^{[p]} \rangle.$$

We can also rewrite

$$\begin{aligned} G(\mathbf{x}_1^{[p]}, \mathbf{x}_1^{[p+1]}; \mathbf{x}_2^{[p]}, \mathbf{x}_2^{[p]}; \ldots; \mathbf{x}_k^{[p]}, \mathbf{x}_k^{[p]}) &= \langle \mathscr{D}(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]}) \circledast_1 (\bigcirc_{i=k}^{2} \mathbf{x}_i^{[p]}), \mathbf{x}_1^{[p+1]} \rangle \\ &= \| \mathscr{D}(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]}) \circledast_1 (\bigcirc_{i=k}^{2} \mathbf{x}_i^{[p]}) \|_2 \end{aligned}$$

where the second equality follows from the definition of $\mathbf{x}_1^{[p+1]}$. Upon comparison, the first inequality follows from the Cauchy-Schwarz inequality.

To prove the second inequality, define

$$\Delta\mathbf{x}_i^{[p]} := \mathbf{x}_i^{[p+1]} - \mathbf{x}_i^{[p]}, \quad i \in [\![k]\!]. \tag{22}$$

Then we can break down the difference as

$$\begin{aligned} &G(\mathbf{x}_1^{[p+1]}, \mathbf{x}_1^{[p+1]}; \mathbf{x}_2^{[p]}, \mathbf{x}_2^{[p]}; \ldots; \mathbf{x}_k^{[p]}, \mathbf{x}_k^{[p]}) - G(\mathbf{x}_1^{[p]}, \mathbf{x}_1^{[p+1]}; \ldots; \mathbf{x}_k^{[p]}, \mathbf{x}_k^{[p]}) \\ &= G(\Delta\mathbf{x}_1^{[p]}, \mathbf{x}_1^{[p+1]}; \mathbf{x}_2^{[p]}, \mathbf{x}_2^{[p]}; \ldots; \mathbf{x}_k^{[p]}, \mathbf{x}_k^{[p]}) \\ &= G(\Delta\mathbf{x}_1^{[p]}, \Delta\mathbf{x}_1^{[p]}; \mathbf{x}_2^{[p]}, \mathbf{x}_2^{[p]}; \ldots; \mathbf{x}_k^{[p]}, \mathbf{x}_k^{[p]}) + G(\Delta\mathbf{x}_1^{[p]}, \mathbf{x}_1^{[p]}; \mathbf{x}_2^{[p]}, \mathbf{x}_2^{[p]}; \ldots; \mathbf{x}_k^{[p]}, \mathbf{x}_k^{[p]}). \end{aligned}$$

The first term on the right side in the last equality is positive because T is SPD. The second term is nonnegative because of the first inequality and the symmetric property (19). The other inequalities can be argued in a similar manner. Since $\{\lambda(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})\}$ is bounded, the sequence converges. □

Thus far, we update $\mathbf{x}_j^{[p]}$ by sweeping $j \in [\![k]\!]$ in the cyclic order and do not update $\mathscr{D}(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})$ until one sweep is over. One possible variant of (17) is to select an integer $j \in [\![k]\!]$ randomly and update the factor $\mathbf{x}_j$ according to the right side of (17), except that only the latest updates of the remaining factors $\mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \mathbf{x}_k$ are used on the right side of (17). This amounts to an asynchronous version of the Gauss-Seidel update [36]. A similar approach applied to the best rank-1 tensor approximation can be found in [20].

More specifically, the scheme (17) for the Gauss-Seidel-type update should be interpreted as follows: Suppose that an index j has been chosen so that $\boldsymbol{\alpha} = \{j\}$, $\boldsymbol{\beta} = [\![k]\!] \backslash \{j\}$) and that the factor $\mathbf{x}_{\alpha_1}^{[p]}$ has been updated to $\mathbf{x}_{\alpha_1}^{[p+1]}$. We immediately recognize $(\mathbf{x}_1^{[p+1]}, \ldots, \mathbf{x}_k^{[p+1]})$ as an update from $(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})$, in which $\mathbf{x}_{\beta_i}^{[p+1]} = \mathbf{x}_{\beta_i}^{[p]}$, $i \in [\![k-1]\!]$ and only $\mathbf{x}_{\alpha_1}^{[p+1]}$ is truly updated. In this way, the variables in the definition of $\mathscr{D}(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})$ are always the most updated ones.

A convergence analysis of this updating scheme in a haphazard order might seem daunting. However, a close examination of the proof used for Theorem 1 shows that the cyclic order is never required. All we need is a fixed integer $j \in [\![k]\!]$ for which $\mathbf{x}_j^{[p]}$ is to be updated to $\mathbf{x}_j^{[p+1]}$ while all other variables are assumed known and stay invariant during the execution of (17). When this is done, the superscripts for all variables are renamed from $^{[p]}$ to $^{[p+1]}$ and we continue the process to a new factor with another randomly selected integer. The iteration exhibits the same interlacing property. The fact that the iteration still maintains convergence under any order of updating strategy is remarkable. For completion, we summarize the result below.

Corollary 1 *Assume that T is SPD. Let the sequence $\{\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]}\}$ be generated by (17) in any acyclic order. Then the variational relationship* (21) *still holds and the sequence $\{\lambda(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})\}$ converges.*

3.2 Convergence of iterates

In addition to the convergence of the objective values, we now argue that under some mild conditions the iterates $\{(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})\}$ also converge.

We first mention the following result from real analysis. The proof can be found in [30, Lemma 4.10]. A refined version is given in [20, Lemma 2.7].

Lemma 2 *Let $\mathbf{z}^*$ be an isolated limit point of the sequence $\{\mathbf{z}^{[p]}\} \subset \mathbb{R}^n$. Assume that for every subsequence $\{\mathbf{z}^{[p_j]}\}$ converging to $\mathbf{z}^*$, $\|\mathbf{z}^{[p_j+1]} - \mathbf{z}^{[p_j]}\|_2 \to 0$ as $p_j \to \infty$. Then $\{\mathbf{z}^{[p]}\}$ converges.*

We next examine the increment between two consecutive iterates in our algorithm.

Lemma 3 *Suppose that T is SPD and that $\{\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]}\}$ is generated by (17) in either cyclic or acyclic order. Then, for each $i\in[\![k]\!]$, the sequence $\{\Delta\mathbf{x}_i^{[p]}\}$ defined in* (22) *converges to zero.*

Proof Without loss of generality, we consider the case when the sequence $\mathbf{x}_i$ is updated in a cyclic order. The following two equations are obvious:

$$\begin{aligned}&G(\mathbf{x}_1^{[p]},\mathbf{x}_1^{[p+1]};\mathbf{x}_2^{[p]},\mathbf{x}_2^{[p]};\ldots;\mathbf{x}_k^{[p]},\mathbf{x}_k^{[p]})-G(\mathbf{x}_1^{[p]},\mathbf{x}_1^{[p]};\mathbf{x}_2^{[p]},\mathbf{x}_2^{[p]};\ldots;\mathbf{x}_k^{[p]},\mathbf{x}_k^{[p]})\\ &=\langle T(\Delta\mathbf{x}_1^{[p]}\otimes\bigotimes_{i=2}^{k}\mathbf{x}_i^{[p]}),\mathbf{x}_1^{[p]}\otimes\bigotimes_{i=2}^{k}\mathbf{x}_i^{[p]}\rangle, \qquad (23)\\ &G(\mathbf{x}_1^{[p+1]},\mathbf{x}_1^{[p+1]};\mathbf{x}_2^{[p]},\mathbf{x}_2^{[p]};\ldots;\mathbf{x}_k^{[p]},\mathbf{x}_k^{[p]})-G(\mathbf{x}_1^{[p]},\mathbf{x}_1^{[p+1]};\mathbf{x}_2^{[p]},\mathbf{x}_2^{[p]};\ldots;\mathbf{x}_k^{[p]},\mathbf{x}_k^{[p]})\\ &=\langle T(\Delta\mathbf{x}_1^{[p]}\otimes\bigotimes_{i=2}^{k}\mathbf{x}_i^{[p]}),\mathbf{x}_1^{[p+1]}\otimes\bigotimes_{i=2}^{k}\mathbf{x}_i^{[p]}\rangle. \qquad (24)\end{aligned}$$

Taking the difference of (23) and (24) and applying the fact that the sequence $\{\lambda(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})\}$ converges, we have

$$\langle T(\Delta\mathbf{x}_1^{[p]}\otimes\bigotimes_{i=2}^{k}\mathbf{x}_i^{[p]}),\Delta\mathbf{x}_1^{[p]}\otimes\bigotimes_{i=2}^{k}\mathbf{x}_i^{[p]}\rangle\to 0.$$

Since all $\mathbf{x}_i^{[p]}$, $i\in[\![k]\!]\backslash\{1\}$ are of unit length, by the positive definiteness of T, it must be that $\Delta\mathbf{x}_1^{[p]}\to 0$. We can carry out a similar argument to show that $\Delta\mathbf{x}_i^{[p]}\to 0$ for $i\in[\![k]\!]\backslash\{1\}$. □

To make use of Lemma 2 for proving convergence, we need to ensure that any limit point is isolated. Toward this end, observer that by continuity any accumulation point of the sequence $\{(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})\}$ must satisfy the system of equations (16) which actually is a polynomial system in the variables $x_1,\cdots,x_k$ with T as the parameter. An algebraic geometry argument can be used to help complete the convergence proof. To begin with, it is known that almost all square systems of polynomial equations over the complex field have finitely many solutions [14]. More specifically, suppose that $F(\mathbf{z};\mathbf{q})$ is a square polynomial system in the variables $\mathbf{z}$ and the parameters $\mathbf{q}$, then the theory on parameter continuation for polynomial systems implies that the number of isolated solutions to this polynomial system is finite for almost all parameters $\mathbf{q}$ [37, Theorem 7.1.1]. The phrase "almost all" means that pathological examples (for the parameter T) can be constructed to negate the assertion. However, those values of parameters that fail to induce finitely many and geometrically isolated solutions form a nowhere dense and measure zero subset in the ambient space. The term "non-generic" is often used to describe collectively this type of special cases. Because of this observation, we conclude that the following condition for the matrix T may be regarded as generic.

Condition P: We say that the matrix T satisfies Condition P if the corresponding polynomial system (16) has finitely many and geometrically isolated real-valued solutions.

Theorem 2 *Assume that T is SPD and satisfies Condition P. Then the sequence $\{(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})\}$ generated by the scheme* (17) *in any order converges to a single limit point which satisfies the system* (16).

Proof Under Condition P, we know that the sequence $\{(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})\}$ has only a finite number of isolated limit points. By Lemma 3, every convergent subsequence $\{(\mathbf{x}_1^{[p_j]}, \ldots, \mathbf{x}_k^{[p_j]})\}$ of $\{(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})\}$ will have diminishing increments. By Lemma 2, we see that the overall converges. □

4 Nonlinear singular value formulation

Thus far, in order to find a solution to the system (16), we have been updating one vector variable at a time. The scheme (17) resembles the conventional power method but is nonlinear in its variables. In this section, we describe another iterative scheme with the goal to update two vectors simultaneously. Toward this goal, we reformulate the approximation problem as a nonlinear singular value decomposition.

4.1 Simultaneous updates via SVD

The first order optimality condition Lemma 1 can be expressed differently as follows.

Lemma 4 *Let* $\boldsymbol{\alpha} := \{\alpha_1, \alpha_2\}$ *and* $\boldsymbol{\beta} := \{\beta_1, \ldots, \beta_{k-2}\}$ *be an arbitrary partition of* $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$. *Then the necessary condition of a local maximizer for* $\lambda(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ *is that*

$$\begin{cases} (\mathscr{D}(\mathbf{x}_1, \ldots, \mathbf{x}_k) \circledast_{\boldsymbol{\alpha}} (\bigcirc_{i=1}^{k-2} \mathbf{x}_{\beta_i})) \mathbf{x}_{\alpha_2} & = \lambda(\mathbf{x}_1, \ldots, \mathbf{x}_k) \mathbf{x}_{\alpha_1}, \\ (\mathscr{D}(\mathbf{x}_1, \ldots, \mathbf{x}_k) \circledast_{\boldsymbol{\alpha}} (\bigcirc_{i=1}^{k-2} \mathbf{x}_{\beta_i}))^\top \mathbf{x}_{\alpha_1} & = \lambda(\mathbf{x}_1, \ldots, \mathbf{x}_k) \mathbf{x}_{\alpha_2}. \end{cases} \tag{25}$$

We may interpret $(\mathbf{x}_{\alpha_1}, \lambda(\mathbf{x}_1, \ldots, \mathbf{x}_k), \mathbf{x}_{\alpha_2})$ as a singular triplet of the matrix $\mathscr{D}(\mathbf{x}_1, \ldots, \mathbf{x}_k) \circledast_{\boldsymbol{\alpha}} (\bigcirc_{i=1}^{k-2} \mathbf{x}_{\beta_i})$. Since our goal is to maximize the $\lambda(\mathbf{x}_1, \ldots, \mathbf{x}_k)$, we may as well search for the dominant singular triplet. We thus propose the scheme that, while varying $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2\}$ through $[\![k]\!]$ to select vectors to be updated, repeat the fixed-point iteration:

$$\begin{aligned} &\left(\operatorname{sgn}\left(x_1^{[p+1]}\right) \mathbf{x}_{\alpha_1}^{[p+1]}, \lambda_{\boldsymbol{\alpha}}^{[p+1]}, \operatorname{sgn}\left(x_1^{[p+1]}\right) \mathbf{x}_{\alpha_2}^{[p+1]} \right) \\ &= \mathsf{svds}(\mathscr{D}(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]}) \circledast_{\boldsymbol{\alpha}} \left(\bigcirc_{i=1}^{k-2} \mathbf{x}_{\beta_i}^{[p]} \right), 1), \end{aligned} \tag{26}$$

where the sign of the first entry $x_1^{[p+1]}$ of the vector $\mathbf{x}_{\alpha_1}^{[p+1]}$ is used to maintain the continuity and svds denotes any route that computes the first dominant singular triplet. If $x_1^{[p+1]} = 0$, then we choose the next nonzero entry of $\mathbf{x}_{\alpha_1}^{[p+1]}$. Since $\mathscr{D}(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})$ varies in p, this nonlinear singular value decomposition approach is of theoretical interest in its own right.

There are several variations worth mentioning. Computing $\mathscr{D}(\mathbf{x}_1^{[p]}, \ldots, \mathbf{x}_k^{[p]})$ and the multiplication $\circledast_{\boldsymbol{\alpha}}$ involves an extensive amount of floating-point arithmetic

operations. To save the overhead, we could keep $\mathscr{D}(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})$ the same throughout the sweeps of $\boldsymbol{\alpha}\subset[\![k]\!]$ per p. This is in the same spirit of the classical Jacobi iteration, but is applied to the SVD in our case. It is also feasible to follow the Gauss-Seidel notion by always using the most updated vectors in the definition of $\mathscr{D}$. This Gauss-Seidel-type updating scheme is implemented in our numerical experimentation. In all, we will show that the order by which $\boldsymbol{\alpha}=\{\alpha_1,\alpha_2\}\subseteq[\![k]\!]$ is selected is immaterial and will not affect the convergence.

4.2 Convergence analysis

So that we can describe the convergence behavior categorically for all possible strategies of selecting the partition $[\![k]\!]=\boldsymbol{\alpha}\cup\boldsymbol{\beta}$, we shall let $(\mathbf{x}_1^{[p+1]},\ldots,\mathbf{x}_k^{[p+1]})$ denote an update from $(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})$ whenever an $\boldsymbol{\alpha}$ is chosen and two vectors $\mathbf{x}_{\alpha_1}^{[p]}$ and $\mathbf{x}_{\alpha_2}^{[p]}$ have been updated according to (26). Thus, in the list of $(\mathbf{x}_1^{[p+1]},\ldots,\mathbf{x}_k^{[p+1]})$, we have exact copies of $\mathbf{x}_{\beta_i}^{[p+1]}=\mathbf{x}_{\beta_i}^{[p]}$, $i\in[\![k-2]\!]$ and only $\mathbf{x}_{\alpha_1}^{[p+1]}$ and $\mathbf{x}_{\alpha_2}^{[p+1]}$ are newly updated. In this way, the update $^{[p+1]}$ already includes a specific choice of $\boldsymbol{\alpha}$. We may thus write $\lambda_{\boldsymbol{\alpha}}^{[p+1]}$ as $\lambda^{[p+1]}$.

We first observe the converges of the λ values.

Theorem 3 *Suppose that T is SPD. If $\{\lambda^{[p]}\}$ is the sequence of dominant singular values generated by* (26) *with randomly selected $\boldsymbol{\alpha}\subseteq[\![k]\!]$, then*

$$\lambda(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})\le\lambda^{[p+1]}\le\lambda(\mathbf{x}_1^{[p+1]},\ldots,\mathbf{x}_k^{[p+1]})\le\lambda^{[p+2]} \tag{27}$$

and the sequence $\{\lambda^{[p]}\}$ converges.

Proof By the definition of λ, we can write

$$\lambda(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})=\langle\mathfrak{D}(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})\circledast_{\boldsymbol{\alpha}}(\bigcirc_{i=1}^{k-2}\mathbf{x}_{\beta_i}^{[p]}),\bigcirc_{i=1}^{2}\mathbf{x}_{\alpha_i}^{[p]}\rangle.$$

Since $\lambda^{[p+1]}$ is the dominant singular values of the matrix $\mathfrak{D}(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})\circledast_{\boldsymbol{\alpha}}(\bigcirc_{i=1}^{k-2}\mathbf{x}_{\beta_i}^{[p]})$, the first inequality follows. Similarly, the third inequality holds. To prove the second inequality, define the abbreviation:

$$\mathbf{a}^{[p]}:=\bigotimes_{i=1}^{k}\mathbf{x}_i^{[p]}. \tag{28}$$

Then, we find that

$$\begin{aligned}
&\lambda(\mathbf{x}_1^{[p+1]},\ldots,\mathbf{x}_k^{[p+1]})-\lambda^{[p+1]}\\
&=\langle T\mathbf{a}^{[p+1]},\mathbf{a}^{[p+1]}\rangle-\langle T\mathbf{a}^{[p]},\mathbf{a}^{[p+1]}\rangle\\
&=\langle T\mathbf{a}^{[p+1]},\mathbf{a}^{[p+1]}-\mathbf{a}^{[p]}\rangle\\
&=\langle T(\mathbf{a}^{[p+1]}-\mathbf{a}^{[p]}),(\mathbf{a}^{[p+1]}-\mathbf{a}^{[p]})\rangle+\langle T\mathbf{a}^{[p]},\mathbf{a}^{[p+1]}-\mathbf{a}^{[p]}\rangle.
\end{aligned}$$

The first term in the last equation is nonnegative because T is SPD. The second term is nonnegative because of the first inequality. Being a bounded monotone sequence, $\{\lambda^{[p]}\}$ must converge. □

The same argument for the generic Condition P imposed on the matrix T for the polynomial system (16) can be applied to the matrix T for the polynomial system (25) [14,37]. Additionally, it is also known that the symmetric matrices with multiply eigenvalues form an algebraic variety of codimension two [9] which, of course, is nowhere dense and of measure zero. This can be translated as that the dominant singular value being simple is generic. Together, the following condition for the matrix T is still generic.

Condition S: We say that the matrix T satisfies Condition S if the corresponding polynomial system (25) has finitely many, geometrically isolated, real-valued solutions and that the associated matrix $\mathscr{D}(\mathbf{x}_1,\ldots,\mathbf{x}_k)\circledast_{\boldsymbol{\alpha}}\left(\mathbf{x}_{\beta_{k-2}}\circ\ldots\circ\mathbf{x}_{\beta_1}\right)$ has a simple dominant singular value.

Finally, we prove the convergence of the SVD-type iteration for generic T.

Theorem 4 *Assume that T is SPD and satisfies Condition S. Then the sequence $\{(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})\}$ generated by the scheme (26) with any order $\boldsymbol{\alpha}\subset[\![k]\!]$ converges.*

Proof In terms of the abbreviation defined in (28), using the assumption that the sequence $\{\lambda^{[p]}\}$ is increasing and the interlacing property (27), we know that as $p\to\infty$,

$$\lambda^{[p+1]}-\lambda(\mathbf{x}_1^{[p]},\ldots,\mathbf{x}_k^{[p]})=\langle\mathbf{a}^{[p+1]},T\mathbf{a}^{[p]}\rangle-\langle\mathbf{a}^{[p]},T\mathbf{a}^{[p]}\rangle\to 0,$$
$$\lambda(\mathbf{x}_1^{[p+1]},\ldots,\mathbf{x}_k^{[p+1]})-\lambda^{[p+1]}=\langle\mathbf{a}^{[p+1]},T\mathbf{a}^{[p+1]}\rangle-\langle\mathbf{a}^{[p+1]},T\mathbf{a}^{[p]}\rangle\to 0.$$

It follows that as $p\to\infty$,

$$\begin{aligned}&\langle\mathbf{a}^{[p+1]}-\mathbf{a}^{[p]},T(\mathbf{a}^{[p+1]}-\mathbf{a}^{[p]})\rangle\\&=(\langle\mathbf{a}^{[p+1]},T\mathbf{a}^{[p+1]}\rangle-\langle\mathbf{a}^{[p+1]},T\mathbf{a}^{[p]}\rangle)-(\langle\mathbf{a}^{[p+1]},T\mathbf{a}^{[p]}\rangle-\langle\mathbf{a}^{[p]},T\mathbf{a}^{[p]}\rangle)\to 0.\end{aligned}$$

Since T is SPD, we see that

$$\lim_{p\to\infty}\|\mathbf{a}^{[p+1]}-\mathbf{a}^{[p]}\|_2=0.$$

By the fact that

$$\|\mathbf{a}^{[p+1]}-\mathbf{a}^{[p]}\|_2^2=2-2\prod_{i=1}^{k}\langle\mathbf{x}_i^{[p]},\mathbf{x}_i^{[p+1]}\rangle,$$

we see that

$$\lim_{p\to\infty}\prod_{i=1}^{k}\langle\mathbf{x}_i^{[p]},\mathbf{x}_i^{[p+1]}\rangle=1.$$

Since $|\langle\mathbf{x}_i^{[p]},\mathbf{x}_i^{[p+1]}\rangle|\le 1$, it must be that

$$\lim_{p\to\infty}\langle\mathbf{x}_i^{[p]},\mathbf{x}_i^{[p+1]}\rangle=1,\quad i\in[\![k]\!].$$

That is, the two consecutive vectors $\mathbf{x}_i^{[p]}$ and $\mathbf{x}_i^{[p+1]}$ must gradually be aligned. The rule of sign selection imposed forces them to gradually point to the same direction as $p\to\infty$. In particular,

$$\Delta\mathbf{x}_i^{[p]}:=\mathbf{x}_i^{[p+1]}-\mathbf{x}_i^{[p]}\to 0,\quad i\in[\![k]\!].$$

We thus complete the proof by Lemma 2 under Condition S. □

5 Numerical experiments

In the above, we have proposed a power-like method and an SVD-like method for tackling the rank-1 k-partite approximation problem. In this section, we carry out some numerical experiments to illustrate the effectiveness of our algorithms. In particular, we want to compare whether our simple iterative methods, even at their rudimentary implementation, are compatible with some existing state-of-the-art optimization techniques. There are many compatible optimization software packages. For demonstration purpose, we use Matlab as the computational platform and limit our comparison to those available in the Matlab Optimization Toolbox. The following experiments are performed on a MacBook Pro laptop with Quad-Core Intel Core i5 @ 2.4GHz processor and 16GB RAM by using MATLAB, version 2020b, as the computing platform.

Example 1. The purpose of our first experiment is to test the efficiency and accuracy of the power-like iteration (17) and the SVD-like iteration (26). To simulate the experiment, we choose $k = 3$ and generate three random vectors $\mathbf{y}_i \in \mathbb{R}^{10}$, $i \in [\![3]\!]$, from the identical and independent Gaussian distribution. Define a rank-1 density matrix

$$T_1 := (\mathbf{x}_1\mathbf{x}_1^\top) \otimes (\mathbf{x}_2\mathbf{x}_2^\top) \otimes (\mathbf{x}_3\mathbf{x}_3^\top) \in \mathbb{R}^{1000\times 1000}, \tag{29}$$

where $\mathbf{x}_i := \frac{y_i}{\|y_i\|_2} \in S^9$, $i \in [\![3]\!]$, as the base matrix which has exact decomposition. We consider the rank-1 3-partite approximation to this perturbed matrix

$$T_\sigma := T_1 + \sigma(B - T_1),$$

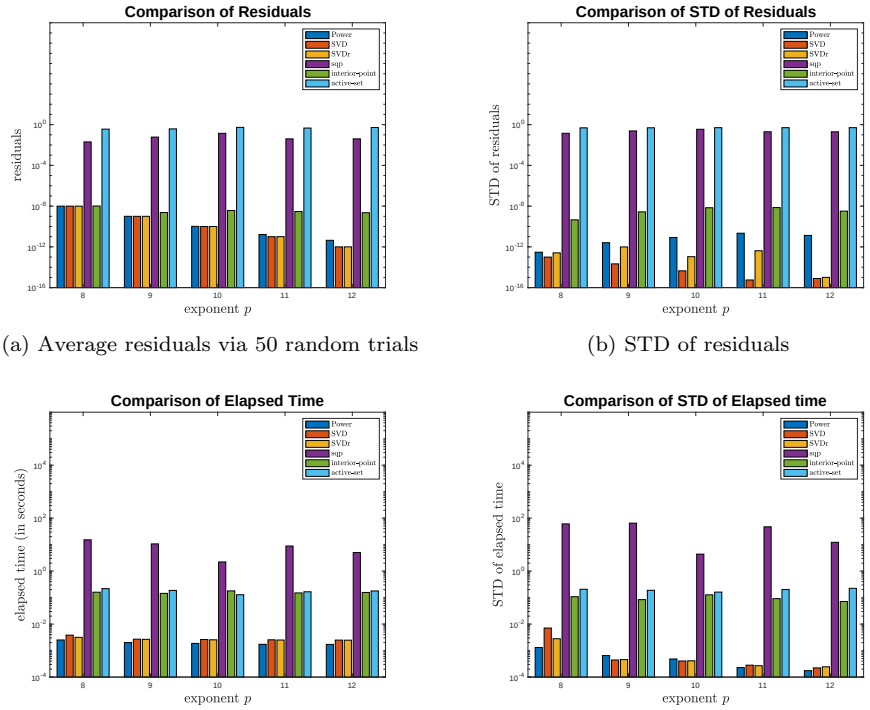
where B is a randomly generated but fixed density matrix and $\sigma = 10^{-p}$, $p = 8, \ldots, 12$, signifies the magnitude of the noise. Since T_σ is a convex combination of two density matrices, it remains to be a density matrix. Nonetheless, even with a small perturbation, the matrix T_σ is generally of full rank.

We compare our algorithms with the conventional Matlab routine fmincon employing three distinct solvers "sqp", "interior-point", and "active-set", respectively. Since counting the number of floating-point arithmetic operations is no longer a reliable means for measuring the computational complexity, we measure the CPU time as a criterion to evaluate the performance. Because each method has its own special characteristics, it makes a term-by-term comparison difficult. For a straightforward comparison, we turn off all other stopping criteria but demand that all methods must meet the same first-order optimality condition:

$$\left\| \left[\mathscr{D}(\mathbf{x}_1, \ldots, \mathbf{x}_k) \circledast_j \bigcirc_{i=1, i\neq j}^{k} \mathbf{x}_i - \lambda \mathbf{x}_j \right]_{j=1,\ldots,k} \right\|_2 < 10^{-10}. \tag{30}$$

before terminating the iteration. For each σ, we repeat our experiments 50 times with randomly generated B and, after taking the standard deviations into account, take the average to represent the general trends. Each time different starting values are generated, but the same values are used for each method. While we want to minimize (8) with $T = T_\sigma$, we are hoping to recover T_1 which is under perturbation. Thus, we measure the final quality of approximation based on the definition:

$$\text{Residual} := \|T_1 - \lambda(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_3)(\hat{\mathbf{x}}_1\hat{\mathbf{x}}_1^\top) \otimes (\hat{\mathbf{x}}_2\hat{\mathbf{x}}_2^\top) \otimes (\hat{\mathbf{x}}_3\hat{\mathbf{x}}_3^\top)\|_F, \tag{31}$$



(a) Average residuals via 50 random trials

(b) STD of residuals

(c) Average CPU Time

(d) STD of elapsed time

Fig. 1: Average residuals, elapsed time, and respective STDs for approximating perturbed T_1.

where the triplet $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_3)$ represents the final result output by each of the specific algorithms, respectively. Finally, since the gradient information is readily available, the option of user-provided analytic gradients for both the objective function and the constraint is turned on for all Matlab routines, which should help the efficiency and the precision if the Hessian is needed.

Under the above rules of setup, depicted in Figure 1(a) are the averages of the residuals computed by six distinct methods. These are the power-like iteration (Power) updated according to the Gauss-Seidel and cyclic rules, the SVD-like iterations with α chosen cyclically (SVD) or randomly (SVDr), the solvers sqp, interior-point, and active-set, respectively. To make sure that taking the average makes sense, we record the standard derivations (STD) of the residuals from the respective means in Figure 1(b). Extremely small variances observed for our methods as well as the relatively small variance for the interior-point method indicate that the means do represent the general mode of approximation. Together with the small residuals observed in Figure 1(a), these empirical data strongly suggest that our techniques, the SVD-like methods in particular, tend to have better approximations.

One possible explanation for the poor performance by the more sophisticated solvers sqp and active-set might be that these methods have terminated prema-

turely under our sole stopping criterion (30). We understand that a robust code should have multi-layer stopping criteria in place. One way to improve the precision in the Matlab built-in solvers is to fine-tune the tolerance of other factors such as the step sizes and constraints. In doing so, however, the required elapsed time will increase dramatically, which is already slow as we can see under the current setting.

Depicted in Figure 1(c) are the averages of elapsed time required by each method to meet the condition (30). We also measure the respective standard deviations of the required elapsed time in Figure 1(d). It manifests that our algorithms are consistently more time efficient at obtaining solutions with smaller residual values. The three routines from the Optimization Toolbox have much larger variances in the needed CPU time, indicating the varying difficulties in achieving (30). Also, while the power-like method is cheaper per iteration, the overall performance of the SVD-based methods, regardless whether $\boldsymbol{\alpha}$ is chosen cyclically or acyclically, seems to consume about the same CPU time while producing slightly smaller residuals. We stress that the implementation of our methods is not as sophisticated as those in the Matlab Optimization Toolbox, yet our experiments seem to suggest that the potential applicability of our power-like or SVD-like methods.

Example 2. The purpose of this experiment is to demonstrate how the numerical calculation helps discover a special property of the so-called a Greenberger-Horne-Zeilinger state (GHZ gate) [13,19]. Consider the quantum system $\mathbb{C}^2$ whose elements, called qubits, serve as the basic units for quantum information science. The standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ of $\mathbb{C}^2$ are often denoted by $|0\rangle$ and $|1\rangle$, respectively. Correspondingly, the 2-qubit element $|0\rangle \otimes |1\rangle$ in the bipartite system $\mathbb{C}^2 \otimes \mathbb{C}^2$ is conveniently abbreviated as $|01\rangle$ and so on. The GHZ gate

$$|GHZ\rangle := \frac{1}{\sqrt{2}}(|0\rangle^{\otimes k} + |1\rangle^{\otimes k})$$

is a pure k-qubit state that involves the entanglement of at least three subsystems, i.e., $k \geq 3$. GHZ states are used in several protocols in quantum communication and cryptography because they exhibit some non-classical properties.

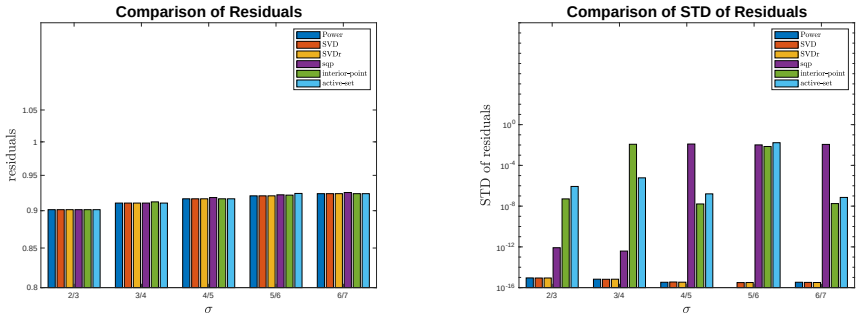
Assume $k = 3$ in this experiment. Consider the mixture of the GHZ state with the white noise state $(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle)^{\otimes 3}$, i.e., a mixed state in the form

$$\Omega_\sigma := (1-\sigma)\,|GHZ\rangle\langle GHZ| + \sigma\frac{1}{8}I_8, \quad 0 \leq \sigma \leq 1,$$

The density matrix Ω_σ, known as the generalized Werner state, has applications in the robustness of entanglement [42], NMR quantum computation [4], and purification schemes for entangled states [31]. In theory, we know that Ω_σ is separable in the sense of (5) if and only if $\frac{4}{5} \leq \sigma \leq 1$ [2,11], but the tensor rank R needed for the decomposition can be high.

On the other hand, the matrix representation of Ω_σ can be expressed as

$$\Omega_\sigma = (1-\sigma)\begin{bmatrix} \frac{1}{2} & 0 & \cdots & 0 & \frac{1}{2} \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{2} & 0 & \cdots & 0 & \frac{1}{2} \end{bmatrix} + \sigma\frac{1}{8}\begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \in \mathbb{R}^{8\times 8}.$$



(a) Average residuals via 50 random trials (b) STD of residuals

Fig. 2: Average residuals and respective STDs for approximating the GHZ gate.

Our problem (8) amounts to finding the best symmetric rank-1 approximation of Ω_σ, but this rank-1 matrix must involve the tensor product of three factors. Even for such a small size problem the specially structured rank-1 approximation is not obvious.

Using the same stopping criterion (30) and repeating the same experiment 50 times for different values of the probability σ, we obtain the empirical results shown in Figure 2. On one hand, we still observe from Figure 2(b) the consistency of our methods in producing the structured rank-1 approximation. On the other hand, Figure 2(a) may seem rather mundane because, for each prescribed σ, all methods produce nearly identical residuals. In order to understand why there is such a coincidence, we take a closer examination. We are surprised to discover that in fact Ω_σ has two best rank-1 separable approximations, i.e., $|000\rangle\langle 000|$ and $|111\rangle\langle 111|$, which are independent of σ. Therefore, up to the error induced by the stopping criterion (30), all methods lead to nearly the same best rank-1 separable approximation, whereas the corresponding optimal λ defined in (11) is given by $\lambda_\sigma = (4 - 3\sigma)/8$.

Example 3. Similar to the GHZ state, we now experiment with another non-biseparable 3-qubit states, i.e., the so-called W state

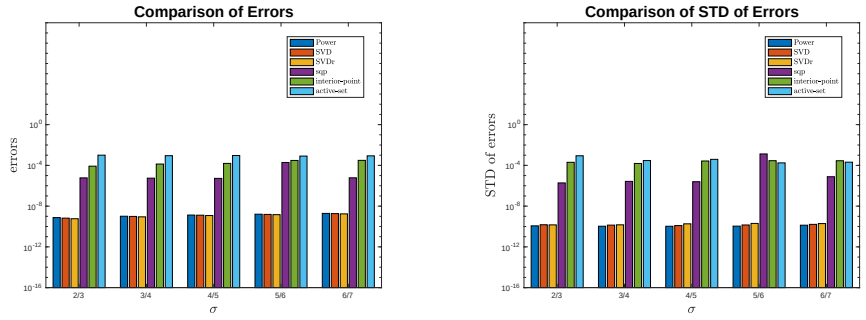
$$|W\rangle = \frac{1}{\sqrt{3}}(|100\rangle + |010\rangle + |001\rangle). \tag{32}$$

In application, constructing the GHZ state is easier than the W state, but the W state is more robustly entangled in the sense that, when a subsystem is traced out, the remaining states is still entangled [12,17,25].

Analogous to the preceding experiment, consider the perturbed state in the form:

$$W_\sigma := (1-\sigma)\,|W\rangle\langle W| + \sigma\frac{1}{8}I_8, \quad 0 \le \sigma \le 1.$$

Having experienced the GHZ gate, we speculate and are able to verify analytically that the density matrix W_σ has three nearest rank-1 separable states. These are the density matrices $|100\rangle\langle 100|$, $|010\rangle\langle 010|$ and $|001\rangle\langle 001|$ which, again, are independent of σ. Let $\hat{\mathbf{x}} := \hat{\mathbf{x}}_1 \otimes \hat{\mathbf{x}}_2 \otimes \hat{\mathbf{x}}_3$ denote the tensor product of the final output triplet $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_3)$ by various algorithms for W_σ. We are curious to know whether



(a) Average errors via 50 random trials (b) STD of Errors

Fig. 3: Average errors, elapsed time, and respective STDs for approximating W_σ.

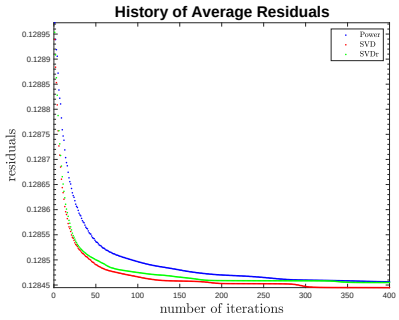
the pure state $\hat{\mathbf{x}}$ approximates any of these 3-qubit entangled states, $|100\rangle$, $|010\rangle$ or $|001\rangle$. Therefore, in this experiment, we take the measurement of

$$\mathsf{Error} := \min\{\| \, |001\rangle - \mathrm{sgn}(\hat{\mathbf{x}}_2)\hat{\mathbf{x}}\|_2, \| \, |010\rangle - \mathrm{sgn}(\hat{\mathbf{x}}_3)\hat{\mathbf{x}}\|_2, \| \, |100\rangle - \mathrm{sgn}(\hat{\mathbf{x}}_5)\hat{\mathbf{x}}\|_2\}. \quad (33)$$

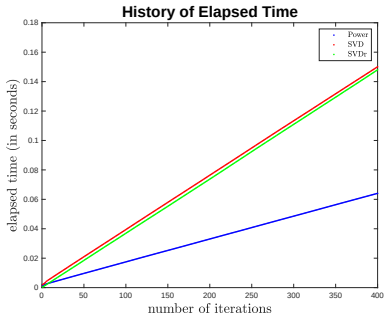
The average of errors measured in terms of (33) and the corresponding STDs after 50 runs for each prescribed algorithm computed with the same stopping condition (30) are shown in Figure 3. The small errors in Figure 3(a) confirmed the working of our power-like and SVD-based methods while, in contrast, the standard optimization routines, e.g., the `sqp`, produce an approximation with much larger errors. The small STDs in Figure 3(b) also confirm that our methods are more consistent than the standard optimization routines in producing the best approximation.

Example 4. The purpose of our last experiment is to assess simultaneously the accuracy and efficiency of our two methods when applied to 4-partite systems. We randomly generate an SPD matrix $T \in \mathbb{R}^{120\times 120}$ with unit trace as the target density matrix and search for unit vectors $\mathbf{x}_1 \in \mathbb{R}^5$, $\mathbf{x}_2 \in \mathbb{R}^4$, $\mathbf{x}_3 \in \mathbb{R}^3$, and $\mathbf{x}_4 \in \mathbb{R}^2$ to minimize (8). With the fixed T, we let each algorithm carry out 400 iterations. We repeat our experiments 50 times, each time with the same random starting values for all methods. We plot the history of the average behavior in Figure 4 to compare the residuals and overhead progressively.

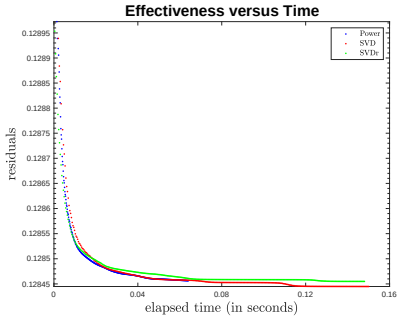
In Figure 4(a), we see the trend that the SVD-like iterations can reduce the residuals more rapidly within a relatively small number of iterations than the power-like iteration. However, the call of `svds` per iteration by the SVD-like schemes requires many iterations within the Lanczos algorithm, whereas the calculation required by the power-like scheme is straightforward and, thus, requires less elapsed time. Figure 4(b) indicates that, on average, the SVD-like scheme takes approximately twice as much time per iteration when comparing to the power-like method. To see how these two conflicting measurements can be mended, we plot in Figure 4(c) the history of residuals versus the elapsed time. It is interesting to find that the improvement of residuals by the SVD-like iteration per unit time with a random choice of $\boldsymbol{\alpha}$ is almost the same as that of the power-like iteration. In other words, if speed and precision are both desired, then the power-like scheme and the acyclic SVD-like scheme might be the methods of choice.



(a) Average residuals via 50 random trials



(b) Average CPU Time



(c) Performance

Fig. 4: Average residuals and elapsed time.

6 Conclusion

Motivated by the structure embedded in quantum entanglement, we consider the real version approximation to entangled multipartite systems. In contrast to the conventional tensor approximation that is multilinear in its factors, the entangled multipartite system involves density matrices of pure states and, hence, is nonlinear with respect to its factors. Generalizing a previous work on bipartite systems, this paper studies the basic rank-1 approximation to multipartite systems.

The first order optimality condition is rewritten as a nonlinear eigenvalue problem and a nonlinear singular value problem. Correspondingly, a power-like iterative scheme and an SVD-like iterative scheme are proposed as means for numerical calculation. Convergence theory is established. Though the schemes appear simple, numerical experiments seem to suggest that they are effective and efficient for tackling the rank-1 multipartite approximation problem. Further work should include the extension of these schemes to the more general low-rank multipartite systems and complex systems.

Declarations

This research concerns the theoretical study and numerical experiment of an open mathematical problem. To our knowledge, it does not involve nor imply any biological applications.

References

1. Aaronson, S.: Quantum Computing since Democritus. Cambridge University Press (2013). URL https://www.scottaaronson.com/democritus/
2. Ben-Aryeh, Y., Mann, A.: Biseparability of 3-qubits density matrices using hilbertschmidt decompositions: Sufficient conditions and explicit expressions. International Journal of Quantum Information **14**(08), 1650044 (2016). URL https://doi.org/10.1142/S0219749916500441
3. Brachat, J., Comon, P., Mourrain, B., Tsigaridas, E.: Symmetric tensor decomposition. Linear Algebra Appl. **433**(11-12), 1851–1872 (2010). URL http://doi.org/10.1016/j.laa.2010.06.046
4. Braunstein, S.L., Caves, C.M., Jozsa, R., Linden, N., Popescu, S., Schack, R.: Separability of very noisy mixed states and implications for nmr quantum computing. Phys. Rev. Lett. **83**, 1054–1057 (1999). URL https://doi.org/10.1103/PhysRevLett.83.1054
5. Chen, K., Wu, L.A.: A matrix realignment method for recognizing entanglement. Quantum Inf. Comput. **3**(3), 193–202 (2003)
6. Chen, L., Aulbach, M., Hajdušek, M.: Comparison of different definitions of the geometric measure of entanglement. Phys. Rev. A **89**, 042305 (2014). URL https://doi.org/10.1103/PhysRevA.89.042305
7. Chu, M.T., Lin, M.M.: Nonlinear power-like and SVD-like iterative schemes with applications to entangled bipartite rank-1 approximation. SIAM J. Sci. Comput. **43**(5), S448–S474 (2021). URL https://doi.org/10.1137/20M1336059
8. Comon, P., Luciani, X., de Almeida, A.L.F.: Tensor decompositions, alternating least squares and other tales. J. Chemometrics **23**, 393–405 (2009). URL https://doi.org/10.1002/cem.1236
9. Dana, M., Ikramov, K.D.: On the codimension of the variety of symmetric matrices with multiple eigenvalues. Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **323**(Chisl. Metody i Vopr. Organ. Vychisl. 18), 34–46, 224 (2005). URL http://doi.org/10.1007/s10958-006-0275-7
10. De Lathauwer, L., De Moor, B., Vandewalle, J.: On the best rank-1 and rank-$(R_1, R_2, \cdots, R_N)$ approximation of higher-order tensors. SIAM J. Matrix Anal. Appl. **21**(4), 1324–1342 (2000). URL https://doi.org/10.1137/S0895479898346995
11. Dür, W., Cirac, J.I.: Classification of multiqubit mixed states: Separability and distillability properties. Phys. Rev. A **61**, 042314 (2000). URL https://doi.org/10.1103/PhysRevA.61.042314
12. Dür, W., Vidal, G., Cirac, J.I.: Three qubits can be entangled in two inequivalent ways. Phys. Rev. A (3) **62**(6), 062314, 12 (2000). URL https://doi.org/10.1103/PhysRevA.62.062314
13. Eltschka, C., Siewert, J.: Entanglement of three-qubit Greenberger-Horne-Zeilinger-symmetric states. Phys. Rev. Lett. **108**, 020502 (2012). URL https://doi.org/10.1103/PhysRevLett.108.020502
14. García, C.B., Li, T.Y.: On the number of solutions to polynomial systems of equations. SIAM J. Numer. Anal. **17**(4), 540–546 (1980). URL https://doi.org/10.1137/0717046
15. Gharibian, S.: Strong np-hardness of the quantum separability problem. Quantum Info. Comput. **10**(3), 343–360 (2010). URL http://dl.acm.org/citation.cfm?id=2011350.2011361
16. Goncharov, A.V., Hrkac, G., Dean, J.S., Schrefl, T.: Kronecker product approximation of demagnetizing tensors for micromagnetics. J. Comput. Phys. **229**(7), 2544–2549 (2010). URL https://doi.org/10.1016/j.jcp.2009.12.004
17. Gorbachev, V.N., Trubilko, A.I., Rodichkina, A.A., Zhiliba, A.I.: Can the states of the W-class be suitable for teleportation? Phys. Lett. A **314**(4), 267–271 (2003). URL https://doi.org/10.1016/S0375-9601(03)00906-X

18. Grasedyck, L.: Existence and computation of low kronecker-rank approximations for large linear systems of tensor product structure. Computing **72**, 247–265 (2004). URL http://doi.org/10.1007/s00607-003-0037-z.
19. Greenberger, D.M., Horne, M.A., Zeilinger, A.: Going beyond Bell's theorem (2007)
20. Guan, Y., Chu, M.T., Chu, D.: Convergence analysis of an SVD-based algorithm for the best rank-1 tensor approximation. Linear Algebra Appl. **555**, 53–69 (2018). URL https://doi.org/10.1016/j.laa.2018.06.006
21. Gurvits, L.: Classical complexity and quantum entanglement. Journal of Computer and System Sciences **69**(3), 448 – 484 (2004). URL http://www.sciencedirect.com/science/article/pii/S0022000004000893
22. Hackbusch, W., Khoromskij, B.N., Tyrtyshnikov, E.E.: Hierarchical Kronecker tensor-product approximations. J. Numer. Math. **13**(2), 119–156 (2005). URL http://doi.org/10.1163/1569395054012767
23. Hiai, F., Petz, D.: Introduction to matrix analysis and applications. Universitext. Springer, Cham; Hindustan Book Agency, New Delhi (2014). URL https://doi.org/10.1007/978-3-319-04150-6
24. Ishteva, M., Absil, P.A., Van Dooren, P.: Jacobi algorithm for the best low multilinear rank approximation of symmetric tensors. SIAM J. Matrix Anal. Appl. **34**(2), 651–672 (2013). URL https://doi.org/10.1137/11085743X
25. Joo, J., Park, Y.J., Oh, S., Kim, J.: Quantum teleportation via a W state. New Journal of Physics **5**, 136–136 (2003). URL https://doi.org/10.1088/1367-2630/5/1/136
26. Kofidis, E., Regalia, P.A.: On the best rank-1 approximation of higher-order supersymmetric tensors. SIAM J. Matrix Anal. Appl. **23**(3), 863–884 (2001). URL http://doi.org/10.1137/S0895479801387413
27. Kolda, T.G.: Numerical optimization for symmetric tensor decomposition. Math. Program. **151**(1, Ser. B), 225–248 (2015). URL http://doi.org/10.1007/s10107-015-0895-0
28. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009). URL http://doi.org/10.1137/07070111X
29. Langville, A.N., Stewart, W.J.: A kronecker product approximate preconditioner for SANs. Numer. Linear Algebra Appl. **11**(8-9), 723–752 (2004). URL http://doi.org/10.1002/nla.344
30. Moré, J.J., Sorensen, D.C.: Computing a trust region step. SIAM J. Sci. Statist. Comput. **4**(3), 553–572 (1983). URL http://doi.org/10.1137/0904038
31. Murao, M., Plenio, M.B., Popescu, S., Vedral, V., Knight, P.L.: Multiparticle entanglement purification protocols. Phys. Rev. A **57**, R4075–R4078 (1998). URL https://doi.org/10.1103/PhysRevA.57.R4075
32. Nagy, J., Kilmer, M.: Kronecker product approximation for preconditioning in three-dimensional imaging applications. Image Processing, IEEE Transactions on **15**(3), 604 –613 (2006).
33. Nakahara, M., Ohmi, T.: Quantum computing. CRC Press, Boca Raton, FL (2008). URL https://doi.org/10.1201/9781420012293. From linear algebra to physical realizations
34. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information: 10th Anniversary Edition. Cambridge University Press (2010).
35. Rezghi, M., Hosseini, S.M., Eldén, L.: Best Kronecker product approximation of the blurring operator in three dimensional image restoration problems. SIAM J. Matrix Anal. Appl. **35**(3), 1086–1104 (2014). URL https://doi.org/10.1137/130917260
36. Silvestre, D., Hespanha, J., Silvestre, C.: A pagerank algorithm based on asynchronous gauss-seidel iterations. In: 2018 Annual American Control Conference (ACC), pp. 484–489 (2018).
37. Sommese, A.J., Wampler II, C.W.: The numerical solution of systems of polynomials arising in engineering and science. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ (2005). URL http://doi.org/10.1142/9789812567727
38. Uschmajew, A.: Local convergence of the alternating least squares algorithm for canonical tensor approximation. SIAM J. Matrix Anal. Appl. **33**(2), 639–652 (2012). URL http://doi.org/10.1137/110843587
39. Van Loan, C.F.: The ubiquitous Kronecker product. J. Comput. Appl. Math. **123**(1-2), 85–100 (2000). URL http://doi.org/10.1016/S0377-0427(00)00393-9
40. Van Loan, C.F., Pitsianis, N.: Approximation with Kronecker products. In: Linear algebra for large scale and real-time applications (Leuven, 1992), *NATO Adv. Sci. Inst. Ser. E Appl. Sci.*, vol. 232, pp. 293–314. Kluwer Acad. Publ., Dordrecht (1993)

41. Vervliet, N., Debals, O., Sorber, L., Van Barel, M., De Lathauwer, L.: Tensorlab 3.0 (2016). URL https://www.tensorlab.net. Available online.
42. Vidal, G., Tarrach, R.: Robustness of entanglement. Phys. Rev. A **59**, 141–155 (1999). URL https://doi.org/10.1103/PhysRevA.59.141
43. Wang, L., Chu, M.T.: On the global convergence of the alternating least squares method for rank-one approximation to generic tensors. SIAM J. Matrix Anal. Appl. **35**(3), 1058–1072 (2014). URL http://doi.org/10.1137/130938207
44. Zhang, T., Golub, G.H.: Rank-one approximation to high order tensors. SIAM J. Matrix Anal. Appl. **23**(2), 534–550 (electronic) (2001). URL http://doi.org/10.1137/S0895479899352045
45. Zhao, X.L., Huang, T.Z., Lv, X.G., Xu, Z.B., Huang, J.: Kronecker product approximations for image restoration with new mean boundary conditions. Appl. Math. Model. **36**(1), 225–237 (2012). URL https://doi.org/10.1016/j.apm.2011.05.050