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

Convergence Analysis of Alternating Direction Methods: A General Framework and Its Applications to Tensor Approximations

Yu Guan · Nan Jiang · Bo Dong · Moody T. Chu

Received: date / Accepted: date

Abstract For problems involving multiple variables, the notion of solving a sequence of simplified problems by fixing all but one variable a time and alternating among the variables has been exploited in a wide range of applications. Finding the various types of low rank tensor approximations, for example, is so complicated that the alternating least squares method has been used as a workhorse for the computation. Though the alternating mechanism often carries an innate monotone property that can be used to explain the limiting behavior of the objective values, a proof of global convergence for the iterates themselves has been mostly elusive. This paper proposes a general framework that can be applied to prove convergence for many types of alternating direction methods. The conditions entailed by this framework are mild and easy to satisfy, so the theory should be of fundamental significance to many algorithms. Its application to a variety of important algorithms is demonstrated. While some existent proofs can be significantly simplified, the convergence results for alternating least squares algorithms designed for the Tucker nearest problem and structured Kronecker approximation problem are new.

Keywords alternating direction iteration · alternating least squares methods · best rank-1 approximation · Tucker nearest problem · structured Kronecker approximation problem · convergence analysis

Mathematics Subject Classification (2000) 15A15 · 15A09 · 15A23

The second author's research was supported in part by the Chinese Scholarship Council. The third author's research was supported in part by the Chinese Scholarship Council and the Fundamental Research Funds for Central Universities. The fourth author's research was supported in part by the National Science Foundation under grant DMS-1316779.

Yu Guan
Department of Mathematics, National University of Singapore, Singapore 119076.
E-mail: a0123906@u.nus.edu

Nan Jiang
Harbin Engineering University, Harbin, China.
E-mail: jngrace@hrbeu.edu.cn

Bo Dong
School of Mathematical Sciences, Dalian University of Technology, Dalian, Liaoning 116024, China.
E-mail: dongbo@dlut.edu.cn

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

1 Introduction

Many algorithms can be cast in the abstract form

$$\begin{cases} \mathbf{x}_{k+1} = f(\mathbf{y}_k), \\ \mathbf{y}_{k+1} = g(\mathbf{x}_{k+1}), \end{cases} \quad k = 0, 1, \ldots, \tag{1}$$

where $f : U \to V$ and $g : V \to U$, referred to henceforth as the generating functions, are maps representing some black-box evaluations or some intermediate numerical procedures. The variables $\mathbf{x}$ and $\mathbf{y}$ can be vectors, matrices, or even functions. The choice of U, V depends on the desired properties of the variables $\mathbf{x}$ and $\mathbf{y}$, which can be, for instance, nonnegative, orthogonal, or stochastic. In this note, we focus only on finite dimensional variables, so the feasible sets U, V are subsets in some Euclidean spaces with suitable dimensions and constraints. We shall give several interesting but nontrivial examples in the later part of this discussion to demonstrate this point. For more complicated problems involving n variables $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$, a similar alternating iteration can be written in this form

$$\begin{cases} \mathbf{x}_{k+1}^{(1)} = f^{(1)}(\mathbf{x}_k^{(2)}, \mathbf{x}_k^{(3)}, \ldots, \mathbf{x}_k^{(n)}), \\ \mathbf{x}_{k+1}^{(2)} = f^{(2)}(\mathbf{x}_{k+1}^{(1)}, \mathbf{x}_k^{(3)}, \ldots, \mathbf{x}_k^{(n)}), \\ \qquad \vdots \\ \mathbf{x}_{k+1}^{(n)} = f^{(n)}(\mathbf{x}_{k+1}^{(1)}, \mathbf{x}_{k+1}^{(2)}, \ldots, \mathbf{x}_{k+1}^{(n-1)}), \end{cases} \quad k = 0, 1, \ldots. \tag{2}$$

Perhaps the simplest algorithm in the form of (2) is the Gauss-Seidel iterative scheme used for solving a linear system where all maps $f^{(\ell)}$ are linear and $\mathbf{x}^{(\ell)}$ are scalars. Another example is the alternating least squares (ALS) method used for low rank tensor approximations of a given order-n tensor [9,24, 30,35,33], where all variables are expected to be of unit length. We shall concentrate on the analysis for (1) in this paper. The generalization to (2) can be accomplished in a similar way.

Obviously, the sequence $\{\mathbf{y}_k\}$ generated by (1) can be obtained from the fixed-point iteration

$$\mathbf{y}_{k+1} = g(f(\mathbf{y}_k)), \quad k = 0, 1, \ldots. \tag{3}$$

If the composite $F := g \circ f$, referred to henceforth as the transition function (of one sweep for $\mathbf{y}_k$), is a contraction map, then the Banach fixed-point theorem asserts that the iterates from (3) converge to a unique fixed point. This is the most impeccable conclusion, but often proving that $g \circ f$ is a contraction map is difficult or impossible. Likewise, if $g \circ f$ is continuous and maps a convex compact set into itself, then the Brouwer fixed-point theorem asserts that there is a fixed-point $\mathbf{y}^*$ such that $g \circ f(\mathbf{y}^*) = \mathbf{y}^*$. In general, however, not much is known about the limiting behavior of the sequence $\{\mathbf{y}_k\}$ itself. For many of the algorithms discussed in the literature and even used in practice, we find that lacking a rigorous convergence analysis for the iterates $\{\mathbf{x}_k\}$ and $\{\mathbf{y}_k\}$ themselves is a serious and widespread shortfall [2,24]. The main contribution of this paper is a general framework for characterizing the limiting behavior of (1) under much easy-to-check criteria. We apply the framework to a variety of alternating direction methods and, in particular, the alternating least squares algorithms to demonstrate how the theory facilitates convergence analysis, some of which are difficult to come by otherwise.

This paper is organized as follows. In Section 2, we build our framework by progressively adding in conditions. The theory works in its most basic form, but more conditions make it easier to draw conclusions. As a demonstration, we apply the theory in Section 3 to a variety of classical results in the literature. In this context, the proof of convergence is not new, but it shows the versatility of our framework. In Section 4, we use our theory to argue the convergence of algorithms for the Tucker nearest problem and the structured Kronecker approximation problem.

2 Basic theory

We begin our theory with the most basic form, namely, checking the difference between every convergent subsequence and its immediate next iterate. The following lemma, originally proved [25, Lemma 4.10] and then reproved in [13, Lemma 2.7], asserts a sufficient condition for convergence.

Lemma 1 *Assume that a^* is an isolated accumulation point of a sequence $\{a_k\}$ such that for every subsequence $\{a_{k_j}\}$ converging to a^*, there is an infinite subsequence $\{a_{k_{j_i}}\}$ such that $|a_{k_{j_i}+1} - a_{k_{j_i}}| \to 0$. Then the whole sequence $\{a_k\}$ converges to a^*.*

To apply Lemma 1 to algorithms such as (1), we follow the steps that

a. Check to see that an accumulation point of a convergent subsequence $\{a_{k_j}\}$ is isolated. (See the remarks following Corollary 1 and Lemma 2.)
b. Search for a subsequence $\{a_{k_{j_i}}\}$ such that after applying the transition map, say F, the difference $|F(a_{k_{j_i}}) - a_{k_{j_i}}|$ diminishes to zero.

For specific applications, see our recent work on the convergence of the ALS algorithm and the SVD-based algorithm for the best rank-1 tensor approximations in [13,33].

By imposing the continuity on the generating function and the finiteness on isolated accumulation points, the following lemma asserts a specific limiting behavior of the resulting iterates.

Theorem 1 *Let $F : U \longrightarrow U$ be a continuous map over a closed subset $U \subset \mathbb{R}^n$. Suppose that the sequence $\{\mathbf{z}_k\}$ generated by iterative scheme $\mathbf{z}_{k+1} = F(\mathbf{z}_k)$ is well defined, bounded, and has finitely many isolated accumulation points. Then*

1. *Either the sequence $\{\mathbf{z}_k\}$ converges, or*
2. *There are disjoint neighborhoods of the accumulation points such that, for k large enough, the consecutive elements $\mathbf{z}_k, \mathbf{z}_{k+1}, \ldots$ visit each neighborhood in a cyclic order.*

Proof Let $\{\mathbf{z}_{k_i}\}$ denote an arbitrary convergent subsequence of $\{\mathbf{z}_k\}$. By continuity, the subsequence $\{\mathbf{z}_{k_i+1}\}$ also converges. Repeating this process, we denote the limiting behavior when $i \to \infty$ as

$$\begin{array}{rl} \mathbf{z}_{k_i} & \longrightarrow \mathbf{z}_0^* \\ \mathbf{z}_{k_i+1} & \longrightarrow \mathbf{z}_1^* = F(\mathbf{z}_0^*) \\ \mathbf{z}_{k_i+2} & \longrightarrow \mathbf{z}_2^* = F(\mathbf{z}_1^*) \\ \vdots & \quad \vdots \end{array} \tag{4}$$

The sequence $\{\mathbf{z}_0^*, \mathbf{z}_1^*, \ldots\}$ is part of the accumulation points of $\{\mathbf{z}_k\}$ and thus must be finite. Let $s \geq 0$ be the smallest integers such that $\mathbf{z}_{s+p}^* = \mathbf{z}_s^*$ for some positive integer p. Then by continuity, we have $\mathbf{z}_{s+p+1}^* = \mathbf{z}_{s+1}^*$, and so on. In this way, elements in $\{\mathbf{z}_0^*, \ldots \mathbf{z}_{s+p-1}^*\}$ are distinct and are the only accumulation points in the process of (4).

As these points are isolated, there exists $\epsilon > 0$ such that the spheres $N_\epsilon(\mathbf{z}_q^*)$ centered at $\mathbf{z}_q^*$ with radius ϵ, $q = 0, 1, \ldots, s+p-1$, are disjoint from each other. For each fixed integer t, all but finitely many points from this sequence $\{\mathbf{z}_{k_i+t}\}$ belong to $N_\epsilon(\mathbf{z}_q^*)$ with

$$q := \begin{cases} t, \text{ if } 0 \leq t \leq s, \\ s + ((t-s) \bmod p), \text{ if } s < t. \end{cases}$$

On the other hand, for a fixed $\mathbf{z}_{k_i}$ with sufficiently large i, write $\mathbf{z}_{k_j} = \mathbf{z}_{k_i+t_j}$ with $t_j := k_j - k_i$ for all $j > i$. Since $\mathbf{z}_{k_j} \in N_\epsilon(\mathbf{z}_0^*)$ when j is sufficiently large, we conclude that the two conditions

$$\begin{cases} s = 0, \\ (k_j - k_i) \bmod p = 0, \quad \text{for all sufficiently large } i, j \end{cases} \tag{5}$$

must hold simultaneously.

Suppose $\{\mathbf{z}_{\ell_j}\}$ is an arbitrary convergent subsequence of $\{\mathbf{z}_k\}$. For each ℓ_j, let k_{i_j} be one of the indices $\{k_i\}$ that is smaller than ℓ_j. Then $\mathbf{z}_{\ell_j} = \mathbf{z}_{k_{i_j}+(\ell_j - k_{i_j})}$ and hence all but finitely many elements in $\{\mathbf{z}_{\ell_j}\}$ must belong to one of these balls $N_\epsilon(\mathbf{z}_q^*)$. In this way, we have proved that all convergent subsequences of $\{\mathbf{z}_k\}$ satisfy (5).

If $p = 1$, then $q = 0$, the sequence $\{\mathbf{z}_k\}$ converges to $\mathbf{z}_0^*$. If $p > 1$, then $\{\mathbf{z}_k\}$ does not converge, but its elements for sufficiently large k must be distributed in such a way as residing alternately among $N_\epsilon(\mathbf{z}_q^*)$ in the order $q = 0, \ldots, p-1$.

It is informative to remark further on the three conditions required by Theorem 1 as follows:

a. *The sequence* $\{\mathbf{z}_k\}$ *being bounded.* This usually poses no additional burden because it is the prerequisite for convergence.
b. *The generating function* F *being continuous.* If F is given in analytic form, then its continuity can easily be checked. However, if F is given as a computational procedure, then cautions should be taken to ensure the continuity. For example, if $F(Y)$ refers to the orthogonal matrix $U(Y)$ in the singular value decomposition of the matrix $Y = U\Sigma V^\top$, then in theory U can be made to be continuously dependent on Y [4,34]. But if U is obtained by a certain SVD algorithm, then the signs of columns of $U(Y_1)$ may differ from those of $U(Y_2)$ even if Y_2 is close to Y_1, leading to discontinuous jumps in the numerical outcomes. An easy fix in the procedure is due.
c. *The accumulation points being finite and geometrically isolated.* This is the most demanding task. Even so, there are multiple avenues to tackle this task. For example, in many algorithm formulations the model (1) is actually a polynomial system in the variables $\mathbf{x}$ and $\mathbf{y}$. The notion of algebraic geometry might be used as a tool for arguing the finite cardinality and isolation of solutions.

The following lemma from the theory of parameter continuation [28, Theorem 7.1.1] is often useful for checking the last condition above.

Lemma 2 *Let* $P(\mathbf{z}; \mathbf{q})$ *be a system of* n *polynomials in variables* $\mathbf{z} \in \mathbb{C}^n$ *and parameters* $\mathbf{q} \in \mathbb{C}^m$. *Let* $\mathcal{N}(\mathbf{q})$ *denote the number of geometrically isolated solutions to* $P(\mathbf{z}; \mathbf{q}) = 0$ *over the algebraically closed complex space. Then,*

1. $\mathcal{N}(\mathbf{q})$ *is finite, and it is the same, say* $\mathcal{N}$, *for almost all* $\mathbf{q} \in \mathbb{C}^m$;
2. *For all* $\mathbf{q} \in \mathbb{C}^m$, $\mathcal{N}(\mathbf{q}) \leq \mathcal{N}$;
3. *The subset of* $\mathbb{C}^m$ *where* $\mathcal{N}(\mathbf{q}) = \mathcal{N}$ *is a Zariski open set. That is, the exceptional subset of* $\mathbf{q} \in \mathbb{C}^m$ *where* $\mathcal{N}(\mathbf{q}) < \mathcal{N}$ *is an affine algebraic set contained within an algebraic set of codimension one.*

Since $\mathbb{R}^n$ (indeed, the closure of any infinite subset) is Zariski dense in $\mathbb{C}^n$, the above statements hold for almost all parameters $\mathbf{q} \in \mathbb{R}^m$, except that the number of real-valued isolated solutions varies as a function of $\mathbf{q}$ and is no longer a constant. For our applications, we only need the fact that the real roots of a polynomial system are finite and geometrically isolated for generic $\mathbf{q}$.

The argument in Theorem 1 can be generalized to multi-level iterative schemes such as (1). Suppose that both functions f and g are continuous and that the sequences $\{\mathbf{x}_k\}$ and $\{\mathbf{y}_k\}$ generated are bounded and have finitely many isolated accumulation points, respectively. Then any convergent subsequence $\{\mathbf{y}_{k_i}\}$ will lead to a process

$$\begin{array}{rl} \mathbf{y}_{k_i} & \longrightarrow \mathbf{y}_0^* \\ \mathbf{x}_{k_i+1} & \longrightarrow \mathbf{x}_1^* = f(\mathbf{y}_0^*) \\ \mathbf{y}_{k_i+1} & \longrightarrow \mathbf{y}_1^* = g(\mathbf{x}_1^*) \\ \mathbf{x}_{k_i+2} & \longrightarrow \mathbf{x}_2^* = f(\mathbf{y}_1^*) \\ \vdots & \quad \vdots \end{array} \tag{6}$$

From this point on, an argument can be made to draw the same conclusion as in Theorem 1 for both $\{\mathbf{x}_k\}$ and $\{\mathbf{y}_k\}$ simultaneously. In this way, we may also interpret Theorem 1 as if $F = g \circ f$ applied to $\mathbf{y}$ for (1) and similarly for the general scheme (2).

An obvious condition for convergence is the exclusion of any possible cyclic behavior. This often can be accomplished if we know additional information such as some monotonicity associated with the iteration.

Corollary 1 *Suppose that the iteration (2) represents an alternating optimization mechanism for an objective function $h(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$. Under the same conditions of Theorem 1 where F denotes the transition function representing one complete sweep of the alternating procedure, the objective function h assumes the same value at all accumulation points.*

Proof By Theorem 1, we only need to consider the case when the sequence $\{\mathbf{z}_k\}$ has cyclic behavior. Without loss of generality, it suffices to consider the scheme (1) which involves only two variables $\mathbf{z} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$. To fix the idea, we assume that the alternating optimization is doing minimization. By the way the sequence $\{(\mathbf{x}_k^{(1)}, \mathbf{x}_k^{(2)})\}$ is generated, we should have the relationship

$$h(\mathbf{x}_{k+1}^{(1)}, \mathbf{x}_{k+1}^{(2)}) \leq h(\mathbf{x}_{k+1}^{(1)}, \mathbf{x}_k^{(2)}) \leq h(\mathbf{x}_k^{(1)}, \mathbf{x}_k^{(2)}).$$

Abbreviate $(\mathbf{x}_k^{(1)}, \mathbf{x}_k^{(2)})$ to $\mathbf{z}_k$. The sequence $\{h(\mathbf{z}_k)\}$ is monotone and must converge. If there are more than one isolated accumulation points, let $\mathbf{z}_0^*$ and $\mathbf{z}_1^*$ denote any two such points. The iterates $\{\mathbf{z}_k\}$ must visit arbitrarily diminishing vicinity of each accumulation point infinitely many times. Suppose $h(\mathbf{z}_0^*) < h(\mathbf{z}_1^*)$. Then there exists a neighborhood $N_\epsilon(\mathbf{z}_0^*)$ of $\mathbf{z}_0^*$ such that the iterates cannot possibly "return" to revisit the higher level $\mathbf{z}_1^*$ again once it has visited $N_\epsilon(\mathbf{z}_0^*)$ because of the non-ascending property mentioned earlier. Similarly, it cannot happen that $h(\mathbf{z}_0^*) > h(\mathbf{z}_1^*)$. Therefore, the objective function must assume the same value at all accumulation points.

Motivated by Corollary 1, we now impose some mild conditions of smoothness on the part of the optimization mechanism. The following observation is handy for applications.

Theorem 2 *Suppose that an alternating optimization method can be cast in form of (2). Write $\mathbf{z} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$ where $\mathbf{x}^{(\ell)} \in U^{(\ell)}$ and $U^{(\ell)} \subset \mathbb{R}^{I_\ell}$. Assume that*

1. *The conditions in Theorem 1 are satisfied where $F(\mathbf{z})$ denotes the transition function of one complete sweep of the alternating optimization, $\mathbf{z}_{k+1} = F(\mathbf{z}_k)$.*
2. *Each generating function $f^{(\ell)}$ represents the optimization mechanism in the ℓ-th direction, is continuously differentiable, and returns the unique global[1] minimizer $\mathbf{x}_{k+1}^{(\ell)}$ of the restricted objective function*
$$h_\ell(\mathbf{w}) := h(\mathbf{x}_{k+1}^{(1)}, \ldots, \mathbf{x}_{k+1}^{(\ell-1)}, \mathbf{w}, \mathbf{x}_k^{(\ell+1)}, \ldots, \mathbf{x}_k^{(n)}).$$
3. *The objective function $h(\mathbf{z})$ is second order continuously differentiable.*
4. *One of the accumulation points $\mathbf{z}_0^*$ of $\{\mathbf{z}_k\}$ is a local minimizer of $h(\mathbf{z})$ at which the Hessian $\nabla^2 h(\mathbf{z}_0^*)$ is symmetric and positive definite.*

Then the sequence $\{\mathbf{z}_k\}$ converges.

Proof Let $\{\mathbf{z}_{k_i}\}$ be an arbitrary convergent subsequence with limit point $\mathbf{z}_0^*$. We claim that the spectral radius $\rho(F'(\mathbf{z}_0^*))$ is strictly less than 1. If this claim is true, then there exists a neighborhood $N_\epsilon(\mathbf{z}_0^*)$ in which F is a contraction. That is, for any convergent subsequence $\{\mathbf{z}_{k_j}\} \subset N_\epsilon(\mathbf{z}_0^*)$, the subsequence

[1] What is really needed in the proof is the continuous differentiability of the transition function F. The uniqueness is to ascertain that $f^{(\ell)}$ unambiguously defines $\mathbf{x}_{k+1}^{(\ell)}$. So long as this map $f^{(\ell)}$ is well defined, the requirement of being a global minimizer is not essential.

$\{F(\mathbf{z}_{k_j})\}$ is also contained in $N_\epsilon(\mathbf{z}_0^*)$. Since $\{\mathbf{z}_{k_j+1}\}$ must also converge by the continuity of F, it converges to $\mathbf{z}_0^*$. By Lemma 1, we know the sequence $\{\mathbf{z}_k\}$ converges.

It only remains to prove that $\rho(F'(\mathbf{z}_0^*)) < 1$. It suffices to consider the case (1) only. The proof can be extended to the general case (2). The following argument is modified from the ideas in [3, Lemma 2]. Define $H : U^{(1)} \times U^{(2)} \times U^{(1)} \times U^{(2)} \to U^{(1)} \times U^{(2)}$ by

$$H(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}; \mathbf{y}^{(1)}, \mathbf{y}^{(2)}) := \begin{bmatrix} \frac{\partial h}{\partial \mathbf{x}^{(1)}}(\mathbf{x}^{(1)}, \mathbf{y}^{(2)}) \\ \frac{\partial h}{\partial \mathbf{x}^{(2)}}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \end{bmatrix}, \tag{7}$$

where the right hand side denotes the partial gradient of H with respect to the variables $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$, but evaluated at different points. Define also $G : U^{(1)} \times U^{(2)} \to U^{(1)} \times U^{(2)}$ by

$$G(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}) := H(F(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}); \mathbf{y}^{(1)}, \mathbf{y}^{(2)}). \tag{8}$$

Given any $(\mathbf{y}_k^{(1)}, \mathbf{y}_k^{(2)})$ near $\mathbf{z}_0^*$, observe that

$$G(\mathbf{y}_k^{(1)}, \mathbf{y}_k^{(2)}) = H(\mathbf{y}_{k+1}^{(1)}, \mathbf{y}_{k+1}^{(2)}, \mathbf{y}_k^{(1)}, \mathbf{y}_k^{(2)}) = \begin{bmatrix} \frac{\partial h}{\partial \mathbf{y}^{(1)}}(\mathbf{y}_{k+1}^{(1)}, \mathbf{y}_k^{(2)}) \\ \frac{\partial h}{\partial \mathbf{y}^{(2)}}(\mathbf{y}_{k+1}^{(1)}, \mathbf{y}_{k+1}^{(2)}) \end{bmatrix} = 0,$$

because $\mathbf{y}_{k+1}^{(1)}$ and $\mathbf{y}_{k+1}^{(2)}$ are the respective global minimizers of the restrictive objective functions h_1 and h_2. We see that $G \equiv 0$ in a neighborhood $N_\epsilon(\mathbf{z}_0^*)$. From (8), the evaluation of the Jacobian of G at $\mathbf{z}_0^*$ yields

$$\left(\frac{\partial H}{\partial(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})} \frac{\partial F}{\partial(\mathbf{y}^{(1)}, \mathbf{y}^{(2)})} + \frac{\partial H}{\partial(\mathbf{y}^{(1)}, \mathbf{y}^{(2)})} \right)\Bigg|_{\mathbf{z}_0^*} = 0, \tag{9}$$

where by (7) we have

$$\frac{\partial H}{\partial(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})} = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}^{(1)}}(\frac{\partial h}{\partial \mathbf{x}^{(1)}}) & 0 \\ \frac{\partial}{\partial \mathbf{x}^{(1)}}(\frac{\partial h}{\partial \mathbf{x}^{(2)}}) & \frac{\partial}{\partial \mathbf{x}^{(2)}}(\frac{\partial h}{\partial \mathbf{x}^{(2)}}) \end{bmatrix},$$

$$\frac{\partial H}{\partial(\mathbf{y}^{(1)}, \mathbf{y}^{(2)})} = \begin{bmatrix} 0 & \frac{\partial}{\partial \mathbf{x}^{(2)}}(\frac{\partial h}{\partial \mathbf{x}^{(1)}}) \\ 0 & 0 \end{bmatrix}.$$

Note that the above two matrices make up

$$\nabla^2 h(\mathbf{z}_0^*) = \left(\frac{\partial H}{\partial(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})} + \frac{\partial H}{\partial(\mathbf{y}^{(1)}, \mathbf{y}^{(2)})} \right)\Bigg|_{\mathbf{z}_0^*},$$

which is assumed to be symmetric and positive definite. It follows from (9) that

$$F'(\mathbf{z}_0^*) = -\left(\frac{\partial H}{\partial(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})}\Bigg|_{\mathbf{z}_0^*} \right)^{-1} \left(\frac{\partial H}{\partial(\mathbf{y}^{(1)}, \mathbf{y}^{(2)})}\Bigg|_{\mathbf{z}_0^*} \right) \tag{10}$$

is well defined. Furthermore, we see in (10) that $F'(\mathbf{z}_0^*)$ is of the form $-(D-L)^{-1}U$ which is precisely the iteration matrix if the (block) Gauss-Seidel scheme is applied to solving a linear equation where the coefficient matrix A is split as $A = D - L - U$ [27, Theorem 7.1.9]. Since the Gauss-Seidel method converges when A is symmetric and positive definite, we know that $\rho(F'(\mathbf{z}_0^*)) < 1$.

Alternating optimization, or more generally alternating direction, is not usually the best approach for solving the underlying problem. However, swapping one complicated problem of many variables with a sequence of simpler problems each of which handles and adjusts one subset of variables a time can sometimes be implemented more easily and offer computational convenience. The above theory outlines a basic convergence analysis framework for these types of alternating direction iterations. In the remaining portion of this paper, we discuss some interesting applications. Some of the convergence results are new.

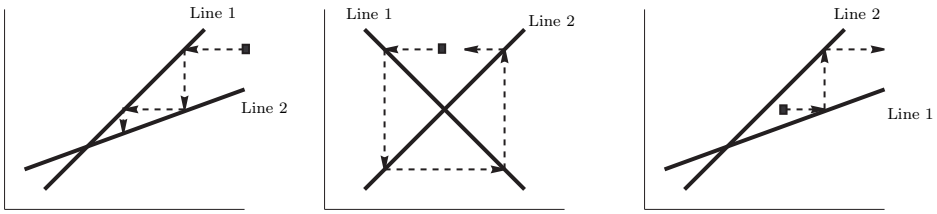

Fig. 1 Converging, cyclic, and diverging behavior of Gauss-Seidel iterations in $\mathbb{R}^2$.

3 Applications to some known cases

The convergence behavior of examples in this section is well understood in the literature. Certainly we are not trying to reinvent the wheels. Rather, we use these known facts to demonstrate the subtleties in dealing with convergence when some of the conditions mentioned in the preceding section are not met. On the other hand, we also demonstrate that our framework offers an alternative and unified proof of convergence which is much simpler than some of those already done in the literature.

3.1 The Gauss-Seidel method for solving a system of linear equations

The classical Gauss-Seidel iteration scheme is of the form (2). It is well known that the method applied to the linear system $A\mathbf{x} = \mathbf{b}$ with non-zero elements on the diagonals does not always produce a convergent result. Convergence is guaranteed only in a few cases such as the matrix A being diagonally dominant or being symmetric and positive definite. In the event that the Gauss-Seidel method fails to converge for a given A, what has happened is that either the iterates become unbounded or the iterates go cyclically, as has been characterized in Theorem 1. The directions of variables are alternated by satisfying one linear equation a time. See Figure 1 for a graphical interpretation of the Gauss-Seidel method applied to a 2-dimensional problem. The scheme itself does not contain any type of optimization in its iteration.

3.2 The power method for finding the dominant eigenvector

Given a matrix $A \in \mathbb{R}^{n\times n}$ and an initial unit vector $\mathbf{y}_0 \in \mathbb{R}^n$, the power method

$$\begin{cases} \mathbf{x}_{k+1} = A\mathbf{y}_k, \\ \mathbf{y}_{k+1} = \frac{\mathbf{x}_{k+1}}{\|\mathbf{x}_{k+1}\|_\infty}, \end{cases} \tag{11}$$

is in the form of (1). The sequences $\{\mathbf{x}_k\}$ and $\{\mathbf{y}_k\}$ are clearly bounded. The functions $f^{(\ell)}$, $\ell = 1, 2$, on the right side of (11) are clearly continuous. Excluding the extraneous zero solution after scaling the second equation by a multiplier $\|\mathbf{x}_{k+1}\|_\infty$, the entire system can be regarded as a polynomial system depending on the parameter A. By Lemma 2, we know that for almost all matrices $A \in \mathbb{R}^{n\times n}$, the stationary points are finite and isolated. By Theorem 1, we conclude that the iterates generated by the power method converge for a generic A. In numerical linear algebra, we know even more specifics when the method fails to converge, e.g., when A has multiple dominant eigenvalues, in which case the matrix A has a multi-dimensional eigenspace and the system (11) has non-isolated stationary points. See also Section 3.5 for more detailed discussion from the perspective of the high-order power method.

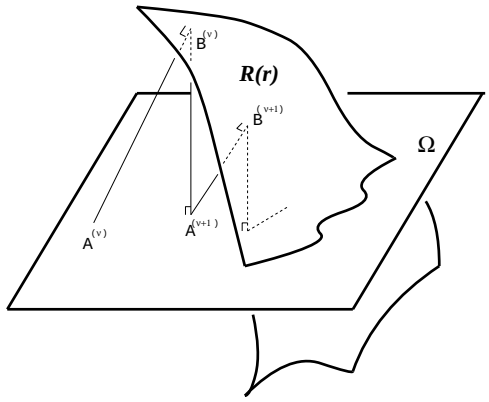

Fig. 2 Alternating projections between lower rank matrices and structured matrices

3.3 The alternating least squares method for computing the QR decomposition

There are efficient algorithms for computing the fundamentally important QR decomposition of a given matrix $A \in \mathbb{R}^{m\times n}$. Surely it is of little value to try to find this decomposition by the alternating least squares approach

$$\begin{cases} R_{k+1} := \underset{R = \text{ upper triangular}}{\arg\min} \|A - Q_k R\|_F, \\ Q_{k+1} := \underset{Q^\top Q = I_n}{\arg\min} \|A - Q R_{k+1}\|_F. \end{cases} \tag{12}$$

Despite its inefficiency, however, the scheme (12) is theoretically doable. Indeed, it can be argued that R_{k+1} is the upper triangular part of the matrix $Q_k^\top A$ and Q_{k+1} is exactly the orthogonal portion in the polar decomposition of $AR_{k+1}^\top$ (which is more expensive than the QR decomposition itself). By construction, the objective values $\|A - Q_k R_k\|_F$ descend and converge, but possibly to a nonzero value. Clearly, the sequences $\{Q_k\}$ and $\{R_k\}$ are bounded and the abstract functions defining them are continuous. The stationary points must satisfy the optimality conditions

$$\begin{cases} R = \mathsf{triu}(Q^\top A), \\ Q^\top A R^\top = R A^\top Q, \end{cases}$$

which is a linear polynomial system in Q with A as the parameter and, by Lemma 2, has finitely many isolated solutions for generic $A \in \mathbb{R}^{m\times n}$. The conditions in Theorem 2 are satisfied, so the iterates $\{Q_k\}$ and $\{R_k\}$ do converge, even though not necessarily they converge to the QR decomposition of A.

3.4 The alternating projection method for finding structured low rank matrices

Let $\mathcal{R}(r)$ denote the set of all rank r matrices and Ω the set of matrices with a prescribed structure, say, Toeplitz or Hankel matrices. Then the desired set of structured rank r matrices can be regarded as the intersection of these two sets. To find a structured low rank matrix, if exist, the idea of alternating projections between these two sets can be employed [5–7]. The process is to satisfy the rank constraint and the structural constraint alternately while the distance in between is being reduced. The geometry of lift and project is depicted in Figure 2. The procedures outlined in Algorithm 1 obviously fits the basic model (1) where both actions of lifting and the projection are continuous. Since $\mathcal{R}(r)$ is not

Algorithm 1 (Lift-and-project algorithm.)

Require: Given an arbitrary $A^{(0)} = A \in \Omega$
Ensure: A pair of matrices that locally minimizes the distance between $\mathcal{R}(r)$ and Ω

1: **repeat**
2: **lift:** Compute the rank r matrix $B^{(\nu)}$ in $\mathcal{R}(r)$ that is nearest to $A^{(\nu)}$.
3: **project:** Compute the projection $A^{(\nu+1)}$ of $B^{(\nu)}$ onto the subspace Ω.
4: **until** the sequence $\{A^{(\nu)}\}$ meets stopping criteria

convex, the iteration might stagnate back and forth between $\mathcal{R}(r)$ and Ω. In that case, an intersection has not been found, but still the iterates converge to a locally nearest location between the two geometric entities by our theory.

3.5 Best rank-one tensor approximation

A tensor of the form

$$\mathbf{u}^{(1)} \circ \ldots \circ \mathbf{u}^{(k)} := [u_{i_1}^{(1)} \ldots u_{i_k}^{(k)}],$$

where elements are the products of entries from vectors $\mathbf{u}^{(j)} \in \mathbb{R}^{I_j}$, $j = 1, \ldots, k$, is said to be of rank one. Given a real-valued tensor of order k

$$T = [\tau_{i_1,\ldots,i_k}] \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_k},$$

one of the most challenging tasks is to find its low-rank approximation in either the general Tucker decomposition form [10,11,29]

$$T \approx \sum_{j_1,j_2,\ldots,j_k} \beta_{j_1,j_2,\ldots,j_k} \mathbf{u}_{j_1}^{(1)} \circ \ldots \circ \mathbf{u}_{j_k}^{(k)} \tag{13}$$

or the CANDECOMP/PARAFAC (CP) decomposition form [12,14,17,20]

$$T \approx \sum_{j} \lambda_j \mathbf{u}_j^{(1)} \circ \ldots \circ \mathbf{u}_j^{(k)}, \tag{14}$$

where the summations involve only a few rank-1 tensors and each column vector is of unit length. It is known that tensors beyond matrices can fail to have best low rank approximations, with the notable exception that the best rank-one approximation always exists for tensors of any order and that when a certain condition of orthogonality is imposed. We discuss the rank-one approximation first. The Tucker nearest problem will be discussed in Section 4.1.

The most popular approach for the best rank-one approximation is the notion of alternating least squares method. The procedures are described in Algorithm 2, where the subscript $\cdot_{[p]}$ indicates the quantities resulting from the p-th iteration, $\widehat{\mathbf{u}}^{(\ell)}$ means to exclude this vector from the list, and

$$T \circledast_\ell S := [\langle \tau_{:,\nu_\ell,:}, S \rangle] \in \mathbb{R}^{I_\ell}, \quad \nu_\ell = 1, \ldots, I_\ell, \tag{15}$$

with $\tau_{:,\nu_\ell,:}$ denoting the ν_ℓ-th "slice" of the tensor T in the ℓ-th direction and $\langle \cdot, \cdot \rangle$ the Frobenius inner product generalized to multi-dimensional arrays.

While the limiting behavior of the objective values $\{\lambda_{[p]}^{(\ell)}\}$ is easy to understand, it has taken tremendous effort to prove the convergence of the iterates $\{\mathbf{u}_{[p]}^{(\ell)}\}$ themselves [30,33]. We now apply our theory to Algorithm 2 to demonstrate how the convergence can be argued in a quick and convenient way.

Algorithm 2 (High-order power method.)

Require: A generic order-k tensor T and k unit vectors $\mathbf{u}_{[0]}^{(1)} \in \mathbb{R}^{I_1}, \ldots, \mathbf{u}_{[0]}^{(k)} \in \mathbb{R}^{I_k}$,
Ensure: A local best rank-1 approximation to T

1: **for** $p = 0, 1, \cdots,$ **do**
2: **for** $\ell = 1, 2, \cdots, k$ **do**
3: $\mathbf{u}_{[p+1]}^{(\ell)} = T \circledast_\ell (\mathbf{u}_{[p+1]}^{(1)} \circ \ldots \circ \mathbf{u}_{[p+1]}^{(\ell-1)} \circ \widehat{\mathbf{u}}^{(\ell)} \circ \mathbf{u}_{[p]}^{(\ell+1)} \ldots \circ \mathbf{u}_{[p]}^{(k)})$
4: $\lambda_{[p+1]}^{(\ell)} := \|\mathbf{u}_{[p+1]}^{(\ell)}\|_2$
5: $\mathbf{u}_{[p+1]}^{(\ell)} := \frac{\mathbf{u}_{[p+1]}^{(\ell)}}{\lambda_{[p+1]}^{(\ell)}}$
6: **end for**
7: **end for**

First, the definition of $\mathbf{u}_{[p+1]}^{(\ell)}$ in Line 3 followed by Line 5 gives rise to precisely the unique global maximizer of the function

$$\lambda_{[p+1]}^{(\ell)}(\mathbf{w}) := \langle T, \mathbf{u}_{[p+1]}^{(1)} \circ \ldots \circ \mathbf{u}_{[p+1]}^{(\ell-1)} \circ \mathbf{w} \circ \mathbf{u}_{[p]}^{(\ell+1)} \ldots \circ \mathbf{u}_{[p]}^{(k)} \rangle$$

which is the restriction of the objective function

$$\lambda(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}) = \langle T, \mathbf{u}^{(1)} \circ \ldots \circ \mathbf{u}^{(k)} \rangle \tag{16}$$

to the ℓ-th direction subject to the constraint of unit length. As a polynomial in variables $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}$, the smoothness of λ and the associated $\lambda_{[p+1]}^{(\ell)}$ is guaranteed. The first order optimality condition for a stationary point of (16) is to satisfy the system of $\sum_{\ell=1}^{k} I_\ell$ polynomials [23,33]

$$T \circledast_\ell (\mathbf{u}^{(1)} \circ \ldots \circ \widehat{\mathbf{u}}^{(\ell)} \circ \ldots \circ \mathbf{u}^{(k)}) = \langle T, \mathbf{u}^{(1)} \circ \ldots \circ \mathbf{u}^{(k)} \rangle \mathbf{u}^{(\ell)}, \quad \ell = 1, \ldots, k, \tag{17}$$

which, by Lemma 2, contains only geometrically isolated solutions for a generic tensor T. Conditions in Theorem 1 are satisfied generically. It is easy to see that the sequence $\{\lambda(\mathbf{u}_{[p]}^{(1)}, \ldots, \mathbf{u}_{[p]}^{(k)})\}$ is monotone non-decreasing. Assuming the generic condition that the Hessian of λ at such a local maximizer is negative definite, then the convergence of the iterates $\{(\mathbf{u}_{[p]}^{(1)}, \ldots, \mathbf{u}_{[p]}^{(k)})\}$ is ensured by Theorem 2.

4 Applications to some new problems

In this section, we apply our theory to two important yet challenging problems in the field — the Tucker nearest problem and the structured Kronecker approximation problem. While numerical algorithms have been proposed and used in practice, we find little discussion of convergence analysis in the literature. This is probably due to the fact that the algorithms usually involve complex algebraic manipulations. Nonetheless, our framework requires fairly mild conditions on these manipulations. We can explain the convergence.

For the ease of later reference, we first introduce the notion of orthogonality which will appear in both problems. Let $\mathscr{S}(p, q)$ denote the Stiefel manifold of matrices in $\mathbb{R}^{p \times q}$ with orthonormal columns and $\mathbb{I}_q$ the identity matrix in $\mathbb{R}^{q \times q}$. The following lemma is essentially the well known polar decomposition [15,16,19], yet its view as the normal bundle of an element Q on $\mathscr{S}(p, q)$ is useful for the subsequent discussion [8].

Lemma 3 *Given a matrix $Q \in \mathscr{S}(p, q)$, then a matrix $Z \in \mathbb{R}^{p \times q}$ whose orthogonal projection to $\mathscr{S}(p, q)$ is precisely Q if and only if $Q^\top Z$ is symmetric.*

Proof Let $Q_\perp$ denote the matrix in $\mathscr{S}(p, p-q)$ so that the augmented matrix $[Q, Q_\perp]$ is orthogonal. It is easy to see that the tangent space $\mathscr{T}_Q\mathscr{S}(p,q)$ at $Q \in \mathscr{S}(p,q)$ is made of matrices in the form

$$H = QK + Q_\perp Q_\perp^\top W,$$

where $K \in \mathbb{R}^{q\times q}$ is skew-symmetric and $W \in \mathbb{R}^{p\times q}$ is arbitrary. For the vector $Z - Q$ to be perpendicular to the surface $\mathscr{S}(p,q)$, it must be such that

$$\mathbf{Proj}_{\mathscr{T}_Q\mathscr{S}(p,q)}(Z-Q) = Q\frac{Q^\top(Z-Q) - (Z-Q)^\top Q}{2} + Q_\perp Q_\perp^\top (Z-Q) = 0. \tag{18}$$

Note that the two terms in the middle equation of (18) are mutually orthogonal. Therefore, each term must be zero by itself. Upon simplification, we see that $Z - Q$ is perpendicular to $\mathscr{S}(p,q)$ if and only if

$$\begin{cases} Q^\top Z = Z^\top Q, \\ Q_\perp^\top Z = 0. \end{cases} \tag{19}$$

Given Q, (19) is a homogeneous linear system of $pq - \frac{q(q+1)}{2}$ independent equations in pq unknowns of Z. So the solutions form a subspace of dimension $\frac{q(q+1)}{2}$. Indeed, if we write the columns of $Z \in \mathbb{R}^{p\times q}$ in terms of the orthonormal basis

$$Z = QS + Q_\perp T,$$

where $S \in \mathbb{R}^{q\times q}$ and $T \in \mathbb{R}^{(p-q)\times q}$, then Z is a solution to (19) if and only if $T = Q_\perp^\top Z = 0$ and $S = Q^\top Z$ is symmetric.

In the above lemma, we look up from a given $Q \in \mathscr{S}(p,q)$ for its normal bundle in $\mathbb{R}^{p\times q}$. Now we look down from a given $Z \in \mathbb{R}^{p\times q}$ for its projection onto $\mathscr{S}(p,q)$.

Corollary 2 *Given an arbitrary $Z \in \mathbb{R}^{p\times q}$, suppose that $Z = UP$ is the polar decomposition of Z where $U \in \mathscr{S}(p,q)$ and $P \in \mathbb{R}^{q\times q}$ is symmetric and positive semi-definite. Then U is the projection of Z onto $\mathscr{S}(p,q)$ and is the nearest matrix in $\mathscr{S}(p,q)$ to Z.*

In the polar decomposition, we stress that the symmetric matrix $P = U^\top Z$ is always unique, but U is unique only if Z is of full column rank.

4.1 Tucker nearest problem

Given the rank parameter $\mathbf{r} = (r_1, \ldots, r_k)$, an order-k tensor in the form

$$A = \sum_{j_1=1}^{r_1} \cdots \sum_{j_k=1}^{r_k} \beta_{j_1,\ldots,j_k} \mathbf{v}_{j_1}^{(1)} \circ \ldots \circ \mathbf{v}_{j_k}^{(k)} \in \mathbb{R}^{I_1\times\ldots\times I_k} \tag{20}$$

with orthonormal vectors $\mathbf{v}_{j_\ell}^{(\ell)} \in \mathbb{R}^{I_\ell}$ is said to be in the Tucker format with core tensor

$$\boldsymbol{\beta} := [\beta_{j_1,\ldots,j_k}] \in \mathbb{R}^{r_1\times\ldots\times r_k}. \tag{21}$$

If we assemble the orthonormal vectors into factor matrices by denoting

$$V^{(\ell)} := [\mathbf{v}_1^{(\ell)}, \ldots, \mathbf{v}_{r_\ell}^{(\ell)}] \in \mathbb{R}^{I_\ell\times r_\ell}, \quad \ell = 1, \ldots, k, \tag{22}$$

then $V^{(\ell)} \in \mathscr{S}(I_\ell, r_\ell)$ and the tensor A in (20) can be written as

$$A = \boldsymbol{\beta} \times_1 V^{(1)} \times_2 V^{(2)} \times_3 \ldots \times_k V^{(k)}, \tag{23}$$

where $\times_d$ denotes the mode-d product[2] [21]. Given an order-k tensor $T \in \mathbb{R}^{I_1 \times \ldots \times I_k}$, the Tucker nearest problem is to find a tenor in the Tucker form (23) with a fixed rank parameter $\mathbf{r}$ such that

$$\widetilde{h}(\boldsymbol{\beta}, V^{(1)}, \ldots, V^{(k)}) := \|\boldsymbol{\beta} \times_1 V^{(1)} \times_2 V^{(2)} \times_3 \ldots \times_k V^{(k)} - T\|_F \tag{24}$$

is minimized.

For an order-k tensor $T \in \mathbb{R}^{I_1 \times \ldots \times I_k}$, let $\mathbf{vec}(T)$ denote the linear array where the entry $\tau_{i_1,\ldots,i_k}$ of T is saved at the location

$$i_1 + \sum_{s=2}^{k} (i_s - 1) \prod_{t=1}^{s-1} I_t \tag{25}$$

of the array. Then it can be verified that (23) is equivalent to [1, Formula (12)]

$$\mathbf{vec}(A) = (V^{(k)} \otimes \ldots \otimes V^{(1)}) \mathbf{vec}(\boldsymbol{\beta}), \tag{26}$$

where $\otimes$ stands for the Kronecker product. The expression above sheds an important insight — entries in $\mathbf{vec}(\boldsymbol{\beta})$ are the coordinates of $\mathbf{vec}(A)$ in terms of the orthonormal columns of $V^{(k)} \otimes \ldots \otimes V^{(1)}$, i.e.,

$$\mathbf{vec}(\boldsymbol{\beta}) = (V^{(k)} \otimes \ldots \otimes V^{(1)})^\top \mathbf{vec}(A). \tag{27}$$

Therefore, given fixed matrices $V^{(\ell)} \in \mathscr{S}(I_\ell, r_\ell)$, $\ell = 1, \ldots, k$, the minimizer $\boldsymbol{\beta}$ in (24) is given by the projection of $\mathbf{vec}(T)$ onto the column space of $V^{(k)} \otimes \ldots \otimes V^{(1)}$, or equivalently,

$$\boldsymbol{\beta} := T \times_1 {V^{(1)}}^\top \times_2 {V^{(2)}}^\top \times_3 \ldots \times_k {V^{(k)}}^\top \in \mathbb{R}^{r_1 \times \ldots \times r_k}. \tag{28}$$

In this way, the Tucker nearest problem is equivalent to the problem of maximizing the Frobenius norm of the tensor

$$\pi(V^{(1)}, \ldots, V^{(k)}) := T \times_1 {V^{(1)}}^\top \times_2 {V^{(2)}}^\top \times_3 \ldots \times_k {V^{(k)}}^\top, \tag{29}$$

subject to the constraint that $V^{(\ell)} \in \mathscr{S}(I_\ell, r_\ell)$, $\ell = 1, \ldots, k$.

The relationship (28) can further be expressed in terms of the mode-d unfolding [1, Formula (11)]

$$\boldsymbol{\beta}_{(d)} = {V^{(d)}}^\top \underbrace{T_{(d)}(V^{(k)} \otimes \ldots \otimes V^{(d+1)} \otimes V^{(d-1)} \otimes \ldots \otimes V^{(1)})}_{Y_{(d)}}, \quad d = 1, \ldots k, \tag{30}$$

where the mode-d unfolding $T_{(d)}$ is simply a rearrangement of T into a matrix of size $I_d \times \prod_{\ell \neq d} I_\ell$ by assigning the element $(T_{(d)})_{i_d, j} := \tau_{i_1,\ldots,i_k}$ with $j = 1 + \sum_{s=1, s\neq d}^{k} (i_s - 1) \prod_{t=1}^{s-1} I_t$. Likewise, $\boldsymbol{\beta}_{(d)}$ is an unfolding of size $r_d \times \prod_{\ell \neq d} r_\ell$. Taking advantage of the form (30) by modifying one factor matrix $V^{(\ell)}$ a time via the singular value decomposition, Algorithm 3 therefore has been proposed in the field as a way for tackling the Tucker nearest problem. By construction, we also know that

$$\begin{aligned} \lambda_{[p]} := \|\pi(V^{(1)}_{[p]}, \ldots, V^{(k)}_{[p]})\|_F &\leq \lambda^{(1)}_{[p+1]} \leq \lambda^{(2)}_{[p+1]} \leq \ldots \leq \lambda^{(k)}_{[p+1]} \\ &= \|\pi(V^{(1)}_{[p+1]}, \ldots, V^{(k)}_{[p+1]})\|_F, \end{aligned} \tag{31}$$

so the convergence of scalars $\{\lambda_{[p]}\}$ is clear. Thus far, however, we have not seen any proof of convergence for the iterates $\{(V^{(1)}_{[p]}, \ldots, V^{(k)}_{[p]})\}$ in the literature. Using our framework, we can establish the convergence as follows.

[2] Given an order-k tensor $T \in \mathbb{R}^{I_1 \times \ldots\ldots \times I_d \times \ldots \times I_k}$ and a matrix $M \in \mathbb{R}^{m \times I_d}$, the mode-$d$ product $\Theta = T \times_d M$ is defined to be the tensor in $\mathbb{R}^{I_1 \times \ldots \times I_{d-1} \times m \times I_{d+1} \times \ldots \times I_k}$ with element $f_{i_1,\ldots,i_{d-1},t,i_{d+1},\ldots i_k} := \sum_{s=1}^{I_d} m_{t,s} \tau_{i_1,\ldots,i_{d-1},s,i_{d+1},\ldots i_k}$.

Algorithm 3 (HOSVD method for Tucker nearest problem.)

Require: A generic order-k tensor T, a fixed rank parameter $\mathbf{r}$, and k initial matrix $V_{[0]}^{(\ell)} \in \mathbb{R}^{I_\ell \times r_\ell}$ with orthonormal columns,

Ensure: A local best Tucker approximation to T

1: **for** $p = 0, 1, \cdots$, **do**
2: **for** $\ell = 1, 2, \cdots, k$ **do**
3: $B_{[p+1]}^{(\ell)} := T_{(\ell)}(V_{[p]}^{(k)} \otimes \ldots \otimes V_{[p]}^{(\ell+1)} \otimes V_{[p+1]}^{(\ell-1)} \otimes \ldots \otimes V_{[p+1]}^{(1)})$ {Of size $I_\ell \times \prod_{j=1, j\neq\ell}^{k} r_j$.}
4: $[U, S, \tilde{}] = \mathsf{svds}(B_{[p+1]}^{(\ell)}, r_\ell)$ {Compute the largest r_ℓ singular values and left singular vectors.}
5: $V_{[p+1]}^{(\ell)} := U$
6: $\lambda_{[p+1]}^{(\ell)} = \|S\|_F$
7: **end for**
8: **end for**

Without loss of generality, consider the objective function to be maximized as

$$h(V^{(1)}, \ldots, V^{(k)}) = \frac{1}{2}\|\pi(V^{(1)}, \ldots, V^{(k)})\|_F^2 = \frac{1}{2}\langle {V^{(d)}}^\top \Upsilon_{(d)}, {V^{(d)}}^\top \Upsilon_{(d)} \rangle \tag{32}$$

which, as indicated in (30), has the same value $\frac{\|\boldsymbol{\beta}\|_F^2}{2}$ for all $d = 1, \ldots, k$. Clearly, h is secondly order continuous differentiable. The definition of $V_{[p+1]}^{(\ell)}$ at Line 5 is the unique global maximizer of the restricted function

$$h_\ell(W) := \frac{1}{2}\|\pi(V_{[p+1]}^{(1)}, \ldots V_{[p+1]}^{(\ell-1)}, W, V_{[p]}^{(\ell+1)}, \ldots, V_{[p]}^{(k)})\|_F^2, \tag{33}$$

subject to the constraint that $W \in \mathscr{S}(I_\ell, r_\ell)$, so Algorithm 3 is an ALS algorithm.

To apply our framework, we need to check out two additional conditions. First, the partial gradient of h with respect to a general $V^{(d)}$ is given by

$$\nabla^{(d)} h(V^{(d)}) := \frac{\partial h}{\partial V^{(d)}} = \Upsilon_{(d)} \Upsilon_{(d)}^\top V^{(d)}, \quad d = 1, \ldots, k. \tag{34}$$

At a stationary point, the projection of $\nabla^{(d)} h(V^{(d)})$ onto the tangent space of $\mathscr{S}(I_d, r_d)$ is zero, implying that

$$\Upsilon_{(d)} \Upsilon_{(d)}^\top V^{(d)} = V^{(d)} {V^{(d)}}^\top \Upsilon_{(d)} \Upsilon_{(d)}^\top V^{(d)}, \quad d = 1, \ldots, k. \tag{35}$$

In other words, the stationary points of the objective function (32) are solutions to a system of $\sum_{d=1}^{k} I_d r_d$ polynomials (35) that is parameterized by T. By Lemma 2, we conclude that for almost all tensors, the accumulation points of Algorithm 3 are finite and geometrically isolated.

Second, each of the constraint $V^{(\ell)} \in \mathscr{S}(I_\ell, r_\ell)$, $\ell = 1, \ldots, k$, is a compact set. The local maximizer for h does exist. The Hessian of h in (32), which depends on T, at its local maximizer is necessarily negative semi-definite. Furthermore, positive definite matrices form an open set whose boundaries consist of positive semi-definite matrices which resides on a submanifold of codimension 1. A small perturbation can easily disrupt the semi-definiteness. We may therefore assume that for almost all tensors, the Hessian of h at one of the stationary point is symmetric and positive definite.

By now, all conditions in Theorem 2 are satisfied. To our knowledge, the following result is new.

Theorem 3 *For almost all order-k tensor T, the iterates $\{(V_{[p]}^{(1)}, \ldots, V_{[p]}^{(k)})\}$ generated by Algorithm 3 converge to a local solution of the Tucker nearest problem.*

4.2 Structured Kronecker approximation

Given $A \in \mathbb{R}^{m\times n}$ with $m = m_1m_2$ and $n = n_1n_2$ and a small enough but fixed integer r, the Kronecker approximation problem concerns finding matrices $B_i \in \mathbb{R}^{m_1\times n_1}$ and $C_i \in \mathbb{R}^{m_2\times n_2}$ such that the objective function

$$\phi_A(B_1,\ldots,B_r,C_1,\ldots,C_r) = \|A - \sum_{i=1}^{r} B_i \otimes C_i\|_F^2, \tag{36}$$

is minimized [26]. The problem is equivalent to a rank-r approximation problem [32]

$$\|A - \sum_{i=1}^{r} B_i \otimes C_i\|_F = \|\mathscr{R}(A) - \sum_{i=1}^{r} \mathbf{vec}(B_i)\mathbf{vec}(C_i)^\top\|_F, \tag{37}$$

where $\mathscr{R}(A) \in \mathbb{R}^{m_1n_1\times m_2n_2}$ is a rearrangement of A as

$$\mathscr{R}(A) := \begin{bmatrix} \mathbf{vec}(A_{1,1})^\top \\ \mathbf{vec}(A_{2,1})^\top \\ \vdots \\ \mathbf{vec}(A_{m_1,n_1})^\top \end{bmatrix},$$

if A is partitioned as a $m_1 \times n_1$ block matrix with blocks $A_{ij} \in \mathbb{R}^{m_2\times n_2}$,

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1,n_1} \\ A_{21} & A_{22} & \cdots & A_{2,n_1} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m_1,1} & A_{m_1,2} & \cdots & A_{m_1,n_1} \end{bmatrix}.$$

The Kronecker approximation problem (36) therefore can be solved effectively by using the truncated singular value decomposition.

It is important to note that the Kronecker product often inherits structures from its factors. For example, the following properties are listed in [31].

$$\text{If } B \text{ and } C \text{ are } \left\{\begin{array}{l} \text{nonsingular} \\ \text{lower(upper) triangular} \\ \text{banded} \\ \text{symmetric} \\ \text{positive definite} \\ \text{stochastic} \\ \text{Toeplitz} \\ \text{permutations} \\ \text{orthogonal} \end{array}\right\}, \text{then } B\otimes C \text{ is } \left\{\begin{array}{l} \text{nonsingular} \\ \text{lower(upper) triangular} \\ \text{banded} \\ \text{symmetric} \\ \text{positive definite} \\ \text{stochastic} \\ \text{Toeplitz} \\ \text{permutations} \\ \text{orthogonal} \end{array}\right\}.$$

Also, with respect to factorizations, the LU-with-partial-pivoting, Cholesky, and QR factorizations of $B \otimes C$ merely require the corresponding factorizations of B and C. An interesting question about the converse then arises, which we refer to as the structured Kronecker approximation problem. Let $\Omega_B \subset \mathbb{R}^{m_1\times n_1}$ and $\Omega_C \subset \mathbb{R}^{m_2\times n_2}$ denote the subsets of desired structures of factors, respectively. How should the approximation (36) be accomplished if it is expected that $B_i \in \Omega_B$ and $C_i \in \Omega_C$, even if the given A is not structured?

In what follows, we consider only the case $r = 1$. Generalizations to general r is possible but with tedious manipulations. See, for example, the work in [18] for the block Toeplitz structure. Once

Algorithm 4 (ALS method for structured Kronecker approximation.)

Require: A generic matrix $A \in \mathbb{R}^{m_1 m_2 \times n_1 n_2}$, two specific structures Ω_B and Ω_C, an initial matrix $C_0 \in \Omega_C$,
Ensure: A local best structured Kronecker approximation to A.

1: **for** $k = 0, 1, \cdots$, **do**
2: $\quad B_{k+1} = \arg\min_{B \in \Omega_B} \|A - B \otimes C_k\|_F$
3: $\quad C_{k+1} = \arg\min_{C \in \Omega_C} \|A - B_{k+1} \otimes C\|_F$
4: **end for**

the procedure such that the generating function is specified, we think that it is possible that our framework is still applicable.

For the case $r = 1$, the following result naturally defines an alternating procedure. In [32, Theorem 4.1], the result can be interpreted as the power method applied to $\mathscr{R}(A)$ for finding the left and right singular vectors associated with its largest singular value.

Lemma 4 *Let $A \in \mathbb{R}^{m \times n}$ with $m = m_1 m_2$ and $n = n_1 n_2$ be given.*

1. *Suppose $C \in \mathbb{R}^{m_2 \times n_2}$ is fixed, then the matrix $B \in \mathbb{R}^{m_1 \times n_1}$ defined by*

$$b_{ij} := \frac{\langle A_{ij}, C \rangle}{\langle C, C \rangle}, \quad 1 \le i \le m_1, \quad 1 \le j \le n_1, \tag{38}$$

   minimizes $\|A - B \otimes C\|_F$.
2. *Suppose $B \in \mathbb{R}^{m_1 \times n_1}$ is fixed, then the matrix $C \in \mathbb{R}^{m_2 \times n_2}$ defined by*

$$c_{ij} := \frac{\langle \widetilde{A}_{ij}, B \rangle}{\langle B, B \rangle}, \quad 1 \le i \le m_2, \quad 1 \le j \le n_2, \tag{39}$$

   where $\widetilde{A}_{ij} = A(i : m_2 : m, j : n_2 : n) \in \mathbb{R}^{m_1 \times n_1}$, minimizes $\|A - B \otimes C\|_F$.

The above lemma can then be exploited to answer a few structured approximation problems, provided that A is similarly structured. We mention, for example, the cases that

$$\text{If } A \text{ and } B \text{ are } \left\{ \begin{array}{l} \text{nonnegative} \\ \text{symmetric} \\ \text{positive definite} \end{array} \right\}, \text{then the minimizer } C \text{ of } \|A - B \otimes C\|_F \text{ is } \left\{ \begin{array}{l} \text{nonnegative} \\ \text{symmetric} \\ \text{positive definite} \end{array} \right\}.$$

For other structures, including the case that the given A does not have any structure at all, the formulas in Lemma 4 does not preserve the structures in general. Some other numerical procedures are needed.

The prototypical ALS procedure proposed in Algorithm 4 is a plausible procedure to tackle the structured Kronecker approximation problem, provided the structured least squares subproblems at Lines 2 and 3 can be resolved. Even so, the nonlinear nature of the Kronecker product would make a formal proof of convergence of the iterates for the general case challenging. Our contribution is that, if the procedures can be checked to satisfy the conditions demanded in Theorem 2, then our framework kicks in and the method will converge.

To demonstrate our point, we concentrate on two special structures — orthogonal factors and stochastic factors — in the subsequent discussion. We propose algorithmic details for computing the structured least squares solutions and carry out the crucial task of checking that the conditions in Theorem 2 are met. At the end, we are able to draw the conclusion of convergence.

Orthogonal factors. To fix the idea, we restate our problem: Given $A \in \mathbb{R}^{m_1 m_2 \times n_1 n_2}$, where $m_1 \ge n_1$ and $m_2 \ge n_2$, find $Q_1 \in \mathscr{S}(m_1, n_1)$ and $Q_2 \in \mathscr{S}(m_2, n_2)$ so that the objective function

$$g(Q_1, Q_2) := \frac{1}{2} \|A - Q_1 \otimes Q_2\|_F^2 \tag{40}$$

is minimized. We shall consider the constraint as the manifold $\mathscr{S}(m_1, n_1) \times \mathscr{S}(m_2, n_2)$ with the product topology.

To find the critical point for the constrained optimization of (40), we compute the projected gradient of $g(Q_1, Q_2)$. We begin with the action of the Fréchet derivative of $g(Q_1, Q_2)$ at a general point $(H_1, H_2) \in \mathbb{R}^{m_1 \times n_1} \times \mathbb{R}^{m_2 \times n_2}$. Under the product topology, we may consider the partial derivatives separately. Thus, the action of the partial derivative of g with respect to Q_1 on H_1 is given by

$$\begin{aligned}\frac{\partial g}{\partial Q_1}.H_1 &= \langle -H_1 \otimes Q_2, A - Q_1 \otimes Q_2 \rangle \\ &= -\langle \mathbf{vec}(H_1), \mathscr{R}(A - Q_1 \otimes Q_2)\mathbf{vec}(Q_2) \rangle \\ &= -\langle H_1, \mathscr{A} \circledast_{(m_1,n_1)} Q_2 - n_2 Q_1 \rangle,\end{aligned}$$

where the block matrix A is considered as an order-4 tensor $\mathscr{A} \in \mathbb{R}^{m_1 \times n_1 \times m_2 \times n_2}$ and, similar to the operation (15),

$$\mathscr{A} \circledast_{(m_1,n_1)} Q_2 := [\langle A_{ij}, Q_2 \rangle] \in \mathbb{R}^{m_1 \times n_1}.$$

Similarly,

$$\frac{\partial g}{\partial Q_2}.H_2 = -\langle H_2, \mathscr{A} \circledast_{(m_2,n_2)} Q_1 - n_1 Q_2 \rangle$$

with

$$\mathscr{A} \circledast_{(m_2,n_2)} Q_1 := \left[\langle \widetilde{A}_{ij}, Q_1 \rangle \right] \in \mathbb{R}^{m_2 \times n_2}.$$

By the Riesz representation theorem, the partial gradients of $g(Q_1, Q_2)$ can be interpreted as

$$\begin{cases} \frac{\partial g}{\partial Q_1} = n_2 Q_1 - \mathscr{A} \circledast_{(m_1,n_1)} Q_2, \\ \frac{\partial g}{\partial Q_2} = n_1 Q_2 - \mathscr{A} \circledast_{(m_2,n_2)} Q_1. \end{cases} \tag{41}$$

We now project the partial gradients onto the tangent spaces of the respective Stiefel spaces. Applying (18) to both partial gradients, we obtain

$$\begin{cases} \mathbf{Proj}_{\mathscr{T}_{Q_1}\mathscr{S}(m_1,n_1)} \frac{\partial g}{\partial Q_1} = Q_1 \frac{(\mathscr{A} \circledast_{(m_1,n_1)} Q_2)^\top Q_1 - Q_1^\top (\mathscr{A} \circledast_{(m_1,n_1)} Q_2)}{2} - (\mathbb{I}_{m_1} - Q_1 Q_1^\top) \mathscr{A} \circledast_{(m_1,n_1)} Q_2, \\ \mathbf{Proj}_{\mathscr{T}_{Q_2}\mathscr{S}(m_2,n_2)} \frac{\partial g}{\partial Q_2} = Q_2 \frac{(\mathscr{A} \circledast_{(m_2,n_2)} Q_1)^\top Q_2 - Q_2^\top (\mathscr{A} \circledast_{(m_2,n_2)} Q_1)}{2} - (\mathbb{I}_{m_2} - Q_2 Q_2^\top) \mathscr{A} \circledast_{(m_2,n_2)} Q_1. \end{cases}$$

We now are ready to characterize the first order optimality condition for the orthogonal Kronecker approximation problem (40).

Lemma 5 *For (Q_1, Q_2) to be a critical point for (40), it must be such that*

1. *Q_1 is the orthogonal portion in the polar decomposition of $\mathscr{A} \circledast_{(m_1,n_1)} Q_2$, and*
2. *Q_2 is the orthogonal portion in the polar decomposition of $\mathscr{A} \circledast_{(m_2,n_2)} Q_1$*

simultaneously.

Proof The first order optimality condition is that the projected gradients should be zero. The conclusion follows from the argument used in proving Corollary 2.

Based on this characterization, we are now able to define the two steps at Lines 2 and 3 in Algorithm 4 more specifically as in Algorithm 5 for the orthogonal Kronecker approximation. Furthermore, using our framework, we are able to argue for the convergence of the algorithm under the following assumptions.

Theorem 4 *Assume that*

Algorithm 5 (Polar method for orthogonal Kronecker approximation.)

Require: A generic matrix $A \in \mathbb{R}^{m_1 m_2 \times n_1 n_2}$, and an initial matrix $Q_2^{(0)} \in \mathscr{S}(m_2, n_2)$,
Ensure: A local best orthogonal Kronecker approximation to A

1: **for** $p = 0, 1, \cdots$, **do**
2: $\quad [Q_1^{(p+1)}, P_1^{(p+1)}] = \mathsf{poldec}(\mathscr{A}\circledast_{(m_1,n_1)} Q_2^{(p)})$ {using polar decomposition.}
3: $\quad [Q_2^{(p+1)} P_2^{(p+1)}] = \mathsf{poldec}(\mathscr{A}\circledast_{(m_2,n_2)} Q_1^{(p+1)})$ {using polar decomposition.}
4: **end for**

1. *The given matrix A is such that the Hessian of the corresponding objective function g defined in (40) is positive definite at one of its local minimizers; and*
2. *The initial matrix $Q_2^{(0)} \in \mathscr{S}(m_2, n_2)$ is such that the subsequent matrices $\{\mathscr{A}\circledast_{(m_1,n_1)} Q_2^{(p)}\}$ and $\{\mathscr{A}\circledast_{(m_2,n_2)} Q_1^{(p+1)}\}$ defined in Algorithm 5 are of full column rank in $\mathbb{R}^{m_1\times n_1}$ and $\mathbb{R}^{m_2\times n_2}$, respectively.*

Then the sequence $\{(Q_1^{(p)}, Q_2^{(p)})\}$ generated by Algorithm 5 converges to a local solution to the orthogonal Kronecker approximation problem.

Proof To apply our framework for convergence, the conditions needed by Theorem 2 should be satisfied by Algorithm 5. We check out two particular conditions, while others are either obvious or assumed.

Observe first that the definitions at Lines 2 and 3 actually represent an ALS optimization mechanism because

$$g(Q_1, Q_2) = \|\mathscr{R}(A) - \mathbf{vec}(Q_1)\mathbf{vec}(Q_2)^\top\|_F^2 = \|\mathscr{A}\circledast_{(m_1,n_1)} Q_2 - Q_1\|_F^2$$

and, by Corollary 2, the nearest $Q \in \mathscr{S}(p, q)$ to a fixed point $Z \in \mathbb{R}^{p\times q}$ comes from the polar decomposition of Z. The polar decomposition is unique for a full rank matrix and is continuous in its parameters.

Observe next that the accumulation points of the iteration must satisfy the system of polynomials [23,33]

$$\begin{cases} Q_1^\top(\mathscr{A}\circledast_{(m_1,n_1)} Q_2) = (\mathscr{A}\circledast_{(m_1,n_1)} Q_2)^\top Q_1, \\ Q_2^\top(\mathscr{A}\circledast_{(m_2,n_2)} Q_1) = (\mathscr{A}\circledast_{(m_2,n_2)} Q_1)^\top Q_2, \\ \mathscr{A}\circledast_{(m_1,n_1)} Q_2 = Q_1 Q_1^\top(\mathscr{A}\circledast_{(m_1,n_1)} Q_2), \\ \mathscr{A}\circledast_{(m_2,n_2)} Q_1 = Q_2 Q_2^\top(\mathscr{A}\circledast_{(m_2,n_2)} Q_1). \end{cases} \tag{42}$$

which, by Lemma 2, contains only geometrically isolated solutions for almost all data matrix A. The iterates $\left\{(Q_1^{(p)}, Q_2^{(p)})\right\}$ are obviously bounded as they are from the Stiefel manifolds. Conditions in Theorem 2 are satisfied.

We remark that the first assumption in Theorem 4 holds for generic A. We conjecture that the second assumption is also true for generic A and $Q_2^{(0)}$ because, otherwise, rank deficient matrices are the union of low dimensional manifolds and are susceptible to perturbations. At present we do not have a formal proof of the genericity, so we state them as assumptions.

Stochastic factors. Again, we first restate the problem: Let $\mathcal{M}(q)$ denote the convex and compact subset of all column stochastic matrices in $\mathbb{R}^{q\times q}$. Given $A \in \mathbb{R}^{n_1 n_2 \times n_1 n_2}$, the stochastic Kronecker approximation concerns finding the factors $B \in \mathscr{M}(n_1)$ and $C \in \mathscr{M}(n_2)$ so that the objective function

$$\psi(B, C) := \frac{1}{2}\|A - B \otimes C\|_F^2 \tag{43}$$

is minimized.

It is worth mentioning that the problem has an interesting interpretation. The entry of $B \otimes C$ has the form $b_{ij}c_{st}$. Thus, the approximation amounts to aggregating the $n_1 n_2$ states into n_1 groups $G_1, \ldots, G_{n_1}$, each of size n_2, such that the transition probability among states within each group is the same. Thus, b_{ij} stands for the probability of transition from group G_j to state G_i while c_{st} stands for the probability of transition from state t to state s within any group.

Each of the two structured least squares subproblems in Algorithm 4 can easily be formulated to take into the stochastic structure. For instance, the subproblem

$$\min_{\mathbf{1}_{n_1}^\top B = \mathbf{1}_{n_1}^\top, B \geq 0} \|A - B \otimes C\|_F^2, \tag{44}$$

where $C \in \mathscr{M}(n_2)$ is fixed and $\mathbf{1}_{n_1} \in \mathbb{R}^{n_1}$ is the column vector of all ones, is a classical constrained linear least squares problem which can be solved via existent optimization software package [22]. Furthermore, the problem (44) is a convex programming problem. If we assume the generic condition that the data are such that the objective function is strictly convex, then the solution to (44) is unique. Replacing the constraints in Algorithm 4 by $\mathscr{M}(n_1)$ and $\mathscr{M}(n_2)$, and equipped with the ability to solve each subproblem of the restricted objective functions, our concern is whether the iteration will converge.

To apply our theory, we need to check in particular the finiteness and isolation of stationary points. The procedure should be quite routine now, except that the feasible sets now have boundaries, i.e., some of the entries of B or C are zero. The projection at the boundaries is equivalent to the KKT conditions. For simplicity, we shall omit the details. We only demonstrate the projected gradient for the problem (44) at an interior point. The partial gradient of ψ with respect to B is

$$\frac{\partial \psi}{\partial B} = B\|C\|_F^2 - (A \circledast_{(n_1,n_1)} C) \in \mathbb{R}^{n_1 \times n_1}. \tag{45}$$

The tangent space of $\mathscr{M}(n_1)$ is made of matrices whose column sum is zero. The projection of any $Z \in \mathbb{R}^{n_1 \times n_1}$ onto the tangent space of $\mathscr{M}(n_1)$ is trivially given by

$$\mathbf{Proj}_{\mathscr{T}_B(\mathscr{M}(n_1))}(Z) = Z - \mathbf{1}_{n_1} \left[\frac{\sum_{i=1}^{n_1} z_{i,1}}{n_1}, \ldots, \frac{\sum_{i=1}^{n_1} z_{i,n_1}}{n_1} \right].$$

So the projected gradient can be calculated. Likewise, the projected gradient of ψ with respect to C can be calculated. In all, setting the projected gradient of $\psi(B, C)$ to zero is equivalent to a system of polynomials which, by Lemma 2, contains finitely many geometrically isolated solutions for a generic A. Without filling in more details, we have sketched a proof by using our theory that the matrices generated by the ALS iteration for the stochastic Kronecker approximation problem converge almost surely.

5 Conclusion

A general theory has been established in this paper as a useful tool for arguing that an alternating optimization method will converge under mild conditions. The conditions are the continuity of the algorithm, the differentiability of the objective function, the boundedness, finiteness, and geometrical isolation of the accumulation points. An array of problems arising from different backgrounds are demonstrated to be under this framework and satisfy these conditions. In particular, algorithms designed for the Tucker nearest problem and the structured Kronecker approximation problems are shown to converge, which is perhaps new in the literature. The theory might serve as an algorithmic foundation for many other methods having the characteristics of iteration by alternating variables.

References

1. Bader, B.W., Kolda, T.G.: Efficient MATLAB computations with sparse and factored tensors. SIAM J. Sci. Comput. **30**(1), 205–231 (2007/08). DOI 10.1137/060676489. URL http://dx.doi.org/10.1137/060676489
2. Bezdek, J.C., Hathaway, R.J.: Some notes on alternating optimization. In: N.R. Pal, M. Sugeno (eds.) Advances in Soft Computing — AFSS 2002: 2002 AFSS International Conference on Fuzzy Systems Calcutta, India,, pp. 288–300. Springer, Berlin, Heidelberg (2002)
3. Bezdek, J.C., Hathaway, R.J.: Convergence of alternating optimization. Neural Parallel Sci. Comput. **11**(4), 351–368 (2003)
4. Bunse-Gerstner, A., Byers, R., Mehrmann, V., Nichols, N.K.: Numerical computation of an analytic singular value decomposition of a matrix valued function. Numer. Math. **60**, 1–40 (1991)
5. Cadzow, J.A.: Signal enhancement: A composite property mapping algorithm. IEEE Trans. on Acoust., Speech, Signal Processing **36**, 39–62 (1988)
6. Chu, M.T., Funderlic, R.E., Plemmons, R.J.: Structured low rank approximation. Linear Algebra Appl. **366**, 157–172 (2003). Special issue on structured matrices: analysis, algorithms and applications (Cortona, 2000)
7. Chu, M.T., Lin, M.M., Wang, L.: A study of singular spectrum analysis with global optimization techniques. J. Global Optim. **60**(3), 551–574 (2014). DOI 10.1007/s10898-013-0117-3. URL http://dx.doi.org/10.1007/s10898-013-0117-3
8. Chu, M.T., Trendafilov, N.T.: The orthogonally constrained regression revisited. J. Comput. Graph. Statist. **10**(4), 746–771 (2001). DOI 10.1198/106186001317243430. URL http://dx.doi.org/10.1198/106186001317243430
9. Comon P., L.X., de Almeida, A.L.F.: Tensor decompositions, alternating least squares and other tales. J. Chemometrics **23**, 393–405 (2009). DOI 10.1002/cem.1236
10. De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. **21**(4), 1253–1278 (electronic) (2000). DOI 10.1137/S0895479896305696. URL http://dx.doi.org/10.1137/S0895479896305696
11. 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 (electronic) (2000). DOI 10.1137/S0895479898346995. URL http://dx.doi.org/10.1137/S0895479898346995
12. Faber, N.K.M., Bro, R., Hopke, P.K.: Recent developments in candecomp/parafac algorithms: a critical review. Chemometrics and Intelligent Laboratory Systems **65**(1), 119 – 137 (2003). DOI 10.1016/S0169-7439(02)00089-8. URL http://www.sciencedirect.com/science/article/pii/S0169743902000898
13. Guan, Y., Chu, M.T., Chu, D.: SVD-based algorithms for the best rank-1 approximation of a symmetric tensor. preprint, North Carolina State University (2017)
14. Harshman, R.: Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis. UCLA Working Papers in Phonetics **16** (1970)
15. Higham, N.J.: Computing the polar decomposition—with applications. SIAM J. Sci. Statist. Comput. **7**(4), 1160–1174 (1986). DOI 10.1137/0907079. URL http://dx.doi.org.prox.lib.ncsu.edu/10.1137/0907079
16. Higham, N.J.: Computing a nearest symmetric positive semidefinite matrix. Linear Algebra Appl. **103**, 103–118 (1988). DOI 10.1016/0024-3795(88)90223-6. URL http://dx.doi.org.prox.lib.ncsu.edu/10.1016/0024-3795(88)90223-6
17. Hitchcock, F.: The Expression of a Tensor Or a Polyadic as a Sum of Products. Contributions from the Department of Mathematics. sn. (1927). URL http://books.google.com/books?id=G7VOHAAACAAJ
18. Kamm, J., Nagy, J.G.: Optimal Kronecker product approximation of block Toeplitz matrices. SIAM J. Matrix Anal. Appl. **22**(1), 155–172 (2000). DOI 10.1137/S0895479898345540. URL http://dx.doi.org/10.1137/S0895479898345540
19. Keller, J.B.: Closest unitary, orthogonal and Hermitian operators to a given operator. Math. Mag. **48**(4), 192–197 (1975). URL https://doi.org/10.2307/2690338
20. Kiers, H.A.L.: Towards a standardized notation and terminology in multiway analysis. Journal of Chemometrics **14**(3), 105–122 (2000). DOI 10.1002/1099-128X(200005/06)14:3¡105::AID-CEM582¿3.0.CO;2-I. URL http://dx.doi.org/10.1002/1099-128X(200005/06)14:3¡105::AID-CEM582¿3.0.CO;2-I
21. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009). DOI 10.1137/07070111X. URL http://dx.doi.org/10.1137/07070111X
22. Lawson, C.L., Hanson, R.J.: Solving least squares problems, *Classics in Applied Mathematics*, vol. 15. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1995). DOI 10.1137/1.9781611971217. URL http://dx.doi.org.prox.lib.ncsu.edu/10.1137/1.9781611971217
23. Lim, L.H.: Singular values and eigenvalues of tensors: a variational approach. In: Computational Advances in Multi-Sensor Adaptive Processing, 2005 1st IEEE International Workshop on, pp. 129 –132 (2005). DOI 10.1109/CAMAP.2005.1574201
24. Mohlenkamp, M.J.: Musings on multilinear fitting. Linear Algebra Appl. **438**(2), 834–852 (2013). DOI 10.1016/j.laa.2011.04.019. URL http://dx.doi.org/10.1016/j.laa.2011.04.019
25. Moré, J.J., Sorensen, D.C.: Computing a trust region step. SIAM J. Sci. Statist. Comput. **4**(3), 553–572 (1983). DOI 10.1137/0904038. URL http://dx.doi.org.prox.lib.ncsu.edu/10.1137/0904038
26. Nagy, J., Kilmer, M.: Kronecker product approximation for preconditioning in three-dimensional imaging applications. Image Processing, IEEE Transactions on **15**(3), 604 –613 (2006). DOI 10.1109/TIP.2005.863112

27. Ortega, J.M.: Numerical analysis, *Classics in Applied Mathematics*, vol. 3, second edn. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1990). DOI 10.1137/1.9781611971323. URL http://dx.doi.org.prox.lib.ncsu.edu/10.1137/1.9781611971323. A second course
28. 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). DOI 10.1142/9789812567727. URL http://dx.doi.org.prox.lib.ncsu.edu/10.1142/9789812567727
29. Tucker, L.: Some mathematical notes on three-mode factor analysis. Psychometrika **31**, 279–311 (1966). URL http://dx.doi.org/10.1007/BF02289464
30. Uschmajew, A.: Local convergence of the alternating least squares algorithm for canonical tensor approximation. SIAM J. Matrix Anal. Appl. **33**(2), 639–652 (2012). DOI 10.1137/110843587. URL http://dx.doi.org/10.1137/110843587
31. Van Loan, C.F.: The ubiquitous Kronecker product. J. Comput. Appl. Math. **123**(1-2), 85–100 (2000). DOI 10.1016/S0377-0427(00)00393-9. URL http://dx.doi.org/10.1016/S0377-0427(00)00393-9. Numerical analysis 2000, Vol. III. Linear algebra
32. 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)
33. 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). DOI 10.1137/130938207. URL http://dx.doi.org.prox.lib.ncsu.edu/10.1137/130938207
34. Wright, K.: Differential equations for the analytic singular value decomposition of a matrix. Numer. Math. **3**(2), 283–295 (1992)
35. Zhang, T., Golub, G.H.: Rank-one approximation to high order tensors. SIAM J. Matrix Anal. Appl. **23**(2), 534–550 (2001). DOI 10.1137/S0895479899352045. URL http://dx.doi.org/10.1137/S0895479899352045