An Approximate Projection onto the Tangent Cone to the Variety of Third-Order Tensors of Bounded Tensor-Train Rank

Charlotte Vermeylen[1] [0000−0002−3615−6571], Guillaume Olikier[2] [0000−0002−2767−0480], and Marc Van Barel[1] [0000−0002−7372−382X]

[1] Department of Computer Science, KU Leuven, Heverlee, Belgium
charlotte.vermeylen@kuleuven.be
[2] ICTEAM Institute, UCLouvain, Louvain-la-Neuve, Belgium

Abstract. An approximate projection onto the tangent cone to the variety of third-order tensors of bounded tensor-train rank is proposed which satisfies a better angle condition than the one proposed by Kutschan (2019). Such an approximate projection enables, e.g., to compute gradient-related directions in the tangent cone, as required by algorithms aiming at minimizing a continuously differentiable function on the variety, a problem appearing notably in tensor completion. We present a numerical experiment indicating that the proven angle condition is pessimistic and that our approximate projection is close to the exact one.

Keywords: Projection · Tangent cone · Angle condition · Tensor-train decomposition.

1 Introduction

Tangent cones play an important role in constrained optimization to describe admissible search directions and to formulate optimality conditions [9, Chap. 6]. In this paper, we focus on the set

$$\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)} := \{X \in \mathbb{R}^{n_1\times n_2\times n_3} \mid \operatorname{rank}_{\mathrm{TT}}(X) \leq (k_1,k_2)\}, \tag{1}$$

where $\operatorname{rank}_{\mathrm{TT}}(X)$ denotes the tensor-train rank of X (see Section 2.2), which is a real algebraic variety [4], and, given $X \in \mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$, we propose an *approximate projection* onto $T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$, i.e., a set-valued mapping $\tilde{\mathcal{P}}_{T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}} : \mathbb{R}^{n_1\times n_2\times n_3} \multimap T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$ such that there exists $\omega \in (0,1]$ such that, for all $Y \in \mathbb{R}^{n_1\times n_2\times n_3}$ and all $\tilde{Y} \in \tilde{\mathcal{P}}_{T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}} Y$,

$$\langle Y, \tilde{Y}\rangle \geq \omega \|\mathcal{P}_{T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}} Y\| \|\tilde{Y}\|, \tag{2}$$

where $\langle\cdot,\cdot\rangle$ is the inner product on $\mathbb{R}^{n_1\times n_2\times n_3}$ given in [1, Example 4.149], $\|\cdot\|$ is the induced norm, and the set

$$\mathcal{P}_{T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}} Y := \operatorname*{argmin}_{Z\in T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}} \|Z-Y\|^2 \tag{3}$$

is the projection of Y onto $T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$. By [10, Definition 2.5], inequality (2) is called an *angle condition*; it is well defined since, as $T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$ is a closed cone, all elements of $\mathcal{P}_{T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}}Y$ have the same norm (see Section 3). Such an approximate projection enables, e.g., to compute a gradient-related direction in $T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$, as required in the second step of [10, Algorithm 1] if the latter is used to minimize a continuously differentiable function $f : \mathbb{R}^{n_1\times n_2\times n_3} \to \mathbb{R}$ on $\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$, a problem that appears notably in tensor completion; see [11] and the references therein.

An approximate projection onto $T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$ satisfying the angle condition (2) with $\omega = \frac{1}{6\sqrt{n_1n_2n_3}}$ was proposed in [5, §5.4]. If X is a singular point of the variety, i.e., $(r_1, r_2) := \mathrm{rank}_{\mathrm{TT}}(X) \neq (k_1, k_2)$, the approximate projection proposed in this paper ensures (see Theorem 1)

$$\omega = \sqrt{\max\left\{\frac{k_1 - r_1}{n_1 - r_1}, \frac{k_2 - r_2}{n_3 - r_2}\right\}}, \tag{4}$$

which is better, and can be computed via SVDs (see Algorithm 1). We point out that no general formula to project onto the closed cone $T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$, which is neither linear nor convex (see Section 2.3), is known in the literature.

This paper is organized as follows. Preliminaries are introduced in Section 2. Then, in Section 3, we introduce the proposed approximate projection and prove that it satisfies (2) with ω as in (4) (Theorem 1). Finally, in Section 4, we present a numerical experiment where the proposed approximate projection preserves the direction better than the one from [5, §5.4].

2 Preliminaries

In this section, we introduce the preliminaries needed for Section 3. In Section 2.1, we recall basic facts about orthogonal projections. Then, in Section 2.2, we review the tensor-train decomposition. Finally, in Section 2.3, we review the description of the tangent to $\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$ given in [4, Theorem 2.6].

2.1 Orthogonal Projections

Given $n, p \in \mathbb{N}$ with $n \geq p$, we let $\mathrm{St}(p, n) := \{U \in \mathbb{R}^{n\times p} \mid U^\top U = I_p\}$ denote the Stiefel manifold. For every $U \in \mathrm{St}(p, n)$, we let $P_U := UU^\top$ and $P_U^\perp := I_n - P_U$ denote the orthogonal projections onto the range of U and its orthogonal complement, respectively. The proof of Theorem 1 relies on the following basic result.

Lemma 1. *Let $A \in \mathbb{R}^{n\times m}$ have rank r. If $\hat{A} = \hat{U}\hat{S}\hat{V}^\top$ is a truncated SVD of rank s of A, with $s < r$, then, for all $U \in \mathrm{St}(s, n)$ and all $V \in \mathrm{St}(s, m)$,*

$$\|P_{\hat{U}}A\| \geq \|P_U A\|, \qquad \|P_{\hat{U}}A\|^2 \geq \frac{s}{r}\|A\|^2, \tag{5}$$

$$\|AP_{\hat{V}}\| \geq \|AP_V\|, \qquad \|AP_{\hat{V}}\|^2 \geq \frac{s}{r}\|A\|^2. \tag{6}$$

Proof. By the Eckart–Young theorem, $\hat{A}$ is a projection of A onto $\mathbb{R}^{n\times m}_{\leq s}$, and, because $\mathbb{R}^{n\times m}_{\leq s}$ is a closed cone, the same conditions as in (14) hold, and because $\hat{S}\hat{V}^\top = \hat{U}^\top A$ and thus $\hat{A} = \hat{U}\hat{U}^\top A$, it holds that

$$\|P_{\hat{U}}A\|^2 = \max_{\substack{A_1 \in \mathbb{R}^{n\times m}_{\leq s} \\ \langle A_1, A\rangle = \|A_1\|^2}} \|A_1\|^2.$$

Furthermore, for all $U \in \mathrm{St}(s,n)$, $\langle P_U A, A\rangle = \left\langle P_U A, P_U A + P_U^\perp A\right\rangle = \|P_U A\|^2$. In other words,

$$\{P_U A \mid U \in \mathrm{St}(s,n)\} \subseteq \{A_1 \in \mathbb{R}^{n\times m}_{\leq s} \mid \langle A_1, A\rangle = \|A_1\|^2\}.$$

Thus,

$$\|P_{\hat{U}}A\|^2 = \max_{U\in\mathrm{St}(s,n)} \|P_U A\|^2.$$

The left inequality in (5) follows, and the one in (6) can be obtained similarly.

By orthogonal invariance of the Frobenius norm and by definition of $\hat{A}$,

$$\|A\|^2 = \|S\|^2 = \sum_{i=1}^{r} \sigma_i^2, \qquad \|\hat{A}\|^2 = \|\hat{S}\|^2 = \sum_{i=1}^{s} \sigma_i^2,$$

where $\sigma_1, \dots, \sigma_r$ are the singular values of A in decreasing order. Moreover, either $\sigma_s^2 \geq \frac{1}{r}\sum_{i=1}^r \sigma_i^2$ or $\sigma_s^2 < \frac{1}{r}\sum_{i=1}^r \sigma_i^2$. In the first case, we have

$$\|\hat{A}\|^2 = \sum_{i=1}^{s} \sigma_i^2 \geq s\sigma_s^2 \geq s\frac{\sum_{i=1}^r \sigma_i^2}{r} = \frac{s}{r}\|A\|^2.$$

In the second case, we have

$$\|\hat{A}\|^2 = \sum_{i=1}^{r} \sigma_i^2 - \sum_{i=s+1}^{r} \sigma_i^2 \geq \|A\|^2 - (r-s)\sigma_s^2 > \|A\|^2 - (r-s)\frac{\sum_{i=1}^r \sigma_i^2}{r} = \frac{s}{r}\|A\|^2.$$

Thus, in both cases, the second inequality in (5) holds. The second inequality in (6) can be obtained in a similar way. □

2.2 The Tensor-Train Decomposition

In this section, we review basic facts about the tensor-train decomposition (TTD) that are used in Section 3; we refer to the original paper [8] and the subsequent works [3,11,12] for more details.

A *tensor-train decomposition* of $X \in \mathbb{R}^{n_1\times n_2\times n_3}$ is a factorization

$$X = X_1 \cdot X_2 \cdot X_3, \tag{7}$$

where $X_1 \in \mathbb{R}^{n_1\times r_1}$, $X_2 \in \mathbb{R}^{r_1\times n_2\times r_2}$, $X_3 \in \mathbb{R}^{r_2\times n_3}$, and '$\cdot$' denotes the contraction between a matrix and a tensor. The minimal (r_1, r_2) for which a TTD of X exists is called the *TT-rank* of X and is denoted by $\mathrm{rank}_{\mathrm{TT}}(X)$. By [2, Lemma 4], the set

$$\mathbb{R}^{n_1\times n_2\times n_3}_{(k_1,k_2)} := \{X \in \mathbb{R}^{n_1\times n_2\times n_3} \mid \mathrm{rank}_{\mathrm{TT}}(X) = (k_1, k_2)\} \tag{8}$$

is a smooth embedded submanifold of $\mathbb{R}^{n_1\times n_2\times n_3}$.

Let $X^{\mathrm{L}} := [X]^{n_1\times n_2 n_3} := \mathrm{reshape}(X, n_1 \times n_2 n_3)$ and $X^{\mathrm{R}} = [X]^{n_1 n_2\times n_3} := \mathrm{reshape}(X, n_1 n_2 \times n_3)$ denote respectively the left and right unfoldings of X. Then, $\mathrm{rank}_{\mathrm{TT}}(X) = (\mathrm{rank}(X^{\mathrm{L}}), \mathrm{rank}(X^{\mathrm{R}}))$ and the minimal rank decomposition can be obtained by computing two successive SVDs of unfoldings; see [8, Algorithm 1]. The contraction interacts with the unfoldings according to the following rules:

$$X_1 \cdot X_2 \cdot X_3 = \left[X_1 (X_2 \cdot X_3)^{\mathrm{L}}\right]^{n_1\times n_2\times n_3}, \qquad X_2 \cdot X_3 = \left[X_2^{\mathrm{R}} X_3\right]^{r_1\times n_2\times n_3}.$$

For every $i \in \{1, 2, 3\}$, if $U_i \in \mathrm{St}(r_i, n_i)$, then the mode-$i$ vectors of $P_{U_1} X^{\mathrm{L}}(P_{U_3} \otimes P_{U_2})$ are the orthogonal projections onto the range of U_i of those of X. A similar property holds for X^{R}. We say that X is *left-orthogonal* if $n_1 \le n_2 n_3$ and $(X^{\mathrm{L}})^\top \in \mathrm{St}(n_1, n_2 n_3)$, and *right-orthogonal* if $n_3 \le n_1 n_2$ and $X^{\mathrm{R}} \in \mathrm{St}(n_3, n_1 n_2)$.

As a TTD is not unique, certain orthogonalization conditions can be enforced, which can improve the stability of algorithms working with TTDs. For every $i \in \{1, 2, 3\}$, the TTD (7) is said to be *i-orthogonal* if the first $i - 1$ factors are right-orthogonal and the last $3 - i$ factors are left-orthogonal.

2.3 The Tangent Cone to the Low-Rank Variety

In [4, Theorem 2.6], a parametrization of the tangent cone to $\mathbb{R}^{n_1\times n_2\times n_3}_{\le(k_1,k_2)}$ is given and, because this parametrization is not unique, corresponding orthogonality conditions are added. The following lemma recalls this parametrization however with slightly different orthogonality conditions that make the computations in the experiments more stable and the proofs in the rest of the paper easier.

Lemma 2. *Let $X \in \mathbb{R}^{n_1\times n_2\times n_3}_{(r_1,r_2)}$ have $X = X_1 \cdot X_2'' \cdot X_3'' = X_1' \cdot X_2' \cdot X_3$ as TTDs, where $(X_2''^{\mathrm{L}})^\top \in \mathrm{St}(r_1, n_2 r_2)$, $X_3''^\top \in \mathrm{St}(r_2, n_3)$, $X_1' \in \mathrm{St}(r_1, n_1)$, and $X_2'^{\mathrm{R}} \in \mathrm{St}(r_2, r_1 n_2)$. Then, $T_X \mathbb{R}^{n_1\times n_2\times n_3}_{\le(k_1,k_2)}$ is the set of all g such that*

$$g = \begin{bmatrix} X_1' & U_1 & W_1 \end{bmatrix} \cdot \begin{bmatrix} X_2' & U_2 & W_2 \\ 0 & Z_2 & V_2 \\ 0 & 0 & X_2'' \end{bmatrix} \cdot \begin{bmatrix} W_3 \\ V_3 \\ X_3'' \end{bmatrix} \tag{9}$$

with $U_1 \in \mathrm{St}(s_1, n_1)$, $W_1 \in \mathbb{R}^{n_1\times r_1}$, $U_2 \in \mathbb{R}^{r_1\times n_2\times s_2}$, $W_2 \in \mathbb{R}^{r_1\times n_2\times r_2}$, $Z_2 \in \mathbb{R}^{s_1\times n_2\times s_2}$, $V_2 \in \mathbb{R}^{s_1\times n_2\times r_2}$, $W_3 \in \mathbb{R}^{r_2\times n_3}$, $V_3^\top \in \mathrm{St}(s_2, n_3)$, $s_i = k_i - r_i$ for all $i \in \{1, 2\}$, and

$$\begin{aligned} U_1^\top X_1' = 0, \quad W_1^\top X_1' = 0, \quad (U_2^{\mathrm{R}})^\top X_2'^{\mathrm{R}} = 0, \\ W_3 X_3''^\top = 0, \quad V_3 X_3''^\top = 0, \quad V_2^{\mathrm{L}} (X_2''^{\mathrm{L}})^\top = 0. \end{aligned} \tag{10}$$

Proof. By [4, Theorem 2.6], $T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$ is the set of all $g\in\mathbb{R}^{n_1\times n_2\times n_3}$ that can be decomposed as

$$g=\begin{bmatrix}X_1' & \dot U_1 & \dot W_1\end{bmatrix}\cdot\begin{bmatrix}X_2' & \dot U_2 & \dot W_2\\ 0 & \dot Z_2 & \dot V_2\\ 0 & 0 & X_2'\end{bmatrix}\cdot\begin{bmatrix}\dot W_3\\ \dot V_3\\ X_3\end{bmatrix},$$

with the orthogonality conditions

$$\begin{aligned}&\dot U_1^\top X_1'=0, && \left(\dot V_2\cdot X_3\right)^{\mathrm{L}}\left((X_2'\cdot X_3)^{\mathrm{L}}\right)^\top=0, && \dot V_3X_3^\top=0,\\ &\dot W_1^\top X_1'=0, && \left(\dot W_2^{\mathrm{R}}\right)^\top X_2'^{\mathrm{R}}=0, && \left(\dot U_2^{\mathrm{R}}\right)^\top X_2'^{\mathrm{R}}=0.\end{aligned}\tag{11}$$

The following invariances hold for all $B\in\mathbb{R}^{s_2\times s_2}, C\in\mathbb{R}^{r_2\times r_2}, Q\in\mathbb{R}^{s_1\times s_1}$, and $R\in\mathbb{R}^{r_1\times r_1}$:

$$\begin{aligned}g&=\begin{bmatrix}X_1' & \dot U_1Q^{-1} & \dot W_1R^{-1}\end{bmatrix}\cdot\begin{bmatrix}X_2' & \dot U_2\cdot B & \dot W_2\cdot C\\ 0 & Q\cdot\dot Z_2\cdot B & Q\cdot\dot V_2\cdot C\\ 0 & 0 & R\cdot X_2'\cdot C\end{bmatrix}\cdot\begin{bmatrix}\dot W_3\\ B^{-1}\dot V_3\\ C^{-1}X_3'\end{bmatrix}\\ &=\begin{bmatrix}X_1' & U_1 & W_1\end{bmatrix}\cdot\begin{bmatrix}X_2' & U_2 & \dot W_2\cdot C\\ 0 & Z_2 & V_2\\ 0 & 0 & X_2''\end{bmatrix}\cdot\begin{bmatrix}\dot W_3\\ V_3\\ X_3''\end{bmatrix},\end{aligned}$$

where we have defined $U_1:=\dot U_1Q^{-1}$, $W_1:=\dot W_1R^{-1}$, $U_2:=\dot U_2\cdot B$, $Z_2:=Q\cdot\dot Z_2\cdot B$, $V_2:=Q\cdot\dot V_2\cdot C$, $X_2'':=R\cdot X_2'\cdot C$, $V_3:=B^{-1}\dot V_3$, and $X_3'':=C^{-1}X_3'$. The matrices B,C,Q, and R can be chosen such that X_2'', X_3'', and V_3 are left-orthogonal, and U_1 is right-orthogonal, e.g., using SVDs. Additionally, $\dot W_3$ can be decomposed as $\dot W_3=\dot W_3X_3''^\top X_3''+W_3$. The two terms involving $\dot W_3$ and $\dot W_2\cdot C$ can then be regrouped as

$$\begin{aligned}&X_1'\cdot X_2'\cdot\dot W_3+X_1'\cdot\dot W_2\cdot CX_3''\\ &=X_1'\cdot\left(X_2'\cdot\dot W_3X_3''^\top+\dot W_2\cdot C\right)\cdot X_3''+X_1'\cdot X_2'\cdot W_3.\end{aligned}$$

Thus, defining $W_2:=X_2'\cdot\dot W_3X_3''^\top+\dot W_2\cdot C$, we obtain the parametrization (9) wich satisfies (10). $\square$

Expanding (9) yields, because of (10), a sum of six mutually orthogonal TTDs:

$$\begin{aligned}g=&\,W_1\cdot X_2''\cdot X_3''+X_1'\cdot X_2'\cdot W_3+X_1'\cdot W_2\cdot X_3''\\ &+U_1\cdot V_2\cdot X_3''+X_1'\cdot U_2\cdot V_3+U_1\cdot Z_2\cdot V_3.\end{aligned}\tag{12}$$

Thus, the following holds:

$$\begin{aligned}&W_1=g^{\mathrm{L}}\left((X_2''\cdot X_3'')^{\mathrm{L}}\right)^\top, && W_2=X_1'^\top\cdot g\cdot X_3''^\top, && W_3=\left((X_1''\cdot X_2'')^{\mathrm{R}}\right)^\top g^{\mathrm{R}},\\ &U_2=X_1'^\top\cdot g\cdot V_3^\top, && V_2=U_1^\top\cdot g\cdot X_3''^\top, && Z_2=U_1^\top\cdot g\cdot V_3^\top.\end{aligned}\tag{13}$$

The first three terms in (12) form the tangent space $T_X\mathbb{R}^{n_1\times n_2\times n_3}_{(r_1,r_2)}$, the projection onto which is described in [7, Theorem 3.1 and Corollary 3.2].

3 The Proposed Approximate Projection

In this section, we prove Proposition 1 and then use it to prove Theorem 1. Both results rely on the following observation. By [6, Proposition A.6], for all $Y \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and all $\hat{Y} \in \mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (k_1,k_2)}} Y$, it holds that $\langle Y - \hat{Y}, \hat{Y} \rangle = 0$ or, equivalently, $\langle Y, \hat{Y} \rangle = \|\hat{Y}\|^2$. Thus, all elements of $\mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (k_1,k_2)}} Y$ have the same norm and (3) can be rewritten as

$$\mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (k_1,k_2)}} Y = \underset{\substack{Z \in T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (k_1,k_2)} \\ \langle Y, Z \rangle = \|Z\|^2}}{\operatorname{argmax}} \|Z\| = \underset{\substack{Z \in T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (k_1,k_2)} \\ \langle Y, Z \rangle = \|Z\|^2}}{\operatorname{argmax}} \left\langle Y, \frac{Z}{\|Z\|} \right\rangle. \tag{14}$$

Proposition 1. *Let X be as in Lemma 2. For every $Y \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and every $\hat{Y} \in \mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (k_1,k_2)}} Y$, if U_1 and V_3 are the parameters of $\hat{Y}$ in (12), then the parameters W_1, W_2, W_3, U_2, V_2, and Z_2 of $\hat{Y}$ can be written as*

$$\begin{aligned}
W_1 &= P^\perp_{X_1'} \left(Y \cdot X_3''^\top \right)^{\mathrm{L}} \left(X_2''^{\mathrm{L}} \right)^\top, & W_3 &= X_2'^{\mathrm{R}} \left(X_1'^\top \cdot Y \right)^{\mathrm{R}} P^\perp_{X_3''^\top}, \\
U_2 &= \left[P^\perp_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot Y \right)^{\mathrm{R}} \right]^{r_1 \times n_2 \times n_3} \cdot V_3^\top, & W_2 &= X_1'^\top \cdot Y \cdot X_3''^\top, \\
V_2 &= U_1^\top \cdot \left[\left(Y \cdot X_3''^\top \right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}} \right)^\top} \right]^{n_1 \times n_2 \times r_2}, & Z_2 &= U_1^\top \cdot Y \cdot V_3^\top.
\end{aligned} \tag{15}$$

Furthermore, $Y_\parallel(U_1, V_3)$ defined as in (12) with the parameters from (15) is a feasible point of (14) for all U_1 and all V_3.

Proof. Observe that any $Y \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ can be decomposed as

$$Y = P_{X_1'} \cdot Y \cdot P_{X_3''^\top} + P_{X_1'} \cdot Y \cdot P^\perp_{X_3''^\top} + P^\perp_{X_1'} \cdot Y \cdot P_{X_3''^\top} + P^\perp_{X_1'} \cdot Y \cdot P^\perp_{X_3''^\top}$$

because $P_U + P_U^\perp = I_m$, for all $U \in \mathrm{St}(r, m)$, $P \cdot Y + Q \cdot Y = (P + Q) \cdot Y$, for all $P, Q \in \mathbb{R}^{n_1 \times n_1}$, and $Y \cdot R + Y \cdot S = Y \cdot (R + S)$, for all $R, S \in \mathbb{R}^{n_3 \times n_3}$. The second and third terms can further be decomposed as

$$\begin{aligned}
Y = \; & P_{X_1'} \cdot Y \cdot P_{X_3''^\top} + X_1' \cdot \left[P_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot Y \right)^{\mathrm{R}} \right]^{r_1 \times n_2 \times n_3} \cdot P^\perp_{X_3''^\top} + \\
& X_1' \cdot \left[P^\perp_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot Y \right)^{\mathrm{R}} \right]^{r_1 \times n_2 \times n_3} \cdot P^\perp_{X_3''^\top} + \\
& P^\perp_{X_1'} \cdot \left[\left(Y \cdot X_3''^\top \right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}} \right)^\top} \right]^{n_1 \times n_2 \times r_2} \cdot X_3'' + \\
& P^\perp_{X_1'} \cdot \left[\left(Y \cdot X_3''^\top \right)^{\mathrm{L}} P_{\left(X_2''^{\mathrm{L}} \right)^\top} \right]^{n_1 \times n_2 \times r_2} \cdot X_3' + P^\perp_{X_1'} \cdot Y \cdot P^\perp_{X_3''^\top}
\end{aligned} \tag{16}$$

because, for all $P \in \mathbb{R}^{n_1 \times n_1}$, $Q \in \mathbb{R}^{n_2 n_3 \times n_2 n_3}$, $R \in \mathbb{R}^{n_3 \times n_3}$, and $S \in \mathbb{R}^{n_1 n_2 \times n_1 n_2}$,

$$\begin{aligned}
P \cdot Y \cdot R &= P \cdot \left[Y^{\mathrm{L}} Q \right]^{n_1 \times n_2 \times n_3} \cdot R + P \cdot \left[B^{\mathrm{L}} \left(I_{mp} - Q \right) \right]^{n_1 \times n_2 \times n_3} \cdot R, \\
&= P \cdot \left[S B^{\mathrm{R}} \right]^{n_1 \times n_2 \times n_3} \cdot R + P \cdot \left[\left(I_{np} - S \right) B^{\mathrm{R}} \right]^{n_1 \times n_2 \times n_3} \cdot R.
\end{aligned}$$

Using U_1 and V_3, the third, fourth, and last terms can be further decomposed as

$$\begin{aligned}
Y = \; & P_{X_1'} \cdot Y \cdot P_{X_3''^\top} + X_1' \cdot \left[P^\perp_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot Y\right)^{\mathrm{R}}\right]^{r_1 \times n_2 \times n_3} \cdot P_{X_3''^\top} + \\
& X_1' \cdot \left[P^\perp_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot Y\right)^{\mathrm{R}}\right]^{r_1 \times n_2 \times n_3} \cdot \left(P^\perp_{X_3''^\top} - P_{V_3^\top}\right) + \\
& X_1' \cdot \left[P^\perp_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot Y\right)^{\mathrm{R}}\right]^{r_1 \times n_2 \times n_3} \cdot P_{V_3^\top} + \\
& P_{U_1} \left[\left(Y \cdot X_3''^\top\right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}}\right)^\top}\right]^{n_1 \times n_2 \times r_2} \cdot X_3'' + \\
& \left(P^\perp_{X_1'} - P_{U_1}\right) \left[\left(Y \cdot X_3''^\top\right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}}\right)^\top}\right]^{n_1 \times n_2 \times r_2} \cdot X_3'' + \\
& P^\perp_{X_1'} \cdot \left[\left(Y \cdot X_3''^\top\right)^{\mathrm{L}} P_{\left(X_2''^{\mathrm{L}}\right)^\top}\right]^{n_1 \times n_2 \times r_2} \cdot X_3' + P_{U_1} \cdot Y \cdot P_{V_3^\top} + \\
& \left(P^\perp_{X_1'} - P_{U_1}\right) \cdot Y \cdot P^\perp_{X_3''^\top} + P_{U_1} \cdot Y \cdot \left(P^\perp_{X_3''^\top} - P_{V_3^\top}\right).
\end{aligned} \tag{17}$$

Because of the projection matrices, these ten terms are mutually orthogonal. On the other hand, based on its definition, $Y_\parallel(U_1, V_3)$ can be written as

$$\begin{aligned}
Y_\parallel(U_1, V_3) = \; & P_{X_1'} \cdot Y \cdot P_{X_3''^\top} + X_1' \cdot \left[P^\perp_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot Y\right)^{\mathrm{R}}\right]^{r_1 \times n_2 \times n_3} \cdot P_{X_3''^\top} + \\
& X_1' \cdot \left[P^\perp_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot Y\right)^{\mathrm{R}}\right]^{r_1 \times n_2 \times n_3} \cdot P_{V_3^\top} + \\
& P_{U_1} \left[\left(Y \cdot X_3''^\top\right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}}\right)^\top}\right]^{n_1 \times n_2 \times r_2} \cdot X_3'' + \\
& P^\perp_{X_1'} \cdot \left[\left(Y \cdot X_3''^\top\right)^{\mathrm{L}} P_{\left(X_2''^{\mathrm{L}}\right)^\top}\right]^{n_1 \times n_2 \times r_2} \cdot X_3' + P_{U_1} \cdot Y \cdot P_{V_3^\top}.
\end{aligned}$$

Thus, $\left\langle Y_\parallel(U_1, V_3), Y - Y_\parallel(U_1, V_3)\right\rangle = 0$ and hence, $Y_\parallel(U_1, V_3)$ is a feasible point of (14). Since $\hat{Y}$ is a solution to (14), $\|Y_\parallel(U_1, V_3)\| \leq \|\hat{Y}\|$. Therefore, if $\hat{Y} = 0$, then $Y_\parallel(U_1, V_3) = 0$ and consequently all parameters in (15) are zero because of (13). When U_1 and V_3 are the parameters of $\hat{Y}$, it holds similarly that $\langle Y, \hat{Y}\rangle = \langle Y_\parallel(U_1, V_3), \hat{Y}\rangle$. By using the Cauchy–Schwarz inequality, we have

$$\|\hat{Y}\|^2 = \langle Y, \hat{Y}\rangle = \langle Y_\parallel(U_1, V_3), \hat{Y}\rangle \leq \|Y_\parallel(U_1, V_3)\|\|\hat{Y}\| \leq \|\hat{Y}\|^2. \tag{18}$$

It follows that the Cauchy–Schwarz inequality is an equality and hence $Y_\parallel(U_1, V_3) = \hat{Y}$. Thus, because of (13), the parameters in (15) are those of $\hat{Y}$. □

Theorem 1. *Let X be as in Lemma 2 with $(r_1, r_2) \neq (k_1, k_2)$. The approximate projection that computes the parameters U_1 and V_3 of $\tilde{Y} \in \tilde{\mathcal{P}}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (k_1, k_2)}} Y$ in (12) with Algorithm 1 and the parameters W_1, W_2, W_3, U_2, V_2, and Z_2 with (15) satisfies (2) with ω as in (4) for all ε and all $i_{\max}$ in Algorithm 1.*

Proof. Let $(s_1, s_2) := (k_1 - r_1, k_2 - r_2)$ and $\hat{Y} \in \mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (k_1,k_2)}} Y$. Thus, $s_1 + s_2 > 0$. Because W_1, W_2, W_3, U_2, V_2, and Z_2 are as in (15), it holds that $\tilde{Y} = Y_{\|}(U_1, V_3)$, thus $\tilde{Y}$ is a feasible point of (14), and hence (2) is equivalent to $\|\tilde{Y}\| \geq \omega \|\hat{Y}\|$.

To compare the norm of $\tilde{Y}$ with the norm of $\hat{Y}$, $\hat{U}_1$ and $\hat{V}_3$ are defined as the parameters of $\hat{Y}$. From (15), because all terms are mutually orthogonal, we have that

$$\begin{aligned} \left\|\tilde{Y}\right\|^2 = \left\|\mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1,r_2)}} Y\right\|^2 + \left\|P_{U_1} \left(Y \cdot X_3''^\top\right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}}\right)^\top} \left(X_3'' \otimes I_{n_2}\right)\right\|^2 \\ + \left\|\left((P_{U_1} \cdot Y)^{\mathrm{R}} + (I_{n_2} \otimes X_1')\, P^\perp_{X_2'^{\mathrm{R}}} (X_1'^\top \cdot Y)^{\mathrm{R}}\right) P_{V_3^\top}\right\|^2 . \end{aligned}$$

Assume now that $s_2/(n_3 - r_2) > s_1/(n_1 - r_1)$ and consider the first iteration of Algorithm 1. Because V_3 is obtained by a truncated SVD of $\big((P_{U_1} \cdot Y)^{\mathrm{R}} + (I_{n_2} \otimes X_1')\, P^\perp_{X_2'^{\mathrm{R}}} (X_1'^\top \cdot Y)^{\mathrm{R}}\big) P^\perp_{X_3''^\top}$ and, by using (6),

$$\begin{aligned} \left\|\tilde{Y}\right\|^2 \geq \left\|\mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1,r_2)}} Y\right\|^2 + \left\|P_{U_1} \left(Y \cdot X_3''^\top\right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}}\right)^\top} \left(X_3'' \otimes I_{n_2}\right)\right\|^2 \\ + \frac{s_2}{n_3 - r_2} \left\|\left((P_{U_1} \cdot Y)^{\mathrm{R}} + (I_{n_2} \otimes X_1')\, P^\perp_{X_2'^{\mathrm{R}}} (X_1'^\top \cdot Y)^{\mathrm{R}}\right) P^\perp_{X_3''^\top}\right\|^2 . \end{aligned}$$

Furthermore, because U_1 is obtained from the truncated SVD of $P^\perp_{X_1'} \big((Y \cdot P^\perp_{X_3''})^{\mathrm{L}} + (Y \cdot X_3''^\top)^{\mathrm{L}} P^\perp_{X_2''^{\mathrm{L}\top}} (X_3'' \otimes I_{n_2})\big)$ and by using (5),

$$\begin{aligned} \left\|\tilde{Y}\right\|^2 \geq \left\|\mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1,r_2)}} Y\right\|^2 + \frac{s_2}{n_3 - r_2} \left\|P_{\hat{U}_1} \left(Y \cdot X_3''^\top\right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}}\right)^\top} \left(X_3'' \otimes I_{n_2}\right)\right\|^2 \\ + \frac{s_2}{n_3 - r_2} \left\|\left((P_{\hat{U}_1} \cdot Y)^{\mathrm{R}} + (I_{n_2} \otimes X_1')\, P^\perp_{X_2'^{\mathrm{R}}} (X_1'^\top \cdot Y)^{\mathrm{R}}\right) P^\perp_{X_3''^\top}\right\|^2 , \end{aligned}$$

where we have used that a multiplication with $\frac{s_2}{n_3 - r_2}$ can only decrease the norm. The same is true for $P_{\hat{V}_3^\top}$ and thus

$$\begin{aligned} \left\|\tilde{Y}\right\|^2 &\geq \frac{s_2}{n_3 - r_2} \left(\left\|\mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1,r_2)}} Y\right\|^2 + \left\|P_{\hat{U}_1} \left(Y \cdot X_3''^\top\right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}}\right)^\top} \left(X_3'' \otimes I_{n_2}\right)\right\|^2 \right. \\ &\quad \left. + \left\|\left((P_{\hat{U}_1} \cdot Y)^{\mathrm{R}} + (I_{n_2} \otimes X_1')\, P^\perp_{X_2'^{\mathrm{R}}} (X_1'^\top \cdot Y)^{\mathrm{R}}\right) P_{\hat{V}_3^\top}\right\|^2\right) \\ &= \frac{s_2}{n_3 - r_2} \left\|\hat{Y}\right\|^2 . \end{aligned}$$

In Algorithm 1, the norm of the approximate projection increases monotonously. Thus, this lower bound is satisfied for any ε and $i_{\max}$. A similar derivation can be made if $s_2/(n_3 - r_2) \leq s_1/(n_1 - r_1)$. □

In Algorithm 1, the instruction "$[U, S, V] \leftarrow \mathrm{SVD}_s(A)$" means that $USV^\top$ is a truncated SVD of rank s of A. Since those SVDs are not necessarily unique, Algorithm 1 can output several (U_1, V_3) for a given input, and hence the approximate projection is set-valued.

Algorithm 1 Iterative Method to Obtain U_1 and V_3 of $\tilde{\mathcal{P}}_{T_X \mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}}$

Input: $Y \in \mathbb{R}^{n_1\times n_2\times n_3}$, $X \in \mathbb{R}^{n_1\times n_2\times n_3}_{(r_1,r_2)}$, $\varepsilon > 0$, $i_{\max}, s_1, s_2 \in \mathbb{N}_0$

$i \leftarrow 0$, $V_3 \leftarrow P^{\perp}_{X_3''^{\top}}$, $U_1 \leftarrow P^{\perp}_{X_1'^{\top}}$, $\eta_1 \leftarrow 0$, $\eta_{\text{new}} \leftarrow \infty$

if $s_2/(n_3 - r_2) > s_1/(n_1 - r_1)$ **then**

  while $i < i_{\max}$ **and** $|\eta_{\text{new}} - \eta_1| \leq \varepsilon$ **do**

    $\eta_1 \leftarrow \eta_{\text{new}}$, $i \leftarrow i + 1$

    $[U_1, \sim, \sim] \leftarrow \mathrm{SVD}_{s_1}\left(P^{\perp}_{X_1'}\left(\left(Y \cdot P_{V_3}\right)^{\mathrm{L}} + \left(Y \cdot X_3''^{\top}\right)^{\mathrm{L}} P^{\perp}_{X_2''^{\mathrm{L}\top}}\left(X_3'' \otimes I_{n_2}\right)\right)\right)$

    $\left[\sim, S, V_3^{\top}\right] \leftarrow \mathrm{SVD}_{s_2}\left(\left(\left(P_{U_1} \cdot Y\right)^{\mathrm{R}} + \left(I_{n_2} \otimes X_1'\right) P^{\perp}_{X_2'^{\mathrm{R}}}\left(X_1'^{\top} \cdot Y\right)^{\mathrm{R}}\right) P^{\perp}_{X_3''^{\top}}\right)$

    $\eta_{\text{new}} \leftarrow \|S\|^2 + \|P_{U_1}\left(Y \cdot X_3''^{\top}\right)^{\mathrm{L}} P^{\perp}_{X_2''^{\mathrm{L}\top}}\left(X_3'' \otimes I_{n_2}\right)\|^2$

else

  while $i < i_{\max}$ **and** $|\eta_{\text{new}} - \eta_1| \leq \varepsilon$ **do**

    $\eta_1 \leftarrow \eta_{\text{new}}$, $i \leftarrow i + 1$

    $\left[\sim, \sim, V_3^{\top}\right] \leftarrow \mathrm{SVD}_{s_2}\left(\left(\left(P_{U_1} \cdot Y\right)^{\mathrm{R}} + \left(I_{n_2} \otimes X_1'\right) P^{\perp}_{X_2'^{\mathrm{R}}}\left(X_1'^{\top} \cdot Y\right)^{\mathrm{R}}\right) P^{\perp}_{X_3''^{\top}}\right)$

    $[U_1, S, \sim] \leftarrow \mathrm{SVD}_{s_1}\left(P^{\perp}_{X_1'}\left(\left(Y \cdot P_{V_3}\right)^{\mathrm{L}} + \left(Y \cdot X_3''^{\top}\right)^{\mathrm{L}} P^{\perp}_{X_2''^{\mathrm{L}\top}}\left(X_3'' \otimes I_{n_2}\right)\right)\right)$

    $\eta_{\text{new}} \leftarrow \|S\|^2 + \|\left(I_{n_2} \otimes X_1'\right) P^{\perp}_{X_2'^{\mathrm{R}}}\left(X_1'^{\top} \cdot Y\right)^{\mathrm{R}} P_{V_3^{\top}}\|^2$

Output: U_1, V_3.

4 A Numerical Experiment

In this section, given a pair (X, Y) with $X \in \mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$ and $Y \in \mathbb{R}^{n_1\times n_2\times n_3}$, we compare the values of $\left\langle \frac{\tilde{Y}}{\|\tilde{Y}\|}, \frac{Y}{\|Y\|} \right\rangle$ obtained by computing the approximate projection $\tilde{Y}$ of Y onto $T_X \mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$ using Theorem 1, the formula from [5, §5.4], and the point output by the built-in MATLAB function `fmincon` applied to (3). The latter can be considered as a benchmark for the exact projection. Since $\|Y\| \geq \|\mathcal{P}_{T_X \mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}} Y\|$, (2) is satisfied if $\left\langle \frac{\tilde{Y}}{\|\tilde{Y}\|}, \frac{Y}{\|Y\|} \right\rangle \geq \omega$. For this experiment, we set $(k_1, k_2) := (3, 3)$ and generates fifty random pairs (X, Y), where $X \in \mathbb{R}^{5\times5\times5}_{(2,2)}$ and $Y \in \mathbb{R}^{5\times5\times5}$, using the built-in MATLAB function `randn`. For such pairs, the ω from (4) equals $\frac{1}{3}$. We use Algorithm 1 with $\varepsilon := 10^{-16}$, which implies that $i_{\max}$ is used as stopping criterion. In the left subfigure of Figure 1, the box plots for this experiment are shown for two values of $i_{\max}$. As can be seen, for both values of $i_{\max}$, the values of $\left\langle \frac{\tilde{Y}}{\|\tilde{Y}\|}, \frac{Y}{\|Y\|} \right\rangle$ obtained by the proposed approximate projection are close to those obtained by `fmincon` and are larger than those obtained by the approximate projection from [5, §5.4]. We observe that $\left\langle \frac{\tilde{Y}}{\|\tilde{Y}\|}, \frac{Y}{\|Y\|} \right\rangle$ is always larger than $\frac{1}{3}$, which suggests that (4) is a pessimistic estimate. The middle subfigure compares ten of the fifty pairs. For one of these pairs, the proposed method obtains a better result than `fmincon`. This is possible since the `fmincon` solver does not necessarily output a global solution because of the nonconvexity of (3). An advantage of the proposed approximate projection is that it requires less computation time than the `fmincon` solver (a fraction of a second for the former and up to ten seconds for the latter). In the rightmost subfigure, the evolution of $\eta_{\text{new}} - \eta_1$ is shown for one of the fifty pairs.

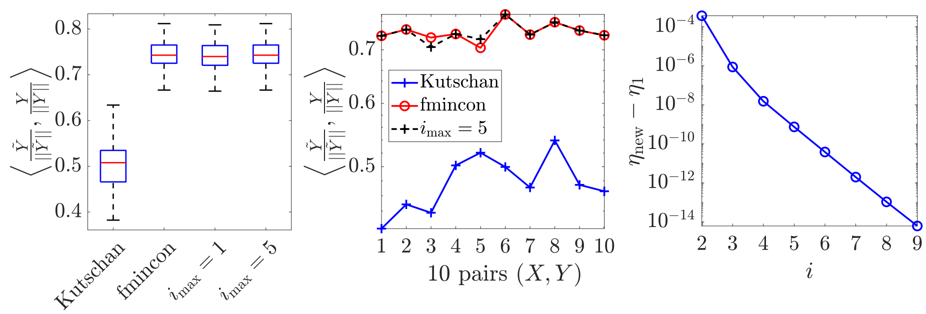

Fig. 1. A comparison of $\left\langle \frac{\tilde{Y}}{\|\tilde{Y}\|}, \frac{Y}{\|Y\|} \right\rangle$ for fifty randomly generated pairs (X, Y), with $X \in \mathbb{R}^{5\times5\times5}_{(2,2)}$, $Y \in \mathbb{R}^{5\times5\times5}$, and $(k_1, k_2) := (3, 3)$, for the approximate projection defined in Theorem 1, the one from [5, §5.4], and the one output by `fmincon`. On the rightmost figure, the evolution of $\eta_{\text{new}} - \eta_1$ is shown for one of the fifty pairs.

References

1. Hackbusch, W.: Tensor Spaces and Numerical Tensor Calculus, Springer Series in Computational Mathematics, vol. 56. Springer Cham, 2nd edn. (2019)
2. Holtz, S., Rohwedder, T., Schneider, R.: On manifolds of tensors of fixed TT-rank. Numerische Mathematik **120**(4), 701–731 (2012), https://doi.org/10.1007/s00211-011-0419-7
3. Kressner, D., Steinlechner, M., Uschmajew, A.: Low-rank tensor methods with subspace correction for symmetric eigenvalue problems. SIAM Journal on Scientific Computing **36**(5), A2346–A2368 (2014). https://doi.org/10.1137/130949919, https://doi.org/10.1137/130949919
4. Kutschan, B.: Tangent cones to tensor train varieties. Linear Algebra and its Applications **544**, 370–390 (2018). https://doi.org/10.1016/j.laa.2018.01.012
5. Kutschan, B.: Convergence of Gradient Methods on Hierarchical Tensor Varieties. Ph.D. thesis, TU Berlin (2019)
6. Levin, E., Kileel, J., Boumal, N.: Finding stationary points on bounded-rank matrices: A geometric hurdle and a smooth remedy. Mathematical Programming (2022). https://doi.org/10.1007/s10107-022-01851-2
7. Lubich, C., Oseledets, I.V., Vandereycken, B.: Time integration of tensor trains. SIAM Journal on Numerical Analysis **53**(2), 917–941 (2015), https://doi.org/10.1137/140976546

8. Oseledets, I.V.: Tensor-train decompositions. Methods and Algorithms for Scientific Computing **33**(5), 2295–2317 (2011), https://doi.org/10.1137/090752286
9. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis, Grundlehren der mathematischen Wissenschaften, vol. 317. Springer-Verlag Berlin Heidelberg (1998), corrected 3rd printing 2009
10. Schneider, R., Uschmajew, A.: Convergence results for projected line-search methods on varieties of low-rank matrices via łojasiewicz inequality. SIAM Journal on Optimization **25**(1), 622–646 (2015), https://doi.org/10.1137/140957822
11. Steinlechner, M.: Riemannian optimization for high-dimensional tensor completion. SIAM Journal on Scientific Computing **38**(5), S461–S484 (2016). https://doi.org/10.1137/15M1010506
12. Steinlechner, M.M.: Riemannian optimization for solving high-dimensional problems with low-rank tensor structure p. 165 (2016). https://doi.org/10.5075/epfl-thesis-6958