Linear convergence of an alternating polar decomposition method for low rank orthogonal tensor approximations

Shenglong Hu\textsuperscript{1} · Ke Ye\textsuperscript{2}

Received: 28 December 2019 / Accepted: 14 July 2022 / Published online: 30 July 2022
© Springer-Verlag GmbH Germany, part of Springer Nature and Mathematical Optimization Society 2022

Abstract
Low rank orthogonal tensor approximation (LROTA) is an important problem in tensor computations and their applications. A classical and widely used algorithm is the alternating polar decomposition method (APD). In this paper, an improved version iAPD of the classical APD is proposed. For the first time, all of the following four fundamental properties are established for iAPD: (i) the algorithm converges globally and the whole sequence converges to a KKT point without any assumption; (ii) it exhibits an overall sublinear convergence with an explicit rate which is sharper than the usual $O(1/k)$ for first order methods in optimization; (iii) more importantly, it converges $R$-linearly for a generic tensor without any assumption; (iv) for almost all LROTA problems, iAPD reduces to APD after finitely many iterations if it converges to a local minimizer.

Keywords Orthogonally decomposable tensors · Low rank orthogonal tensor approximation · $R$-linear convergence · Sublinear convergence · Global convergence

We claim that parts of the paper have not been published or were submitted elsewhere. This work is partially supported by National Science Foundation of China (Grant No. 11771328). The first author is also partially supported by National Science Foundation of China (Grant No. 12171128), Young Elite Scientists Sponsorship Program by Tianjin and the Natural Science Foundation of Zhejiang Province, China (Grant No. LD19A010002 and Grant No. LY22A010022). The second author is also partially supported by National Science Foundation of China (Grant No. 11801548 and Grant No. 11688101), National Key R&D Program of China (Grant No. 2018YFA0306702 and Grant No. 2020YFA0712300) and CAS Project for Young Scientists in Basic Research (Grant No. YSBR-008).

\textsuperscript{1}Department of Mathematics, School of Science, Hangzhou Dianzi University, Hangzhou 310018, China
\textsuperscript{2}KLMM, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

Ke Ye
keyk@amss.ac.cn
Shenglong Hu
shenglonghu@hdu.edu.cn
1 Introduction

As higher order generalizations of matrices, tensors (a.k.a. hypermatrices or multi-way arrays) are ubiquitous and inevitable in mathematical modeling and scientific computing [16, 27, 39, 50, 55]. Among numerous tensor problems studied in recent years, tensor approximation and its related topics have been becoming the main focus, see [15, 36, 39] and references therein. Applications of tensor approximation are diverse and broad, including signal processing [16], computational complexity [39], pattern recognition [2], principal component analysis [14], etc. We refer interested readers to surveys [15, 25, 36, 46] and books [27, 39, 55] for more details.

Singular value decomposition (SVD) of matrices is both a theoretical foundation and a computational workhorse for linear algebra, with applications spreading throughout scientific computing and engineering [24]. SVD of a given matrix is an orthogonal decomposition of the matrix [24, 30], and a truncated SVD according to the non-increasing singular values is a low rank orthogonal approximation of that matrix by the well-known Eckart-Young theorem [21]. While a higher order tensor cannot be diagonalized by orthogonal matrices in general [39], there are several generalizations of SVD from matrices to tensors, such as higher order SVD [18], orthogonally decomposable tensor decompositions and approximations and their variants, see [13, 23, 34, 53, 58, 64] and references therein. In this paper, we focus on the low rank orthogonal tensor approximation (abbreviated as LROTA) problem. It is a low rank tensor approximation problem with all the factor matrices being orthonormal [15, 19, 36]. This problem is of crucial importance in applications, such as blind source separation in signal processing and statistics [14–16, 50].

In the literature, several numerical methods have been proposed to solve this problem, such as Jacobi-type methods [14], for which the tensor considered usually has a symmetric structure. Interested readers are referred to [32, 43–45, 49, 62]. A more general problem is studied, where some factor matrices are orthonormal and the rest of them are unconstrained. We denote these low rank orthogonal tensor approximation problems by LROTA-s with s the number of orthonormal factor matrices. For simplicity, LROTA denotes the problem where all the factor matrices are orthonormal. For LROTA-s, a commonly adopted algorithmic framework is the alternating minimization method (AMM) [7], under which the alternating polar decomposition (APD) method is proposed and widely employed [13, 64]. Under a regularity condition that all matrices in certain iterative sequence are of full rank, it is proved that every converging subsequence generated by this method for LROTA converges to a stationary point of the objective function by Chen and Saad [13] in 2009. In 2012, Uschmajew established a local convergence result under some appropriate assumptions [60]. In 2015, Wang, Chu and Yu proposed an AMM with a modified polar decomposition for LROTA-1, and established the global convergence without any further assumption for a generic tensor [64]. In 2019, Guan and Chu [26] established the global convergence for LROTA-s with general s under a similar regularity condition as [13]. Very recently, Yang proposed an epsilon-alternating least square method to solve the
Linear convergence of iAPD

problem LROTA-s with general s and established its global convergence without any assumption [65]. On the other hand, the special case of rank-one tensor approximation was systematically studied since the work of De Lathauwer, De Moor, and Vandewalle [18]. A higher order power method, which is essentially an application of AMM, was proposed and global convergence results were established, see [52, 61, 63]. Moreover, the convergence rate was also estimated in [31, 60, 66].

Motivated by the development of the convergence analysis of the rank-one case and the general low rank case, a fundamental question is: *is there an algorithm for LROTA such that all the favorable convergence properties in the rank-one case also hold for the general low rank case?* Given the existence of such an algorithm, one would also hope that this algorithm should be as close as possible to the widely used and state of the art APD method, so that several questions raised in the literature can be addressed [13, 26, 64]. To achieve this, we propose the iAPD method, an improved version of the APD method. The new method is simply the APD method modified with adaptive proximal corrections and truncations. It turns out that proximal corrections guarantee the sufficient increase property of the iAPD method and truncations ensure that the iAPD method converges to a desired critical point. It is also worth mentioning that with some reasonable assumptions, we can prove that in iAPD, only finitely many corrections and truncations are necessary. This is also demonstrated by numerical experiments. Listed below are main contributions of this paper:

1. We prove the global convergence of iAPD without any assumption. This is exactly the content of the following theorem to be proved in Sect. 4.3.

**Theorem 2** (Global Convergence) Any sequence \( \{U[p]\} \) generated by iAPD (i.e., Algorithm 1) is bounded and converges to a KKT point of the LROTA problem (i.e., (9)).

2. We establish an overall sublinear convergence of iAPD and present an explicit eventual convergence rate in terms of the dimension and the order of target tensors. This result is formally stated in the next theorem, which is proved in Sect. 4.4.

**Theorem 3** (Sublinear Convergence Rate) Let \( \{U[p]\} \) be a sequence generated by Algorithm 1 for a given nonzero tensor \( A \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} \) and let \( \tau \) be defined by (50). The following statements hold:

(i) the sequence \( \{f(U[p])\} \) converges to \( f^* \), with sublinear convergence rate at least \( O(p^{-2\tau}) \), that is, there exist \( M_1 > 0 \) and \( p_0 \in \mathbb{N} \) such that for all \( p \geq p_0 \)

\[
    f^* - f(U[p]) \leq M_1 p^{-\frac{1}{2\tau}};
\]

(ii) \( \{U[p]\} \) converges to \( U^* \) globally with the sublinear convergence rate at least \( O(p^{-\frac{\tau - 1}{2\tau}}) \), that is, there exist \( M_2 > 0 \) and \( p_0 \in \mathbb{N} \) such that for all \( p \geq p_0 \)

\[
    \|U[p] - U^*\|_F \leq M_2 p^{\frac{\tau - 1}{2\tau}}.
\]

---

1 Yang’s paper is posted during our final preparation of this paper. We can see that we both employ the proximal technique. A difference is that the proximal correction in our algorithm is adaptive, and a theoretical investigation is also given (cf. Sect. 5) for the execution of the proximal correction.
The derived convergence rate is sharper than the usual convergence rate \( O(1/k) \) established for first order methods in the literature [6].

3. We prove that iAPD is linearly convergent for a generic tensor without any assumption. The following theorem, whose proof is provided in Sect. 4.5, is a rigorous statement of the aforementioned linear convergence result.

**Theorem 5 (Generic Linear Convergence)** If \( \{U[p]\} \) is a sequence generated by Algorithm 1 for a generic tensor \( A \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} \), then the sequence \( \{U[p]\} \) converges \( R \)-linearly to a KKT point of (9).

4. We prove in Theorem 8 that for almost all LROTA problems, iAPD reduces to APD after finitely many iterations if it converges to a local minimizer. In particular, this relaxes the requirement for each iterative matrix being of full rank in the literature, such as [13, 26], to a simple requirement on the limit point.

Other than these, we also prove in Theorem 1 that every essential KKT point of LROTA is nondegenerate for a generic tensor, which might be of independent interests.

The rest of this paper is organized as follows: basic notions and facts on tensors that will be encountered repeatedly in the sequel are included in Sect. 2. In particular, the LROTA problem is stated in Sect. 2.2. Section 3 is devoted to the analysis of the manifold structures of the set of orthogonally decomposable tensors. In this section, the connection between KKT points of LROTA and critical points of the projection function on manifolds is established. It is shown that every essential KKT point of LROTA for a generic tensor is nondegenerate. The new algorithm iAPD is presented in Sect. 4 and detailed convergence analysis for this algorithm is given. The overall sublinear convergence and generic linear convergence are also proved. Section 5 proves that for almost all LROTA problems, iAPD reduces to APD after finitely many iterations if it converges to a local minimizer. Numerical examples are presented in Sect. 6. Some final remarks are given in Sect. 7. Preliminaries on differential geometry and optimization theory, necessary for our convergence analysis, are provided in Appendix A. Technical lemmas are stated when they are needed and proofs are supplied in Appendices B and C.

### 2 Low rank orthogonal tensor approximation problem

In this section, we provide a quick review on basic notions and facts of tensors. Given positive integers \( k \geq 2 \) and \( n_1, \ldots, n_k \), the tensor space consisting of real tensors of dimension \( n_1 \times \cdots \times n_k \) is denoted by \( \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} \). In this vector space, an inner product and hence a norm can be defined. The Hilbert-Schmidt inner product of two given tensors \( A, B \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} \) is defined by

\[
\langle A, B \rangle := \sum_{i_1=1}^{n_1} \cdots \sum_{i_k=1}^{n_k} a_{i_1 \ldots i_k} b_{i_1 \ldots i_k}.
\]
The Hilbert-Schmidt norm $\|A\|$ is then given by (cf. [46])

$$\|A\| := \sqrt{\langle A, A \rangle}.$$ 

In particular, if $k = 2$, then an element in $\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$ is simply an $n_1 \times n_2$ matrix $A$, whose Hilbert-Schmidt norm reduces to the Frobenius norm $\|A\|_F$.

Given a positive integer $r$ and $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$, we denote by $\text{diag}(\lambda_1, \ldots, \lambda_r)$ the diagonal tensor in $\mathbb{R}^r \otimes \cdots \otimes \mathbb{R}^r$ with the order being understood from the context. To be more precise, we have

$$(\text{diag}(\lambda_1, \ldots, \lambda_r))_{i_1 \ldots i_k} = \begin{cases} 
\lambda_j, & \text{if } i_1 = \cdots = i_k = j \in \{1, \ldots, r\}, \\
0, & \text{otherwise.}
\end{cases}$$

For a given positive integer $k$, we may regard the tensor $\text{diag}(\lambda_1, \ldots, \lambda_r)$ as the image of $(\lambda_1, \ldots, \lambda_r)$ under the map $\text{diag} : \mathbb{R}^r \rightarrow \otimes^k \mathbb{R}^r$ defined in an obvious way. We also define the map $\text{Diag} : \otimes^k \mathbb{R}^r \rightarrow \mathbb{R}^r$ by taking the diagonal of a $k$-th order tensor. By definition, $\text{Diag} \circ \text{diag} : \mathbb{R}^r \rightarrow \mathbb{R}^r$ is the identity map.

We define a map $\tau : \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k} \rightarrow \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k}$ by

$$\tau(x) := x_1 \otimes \cdots \otimes x_k,$$  

where $x$ is a block vector

$$x := (x_1, \ldots, x_k) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k} \simeq \mathbb{R}^{n_1 + \cdots + n_k}$$

with $x_i \in \mathbb{R}^{n_i}$ for all $i = 1, \ldots, k$.

More precisely, if we write $x_i = (x_{i1}, \ldots, x_{i n_i})^T$ for each $i = 1, \ldots, k$, then we simply have

$$(\tau(x))_{j_1 \ldots j_k} = x_{1 j_1} \cdots x_{k j_k}, \quad 1 \leq j_i \leq n_i, \ i = 1, \ldots, k.$$ 

For each $i \in \{1, \ldots, k\}$, we define a map $\tau_i : \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k} \rightarrow \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_{i-1}} \otimes \mathbb{R}^{n_{i+1}} \otimes \cdots \otimes \mathbb{R}^{n_k}$ by

$$\tau_i(x) := x_1 \otimes \cdots \otimes x_{i-1} \otimes x_{i+1} \otimes \cdots \otimes x_k, \quad x \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}.$$ 

Given a tensor $\mathcal{A} \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k}$ and a block vector $x$ as above, $\mathcal{A}\tau(x)$ is defined by

$$\mathcal{A}\tau(x) := \langle \mathcal{A}, \tau(x) \rangle$$

and $\mathcal{A}\tau_i(x) \in \mathbb{R}^{n_i}$ is a vector defined implicitly by the relation:

$$\langle \mathcal{A}\tau_i(x), x_i \rangle = \mathcal{A}\tau(x)$$

@ Springer
for any block vector \( x \). Moreover, given \( k \) matrices \( B^{(i)} \in \mathbb{R}^{m_i \times n_i} \) for \( i \in \{1, \ldots, k\} \), the matrix-tensor product \( (B^{(1)}, \ldots, B^{(k)}) \cdot A \) is a tensor in \( \mathbb{R}^{m_1 \otimes \cdots \otimes \mathbb{R}^{m_k}} \), defined entry-wisely as

\[
[(B^{(1)}, \ldots, B^{(k)}) \cdot A]_{i_1 \ldots i_k} := \sum_{j_1=1}^{n_1} \cdots \sum_{j_k=1}^{n_k} b^{(1)}_{i_1j_1} \cdots b^{(k)}_{i_kj_k} a_{j_1 \ldots j_k} \quad (2)
\]

for all \( i_t \in \{1, \ldots, m_t\} \) and \( t \in \{1, \ldots, k\} \).

### 2.1 Orthogonally decomposable tensor

In this subsection, we discuss a special type of tensors defined by orthogonality. Given two positive integers \( 1 \leq m \leq n \), we denote by \( V(m, n) \) the set of \( n \times m \) matrices \( U \) such that \( U^T U = I \), where \( I \) is the identity matrix of matching size. In the literature, \( V(m, n) \) is called the Stiefel manifold of orthonormal \( n \times m \) matrices. We refer readers to Appendix A.1 for basic properties and constructions concerning \( V(m, n) \). A tensor \( A \in \mathbb{R}^{n_1 \otimes \cdots \otimes \mathbb{R}^{n_k}} \) is called orthogonally decomposable (cf. [23, 34, 35, 66]) if there exist orthonormal matrices

\[
U^{(i)} = \begin{bmatrix} u_{1}^{(i)} & \cdots & u_{r}^{(i)} \end{bmatrix} \in V(r, n_i), \ i = 1, \ldots, k
\]

and numbers \( \lambda_j \in \mathbb{R} \) for \( 1 \leq j \leq r \leq \min\{n_1, \ldots, n_k\} \) such that

\[
A = \sum_{j=1}^{r} \lambda_j u_j^{(1)} \otimes \cdots \otimes u_j^{(k)}. \quad (3)
\]

Here for each \( i = 1, \ldots, k \) and \( j = 1, \ldots, r \), the vector \( u_j^{(i)} \in \mathbb{R}^{n_i} \) is the \( j \)-th column vector of the orthonormal matrix \( U^{(i)} \). Without loss of generality, we may assume that \( \lambda_j \geq 0 \) for all \( j = 1, \ldots, r \). Throughout this paper, we always assume that \( k \geq 3 \). A decomposition of the form (3) is called an orthogonal decomposition of \( A \). We call an orthogonal decomposition of the form (3) with the smallest \( r \) an orthogonal rank decomposition of \( A \). We notice that a general tensor may not admit a decomposition of the form (3), thus the notion of orthogonal rank decomposition only makes sense for orthogonally decomposable tensors.

Let \( r, k \) be positive integers and let \( n := (n_1, \ldots, n_k) \) be a \( k \)-dimensional vector of integers. We denote by \( C(n, r) \subseteq \mathbb{R}^{n_1 \otimes \cdots \otimes \mathbb{R}^{n_k}} \) the set of all orthogonally decomposable tensors of rank at most \( r \), i.e.,
\[ C(n, r) := \left\{ A \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} : A = (U^{(1)}, \ldots, U^{(k)}) \cdot \text{diag}(\lambda_1, \ldots, \lambda_r), \right. \]
\[ \left. U^{(i)} \in V(r, n_i) \text{ for all } i \in \{1, \ldots, k\}, \lambda_j \in \mathbb{R} \text{ for all } j \in \{1, \ldots, r\} \right\}. \]  

(4)

We also let \( D(n, r) \subseteq \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} \) be the set of all orthogonally decomposable tensors of rank \( r \), i.e.,

\[ D(n, r) := \left\{ A \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} : A = (U^{(1)}, \ldots, U^{(k)}) \cdot \text{diag}(\lambda_1, \ldots, \lambda_r), \right. \]
\[ \left. U^{(i)} \in V(r, n_i) \text{ for all } i \in \{1, \ldots, k\}, \lambda_j \neq 0 \text{ for all } j \in \{1, \ldots, r\} \right\}. \]  

(5)

For a general tensor \( T \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} \), the rank of \( T \) is the smallest integer \( r \) such that

\[ T = \sum_{j=1}^{r} \lambda_j u^{(1)}_j \otimes \cdots \otimes u^{(k)}_j \]  

(6)

for some \( \lambda_j \in \mathbb{R} \) and \( u^{(i)}_j \in \mathbb{R}^{n_i} \) with \( \|u^{(i)}_j\| = 1 \), where \( i = 1, \ldots, k \) and \( j = 1, \ldots, r \). We notice that unlike (3), components \( u^{(i)}_1, \ldots, u^{(i)}_r \) in (6) are not required to be orthogonal to each other. The rank of \( T \) is denoted by \( \text{rank}(T) \) and a rank decomposition of \( T \) is a decomposition of the form (6) with \( r = \text{rank}(T) \). We say that the rank decomposition of \( T \) is unique if

\[ T = \sum_{j=1}^{\text{rank}(T)} \lambda_j u^{(1)}_j \otimes \cdots \otimes u^{(k)}_j = \sum_{j=1}^{\text{rank}(T)} \mu_j v^{(1)}_j \otimes \cdots \otimes v^{(k)}_j \]

implies that there exist some permutation \( \sigma \) on the set \( \{1, \ldots, \text{rank}(T)\} \) and nonzero numbers \( \alpha^{(i)}_j \) such that

\[ v^{(i)}_{\sigma(j)} = \alpha^{(i)}_j u^{(i)}_j, \quad \left( \prod_{s=1}^{k} \alpha^{(s)}_j \right) \mu_{\sigma(j)} = \lambda_j \]

for each \( i = 1, \ldots, k \) and \( j = 1, \ldots, \text{rank}(T) \).

In particular, an orthogonally decomposable tensor \( A \) a priori admits both orthogonal rank decomposition and rank decomposition. Thus it is natural to compare the two notions of decompositions for orthogonally decomposable tensors. Fortunately, the two notions coincide, according to the next lemma.

**Lemma 1** (Unique Decomposition) For each \( A \in D(n, r) \), the rank of \( A \) is \( r \) and the rank decomposition of \( A \) is unique. In particular, the orthogonal rank decomposition of \( A \) is unique.

**Proof** It follows from Kruskal’s inequality [36, 37, 39] immediately. A direct proof can also be found in [66].
Lemma 1 implies that for any \( A \in D(n, r) \), we have \( \text{rank}(A) = r \) and its rank decomposition coincides with its orthogonal rank decomposition, which is unique. Therefore it is not necessary to distinguish rank decomposition and orthogonal rank decomposition for a tensor in \( D(n, r) \). As a direct corollary of Lemma 1, we have the following characterization of \( D(n, r) \).

**Corollary 1** The set \( D(n, r) \) consists of all \( A \in C(n, r) \) whose rank is \( r \).

**Proof** The fact that each \( A \in D(n, r) \subset C(n, r) \) has rank \( r \) follows from Lemma 1. Conversely, if \( A \in C(n, r) \), then \( A \) admits an orthogonal decomposition (cf. (4))

\[
A = (U^{(1)}, \ldots, U^{(k)}) \cdot \text{diag}(\lambda_1, \ldots, \lambda_r)
\]

for some \( U^{(i)} \in V(r, n_i) \) and \( \lambda_j \in \mathbb{R}, i = 1, \ldots, k, j = 1, \ldots, r \). If furthermore \( \text{rank}(A) = r \), then all \( \lambda_j \)'s in (7) must be nonzero by the definition of rank, i.e., \( A \in D(n, r) \).

\( \square \)

### 2.2 Low rank orthogonal tensor approximation

The problem considered in this paper can be described as follows: given a tensor \( A \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} \), find an orthogonally decomposable tensor \( B \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} \) of rank at most \( r \leq \min\{n_1, \ldots, n_k\} \) such that the residual \( \|A - B\| \) is minimized. More precisely, we consider the following optimization problem:

\[
\text{(LROTA(r))} \quad \min \|A - (U^{(1)}, \ldots, U^{(k)}) \cdot Y\|^2 \quad \text{s.t.} \quad Y = \text{diag}(v_1, \ldots, v_r), \quad v_j \in \mathbb{R}, \quad (U^{(i)})^T U^{(i)} = I \text{ for all } 1 \leq i \leq k.
\]

**Proposition 1** (Maximization Equivalence) The optimization problem (8) is equivalent to

\[
\text{(mLROTA(r))} \quad \max \sum_{j=1}^r \left( \left( (U^{(1)})^T, \ldots, (U^{(k)})^T \right) \cdot A \right)^2_{j \ldots j} \quad \text{s.t.} \quad (U^{(i)})^T U^{(i)} = I \text{ for all } 1 \leq i \leq k
\]

in the following sense

1. if \( (U_*, Y_*) := ((U^{(1)}_*), \ldots, U^{(k)}_*), \text{diag}((v_1)_*, \ldots, (v_r)_*) \) is an optimizer of (8) with the optimal value \( \|A\|^2 - \sum_{j=1}^r (v_j)_*^2 \), then \( U_* \) is an optimizer of (9) with the optimal value \( \sum_{j=1}^r (v_j)_*^2 \);
2. conversely, if \( U_* \) is an optimizer of (9), then \( (U_*, Y_*) \) is an optimizer of (8) where \( Y_* = \text{diag} \left( \text{Diag} \left( ((U^{(1)}_*)^T, \ldots, (U^{(k)}_*)^T) \cdot A \right) \right) \).
Proof. By a direct calculation we may obtain (cf. (2))

$$
\|A - (U^{(1)}, \ldots, U^{(k)}) \cdot \Upsilon \|^2 = \|A\|^2 + \sum_{j=1}^{r} v_j^2 - 2 \langle A, (U^{(1)}, \ldots, U^{(k)}) \cdot \Upsilon \rangle \\
= \|A\|^2 + \sum_{j=1}^{r} v_j^2 - 2 \langle ((U^{(1)})^\top, \ldots, (U^{(k)})^\top) \cdot A, \Upsilon \rangle \\
= \|A\|^2 + r \sum_{j=1}^{r} v_j^2 - 2 \sum_{j=1}^{r} v_j \left[ ((U^{(1)})^\top, \ldots, (U^{(k)})^\top) \cdot A \right]_{j\ldots j}.
$$

Note that $v_j$ in the minimization problem (8) is unconstrained for all $j \in \{1, \ldots, r\}$, and they are mutually independent. Thus, at an optimizer $(U_*, \Upsilon_*):=(U_*^{(1)}, \ldots, U_*^{(k)}), \Upsilon_*$) of (8), we must have

$$(v_*)_j = \left[ ((U_*^{(1)})^\top, \ldots, (U_*^{(k)})^\top) \cdot A \right]_{j\ldots j} \text{ for all } 1 \leq j \leq r \quad (10)$$

and the optimal value is

$$
\|A\|^2 - \sum_{j=1}^{r} (v_*)_j^2.
$$

Therefore, problem (8) is equivalent to (9). □

3 KKT points via projections onto $C(n, r)$

On the one hand, a numerical algorithm solving the optimization problem (8) (or equivalently its maximization reformulation (9)) is usually designed in the parameter space $V_{n,r}:=V(r, n_1) \times \cdots \times V(r, n_k) \times \mathbb{R}^r$. See for example, [13, 26, 64, 65]. On the other hand, from a more geometric perspective, we can also regard problem (8) as the projection of a given tensor $A$ onto $C(n, r)$. A key ingredient in our study of problem (8) is the relation between these two viewpoints. Once such a connection is understood, we are able to derive an algorithm in $V_{n,r}$ but analyse it in $C(n, r)$. To be more precise, we study both the problem of the projection

$$
\min \|A - B\|^2 \quad \text{s.t. } B \in D(n, r), \quad (11)
$$

and its parametrization

$$
(LROTA-P) \quad \min g(U, x):=\frac{1}{2} \|A - (U^{(1)}, \ldots, U^{(k)}) \cdot \text{diag}(x)\|^2 \quad \text{s.t. } (U^{(i)})^\top U^{(i)} = I \text{ for all } 1 \leq i \leq k, \quad x \in \mathbb{R}^r^*, \quad (12)
$$
where $\mathbb{R}_*: = \mathbb{R} \setminus \{0\}$ and $\mathbb{U}: = (U^{(1)}, \ldots, U^{(k)})$. We remark that LROTA-P described in (12) has the same (up to a constant factor $1/2$) objective function as LROTA(r) described in (8), but the feasible domain of LROTA-P is $V(r, n_1) \times \cdots \times V(r, n_k) \times \mathbb{R}^*_{r}$, which is a proper subset of the feasible domain $V(r, n_1) \times \cdots \times V(r, n_k) \times \mathbb{R}^r$ of LROTA(r). Geometrically, LROTA-P corresponds to the projection of a tensor onto $D(n, r)$ while LROTA(r) corresponds to $C(n, r)$.

We first study properties of $C(n, r)$ and $D(n, r)$ and then discuss critical points of problem (11) in Sect. 3.1. In particular, we prove that $D(n, r)$ is a smooth manifold while $C(n, r)$ is an algebraic variety with singularities. KKT points of (9) and hence those of (8) are discussed in Sect. 3.2. The connection between them is studied in Sect. 3.3, in which a Łojasiewicz inequality for KKT points of (9) is given. We refer to [20, 28, 51, 57] for basic facts of differential geometry, algebraic geometry and algebraic topology that are used in the sequel.

### 3.1 Geometry of orthogonally decomposable tensors

Let

$$U_{n, r} := V(r, n_1) \times \cdots \times V(r, n_k) \times \mathbb{R}^*_{r}, \quad U_{n, r}^+ := V(r, n_1) \times \cdots \times V(r, n_k) \times \mathbb{R}^r,$$

(13)

where $\mathbb{R}^r_+$ denotes the set of positive real numbers. It is clear that $U_{n, r}$ has $2^r$ components partitioned by $\mathbb{R}^*_{r}$ and $U_{n, r}^+$ is one of them. Moreover, any other component of $U_{n, r}$ is diffeomorphic to $U_{n, r}^+$. $U_{n, r}^+$ can be further divided into finitely many connected components, depending on whether $r = n_i$ for some $i$ or not, and these connected components are again diffeomorphic.

Let

$$\mathcal{D}_r := \{ \text{diag}(d) : d \in \{\pm 1\}^r \}$$

be the group of diagonal matrices with $\pm 1$ diagonal elements and $\mathcal{S}_r$ be the permutation group on $r$ elements. Let

$$\mathcal{D}_{r, k} := \{ E := (E_1, \ldots, E_k) \in \mathcal{D}_r \times \cdots \times \mathcal{D}_r : \prod_{i=1}^k E_i = I \}$$

be the subgroup of $\mathcal{D}_r \times \cdots \times \mathcal{D}_r$ such that the product of the component matrices is the identity.

By the next proposition, $U_{n, r}^+$ parameterizes the manifold $D(n, r)$.

**Proposition 2** For each positive integer $r \leq \min\{n_1, \ldots, n_k\}$, the map

$$\varphi_{n, r} : V(r, n_1) \times \cdots \times V(r, n_k) \times \mathbb{R}^r \to C(n, r),$$

$$(U^{(1)}, \ldots, U^{(k)}, (\lambda_1, \ldots, \lambda_r)) \mapsto (U^{(1)}, \ldots, U^{(k)}) \cdot \text{diag}(\lambda_1, \ldots, \lambda_r)$$

is a surjective map and we have the following:

 Springer
variety is a closed subset of some $\mathbb{R}^3$. This is due to the fact that $\dim_2$ an algebraic variety here means a set of common roots of some polynomials. In other words, an algebraic $\sigma$ action by permuting columns. In other words, an element $(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(r)})$. Now it is straightforward to verify that $\varphi_{n,r}$ is $\mathcal{O}_{r,k} \times \mathcal{S}_r$-invariant.

By (5), it is obvious that $\varphi_{n,r}(U_{n,r}) \subseteq D(n, r)$ and hence $U_{n,r} \subseteq \varphi_{n,r}^{-1}(D(n, r))$. For the reverse inclusion, we notice that if $(U^{(1)}, \ldots, U^{(k)}, (\lambda_1, \ldots, \lambda_r)) \in \varphi_{n,r}^{-1}(D(n, r))$, then we have

$$(U^{(1)}, \ldots, U^{(k)}) \cdot \text{diag}(\lambda_1, \ldots, \lambda_r) \in D(n, r).$$

By Corollary 1, we may conclude that all $\lambda_j$’s are nonzero and this implies that

$$(U^{(1)}, \ldots, U^{(k)}, (\lambda_1, \ldots, \lambda_r)) \in U_{n,r}.$$
It is clear that $U_{n,r}$ is an open subset and it is also an open submanifold of $V_{n,r}$. Since $U_{n,r}$ is a union of connected components of $U_{n,r}$, we conclude that $U_{n,r}$ is also an open submanifold of $V_{n,r}$.

We notice that $U_{n,r}^+$ admits an $\mathcal{D}_{r,k} \times \mathcal{G}_r$-action by the restriction of that on $V_{n,r}$ and the fiber $U_{n,r}^+ \cap \varphi_{n,r}^{-1}(T) \simeq \mathcal{D}_{r,k} \times \mathcal{G}_r$ if $T \in D(n, r)$ by Lemma 1. This implies that $U_{n,r}^+ / (\mathcal{D}_{r,k} \times \mathcal{G}_r) \simeq D(n, r)$.

Since $\mathcal{D}_{r,k} \times \mathcal{G}_r$ is a finite group acting on $U_{n,r}$ freely and properly discontinuously, we conclude that $D(n, r) \simeq U_{n,r}^+ / (\mathcal{D}_{r,k} \times \mathcal{G}_r)$ is a smooth manifold whose dimension is

$$\dim D(n, r) = \dim U_{n,r}^+ = \dim V_{n,r} = \sum_{j=1}^k \dim V(r, n_j) + \dim \mathbb{R}^r.$$ 

Observing that $\dim V(r, n) = r(n - r) + \binom{r}{2}$, we obtain the desired formula for $\dim D(n, r)$.

The fact that $C(n, r)$ is an algebraic variety follows directly from [12, Theorem 4], which states that $\bigcup_{s=0}^\min\{n_i\} D(n, s)$ is an algebraic variety. Indeed, if we denote by $M_r$ the set of tensors $A$ in $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k}$ such that the mode-1 unfolding of $A$ has rank at most $r$, then we have that $C(n, r) = M_r \cap \bigcup_{s=0}^\min\{n_i\} D(n, s)$. Since $M_r$ is an algebraic variety, we may conclude that $C(n, r)$ is an algebraic variety.

Since $C(n, r)$ is the image of the irreducible algebraic variety $V(n, r)$ under the map $\varphi_{n,r}$, we may conclude that $C(n, r)$ is irreducible.

Let $r < \max\{n_i : i = 1, \ldots, k\}$. To check that $\bigcup_{t=0}^{r-1} D(n, t)$ is the singular locus of $C(n, r)$, we calculate the dimension of $T_{C(n,r)}(B)$ for some $B \in \bigcup_{t=0}^{r-1} D(n, t)$. Without loss of generality, we may suppose that $B \in D(n, r-1)$. If $r = 1$ then $B = 0$ and it is the vertex and hence is the singularity of the cone $C(n, 1) = D(n, 1) \bigcup \{0\} \subseteq \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k}$. Next we assume that $r \geq 2$. A curve $B(t)$ in $C(n, r)$ such that $B(0) = B$ can be written as

$$B(t) = \left(U^{(1)}(t), \ldots, U^{(k)}(t)\right) \cdot \mathrm{diag}(\lambda_1(t), \ldots, \lambda_r(t)), \quad t \in (-\epsilon, \epsilon)$$

for some small $\epsilon > 0$. We may also assume that $\lambda_1(0) \geq \cdots \geq \lambda_r(0) \geq 0$. Here we must have $\lambda_r(0) = 0$.

By Lemma 1 and the finiteness of the acting group, it follows that there are finitely many points in $V_{n,r} \cap \varphi_{n,r}^{-1}(B(t))$ for each $t \neq 0$. Let us consider the differential $d(\varphi_{n,r})$ of $\varphi_{n,r}$ at the preimage $(U^{(1)}, \ldots, U^{(k)})$, $\lambda := ((U^{(1)}(0), \ldots, U^{(k)}(0), \lambda(0))$ of $B$, where $\lambda := \lambda(0) := (\lambda_1(0), \ldots, \lambda_r(0))^T$. In coordinates, the differential is represented by block matrices of the form

$$\begin{bmatrix}
\lambda_i \mathrm{id} \otimes u_i^{(2)} \otimes \cdots \otimes u_i^{(k)} & \ldots & \lambda_i u_i^{(1)} \otimes \cdots \otimes u_i^{(k-1)} \otimes \mathrm{id} u_i^{(1)} \otimes \cdots \otimes u_i^{(k)}
\end{bmatrix},$$

$i = 1, \ldots, r$.
where id is the corresponding identity mapping. Note that \( \lambda_r = 0 \), so the \( r \)-th block matrix has \( k \) zero blocks. It is easy to see that

\[
T_{C(n,r)}(B) = d(\phi_{n,r})(((U^{(1)}, \ldots, U^{(k)}), \lambda)(T_{V(r,n_1)}(U^{(1)}) \times \cdots \times T_{V(r,n_k)}(U^{(k)}) \\
\times T_{\mathbb{R}^r}(\lambda)).
\]

By the assumption, there is some \( n_i > r \). It follows that the tangent space at \( A \in V(r,n_i) \) of the Stiefel manifold \( V(r,n_i) \) is \( A \), skew \( (\mathbb{R}^{r\times r}) + A^\perp \mathbb{R}^{(n_i-r)\times r} \), where \( A^\perp \mathbb{R}^{n_i\times (n_i-r)} \) is any matrix formed by orthonormal columns that are complement to those of \( A \). Then, it follows that the dimension of the tangent space \( T_{C(n,r)}(B) \) would be strictly less than \( r + \sum_{i=1}^k \dim(T_{V(r,n_i)}(U^{(i)})) = d_{n,r} \), since some bases determined by the part \( A^\perp \mathbb{R}^{(n_i-r)\times r} \) are mapped to zero by the differential \( d(\phi_{n,r}) \) at \( ((U^{(1)}, \ldots, U^{(k)}), \lambda) \). Hence, \( B \) cannot be a smooth point. Therefore we may conclude that \( \bigcup_{i=1}^{n-2} D(n,i) \) is the singular locus of \( C(n,r) \).

If \( n_1 = \cdots = n_k = n \), a similar argument shows that the tangent space of each \( B \in D(n,n-1) \bigcup D(n,n) \) has the maximal dimension \( d_{n,n} \) and the singular locus is \( \bigcup_{i=0}^{n-2} D(n,i) \).

We show in the next lemma that \( U_{n,r} \) is locally diffeomorphic to \( D(n,r) \).

Lemma 2 (Local Diffeomorphism) For any positive integers \( n_1, \ldots, n_k \) and \( r \leq \min(n_1, \ldots, n_k) \), the set \( U_{n,r} \) is a smooth manifold and is locally diffeomorphic to the manifold \( D(n,r) \).

Proof We recall from Proposition 2 that \( U_{n,r}^+ \) is a principle \( O_{r,k} \times \mathbb{G}_r \)-bundle on \( D(n,r) \) and all connected components of \( U_{n,r} \) are diffeomorphic to any connected component of \( U_{n,r}^+ \). In particular, since \( O_{r,k} \times \mathbb{G}_r \) is a finite group, for any \( \mathcal{T} \in D(n,r) \), the fiber \( \phi_{n,r}(\mathcal{T}) \) of the map

\[
\phi_{n,r} : U_{n,r} \rightarrow D(n,r)
\]

consists of \( s := 2^{rk} r! \) points. Therefore, for a small enough neighbourhood \( W \subseteq D(n,r) \) of \( \mathcal{T} \), the inverse image \( \phi_{n,r}^{-1}(W) \) is the disjoint union of \( s \) open subsets \( W_1, \ldots, W_s \subseteq U_{n,r} \) and for each \( j = 1, \ldots, s \), we have that

\[
(\phi_{n,r})|_{W_j} : W_j \rightarrow W
\]

is a diffeomorphism. \( \square \)

By Lemma 2 and Proposition 11, problems on \( D(n,r) \) can be studied via problems on \( U_{n,r} \). To that end, the tangent space of \( U_{n,r} \) is given at first. The following result can be checked directly, see [1, 22].

Proposition 3 (Tangent Space of \( U_{n,r} \)) At any point \( (\mathbb{U}, x) \in U_{n,r} \), the tangent space of \( U_{n,r} \) at \( (\mathbb{U}, x) \) is

\[
T_{U_{n,r}}(\mathbb{U}, x) = T_{V(r,n_1)}(U^{(1)}) \times \cdots \times T_{V(r,n_k)}(U^{(k)}) \times \mathbb{R}^r,
\]

(14)
where $T_{V(r,n_i)}(U^{(i)})$ is the tangent space of the Stiefel manifold $V(r,n_i)$ at $U^{(i)}$, which is

$$T_{V(r,n_i)}(U^{(i)}) = \{Z \in \mathbb{R}^{n_i \times r} : (U^{(i)})^T Z + Z^T U^{(i)} = 0\},$$

for all $i = 1, \ldots, k$.

We can embed $U_{n,r}$ into $\mathbb{R}^{n_1 \times r} \times \cdots \times \mathbb{R}^{n_k \times r} \times \mathbb{R}^r$ in an obvious way and hence $U_{n,r}$ becomes an embedded submanifold of the latter. For a differentiable function $f : U_{n,r} \subseteq \mathbb{R}^{n_1 \times r} \times \cdots \times \mathbb{R}^{n_k \times r} \times \mathbb{R}^r \rightarrow \mathbb{R}$, a critical point $(U, x)$ is a point at which the Riemannian gradient $\text{grad}(f)(U, x)$ of $f$ at $(U, x)$ is zero, which is equivalent to the fact that the projection of the Euclidean gradient $\nabla f(U, x)$ onto the tangent space of $U_{n,r}$ at $(U, x)$ is zero. More explicitly, we have the following characterization.

**Lemma 3** Let $A \in V(r,n)$ and let $f : V(r,n) \subseteq \mathbb{R}^{n \times r} \rightarrow \mathbb{R}$ be a smooth function. Then $\text{grad}(f)(A) = 0$ if and only if

$$\nabla f(A) = A(\nabla f(A))^T A,$$

which is also equivalent to $\nabla f(A) = AP$ for some symmetric matrix $P \in S^{r \times r}$. In particular, $A^T \nabla f(A)$ is a symmetric matrix.

A proof of the first equivalence can be found in [47, Proposition 1]. For completeness, we provide a proof using the projection formula (67).

**Proof** By (67) we have

$$\text{grad}(f)(A) = (I - \frac{1}{2} AA^T)(\nabla f(A)) - A \nabla f(A)^T A),$$

from which we may conclude that $\text{grad}(f)(A) = 0$ if and only if $\nabla f(A) = A \nabla f(A)^T A$, since $I - \frac{1}{2} AA^T$ is of full rank.

For the second, we notice that from (16)

$$A^T \nabla f(A) = A^T A(\nabla f(A))^T A = \nabla f(A)^T A$$

and this proves that $A^T \nabla f(A)$ is symmetric. Now if we set $P := A^T \nabla f(A)$ then (16) can be written as

$$\nabla f(A) = AP^T = AP.$$

Conversely, if $\nabla f(A) = AP$ for some symmetric matrix $P$, then (16) obviously holds by the symmetry of $P$ and the fact that $A^T A = I$. \qed

We recall that the objective function of (12) is

$$g(U, x) = \frac{1}{2} \|A - (U^{(1)}, \ldots, U^{(k)}) \cdot \text{diag}(x)\|^2.$$

Since the feasible set $D(n, r)$ (resp. $U_{n,r}$) of (11) (resp. (12)) is a smooth manifold, we may apply Proposition 11, Lemmas 11 and 2 to obtain the following
Proposition 4  For a generic tensor $A$, each critical point of the function $g$ on $U_{n,r}$ is nondegenerate.

We conclude this subsection by a remark on how we use the results obtained so far to solve and analyse (8). Our elaboration consists of two parts:

- On the one hand, (8) is an optimization problem with constraints so characterizing its KKT points is a necessary step to solve it. On the other hand, from a geometric point of view, one can interpret (8) as a projection of a tensor $A$ onto $C(n, r)$ and hence critical points of the squared distance function from $A$ to $C(n, r)$ should be studied carefully. It turns out that the two perspectives have their own advantages. To be more precise, it turns out that KKT points of (8) can be characterized by simple equations, based on which an efficient algorithm can be designed. To analyse the convergent behaviour of the algorithm, the geometric point of view is more convenient because powerful tools can be borrowed from differential/algebraic geometry. However, since $C(n, r)$ is not smooth, differential geometric tools cannot be applied directly. To remedy this, it is unavoidable to investigate $C(n, r)$ more closely. Actually, it is the simple stratified structure $C(n, r) = \bigsqcup_{t=0}^{r} D(n, t)$ which is essential in our convergent analysis.

- Although it is much more efficient and convenient to work in the parameter space $V_{n,r}$ (resp. $U_{n,r}$) if we want to implement our algorithm, the analysis of the algorithm requires us to work in the tensor space $C(n, r)$ (resp. $D(n, r)$) itself. This discrepancy is a natural impetus for the study of the relation between $V_{n,r}$ (resp. $U_{n,r}$) and $C(n, r)$ (resp. $D(n, r)$). Eventually this connection enables us to translate properties of critical points into properties of KKT points.

3.2 KKT points of LROTA

In this subsection, we derive the KKT system of the optimization problem (9) and study its properties.

3.2.1 Existence

Since (9) is a nonlinear optimization problem with equality constraints of continuously differentiable functions, the classical Karush-Kuhn-Tucker condition (KKT condition) for (9) can be determined. Let us consider a general optimization problem

$$\max f(x)$$

s.t. $g_i(x) = 0$, $i = 1, \ldots, p,$

(18)

where $f, g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ are continuously differentiable functions. A feasible point $x$ of (18) is a Karush-Kuhn-Tucker point (KKT point) if the following KKT condition is satisfied (cf. [7])

$$\nabla f(x) + \sum_{i=1}^{p} u_i \nabla g_i(x) = 0$$

(19)
for some Lagrange multiplier vector \( \mathbf{u} = (u_1, \ldots, u_p)^T \in \mathbb{R}^p \). A fundamental result in optimization theory is that under some constraint qualifications, an optimizer of (18) must be a KKT point [7]. This general theory applies to (9) and the KKT condition can be explicitly obtained, which is incorporated in the following definition for easy reference. To do so, we need the following notations.

Let \( \mathbb{U} = (U^{(1)}, \ldots, U^{(k)}) \) be the collection of the variable matrices in (9) and for each \( 1 \leq i \leq k \) and \( 1 \leq j \leq r \), let \( \mathbf{u}_j^{(i)} \) be the \( j \)-th column of the matrix \( U^{(i)} \) and let

\[
\mathbf{x}_j := (\mathbf{u}_j^{(1)}, \ldots, \mathbf{u}_j^{(k)}) \text{ and } \mathbf{v}_j^{(i)} := A\tau_i(\mathbf{x}_j).
\]

For each \( 1 \leq i \leq k \), we define a matrix

\[
V^{(i)} := \begin{bmatrix} \mathbf{v}_1^{(i)} & \cdots & \mathbf{v}_r^{(i)} \end{bmatrix},
\]

and a diagonal matrix

\[
\Lambda := \text{diag}(A\tau(\mathbf{x}_1), \ldots, A\tau(\mathbf{x}_r)).
\]

For each \( 1 \leq j \leq r \), we also set (cf. (1))

\[
\lambda_j(\mathbb{U}) := \left( (U^{(1)})^T, \ldots, (U^{(k)})^T \right) \cdot A = A\tau(\mathbf{x}_j) = \langle A, \mathbf{u}_j^{(1)} \otimes \cdots \otimes \mathbf{u}_j^{(k)} \rangle.
\]

Now the objective function of (9) can be re-written as

\[
f(\mathbb{U}) := \sum_{j=1}^r \left( ((U^{(1)})^T, \ldots, (U^{(k)})^T) \cdot A \right)^2_{j \cdots j} = \sum_{j=1}^r \lambda_j(\mathbb{U})^2.
\]

**Definition 1 (KKT Point)** Let \( \mathbb{U} = (U^{(1)}, \ldots, U^{(k)}) \) be a feasible point of (9). If there exists \( \mathbb{P} = (P_1, \ldots, P_k) \) where \( P_i \in S^{r \times r} \) for each \( 1 \leq i \leq k \) such that the system

\[
V^{(i)} \Lambda - U^{(i)} P_i = 0, \quad 1 \leq i \leq k,
\]

is satisfied, then \( \mathbb{U} \) is called a KKT point and \( \mathbb{P} \) is called a Lagrange multiplier associated to \( \mathbb{U} \). The set of all multipliers associated to \( \mathbb{U} \) is denoted by \( M(\mathbb{U}) \).

It follows immediately from the system (24) that for all \( 1 \leq i \leq k \),

\[
(U^{(i)})^T V^{(i)} A = (V^{(i)} A)^T U^{(i)}.
\]

For an equality constrained optimization problem, we say that a feasible point satisfies the **linear independence constraint qualification** (LICQ) if at this point all the gradients of the constraints are linearly independent.

**Proposition 5 (LICQ)** At any feasible point of the problem (9), LICQ is satisfied. Thus, at any local maximizer of (9), the system of KKT condition holds and \( M(\mathbb{U}) \) is a singleton.

 Springer
Suppose on the contrary that LICQ is not satisfied at a feasible point \( \mathbf{U} = (\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(k)}) \) of (9). Let \( P_i \in \mathbb{S}^r \times r \) for \( i = 1, \ldots, k \) be the corresponding multipliers for the equality constraints in (9) such that they are not all zero. To be more precise, \( P_i \)'s are defined by

\[
\nabla_{\mathbf{U}} \left( \sum_{i=1}^{k} \left( (\mathbf{U}^{(i)})^T \mathbf{U}^{(i)} - I, P_i \right) \right) = 0.
\]

Aligning along the natural block partition as \( \mathbf{U} \), we must have

\[
\langle \mathbf{U}^{(i)} P_i, M^{(i)} \rangle = 0, \quad \text{for all } M^{(i)} \in \mathbb{R}^{n_i \times r}, \ 1 \leq i \leq k.
\]

Now from (26) we obtain \( \mathbf{U}^{(i)} P_i = 0 \) and hence \( P_i = 0 \) by the orthonormality of \( \mathbf{U}^{(i)} \) for all \( 1 \leq i \leq k \), and this contradicts the assumption that not all \( P_i \)'s are zero. Therefore, LICQ is satisfied, which implies the uniqueness of the multiplier. The rest conclusion follows from the classical theory of KKT condition [7].

**Lemma 4** A feasible point \((\mathbf{U}, \mathbf{Y})\) is a KKT point of problem (8) with multiplier \( \mathbb{P} = (P^{(1)}, \ldots, P^{(k)}) \) if and only if \( \mathbf{U} \) is a KKT point of problem (9) with multiplier \( \mathbb{P} \) and \( \mathbf{Y} = \text{diag} \left( \text{Diag} \left( ((\mathbf{U}^{(1)})^T, \ldots, (\mathbf{U}^{(k)})^T) \cdot A \right) \right) \).

**Proof** Noticing that variables \( v_j \)'s in (8) are unconstrained and also the objective function (8) is quadratic with leading coefficient 1 for each \( v_j \), we may conclude that for fixed \( \mathbf{U}^{(i)} \)'s, the objective function of (8) has a unique critical point for \( v_j \)'s. Thus, when applying the general theory of calculating the KKT points of (8), we see that \( \mathbf{Y} \) is uniquely determined by \( \mathbf{U} \). The desired correspondence between KKT points of (8) and (9) then follows.

### 3.2.2 Primitive KKT points and essential KKT points

It is possible that for some \( j \in \{1, \ldots, r\} \), \( v_j \) approaches to zero along iterations of an algorithm solving the problem (8). In this case, the resulting orthogonally decomposable tensor is of rank strictly smaller than \( r \). We discuss this degenerate case in this section.

**Proposition 6** (KKT Reduction) Let \( \mathbf{U} = (\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(k)}) \in V(r, n_1) \times \cdots \times V(r, n_k) \) be a KKT point of problem mLROTA\((r)\) defined in (9) and let \( j \in \{1, \ldots, r\} \) be a fixed integer. Set

\[
\hat{\mathbf{U}} := (\hat{\mathbf{U}}^{(1)}, \ldots, \hat{\mathbf{U}}^{(k)}) \in V(r - 1, n_1) \times \cdots \times V(r - 1, n_k),
\]

where for each \( 1 \leq i \leq k \), \( \hat{\mathbf{U}}^{(i)} \) is the matrix obtained by deleting the \( j \)-th column of \( \mathbf{U}^{(i)} \). If \( A \tau (\mathbf{x}_j) = 0 \), then \( \hat{\mathbf{U}} \) is a KKT point of the problem mLROTA\((r-1)\):

\[
\begin{align*}
\max & \quad \| \text{Diag} \left( ((\mathbf{U}^{(1)})^T, \ldots, (\mathbf{U}^{(k)})^T) \cdot A \right) \|^2 \\
\text{s.t.} & \quad (\mathbf{U}^{(i)})^T \mathbf{U}^{(i)} = I, \quad \mathbf{U}^{(i)} \in \mathbb{R}^{n_i \times (r-1)}, \ 1 \leq i \leq k.
\end{align*}
\]
Proof By (24), the KKT system of problem (9) is

\[ V^{(i)} A = U^{(i)} P_i \text{ for all } i = 1, \ldots, k, \]

where \((P_1, \ldots, P_k) \in S^{r \times r} \times \cdots \times S^{r \times r}\) is the associated Lagrange multiplier. Without loss of generality, we may assume that \(j = r\), which implies that the last diagonal element of \(A\) is zero. Thus,

\[
(U^{(i)})^\top \begin{bmatrix} v_1^{(i)} & \ldots & v_{r-1}^{(i)} & v_r^{(i)} \end{bmatrix} \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix} = P_i, \quad 1 \leq i \leq k,
\]

where \(\hat{A}\) is the leading \((r-1) \times (r-1)\) principal submatrix of \(A\). This implies that the last column of \(P_i\) is zero. By the symmetry of \(P_i\), we conclude that \(P_i\) is in a block diagonal form with

\[
P_i = \begin{bmatrix} \hat{P}_i & 0 \\ 0 & 0 \end{bmatrix}, \quad 1 \leq i \leq k.
\]

Therefore we have

\[
\begin{bmatrix} v_1^{(i)} & \ldots & v_{r-1}^{(i)} & 0 \end{bmatrix} \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \hat{U}^{(i)} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{P}_i & 0 \\ 0 & 0 \end{bmatrix}, \quad 1 \leq i \leq k,
\]

which implies

\[
\begin{bmatrix} v_1^{(i)} & \ldots & v_{r-1}^{(i)} \end{bmatrix} \hat{A} = \hat{U}^{(i)} \hat{P}_i, \quad 1 \leq i \leq k.
\]

Consequently, we may conclude that \(\hat{U}\) is a KKT point of (27). \(\blacksquare\)

A KKT point \(U = (U^{(1)}, \ldots, U^{(k)}) \in V(r, n_1) \times \cdots \times V(r, n_k)\) of problem mLROTA(r) (cf. (9)) with \(A_\tau(x_j) \neq 0\) for all \(1 \leq j \leq r\) is called a primitive KKT point of mLROTA(r). Iteratively applying Proposition 6, we obtain the following

**Corollary 2** Let \(S\) be a subset of \(\{1, \ldots, r\}\) with cardinality \(s := |S| < r\) and let \(U = (U^{(1)}, \ldots, U^{(k)}) \in V(r, n_1) \times \cdots \times V(r, n_k)\) be a KKT point of mLROTA(r). Set

\[
\hat{U} := (\hat{U}^{(1)}, \ldots, \hat{U}^{(k)}) \in V(r-s, n_1) \times \cdots \times V(r-s, n_k),
\]

where for each \(1 \leq i \leq k\), \(\hat{U}^{(i)}\) is obtained by deleting those columns indexed by \(S\). If \(A_\tau(x_j) = 0\) exactly for \(j \in S\), then \(\hat{U}\) is a primitive KKT point of mLROTA(r-s).

It would happen that several KKT points of mLROTA(r) reduce in this way to the same primitive KKT point of mLROTA(r-s). We call the set of such KKT points an essential KKT point. KKT points with \(\lambda_j(U) = 0\) for all \(j \in \{1, \ldots, r\}\) are of less interest since they can always be avoided by algorithms like Algorithm 1, which keeps
the objective function in (9) positively bounded below. Thus, in this paper, we only discuss KKT points with at least one positive \( \lambda_j (\mathbb{U}) \). Therefore, there is a one to one correspondence between essential KKT points of mLROTA(\( r \)) and all primitive KKT points of mLROTA(\( s \)) for \( s \in \{1, \ldots, r\} \).

### 3.3 Critical points are KKT points

In this subsection, we establish the relation between KKT points of problem (9) and critical points of \( g \), which is the objective function defined in (17), on the manifold \( U_{n,r} \). To do this, we recall from (67) that the gradient of \( g \) at a point \( (\mathbb{U}, \mathbf{x}) \in U_{n,r} \) is given by

\[
\nabla_{\mathbf{U}} g (\mathbb{U}, \mathbf{x}) = -\left( I - \frac{1}{2} U^{(i)} (U^{(i)})^\top \right) \left( V^{(i)} \Gamma - U^{(i)} (V^{(i)})^\top U^{(i)} \right),
\]

\[
\nabla_{\mathbf{x}} g (\mathbb{U}, \mathbf{x}) = \mathbf{x} - \text{Diag} \left( ((U^{(1)})^\top, \ldots, (U^{(k)})^\top) \cdot \mathcal{A} \right),
\]

where \( i = 1, \ldots, k \) and \( \Gamma := \text{diag}(\mathbf{x}) \) is the diagonal matrix formed by the vector \( \mathbf{x} \).

**Proposition 7** A point \( (\mathbb{U}, \mathbf{x}) \in U_{n,r} \) is a critical point of \( g \) defined in (17) if and only if \( (\mathbb{U}, \text{diag}(\mathbf{x})) \) is a KKT point of problem (8).

**Proof** We recall that a critical point \( (\mathbb{U}, \mathbf{x}) \in U_{n,r} \) of \( g \) is defined by \( \nabla g (\mathbb{U}, \mathbf{x}) = 0 \). It follows from (28) and (29), and Proposition 3 that these critical points are defined by

\[
\nabla_{\mathbf{U}} g (\mathbb{U}, \mathbf{x}) = -V^{(i)} \Lambda + U^{(i)} \Lambda^2,
\]

where \( P_i \) is some \( r \times r \) symmetric matrix and

\[
\nabla_{\mathbf{x}} g (\mathbb{U}, \mathbf{x}) = 0.
\]

By (29), we have

\[
\mathbf{x} = \text{Diag} \left( ((U^{(1)})^\top, \ldots, (U^{(k)})^\top) \cdot \mathcal{A} \right),
\]

and according to (17), we obtain

\[
\nabla_{\mathbf{U}} g (\mathbb{U}, \mathbf{x}) = -V^{(i)} \Lambda + U^{(i)} \Lambda^2.
\]

Therefore, by (24) a critical point of \( g \) on \( U_{n,r} \) must come from a KKT point of problem (8). The converse is obvious and this completes the proof. \(\square\)

**Definition 2** (Nondegenerate KKT Point) A KKT point \( (\mathbb{U}, \mathcal{Y}) \) of problem (8) is non-degenerate if \( (\mathbb{U}, \mathbf{x}) \in U_{n,r} \) is a nondegenerate critical point of \( g \) with \( \text{diag}(\mathbf{x}) = \mathcal{Y} \).
**Theorem 1** (Finite Essential Critical Points) For a generic tensor, there are only finitely many essential KKT points for problem (9), and for any positive integers $r \geq s > 0$, a primitive KKT point of the problem mlROTA(s) corresponding to an essential KKT point of the problem mlROTA(r) is nondegenerate.

**Proof** To prove the finiteness of essential KKT points, it is sufficient to show that there are only finitely many primitive KKT points on $U_{n,r}$, and the finiteness follows from the layer structure of the set $V_{n,r}$ (or equivalently $C(n,r)$, cf. Proposition 2) and Proposition 6. We first recall that KKT points on $U_{n,r}$ are defined by (24), which is a system of polynomial equations, which implies that the set $K_{n,r}$ of KKT points of problem (9) on $U_{n,r}$ is a closed subvariety of the quasi-variety $U_{n,r}$. We also note that there are finitely many irreducible components of $K_{n,r}$ [28] and hence it suffices to prove that each irreducible component of $K_{n,r}$ is a singleton. Now let $Z \subseteq K_{n,r}$ be an irreducible component of $K_{n,r}$. If $Z$ contains infinitely many points, then dim $Z \geq 1$ [28]. However, each point in $Z$ determines a critical point of the function $g$ defined in (17) on the manifold $U_{n,r}$ (cf. Proposition 7). This implies that the set of critical points of $g$ on $U_{n,r}$ has a positive dimension, which contradicts Lemma 12 and Proposition 4.

Next, by Corollary 2, given a non-primitive KKT point $\hat{U}$ of the problem mlROTA(r), we can get a primitive KKT point $\hat{U}$ of problem mlROTA(s) with $s < r$ and hence we have $(\hat{U}, x) \in U_{n,s}$ where $x$ is determined by $\lambda_j(\hat{U})$’s. Since for a generic tensor the function $g$ has only nondegenerate critical points on $U_{n,s}$ by Proposition 4, the second assertion follows from Proposition 7 and Corollary 2. \hfill $\square$

In the following, when we talk about the nondegeneracy of a KKT point, we always refer to the nondegeneracy of the underlying unique primitive KKT point. For simplicity, we abbreviate $\nabla_U f(\hat{U})$ as $\nabla_i f(\hat{U})$ for each $1 \leq i \leq k$. We also define for each tuple $(\mathcal{W}, \mathcal{Y}) := (W^{(1)}, \ldots, W^{(k)}, \mathcal{Y}) \in \mathbb{R}^{n_1 \times r} \times \cdots \times \mathbb{R}^{n_k \times r} \times \otimes^k \mathbb{R}^r$ the norms

$$\|\mathcal{W}\|_F^2 := \sum_{i=1}^k \|W^{(i)}\|_F^2, \text{ and } \|(\mathcal{W}, \mathcal{Y})\|_F^2 := \sum_{i=1}^k \|W^{(i)}\|_F^2 + \|\mathcal{Y}\|^2.$$ 

The following result is crucial to the linear convergence analysis in the sequel.

**Lemma 5** (Łojasiewicz’s Inequality) If $(\hat{U}^*, \hat{\mathcal{Y}}^*)$ is a nondegenerate KKT point of problem (8), then there exist $\mu > 0$ and $\epsilon > 0$ such that

$$\sum_{i=1}^k \|\nabla_i f(\hat{U}) - U^{(i)} \nabla_i (f(\hat{U}))^\top U^{(i)}\|_F^2 \geq \mu |f(\hat{U}) - f(\hat{U}^*)|$$

(30)

for any $(\hat{U}, x) \in V_{n,r}$ such that $\|\hat{U} - \hat{U}^*\|_F \leq \epsilon$. Here $f$ is the objective function of problem (9) defined by (23).

**Proof** Let $\delta > 0$ be the radius of the neighborhood given by Proposition 12. Since $(\hat{U}^*, \hat{\mathcal{Y}}^*)$ is a KKT point of (8), with $\hat{U}^* = (U^{(*,1)}, \ldots, U^{(*,k)})$, we have

$$\text{Diag}(\hat{\mathcal{Y}}^*) = \text{Diag} \left( ((U^{(*,1)})^\top, \ldots, (U^{(*,k)})^\top) \cdot A \right).$$

$\square$ Springer
For a given $\mathbb{U}$, let $\mathcal{Y}$ be defined as

$$\text{Diag}(\mathcal{Y}) := \text{Diag} \left( ((U^{(1)})^\top, \ldots, (U^{(k)})^\top) \cdot \mathcal{A} \right).$$

Then, there exists $\epsilon > 0$ such that

$$\| (\mathbb{U}, \mathcal{Y}) - (\mathbb{U}^*, \mathcal{Y}^*) \|_F \leq \delta$$

whenever $\| \mathbb{U} - \mathbb{U}^* \|_F \leq \epsilon$.

By Proposition 7, Proposition 12 is applicable to $(\mathbb{U}^*, \mathbf{x}^* := \text{Diag}(\mathcal{Y}^*)) \in U_{n,r}$ for the function $g$. Thus, there exists $\mu_0 > 0$ such that

$$\| \text{grad}(g)(\mathbb{U}, \mathbf{x}) \|^2 \geq \mu_0 |g(\mathbb{U}, \mathbf{x}) - g(\mathbb{U}^*, \mathbf{x}^*)|$$

for all $(\mathbb{U}, \mathbf{x}) \in V_{n,r}$ such that $\| \mathbb{U} - \mathbb{U}^* \|_F \leq \epsilon$, where $\mathbf{x} = \text{Diag}(\mathcal{Y})$ is formed by the diagonal elements of $\mathcal{Y}$.

We first have

$$|g(\mathbb{U}, \mathbf{x}) - g(\mathbb{U}^*, \mathbf{x}^*)| = \frac{1}{2} |f(\mathbb{U}) - f(\mathbb{U}^*)|,$$

since $f(\mathbb{U}) = \| \mathbf{x} \|^2$ and

$$g(\mathbb{U}, \mathbf{x}) = \frac{1}{2} \| \mathcal{A} - (U^{(1)}, \ldots, U^{(k)}) \cdot \text{diag}(\mathbf{x}) \|^2$$

$$= \frac{1}{2} \| \mathcal{A} \|^2 - \langle \mathcal{A}, (U^{(1)}, \ldots, U^{(k)}) \cdot \text{diag}(\mathbf{x}) \rangle + \frac{1}{2} \| \mathbf{x} \|^2$$

$$= \frac{1}{2} \| \mathcal{A} \|^2 - \langle \text{Diag} \left( ((U^{(1)})^\top, \ldots, (U^{(k)})^\top) \cdot \mathcal{A} \right), \mathbf{x} \rangle + \frac{1}{2} \| \mathbf{x} \|^2$$

$$= \frac{1}{2} \left( \| \mathcal{A} \|^2 - \| \mathbf{x} \|^2 \right).$$

By (29) and the definition of $\mathbf{x}$, we also have

$$\text{grad}_x g(\mathbb{U}, \mathbf{x}) = \mathbf{x} - \text{Diag} \left( ((U^{(1)})^\top, \ldots, (U^{(k)})^\top) \cdot \mathcal{A} \right) = 0.$$

Since

$$\text{grad}_{U^{(i)}} g(\mathbb{U}, \mathbf{x}) = -(I - \frac{1}{2} U^{(i)}(U^{(i)})^\top)(V^{(i)} \Gamma - U^{(i)}(V^{(i)} \Gamma)^\top U^{(i)}),$$

for all $i = 1, \ldots, k,$

and

$$\nabla_i f(\mathbb{U}) = 2V^{(i)} \Gamma,$$ for all $i = 1, \ldots, k,$
where $\Gamma = \text{diag}(\mathbf{x})$ is the diagonal matrix formed by the vector $\mathbf{x}$, the assertion will follow if we can show that

$$
\|I - \frac{1}{2} U^{(i)} (U^{(i)})^T\|_F \leq \mu_1
$$

is uniformly bounded by $\mu_1 > 0$ over $\|\mathbb{U} - \mathbb{U}^*\|_F \leq \epsilon$. This is obviously true. The proof is then complete.

4 iAPD algorithm and convergence analysis

4.1 Description of iAPD algorithm

In [26], an alternating polar decomposition (APD) algorithm is proposed to solve the optimization problem (9). Algorithm 1 below is an improved version of the classical APD, which we call iAPD. It is an alternating polar decomposition method with adaptive proximal corrections and truncations. We notice that since both APD and iAPD are methods designed to solve (9), the initial value of $r$ is pre-assigned by (9) itself to characterize the low rank structure. Also, in the truncation step of Algorithm 1, $r$ may decrease. This is because $D(\mathbf{n}, r)$ is not closed and a maximizer of (9) may not lie in $D(\mathbf{n}, r)$. In this case, we need to search a maximizer on the boundary $C(\mathbf{n}, r - 1)$ of $D(\mathbf{n}, r)$ via the truncation step. Moreover, the constant parameter $\epsilon$ is used in the proximal correction step to avoid possible numerical issues raised by singular values being too small. Therefore, in practice we choose $\epsilon$ to be some positive constant with a moderate magnitude. Because the aforementioned reasons, values of $r$ and $\epsilon$ should be considered as a part of the input in iAPD. An iteration step in iAPD with a truncation is called a truncation iteration. Obviously, there are at most $r$ truncation iterations.

In the following, we briefly explain the rationale of Algorithm 1. First we recall that the problem we want to solve is (cf. (23))

$$
\max f(\mathbb{U}) \quad \text{s.t. } (U^{(i)})^T U^{(i)} = I \text{ for all } i = 1, \ldots, k. \quad (32)
$$

The guiding principle to solve (32) is maximizing $U^{(i)}$ alternatively for $i = 1, \ldots, k$ in a cyclic order. More precisely, each time we fix all but one $U^{(i)}$ in the block variable $\mathbb{U}$, and optimize $U^{(i)}$ to get a better objective function value. We notice that (cf. (20) and (21))

$$
f(\mathbb{U}) = \langle U^{(i)}, V^{(i)} \Lambda \rangle
$$

and $\Lambda$ depends on $U^{(i)}$ as well. If we further fix the matrix $\Lambda$ as a constant, then the resulting subproblem

$$
\max \langle U^{(i)}, V^{(i)} \Lambda \rangle \quad \text{s.t. } (U^{(i)})^T U^{(i)} = I \quad (33)
$$

is uniformly bounded by $\mu_1 > 0$ over $\|\mathbb{U} - \mathbb{U}^*\|_F \leq \epsilon$. This is obviously true. The proof is then complete.
can be solved by polar decomposition efficiently. Substep 1 in Algorithm 1 computes a solution for (33), and it forms one of the $k$ inner loops of Step 1 in Algorithm 1. If this procedure is repeated until a termination criterion is fulfilled, it is exactly APD proposed in [26]. Hence APD is simply iAPD (i.e., Algorithm 1) without Substep 2 and Step 2. Secondly, Substep 2 and Step 2 are designed for the desired convergence properties of iAPD. The convergence of APD relies on the error bound for the problem (33), which is presented in Theorem 9. We can see from (75) that this bound is determined by the smallest singular value of a matrix. If it is zero, then we do not have a sufficient increase of the objective function value to carry out the convergence analysis. Therefore, it requires the full rank condition for a certain matrix sequence to establish the convergence of APD in [26]. In order to remove this additional stringent requirement, we employ an adaptive proximal step (i.e., Substep 2). Instead of (33), we solve the following problem

$$\max \langle U^{(i)}, V^{(i)} A \rangle - \frac{\epsilon}{2} \|U^{(i)} - U_0^{(i)}\|^2$$

s.t. $(U^{(i)})^T U^{(i)} = I$, \hspace{1cm} (34)

where $U_0^{(i)}$ is obtained from the previous iteration. Substep 2 computes a solution for problem (34). By doing this, we have the desired sufficient increase property (cf. Proposition 8). Thirdly, in order to obtain the linear convergence, a truncation step is applied (i.e., Step 2). This step removes the columns of the iteration matrices $U^{(i)}$ corresponding to small $\lambda_j(U)'s$ (cf. (22)). Actually, by doing this, iterations from iAPD converge to a primitive KKT point and hence we can apply the analysis in Sect. 3.3 to derive the linear convergence. Lastly, if we set $r = 1$, then there will be no proximal correction and truncation steps in Algorithm 1. Hence iAPD actually reduces to APD for $r = 1$, which generates the same iterative sequence up to scaling as the higher order power method (HOPM) discussed in full details in [31].

The main storage of Algorithm 1 consists of those for the given tensor $A$, two adjacent iterations $U_{[p-1]}$ and $U_{[p]}$, the matrix $V_{[p]}^{(i)}$, a vector for $\lambda_{j,[p]}^{i-1}$, and the three matrices for SVD in Substep 1. Therefore, the total storage is at most the aggregation of that for a tensor of size $n_1 \times \cdots \times n_k$ and $2k + 4$ matrices of size $\max\{n_i\} \times r$. We also remark that the storage of $A$ could be quite problem dependent.

We conclude this subsection with a few remarks on the computational cost of Algorithm 1. The most important ingredient of APD and iAPD is the polar decomposition of a matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$. It is known that the polar decomposition can be computed either by the singular value decomposition (SVD) [24] with complexity cost $O(mn^2)$ flops or Newton’s method [29]. In Algorithm 1 we choose the SVD based method for the polar decomposition because it is more numerically stable and is readily accessible in numerical linear algebra softwares library like LAPACK [3]. For easy reference, we summarize some basic properties of the polar decomposition required for our convergence analysis of iAPD in Appendix B.1. Interested readers are referred to [24] for more details.

The basic computation in iAPD is the evaluation of $A \tau_i(x_{j,[p]}^i)$ for a block vector $x_{j,[p]}^i$. The (multiplication) complexity is quite problem dependent. For example, if $A$ is a dense tensor, then it would be $(k - 1) \prod_{i=1}^k n_i$. While it would be $\sum_{i=1}^k n_i + k - 2$
Algorithm 1 iAPD for Low Rank Orthogonal Tensor Approximation

Input: a nonzero tensor $A \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k}$, a positive integer $r$, and a proximal parameter $\epsilon$.

1: **Step 0** [Initialization]: choose $U_{[0]} := (U_{[0]}^{(1)}, \ldots, U_{[0]}^{(k)}) \in V(r, n_1) \times \cdots \times V(r, n_k)$ such that $f(U_{[0]}) > 0$, and a truncation parameter $\kappa \in (0, \sqrt{f(U_{[0]})}/r)$. Let $p := 1$.

2: **Step 1** [Alternating Polar Decompositions-APD]: Let $i := 1$.

3: Substep 1 [Polar Decomposition]: If $i > k$, go to Step 2. Otherwise, for all $j = 1, \ldots, r$, let

$$
x_{j,[p]}^{(i)} := (u_{j,[p]}^{(1)}, \ldots, u_{j,[p-1]}^{(i)}),
$$

where $u_{j,[p]}^{(i)}$ is the $j$-th column of the factor matrix $U_{[p]}^{(i)}$.

Compute the matrix $A_{[p]}^{(i)}$ as

$$
A_{[p]}^{(i)} := \text{diag}(\lambda_{i,1,[p]}^{(i)}, \ldots, \lambda_{i,r,[p]}^{(i)}) \text{ with } \lambda_{i,j,[p]}^{(i)} := \sigma_{j,[p]}^{(i)}(x_{j,[p]}^{(i)}) \text{ for } j = 1, \ldots, r,
$$

and the matrix $V_{[p]}^{(i)}$ as

$$
V_{[p]}^{(i)} := \left[ v_{1,[p]}^{(i)} \cdots v_{r,[p]}^{(i)} \right] \text{ with } v_{j,[p]}^{(i)} := \sigma_{j,[p]}^{(i)}(x_{j,[p]}^{(i)}) \text{ for } j = 1, \ldots, r.
$$

Compute the singular value decomposition of the matrix $V_{[p]}^{(i)} A_{[p]}^{(i)}$ as

$$
V_{[p]}^{(i)} A_{[p]}^{(i)} = G_{[p]}^{(i)} \Sigma_{[p]}^{(i)} (H_{[p]}^{(i)})^\top, \quad G_{[p]}^{(i)} \in V(r, n_j) \text{ and } H_{[p]}^{(i)} \in O(r),
$$

where the singular values $\sigma_{j,[p]}^{(i)}$’s are ordered nonincreasingly in the diagonal matrix $\Sigma_{[p]}^{(i)}$. Then the polar decomposition of the matrix $V_{[p]}^{(i)} A_{[p]}^{(i)}$ is

$$
V_{[p]}^{(i)} A_{[p]}^{(i)} = U_{[p]}^{(i)} S_{[p]}^{(i)} \text{ with } U_{[p]}^{(i)} := G_{[p]}^{(i)} (H_{[p]}^{(i)})^\top \text{ and } S_{[p]}^{(i)} := (H_{[p]}^{(i)})^{-1} (H_{[p]}^{(i)})^\top.
$$

4: Substep 2 [Proximal Correction]: If $\sigma_{r,[p]}^{(i)} < \epsilon$, then compute the polar decomposition of the matrix $V_{[p]}^{(i)} A_{[p]}^{(i)} + \epsilon U_{[p-1]}^{(i)}$ as

$$
V_{[p]}^{(i)} A_{[p]}^{(i)} + \epsilon U_{[p-1]}^{(i)} = \hat{U}_{[p]}^{(i)} \hat{S}_{[p]}^{(i)}
$$

for an orthonormal matrix $\hat{U}_{[p]}^{(i)} \in V(r, n_j)$ and a symmetric positive semidefinite matrix $\hat{S}_{[p]}^{(i)}$. Update $U_{[p]}^{(i)} := \hat{U}_{[p]}^{(i)}$ and $S_{[p]}^{(i)} := \hat{S}_{[p]}^{(i)}$. Set $i := i + 1$ and go to Substep 1.

5: **Step 2** [Truncation]: If $\lambda_{j,[p+1]}^{(i)} = \langle (U_{[p]}^{(k)})^\top V_{[p]}^{(k)} \rangle_{jj} < \kappa$ for some $j \in \{1, \ldots, r\}$, formulate matrices $\hat{U}_{[p]}^{(i)}$’s, where $\hat{U}_{[p]}^{(i)}$ is an $n_j \times (r - |J|)$ matrix formed by the columns of $U_{[p]}^{(i)}$ corresponding to $\{1, \ldots, r\} \setminus J$, for all $i \in \{1, \ldots, k\}$. Update $r := r - |J|$, and $U_{[p]}^{(i)} := \hat{U}_{[p]}^{(i)}$ for all $i \in \{1, \ldots, k\}$. Go to Step 3.

6: **Step 3**: Unless a termination criterion is satisfied, let $p := p + 1$ and go back to Step 1.
if it is a rank-one tensor and therefore \( O(\sum_{i=1}^{k} n_i) \) if it is a low rank tensor. Due to the dependence on the structure of the given data, we uniformly denote this complexity by \( \omega \). In Substep 1, the computational cost of the matrix \( V_{[p]}^{(i)} \) is \( r \omega \), and that of \( \lambda_{j,[p]}^{i-1} \)'s is an additional \( r n_i \). The rest computational cost in Substep 1 is the polar decomposition, which is \( O(n_i r^2) \). If the proximal correction step is executed, another polar decomposition is implemented and the cost is \( O(n_i r^2) \). The computational cost for the truncation step is negligible. Therefore, the total computational cost for one iteration in Algorithm 1 is \( k r \omega + O(\sum_{i=1}^{k} n_i r^2) \).

### 4.2 Properties of iAPD

In this section, we derive some inequalities for the increments of the objective function during iterations. To do this, we define

\[
U_{i,[p]} := (U_{[p]}^{(1)}, \ldots, U_{[p]}^{(i)}, U_{[p-1]}^{(i+1)}, \ldots, U_{[p-1]}^{(k)})
\]  

for each \( i \in \{0, \ldots, k\} \), and

\[
U_{[p]} := (U_{[p]}^{(1)}, \ldots, U_{[p]}^{(k)}),
\]

which is equal to \( U_{k,[p]} = U_{0,[p+1]} \) for each \( p \in \mathbb{N} \). We remark that the \( j \)-th column of a factor matrix \( U_{[p]}^{(i)} \) is denoted by \( u_{j,[p]}^{(i)} \) for each \( j \in \{1, \ldots, r\} \) while the superscript \( i \) for the block vector \( x_{j,[p]}^{i} \) is not bracketed. For each \( j \in \{1, \ldots, r\} \) and \( p \in \mathbb{N} \), we also denote

\[
\lambda_{j,[p-1]}^{i-1} := \lambda_{j,[p]}^{0},
\]

where \( \lambda_{j,[p]}^{i-1} \) is defined in (36) for the \( i \)-th inner iteration. One immediate observation is that if the \( p \)-th iteration in Algorithm 1 is not a truncation iteration, then the sizes of the matrices in \( U_{[p]} \) and those in \( U_{[p-1]} \) are the same. Also if the number of iterations in Algorithm 1 is infinite, then there is a sufficiently large \( N_0 \) such that the \( p \)-th iteration is not a truncation iteration for any \( p \geq N_0 \). The proof of the next lemma can be found in Appendix C.1.

**Lemma 6** (Monotonicity of iAPD) *If the \( p \)-th iteration in Algorithm 1 is not a truncation iteration, then for each \( 0 \leq i \leq k - 1 \), we have*

\[
f(U_{i+1,[p]}) - f(U_{i,[p]}) \geq \frac{\epsilon}{2} \|U_{[p]}^{(i+1)} - U_{[p-1]}^{(i+1)}\|_F^2.
\]

**Proposition 8** (Sufficient Increase) *If the \( p \)-th iteration in Algorithm 1 is not a truncation iteration, then we have*

\[
f(U_{[p]}) - f(U_{[p-1]}) \geq \frac{\epsilon}{2} \|U_{[p]} - U_{[p-1]}\|_F^2.
\]
Proof We have
\[
\begin{align*}
f(U[p]) - f(U[p-1]) &= \sum_{i=0}^{k-1} \left( f(U[i+1,p]) - f(U[i,p]) \right) \\
&\geq \frac{\epsilon}{2} \sum_{i=0}^{k-1} \| U[i+1,p] - U[i,p] \|_F^2 \\
&= \frac{\epsilon}{2} \| U[p] - U[p-1] \|_F^2,
\end{align*}
\]
where the inequality follows from (44) in Lemma 6.

At each truncation iteration, the number of columns of the matrices in $U[p]$ is decreased strictly. The first issue we have to address is that the iteration $U[p]$ is not vacuous, i.e., the numbers of the columns of the matrices in $U[p]$ are stable and positive. We have the following proposition, which is recorded for latter reference.

**Proposition 9** The number of columns of $U[i]^{(p)}$’s will be stable at a positive integer $s \leq r$ and there exists $N_0$ such that $f(U[p])$ is nondecreasing for all $p \geq N_0$.

**Proof** Since the initial number $r$ of columns is finite, the truncation occurs at most $r$ times and the total decreased number of columns of matrices in $U[p]$ is bounded above by $r$. It follows from Step 2 of Algorithm 1 and Lemma 6 that if the $p$-th iteration is a truncation iteration and the number of columns of the matrices in $U[p]$ is decreased from $r_1$ to $r_2 < r_1$, then we have
\[
f(U[p]) \geq f(U[p-1]) - (r_1 - r_2)\kappa^2.
\]
By the truncation strategy in Algorithm 1, after all the truncation iterations, the value of the objective function decreases by at most $r\kappa^2$. Moreover, at each iteration without truncation, the value of the objective function is nondecreasing by Lemma 6 and $r\kappa^2 < f(U[0])$. Hence, $U[p]$ cannot be vacuous. As there can only be a finite number of truncations, there exists $N_0$ such that for any $p \geq N_0$, the $p$-th iteration is not a truncation iteration, and the conclusion then follows.

Let us consider the following optimization problem
\[
\min_U h(U) := -f(U) + \sum_{i=1}^k \delta_{V(r,n_i)}(U^{(i)}), 
\]
where $\delta_{V(r,n_i)}(U^{(i)})$ is the indicator function for the set $V(r,n_i)$ defined by (68). By [56, Theorem 8.2 and Proposition 10.5], we have
\[
\partial \left( \sum_{i=1}^k \delta_{V(r,n_i)}(U^{(i)}) \right) = \partial \delta_{V(r,n_1)}(U^{(1)}) \times \cdots \times \partial \delta_{V(r,n_k)}(U^{(k)}).
\]
Moreover, according to [56, Exercise 8.8] and (69) we also obtain that
\[
\partial h(U) = -\nabla f(U) + \partial \left( \sum_{i=1}^k \delta_{V(r,n_i)}(U^{(i)}) \right).
\]
We recall that $\nabla_i f(U)$ is the gradient $\nabla_{U^{(i)}} f(U)$. With the notation in Sect. 3.2.1, we have $\nabla_i f(U) = 2V^{(i)} A$ for all $i = 1, \ldots, k$. Thus, it follows from Definition 1, (65), (69) and (47) that critical points of $h$ are exactly KKT points of problem (9).

It is straightforward to verify that $h$ is a KL function according to Lemma 13, and (46) is an unconstrained reformulation of problem (9) in the sense that

$$(9) \iff - \min_U h(U).$$

Readers can find the proof of the next lemma in Appendix C.2.

**Lemma 7 (Subdifferential Bound)** If the $(p + 1)$-st iteration is not a truncation iteration, then there exists a subgradient $W_{[p+1]} \in \partial h(U_{[p+1]})$ such that

$$\|W_{[p+1]}\|_F \leq 2\sqrt{k}(2r\sqrt{k}\|A\|^2 + \epsilon)\|U_{[p+1]} - U_{[p]}\|_F.$$ (48)

### 4.3 Global convergence

The following classical result can be found in [5].

**Lemma 8 (Abstract Convergence)** Let $p : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm \infty\}$ be a proper lower semicontinuous function and let $\{x^{(k)}\} \subseteq \mathbb{R}^n$ be a sequence satisfying the following properties

1. there is a constant $\alpha > 0$ such that

$$p(x^{(k)}) - p(x^{(k+1)}) \geq \alpha\|x^{(k+1)} - x^{(k)}\|^2,$$

2. there is a constant $\beta > 0$ and a $w^{(k+1)} \in \partial p(x^{(k+1)})$ such that

$$\|w^{(k+1)}\| \leq \beta\|x^{(k+1)} - x^{(k)}\|,$$

3. there is a subsequence $\{x^{(k_i)}\}$ of $\{x^{(k)}\}$ and $x^* \in \mathbb{R}^n$ such that

$$x^{(k_i)} \rightarrow x^* \text{ and } p(x^{(k_i)}) \rightarrow p(x^*) \text{ as } i \rightarrow \infty.$$

If $p$ has the Kurdyka-Łojasiewicz property at the point $x^*$, then the whole sequence $\{x^{(k)}\}$ converges to $x^*$, and $x^*$ is a critical point of $p$.

As our domain of optimization is the product Stiefel manifold which is compact, the iteration sequence must be bounded. Hence the condition 3 in Lemma 8 is satisfied automatically.

Regarding the global convergence of Algorithm 1, by Proposition 9, we can assume without loss of generality that there is no truncation iteration in the sequence $\{U_{[p]}\}$ generated by Algorithm 1.

**Proposition 10** Given a sequence $\{U_{[p]}\}$ generated by Algorithm 1, the sequence $\{f(U_{[p]})\}$ increases monotonically and hence converges.
Proof Since the sequence \( \{U[p]\} \) is bounded, \( \{f(U[p])\} \) is bounded as well by the definition (cf. (23)). The convergence then follows from Proposition 9.

Now we are ready to prove the global convergence of Algorithm 1, which is the first main theorem in our convergence analysis.

**Theorem 2** (Global Convergence) Any sequence \( \{U[p]\} \) generated by Algorithm 1 is bounded and converges to a KKT point of the problem (9).

Proof Obviously, the sequence \( \{U[p]\} \) is bounded and the function \( h \) is continuous on the product of the Stiefel manifolds. Thus the convergence follows from Proposition 8, Lemma 7, Lemma 8, and Proposition 10.

It is clear that the converged KKT point is primitive by the truncation strategy. We remark that Theorem 2 can only ensure the global convergence to a KKT point of (9), which is not necessarily a maximizer. Even if the tensor \( \mathcal{A} \) is orthogonally decomposable, there is no guarantee for a global convergence to an orthogonal decomposition of \( \mathcal{A} \). However, for a generic tensor \( \mathcal{A} \), there is a neighborhood of a global maximizer such that for any initialization in this neighborhood the sequence generated by Algorithm 1 is guaranteed to converge to this global maximizer. This can be easily seen by the second order optimality condition, the nondegeneracy of the KKT point in the generic case (Theorem 1) and the global convergence of iAPD (Theorem 2). In particular, if this \( \mathcal{A} \) is further orthogonally decomposable, then Algorithm 1 is assured to converge to an orthogonal decomposition of \( \mathcal{A} \) provided that the initial point is chosen appropriately.

Moreover, it is proved in [40] that the gradient descent algorithm can avoid converging to a strict saddle point for almost all initializations. We also recall from Theorem 1 that all KKT points of (8) are nondegenerate for a generic \( \mathcal{A} \). In particular, a KKT point is either a strict saddle point or a strict local minimizer or a strict local maximizer. Based on the above observations, it is seemingly plausible to strengthen Theorem 2 for generic \( \mathcal{A} \) by a result proved in [40]. Unfortunately, a closer investigation indicates that Algorithm 1 is quite different from the gradient descent algorithm considered in [40]. To be more specific, since Algorithm 1 is based on the alternating method, we only need to update our candidate with respect to the partial gradient in each iteration, while the algorithm discussed in [40] employs the whole gradient in every iteration. Therefore results obtained in [40] are not directly applicable to our situation. However, we believe that the argument used in [40, 41] can be adapted to our situation and we carry out some further studies along this line in the future.

Lastly, although we have no guarantee on the convergence to a local maximizer, in Sect. 5, we do investigate some properties (Theorems 6 and 7 and their respective corollaries) of local maximizers of (9), or equivalently of local minimizers of (8).
4.4 Sublinear convergence rate

We consider the following function

\[ q(U, P) := f(U) - \sum_{i=1}^{k} (P^{(i)} \cdot (U^{(i)})^T U^{(i)} - I), \]  

(49)

which is a polynomial of degree 2 in \( N := \sum_{i=1}^{k} (rn_i + \binom{r+1}{2}) \) variables:

\[ (U, P) = (U^{(1)}, \ldots, U^{(k)}, P^{(1)}, \ldots, P^{(k)}) \in \mathbb{R}^{n_1 \times r} \times \cdots \times \mathbb{R}^{n_k \times r} \times S^{r \times r} \times \cdots \times S^{r \times r}. \]

Let

\[ \tau := 1 - \frac{1}{2k(6k - 3)^{N-1}}, \]  

(50)

which is the Lojasiewicz exponent of the polynomial \( q \) obtained by Lemma 14. We suppose that \( U^* \) is a KKT point of (9) with the multiplier \( P^* \). For

\[ \hat{q}(U, P) := q(U, P) - q(U^*, P^*), \]  

(51)

we must have

\[ \hat{q}(U^*, P^*) = 0, \quad \nabla \hat{q}(U^*, P^*) = 0. \]

Thus according to Lemma 14, there exist some \( \gamma, \mu > 0 \) such that

\[ \| \nabla \hat{q}(U, P) \|_F \geq \mu |\hat{q}(U, P)|^\tau \text{ whenever } \| (U, P) - (U^*, P^*) \|_F \leq \gamma. \]

Therefore,

\[ \sum_{i=1}^{k} \| \nabla f_i(U) - 2U^{(i)} P^{(i)} \|_F^2 \geq \mu^2 (f(U) - f(U^*))^{2\tau} \]  

(52)

for any feasible point \( U \) of (9) satisfying \( \| (U, P) - (U^*, P^*) \|_F \leq \gamma. \)

With all the above preparations, we are able to derive the second main result in our convergence analysis, which is concerned with the sublinear convergence rate of Algorithm 1.

Theorem 3 (Sublinear Convergence Rate) Let \( \{U[p]\} \) be a sequence generated by Algorithm 1 for a given nonzero tensor \( A \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} \) and let \( \tau \) be defined by (50). The following statements hold:

(i) the sequence \( \{ f(U[p]) \} \) converges to \( f^* \), with sublinear convergence rate at least \( O(p^{1-2\tau}) \), that is, there exist \( M_1 > 0 \) and \( p_0 \in \mathbb{N} \) such that for all \( p \geq p_0 \)

\[ f^* - f(U[p]) \leq M_1 p^{\frac{1}{1-2\tau}}; \]  

(53)
(ii) \( \{U_{[p]}\} \) converges to \( U^* \) globally with the sublinear convergence rate at least \( O(p^{\frac{2}{p+1}}) \), that is, there exist \( M_2 > 0 \) and \( p_0 \in \mathbb{N} \) such that for all \( p \geq p_0 \)

\[
\|U_{[p]} - U^*\|_F \leq M_2 \, p^{\frac{2}{p+1}}.
\]

**Proof** In the following, we consider the sequence \( \{U_{[p]}\} \) generated by Algorithm 1. Let

\[
P_{[p]}^{(i)} := S_{[p]}^{(i)} - \alpha_{i,[p]}I := \begin{cases} 
S_{[p]}^{(i)} - \epsilon I & \text{if proximal correction is executed,} \\
S_{[p]}^{(i)} & \text{otherwise,}
\end{cases}
\]

where \( \alpha_{i,[p]} \in \{0, \epsilon\} \). We also have that

\[
S_{[p]}^{(i)} = \begin{cases} 
(U_{[p]}^{(i)})^T(V_{[p]}^{(i)}A_{[p]}^{(i)} + \epsilon U_{[p-1]}^{(i)}) & \text{if proximal correction is executed,} \\
(U_{[p]}^{(i)})^TV_{[p]}^{(i)}A_{[p]}^{(i)} & \text{otherwise.}
\end{cases}
\]

Note that \( \{U_{[p]}\} \) converges by Theorem 2 and hence \( \{V_{[p]}^{(i)}A_{[p]}^{(i)}\} \) converges. Recall that the proximal correction step is determined by singular values of the matrices \( V_{[p]}^{(i)}A_{[p]}^{(i)} \)'s. Thus, for sufficiently large \( p \) (say \( p \geq p_0 \), \( \alpha_{i,[p]} \) will be stable at \( \alpha_i \) for all \( p \) and \( 1 \leq i \leq k \). By Lemma 7, (39), (40) and (95), we have

\[
\|\nabla_i f(U_{[p+1]}^{(i)}) - 2U_{[p+1]}^{(i)}P_{[p+1]}^{(i)}\|_F \leq \|\nabla_i f(U_{[p+1]}^{(i)}) - 2U_{[p+1]}^{(i)}S_{[p+1]}^{(i)} + 2\alpha_i U_{[p+1]}^{(i)}\|_F \\
\leq B\|U_{[p+1]}^{(i)} - U_{[p]}^{(i)}\|_F,
\]

where \( B := 2(2r\sqrt{\kappa}\|A\|^2 + \epsilon) \). Let \( \hat{P}_{[p]}^{(i)} := \frac{1}{2}((U_{[p]}^{(i)})^T\nabla_i f(U_{[p]}^{(i)}) + (\nabla_i f(U_{[p]}^{(i)}))^TU_{[p]}^{(i)}) \) for all \( i = 1, \ldots, k \). Then \( \hat{P}_{[p]}^{(i)} \) is symmetric. In the notation of the proof for Lemma 7, we have \( \nabla_i f(U_{[p+1]}^{(i)}) = 2V_{[p]}^{(i)}A \) (cf. (92) and (93)). Thus, we have

\[
\|U_{[p+1]}^{(i)}P_{[p+1]}^{(i)}(U_{[p+1]}^{(i)})^T\hat{P}_{[p]}^{(i)}\|_F = \|S_{[p+1]}^{(i)} - \alpha_i I - \frac{1}{2}((U_{[p+1]}^{(i)})^TV_{[p]}^{(i)}A + (V_{[p]}^{(i)})^TU_{[p+1]}^{(i)})\|_F \\
\leq \frac{1}{2}\|((U_{[p+1]}^{(i)})^TV_{[p]}^{(i)}A_{[p+1]}^{(i)} + \alpha_i U_{[p]}^{(i)}) - \alpha_i I - (U_{[p+1]}^{(i)})^TV_{[p]}^{(i)}A\|_F \\
+ \|V_{[p+1]}^{(i)}A_{[p+1]}^{(i)} + \alpha_i U_{[p]}^{(i)}\|_F^2 - \|V_{[p+1]}^{(i)}A_{[p+1]}^{(i)} - \alpha_i I - (U_{[p+1]}^{(i)})^TV_{[p]}^{(i)}A\|_F \\
\leq \|((U_{[p+1]}^{(i)})^TV_{[p]}^{(i)}A_{[p+1]}^{(i)} + \alpha_i U_{[p]}^{(i)}) - \alpha_i I - (U_{[p+1]}^{(i)})^TV_{[p]}^{(i)}A\|_F \\
+ \|U_{[p+1]}^{(i)}^TV_{[p]}^{(i)}A_{[p+1]}^{(i)} - V_{[p]}^{(i)}A\|_F + \epsilon\|U_{[p+1]}^{(i)} - U_{[p]}^{(i)}\|_F \\
\leq C_0\|V_{[p+1]}^{(i)}A_{[p+1]}^{(i)} - V_{[p]}^{(i)}A\|_F + \epsilon\|U_{[p+1]}^{(i)} - U_{[p]}^{(i)}\|_F \\
\leq C\|U_{[p+1]}^{(i)} - U_{[p]}^{(i)}\|_F,
\]

where the second equality follows from the fact that the matrix \( S_{[p+1]}^{(i)} = (U_{[p+1]}^{(i)})^TV_{[p+1]}^{(i)}A_{[p+1]}^{(i)} + \alpha_i U_{[p]}^{(i)} \) is symmetric, and \( C_0, C > 0 \) are some constants depending on \( A \) only (cf. the proof for Lemma 7).
Therefore, by (54) and (55), we have

\[
\|\nabla_i f(\mathbb{U}_{[p+1]}) - 2U_{[p+1]}^{(i)}\hat{P}_{[p+1]}^{(i)}\|_F \\
= \|\nabla_i f(\mathbb{U}_{[p+1]}) - 2U_{[p+1]}^{(i)}P_{[p+1]}^{(i)} + 2U_{[p+1]}^{(i)}P_{[p+1]}^{(i)} - 2U_{[p+1]}^{(i)}\hat{P}_{[p+1]}^{(i)}\|_F \\
\leq \|\nabla_i f(\mathbb{U}_{[p+1]}) - 2U_{[p+1]}^{(i)}P_{[p+1]}^{(i)}\|_F \\
+ 2\|U_{[p+1]}^{(i)}P_{[p+1]}^{(i)} - U_{[p+1]}^{(i)}\hat{P}_{[p+1]}^{(i)}\|_F \\
\leq (B + 2C)\|\mathbb{U}_{[p+1]} - \mathbb{U}_{[p]}\|_F. \tag{56}
\]

Since \{\mathbb{U}_{[p]}\} converges by Theorem 2, we see that

\[
\lim_{p \to \infty} \hat{P}_{[p]}^{(i)} = (U^{(*,i)})^T V^{(*,i)} A^* := P^{(*,i)}.
\]

Hence for sufficiently large \(p\), we may conclude that

\[
\|\mathbb{U}_{[p]}, \hat{P}_{[p]}^{(i)} - (\mathbb{U}^*, \mathbb{P}^*)\|_F \leq \gamma.
\]

This implies

\[
\mu^2(f(\mathbb{U}_{[p]}) - f(\mathbb{U}^*))^{2\tau} \leq \sum_{i=1}^{k} \|\nabla_i f(\mathbb{U}_{[p]}^{(i)}) - 2U_{[p]}^{(i)}\hat{P}_{[p]}^{(i)}\|_F^2 \\
\leq 2\sum_{i=1}^{k} \|\nabla_i f(\mathbb{U}_{[p+1]}^{(i)}) - 2U_{[p+1]}^{(i)}\hat{P}_{[p+1]}^{(i)}\|_F^2 \\
+ 2\sum_{i=1}^{k} \|\nabla_i f(\mathbb{U}_{[p]}^{(i)}) - 2U_{[p]}^{(i)}\hat{P}_{[p]}^{(i)} - (\nabla_i f(\mathbb{U}_{[p+1]}^{(i)}) - 2U_{[p+1]}^{(i)}\hat{P}_{[p+1]}^{(i)})\|_F^2 \\
\leq (2k(B + 2C)^2 + L)(f(\mathbb{U}_{[p+1]}^{(i)}) - f(\mathbb{U}_{[p]}^{(i)})). \tag{57}
\]

where the first inequality follows from (52), the third from (56), and the fact that the function \(\nabla_i f(\mathbb{U}_{[p]}^{(i)}) - 2U_{[p]}^{(i)}\hat{P}_{[p]}^{(i)}\) in (57) is Lipschitz continuous on the product of Stiefel manifolds, and the last one follows from Proposition 8. Here \(L > 0\) is a constant determined by the Lipschitz constant of the function in (57).

If we set \(\beta_p := f(\mathbb{U}^*) - f(\mathbb{U}_{[p]}^{(i)})\), then we have

\[
\beta_p - \beta_{p+1} \geq M\beta_p^{2\tau}
\]
for some constant $M > 0$, from which we can show
\[ \beta_{p+1}^{1-2\tau} - \beta_p^{1-2\tau} \geq (2\tau - 1)M. \]

Thus,
\[ \beta_p^{1-2\tau} \geq M(2\tau - 1)(p - p_0) + \beta_p^{1-2\tau} \]
and the conclusion follows since $\tau \in (\frac{1}{2}, 1)$. For a more detailed analysis on the sequence $\{\beta_p\}$, we refer readers to [31, Section 3.4]. \hfill \Box

We remark that the convergence rate in (53) is faster than the classical $O(1/p)$ for first order methods in optimization [6], while the optimal rate is $O(1/p^2)$ for convex problems by the celebrated work of Nesterov [54].

4.5 Linear convergence

In this subsection, we establish the linear convergence of Algorithm 1. The proof of the next lemma is available in Appendix C.3.

**Lemma 9** (Relative Error) There exists a constant $\gamma > 0$ such that whenever the $(p+1)$-st iteration is not a truncation iteration, it holds
\[ \|\nabla_i f(U[p+1]) - U^{(i)}[p+1](\nabla_i f(U[p+1]))^TU^{(i)}[p+1]\|_F \leq \gamma\|U[p] - U[p+1]\|_F \]
for all $1 \leq i \leq k$.

**Theorem 4** (Linear Convergence Rate) Let $\{U[p]\}$ be a sequence generated by Algorithm 1 for a given nonzero tensor $A \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k}$. If $U[p] \to U^*$ with $U^*$ a nondegenerate KKT point of (8), then the whole sequence $\{U[p]\}$ converges $R$-linearly to $U^*$.

**Proof** By Theorem 2, the sequence $\{U[p]\}$ converges globally to $U^*$, which together with
\[ x^* := \text{Diag} \left( (U^{(s,1)})^T, \ldots, (U^{(s,k)})^T \right) \cdot A \]
is a nondegenerate critical point of the function $g$ on $U_{n,r}$. Hence for a sufficiently large $p$, Lemma 5 implies that
\[ \sum_{i=1}^k \|\nabla_i f(U[p]) - U^{(i)}[p]\nabla_i (f(U[p]))^TU^{(i)}[p]\|_F^2 \geq \mu |f(U[p]) - f(U^*)|. \]
On the other hand, by Lemma 9, we have
\[ \sum_{i=1}^{k} \| \nabla_i f(U[p]) - U^{(i)}[p] \nabla_i (f(U[p]))^T U^{(i)}[p] \|_F^2 \leq k \gamma^2 \| U[p] - U[p-1] \|_F^2. \]

Thus,
\[ f(U[p]) - f(U[p-1]) \geq \frac{\epsilon}{2} \| U[p] - U[p-1] \|_F^2 \]
\[ \geq \frac{\mu \epsilon}{2k \gamma^2} (f(U^*) - f(U[p])), \]
where the first inequality follows from Proposition 8, and the second follows from the preceding two inequalities and Proposition 10. Therefore, for a sufficiently large \( p \), we have
\[ f(U^*) - f(U[p]) \leq \frac{2k \gamma^2}{2k \gamma^2 + \mu \epsilon} (f(U^*) - f(U[p-1])), \tag{58} \]
which establishes the local \( Q \)-linear convergence of the sequence \( \{ f(U[p]) \} \). Consequently, we have
\[ \| U[p] - U[p-1] \|_F \leq \sqrt{\frac{2}{\epsilon} \sqrt{f(U[p]) - f(U[p-1])}} \]
\[ \leq \sqrt{\frac{2}{\epsilon} \sqrt{f(U^*) - f(U[p-1])}} \]
\[ \leq \sqrt{\frac{2}{\epsilon} \left( \frac{2k \gamma^2}{2k \gamma^2 + \mu \epsilon} \right)^{p-1} \sqrt{f(U^*) - f(U[0])}}. \]

As \( U[p] \to U^* \), we have
\[ \| U[p] - U^* \|_F \leq \sum_{s=p}^{\infty} \| U[s+1] - U[s] \|_F. \]

Hence, we obtain
\[ \| U[p] - U^* \|_F \leq \sqrt{\frac{2}{\epsilon} \sqrt{f(U^*) - f(U[0])}} \frac{1}{1 - \sqrt{\frac{2k \gamma^2}{2k \gamma^2 + \mu \epsilon}}} \left( \frac{2k \gamma^2}{2k \gamma^2 + \mu \epsilon} \right)^{p}, \]
which is the claimed \( R \)-linear convergence of the sequence \( \{ U[p] \} \) and this completes the proof. \( \square \)

By Theorems 1, 2 and 4, we eventually arrive at the third main result in our convergence analysis.
Theorem 5 (Generic Linear Convergence) If $\{U[p]\}$ is a sequence generated by Algorithm 1 for a generic tensor $A \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k}$, then the sequence $\{U[p]\}$ converges $R$-linearly to a KKT point of (9).

5 Discussions on proximal correction and truncation

In this section, we carry out a further study of proximal corrections and truncation iterations in Algorithm 1. We prove that if we make an appropriate assumption on the limiting point, then the truncation iteration is unnecessary and proximal corrections are only needed for finitely many times. Thus, our iAPD reduces to the classical APD proposed in [26] after finitely many iterations. Remarkably, the assumption on the whole iteration sequence (cf. [26, Assumption A]) is vastly relaxed to a requirement on the limiting point. Together with conclusions in Sect. 3 about KKT points, our results in this section can shed some light on the further understanding of APD and iAPD.

5.1 Proximal correction

In this subsection, we prove that in most situations, the proximal correction in Algorithm 1 is unnecessary. Before we proceed, we introduce the notion of regular KKT points.

Definition 3 (Regular KKT Point) A KKT point $U := (U^{(1)}, \ldots, U^{(k)}) \in \mathbb{R}^{n_1 \times r} \times \cdots \times \mathbb{R}^{n_k \times r}$ of (9) is called a regular KKT point if the matrix $V^{(i)} \Lambda$ (cf. (20) and (21)) is of rank at least $\min\{r, n_i - 1\}$ for each $1 \leq i \leq k$.

The requirement of a regular KKT point in Definition 3 is natural. The matrix $U^{(i)}$ is orthonormal and hence it is of full rank, for each $1 \leq i \leq k$. On the other hand, these matrices are polar orthonormal factor matrices of $V^{(i)} \Lambda$’s by Algorithm 1. If for some $1 \leq i \leq k$, the matrix $V^{(i)} \Lambda$ is of defective rank, then the best rank $r$ approximation of the $i$-th factor matrix is not unique. The case that $r = n_i$ is an exceptional case, see Lemma 21. With Lemma 21, we have a revised proximal correction step, which is described in Algorithm 2.

If we replace the Substep 2 in Algorithm 1 by the revised version described in Algorithm 2, then the sequence $\{U[p]\}$ still has the sufficient increasing property. This is the content of the following lemma.
Algorithm 2 Revised Proximal Step

\( \tau > \epsilon \) is a given constant.

Substep 2 [Revised Proximal Correction]: If \( \sigma_{r,\{p\}}^{(i)} < \epsilon \), then consider the following two cases.

(i) If \( r = n_i \) and \( \sigma_{r-1,\{p\}}^{(i)} \geq \tau \), then define a vector

\[
\hat{g}_r^{(i)} := \begin{cases} -g_r^{(i)} & \text{if } g_r^{(i)} \cdot (U_{\{p\}}^{(i)} H_{\{p\}}^{(i)})_r < 0 \\ g_r^{(i)} & \text{otherwise} \end{cases}
\]

where \( g_r^{(i)} \) is the \( r \)-th column of the matrix \( G_{\{p\}}^{(i)} \) and similar for \( (U_{\{p\}}^{(i)} H_{\{p\}}^{(i)})_r \). Form a matrix \( \hat{G}_{\{p\}}^{(i)} \) from \( G_{\{p\}}^{(i)} \) by replacing the last column with \( \hat{g}_r^{(i)} \). Let \( U_{\{p\}}^{(i)} := \hat{G}_{\{p\}}^{(i)} H_{\{p\}}^{(i)} \) and \( S_{\{p\}}^{(i)} := (U_{\{p\}}^{(i)})^T V_{\{p\}} A_{\{p\}}^{(i)} \).

(ii) For the other cases, compute the polar decomposition of the matrix \( V_{\{p\}} A_{\{p\}}^{(i)} + \epsilon U_{\{p\}}^{(i-1)} \) as

\[
V_{\{p\}} A_{\{p\}}^{(i)} + \epsilon U_{\{p\}}^{(i-1)} = \hat{G}_{\{p\}}^{(i)} \hat{\tilde{g}}^{(i)}
\]

for an orthonormal matrix \( \hat{G}_{\{p\}}^{(i)} \in V(r, n_i) \) and a symmetric positive semidefinite matrix \( \hat{\tilde{S}}_{\{p\}}^{(i)} \).

Update \( U_{\{p\}}^{(i)} := \hat{U}_{\{p\}}^{(i)} \) and \( S_{\{p\}}^{(i)} := \hat{\tilde{S}}_{\{p\}}^{(i)} \).

Set \( i := i + 1 \) and go to Substep 1.

Lemma 10 (Revised Version) Suppose that \( \tau > \epsilon > 0 \). For any \( p \in \mathbb{N} \) such that the \( p \)-th iteration is not a truncation iteration, we have

\[
f(U_{i+1,\{p\}}) - f(U_{i,\{p\}}) \geq \frac{1}{4} \min\{\epsilon, \tau - \epsilon\} \| U_{\{p\}}^{(i+1)} - U_{\{p\}}^{(i+1)} \|_F^2, \quad 0 \leq i \leq k-1. \quad (60)
\]

Moreover, the matrix \( S_{\{p\}}^{(i)} \) defined in (59) is symmetric.

Proof The matrix \( S_{\{p\}}^{(i)} \) is obviously symmetric by a direct calculation. For the increasing property, it is sufficient to prove the result for Case (i) in Algorithm 2. We suppose that at iteration \( p \) and \( i \), Case (i) of Algorithm 2 is executed. In this case, \( g_r^{(i)} \) is totally determined (up to sign) by the first \( r-1 \) columns of the matrix \( G_{\{p\}}^{(i)} \). It follows from Lemma 21 and the choice of \( \hat{g}_r^{(i)} \) that

\[
\| \hat{G}_{\{p\}}^{(i)} - (U_{\{p\}}^{(i)} H_{\{p\}}^{(i)})_r \| = \min\{\| g_r^{(i)} - (U_{\{p\}}^{(i)} H_{\{p\}}^{(i)})_r \|, \| g_r^{(i)} + (U_{\{p\}}^{(i)} H_{\{p\}}^{(i)})_r \| \}
\]

\[
\leq \| \hat{G}_{\{p\}}^{(i)} \|_{1:r-1} - (U_{\{p\}}^{(i)} H_{\{p\}}^{(i)})_r \|_F
\]

\[
= \| \hat{G}_{\{p\}}^{(i)} \|_{1:r-1} - (U_{\{p\}}^{(i)} H_{\{p\}}^{(i)})_r \|_F, \quad (61)
\]

where \((A)_{1:r-1}\) represents the \( n_i \times (r-1) \) submatrix formed by the first \( r-1 \) columns of a given \( n_i \times r \) matrix \( A \), and the last equality follows from the definition of \( \hat{G}_{\{p\}}^{(i)} \) and hence \( \hat{\tilde{G}}_{\{p\}}^{(i)} \).
We then have

\[
\sum_{j=1}^{\mathcal{r}} \lambda_{j,[p]}^{i-1} (\lambda_{j,[p]} - \lambda_{j,[p]}^i) = \text{Tr}((U_{[p]}^{(i)} V_{[p]}^{(i)} A_{[p]}^{(i)}) - \text{Tr}((U_{[p-1]}^{(i)} V_{[p]}^{(i)} A_{[p]}^{(i)})
\]

\[
= (G_{[p]}^{(i)} \Sigma_{[p]}^{(i)} (H_{[p]}^{(i)}), U_{[p]}^{(i)} - U_{[p-1]}^{(i)})
\]

\[
= (G_{[p]}^{(i)} \Sigma_{[p]}^{(i)}, \hat{G}_{[p]}^{(i)} - U_{[p-1]}^{(i)} H_{[p]}^{(i)})
\]

\[
\geq (G_{[p]}^{(i)} 1:r-1 \Sigma_{[p]}^{(i)}, (\hat{G}_{[p]}^{(i)} - U_{[p-1]}^{(i)} H_{[p]}^{(i)}) 1:r-1)
\]

\[
- \epsilon \| \hat{g}_r^{(i)} - (U_{[p-1]}^{(i)} H_{[p]}^{(i)})_r \|
\]

\[
\geq \frac{\tau}{2} \| (\hat{G}_{[p]}^{(i)}) 1:r-1 - (U_{[p-1]}^{(i)} H_{[p]}^{(i)}) 1:r-1 \|_F^2
\]

\[
- \frac{\epsilon}{2} \| \hat{g}_r^{(i)} - (U_{[p-1]}^{(i)} H_{[p]}^{(i)})_r \|^2
\]

\[
\geq \frac{\tau - \epsilon}{4} \| \hat{G}_{[p]}^{(i)} - U_{[p-1]}^{(i)} H_{[p]}^{(i)} \|_F^2
\]

\[
= \frac{\tau - \epsilon}{4} \| U_{[p]}^{(i)} - U_{[p-1]}^{(i)} \|_F^2,
\]

where \( \Sigma_{[p]}^{(i)} \) is the \((r-1) \times (r-1)\) leading principal submatrix of \( \Sigma_{[p]}^{(i)} \), the first inequality follows from \( \sigma_{r,[p]}^{(i)} < \epsilon \), the second inequality follows from \( \sigma_{r-1,[p]} \geq \tau \) and Theorem 9, the last two inequalities both follow from (61). With (62), the rest of the proof is the same as that of Lemma 6 and the conclusion follows.

With Lemma 10, all the convergence results established in Sect. 4 hold as well.

**Theorem 6 (Regular KKT point)** Let \( \mathcal{A} \) be a generic tensor. If \( \mathcal{U} \) is a local maximizer of problem (9) where each entry of \( \text{Diag}(\mathcal{Y}) \) is nonzero, then \( \mathcal{U}, \text{Diag}(\mathcal{Y}) \) is a nondegenerate critical point of \( g \) defined in (17) and \( \mathcal{U} \) is a regular KKT point of problem (9).

**Proof** If \( \text{Diag}(\mathcal{Y}) \) is a vector with all nonzero components, then we must have \( \mathcal{U}, \text{Diag}(\mathcal{Y})) \in U_{n,r} \). Moreover, Proposition 7 implies that \( \mathcal{U}, \text{Diag}(\mathcal{Y}) \) is a critical point of \( g \) and since \( g \) is a Morse function on \( U_{n,r} \) for a generic tensor \( \mathcal{A} \) by Proposition 4, it is actually a nondegenerate critical point. According to Lemma 12, we may conclude that \( \mathcal{U}, \text{Diag}(\mathcal{Y}) \) is isolated, i.e., \( g \) has no other nondegenerate critical point near \( \mathcal{U}, \text{Diag}(\mathcal{Y}) \).

In the following, we prove that the matrix \( V^{(i)} \mathcal{A} \) defined by (20) and (21) is of rank at least \( \min\{r, n_i - 1\} \) for all \( i = 1, \ldots, k \). Suppose on the contrary that there exists some \( i \in \{1, \ldots, k\} \) such that the matrix \( V^{(i)} \mathcal{A} \) has rank \( s < \min\{r, n_i - 1\} \). We consider the singular value decomposition of \( V^{(i)} \mathcal{A} \)

\[
V^{(i)} \mathcal{A} = U \Sigma V^T
\]
with $U := [U_1 \, U_2] \in V(r, n_i)$, $U_1 \in \mathbb{R}^{n_i \times s}$, $\Sigma \in \mathbb{R}^{s \times r}$, $V = [V_1 \, V_2] \in O(r)$ and $V_1 \in \mathbb{R}^{r \times s}$. Hence

$$V^{(i)} A = (U V^T)(V \Sigma V^T)$$

is a polar decomposition of $V^{(i)} A$ with the polar orthonormal factor matrix $U V^T$. Since the rank of $V^{(i)} A$ is $s < \min\{r, n_i - 1\}$, the last $(r - s)$ diagonal elements of $\Sigma$ are zero. Hence we may partition $\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$, where $\Sigma_1$ is some $s \times s$ diagonal matrix. This indicates that polar decomposition of $V^{(i)} A$ is not unique. The essential uniqueness of the singular value decomposition for $U_1 \Sigma_1 V_1^T = V^{(i)} A$ implies that the polar decomposition $V^{(i)} A = P(V \Sigma V^T)$ has the form

$$P = U_1 V_1^T + U_2 V_2^T = [U_1 \, U_2] [V_1 \, V_2]^T \in V(r, n_i)$$

whenever $[U_1 \, U_2] \in V(r, n_i)$.

Next we consider the set

$$C := \{W_2 \in \mathbb{R}^{n_i \times (r-s)}: [U_1 \, W_2] \in V(r, n_i)\}.$$

Since $U_1$ is a fixed element in $V(s, n_i)$, $C$ is isomorphic to $V(r - s, n_i - s)$. The assumption $\min\{r, n_i - 1\} > s$ implies $n_i - s \geq 2$ and thus $C \subseteq \mathbb{R}^{n_i \times (r-s)}$ contains an irreducible closed subvariety of dimension

$$\dim C = \frac{1}{2} (r - s) ((n_i - r) + (n_i - s - 1)) \geq 1.$$

Therefore, in any small neighborhood of $P$, there exists an orthonormal matrix $\tilde{P}$ such that $\tilde{P}$ also gives a polar orthonormal factor matrix of $V^{(i)} A$. Now if we fix the other $U^{(i)}$’s and $\Upsilon$, then

$$2g((U^{(1)}, \ldots, U^{(i-1)}, \tilde{P}, U^{(i+1)}, \ldots, U^{(k)}), \text{Diag}(\Upsilon)) = \|A\|_2^2 - 2 \langle V^{(i)} A, \tilde{P} \rangle + \|\Upsilon\|_2^2$$

$$= \|A\|_2^2 - 2 \langle V^{(i)} A, P \rangle + \|\Upsilon\|_2^2,$$

where the second equality follows from the fact that $\tilde{P}$ is also a polar orthonormal factor matrix of $V^{(i)} A$. Therefore, $g$ is constant at such a point

$$((U^{(1)}, \ldots, U^{(i-1)}, \tilde{P}, U^{(i+1)}, \ldots, U^{(k)}), \text{Diag}(\Upsilon)).$$

Since $(\bar{U}, \Upsilon)$ is a local minimizer of (8) (or equivalently, $\bar{U}$ is a local maximizer of (9)), we may conclude that such a point is also a local minimizer of (8). In particular, each such point corresponds to a critical point of $g$ on $U_{n,r}$ by Proposition 5 and Proposition 7. However, this contradicts to the fact that $(\bar{U}, \text{Diag}(\Upsilon))$ is an isolated critical point of $g$ on $U_{n,r}$.

$\square$
Corollary 3 For a generic tensor \( A \), there exist \( \tau > \epsilon > 0 \) such that if Substep 2 in Algorithm 1 is replaced by Algorithm 2, then the proximal step (i.e., Case (ii) in Algorithm 2) will only be executed finitely many times if the sequence generated by the algorithm converges to a local maximizer of (9).

Proof For a generic tensor, there are finitely many essential KKT points whose corresponding primitive KKT points are all nondegenerate by Theorem 1. Therefore, there exists a constant \( \tau > 0 \) such that it is strictly smaller than the smallest positive singular values of all \( V(i)A \)'s determined by these primitive KKT points.

We take \( 0 < \epsilon < \tau \) and let \( \{ U[p]\} \) be a sequence generated by the modified algorithm which converges to \( U^* \). Note that the convergence is guaranteed by Lemma 10 and results in Sect. 4. Let \( \Lambda^* \) be the limit of \( \Lambda[p] \) defined in (36) and \( V^{(i)} \) the limit of \( V[p] \) defined in (37). The truncation iteration ensures that \( \text{Diag}(\Lambda^*) \) is a vector with each component nonzero. Thus, by Theorem 6, the limit point \( \Lambda^* \) is a regular KKT point of (9). Consequently, the rank of the matrix \( V^{(i)} \Lambda^* \) is either of full rank and hence the proximal correction step will not be executed by the choice of \( \epsilon \) and \( \tau \), or the rank of the matrix \( V^{(i)} \Lambda^* \) is of rank \( (r-1) \) when \( r = n_i \), in which case Case (i) in Algorithm 2 will be executed by the choice of \( \epsilon \) and \( \tau \). Therefore, after finitely many iterations, Case (ii) in Algorithm 2 will not be executed.

\[ \square \]

5.2 Truncation

In this subsection, we prove that for almost all LROTA problems, local minimizers of (8) are actually contained in the manifold \( D(n, r) \). Therefore, if Algorithm 1 converges to a local minimizer of (8), we can choose a suitable \( \kappa > 0 \) such that the truncation step (i.e., Step 2) in Algorithm 1 is unnecessary.

Theorem 7 If the sequence \( n = (n_1, \ldots, n_k) \) and the positive integer \( r \leq \min\{n_1, \ldots, n_k\} \) satisfy the relation

\[ d_{n, r-1} < \prod_{i=1}^{k} (n_i - r + 1), \quad (63) \]

where \( d_{n, r-1} := (r-1) \left[ \sum_{i=1}^{k} n_i - \frac{kr}{2} + 1 \right] \), then for a generic \( A \in \mathbb{R}^{n_1 \times \cdots \times n_k} \), each local minimizer of problem (8) is contained in \( V(r, n_1) \times \cdots \times V(r, n_k) \times \mathbb{R}^r \).

The idea of the proof of Theorem 7 is to first construct a set \( Z \) consisting of tensors obtained from those in \( C(n, r-1) \) by adding tensors with “restricted” directions. In particular, we can prove by contradiction that for each \( A \notin Z \), any local minimizer must lie in \( D(n, r) \). Then we can estimate the dimension of \( Z \) and the assumption (63) is a sufficient condition to guarantee that dim \( Z \) is strictly smaller than the dimension \( \prod_{i=1}^{k} n_i \) of the ambient space \( \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_k} \).

Proof We consider the subset \( Z \subseteq \mathbb{R}^{n_1 \times \cdots \times n_k} \) consisting of tensors of the form

\[ (V^{(1)}, \ldots, V^{(k)}) \cdot \text{diag}(\mu_1, \ldots, \mu_{r-1}) + \mathcal{X}. \]
Here \( V^{(i)} \in V(r - 1, n_i) \), \( \mu_j \in \mathbb{R} \), and \( \mathcal{X} \) is a linear combination of decomposable tensors \( v_1 \otimes \cdots \otimes v_k \) where for each \( i \in \{1, \ldots, k\} \),

1. \( v_i \in \mathbb{R}^{n_i} \) is a unit norm vector;
2. if \( v_i \) is not a column vector of \( V^{(i)} \), then \( v_i^T V^{(i)} = 0 \);
3. there exists some \( 1 \leq j \leq k \) such that \( v_j \) is a column vector of \( V^{(j)} \).

We notice that

\[
(V^{(1)}, \ldots, V^{(k)}) \cdot \text{diag}(\mu_1, \ldots, \mu_{r-1}) \in C(n, r - 1)
\]

and \( \mathcal{X} \) is contained in a vector space of dimension at most

\[
\prod_{i=1}^{k} n_i - \prod_{i=1}^{k} (n_i - r + 1).
\]

This implies that the dimension of \( Z \) is bounded above by

\[
\dim C(n, r - 1) + \left( \prod_{i=1}^{k} n_i - \prod_{i=1}^{k} (n_i - r + 1) \right) = d_{n, r-1} + \left( \prod_{i=1}^{k} n_i - \prod_{i=1}^{k} (n_i - r + 1) \right),
\]

since \( \dim C(n, r - 1) = d_{n, r-1} \) by Proposition 2. In particular, we have

\[
\dim \overline{Z} = \dim Z < \prod_{i=1}^{k} n_i,
\]

where \( \overline{Z} \) is the Zariski closure of \( Z \). Next we suppose that \( A \in U := \mathbb{R}^{n_1 \times \cdots \times n_k} \setminus \overline{Z} \) and there exist \((V^{(1)}, \ldots, V^{(k)}) \in V(r - 1, n_1) \times \cdots \times V(r - 1, n_k) \) and \((\mu_1, \ldots, \mu_{r-1}) \in \mathbb{R}^{r-1} \) such that the point \((V^{(1)}, \ldots, V^{(k)}, (\mu_1, \ldots, \mu_{r-1}))\) is a local minimizer of (8).

We can write

\[
\mathcal{X}^\prime := A - (V^{(1)}, \ldots, V^{(k)}) \cdot \text{diag}(\mu_1, \ldots, \mu_{r-1})
\]

as a linear combination of decomposable tensors \( v_1 \otimes \cdots \otimes v_k \), such that

i. \( v_i \in \mathbb{R}^{n_i} \) is a unit norm vector;
ii. either of the following occurs:
   (a) for each \( i = 1, \ldots, k \), \( v_i \) is a column vector of \( V^{(i)} \);
   (b) for each \( i = 1, \ldots, k \), \( v_i^T V^{(i)} = 0 \).

According to the choice of \( A \), there exists \( v_1 \otimes \cdots \otimes v_k \) satisfying (i) and (iib) such that

\[
\langle \mathcal{X}^\prime, v_1 \otimes \cdots \otimes v_k \rangle > 0.
\]
Now we set
\[ Y := (V^{(1)}, \ldots, V^{(k)}) \cdot \text{diag}(\mu_1, \ldots, \mu_{r-1}) + \epsilon v_1 \otimes \cdots \otimes v_k \in D(\mathbf{n}, r), \]
for a sufficiently small positive number \( \epsilon \). It is straightforward to verify that
\[ \|A - Y\| = \|X - \epsilon v_1 \otimes \cdots \otimes v_k\| < \|X\|. \]
This contradicts the assumption that \((V^{(1)}, \ldots, V^{(k)}, (\mu_1, \ldots, \mu_{r-1}))\) is a local minimizer of problem (8).

As a special case, we suppose that \( n_1 = \cdots = n_k = n \) so (63) is written as
\[ (r - 1) \left( k(n - \frac{r}{2}) + 1 \right) < (n - r + 1)^k. \]
We set \( r - 1 = (1 - \alpha)n \) for \( \alpha \in [\frac{1}{n}, 1] \), hence we have
\[ (1 - \alpha)n \left( \frac{kn(1 + \alpha)}{2} + 1 - \frac{k}{2} \right) < \alpha^k n^k. \]
Therefore, to guarantee (63) in this case, it is sufficient to require
\[ \frac{2n^k - 2}{k} \alpha^k + 1 - 1 > 0. \]

**Corollary 4** If \( n_1 = \cdots = n_k = n \) and
\[ 1 \leq r \leq \left( 1 - \left( \frac{k}{2n^{k-2}} \right)^{\frac{1}{r}} \right) n + 1, \]
then for a generic \( A \in \mathbb{R}^{n \times \cdots \times n} \), any local minimizer of the problem (8) lies in \( D(\mathbf{n}, r) \) (or equivalently \( U_{\mathbf{n}, r} \)). In particular, for any fixed \( k \) and \( r \), there exists \( n_0 \) such that whenever \( n \geq n_0 \) and \( A \in \mathbb{R}^{n \times \cdots \times n} \) is generic, any local minimizer of problem (8) lies in \( D(\mathbf{n}, r) \).

**Proof** We observe that for any \((k/2n^{k-2})^{1/k} \leq \alpha \leq 1\), (64) and hence (63) is satisfied by \( n_1 = \cdots = n_k = n \) and
\[ r = (1 - \alpha)n + 1. \]
This implies that for
\[ 1 \leq r \leq \left( 1 - \left( \frac{k}{2n^{k-2}} \right)^{\frac{1}{r}} \right) n + 1, \]
any local minimizer of the problem (8) lies in $D(n, r)$ for a generic $A$. In particular, if $k \geq 3$ is a fixed integer, then

$$\lim_{n \to \infty} \left( 1 - \left( \frac{k}{2n^{k-2}} \right)^{\frac{1}{k}} \right)^{\frac{1}{k}} n + 1 = 1.$$ 

This implies that there exists some integer $n_0$ such that for any $n \geq n_0$ and a generic $A$, local minimizers of problem (8) must all lie in $D(n, r)$. □

If (63) is fulfilled, Theorem 7 implies that all local maximizers of (9) for a generic tensor are in $D(n, r)$. Recall from Theorem 1 that these local maximizers are finite. Thus, we have the next corollary.

**Corollary 5** Let $n_1, \ldots, n_k$ and $r$ be positive integers satisfying (63). For a generic tensor $A$, if Algorithm 1 converges to a local maximizer, then the truncation step in Algorithm 1 will not be executed when a suitable $\kappa$ is chosen.

Combining Corollaries 3 and 5, we have the following conclusion.

**Theorem 8** For almost all LROTA problems, there exist $\kappa$, $\epsilon$ and $\tau$ such that iAPD with Substep 2 being replaced by Algorithm 2 reduces to APD after finitely many iterations if it converges to a local minimizer.

### 6 Numerical illustration

In this section, we implement APD and iAPD with some simple illustrative examples. The purposes of this section are two folds: to numerically verify the convergent properties established in Theorems 2 and 5, and to provide a concrete example for the fact that iAPD reduces to APD in most cases, which is proved in Theorem 8. For the reader’s convenience, we summarize in the following list the specific purpose of each example.

- Example 1 shows that (i) the truncation is dominated by the truncation parameter $\kappa$. When it is chosen appropriately, iAPD reduces to APD. (ii) iAPD finds the same essential KKT point with or without truncation, which can be seen from Fig. 1 on the objective function values. (iii) iAPD exhibits linear convergence which can be seen from Fig. 2.
- Example 2 shows that iAPD is identical to APD for randomly generated examples, i.e., generic tensors.
- Example 3 strengthens the phenomena exhibited by Example 1 through thousands of designed randomly generated examples.

We also remark that for a generic tensor, there always exists a threshold for the truncation parameter such that iAPD performs a correct truncation, i.e., it converges to the unique primitive KKT point corresponding to an essential KKT point. This fact follows from Theorem 1 on finitely many primitive KKT points and Theorem 4 on the linear convergence.
The criterion for termination of both algorithms is that the optimality of problem (9) is within $10^{-6}$, i.e.,

$$\sum_{i=1}^{k} \| \text{grad}_{U(i)} f(U) \| \leq 10^{-6}.$$  

All the tests were conducted on a Lenovo laptop with 128GB RAM and 2.8GHz E-2276M CPU running 64bit Windows operation system. All codes were written in MATLAB.

**Example 1** We consider a tensor $A$ in $\mathbb{R}^{10} \otimes \mathbb{R}^{12} \otimes \mathbb{R}^{14} \otimes \mathbb{R}^{16}$ given by

$$A = \frac{1}{5} \sum_{i=1}^{6} \lambda_i a_i^{(1)} \otimes a_i^{(2)} \otimes a_i^{(3)} \otimes a_i^{(4)} + \epsilon B,$$

where $A^{(i)} = [a_{i1}^{(i)} \ldots a_{i6}^{(i)}] \in \mathbb{R}^{n_i \times 6}$ is a matrix with entries randomly chosen from $[0, 1]$ for $n_i = 10, 12, 14, 16, \lambda_i \in [0, 1]$ is randomly generated for $i = 1, 2, 3, 4, \lambda_5 = 0.01, \lambda_6 = 10^{-3}, \epsilon = 10^{-3}$ and $B \in \mathbb{R}^{10} \otimes \mathbb{R}^{12} \otimes \mathbb{R}^{14} \otimes \mathbb{R}^{16}$ a tensor with entries randomly chosen from $[0, 1]$. Here the choice of $\lambda_5$ and $\lambda_6$ is for the purpose of occurrence of truncation for some tested truncation parameters $\kappa$. The best rank seven orthogonal approximation is computed. The computation is implemented by Algorithm 1 (iAPD) with different values of $\kappa = 0.5, 0.1, 0.001, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}$. We also compare iAPD with APD, whose efficiency is exhibited by various numerical examples in the literature [13, 26, 65]. All computations are started from the same initial point, which is taken from the polar orthonormal factors of randomly generated matrices.

Results are presented in Figs. 1 and 2. Figure 1 depicts the objective function values along the iterations, and Fig. 2 shows the convergence of the iteration sequence $\{U[p]\}$ of iAPD. We notice that in Fig. 1 truncations indeed occur for $\kappa = 0.5, 0.1, 0.01$ (there are more “∗” than “⋄” in these cases), indicating that truncation steps are necessary in iAPD. We see that in all the tested cases, the algorithm finds points with almost the same objective function values, which are $12.0521$ for $\kappa = 0.5$ and $\kappa = 0.1$, $12.0550$ for $\kappa = 0.01$, and $12.0551$ for the others. We also see from Fig. 2 that the linear convergence is verified by this example. Another property verified by this example is that iAPD reduces to APD for $\kappa \leq 10^{-3}$.

**Example 2** We test our algorithms for tensors in $\mathbb{R}^{n} \otimes \mathbb{R}^{n} \otimes \mathbb{R}^{n}$ with $n \in \{10, \ldots, 20\}$. For each case, 100 simulations are generated. We compute the best rank $r = n - 2$ orthogonal approximation. Parameters are chosen as $\epsilon = \kappa = 10^{-12}$. We record in Table 1 the number of simulations for which iAPD identically coincides with APD, in the sense that with the same number of iterations, objective function values and approximation tensors obtained by both algorithms are differed by at most $10^{-6}$.  

\(\text{Springer}\)
Fig. 1 Objective function values along the iterations at several truncation parameters

Fig. 2 Convergence of the iterative sequence \( \{U_p\} \) of iAPD at several truncation parameters

Table 1 The numbers of identical iterations for iAPD and APD

| \( n \) | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| number | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Example 3 We test for orthogonally decomposable tensors in \( \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n \) with \( 10 \leq n \leq 20 \). The tensors are generated as follows with \( r = n - 2 \),

\[
A = \sum_{i=1}^{r} \lambda_i u_i \otimes v_i \otimes w_i
\]
Table 2 The numbers without truncation

| κ      | $10^{-7}$ | $10^{-8}$ | $10^{-9}$ | $10^{-10}$ | $10^{-11}$ | $10^{-12}$ | $10^{-13}$ | $10^{-14}$ |
|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| num    | 346       | 359       | 338       | 351       | 797       | 1024      | 1087      | 1099      |

Fig. 3 Statistics of the approximation quality by iAPD and APD at several truncation parameters

where $[u_1 \ldots u_r], [v_1 \ldots v_r], [w_1 \ldots w_r] \in V(r, n)$ are the polar orthonormal factor matrices of randomly generated matrices in $\mathbb{R}^{n \times r}$, $\lambda_1, \ldots, \lambda_{r-3} \in [0, 1]$ are randomly generated and $\lambda_{r-2} = \lambda_{r-1} = \lambda_r = 10^{-9}$. Truncation parameters are chosen to be $\kappa = 10^{-7}, 10^{-8}, 10^{-9}, 10^{-10}, 10^{-11}, 10^{-12}, 10^{-13}, 10^{-14}$. For each pair $(n, \kappa)$, 100 simulations are generated and the total number of simulations is 8800. We compute the best rank $r = n - 2$ orthogonal approximation with $\epsilon = 10^{-12}$ and results are presented in Table 2 and Fig. 3.

In Table 2, the number of simulations for which no truncation occurs is listed for each value of $\kappa$. Figure 3 displays the number of simulations (y-axis) for which the difference between the computed approximation tensors obtained by iAPD and APD (labelled as “difference”), the difference between the computed approximation tensor obtained by iAPD and the ground truth $\mathcal{A}$ (labelled as “iAPD-app”), and the difference between the computed approximation tensor obtained by APD and the ground truth $\mathcal{A}$ (labelled as “APD-app”) within the residual level (x-axis). Here the residual level is labelled by $i = 1, \ldots, 12$ which corresponds to the number $\epsilon_i$ defined by

$$
\epsilon_i = \begin{cases} 
10^{-14+i} & \text{if } 1 \leq i \leq 10, \\
5 \times 10^{-3} & \text{if } i = 11, \\
10^{-3} & \text{if } i = 12.
\end{cases}
$$
7 Conclusions

In this paper, we propose an improved alternating polar decomposition algorithm (iAPD) with adaptive proximal correction and truncation for approximating a given tensor by a low rank orthogonally decomposable tensor. Without any assumption we prove that this algorithm has global convergence and overall sublinear convergence with an explicit convergence rate. For a generic tensor, this algorithm converges $R$-linearly without any further assumption. For the first time, the convergence rate analysis for the problem of low rank orthogonal tensor approximations is accomplished. We also prove that for generic $A$, truncation step is not required and only finitely many proximal corrections are needed and after that, our iAPD reduces to the well-known APD algorithm if it converges to a local minimizer.

Acknowledgements We would like to acknowledge the anonymous associate editor and two referees for their thoughtful suggestions and helpful comments on our manuscript.

A Preliminaries

This section consists of preliminaries on differential geometry and optimization theory, which are necessary to get through the details of proofs in this paper. To be more precise, we discuss Stiefel manifolds and their basic properties in Appendix A.1. Morse functions and Kurdyka-Łojasiewicz property are introduced in Appendix A.2 and A.3, respectively.

A.1 Stiefel manifolds

Let $m \leq n$ be two positive integers and let $V(m, n) \subseteq \mathbb{R}^{n \times m}$ be the set of all $n \times m$ orthonormal matrices, i.e.,

$$V(m, n) := \{ U \in \mathbb{R}^{n \times m} : U^T U = I \},$$

where $I$ is the identity matrix of matching size. Indeed, $V(m, n)$ admits a smooth manifold structure and is called the Stiefel manifold of orthonormal $m$-frames in $\mathbb{R}^n$. In particular, if $m = n$ then $V(n, n)$ simply reduces to the orthogonal group $O(n)$.

In general, if $M$ is a Riemannian submanifold of $\mathbb{R}^k$, then the normal space $N_M(x)$ of $M$ at a point $x \in M$ is defined to be the orthogonal complement of its tangent space $T_M(x)$ in $\mathbb{R}^k$, i.e., $N_M(x) := T_M(x)^\perp$. By definition, $V(m, n)$ is naturally a Riemannian submanifold of $\mathbb{R}^{n \times m}$ and hence its normal space is well-defined. Indeed, it follows from [56, Chapter 6.C] and [1, 22] that

$$N_{V(m,n)}(A) = \{ AX : X \in S^{m \times m} \},$$  \hspace{1cm} (65)

where $S^{m \times m} \subseteq \mathbb{R}^{m \times m}$ is the subspace of $m \times m$ symmetric matrices.
Given a matrix \( B \in \mathbb{R}^{n \times m} \), the projection of \( B \) onto the normal space of \( V(m, n) \) at a point \( A \in V(m, n) \) is
\[
\pi_{N_{V(m,n)}}(A)(B) = A \left( \frac{A^T B + B^T A}{2} \right).
\]
Likewise, given a matrix \( B \in \mathbb{R}^{n \times m} \), the projection of \( B \) onto the tangent space of \( V(m, n) \) at \( A \) is given by
\[
\pi_{T_{V(m,n)}}(A)(B) = A \text{skew}(A^T B) + (I - AA^T)B,
\]
where \( \text{skew}(C) := \frac{C - C^T}{2} \) for a square matrix \( C \in \mathbb{R}^{m \times m} \). A more explicit formula is given as
\[
\pi_{T_{V(m,n)}}(A)(B) = (I - \frac{1}{2} AA^T)(B - AB^TA).
\]
Given a function \( f : \mathbb{R}^k \to \mathbb{R} \cup \{ \infty \} \), the subdifferential of \( f \) at \( x \in \mathbb{R}^k \) is defined as (cf. [56])
\[
\partial f(x) := \{ v \in \mathbb{R}^k : \liminf_{x \neq y \to x} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|y - x\|} \geq 0 \}.
\]
If \( 0 \in \partial f(x) \), then \( x \) is a critical point of \( f \).

The indicator function \( \delta_{V(m,n)} : \mathbb{R}^{n \times m} \to \mathbb{R} \cup \{ \infty \} \) of \( V(m, n) \) is defined by
\[
\delta_{V(m,n)}(X) := \begin{cases} 
0 & \text{if } X \in V(m, n), \\
+\infty & \text{otherwise}.
\end{cases}
\]
An important fact about the normal space \( N_{V(m,n)}(A) \) and the subdifferential of the indicator function \( \delta_{V(m,n)} \) of \( V(m, n) \) at \( A \) is (cf. [56])
\[
\partial \delta_{V(m,n)}(A) = N_{V(m,n)}(A).
\]

A.2 Morse functions

In the following, we introduce the notion of Morse functions and recall some of its basic properties. On a smooth manifold \( M \), a smooth function \( f : M \to \mathbb{R} \) is called a Morse function if each critical point of \( f \) on \( M \) is nondegenerate, i.e., the Hessian matrix of \( f \) at each critical point is non-singular. The following result is well-known, see for example [51, Theorem 6.6].

Lemma 11 (Projection is Generically Morse) Let \( M \) be a submanifold of \( \mathbb{R}^n \). For a generic \( a = (a_1, \ldots, a_n)^T \in \mathbb{R}^n \), the Euclidean distance function
\[
f(x) = \|x - a\|^2
\]
is a Morse function on $M$.

We also need the following property (cf. [51, Corollary 2.3]) of nondegenerate critical points in the sequel.

**Lemma 12** Let $M$ be a manifold and let $f : M \to \mathbb{R}$ be a smooth function. Nondegenerate critical points of $f$ are isolated.

To conclude this subsection, we briefly discuss how critical points behave under a local diffeomorphism. For this purpose, we recall that a smooth manifold $M_1$ is called **locally diffeomorphic** to a smooth manifold $M_2$ if there is a smooth map $\varphi : M_1 \to M_2$ such that for each point $x \in M_1$ there exists a neighborhood $U \subseteq M_1$ of $x$ and a neighborhood $V \subseteq M_2$ of $\varphi(x)$ such that the restriction $\varphi|_U : U \to V$ is a diffeomorphism [20]. In this case, the corresponding $\varphi$ is called a **local diffeomorphism** from $M_1$ to $M_2$. It is clear from the definition that whenever $M_1$ is locally diffeomorphic to $M_2$ then the two manifolds must have the same dimension. Moreover, we have the following result, whose proof can be found in [31, Proposition 5.2].

**Proposition 11** Let a smooth manifold $M_1$ be locally diffeomorphic to another smooth manifold $M_2$ and let $\varphi : M_1 \to M_2$ be the corresponding local diffeomorphism. Let $f : M_2 \to \mathbb{R}$ be a smooth function. Then $x \in M_1$ is a (nondegenerate) critical point of $f \circ \varphi$ on $M_1$ if and only if $\varphi(x)$ is a (nondegenerate) critical point of $f$ on $M_2$.

When $f$ is a smooth function on $\mathbb{R}^n$ and $M$ is a submanifold of $\mathbb{R}^n$, we denote by $\nabla f$ the gradient of $f$ as a vector field on $\mathbb{R}^n$, while we denote by grad$(f)$ the Riemannian gradient of $f$ as a vector field on $M$.

### A.3 Kurdyka-Łojasiewicz property

In this subsection, we review some basic facts about the Kurdyka-Łojasiewicz property, which was first established for a smooth definable function in [38] by Kurdyka and later was even generalized to non-smooth definable case in [9] by Bolte, Daniilidis, Lewis and Shiota. Interested readers are also referred to [4, 5, 10, 42] for more discussions and applications of the Kurdyka-Łojasiewicz property.

Let $p$ be an extended real-valued function and let $\partial p(x)$ be the set of subgradients of $p$ at $x$ (cf. [56]). We define $\text{dom}(\partial p) := \{ x : \partial p(x) \neq \emptyset \}$ and take $x^* \in \text{dom}(\partial p)$. If there exist some $\eta \in (0, +\infty)$, a neighborhood $U$ of $x^*$, and a continuous concave function $\varphi : [0, \eta) \to \mathbb{R}_+$, such that

1. $\varphi(0) = 0$ and $\varphi$ is continuously differentiable on $(0, \eta)$,
2. for all $s \in (0, \eta)$, $\varphi'(s) > 0$, and
3. for all $x \in U \cap \{ y : p(x^*) < p(y) < p(x^*) + \eta \}$, the Kurdyka-Łojasiewicz inequality holds

$$\varphi'(p(x) - p(x^*)) \text{dist}(0, \partial p(x)) \geq 1,$$
then we say that $p$ has the Kurdyka-Łojasiewicz (abbreviated as KL) property at $x^*$. Here $\text{dist}(0, \partial p(x))$ denotes the distance from $0$ to the set $\partial p(x)$.

If $p$ is proper, lower semicontinuous, and has the KL property at each point of $\text{dom}(\partial p)$, then $p$ is said to be a KL function. Examples of KL functions include real subanalytic functions and semi-algebraic functions [11]. In this paper, semi-algebraic functions are involved. We refer to [8] and references herein for more details on such functions. In particular, polynomial functions are semi-algebraic functions and hence KL functions. Another important fact is that the indicator function of a semi-algebraic set is also a semi-algebraic function [8, 11]. Also, a finite sum of semi-algebraic functions is again semi-algebraic. We assemble these facts to derive the following lemma which is crucial to the analysis of the global convergence of our algorithm.

Lemma 13 A finite sum of polynomial functions and indicator functions of semi-algebraic sets is a KL function.

While KL-property is used for global convergence analysis, the Łojasiewicz inequality discussed in the rest of this subsection is for convergence rate analysis. The classical Łojasiewicz inequality for analytic functions is stated as follows (cf. [48]):

(Classical Łojasiewicz’s Gradient Inequality) If $f$ is an analytic function with $f(0) = 0$ and $\nabla f(0) = 0$, then there exist positive constants $\mu, \kappa$, and $\epsilon$ such that

$$\|\nabla f(x)\| \geq \mu |f(x)|^\kappa$$

for all $\|x\| \leq \epsilon$. (70)

As pointed out in [4, 10], it is often difficult to determine the corresponding exponent $\kappa$ in Łojasiewicz’s gradient inequality, and it is unknown for a general function. However, an estimate of the exponent $\kappa$ in the gradient inequality is obtained by D’Acunto and Kurdyka in [17, Theorem 4.2] when $f$ is a polynomial function. We record this fundamental result in the next lemma, which plays a key role in our sublinear convergence rate analysis.

Lemma 14 (Łojasiewicz’s Gradient Inequality for Polynomials) Let $f$ be a real polynomial of degree $d$ in $n$ variables. Suppose that $f(0) = 0$ and $\nabla f(0) = 0$. There exist constants $\mu, \epsilon > 0$ such that for all $\|x\| \leq \epsilon$, we have

$$\|\nabla f(x)\| \geq \mu |f(x)|^\kappa$$

with $\kappa = 1 - \frac{1}{d(3d - 3)^n - 1}$.

Below is a more precise version of the classical Łojasiewicz gradient inequality (70), in which we can take the exponent $\kappa$ to be $1/2$ near a nondegenerate critical point. For the sake of subsequent analysis, it is stated for a smooth manifold [20].

---

4 In the general definition of KL property, $\partial p(x)$ is actually the limiting subdifferential of $p$ at $x$. However, for functions considered in this paper, the two notions of subdifferentials coincide. For this reason, we do not distinguish them in the paper.
Proposition 12 (Łojasiewicz’s Gradient Inequality) Let $M$ be a smooth manifold and let $f : M \to \mathbb{R}$ be a smooth function for which $z^*$ is a nondegenerate critical point. Then there exist a neighborhood $U$ in $M$ of $z^*$ and some $\mu > 0$ such that for all $z \in U$

$$\| \text{grad}(f)(z) \| \geq \mu | f(z) - f(z^*) |^{\frac{1}{2}}.$$ 

B Properties on orthonormal matrices

B.1 Polar decomposition

In this section, we establish an error bound analysis for the polar decomposition. For a positive semidefinite matrix $H \in S_n^+$, there exists a unique positive semidefinite matrix $P \in S_n^+$ such that $P^2 = H$. In the literature, this matrix $P$ is called the square root of the matrix $H$ and denoted as $P = \sqrt{H}$. If $H = U \Sigma U^T$ is the eigenvalue decomposition of $H$, then we have $\sqrt{H} = U \sqrt{\Sigma} U^T$, where $\sqrt{\Sigma}$ is the diagonal matrix whose diagonal elements are square roots of those of $\Sigma$. The next result is classical, which can be found in [24, 29].

Lemma 15 (Polar Decomposition) Let $A \in \mathbb{R}^{m \times n}$ with $m \geq n$. Then there exist an orthonormal matrix $U \in V(n, m)$ and a unique symmetric positive semidefinite matrix $H \in S_n^+$ such that $A = UH$ and

$$U \in \arg\max \{ \langle Q, A \rangle : Q \in V(n, m) \}. \quad (71)$$

Moreover, if $A$ is of full rank, then the matrix $U$ is uniquely determined and $H$ is positive definite.

The matrix decomposition $A = UH$ as in Lemma 15 is called the polar decomposition of the matrix $A$ [24]. For convenience, the matrix $U$ is referred as a polar orthonormal factor matrix and the matrix $H$ is the polar positive semidefinite factor matrix. The optimization reformulation (71) comes from the approximation problem

$$\min_{Q \in V(n, m)} \| B - QC \|^2_F$$

for two given matrices $B$ and $C$ of proper sizes. In the following, we give a global error bound for this problem. To this end, the next lemma is useful.

Lemma 16 (Error Reformulation) Let $p, m, n$ be positive integers with $m \geq n$ and let $B \in \mathbb{R}^{m \times p}$, $C \in \mathbb{R}^{n \times p}$ be two given matrices. We set $A := BC^T \in \mathbb{R}^{m \times n}$ and suppose that $A = WH$ is a polar decomposition of $A$. We have

$$\| B - QC \|^2_F - \| B - WC \|^2_F = \| W\sqrt{H} - Q\sqrt{H} \|^2_F \quad (72)$$

for any orthonormal matrix $Q \in V(n, m)$. 

Springer
Proof We have
\[
\| B - QC \|_F^2 - \| B - WC \|_F^2 = 2 \langle B, WC - QC \rangle \\
= 2 \langle A, W - Q \rangle \\
= 2 \langle WH, W - Q \rangle \\
= \langle WH, W \rangle - 2 \langle WH, Q \rangle + \langle QH, Q \rangle \\
= \| W\sqrt{H} - Q\sqrt{H} \|_F^2,
\]
where both the first and the fourth equalities follow from the fact that both $Q$ and $W$ are in $V(n, m)$, and the last one is derived from the fact that $H$ is symmetric and positive semidefinite by Lemma 15.

Given an $n \times n$ symmetric positive semidefinite matrix $H$, we can define a symmetric bilinear form on $\mathbb{R}^{m \times n}$ by
\[
\langle P, Q \rangle_H := \langle PH, Q \rangle
\]
for all $m \times n$ matrices $P$ and $Q$. It can also induce a seminorm
\[
\| A \|_H := \sqrt{\langle A, A \rangle_H} = \| A\sqrt{H} \|_F.
\]
In particular, if $H$ is positive definite, then $\| \cdot \|_H$ is a norm on $\mathbb{R}^{m \times n}$. Thus, the error estimation in (72) can be viewed as a distance estimation between $W$ and $Q$ with respect to the distance induced by this norm. Moreover, if $H$ is the identity matrix, then $\| \cdot \|_H$ is simply the Frobenius norm which induces the Euclidean distance on $\mathbb{R}^{m \times n}$. By Lemma 16 it is easy to see that the optimizer in (71) is unique whenever $A$ is of full rank.

The following result establishes the error estimation with respect to the Euclidean distance. Given a matrix $A \in \mathbb{R}^{m \times n}$, let $\sigma_{\min}(A)$ be the smallest singular value of $A$. If $A$ is of full rank, then $\sigma_{\min}(A) > 0$.

**Theorem 9** (Global Error Bound in Frobenius Norm) Let $p, m, n$ be positive integers with $m \geq n$ and let $B \in \mathbb{R}^{m \times p}$ and $C \in \mathbb{R}^{n \times p}$ be two given matrices. We set $A := BC^T \in \mathbb{R}^{m \times n}$ and suppose that $A$ is of full rank with the polar decomposition $A = WH$. We have that
\[
\| B - QC \|_F^2 - \| B - WC \|_F^2 \geq \sigma_{\min}(A) \| W - Q \|_F^2
\]
for any orthonormal matrix $Q \in V(n, m)$.

Proof We know that in this case
\[
\sqrt{H} - \sqrt{\sigma_{\min}(A)} I \in S_+^n.
\]
Therefore we may conclude that
\[
\| W\sqrt{H} - Q\sqrt{H} \|_F^2 = \| W - Q \|_\sqrt{H}^2 \geq \| W - Q \|_\sqrt{\sigma_{\min}(A)} I^2
\]
\[
\|W\sqrt{\sigma_{\min}(A)}I - Q\sqrt{\sigma_{\min}(A)}I\|_F^2 = \sigma_{\min}(A)\|W - Q\|_F^2.
\]

According to Lemma 16, we can derive the desired inequality. \qed

Theorem 9 is a refinement of Sun and Chen’s result (cf. [59, Theorem 4.1]), in which the right hand side of (75) has an extra factor \(\frac{1}{4}\).

### B.2 Principal angles between subspaces

Given two linear subspaces \(U\) and \(V\) of dimension \(r\) in \(\mathbb{R}^n\), the principal angles \(\{\theta_i : i = 1, \ldots, r\}\) between \(U\) and \(V\) and the associated principal vectors \(\{(u_i, v_i) : i = 1, \ldots, r\}\) are defined recursively by

\[
\cos(\theta_i) = \langle u_i, v_i \rangle = \max_{u \in U} \max_{[u, u_1, \ldots, u_{i-1}] \in V(i,n)} \max_{v \in V} \max_{[v, v_1, \ldots, v_{i-1}] \in V(i,n)} \langle u, v \rangle.
\]

The following result is standard, whose proof can be found in [24, Section 6.4.3].

**Lemma 17** For any orthonormal matrices \(U, V \in V(r, n)\) of which subspaces spanned by column vectors are \(U\) and \(V\) respectively, we have

\[
\sigma_i(U^T V) = \cos(\theta_i) \text{ for all } i = 1, \ldots, r,
\]

where \(\theta_i\)'s are the principal angles between \(U\) and \(V\), and \(\sigma_i(U^T V)\) is the \(i\)-th largest singular value of the matrix \(U^T V\).

**Lemma 18** For any orthonormal matrices \(U, V \in V(r, n)\), we have

\[
\langle U, V \rangle \leq \sum_{j=1}^{r} \sigma_j(U^T V).
\]

**Proof** We recall from [24, pp. 331] that

\[
\min_{Q \in O(r)} \|A - BQ\|_F^2 = \sum_{j=1}^{r} (\sigma_j(A))^2 - 2 \sigma_j(B^T A) + \sigma_j(B)^2,
\]

for any \(n \times r\) matrices \(A, B\). In particular, if \(U, V \in V(r, n)\) then

\[
\sigma_j(U) = \sigma_j(V) = 1 \text{ for all } j = 1, \ldots, r, \quad \|U\|_F^2 = \|V\|_F^2 = r.
\]

This implies that

\[
2r - 2 \sum_{j=1}^{r} \sigma_j(U^T V) = \min_{Q \in O(r)} \|U - VQ\|_F^2 \leq \|U - V\|_F^2 = 2r - 2\langle U, V \rangle,
\]

and the desired inequality follows immediately. \qed

\(\Box\) Springer
Lemma 19  For any orthonormal matrices \( U, V \in V(r, n) \), we have
\[
\| U^T V - I \|_F^2 \leq \| U - V \|_F^2.
\] (79)

Proof  We have
\[
\| U^T V - I \|_F^2 = r + \sum_{i=1}^{r} \cos^2 \theta_i - 2 \text{tr}(U^T V) \leq 2r - 2 \text{tr}(U^T V) = \| U - V \|_F^2,
\]
where the first equality follows from Lemma 17. \( \square \)

Let \( U^\perp \) be the orthogonal complement subspace of a given linear subspace \( U \) in \( \mathbb{R}^n \). A useful fact about principal angles between two linear subspaces \( U \) and \( V \) and those between \( U^\perp \) and \( V^\perp \) is stated as follows. The proof can be found in [33, Theorem 2.7].

Lemma 20  Let \( U \) and \( V \) be two linear subspaces of the same dimension and let \( \frac{\pi}{2} \geq \theta_s \geq \cdots \geq \theta_1 > 0 \) be the nonzero principal angles between \( U \) and \( V \). Then the nonzero principal angles between \( U^\perp \) and \( V^\perp \) are \( \frac{\pi}{2} \geq \theta_s \geq \cdots \geq \theta_1 > 0 \).

The following result is for the general case, which might be of independent interests.

Lemma 21  Let \( m \geq n \) be positive integers and let \( V := [V_1 \ V_2] \in O(m) \) with \( V_1 \in V(n, m) \) and \( U \in V(n, m) \) be two given orthonormal matrices. Then, there exists an orthonormal matrix \( W \in V(m - n, m) \) such that \( P := [U \ W] \in O(m) \) and
\[
\| P - V \|_F^2 \leq 2 \| U - V_1 \|_F^2.
\] (80)

Proof  By a simple computation, it is straightforward to verify that (80) is equivalent to
\[
\| W - V_2 \|_F^2 \leq \| U - V_1 \|_F^2.
\] (81)
To that end, we let \( U_2 \in V(m - n, m) \) be an orthonormal matrix such that \( [U_2 \ V_2] \in O(m) \). Then, we have that the linear subspace \( U_2^\perp \) spanned by column vectors of \( U_2 \) is the orthogonal complement of \( U_1^\perp \), which is spanned by column vectors of \( U \). Likewise, let \( V_1^\perp \) and \( V_2^\perp \) be linear subspaces spanned by column vectors of \( V_1 \) and \( V_2 \) respectively.

Let \( \frac{\pi}{2} \geq \theta_s \geq \cdots \geq \theta_1 > 0 \) be the nonzero principal angles between \( U_2 \) and \( V_2 \) for some nonnegative integer \( s \leq m - n \). We have by Lemmas 17, and 18 that
\[
\langle U_2, V_2 \rangle \leq \sum_{i=1}^{m-n} \sigma_i(U_2^T V_2) = \sum_{i=1}^{s} \cos(\theta_i) + (m - n) - s.
\] (82)
Let \( Q \in O(m - n) \) be a polar orthogonal factor matrix of the matrix \( U_2^T V_2 \). It follows from the property of polar decomposition that
\[
\langle U_2 Q, V_2 \rangle = \sum_{i=1}^{m-n} \sigma_i(U_2^T V_2).
\] (83)
On the other hand, nonzero principal angles between $U_1$ and $V_1$ are $\frac{\pi}{2} \geq \theta_s \geq \cdots \geq \theta_1 > 0$ by Lemma 20. Therefore, by Lemmas 17 and 18, we have that

$$
\langle U, V_1 \rangle \leq \sum_{i=1}^{n} \sigma_i(U^T V_1) = \sum_{i=1}^{s} \cos(\theta_i) + n - s. \quad (84)
$$

In a conclusion, if we set $W := U_2 Q$, then we have the following:

$$
\|W - V_2\|_F^2 = 2(m - n) - 2\langle U_2 Q, V_2 \rangle
$$

$$
= 2(m - n) - 2\left(\sum_{i=1}^{s} \cos(\theta_i) + m - n - s\right)
$$

$$
= 2n - 2\left(\sum_{i=1}^{s} \cos(\theta_i) + n - s\right)
$$

$$
\leq 2n - 2\langle U, V_1 \rangle
$$

$$
= \|U - V_1\|_F^2,
$$

where the second equality follows from (82) and (83) and the inequality follows from (84).

\[\square\]

C Proofs of technical lemmas in section 4

C.1 Proof of Lemma 6

Proof For each $i \in \{0, \ldots, k - 1\}$, we have (cf. (23), (36) and (41))

$$
f(U_i+1,[p]) - f(U_i,[p]) = \sum_{j=1}^{r} (\lambda_{j,[p]}^{i+1})^2 - \sum_{j=1}^{r} (\lambda_{j,[p]}^{i})^2
$$

$$
= \sum_{j=1}^{r} (\lambda_{j,[p]}^{i+1} + \lambda_{j,[p]}^{i})(\lambda_{j,[p]}^{i+1} - \lambda_{j,[p]}^{i})
$$

$$
= \sum_{j=1}^{r} \lambda_{j,[p]}^{i+1}(\lambda_{j,[p]}^{i+1} - \lambda_{j,[p]}^{i}) + \sum_{j=1}^{r} \lambda_{j,[p]}^{i}(\lambda_{j,[p]}^{i+1} - \lambda_{j,[p]}^{i}). \quad (85)
$$

We first analyze the second summand in (85) and by considering the following two cases:
1. If $\sigma_{r,[p]}^{(i+1)} \geq \epsilon$, then there is no proximal step in Algorithm 1 and we have that $\sigma_{\min}(S_{[p]}^{(i+1)}) = \sigma_{r,[p]}^{(i+1)} \geq \epsilon$, where $S_{[p]}^{(i+1)}$ is the polar positive semidefinite factor matrix of $V_{[p]}^{(i+1)} A_{[p]}^{(i+1)}$ obtained in (39). From (35), (36) and (37) we notice that

\[
((U_{[p]}^{(i+1)})^T V_{[p]}^{(i+1)} A_{[p]}^{(i+1)})_{jj} = ((u_{l,j,[p]}^{(i+1)})^T v_{l,j,[p]}^{(i+1)})\lambda_{j,j,[p]}^{(i+1)} = ((u_{l,j,[p]}^{(i+1)})^T \Lambda_{j,j,[p]}^{(i+1)})\lambda_{j,j,[p]}^{(i+1)}
\]

and similarly $((U_{[p-1]}^{(i+1)})^T V_{[p]}^{(i+1)} A_{[p]}^{(i+1)})_{jj} = \lambda_{j,j,[p]}^{(i+1)}$. Hence by Lemma 16, we obtain

\[
\sum_{j=1}^{r} \lambda_{j,j,[p]}^{(i+1)} = \text{Tr}((U_{[p]}^{(i+1)})^T V_{[p]}^{(i+1)} A_{[p]}^{(i+1)}) = \text{Tr}((U_{[p-1]}^{(i+1)})^T V_{[p]}^{(i+1)} A_{[p]}^{(i+1)})
\]

\[
= \frac{1}{2} \left\| (U_{[p]}^{(i+1)} - U_{[p-1]}^{(i+1)}) \sqrt{S_{[p]}^{(i+1)}} \right\|_F^2
\]

\[
\geq \frac{\epsilon}{2} \left\| U_{[p]}^{(i+1)} - U_{[p-1]}^{(i+1)} \right\|_F^2
\]

\[
\geq 0.
\]

(86)

2. If $\sigma_{r,[p]}^{(i+1)} < \epsilon$, we consider the following matrix optimization problem

\[
\text{max} \quad \langle V_{[p]}^{(i+1)} A_{[p]}^{(i+1)}, U \rangle - \frac{\epsilon}{2} \left\| U - U_{[p-1]}^{(i+1)} \right\|_F^2
\]

s.t. $U \in V(r, n_{i+1})$.

(87)

Since $U, U_{[p-1]}^{(i+1)} \in V(r, n_{i+1})$, we must have

\[
\frac{\epsilon}{2} \left\| U - U_{[p-1]}^{(i+1)} \right\|_F^2 = \epsilon r - \epsilon \langle U_{[p-1]}^{(i+1)}, U \rangle.
\]

Thus, by Lemma 15, a global maximizer of (87) is given by a polar orthonormal factor matrix of the matrix $V_{[p]}^{(i+1)} A_{[p]}^{(i+1)} + \epsilon U_{[p-1]}^{(i+1)}$. By Step 2 of Algorithm 1, $U_{[p]}^{(i+1)}$ is a polar orthonormal factor matrix of the matrix $V_{[p]}^{(i+1)} A_{[p]}^{(i+1)} + \epsilon U_{[p-1]}^{(i+1)}$, and hence a global maximizer of (87). Thus, by the optimality of $U_{[p]}^{(i+1)}$ for (87), we have

\[
\langle V_{[p]}^{(i+1)} A_{[p]}^{(i+1)}, U_{[p]}^{(i+1)} \rangle - \frac{\epsilon}{2} \left\| U_{[p]}^{(i+1)} - U_{[p-1]}^{(i+1)} \right\|_F^2 \geq \langle V_{[p]}^{(i+1)} A_{[p]}^{(i+1)}, U_{[p-1]}^{(i+1)} \rangle.
\]

Therefore, the inequality (86) in case (1) also holds in this case.

Consequently, we have

\[
0 \leq \sum_{j=1}^{r} \lambda_{j,j,[p]}^{i+1} \lambda_{j,j,[p]}^{i} - \sum_{j=1}^{r} \lambda_{j,j,[p]}^{i+1} \lambda_{j,j,[p]}^{i} - \sum_{j=1}^{r} (\lambda_{j,j,[p]}^{i})^2.
\]

(88)
which together with Cauchy-Schwartz inequality implies that

\[
(\sum_{j=1}^{r} \lambda_{j,[p]}^i)^2 \leq (\sum_{j=1}^{r} \lambda_{j,[p]}^{i+1})^2 \leq \sum_{j=1}^{r} (\lambda_{j,[p]}^i)^2 \sum_{j=1}^{r} (\lambda_{j,[p]}^{i+1})^2.
\]  

(89)

Since \( f(U_{[0]}) > 0 \), we conclude that \( f(U_{0,[1]}) = \sum_{j=1}^{r} \lambda_{j,[1]}^0 \) > 0 and hence \( \sum_{j=1}^{r} \lambda_{j,[1]}^i \lambda_{j,[1]}^0 > 0 \) by (88). Thus, we conclude that

\[
\sum_{j=1}^{r} \lambda_{j,[1]}^i \lambda_{j,[1]}^0 \leq \sum_{j=1}^{r} (\lambda_{j,[1]}^i)^2 \sum_{j=1}^{r} \lambda_{j,[1]}^0 \lambda_{j,[1]}^0 \leq \sum_{j=1}^{r} (\lambda_{j,[1]}^i)^2,
\]  

(90)

where the first inequality follows from (89) and the second from (88). Combining (90) with (85) and (86), we may obtain (44) for \( i = 0 \) and \( p = 1 \).

On the one hand, from (89) we obtain

\[
0 \leq \sum_{j=1}^{r} (\lambda_{j,[p]}^i)^2 (\sum_{j=1}^{r} (\lambda_{j,[p]}^{i+1})^2 - \sum_{j=1}^{r} (\lambda_{j,[p]}^i)^2).
\]

Since \( \sum_{j=1}^{r} (\lambda_{j,[p]}^i)^2 > 0 \) if there is no truncation, we must have

\[
0 \leq \sum_{j=1}^{r} (\lambda_{j,[p]}^{i+1})^2 - \sum_{j=1}^{r} (\lambda_{j,[p]}^i)^2 = f(U_{i+1,[p]}) - f(U_{i,[p]}),
\]

i.e., the objective function \( f \) is monotonically increasing during the APD iteration as long as there is no truncation. On the other hand, there are at most \( r \) truncations and hence the total loss of \( f \) by the truncation is at most \( r \kappa^2 < f(U_{[0]}) \). Therefore, \( f \) is always positive and according to (88), we may conclude that \( \sum_{j=1}^{r} \lambda_{j,[p]}^{i+1} \lambda_{j,[p]}^i > 0 \) along iterations. By induction on \( p \), we obtain

\[
\sum_{j=1}^{r} \lambda_{j,[p]}^{i+1} \lambda_{j,[p]}^i \leq \sum_{j=1}^{r} (\lambda_{j,[p]}^{i+1})^2 \sum_{j=1}^{r} (\lambda_{j,[p]}^i)^2 \sum_{j=1}^{r} \lambda_{j,[p]}^{i+1} \lambda_{j,[p]}^i \leq \sum_{j=1}^{r} (\lambda_{j,[p]}^i)^2,
\]

which together with (85) and (86), implies (44) for arbitrary nonnegative integer \( p \). □

C.2 Proof for Lemma 7

Proof The set of subgradients of \( h \) can be partitioned as follows:

\[
\partial h(U) = (-\nabla V_{(r,n_1)}(U^{(1)})) \times \cdots \times (-\nabla V_{(r,n_k)}(U^{(k)})) \times \nabla f(U) + \partial \delta V_{(r,n_1)}(U^{(1)})) \times \cdots \times \partial \delta V_{(r,n_k)}(U^{(k)}))
\]  

(91)
Following the notation of Algorithm 1, we set

\[ x_j := (u^{(1)}_{j,[p+1]}, \ldots, u^{(k)}_{j,[p+1]}) \]

for all \( j = 1, \ldots, r \),

where \( u^{(i)}_{j,[p+1]} \) is the \( j \)-th column of the matrix \( U^{(i)}_{[p+1]} \) for all \( i \in \{1, \ldots, k\} \),

\[ V^{(i)} := \begin{bmatrix} v_{1}^{(i)} & \ldots & v_{r}^{(i)} \end{bmatrix} \]

with \( v_{j}^{(i)} := A\tau_i(x_j) \) for all \( j = 1, \ldots, r \), \( i \in \{1, \ldots, k\} \) and

\[ \Lambda := \text{diag}(\lambda_1, \ldots, \lambda_r) \] with \( \lambda_j := A\tau(x_j) \).

By (39) and (87), we have

\[ V^{(i)}_{[p+1]}A^{(i)}_{[p+1]} + \alpha U^{(i)}_{[p]} = U^{(i)}_{[p+1]}S^{(i)}_{[p+1]}, \]

where \( \alpha \in \{0, \epsilon\} \) depending on whether or not there is a proximal correction. According to (69) and (94), we have

\[ -U^{(i)}_{[p+1]} \in \partial \delta_{V(r,n)}(U^{(i)}_{[p+1]}), \quad V^{(i)}_{[p+1]}A^{(i)}_{[p+1]} + \alpha U^{(i)}_{[p]} \in \partial \delta_{V(r,n)}(U^{(i)}_{[p+1]}). \]

which implies that \( V^{(i)}_{[p+1]}A^{(i)}_{[p+1]} + \alpha(U^{(i)}_{[p]} - U^{(i)}_{[p+1]}) \in \partial \delta_{V(r,n)}(U^{(i)}_{[p+1]}). \) If we take

\[ W^{(i)}_{[p+1]} := -2V^{(i)}\Lambda + 2V^{(i)}_{[p+1]}A^{(i)}_{[p+1]} + 2\alpha(U^{(i)}_{[p]} - U^{(i)}_{[p+1]}), \]

then we have

\[ W^{(i)}_{[p+1]} \in -2V^{(i)}\Lambda + \partial \delta_{V(r,n)}(U^{(i)}_{[p+1]}). \]

On the other hand,

\[
\frac{1}{2} \|W^{(i)}_{[p+1]}\|_F = \|V^{(i)}\Lambda - V^{(i)}_{[p+1]}A^{(i)}_{[p+1]} - \alpha(U^{(i)}_{[p]} - U^{(i)}_{[p+1]})\|_F
\leq \|V^{(i)}\Lambda - V^{(i)}_{[p+1]}A\|_F + \|V^{(i)}_{[p+1]}A^{(i)}_{[p+1]} - \alpha(U^{(i)}_{[p]} - U^{(i)}_{[p+1]})\|_F + \|U^{(i)}_{[p]} - U^{(i)}_{[p+1]}\|_F
\leq \|\Lambda\|_F \|A\|_F \left( \sum_{j=1}^r \|\tau_j(x_j) - \tau_j(x_{j,[p+1]})\| \right)
+ \|V^{(i)}_{[p+1]}\|_F \|A\| \left( \sum_{j=1}^r \|\tau(x_j) - \tau(x_{j,[p+1]})\| \right) + \alpha\|U^{(i)}_{[p]} - U^{(i)}_{[p+1]}\|_F
\leq \sqrt{r}\|\Lambda\|_F \left( \sum_{j=1}^r \sum_{i=1}^k \|u^{(i)}_{j,[p+1]} - u^{(i)}_{j,1}\| \right)
\[\frac{1}{2} \| \nabla_i f(\mathbb{U}_{[p+1]}) - U_{[p+1]}^{(i)} (\nabla_i f(\mathbb{U}_{[p+1]}))^T U_{[p+1]}^{(i)} \| F = \| V^{(i)} A - U_{[p+1]}^{(i)} V^{(i)} A^T U_{[p+1]}^{(i)} \| F = \| V^{(i)} A - U_{[p+1]}^{(i)} V^{(i)} A^T U_{[p+1]}^{(i)} \| F \]

where the third inequality follows from the fact that

\[V^{(i)} - V_{[p+1]}^{(i)} = \left[ \mathcal{A}(\tau_i(x_1) - \tau_i(x_{i,1}^{[p+1]})) \ldots \mathcal{A}(\tau_i(x_r) - \tau_i(x_{r,1}^{[p+1]})) \right],\]

and a similar formula for \( \Lambda - \Lambda_{[p+1]}^{(i)} \), the fourth follows from the fact that

\[|\mathcal{A} \tau(x)| \leq \| \mathcal{A} \| \]

for any vector \( x := (x_1, \ldots, x_k) \) with \( \|x_i\| = 1 \) for all \( i = 1, \ldots, k \) and the last one follows from \( \alpha \leq \epsilon \). This, together with (91), implies (48). \( \square \)

C.3 Proof for Lemma 9

**Proof** We let

\[W_{[p+1]}^{(i)} := V^{(i)} A - \underbrace{V_{[p+1]}^{(i)} A_{[p+1]}^{(i)}}_{\alpha \in \{0, \epsilon\} \text{ depending on whether there is a proximal correction or not (cf. the proof for Lemma 7)}} - \alpha \left( U_{[p]}^{(i)} - U_{[p+1]}^{(i)} \right),\]

where \( \alpha \in \{0, \epsilon\} \) depending on whether there is a proximal correction or not (cf. the proof for Lemma 7). It follows from Lemma 7 that

\[\| W_{[p+1]}^{(i)} \|_F \leq \gamma_0 \| \mathbb{U}_{[p]} - \mathbb{U}_{[p+1]} \|_F\]

for some constant \( \gamma_0 > 0 \). By Algorithm 1, we have

\[V_{[p+1]}^{(i)} A_{[p+1]}^{(i)} + \alpha U_{[p]}^{(i)} = U_{[p+1]}^{(i)} S_{[p+1]}^{(i)} \quad \quad (96)\]

where \( S_{[p+1]}^{(i)} \) is a symmetric positive semidefinite matrix. Since \( U_{[p+1]}^{(i)} \in V(r, n_i) \) is an orthonormal matrix, we have

\[S_{[p+1]}^{(i)} = (U_{[p+1]}^{(i)})^T (V_{[p+1]}^{(i)} A_{[p+1]}^{(i)} + \alpha U_{[p]}^{(i)}) = (V_{[p+1]}^{(i)} A_{[p+1]}^{(i)} + \alpha U_{[p]}^{(i)})^T U_{[p+1]}^{(i)}, \quad \quad (97)\]

where the second equality follows from the symmetry of the matrix \( S_{[p+1]}^{(i)} \).

Consequently, we have

\[\frac{1}{2} \| \nabla_i f(\mathbb{U}_{[p+1]}) - U_{[p+1]}^{(i)} (\nabla_i f(\mathbb{U}_{[p+1]}))^T U_{[p+1]}^{(i)} \| F = \| V^{(i)} A - U_{[p+1]}^{(i)} V^{(i)} A^T U_{[p+1]}^{(i)} \| F = \| W_{[p+1]}^{(i)} + V_{[p+1]}^{(i)} A_{[p+1]}^{(i)} + \alpha (U_{[p]}^{(i)} - U_{[p+1]}^{(i)}) - U_{[p+1]}^{(i)} V^{(i)} A^T U_{[p+1]}^{(i)} \| F\]
where $\gamma_1 = \gamma_0 + \epsilon$. Next, we derive an estimation for the second summand of the right hand side of (98). To do this, we notice that

$$\| V_{[p+1]}^{(i)} A_{[p+1]}^{(i)} - U_{[p+1]}^{(i)} (V^{(i)} A)^T U_{[p+1]}^{(i)} \|_F \leq \gamma_1 \| U_{[p]}^{(i)} - U_{[p+1]}^{(i)} \|_F + \| V_{[p+1]}^{(i)} A_{[p+1]}^{(i)} - U_{[p+1]}^{(i)} (V^{(i)} A)^T U_{[p+1]}^{(i)} \|_F,$$

(98)

where $\gamma_2 > 0$ is some constant, the first equality follows from (96), the second from (97), the second inequality follows from the fact that $U_{[p+1]}^{(i)} \in V(r, n_i)$ and the last inequality from the relation

$$\| (U_{[p]}^{(i)})^T U_{[p+1]}^{(i)} - I \|_F \leq \| U_{[p]}^{(i)} - U_{[p+1]}^{(i)} \|_F,$$

which is obtained by Lemma 19. The desired inequality can be derived easily from (98) and (99).

References

1. Absil, P.-A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton, USA (2008)
2. Anandkumar, A., Ge, R., Hsu, D., Kakade, S.M., Telgarsky, M.: Tensor decompositions for learning latent variable models. J. Mach. Learn. Res. 15, 2773–2832 (2014)
3. Anderson, E., Bai, Z.Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., Sorensen, D.: LAPACK Users’ Guide, 3rd edn. SIAM, Philadelphia, PA (1999)
4. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. Math. Oper. Res. 35, 438–457 (2010)
5. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Math. Program. 137(1–2), 91–129 (2013)
6. Beck, A.: First-Order Methods in Optimization. MOS-SIAM Series on Optimization, SIAM (2017)
7. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmont, USA (1999)

---

Since $U$ is orthonormal, we must have

$$\| U A U \|_F^2 = \| A U \|_F^2 = \langle A U^T, A \rangle \leq \| A \|_F^2,$$
8. Bochnak, J., Coste, M., Roy, M.-F.: Real Algebraic Geometry. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 36. Springer, Berlin (1998)
9. Bolte, J., Daniilidis, A., Lewis, A.S., Shiota, M.: Clarke subgradients of stratifiable manifolds. SIAM J. Optim. 18(2), 556–572 (2007)
10. Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of Łojasiewicz inequalities and applications: subgradient flows, talweg, convexity. Trans. Am. Math. Soc. 362, 3319–3363 (2010)
11. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. 146(1–2), 459–494 (2014)
12. Boralevi, A., Draisma, J., Horobet, E., Robeva, E.: Orthogonal and unitary tensor decomposition from an algebraic perspective. Isr. J. Math. 222, 223–260 (2017)
13. Chen, J., Saad, Y.: On the tensor SVD and the optimal low rank orthogonal approximation of tensors. SIAM J. Matrix Anal. Appl. 30, 1709–1734 (2009)
14. Comon, P.: Independent component analysis, a new concept? Signal Process. 36, 287–314 (1994)
15. Comon, P.: Tensors: a brief introduction. IEEE Signal Proc. Mag. 31, 44–53 (2014)
16. Comon, P., Kressner, D.: A literature survey of low-rank tensor approximation techniques. GAMM-Mitt. 36, 53–78 (2013)
17. Eckart, C., Young, G.: The approximation of one matrix by another of lower rank. Psychometrika 1, 211–218 (1936)
18. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
19. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
20. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
21. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
22. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
23. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
24. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
25. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
26. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
27. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
28. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
29. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
30. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
31. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
32. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
33. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
34. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
35. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
36. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
37. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
38. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
39. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
40. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM. J. Matrix Anal. Appl. 20(2), 303–353 (1998)
41. Lee, J.D., Simchowitz, M., Jordan, M.I., Recht, B.: Gradient descent only converges to minimizers. P. Mach. Learn. Res. 49, 1–12 (2016)
42. Li, G., Pong, T.K.: Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods. Found. Comput. Math. 18, 1199–1232 (2018)
43. Li, J., Usevich, K., Comon, P.: Globally convergent jacobi-type algorithms for simultaneous orthogonal symmetric tensor diagonalization. SIAM J. Matrix Anal. Appl. 39, 1–22 (2018)
44. Li, J., Usevich, K., Comon, P.: On approximate diagonalization of third order symmetric tensors by orthogonal transformations. Linear Algebra. Appl. 576, 324–351 (2019)
45. Li, J., Usevich, K., Comon, P.: Jacobi-type algorithm for low rank orthogonal approximation of symmetric tensors and its convergence analysis. arXiv: 1911.00659
46. Lim, L.-H.: Tensors and hypermatrices. Chapter 15 in Handbook of Linear Algebra. In: Hogben, L. (eds.) (2013)
47. Liu, H., Wu, W., So, A.M.-C.: Quadratic optimization with orthogonality constraints: explicit Łojasiewicz exponent and linear convergence of line-search methods. In: Proceedings of the 33rd International Convergnece on Machine Learning, New York, USA (2016)
48. Łojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels, Les Équations aux Dérivées Partielles. Éditions du centre National de la Recherche Scientifique, Paris, pp. 87–89 (1963)
49. Martin, C.D.M., Van Loan, C.F.: A Jacobi-type method for computing orthogonal tensor decompositions. SIAM J. Matrix Anal. Appl. 30, 1219–1232 (2008)
50. McCullagh, P.: Tensor Methods in Statistics. Chapman and Hall, London (1987)
51. Milnor, J.: Morse Theory. Annals Math. Studies, 51. Princeton Univ. Press, Princeton, NJ (1963)
52. Mohlenkamp, M.J.: Musings on multilinear fitting. Linear Algebra. Appl. 438, 834–852 (2013)
53. Mu, C., Hsu, D., Goldfarb, D.: Successive rank-one approximations for nearly orthogonally decomposable symmetric tensors. SIAM J. Matrix Anal. Appl. 36, 1638–1659 (2015)
54. Nesterov, Y.: Introductory Lectures on Convex Optimization-A Basic Course. Kluwer Academic Publishers, London (2004)
55. Qi, L., Luo, Z.: Tensor Analysis: Spectral Theory and Special Tensors. SIAM (2017)
56. Rockafellar, R.T., Wets, R.: Variational Analysis. Grundlehren der Mathematischen Wissenschaften, vol. 317. Springer, Berlin (1998)
57. Shafarevich, I.R.: Basic Algebraic Geometry. Springer-Verlag, Berlin (1977)
58. Sørensen, M., De Lathauwer, L., Comon, P., Jcart, S., Deneire, L.: Canonical polyadic decomposition with a columnwise orthonormal factor matrix. SIAM J. Matrix Anal. Appl. 33, 1190–1213 (2012)
59. Sun, J., Chen, C.: Generalized polar decomposition. Math. Numer. Sinica 11, 262–273 (1989)
60. Uschmajew, A.: Local convergence of the alternating least squares algorithm for canonical tensor approximation. SIAM J. Matrix Anal. Appl. 33, 639–652 (2012)
61. Uschmajew, A.: A new convergence proof for the high-order power method and generalizations. Pacific J. Optim. 11, 309–321 (2015)
62. Usevich, K., Li, J., Comon, P.: Approximate matrix and tensor diagonalization by unitary transformations: convergence of Jacobi-type algorithms. SIAM J. Optim. 30, 2998–3028 (2020)
63. Wang, L., Chu, M.T.: On the global convergence of the alternating least squares method for rank-one tensor approximation. SIAM J. Matrix Anal. Appl. 35(3), 1058–1072 (2014)
64. Wang, L., Chu, M.T., Yu, B.: Orthogonal low rank tensor approximation: alternating least squares method and its global convergence. SIAM J. Matrix Anal. Appl. 36, 1–19 (2015)
65. Yang, Y.: The epsilon-alternating least squares for orthogonal low-rank tensor approximation and its global convergence. SIAM J. Matrix Anal. Appl. 41, 1797–1825 (2020)
66. Zhang, T., Golub, G.H.: Rank-one approximation to high order tensors. SIAM J. Matrix Anal. Appl. 23, 534–550 (2001)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.