POLAR DECOMPOSITION BASED ALGORITHMS ON THE PRODUCT OF STIEFEL MANIFOLDS WITH APPLICATIONS IN TENSOR APPROXIMATION

JIANZE LI AND SHUZHONG ZHANG

Abstract. In this paper, based on the matrix polar decomposition, we propose a general algorithmic framework to solve a class of optimization problems on the product of Stiefel manifolds. We establish the weak convergence and global convergence of this general algorithmic approach based on the Łojasiewicz gradient inequality. This general algorithm and its convergence results are applied to the best rank-1 approximation, low rank orthogonal approximation and low multilinear rank approximation for higher order tensors. We also present a symmetric variant of this general algorithm to solve a symmetric variant of this class of optimization models, which essentially optimizes over a single Stiefel manifold. We establish its weak convergence and global convergence in a similar way. This symmetric variant and its convergence results are applied to the best symmetric rank-1 approximation and low rank symmetric orthogonal approximation for higher order tensors. It turns out that well-known algorithms such as HOPM, S-HOPM, LROAT, S-LROAT are all special cases of this general algorithmic framework and its symmetric variant.

1. Introduction

Theory and algorithms on optimization over manifolds have been developed and applied widely because of its practical relevance; see [3]. In particular, methods such as the Newton-type and trust-region algorithms have been proposed for optimization over Stiefel manifolds, which represents the orthogonal constraints [1, 3, 4, 18].

Let \( \text{St}(r, n) \subseteq \mathbb{R}^{n \times r} \) be the Stiefel manifold with \( 1 \leq r \leq n \). In this paper, we mainly study the optimization problem on the product of Stiefel manifolds, which is to maximize a smooth function

\[
f : \Omega \rightarrow \mathbb{R}^+, \tag{1}
\]

where

\[
\Omega \triangleq \text{St}(r_1, n_1) \times \text{St}(r_2, n_2) \times \cdots \times \text{St}(r_d, n_d)
\]
is the product of \( d \) Stiefel manifolds \((d > 1)\). Let \( \omega \overset{\text{def}}{=} (U^{(1)}, U^{(2)}, \ldots, U^{(d)}) \in \Omega \). If \( r_i = r \) and \( n_i = n \) for \( 1 \leq i \leq d \), then \( \Omega \) becomes
\[
\Omega_s \overset{\text{def}}{=} \text{St}(r, n) \times \text{St}(r, n) \times \cdots \times \text{St}(r, n).
\]
Assume that the cost function \((1)\) is symmetric on \( \Omega_s \), that is,
\[
f(\omega) = f(\pi(\omega))
\]
for any \( \omega \in \Omega_s \) and permutation \( \pi \). In this paper, we also study the symmetric variant of problem \((1)\), which is to maximize the smooth function
\[
g : \text{St}(r, n) \longrightarrow \mathbb{R}^+, \quad U \mapsto f(U, U, \ldots, U).
\]

Many low rank approximation problems for higher order tensors \([8, 12, 29, 44]\), such as the best rank-1 approximation \([15, 16, 17, 50]\), low rank orthogonal approximation \([7, 28]\) and low multilinear rank approximation \([15, 17, 23]\), can be formulated in the form of problem \((1)\). These approximations have been widely used in various fields, including signal processing \([12, 14, 20, 26, 38]\), numerical linear algebra \([41, 42]\) and data analysis \([5, 8, 29, 44]\). In particular, the symmetric variant of best rank-1 approximation is corresponding to finding the largest Z-eigenvalue \([42]\). When the rank is equal to the dimension, the low rank orthogonal approximation boils down to the approximate orthogonal diagonalization \([11, 34, 36, 37, 47]\) of symmetric tensors, which is central in Independent Component Analysis (ICA) \([9, 10]\). The low multilinear rank approximation is equivalent to the well-known Tucker decomposition \([17, 29]\), which has been a popular method for data reduction in signal processing and machine learning. Therefore, it is desirable to develop a general algorithmic scheme to solve problem \((1)\) and its symmetric version \((3)\).

**Contribution.** In this paper, based on the matrix polar decomposition, we propose an approach to be called APDOI (alternating polar decomposition based orthogonal iteration) (Algorithm 1) to solve problem \((1)\). Then we establish its weak convergence\(^1\) and global convergence\(^2\) based on the Lojasiewicz gradient inequality. We apply APDOI and establish its convergence properties to the best rank-1 approximation (Section 6.1); the low rank orthogonal approximation (Section 7.1); the low multilinear rank approximation (Section 8) for higher order tensors. It turns out that the well-known methods HOPM \([16, 17]\) (for best rank-1 approximation) and LROAT \([7]\) (for low rank orthogonal approximation) are both special cases of APDOI. Our convergence results subsume the results found in the literature designed for those special cases.

Algorithm APDOI for the low multilinear rank approximation will be called LMPD (low multilinear rank approximation based on polar decomposition) in this paper. For this particular algorithm, we propose its shifted variant, LMPD-S, which has a better

\(^1\)every accumulation point is critical.
\(^2\)for any starting point, the iterations converge as a whole sequence.
convergence property. Then we conduct experiments to show that LMPD and LMPD-S have comparable speed with the well-known HOOI [17] algorithm, and LMPD-S has a even much better convergence performance.

As the symmetric variant of APDOI, we propose a PDOI (polar decomposition based orthogonal iteration) approach (Algorithm 2) to solve problem (3). We establish its weak convergence and global convergence in a similar way. Then Algorithm PDOI and its convergence results are applied to the best symmetric rank-1 approximation (Section 6.2) and low rank symmetric orthogonal approximation (Section 7.2) for higher order tensors. It turns out that the well-known algorithms S-HOPM [17] (for best symmetric rank-1 approximation) and S-LROAT [7] (for low rank symmetric orthogonal approximation) are both special cases of PDOI. Our convergence results also subsume the results found in the literature designed for those special cases.

Organization. The paper is organized as follows. In Section 2, we introduce notations and preliminaries. In Section 3, we show the condition under which problems (1) and (3) will be studied in this paper. Then we prove that this condition is satisfied in the best rank-1 approximation, low rank orthogonal approximation and low multilinear rank approximation for higher order tensors. In Section 4, we propose Algorithm APDOI to solve problem (1), and establish its convergence. In Section 5, as a symmetric variant of APDOI, we propose Algorithm PDOI to solve problem (3), and establish its convergence. In Sections 6 to 8, we apply these convergence results to the best rank-1 approximation, low rank orthogonal approximation and low multilinear rank approximation for higher order tensors, respectively.

2. Notations and preliminaries

2.1. Notations. Let $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ be the linear space of $d$-th order real tensors and $\text{symm} (\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}) \subseteq \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ be the set of symmetric ones, whose entries do not change under any permutation of indices [13, 41]. Tensor arrays, matrices, and vectors, will be respectively denoted by bold calligraphic letters, e.g., $\mathcal{A}$, with bold uppercase letters, e.g., $\mathbf{M}$, and with bold lowercase letters, e.g., $\mathbf{u}$; corresponding entries will be denoted by $A_{ijk}$, $M_{ij}$, and $u_i$. Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ and $\mathbf{M} \in \mathbb{R}^{m \times n}$, we use the $i$-mode product defined by

$$(\mathcal{A} \cdot_i \mathbf{M})_{p_1 \cdots j \cdots p_d} = \sum_{p_i} A_{p_1 \cdots p_d} M_{jp_i}.$$

The $i$-mode unfolding of $\mathcal{A}$ is denote by $\mathcal{A}_{(i)}$, which is the matrix obtained by reordering the $i$-mode fibers\(^3\) of $\mathcal{A}_{(i)}$ in a fixed way [29]. We denote by $\text{rank}_i(\mathcal{A})$ the $i$-rank, that is, the rank of $\mathcal{A}_{(i)}$. The vector $(\text{rank}_1(\mathcal{A}), \cdots, \text{rank}_d(\mathcal{A}))$ is called the multilinear rank of

---

\(^3\)the fiber of a tensor is defined by fixing every index but one.
A. The diagonal of a tensor $A \in \mathbb{R}^{n \times n \times \cdots \times n}$ is defined by
$$\text{diag}\{A\} \overset{\text{def}}{=} (A_{11}, \cdots, A_{nn})^T.$$

Let $O_n \subseteq \mathbb{R}^{n \times n}$ be the orthogonal group. We denote by $\| \cdot \|$ the Frobenius norm of a tensor or a matrix, or the Euclidean norm of a vector. We denote by $S^{n-1}$ the unit sphere of $\mathbb{R}^n$. We denote by $\sigma_{\text{min}}$ the minimal singular value of a matrix. Let $\text{sym}(X) \overset{\text{def}}{=} \frac{1}{2}(X + X^T)$ for $X \in \mathbb{R}^{r \times r}$.

2.2. Łojasiewicz gradient inequality. Let $M \subseteq \mathbb{R}^n$ be a Riemannian submanifold, and $f : M \rightarrow \mathbb{R}$ be a differentiable function. In this paper, for simplicity, we also denote by $f$ the natural extension of itself to $\mathbb{R}^n$. Let $x \in M$ and $T_xM$ be the tangent space of $M$ at $x$. We denote by $\nabla f(x)$ the Euclidean gradient of $f$ at $x$, and $\text{grad} f(x)$ the Riemannian gradient of $f$ at $x$.

**Definition 2.1.** ([43, Definition 2.1]) The function $f : M \rightarrow \mathbb{R}$ is said to satisfy a Łojasiewicz gradient inequality at $x \in M$, if there exist $\sigma > 0$, $\zeta \in (0, \frac{1}{2}]$ and a neighborhood $U$ in $M$ of $x$ such that for all $y \in U$, it follows that
$$|f(y) - f(x)|^{1-\zeta} \leq \sigma \|\text{grad} f(y)\|.$$  \hfill (4)

**Lemma 2.2** ([43, Proposition 2.2]). Let $M \subseteq \mathbb{R}^n$ be an analytic submanifold\(^5\) and $f : M \rightarrow \mathbb{R}$ be a real analytic function. Then for any $x \in M$, $f$ satisfies a Łojasiewicz gradient inequality (4) in the $\delta$-neighborhood of $x$, for some\(^6\) $\delta, \sigma > 0$ and $\zeta \in (0, \frac{1}{2}]$.

Based on the discrete-time analogue of classical Łojasiewicz’s theorem [2, 33, 39], the following result was proved [43, Theorem 2.3].

**Theorem 2.3.** Let $M \subseteq \mathbb{R}^n$ be an analytic submanifold and $\{x_k\}_{k \geq 1} \subseteq M$. Suppose that $f$ is real analytic and, for large enough $k$,
(i) there exists $\sigma > 0$ such that
$$|f(x_{k+1}) - f(x_k)| \geq \sigma \|\text{grad} f(x_k)\||x_{k+1} - x_k|;$$
(ii) $\text{grad} f(x_k) = 0$ implies that $x_{k+1} = x_k$.
Then any accumulation point $x_*$ of $\{x_k\}_{k \geq 1}$ must in fact be the same.

**Remark 2.4.** It can be verified, after checking the proof in [43], that condition (ii) in Theorem 2.3 can be replaced by that $f(x_{k+1}) = f(x_k)$ implies $x_{k+1} = x_k$.

\(^4\)See [3, Section 3.6] for a detailed definition.
\(^5\)See [31, Definition 2.7.1] or [36, Definition 5.1] for a definition of an analytic submanifold.
\(^6\)The values of $\delta, \sigma, \zeta$ depend on a specific point.
3. Problem statement and tensor approximation examples

3.1. Problem statement. We first discuss the condition, under which problems (1) and (3) will be studied in this paper. This condition is motivated by the tensor approximation problems in Section 3.2.1, Section 3.3.1 and Section 3.4.

3.1.1. General case. Let

\[ \Omega^{(i)} \triangleq \text{St}(r_1, n_1) \times \cdots \times \text{St}(r_{i-1}, n_{i-1}) \times \text{St}(r_{i+1}, n_{i+1}) \times \cdots \times \text{St}(r_d, n_d) \]

be the product of \(d-1\) Stiefel manifolds and \(\nu^{(i)} \in \Omega^{(i)}\). Let \(U \in \text{St}(r_i, n_i)\). Define

\[ h_{(i)} : \text{St}(r_i, n_i) \rightarrow \mathbb{R}^+, \]

\[ U \mapsto f(\nu^{(i)}, U) \triangleq f(U^{(1)}, \ldots, U^{(i-1)}, U, U^{(i+1)}, \ldots, U^{(d)}). \tag{5} \]

In this paper, we assume that problem (1) always satisfies the restricted \(\frac{1}{\alpha}\)-homogeneity condition, which means that there is some fixed \(\alpha \in (0, 1)\) such that, for any \(\nu^{(i)} \in \Omega^{(i)}\) and \(1 \leq i \leq d\), it holds that

\[ h_{(i)}(U) = \alpha \langle U, \nabla h_{(i)}(U) \rangle \] \tag{6}

for any \(U \in \text{St}(r_i, n_i)\). By the gradient inequality [6, (3.2)], it follows that the function (5) is convex if and only if

\[ \langle U' - U, \nabla h_{(i)}(U) \rangle \leq h_{(i)}(U') - h_{(i)}(U), \]

for any \(U, U' \in \mathbb{R}^{n_i \times r_i}\), and then by condition (6), if and only if

\[ \langle U', \nabla h_{(i)}(U) \rangle \leq (1 - \alpha) \langle U, \nabla h_{(i)}(U) \rangle + \alpha \langle U', \nabla h_{(i)}(U') \rangle, \tag{7} \]

for any \(U, U' \in \mathbb{R}^{n_i \times r_i}\). We say that the cost function (1) is partially convex if the function (5) is convex for any \(\nu^{(i)} \in \Omega^{(i)}\) and \(1 \leq i \leq d\).

3.1.2. Symmetric case. In this paper, we assume that problem (3) always satisfies the \(\frac{1}{\beta}\)-homogeneity condition, which means that there is some fixed \(\beta \in (0, 1)\) such that

\[ g(U) = \beta \langle U, \nabla g(U) \rangle \] \tag{8}

for any \(U \in \text{St}(r, n)\). By the gradient inequality [6, (3.2)], it follows that the cost function (3) is convex if and only if

\[ \langle U' - U, \nabla g(U) \rangle \leq g(U') - g(U), \]

for any \(U, U' \in \mathbb{R}^{n \times r}\), and then by condition (8), if and only if

\[ \langle U', \nabla g(U) \rangle \leq (1 - \beta) \langle U, \nabla g(U) \rangle + \beta \langle U', \nabla g(U') \rangle, \]

for any \(U, U' \in \mathbb{R}^{n \times r}\).
Remark 3.1. Suppose that the cost function (1) is symmetric on \( \Omega_s \) and satisfies condition (6). Then the corresponding cost function (3) satisfies condition (8) with \( \beta = \frac{\alpha}{d} \).

In fact, for any \( U_0 \in \text{St}(r,n) \), we define
\[
h : \text{St}(r,n) \longrightarrow \mathbb{R}^+, \ U \longmapsto f(U_0, \ldots, U_0, U).
\]

Then we get that
\[
\nabla g(U_0) = d \nabla h(U_0)
\]
by (2). It follows from condition (6) that
\[
g(U_0) = h(U_0) = \alpha \langle U_0, \nabla h(U_0) \rangle = \frac{\alpha}{d} \langle U_0, \nabla g(U_0) \rangle.
\]
Therefore, the cost function (3) satisfies condition (8) with \( \beta = \frac{\alpha}{d} \).

3.2. Example: best rank-1 approximation.

3.2.1. General case. Let \( A \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \). The best rank-1 approximation problem [15, 16, 17, 50] is to find
\[
B^* = \arg \min \| A - B \|,
\]
where \( B = \lambda u^{(1)} \otimes u^{(2)} \otimes \cdots \otimes u^{(d)} \) with \( \lambda \in \mathbb{R} \) and \( u^{(i)} \in \mathbb{S}^{n_i-1} \). It was proved [15, 17] that problem (10) is equivalent to the maximization of
\[
f(u^{(1)}, u^{(2)}, \ldots, u^{(d)}) = |A \bullet_1 u^{(1)} \bullet_2 u^{(2)} \cdots \bullet_d u^{(d)}|^2,
\]
where \( u^{(i)} \in \mathbb{S}^{n_i-1} \) for \( 1 \leq i \leq d \).

Lemma 3.2. The cost function (11) satisfies (6) with \( \alpha = \frac{1}{2} \) and is partially convex.

Proof. Let \( 1 \leq i \leq d \). Denote that
\[
v^{(i)} \overset{\text{def}}{=} A \bullet_1 u^{(1)} \bullet_2 u^{(2)} \cdots \bullet_{i-1} u^{(i-1)} \bullet_{i+1} u^{(i+1)} \cdots \bullet_d u^{(d)}.
\]

Let \( V^{(i)} \overset{\text{def}}{=} v^{(i)} v^{(i)} \) and \( u \in \mathbb{R}^{n_i} \). Then we have that
\[
h^{(i)}(u) = |\langle v^{(i)}, u \rangle|^2 = \langle u, V^{(i)} u \rangle,
\]
\[
\nabla h^{(i)}(u) = 2 V^{(i)} u = 2 \langle v^{(i)}, u \rangle v^{(i)}.
\]
(12)

It is easy to see that the cost function (11) satisfies condition (6) with \( \alpha = \frac{1}{2} \). Moreover, by the Cauchy-Schwarz inequality, we have that
\[
\langle u', \nabla h^{(i)}(u) \rangle \leq \frac{1}{2} (\langle u, \nabla h^{(i)}(u) \rangle + \langle u', \nabla h^{(i)}(u') \rangle)
\]
for any \( u, u' \in \mathbb{R}^{n_i} \). Then the cost function (11) is partially convex by (7). The proof is complete. \( \square \)
3.2.2. Symmetric case. Let $A \in \text{symm}(\mathbb{R}^{n \times n \times \cdots \times n})$. The best symmetric rank-1 approximation problem [17, 27, 30, 42, 51] is to find

$$B_* = \arg \min ||A - B||,$$

where $B = \lambda u \otimes u \otimes \cdots \otimes u$ with $\lambda \in \mathbb{R}$ and $u \in \mathbb{S}^{n-1}$. It was proved [15, 17] that problem (13) is equivalent to the maximization of

$$g(u) \overset{\text{def}}{=} f(u, u, \cdots, u) = |A \circ_1 u^T \circ_2 u^T \cdots \circ_d u^T|^2,$$

where $u \in \mathbb{S}^{n-1}$. By Remark 3.1 and Lemma 3.2, we see that the cost function (14) satisfies condition (8) with $\beta = \frac{1}{27}$.

**Lemma 3.3.** Suppose that $d$ is even. Let

$$C = D \circ_1 H^T \circ_2 H^T \cdots \circ_d H^T$$

where $H \in \mathbb{R}^{m \times n}$ and $D \in \mathbb{R}^{m \times m \times \cdots \times m}$ is a diagonal tensor satisfying that $\text{diag}(D) = (\mu_1, \cdots, \mu_m)^T$ with $\mu_j \geq 0$. Let $X \in \mathbb{R}^{n \times r}$ and $W = C \circ_1 X^T \circ_2 X^T \cdots \circ_d X^T$. Then, for any $1 \leq i \leq r$, the function

$$\tau_i : \mathbb{R}^{n \times r} \to \mathbb{R}^+, \ X \mapsto |W_{ii} \cdots |^2$$

is convex.

**Proof.** Let $\psi(X) \overset{\text{def}}{=} HX$ be the linear map and

$$\gamma_i : \mathbb{R}^{m \times r} \to \mathbb{R}^+, \ Y \mapsto |W_{ii} \cdots |^2,$$

where $W = D \circ_1 Y^T \circ_2 Y^T \cdots \circ_d Y^T$. Note that $\gamma_i(Y) = \sum_{j=1}^m \mu_j y_{ji}$. We see that $\gamma_i$ is convex, and thus $\tau_i = \gamma_i \circ \psi$ is also convex. \[\square\]

**Remark 3.4.** By Lemma 3.3, if $A$ takes the form in (15), then the cost function (14) is convex. This fact has also been proved by the method of square matrix unfolding in [27, (4.2)].

3.3. Example: low rank orthogonal approximation.

3.3.1. General case. Let $A \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ and $r \leq n_i$ for $1 \leq i \leq d$. The low rank orthogonal approximation problem [7, 22, 28, 49] is to find

$$B_* = \arg \min ||A - B||,$$

where

$$B = \sum_{\ell=1}^r \mu_\ell u_\ell^{(1)} \otimes \cdots \otimes u_\ell^{(d)}$$

satisfies that $\langle u_\ell^{(i)}, u_\ell^{(i)} \rangle = 0$ for any $1 \leq i \leq d$ and $\ell_1 \neq \ell_2$. Let $U^{(i)} = [u_1^{(i)}, \cdots, u_r^{(i)}] \in \text{St}(r, n_i)$ for $1 \leq i \leq d$. It is clear that

$$B = D \circ_1 U^{(1)} \circ_2 U^{(2)} \circ_3 \cdots \circ_d U^{(d)}.$$
where $D \in \mathbb{R}^{r \times r \times \cdots}$ is the diagonal tensor satisfying that $\text{diag}\{D\} = (\mu_1, \cdots, \mu_r)^T$. It was proved [7] that problem (16) is equivalent to the maximization of

$$f(U^{(1)}, U^{(2)}, \ldots, U^{(d)}) = \|\text{diag}\{W\}\|^2,$$

(17)

where $W = A \cdot_1 U^{(1)} T \cdot_2 U^{(2)} T \cdots \cdot_d U^{(d)} T \in \mathbb{R}^{r \times r \times \cdots}$.

**Lemma 3.5.** The cost function (17) satisfies (6) with $\alpha = \frac{1}{2}$ and is partially convex.

**Proof.** Let $1 \leq i \leq d$. Denote that

$$V^{(i)}(i) \stackrel{\text{def}}{=} A \cdot_1 U^{(1)} T \cdots \cdot_{i-1} U^{(i-1)} T \cdot_{i+1} U^{(i+1)} T \cdots \cdot_d U^{(d)} T,$$

$$v^{(i,q)}(i) \stackrel{\text{def}}{=} A \cdot_1 u^{(1)} q T \cdots \cdot_{i-1} u^{(i-1)} q T \cdot_{i+1} u^{(i+1)} q T \cdots \cdot_d u^{(d)} q T \in \mathbb{R}^{n_i},$$

for $1 \leq q \leq r$. Let $U = [u_1, \cdots, u_r] \in \text{St}(r, n_i)$ and $W = V^{(i)}(i) \cdot U^T$. We can calculate that

$$h^{(i)}(U) = \|\text{diag}\{W\}\|^2 = \sum_{q=1}^{r} \langle v^{(i,q)}, u_q \rangle^2,$$

(18)

$$\nabla h^{(i)}(U) = 2[v^{(i,1)}, \cdots, v^{(i,r)}] \begin{bmatrix} W_{1-1} & & & \\ & \ddots & & \\ & & W_{r-r} & \end{bmatrix}.$$ 

Then it can be verified that

$$h^{(i)}(U) = \frac{1}{2} \langle U, \nabla h^{(i)}(U) \rangle.$$

Let $U' \in \text{St}(r, n_i)$ and $W' = V^{(i)}(i) \cdot U'^T$. By (18) and the Cauchy-Schwarz inequality, we get

$$2\langle U', \nabla h^{(i)}(U) \rangle = 2 \sum_{q=1}^{r} W'_{q-q} W_{q-q} \leq \sum_{q=1}^{r} W'^2_{q-q} + \sum_{q=1}^{r} W^2_{q-q} = \langle U', \nabla h^{(i)}(U') \rangle + \langle U, \nabla h^{(i)}(U) \rangle.$$

Then the cost function (17) is partially convex by (7). The proof is complete. \qed

**3.3.2. Symmetric case.** Let $A \in \text{symm}(\mathbb{R}^{n \times n \times \cdots \times n})$ and $r \leq n$. Then *low rank symmetric orthogonal approximation* problem [7, 35, 40] is to find

$$B_* = \arg \min \|A - B\|,$$

(19)

where

$$B = \sum_{\ell=1}^{r} \mu_\ell u_\ell \otimes \cdots \otimes u_\ell$$

...
satisfies that \(\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = 0\) for \(\ell_1 \neq \ell_2\). It was proved [7] that problem (19) is equivalent to the maximization of
\[
g(\mathbf{U}) \overset{\text{def}}{=} f(\mathbf{U}, \mathbf{A}^{(1)} \cdot \mathbf{U}^{(2)} \cdot \cdots \cdot \mathbf{U}^{(d)}) = \| \text{diag}\{\mathbf{W}\} \|^2, \tag{20}
\]
where \(\mathbf{W} = \mathbf{A} \cdot_1 \mathbf{U}_1^T \cdot_2 \mathbf{U}_2^T \cdots \cdot_d \mathbf{U}_d^T \in \mathbb{R}^{r \times \cdots \times r} \).

**Remark 3.6.** By Lemma 3.3, if \(\mathbf{A}\) takes the form in (15), then the cost function (20) is convex.

3.4. Example: low multilinear rank approximation. Let \(\mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}\) and \(r_i \leq n_i\) for \(1 \leq i \leq d\). The low multilinear rank approximation problem [15, 17, 32] is to find
\[
\mathbf{B}_s = \arg \min \| \mathbf{A} - \mathbf{B} \|,
\tag{21}
\]
where rank\(_i\)\((\mathbf{B}) \leq r_i\) for \(1 \leq i \leq d\). Note that any such \(\mathbf{B}\) can be decomposed as
\[
\mathbf{B} = \mathbf{C} \cdot_1 \mathbf{U}^{(1)} \cdot_2 \mathbf{U}^{(2)} \cdot_3 \cdots \cdot_d \mathbf{U}^{(d)},
\]
with \(\mathbf{C} \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_d}\) and \(\mathbf{U}^{(i)} \in \text{St}(r_i, n_i)\) for \(1 \leq i \leq d\). Problem (21) is in fact equivalent to the following Tucker decomposition [45, 15, 17] problem
\[
\min \| \mathbf{A} - \mathbf{C} \cdot_1 \mathbf{U}^{(1)} \cdot_2 \mathbf{U}^{(2)} \cdot_3 \cdots \cdot_d \mathbf{U}^{(d)} \|. \tag{22}
\]
It was proved [15, 17] that problem (22) is equivalent to the maximization of
\[
f(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(d)}) = \| \mathbf{W} \|^2, \tag{23}
\]
where \(\mathbf{W} = \mathbf{A} \cdot_1 \mathbf{U}_1^{(1)}^T \cdot_2 \mathbf{U}_2^{(2)}^T \cdots \cdot_d \mathbf{U}_d^{(d)}^T\) and \(\mathbf{U}_i^{(i)} \in \text{St}(r_i, n_i)\) for \(1 \leq i \leq d\).

**Lemma 3.7.** The cost function (23) satisfies (6) with \(\alpha = \frac{1}{2}\) and is partially convex.

**Proof.** Let \(1 \leq i \leq d\). Denote that
\[
\mathbf{V}^{(i)} = \mathbf{A} \cdot_1 \mathbf{U}_1^{(1)}^T \cdots \cdot_{i-1} \mathbf{U}_1^{(i-1)}^T \cdot_i \mathbf{U}_i^{(i)} \cdot_{i+1} \mathbf{U}_1^{(i+1)}^T \cdots \cdot_d \mathbf{U}_d^{(d)}^T.
\]
Let \(\mathbf{U} \in \text{St}(r_i, n_i)\) and \(\mathbf{W} = \mathbf{V}^{(i)} \cdot_1 \mathbf{U}^T\). By [29, Section 2.5], we see that
\[
h_{(i)}(\mathbf{U}) = \| \mathbf{W} \|^2 = \| \mathbf{U}^T \mathbf{V}^{(i)} \|^2,
\]
where \(\mathbf{V}^{(i)} = [\mathbf{v}_1^{(i)}, \mathbf{v}_2^{(i)}, \ldots, \mathbf{v}_R^{(i)}] \in \mathbb{R}^{n_i \times R_i}\) with \(R_i = r_1 r_2 \cdots r_{i-1} r_{i+1} \cdots r_d\). Let \(\mathbf{V}^{(i)} \overset{\text{def}}{=} \sum_{\ell=1}^{R_i} \mathbf{v}_\ell^{(i)} \mathbf{v}_\ell^{(i)T}\). It can be verified that
\[
h_{(i)}(\mathbf{U}) = \langle \mathbf{U}, \mathbf{V}^{(i)} \mathbf{U} \rangle, \quad \nabla h_{(i)}(\mathbf{U}) = 2 \mathbf{V}^{(i)} \mathbf{U}, \tag{24}
\]
and thus
\[
h_{(i)}(\mathbf{U}) = \frac{1}{2} \langle \mathbf{U}, \nabla h_{(i)}(\mathbf{U}) \rangle.
\]
Let $U' \in \text{St}(r_i, n_i)$. By the Cauchy-Schwarz inequality, it is easy to see that
\[
\langle U', V^{(i)} U \rangle \leq \frac{1}{2}(\langle U, V^{(i)} U \rangle + \langle U', V^{(i)} U' \rangle).
\]
Then the cost function (23) is partially convex by (7) and (24). The proof is complete. \hfill \Box

Remark 3.8. By (24), we see that the cost function (23) satisfies that $U^T \nabla h^{(i)}(U)$ is always positive semidefinite for any $U \in \text{St}(r_i, n_i)$ and $1 \leq i \leq d$.

4. Algorithm APDOI on the product of Stiefel manifolds

In this section, based on the matrix polar decomposition, we propose a general algorithm to solve problem (1), and establish its convergence. The algorithm and convergence results in this section apply to the best rank-1 approximation in Section 3.2.1, low rank orthogonal approximation in Section 3.3.1 and low multilinear rank approximation in Section 3.4.

4.1. Algorithm APDOI. By [3, (3.35)], the Riemannian gradient of (5) at $U$ is
\[
\text{grad } h^{(i)}(U) = \nabla h^{(i)}(U) - U \text{sym}(U^T \nabla h^{(i)}(U)).
\]
Therefore, if $U^{(i)}_*$ is a maximal point of (5), it should satisfy that
\[
\nabla h^{(i)}(U^{(i)}_*) = U^{(i)}_* \text{sym}(U^{(i)}_*^T \nabla h^{(i)}(U^{(i)}_*)),
\]
which is close to a polar decomposition form of $\nabla h^{(i)}(U^{(i)}_*)$.

Lemma 4.1. ([19, Theorem 9.4.1], [7, Lemma 5.3]) Let $X \in \mathbb{R}^{n \times r}$ with $r \leq n$. Then there exist $U \in \text{St}(r, n)$ and positive semidefinite matrix $P \in \text{symm}(\mathbb{R}^{r \times r})$ such that $X$ has the polar decomposition $X = UP$. We say that $U$ is the orthogonal polar factor and $P$ is the symmetric polar factor. Moreover,
(i) for any $U' \in \text{St}(r, n)$, we have that
\[
\langle U, X \rangle \geq \langle U', X \rangle;
\]
(ii) $U$ is the best orthogonal approximation to $X$, that is,
\[
\|X - U\| \leq \|X - U'\|
\]
for any $U' \in \text{St}(r, n)$;
(iii) if rank$(X) = R$, then $U$ and $P$ are both unique.

Let $k \geq 1$ and $1 \leq i \leq d$. We denote that
\[
\nu^{(k,i)} \overset{\text{def}}{=} (U^{(1)}_k, \ldots, U^{(i-1)}_k, U^{(i+1)}_k, \ldots, U^{(d)}_k) \in \Omega^{(i)},
\]
\[
\omega^{(k,i)} \overset{\text{def}}{=} (U^{(1)}_k, \ldots, U^{(i)}_k, U^{(i+1)}_k, \ldots, U^{(d)}_k) \in \Omega,
\]
and \( \omega^{(k)} \equiv \omega^{(k,d)} = \omega^{(k+1,0)} \). Define

\[
\omega^{(k)}(d) = \omega^{(k+1)}(0).
\]

Define \( h(k,i) : St(r_i, n_i) \rightarrow \mathbb{R}^+ \),

\[
U \mapsto f(\nu^{(k,i)}, U) \equiv f(U^{(1)}_k, \ldots, U^{(i-1)}_k, U, U^{(i+1)}_{k-1}, \ldots, U^{(d)}_{k-1}).
\]

Inspired by the decomposition form in (25), we propose the following alternating polar decomposition based orthogonal iteration (APDOI) algorithm to solve problem (1). We assume that the starting point \( \omega^{(0)} \) satisfies that \( f(\omega^{(0)}) > 0 \) without loss of generality.

**Algorithm 1. (APDOI)**

**Input:** starting point \( \omega^{(0)} \).

**Output:** \( \omega^{(k,i)}, k \geq 1, 1 \leq i \leq d \).

- For \( k = 1, 2, \ldots \),
- For \( i = 1, 2, \ldots, d \)
- Compute \( \nabla h(k,i)(U^{(i)}_{k-1}) \).
- Compute the polar decomposition of \( \nabla h(k,i)(U^{(i)}_{k-1}) \).
- Update \( U^{(i)}_k \) to be the orthogonal polar factor.
- End for
- Until convergence

**Remark 4.2.** Note that (1) is smooth and \( \Omega \) is compact. Let \( \Delta_* \equiv \max_{\omega \in \Omega} \|\nabla f(\omega)\| \). Then \( \|\nabla h(k,i)(U^{(i)}_{k-1})\| \leq \Delta_* \) always holds in Algorithm 1.

**Lemma 4.3.** Let \( h : St(r, n) \rightarrow \mathbb{R} \) be a differentiable function and \( U \in St(r, n) \). Suppose that \( U_* \) is the orthogonal polar factor of \( \nabla h(U) \).

(i) If \( U_* = U \), then \( \text{grad} h(U) = 0 \).

(ii) If \( \text{grad} h(U) = 0 \), then \( U^T \nabla h(U) \) is symmetric. Furthermore, if \( \nabla h(U) \) is of full rank and \( U^T \nabla h(U) \) is positive semidefinite, then \( U_* = U \).

**Proof.** (i) Since \( U_* = U \), we see that \( \nabla h(U) = UP \) is a polar decomposition. It follows that \( U^T \nabla h(U) = P \) is symmetric, and thus

\[
\nabla h(U) = UP = U\text{sym}(U^T \nabla h(U)).
\]

Therefore, by [3, (3.35)], we get that \( \text{grad} h(U) = 0 \).

(ii) By [3, (3.35)], we see that \( \nabla h(U) = U\text{sym}(U^T \nabla h(U)) \) and thus

\[
2U^T \nabla h(U) = U^T \nabla h(U) + \nabla h(U)^T U,
\]

which means that \( U^T \nabla h(U) \) is symmetric. Note that \( U^T \nabla h(U) \) is positive semidefinite and \( \nabla h(U) \) is of full rank. Then \( \nabla h(U) = UU^T \nabla h(U) \) is the unique polar decomposition of \( \nabla h(U) \) by Lemma 4.1(iii). Therefore, we get that \( U_* = U \). \qed
Remark 4.4. By Lemma 4.3, we see that, in Algorithm 1, if $\nabla h_{(k,i)}(U_{k-1}^{(i)})$ is of full rank and $\text{sym}(U_{k-1}^{(i)^T} \nabla h_{(k,i)}(U_{k-1}^{(i)}))$ is positive semidefinite, then $U_k^{(i)} = U_{k-1}^{(i)}$ if and only if $\text{grad} h_{(k,i)}(U_{k-1}^{(i)}) = 0$.

4.2. Convergence analysis. For Algorithm 1, we mainly prove the following results about its weak convergence and global convergence.

Theorem 4.5. Suppose that the cost function (1) is partially convex. If there exists $\delta > 0$ such that

$$\sigma_{\min}(\nabla h_{(k,i)}(U_{k-1}^{(i)})) > \delta \quad (26)$$

always holds in Algorithm 1, then every accumulation point of the iterations $\omega^{(k,i)}$ is critical.

Theorem 4.6. Suppose that the cost function (1) is partially convex and real analytic. In Algorithm 1, for any starting point, if there exists $\delta > 0$ such that (26) always holds, then the iterations $\omega^{(k,i)}$ converge to a critical point as a whole sequence.

Let $\nabla_i f(\omega) \in \mathbb{R}^{n_i \times r_i}$ be the partial derivative of the cost function (1) with respect to the $i$-th block matrix $U^{(i)}$ at $\omega$ point. Note that (1) is smooth and $\Omega$ is compact. There exists $\rho > 0$ such that

$$\|\nabla_i f(\omega) - \nabla_i f(\omega')\| \leq \rho \|\omega - \omega'\| \quad (27)$$

for any $\omega, \omega' \in \Omega$ and $1 \leq i \leq d$. Define

$$p_{(k,i)} : \text{St}(r_i, n_i) \rightarrow \mathbb{R}^+, \quad U \mapsto f(U_{k-1}^{(1)}; \cdots; U_{k-1}^{(i-1)}; U; U_{k-1}^{(i+1)}; \cdots; U_{k-1}^{(d)}).$$

Let $r_{\text{max}} \overset{\text{def}}{=} \max(r_1, \cdots, r_d)$. Now we need some lemmas before proving Theorem 4.5 and Theorem 4.6.

Lemma 4.7. Suppose that the cost function (1) is partially convex. Then, for any starting point, the cost function (1) converges increasingly in Algorithm 1.

Proof. By Lemma 4.1(i) and (6), we get that

$$\frac{1}{\alpha} h_{(k,i)}(U_{k-1}^{(i)}) = \langle U_{k-1}^{(i)}, \nabla h_{(k,i)}(U_{k-1}^{(i)}) \rangle \leq \langle U_k^{(i)}, \nabla h_{(k,i)}(U_{k-1}^{(i)}) \rangle. \quad (28)$$

Then, it follows from (7) that

$$\langle U_k^{(i)}, \nabla h_{(k,i)}(U_{k-1}^{(i)}) \rangle \leq \langle U_k^{(i)}, \nabla h_{(k,i)}(U_{k-1}^{(i)}) \rangle = \frac{1}{\alpha} h_{(k,i)}(U_k^{(i)}). \quad (29)$$

Note that $h_{(k,i)}$ is continuous, and thus upper bounded. The proof is complete. \qed
Lemma 4.8. Let $U, U' \in \text{St}(r,n)$ and $P \in \mathbb{R}^{r \times r}$ be positive semidefinite. Suppose that $\sigma_{\min}(P) > 0$. Then

$$\langle U - U', UP \rangle \geq \frac{1}{2} \sigma_{\min}(P) \|U - U'\|^2.$$ 

Proof. Let $P = Q^TDQ$ be the spectral decomposition with $\text{diag}\{D\} = (\sigma_1, \cdots, \sigma_r)^T$. Denote that $UQ^T = [u_1, \cdots, u_r]$ and $U'Q^T = [u'_1, \cdots, u'_r]$. Then

$$\langle U - U', UP \rangle = \langle UQ^T - U'Q^T, UQ^T \rangle = \sum_{i=1}^r \sigma_i(1 - \langle u_i, u'_i \rangle).$$

Note that $\|U - U'\|^2 = 2 \sum_{i=1}^r (1 - \langle u_i, u'_i \rangle)$. The proof is complete. \hfill \Box

Lemma 4.9. Suppose that the cost function (1) is partially convex. If there exists $\delta > 0$ such that (26) always holds in Algorithm 1, then

$$h_{(k,i)}(U_{k-1}^{(i)}) - h_{(k,i)}(U_{k-1}^{(i)}) \geq \frac{\alpha \delta}{2} \|U_k^{(i)} - U_{k-1}^{(i)}\|^2.$$ 

Proof. Let $\nabla h_{(k,i)}(U_{k-1}^{(i)}) = U_k^{(i)}P$ be the polar decomposition. Then, by (28), (29) and Lemma 4.8, we see that

$$h_{(k,i)}(U_{k}^{(i)}) - h_{(k,i)}(U_{k-1}^{(i)}) \geq \alpha ((U_k^{(i)}, \nabla h_{(k,i)}(U_{k-1}^{(i)})) - (U_{k-1}^{(i)}, \nabla h_{(k,i)}(U_{k-1}^{(i)})))$$

$$= \alpha \langle U_k^{(i)} - U_{k-1}^{(i)}, U_k^{(i)}P \rangle$$

$$\geq \frac{\alpha \delta}{2} \|U_k^{(i)} - U_{k-1}^{(i)}\|^2.$$ 

The proof is complete. \hfill \Box

Corollary 4.10. Suppose that the cost function (1) is partially convex. In Algorithm 1, if $\sigma_{\min}(\nabla h_{(k,i)}(U_{k-1}^{(i)})) > 0$, then $f(\omega^{(k,i)}) = f(\omega^{(k,i-1)})$ implies that $\omega^{(k,i)} = \omega^{(k,i-1)}$.

Corollary 4.11. Suppose that the cost function (1) is partially convex. If there exists $\delta > 0$ such that (26) always holds in Algorithm 1, then

$$f(\omega^{(k)}) - f(\omega^{(k-1)}) \geq \frac{\alpha \delta}{2} \|\omega^{(k)} - \omega^{(k-1)}\|^2.$$ 

Lemma 4.12. Let $h : \text{St}(r,n) \to \mathbb{R}$ be a differentiable function and $U \in \text{St}(r,n)$. Suppose that $\nabla h(U) \neq 0$ and $\nabla h(U) = U_P$ is the polar decomposition. Then

$$\|U_* - U\| \geq \frac{\|\nabla h(U)\|}{(r + 1)\|\nabla h(U)\|}.$$
Proof. Note that \(\|\nabla h(U)\| = \|P\|\). By [3, (3.35)], we see that
\[
\|\text{grad} h(U)\| = \frac{1}{2} \| (\nabla h(U) - UU^T \nabla h(U)) + (\nabla h(U) - U \nabla h(U)^T U) \|
\leq \frac{1}{2} \| (U_* - U) + UU^T (U - U_*) \| \|P\|
+ \frac{1}{2} \| (U_* - U) P + UP(U^T - U_*)^T U \|
\leq (r + 1) \|\nabla h(U)\| \|U_* - U\|.
\]
The proof is complete. \(\square\)

Lemma 4.13. Suppose that the cost function (1) is partially convex. If there exists \(\delta > 0\) such that (26) always holds in Algorithm 1, then
\[\|\omega^{(k)} - \omega^{(k-1)}\| \geq \frac{\|\text{grad} f(\omega^{(k-1)})\|}{\sqrt{d}(1 + r_{\max})(\rho + \Delta_*}).\]

Proof. By [3, (3.35)], (27) and Lemma 4.12, for any \(1 \leq i \leq d\), we get that
\[
\|\text{grad} p_{(k,i)}(U_{k-1}^{(i)})\| = \|\nabla p_{(k,i)}(U_{k-1}^{(i)}) - U_{k-1}^{(i)} \text{sym}(U_{k-1}^{(i)^T} \nabla p_{(k,i)}(U_{k-1}^{(i)}))\|
\leq \|\nabla p_{(k,i)}(U_{k-1}^{(i)}) - \nabla h_{(k,i)}(U_{k-1}^{(i)})\|
+ \|\nabla h_{(k,i)}(U_{k-1}^{(i)}) - U_{k-1}^{(i)} \text{sym}(U_{k-1}^{(i)^T} \nabla h_{(k,i)}(U_{k-1}^{(i)}))\|
+ \|U_{k-1}^{(i)} \text{sym}(U_{k-1}^{(i)^T} \nabla h_{(k,i)}(U_{k-1}^{(i)})) - U_{k-1}^{(i)} \text{sym}(U_{k-1}^{(i)^T} \nabla p_{(k,i)}(U_{k-1}^{(i)}))\|
\leq (1 + \|U_{k-1}^{(i)}\|^2) \|\nabla_i f(\omega^{(k-1)}) - \nabla_i f(\omega^{(k-1)})\| + \|\text{grad} h_{(k,i)}(U_{k-1}^{(i)})\|
\leq (1 + r_{\max}) \rho \|\omega^{(k-1)} - \omega^{(k)}\| + (1 + r_{\max}) \Delta_* \|U_{k-1}^{(i)} - U_k^{(i)}\|
\leq (1 + r_{\max})(\rho + \Delta_*) \|\omega^{(k-1)} - \omega^{(k)}\|.
\]
It follows that
\[\|\text{grad} f(\omega^{(k-1)})\| \leq \sqrt{d}(1 + r_{\max})(\rho + \Delta_*) \|\omega^{(k)} - \omega^{(k-1)}\|.
\]
The proof is complete. \(\square\)

The following result is direct by Lemma 4.9, Lemma 4.12 and Remark 4.2.

Proposition 4.14. Suppose that the cost function (1) is partially convex. If there exists \(\delta > 0\) such that (26) always holds in Algorithm 1, then
\[h_{(k,i)}(U_{k}^{(i)}) - h_{(k,i)}(U_{k-1}^{(i)}) \geq \frac{\alpha \delta}{2(1 + r_{\max})^2 \Delta_*^2} \|\text{grad} h_{(k,i)}(U_{k-1}^{(i)})\|^2.
\]
The following result is direct by Corollary 4.11 and Lemma 4.13.
Proposition 4.15. Suppose that the cost function (1) is partially convex. If there exists \( \delta > 0 \) such that (26) always holds in Algorithm 1, then
\[
\left. f(\omega^{(k)}) - f(\omega^{(k-1)}) \geq \frac{\alpha \delta}{2\sqrt{d}(1 + r_{\max})(\rho + \Delta)} \left\| \omega^{(k)} - \omega^{(k-1)} \right\| \| \text{grad} f(\omega^{(k-1)}) \right. ,
\]
(30)

Proof of Theorem 4.5. Let \( \omega_s = (U_s^{(1)}, U_s^{(2)}, \ldots, U_s^{(d)}) \) be an accumulation point in Algorithm 1. Assume that
\[
\text{grad} f(\omega_s) = (\text{grad} h_{(1)}(U_s^{(1)}), \ldots, \text{grad} h_{(d)}(U_s^{(d)})) \neq 0.
\]
We assume that \( \| \text{grad} h_{(d)}(U_s^{(d)}) \| \neq 0 \) without loss of generality. Note that \( \omega_s \) is an accumulation point. There exists a subsequence of the iterations \( \omega^{(k,i)} \) converging to \( \omega_s \). In this subsequence, we can choose \( i_0 \) such that the number of elements with \( i_0 \) is infinite. We assume that \( i_0 = 1 \) without loss of generality, and denote these elements by \( \omega^{(k_1,1)} \). Note that \( \| U_k^{(1)} - U_k^{(1)} \| \to 0 \) by Lemma 4.9. We see that \( \omega^{(k_1,d-1)} \) also converges to \( \omega_s \). Therefore, we get that
\[
\| \text{grad} h_{(k_1,d)}(U_{k_1-1}^{(d)}) \| \to \| \text{grad} h_{(d)}(U_s^{(d)}) \| \neq 0,
\]
which contradicts with Proposition 4.14. The proof is complete. \( \square \)

Proof of Theorem 4.6. It can be seen that (30) holds by Proposition 4.15, and \( f(\omega^{(k+1)}) = f(\omega^{(k)}) \) implies that \( \omega^{(k+1)} = \omega^{(k)} \) by Corollary 4.11. Then, by Theorem 2.3 and Remark 2.4, we deduce that the iterations \( \omega^{(k)} \) converge to \( \omega_s \) as a whole sequence. Moreover, by Theorem 4.5, we see that \( \omega_s \) is critical. Note that \( \| \omega^{(k,i)} - \omega^{(k,i-1)} \| \to 0 \) by Lemma 4.9. It is not difficult to see that the iterations \( \omega^{(k,i)} \) also converge to \( \omega_s \). The proof is complete. \( \square \)

5. Algorithm PDOI on the Stiefel Manifold

In this section, based on the matrix polar decomposition, we propose a symmetric variant of Algorithm 1 to solve problem (3), and establish its convergence. The algorithm and convergence results in this section apply to the best symmetric rank-1 approximation in Section 3.2.2 and the low rank symmetric orthogonal approximation in Section 3.3.2.

5.1. Algorithm PDOI. By the similar deduction in (25), if \( U_s \) is a maximal point of (3), it should satisfy that
\[
\nabla g(U_s) = U_s \text{sym}(U_s^T \nabla g(U_s)).
\]
(31)

Note that \( U_s \) is columnly orthogonal and \( \text{sym}(U_s^T \nabla g(U_s)) \) is symmetric. It can be seen that (31) is close to a polar decomposition form of \( \nabla g(U_s) \). Inspired by the decomposition form in (31), we propose the following polar decomposition based orthogonal iteration (PDOI) algorithm to solve problem (3). We assume that the starting point \( U_0 \) satisfies that \( g(U_0) > 0 \) without loss of generality.
Algorithm 2. (PDOI)

Input: starting point \(U_0\).
Output: \(U_k\), \(k \geq 1\).

- For \(k = 1, 2, \ldots\),
  - Compute \(\nabla g(U_{k-1})\).
  - Compute the polar decomposition of \(\nabla g(U_{k-1})\).
  - Update \(U_k\) to be the orthogonal polar factor.

Until convergence

Remark 5.1. Note that the cost function (3) is smooth and \(\text{St}(r, n)\) is compact. We denote \(\Delta \overset{\text{def}}{=} \max_{U \in \text{St}(r, n)} \|\nabla g(U)\|\).

Remark 5.2. By Lemma 4.3, we know that, in Algorithm 2, if \(\nabla g(U_{k-1})\) is of full rank and \(\text{sym}(U_{k-1}^T \nabla g(U_{k-1}))\) is positive semidefinite, then \(U_k = U_{k-1}\) if and only if \(\text{grad} g(U_{k-1}) = 0\).

5.2. Convergence analysis. For Algorithm 2, we mainly prove the following results about its weak convergence and global convergence.

Theorem 5.3. Suppose that the cost function (3) is convex. If there exists \(\delta > 0\) such that
\[
\sigma_{\min}(\nabla g(U_{k-1})) > \delta
\]
always holds in Algorithm 2, then every accumulation point of the iterations \(U_k\) is critical.

Theorem 5.4. Suppose that the cost function (3) is convex and real analytic. In Algorithm 2, for any starting point, if there exists \(\delta > 0\) such that (32) always holds, then the iterations \(U_k\) converge to a critical point as a whole sequence.

Before proving Theorem 5.3 and Theorem 5.4, we need some lemmas, which can be deduced by the similar methods as in Section 4.2.

Lemma 5.5. Suppose that the cost function (3) is convex. Then, for any starting point, the cost function (3) converges increasingly in Algorithm 2.

Lemma 5.6. Suppose that the cost function (3) is convex. If there exists \(\delta > 0\) such that (32) always holds in Algorithm 2, then
\[
g(U_k) - g(U_{k-1}) \geq \frac{\beta \delta}{2} \|U_k - U_{k-1}\|^2.
\]

Corollary 5.7. Suppose that the cost function (3) is convex. In the \(k\)-th iteration of Algorithm 2, if \(\sigma_{\min}(\nabla g(U_{k-1})) > 0\), then \(g(U_k) = g(U_{k-1})\) implies that \(U_k = U_{k-1}\).
Proposition 5.8. Suppose that the cost function (3) is convex. If there exists $\delta > 0$ such that (32) always holds in Algorithm 2, then
(i) we have that
\[
g(U_k) - g(U_{k-1}) \geq \frac{\beta \delta}{2(r+1)\Delta} \| U_k - U_{k-1} \| \| \text{grad } g(U_{k-1}) \|. \tag{33}
\]
(ii) Moreover, it holds that
\[
g(U_k) - g(U_{k-1}) \geq \frac{\beta \delta}{2(r+1)^2\Delta^2} \| \text{grad } g(U_{k-1}) \|^2. \tag{34}
\]
Proof of Theorem 5.3. Assume that $U_*$ is an accumulation point and $\| \text{grad } g(U_*) \| \neq 0$. Then there exists a subsequence $\{U_{k_\ell}\}$ converging to $U_*$. Let $\ell \to \infty$. We see that, in (34), the left side converges to 0, while the right side converges to $\frac{\beta \delta}{2(r+1)^2\Delta^2} \| \text{grad } g(U_*) \|^2 > 0$. This is impossible, and thus $U_*$ is critical. The proof is complete. \hfill \Box

Proof of Theorem 5.4. It can be seen that (33) holds by Proposition 5.8(i), and $g(U_{k+1}) = g(U_k)$ implies that $U_{k+1} = U_k$ by Lemma 5.6. Then, by Theorem 2.3 and Remark 2.4, we can deduce that the iterations converge to $U_*$ as a whole sequence. Moreover, by Theorem 5.3, we see that $U_*$ is critical. \hfill \Box

6. Algorithms for the best rank-1 approximation

6.1. Algorithm HOPM. Let $k \geq 1$ and
\[
v^{(k,i)} \overset{\text{def}}{=} \mathcal{A} \cdot u^{(1)}_k \cdots \cdot u^{(i-1)}_k \cdot u^{(i+1)}_k \cdots \cdot u^{(d)}_k T.
\]
Note that
\[
\nabla h^{(k,i)}(u^{(i)}_{k-1}) = 2\langle v^{(k,i)}, u^{(i)}_{k-1} \rangle v^{(k,i)} \tag{35}
\]
by (12). Then Algorithm 1 for (11) can be written as follows, which is the well-known higher order power method (HOPM) [16, 17] algorithm.

Algorithm 3. (HOPM)
Input: $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, starting point $\omega^{(0)} = (u^{(1)}_0, u^{(2)}_0, \ldots, u^{(d)}_0)$.
Output: $\omega^{(k,i)} = (u^{(1)}_k, \ldots, u^{(i)}_k, u^{(i+1)}_k, \ldots, u^{(d)}_k)$, $k \geq 1$, $1 \leq i \leq d$.

- For $k = 1, 2, \ldots$,
- For $i = 1, 2, \ldots, d$
- Compute $v^{(k,i)}$.
- Update $u^{(i)}_k = \frac{v^{(k,i)}}{\| v^{(k,i)} \|}$.
- End for
- Until convergence
The equation \[ (3.9) \] obtained by Lagrange multipliers is in fact a special case of the decomposition form in \((25)\). In Algorithm 3, since \(|\langle v^{(k,i)} , u_{k-1}^{(i)} \rangle| = f(\omega^{(k,i-1)})^{1/2}\) is increasing by Lemma 4.7, we can deduce that there exists \(\varsigma > 0\) such that \(\|v^{(k,i)}\| > \varsigma\). Therefore, by \((35)\), it always holds that
\[
\|\nabla h^{(k,i)}(u_{k-1}^{(i)})\| = 2f(\omega^{(k-1,i)})^{1/2}\|v^{(k,i)}\| \geq 2\varsigma f(\omega^{(0)})^{1/2} > 0. \tag{36}
\]

The following result is direct by Theorem 4.6 and \((36)\).

**Corollary 6.1.** In Algorithm 3, for any starting point, the iterations \(\omega^{(k,i)}\) converge to a critical point as a whole sequence.

**Remark 6.2.** It was proved in [46, Theorem 11], [21, Theorem 4.4] and [21, Corollary 5.4] that the iterations \(\omega^{(k)}\) in Algorithm 3 converge to a critical point as a whole sequence, and the convergence rate is linear for almost all tensors in \(\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}\).

### 6.2. Algorithm S-HOPM

Let \(k \geq 1\) and \(v^{(k)} \equiv A \cdot_1 u_{k-1}^T \cdots \cdot_{d-1} u_{k-1}^T\). Note that
\[
\nabla g(u_{k-1}) = 2d\langle v^{(k)}, u_{k-1} \rangle v^{(k)} \tag{37}
\]
by \((9)\) and \((12)\). Then Algorithm 2 for \((14)\) can be written as follows, which is the well-known symmetric higher order power method (S-HOPM) [17, 27].

**Algorithm 4.** (S-HOPM)

**Input:** \(A \in \text{symm}(\mathbb{R}^{n \times n \times \cdots \times n})\), starting point \(u_0\).

**Output:** \(u_k, k \geq 1\).

- For \(k = 1, 2, \ldots\),
- Compute \(v^{(k)}\),
- Update \(u_k = v^{(k)} / \|v^{(k)}\|\).
- Until convergence

Suppose that \((14)\) is convex. In Algorithm 4, because \(|\langle v^{(k)}, u_{k-1} \rangle| = g(u_{k-1})^{1/2}\) is increasing by Lemma 5.5, we can deduce that there exists \(\varsigma > 0\) such that \(\|v^{(k)}\| > \varsigma\). Therefore, by \((37)\), it always holds that
\[
\|\nabla g(u_{k-1})\| = 2dg(u_{k-1})^{1/2}\|v^{(k)}\| \geq 2\varsigma g(u_0)^{1/2} > 0. \tag{38}
\]

The following result is direct by Theorem 5.4 and \((38)\).

**Corollary 6.3.** If the cost function \((14)\) is convex, then in Algorithm 4, for any starting point, the iterations \(u_k\) converge to a critical point as a whole sequence.

**Remark 6.4.** The weak convergence of Algorithm 4 was studied in [27, Theorem 4].
7. Algorithms for the Low Rank Orthogonal Approximation

7.1. Algorithm LROAT. Let \( k \geq 1 \) and \( 1 \leq i \leq d \). Denote \( U_k^{(i)} = [u_1^{(k,i)}, \ldots, u_r^{(k,i)}] \in \text{St}(r, n_i) \) and

\[
\begin{align*}
\mathcal{V}^{(k,i)} & \equiv \mathcal{A} \cdot_1 U_1^{(1)} \cdots \cdot_{i-1} U_k^{(i-1)} \cdot_{i+1} U_{k-1}^{(i+1)} \cdots \cdot_d U_{k-1}^{(d)} \cdot \mathcal{U}_k^{(i)} , \\
v^{(k,i,q)} & \equiv \mathcal{A} \cdot_1 u_1^{(k,1)} \cdots \cdot_{i-1} u_q^{(k,i-1)} \cdot_{i+1} u_{q+1}^{(k+1,i+1)} \cdots \cdot_d u_d^{(d)} \mathcal{U}_k^{(i)} \in \mathbb{R}^{n_i},
\end{align*}
\]

for \( 1 \leq q \leq r \). Let \( \mathcal{W}^{(k,i-1)} \equiv \mathcal{V}^{(k,i)} \cdot \mathcal{U}_k^{(i)} \). Note that

\[
\nabla h^{(k,i)}(U_k^{(i)}) = 2[v^{(k,i,1)}, \ldots, v^{(k,i,r)}] \begin{bmatrix}
\mathcal{W}_1^{(k,i-1)} \\
\vdots \\
\mathcal{W}_r^{(k,i-1)}
\end{bmatrix}
\]

by (18). Then Algorithm 1 for (17) can be written as follows, which is the low rank orthogonal approximation of tensors (LROAT) algorithm in [7, Section 5.4].

Algorithm 5. (LROAT)
Input: starting point \( \omega^{(0)} \).
Output: \( \omega^{(k,i)}, k \geq 1, 1 \leq i \leq d \).

- For \( k = 1, 2, \ldots \),
- For \( i = 1, 2, \ldots, d \),
- Compute \( \nabla h^{(k,i)}(U_k^{(i)}) \) by (39).
- Compute the polar decomposition of \( \nabla h^{(k,i)}(U_k^{(i)}) \).
- Update \( U_k^{(i)} \) to be the orthogonal polar factor.
- End for
- Until convergence

The decomposition form in (25) for cost function (17) is the same as [7, (5.16)]. The following result is direct by Theorem 4.5.

Corollary 7.1. In Algorithm 5, for any starting point, if there exists \( \delta > 0 \) such that (26) always holds, then the iterations \( \omega^{(k,i)} \) converge to a critical point as a whole sequence.

Remark 7.2. In [7, Theorem 5.7], the weak convergence of iterations \( \omega^{(k)} \) in Algorithm 5 was proved under the condition that \( \nabla h^{(k,i)}(U_k^{(i)}) \) is always of full rank. Very recently, we found that an epsilon-alternating least square method was proposed for solving a problem more general than (16), and its global convergence was established [49]. Meanwhile, the global convergence of an improved version of Algorithm 5 was established and its convergence rate was studied [22].
7.2. Algorithm S-LROAT. Let $k \geq 1$. Denote $U_k = [u_1^{(k)}, \ldots, u_r^{(k)}] \in \text{St}(r, n)$ and

$$V^{(k)} \overset{\text{def}}{=} A \cdot_1 U_{k-1}^T \cdots \cdot_{d-1} U_{k-1}^T,$$

$$v^{(k,q)} \overset{\text{def}}{=} A \cdot_1 u_{q}^{(k-1)^T} \cdots \cdot_{d-1} u_{q}^{(k-1)^T} \in \mathbb{R}^n,$$

for $1 \leq q \leq r$. Let $W^{(k-1)} \overset{\text{def}}{=} V^{(k)} \cdot_{1} U_{k-1}^T$. Note that

$$\nabla g(U_{k-1}) = 2d[v^{(k,1)}, \ldots, v^{(k,r)}] W^{(k-1)} \cdots W^{(k-1)}$$

by (9) and (18). Then Algorithm 2 for (20) can be written as follows, which is the symmetric variant of LROAT (S-LROAT) in [7, Section 5.6].

Algorithm 6. (S-LROAT)
Input: starting point $U_0$.
Output: $U_k$, $k \geq 1$.

- For $k = 1, 2, \ldots$, 
- Compute $\nabla g(U_{k-1})$ by (40).
- Compute the polar decomposition of $\nabla g(U_{k-1})$.
- Update $U_k$ to be the orthogonal polar factor.
- Until convergence

The following result is direct by Theorem 5.4.

Corollary 7.3. Suppose that the cost function (20) is convex. Then in Algorithm 6, for any starting point, if there exists $\delta > 0$ such that (32) always holds, then the iterations $U_k$ converge to a critical point as a whole sequence.

Remark 7.4. Problem (20) has also been studied in [35, 40]. In [40], based on the decomposition form [40, (11)] which is similar to [7, (5.16)], an iterative algorithm was proposed. Then the convergence of its shifted variant was studied. In [35], problem (20) is transformed to an equivalent problem on the orthogonal group $O_n$. Then the Jacobi low rank orthogonal approximation algorithm (JLROA) was proposed, which includes the well-known Jacobi CoM algorithm [10, 14] as a special case.

8. Algorithms for the low multilinear rank approximation

8.1. Algorithm LMPD. Let $k \geq 1$ and $1 \leq i \leq d$. Denote that

$$V^{(k,i)} \overset{\text{def}}{=} A \cdot_1 U_{k}^{(i)^T} \cdots \cdot_{i-1} U_{k}^{(i-1)^T} \cdot_{i+1} U_{k-1}^{(i+1)^T} \cdots \cdot_d U_{k-1}^{(d)^T}.$$
Let \( \mathbf{V}_{(k,i)}^{(k,i)} = [\mathbf{v}_1^{(k,i)}, \mathbf{v}_2^{(k,i)}, \ldots, \mathbf{v}_{R_i}^{(k,i)}] \in \mathbb{R}^{n_i \times R_i} \) and \( \mathbf{V}^{(k,i)} \equiv \sum_{t=1}^{R_i} \mathbf{v}_t^{(k,i)} \mathbf{v}_t^{(k,i)T} \) as in Section 3.4. Then
\[
\nabla h_{(k,i)}(U_{k-1}^{(i)}) = 2\mathbf{V}^{(k,i)} U_{k-1}^{(i)},
\]
by (24). We call Algorithm 1 for (23) the low multilinear rank approximation based on polar decomposition (LMPD) algorithm, which can be shown as follows.

Algorithm 7. (LMPD)

**Input:** \( \mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \), \((r_1, \cdots, r_d)\), starting point \( \omega^{(0)} \).

**Output:** \( \omega^{(k,i)}, k \geq 1, 1 \leq i \leq d \).

- **For** \( k = 1, 2, \ldots \),
  - **For** \( i = 1, 2, \ldots, d \)
    - Compute \( \nabla h_{(k,i)}(U_{(k-1)}^{(i)}) \) by (41).
    - Compute the polar decomposition of \( \nabla h_{(k,i)}(U_{(k-1)}^{(i)}) \).
    - Update \( U_{k}^{(i)} \) to be the orthogonal polar factor.
  - **End for**
- **Until** convergence

For the cost function (23), we can also derive the decomposition form in (25) by the similar method as in [7, (5.16)]. We omit the details here. The following result is direct by Theorem 4.6.

**Corollary 8.1.** In Algorithm 7, for any starting point, if there exists \( \delta > 0 \) such that (26) always holds, then the iterations \( \omega^{(k,i)} \) converge to a critical point as a whole sequence.

8.2. Algorithm LMPD-S. Inspired by condition (26), we propose the following shifted variant of Algorithm 7, which is called Algorithm LMPD-S in this paper.

Algorithm 8. (LMPD-S)

**Input:** \( \mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \), \((r_1, \cdots, r_d)\), starting point \( \omega^{(0)} \), \( \gamma > 0 \).

**Output:** \( \omega^{(k,i)}, k \geq 1, 1 \leq i \leq d \).

- **For** \( k = 1, 2, \ldots \),
  - **For** \( i = 1, 2, \ldots, d \)
    - Compute \( \nabla h_{(k,i)}(U_{(k-1)}^{(i)}) \) by (41).
    - Compute the polar decomposition of \( \nabla h_{(k,i)}(U_{(k-1)}^{(i)}) + \gamma U_{k-1}^{(i)} \).
    - Update \( U_{k}^{(i)} \) to be the orthogonal polar factor.
  - **End for**
- **Until** convergence

Note that \( V^{(k,i)} \) is positive semidefinite in (41). Then \( \sigma_{\min}(\nabla h_{(k,i)}(U_{(k-1)}^{(i)}) + \gamma U_{k-1}^{(i)}) \geq \gamma \) always holds in Algorithm 8. We can prove the following convergence result.
Theorem 8.2. In Algorithm 8, for any starting point, the iterations \( \omega^{(k,i)} \) converge to a critical point as a whole sequence.

Since the proof of Theorem 8.2 is very similar to the deductions in Section 4, we omit the details, and only present some important intermediate lemmas.

Lemma 8.3. In Algorithm 8, for any starting point, the cost function (23) converges increasingly.

Lemma 8.4. Let \( h: \text{St}(r,n) \rightarrow \mathbb{R} \) be a differentiable function and \( U \in \text{St}(r,n) \). Suppose that \( \nabla h(U) + \gamma U = U^*P \) is the polar decomposition. Then

\[
\|U^* - U\| \geq \frac{\|\text{grad} f(U)\|}{(r + 1)\|\nabla h(U) + \gamma U\|}.
\]

Lemma 8.5. In Algorithm 8, for any \( k \geq 1 \) and \( 1 \leq i \leq d \), we have that

(i) \( h_{(k,i)}(U^{(i)}_k) - h_{(k,i)}(U^{(i)}_{k-1}) \geq \frac{\alpha \gamma}{2} \|U^{(i)}_k - U^{(i)}_{k-1}\|^2 \).

(ii) \( \|\omega^{(k)} - \omega^{(k-1)}\| \geq \frac{\|\text{grad} f(\omega^{(k-1)})\|}{\sqrt{d(1 + r_{\max})(\rho + \Delta_s + \gamma \sqrt{r})}} \).

(iii) \( f(\omega^{(k)}) - f(\omega^{(k-1)}) \geq \frac{\alpha \gamma}{2\sqrt{d(1 + r_{\max})(\rho + \Delta_s + \gamma \sqrt{r})}}\|\omega^{(k)} - \omega^{(k-1)}\|\|\text{grad} f(\omega^{(k-1)})\| \).

8.3. Experiments. To solve problem (21), many other methods [23] have been developed, such as the HOOI algorithm [17, 48], the Riemannian trust-region algorithm [25] and the Jacobi-type algorithm [24]. In this subsection, we conduct experiments to compare the performances of HOOI, LMPD, and LMPD-S.

Example 8.6. We randomly generate one tensor in \( \mathbb{R}^{5 \times 5 \times 5} \), and run HOOI, LMPD and LMPD-S (\( \gamma = 0.01 \)). The results of cost function value (23) are shown in Figure 1. It can be seen that LMPD and LMPD-S have comparable speed with HOOI.

Example 8.7. We randomly generate two tensors in \( \mathbb{R}^{10 \times 10 \times 10} \), and run HOOI, LMPD and LMPD-S (\( \gamma = 0.01 \)) with \((r_1, r_2, r_3) = (1,1,2)\). The distances of successive iterations \( \|\omega^{(k,i+1)} - \omega^{(k,i)}\| \) are shown in Figure 2. It can be seen that HOOI and LMPD fail to converge globally, while LMPD-S has a much better convergence performance.

9. Conclusions

Motivated by the works [7, 21, 27, 30, 34, 46, 47], in this paper, we propose a general algorithmic framework to solve a class of optimization problems on the product of Stiefel manifolds satisfying (6) or its symmetric variant. In Example 8.7, we observe that LMPD fails to converge globally. Therefore, condition (26) is necessary in the global convergence assurance. A natural question is whether or not condition (26) can be relaxed further, say to only require \( \nabla h_{(k,i)}(U^{(i)}_{k-1}) \) to have a full rank.
Figure 1. Results for the tensor randomly generated in Example 8.6.
Figure 2. Results for the tensors randomly generated in Example 8.7.
REFERENCES

[1] P.-A. Absil, C. G. Baker, and K. A. Gallivan, Trust-region methods on riemannian manifolds, Foundations of Computational Mathematics, 7 (2007), pp. 303–330.

[2] P. A. Absil, R. Mahony, and B. Andrews, Convergence of the iterates of descent methods for analytic cost functions, SIAM Journal on Optimization, 16 (2005), pp. 531–547.

[3] P.-A. Absil, R. Mahony, and R. Sepulchre, Optimization algorithms on matrix manifolds, Princeton University Press, 2009.

[4] R. L. Adler, J.-P. Dedieu, J. Y. Margulies, M. Martens, and M. Shub, Newton’s method on riemannian manifolds and a geometric model for the human spine, IMA Journal of Numerical Analysis, 22 (2002), pp. 359–390.

[5] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky, Tensor decompositions for learning latent variable models, Journal of Machine Learning Research, 15 (2014), pp. 2773–2832.

[6] S. Boyd and L. Vandenberghe, Convex optimization, Cambridge university press, 2004.

[7] J. Chen and Y. Saad, On the tensor SVD and the optimal low rank orthogonal approximation of tensors, SIAM Journal on Matrix Analysis and Applications, 30 (2009), pp. 1709–1734.

[8] A. Cichocki, D. Mandic, L. D. Lathauwer, G. Zhou, Q. Zhao, C. Cai, and H. A. PHAN, Tensor decompositions for signal processing applications: From two-way to multiway component analysis, IEEE Signal Processing Magazine, 32 (2015), pp. 145–163.

[9] P. Comon, Independent Component Analysis, in Higher Order Statistics, J.-L. Lacoume, ed., Elsevier, Amsterdam, London, 1992, pp. 29–38.

[10] P. Comon, Independent component analysis, a new concept?, Signal Processing, 36 (1994), pp. 287–314.

[11] P. Comon, Tensor Diagonalization, A useful Tool in Signal Processing, in 10th IFAC Symposium on System Identification (IFAC-SYSID), M. Blanke and T. Soderstrom, eds., vol. 1, Copenhagen, Denmark, July 1994, IEEE, pp. 77–82.

[12] P. Comon, Tensors: a brief introduction, IEEE Signal Processing Magazine, 31 (2014), pp. 44–53.

[13] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, Symmetric tensors and symmetric tensor rank, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1254–1279.

[14] P. Comon and C. Jutten, eds., Handbook of Blind Source Separation, Academic Press, Oxford, 2010.

[15] L. De Lathauwer, Signal processing based on multilinear algebra, Katholieke Universiteit Leuven, 1997.

[16] L. De Lathauwer, P. Comon, B. De Moor, and J. Vandewalle, Higher-order power method - application in independent component analysis, in Proc. of the International Symposium on Nonlinear Theory and its Applications (NOLTA’95), 1995, pp. 91–96.

[17] L. De Lathauwer, B. De Moor, and J. Vandewalle, On the best rank-1 and rank-(r1 ,r2 ,.. .,rn) approximation of higher-order tensors, SIAM Journal on Matrix Analysis and Applications, 21 (2000), pp. 1324–1342.

[18] A. Edelman, T. A. Arias, and S. T. Smith, The geometry of algorithms with orthogonality constraints, SIAM journal on Matrix Analysis and Applications, 20 (1998), pp. 303–353.

[19] G. H. Golub and C. F. Van Loan, Matrix computations, vol. 3, JHU press, 2012.

[20] G. Hu, Y. Hua, Y. Yuan, Z. Zhang, Z. Lu, S. S. Mukherjee, T. M. Hospedales, N. M. Robertson, and Y. Yang, Attribute-enhanced face recognition with neural tensor fusion networks, in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 3744–3753.

[21] S. Hu and G. Li, Convergence rate analysis for the higher order power method in best rank one approximations of tensors, Numerische Mathematik, 140 (2018), pp. 993–1031.
[22] S. Hu and K. Ye, Linear convergence of an alternating polar decomposition method for low rank orthogonal tensor approximations, arXiv:1912.04085, (2019).
[23] M. Ishteva, Numerical methods for the best low multilinear rank approximation of higher-order tensors, PhD thesis, Department of Electrical Engineering, Katholieke Universiteit Leuven, (2009).
[24] M. Ishteva, P.-A. Absil, and P. Van Dooren, Jacobi algorithm for the best low multilinear rank approximation of symmetric tensors, SIAM J. Matrix Anal. Appl., 2 (2013), pp. 651–672.
[25] M. Ishteva, P.-A. Absil, S. Van Huffel, and L. De Lathauwer, Best low multilinear rank approximation of higher-order tensors, based on the riemannian trust-region scheme, SIAM Journal on Matrix Analysis and Applications, 32 (2011), pp. 115–135.
[26] A. Karami, M. Yazdi, and G. Mercier, Compression of hyperspectral images using discrete wavelet transform and tucker decomposition, IEEE journal of selected topics in applied earth observations and remote sensing, 5 (2012), pp. 444–450.
[27] E. Kofidis and P. A. Regalia, On the best rank-1 approximation of higher-order supersymmetric tensors, SIAM Journal on Matrix Analysis and Applications, 23 (2002), pp. 863–884.
[28] T. G. Kolda, Orthogonal tensor decompositions, SIAM Journal on Matrix Analysis and Applications, 23 (2001), pp. 243–255.
[29] T. G. Kolda and B. W. Bader, Tensor decompositions and applications, SIAM review, 51 (2009), pp. 455–500.
[30] T. G. Kolda and J. R. Mayo, Shifted power method for computing tensor eigenpairs, SIAM Journal on Matrix Analysis and Applications, 32 (2011), pp. 1095–1124.
[31] S. Krantz and H. Parks, A Primer of Real Analytic Functions, Birkhäuser Boston, 2002.
[32] J. B. Kruskal, Rank, decomposition, and uniqueness for 3-way and n-way arrays, Multiway data analysis, (1989), pp. 7–18.
[33] S. Łojasiewicz, Ensembles semi-analytiques, IHES notes, (1965).
[34] J. Li, K. Usevich, and P. Comon, Globally convergent jacobi-type algorithms for simultaneous orthogonal symmetric tensor diagonalization, SIAM Journal on Matrix Analysis and Applications, 39 (2018), pp. 1–22.
[35] J. Li, K. Usevich, and P. Comon, Jacobi-type algorithm for low rank orthogonal approximation of symmetric tensors and its convergence analysis, arXiv:1911.00659, (2019).
[36] J. Li, K. Usevich, and P. Comon, On approximate diagonalization of third order symmetric tensors by orthogonal transformations, Linear Algebra and its Applications, 576 (2019), pp. 324–351.
[37] J. Li, K. Usevich, and P. Comon, On the convergence of jacobi-type algorithms for independent component analysis, arXiv:1912.07194, (2019).
[38] J. Li, X.-P. Zhang, and T. Tran, Point cloud denoising based on tensor tucker decomposition, in 2019 26th IEEE International Conference on Image Processing (ICIP), IEEE, 2019.
[39] S. Łojasiewicz, Sur la géométrie semi-et sous-analytique, Annales de l’institut Fourier, 43 (1993), pp. 1575–1595.
[40] J. Pan and M. K. Ng, Symmetric orthogonal approximation to symmetric tensors with applications to image reconstruction, Numerical Linear Algebra with Applications, 25 (2018), p. e2180.
[41] L. Qi and Z. Luo, Tensor analysis: Spectral theory and special tensors, SIAM, 2017.
[42] L. Qi, F. Wang, and Y. Wang, Z-eigenvalue methods for a global polynomial optimization problem, Mathematical Programming, 118 (2009), pp. 301–316.
[43] R. Schneider and A. Uschmajew, Convergence results for projected line-search methods on varieties of low-rank matrices via Łojasiewicz inequality, SIAM Journal on Optimization, 25 (2015), pp. 622–646.
[44] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, Tensor decomposition for signal processing and machine learning, IEEE Transactions on Signal Processing, 65 (2017), pp. 3551–3582.

[45] L. R. Tucker, Some mathematical notes on three-mode factor analysis, Psychometrika, 31 (1966), pp. 279–311.

[46] A. Uschmajew, A new convergence proof for the higher-order power method and generalizations, Pac. J. Optim., 11 (2015), pp. 309–321.

[47] K. Usevich, J. Li, and P. Comon, Approximate matrix and tensor diagonalization by unitary transformations: convergence of jacobi-type algorithms, arXiv:1905.12295, (2019).

[48] Y. Xu, On the convergence of higher-order orthogonal iteration, Linear and Multilinear Algebra, 66 (2018), pp. 2247–2265.

[49] Y. Yang, The epsilon-alternating least squares for orthogonal low-rank tensor approximation and its global convergence, arXiv:1911.10921, (2019).

[50] T. Zhang and G. H. Golub, Rank-one approximation to high order tensors, SIAM Journal on Matrix Analysis and Applications, 23 (2001), pp. 534–550.

[51] X. Zhang, C. Ling, and L. Qi, The best rank-1 approximation of a symmetric tensor and related spherical optimization problems, SIAM Journal on Matrix Analysis and Applications, 33 (2012), pp. 806–821.

(Jianze Li) Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen, China
E-mail address: lijianze@gmail.com

(Shuzhong Zhang) Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN 55455, USA.
E-mail address: zhangs@umn.edu