Rank Estimation for Third-Order Tensor Completion in the Tensor-Train Format

Charlotte Vermeylen
Dept. of Computer Science
KU Leuven
Heverlee, Belgium
charlotte.vermeylen@kuleuven.be

Guillaume Olikier
ICTEAM Institute
UCLouvain
Louvain-la-Neuve, Belgium
guillaume.olikier@uclouvain.be

P.-A. Absil
ICTEAM Institute
UCLouvain
Louvain-la-Neuve, Belgium
pa.absil@uclouvain.be

Marc Van Barel
Dept. of Computer Science
KU Leuven
Heverlee, Belgium
marc.vanbarel@kuleuven.be

***Abstract*—We propose a numerical method to obtain an adequate value for the upper bound on the rank for the tensor completion problem on the variety of third-order tensors of bounded tensor-train rank. The method is inspired by the parametrization of the tangent cone derived by Kutschan (2018). A proof of the adequacy of the upper bound for a related low-rank tensor approximation problem is given and an estimated rank is defined to extend the result to the low-rank tensor completion problem. Some experiments on synthetic data illustrate the approach and show that the method is very robust, e.g., to noise on the data.**

***Index Terms*—tensor-train, tensor completion, rank estimation, tangent cone**

I. Introduction

We consider the *low-rank tensor completion problem (LRTCP)* formulated as a least squares optimization problem on the algebraic variety $\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$ [1, Definition 1.4] of $n_1 \times n_2 \times n_3$ real third-order tensors of tensor-train (TT) rank at most (k_1, k_2):

$$\min_{X\in\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}} \underbrace{\frac{1}{2}\|X_\Omega - A_\Omega\|^2}_{=:f_\Omega(X)}, \tag{1}$$

where $A \in \mathbb{R}^{n_1\times n_2\times n_3}$, $\Omega \subseteq \{1,\dots,n_1\} \times \{1,\dots,n_2\} \times \{1,\dots,n_3\}$ is called the sampling set,

$$Z_\Omega(i_1,i_2,i_3) := \begin{cases} Z(i_1,i_2,i_3) & \text{if } (i_1,i_2,i_3)\in\Omega,\\ 0 & \text{otherwise,}\end{cases}$$

for all $Z \in \mathbb{R}^{n_1\times n_2\times n_3}$, and the norm is induced by the inner product

$$\langle Y, X\rangle = \langle \mathrm{vec}(Y), \mathrm{vec}(X)\rangle, \quad \forall X, Y \in \mathbb{R}^{n_1\times n_2\times n_3}. \tag{2}$$

A *tensor-train decomposition* (TTD) of a third-order tensor $X \in \mathbb{R}^{n_1\times n_2\times n_3}$ is a factorization $X = X_1 \cdot X_2 \cdot X_3$, where $X_1 \in \mathbb{R}^{n_1\times r_1}$, $X_2 \in \mathbb{R}^{r_1\times n_2\times r_2}$, and $X_3 \in \mathbb{R}^{r_2\times n_3}$ [2]. The '$\cdot$' indicates the contraction between a matrix and a tensor. They interact with the left and right unfolding of X_2,

$$X_2^{\mathrm{L}} := [X_2]^{r_1\times n_2 r_2} := \mathrm{reshape}(X_2, r_1\times n_2 r_2),$$
$$X_2^{\mathrm{R}} := [X_2]^{r_1 n_2\times r_2} := \mathrm{reshape}(X_2, r_1 n_2\times r_2),$$

in the following way:

$$X_1\cdot X_2 = \left[X_1 X_2^{\mathrm{L}}\right]^{n_1\times n_2\times r_2}, X_2\cdot X_3 = \left[X_2^{\mathrm{R}} X_3\right]^{r_1\times n_2\times n_3}.$$

The minimal r_1 and r_2 for which a TTD of X exists, is called the *TT-rank* of X. For second-order tensors (matrices), the TT-rank reduces to the usual matrix rank, and since no other definition of tensor rank is used in this paper, it is simply denoted by $\operatorname{rank} X$ and can be determined as

$$\operatorname{rank} X = \left(\operatorname{rank} X^{\mathrm{L}}, \operatorname{rank} X^{\mathrm{R}}\right) =: (r_1, r_2). \tag{3}$$

The minimal rank decomposition can be obtained by computing successive singular value decompositions (SVDs) of the unfoldings [2, Algorithm 1].

The low-rank variety can then be defined as

$$\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)} := \{X \in \mathbb{R}^{n_1\times n_2\times n_3} \mid \operatorname{rank} X \leq (k_1,k_2)\} \tag{4}$$

and the fixed-rank smooth manifold [3] as

$$\mathbb{R}^{n_1\times n_2\times n_3}_{(r_1,r_2)} := \{X \in \mathbb{R}^{n_1\times n_2\times n_3} \mid \operatorname{rank} X = (r_1,r_2)\}. \tag{5}$$

In practical LRTCPs, the rank of A is not known or A has full rank due to noise. This is notably the case for movie rating recommendation systems [4], where, e.g., the ratings of different users over time (or any other variable of interest) form samples of a large third-order tensor. Evaluating a solution to (1) in elements outside the set Ω allows us to recommend movies with a high estimated rating. Note that evaluating one element of a third-order TTD corresponds to performing a vector-matrix-vector multiplication and can be done efficiently in $\mathcal{O}(r_1 r_2)$ operations.

When k_1 and k_2 are set too high however, the complexity of an algorithm to solve (1) is unnecessarily high and furthermore overfitting can occur, i.e., X approximates A_Ω well but not the full tensor A. To detect overfitting, usually a test data set Γ is used [5]. When the error on this test set increases during optimization while the error of (1) decreases overfitting has occurred and the algorithm should be stopped or the rank decreased. On the other hand, when k_1 or k_2 are set too low, the search space may not contain a sufficiently good approximation of A. It is thus important to choose adequate values for k_1 and k_2.

Intuitively the smaller $|\Omega|$, the more difficult it is to recover A from A_Ω by solving (1). However, the minimal number of samples needed is not known [6].

In this work, a method to estimate the rank of A from A_Ω is proposed. When A is not exactly low rank, a good value for a

low-rank approximation is obtained. This method can then be used, e.g., in a rank-adaptive optimization algorithm to solve (1).

The paper is organized as follows. First, in section II, the preliminaries for section IV and section V are given. This includes some basic facts concerning orthogonal projections onto vector spaces and arithmetic rules for TTDs. For a more extensive overview of properties of the TTD, we refer to the original paper [2] and the notation introduced in [7], which is also used in [5], [8]. In section III, the auxiliary low-rank tensor approximation problem (LRTAP) is defined. In section IV, the parametrization of the tangent cone to the low-rank variety is given [1]. New orthogonality conditions are derived to ensure that in Proposition 2 no matrix inverse is needed, which improves the stability of the proposed method, and makes the proofs in the rest of the paper easier. In section V, the main proposition is derived, and an estimated rank is defined to extend our result to the LRTCP. Lastly, in section VI, some experiments illustrate the use and advantages of the proposed rank estimation method for the LRTCP.

II. Preliminaries

A TTD is not unique. Orthogonality conditions can be enforced, which can improve the stability of algorithms working with TTDs. Those used in this paper are introduced in section IV.

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. A tensor is said to be *left-orthogonal* if $n_1 \leq n_2 n_3$ and $\left(X^{\mathrm{L}}\right)^\top \in \mathrm{St}(n_1, n_2 n_3)$, and *right-orthogonal* if $n_3 \leq n_1 n_2$ and $X^{\mathrm{R}} \in \mathrm{St}(n_3, n_1 n_2)$.

The following properties and arithmetic rules are used frequently in the rest of the paper.

- For all matrices Y and Z, it holds that $Y \cdot (X_1 \cdot X_2 \cdot X_3) \cdot Z = Y X_1 \cdot X_2 \cdot X_3 Z$.
- The left and right unfoldings of $X = X_1 \cdot X_2 \cdot X_3$ can be rewritten as:

$$\begin{aligned} X^{\mathrm{L}} &= X_1 \left(X_2 \cdot X_3\right)^{\mathrm{L}} = X_1 X_2^{\mathrm{L}} \left(X_3 \otimes I_{n_2}\right), \\ X^{\mathrm{R}} &= \left(X_1 \cdot X_2\right)^{\mathrm{R}} X_3 = \left(I_{n_2} \otimes X_1\right) X_2^{\mathrm{R}} X_3, \end{aligned} \tag{6}$$

  where '$\otimes$' denotes the Kronecker product.
- From (3) and (6) it can be deduced that

$$\begin{aligned} \mathrm{rank}(X_1) &= \mathrm{rank}\left(X_2^{\mathrm{L}} \left(X_3 \otimes I_{n_2}\right)\right) = r_1, \\ \mathrm{rank}(X_3) &= \mathrm{rank}\left(\left(I_{n_2} \otimes X_1\right) X_2^{\mathrm{R}}\right) = r_2, \end{aligned} \tag{7}$$

  and because the ranks of $I_{n_2} \otimes X_1$ and $X_3 \otimes I_{n_2}$ are $n_2 r_1$ and $r_2 n_2$, respectively, (7) can be simplified to $\mathrm{rank}\left(X_2^{\mathrm{L}}\right) = r_1$ and $\mathrm{rank}\left(X_2^{\mathrm{R}}\right) = r_2$.
- Orthogonality between TTDs is exploited frequently in the parametrization of the tangent cone in section IV and the proofs in section V. If $Y = Y_1 \cdot Y_2 \cdot Y_3$ and $Z = Z_1 \cdot Z_2 \cdot Z_3$, then by using (6), the inner product $\langle Y, Z\rangle$ is zero if at least one of the following equalities holds:

$$\begin{aligned} Y_1^\top Z_1 &= 0, & \left(Y_2 \cdot Y_3\right)^{\mathrm{L}} \left(\left(Z_2 \cdot Z_3\right)^{\mathrm{L}}\right)^\top &= 0, \\ Y_3 Z_3^\top &= 0, & \left(\left(Y_1 \cdot Y_2\right)^{\mathrm{R}}\right)^\top \left(Z_1 \cdot Z_2\right)^{\mathrm{R}} &= 0. \end{aligned} \tag{8}$$

III. Low-Rank Tensor Approximation

The *low-rank tensor approximation problem* (LRTAP) is defined as:

$$\min_{X \in \mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}} \underbrace{\frac{1}{2}\|X - A\|^2}_{=:f(X)}. \tag{9}$$

This problem is related to the LRTCP (1) because $f_\Omega(X) = f(X)$ for $\Omega = \{1,\ldots,n_1\} \times \{1,\ldots,n_2\} \times \{1,\ldots,n_3\}$. Remark that, as for (1), a global minimizer is, in general, not unique because $\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$ is nonconvex and NP-hard to obtain [9]. This problem is used in section V.

IV. Tangent Cone

The set of all tangent vectors to $\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$ at $X = X_1' \cdot X_2' \cdot X_3 \in \mathbb{R}^{n_1\times n_2\times n_3}_{(r_1,r_2)}$, where $X_1' \in \mathrm{St}(r_1, n_1)$, ${X_2'}^{\mathrm{R}} \in \mathrm{St}(r_2, r_1 n_2)$, and $k_1 \geq r_1$, $k_2 \geq r_2$, is a closed cone called the tangent cone to $\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$ at X and denoted by $T_X\mathbb{R}^{n_1\times n_2\times n_3}_{\leq(k_1,k_2)}$. By [1, 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

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

where $U_1 \in \mathbb{R}^{n_1\times s_1}$, $W_1 \in \mathbb{R}^{n_1\times r_1}$, $U_2 \in \mathbb{R}^{r_1\times n_2\times s_2}$, $W_2 \in \mathbb{R}^{r_1\times n_2\times r_2}$, $Z_2 \in \mathbb{R}^{s_1\times n_2\times s_2}$, $V_2 \in \mathbb{R}^{s_1\times n_2\times r_2}$, $W_3 \in \mathbb{R}^{r_2\times n_3}$, $V_3 \in \mathbb{R}^{s_2\times n_3}$, and $s_i = k_i - r_i$. As for the TTD itself, this parametrization is not unique. In [1], the following orthogonality conditions are derived:

$$\begin{aligned} U_1^\top X_1' &= 0, & W_1^\top X_1' &= 0, \\ \left(U_2^{\mathrm{R}}\right)^\top {X_2'}^{\mathrm{R}} &= 0, & \left(V_2 \cdot X_3\right)^{\mathrm{L}} \left(\left(X_2' \cdot X_3\right)^{\mathrm{L}}\right)^\top &= 0, \\ \left(W_2^{\mathrm{R}}\right)^\top {X_2'}^{\mathrm{R}} &= 0, & V_3 X_3^\top &= 0. \end{aligned}$$

We change these orthogonality conditions slightly to make the proofs in the rest of this paper easier and the computations in the experiments more stable. To do so, we notice that

$$G = \begin{bmatrix} X_1' & U_1 & \dot{W}_1 \end{bmatrix} \cdot \begin{bmatrix} X_2' & U_2 & \dot{W}_2 \\ 0 & Z_2 & \dot{V}_2 \\ 0 & 0 & X_2'' \end{bmatrix} \cdot \begin{bmatrix} W_3 \\ V_3 \\ X_3'' \end{bmatrix},$$

where we have defined $\dot{W}_1 := W_1 R^{-1}$, $\dot{W}_2 := W_2 \cdot C$, $\dot{V}_2 := V_2 \cdot C$, $X_2'' := R \cdot X_2' \cdot C$, $X_3'' := C^{-1} X_3$, and $R \in \mathbb{R}^{r_1\times r_1}$ and $C \in \mathbb{R}^{r_2\times r_2}$ are chosen such that X_2'' and X_3'' are left-orthogonal. Thus, we can also define $X_1 := X_1' R^{-1}$ and $X_2 := X_2' \cdot C$, such that $X = X_1 \cdot X_2'' \cdot X_3'' = X_1' \cdot X_2 \cdot X_3''$. Additionally, W_3 can be decomposed as $W_3 = W_3 {X_3''}^\top X_3'' + \hat{W}_3$.

The two terms involving W_3 and $\dot{W}_2$ can then be regrouped as

$$X_1' \cdot \left(X_2' \cdot W_3 X_3''^\top + \dot{W}_2 \right) \cdot X_3'' + X_1' \cdot X_2' \cdot \hat{W}_3.$$

Defining $\tilde{W}_2 := X_2' \cdot W_3 X_3''^\top + \dot{W}_2$, we obtain

$$G = \begin{bmatrix} X_1' & U_1 & \dot{W}_1 \end{bmatrix} \cdot \begin{bmatrix} X_2' & U_2 & \tilde{W}_2 \\ 0 & Z_2 & \dot{V}_2 \\ 0 & 0 & X_2'' \end{bmatrix} \cdot \begin{bmatrix} \hat{W}_3 \\ V_3 \\ X_3'' \end{bmatrix}, \quad (10)$$

with the modified orthogonality conditions

$$\begin{aligned} U_1^\top X_1' = 0, \quad \dot{W}_1^\top X_1' = 0, \quad \left(U_2^{\mathrm{R}}\right)^\top X_2'^{\mathrm{R}} = 0, \\ \hat{W}_3 X_3''^\top = 0, \quad V_3 X_3''^\top = 0, \quad \dot{V}_2^{\mathrm{L}} \left(X_2''^{\mathrm{L}}\right)^\top = 0. \end{aligned} \quad (11)$$

Expanding (10), a sum of 6 mutually orthogonal TTDs is obtained because of (8) and (11).

The projection onto the closed cone $T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (k_1, k_2)}$ is not known and, in general, difficult to obtain because it is nonlinear and nonconvex. However, in what follows we show that any tensor $Y \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an element of $T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (n_1, n_3)}$. Straightforward computations show that

$$\begin{aligned} Y &= P_{X_1'} \cdot Y \cdot P_{X_3''^\top} + P^\perp_{X_1'} \cdot Y \cdot P_{X_3''^\top} \\ &\quad + P_{X_1'} \cdot Y \cdot P^\perp_{X_3''^\top} + P^\perp_{X_1'} \cdot Y \cdot P^\perp_{X_3''^\top} \\ &= P_{X_1'} \cdot Y \cdot P_{X_3''^\top} \\ &\quad + \left[P^\perp_{X_1'} \left(Y \cdot X_3''^\top \right)^{\mathrm{L}} P_{\left(X_2''^{\mathrm{L}}\right)^\top} \right]^{n_1 \times n_2 \times r_2} \cdot X_3'' \\ &\quad + \left[P^\perp_{X_1'} \left(Y \cdot X_3''^\top \right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}}\right)^\top} \right]^{n_1 \times n_2 \times r_2} \cdot X_3'' \\ &\quad + X_1' \cdot \left[P_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot Y \right)^{\mathrm{R}} P^\perp_{X_3''^\top} \right]^{r_1 \times n_2 \times n_3} \\ &\quad + X_1' \cdot \left[P^\perp_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot Y \right)^{\mathrm{R}} P^\perp_{X_3''^\top} \right]^{r_1 \times n_2 \times n_3} \\ &\quad + P^\perp_{X_1'} \cdot Y \cdot P^\perp_{X_3''^\top}. \end{aligned} \quad (12)$$

Thus, $Y = Y_1 \cdot Y_2 \cdot Y_3 \in \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (n_1, n_3)}$ can be parameterized as in (10), with

$$\begin{aligned} U_1 &= P^\perp_{X_1'} Y_1, \\ U_2 &= \left[P^\perp_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot Y_1 \cdot Y_2 \right)^{\mathrm{R}} \right]^{r_1 \times n_2 \times n_3}, \\ \dot{V}_2 &= \left[\left(Y_2 \cdot Y_3 \cdot X_3''^\top \right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}}\right)^\top} \right]^{n_1 \times n_2 \times r_2}, \\ V_3 &= Y_3 P^\perp_{X_3''^\top}, \\ \dot{W}_1 &= \left(P^\perp_{X_1'} \cdot Y \cdot X_3''^\top \right)^{\mathrm{L}} \left(X_2''^{\mathrm{L}} \right)^\top, \\ \tilde{W}_2 &= X_1'^\top \cdot Y \cdot X_3''^\top, \\ \hat{W}_3 &= \left(X_2'^{\mathrm{R}} \right)^\top \left(X_1'^\top \cdot Y \cdot P^\perp_{X_3''^\top} \right)^{\mathrm{R}}, \\ Z_2 &= Y_2, \end{aligned} \quad (13)$$

which satisfies (11). Furthermore, U_1 and V_3 have rank at most $n_1 - r_1$ and $n_3 - r_2$, respectively, and thus the matrix $U_1 \dot{V}_2^{\mathrm{L}}$ has rank $n_1 - r_1$, and U_1 and $\dot{V}_2^{\mathrm{L}}$ can be reduced to size $n_1 \times (n_1 - r_1)$ and $(n_1 - r_1) \times n_2 r_2$, respectively, e.g., by computing the SVD of $U_1 \dot{V}_2^{\mathrm{L}}$. Similarly, this can be done for $U_2^{\mathrm{R}} V_3$ to obtain U_2^{R} and V_3 of size $r_1 n_2 \times (n_3 - r_2)$ and $(n_3 - r_2) \times n_3$, respectively. Then, Z_2 can be changed accordingly to $U_1^\top \cdot Y_2 \cdot V_3^\top$, which is the same result as would be obtained by the TT-rounding algorithm [2, Algorithm 2], except for the orthogonality conditions. Thus, Y can be written in the form (10) with $s_1 = n_1 - r_1$, $s_2 = n_3 - r_2$ and hence by definition $Y \in T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (n_1, n_3)}$.

The Riemannian gradient of (9) at $X \in \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1, r_2)}$ is defined as the projection of the Euclidean gradient $\nabla f(X) = X - A$ onto the tangent space [3]:

$$\begin{aligned} &T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1, r_2)} := \\ &\left\{ \; X_1' \cdot X_2' \cdot \hat{W}_3 + X_1' \cdot \tilde{W}_2 \cdot X_3'' + \dot{W}_1 \cdot X_2'' \cdot X_3'' \; \right\}. \end{aligned}$$

By replacing Y by $X - A$ in (13), the parameters of $\mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1, r_2)}} \nabla f(X)$ are:

$$\begin{aligned} \dot{W}_1 &= - \left(P^\perp_{X_1'} \cdot A \cdot X_3''^\top \right)^{\mathrm{L}} \left(X_2''^{\mathrm{L}} \right)^\top, \\ \tilde{W}_2 &= X_2 - X_1'^\top \cdot A \cdot X_3''^\top, \\ \hat{W}_3 &= - \left(X_2'^{\mathrm{R}} \right)^\top \left(X_1'^\top \cdot A \cdot P^\perp_{X_3''^\top} \right)^{\mathrm{R}}. \end{aligned} \quad (14)$$

A similar projection onto the tangent space was used in [5] and [10], but with different orthogonality conditions.

V. Rank Estimation

Proposition 2 states the main result for the LRTAP (9). Afterwards, the estimated rank (16) is defined to extend this result to the LRTCP (1). To prove Proposition 2, the following auxiliary lemma is used.

Lemma 1. *Let* $X = X_1' \cdot X_2 \cdot X_3'' = X_1' \cdot X_2' \cdot X_3 = X_1 \cdot X_2'' \cdot X_3'' \in \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1, r_2)}$, *with* $X_1' \in \mathrm{St}(r_1, n_1)$, $X_2'^{\mathrm{R}} \in \mathrm{St}(r_2, r_1 n_2)$, $\left(X_2''^{\mathrm{L}}\right)^\top \in \mathrm{St}(r_1, n_2 r_2)$, *and* $X_3''^\top \in \mathrm{St}(r_2, n_3)$, *and let* $A = A_1' \cdot A_2 \cdot A_3'' \in \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1', r_2')}$, *with* $A_1' \in \mathrm{St}(r_1', n_1)$ *and* $A_3''^\top \in \mathrm{St}(r_2', n_3)$. *If* $\mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1, r_2)}} \nabla f(X) = 0$, *then*

$$X_1' = A_1' B_1, \quad X_3'' = B_3 A_3'', \quad X_2 = B_1^\top \cdot A_2 \cdot B_3^\top,$$

for some $B_1 \in \mathrm{St}(r_1, r_1')$ *and* $B_3^\top \in \mathrm{St}(r_2, r_2')$.

Proof. From (14), $\mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1, r_2)}} \nabla f(X) = 0$ if and only if $\dot{W}_1 = 0$, $\tilde{W}_2 = 0$, and $\hat{W}_3 = 0$. From the second equation in (14), it is clear that $\tilde{W}_2$ can only be zero if $X_2 = X_1'^\top \cdot A \cdot X_3''^\top$. The matrices X_1' and X_3'' are decomposed as

$$\begin{aligned} X_1' &= \begin{bmatrix} A_1' & A_1'^\perp \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = A_1' B_1 + A_1'^\perp B_2, \\ X_3'' &= \begin{bmatrix} B_3 & B_4 \end{bmatrix} \begin{bmatrix} A_3'' \\ A_3''^\perp \end{bmatrix} = B_3 A_3'' + B_4 A_3''^\perp, \end{aligned} \quad (15)$$

where $\begin{bmatrix} A_1' & A_1'^\perp \end{bmatrix} \in \mathrm{St}(n_1, n_1)$, $\begin{bmatrix} A_3'' \\ A_3''^\perp \end{bmatrix} \in \mathrm{St}(n_3, n_3)$, $\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \in \mathrm{St}(r_1, r_1 + r_1')$, and $\begin{bmatrix} B_3^\top \\ B_4^\top \end{bmatrix} \in \mathrm{St}(r_2, r_2 + r_2')$. Substituting (15) into the equation for X_2, we obtain $X_2 = B_1^\top \cdot A_2 \cdot B_3^\top$.

Substituting (15) into the equation for $\dot{W}_1$ in (14), we get

$$\dot{W}_1 = 0 \Leftrightarrow \Big(\Big(I_{n_1} - A_1' B_1 B_1^\top A_1'^\top - A_1'^\perp B_2 B_1^\top A_1'^\top - A_1' B_1 B_2^\top \left(A_1'^\perp\right)^\top - A_1'^\perp B_2 B_2^\top \left(A_1'^\perp\right)^\top\Big) A_1' \cdot A_2 \cdot A_3'' \Big(A_3''^\top B_3^\top + \left(A_3''^\perp\right)^\top B_4^\top\Big)\Big)^{\mathrm{L}} \left(X_2''^{\mathrm{L}}\right)^\top = 0 \Leftrightarrow \left(\left(A_1' - A_1' B_1 B_1^\top - A_1'^\perp B_2 B_1^\top\right) \cdot A_2 \cdot B_3^\top\right)^{\mathrm{L}} \left(X_2''^{\mathrm{L}}\right)^\top = 0.$$

Multiplying both sides by $\left(A_1'^\perp\right)^\top$, we obtain

$$\begin{aligned}\dot{W}_1 = 0 &\Rightarrow B_2 B_1^\top A_2^{\mathrm{L}} \left(B_3^\top \otimes I_{n_2}\right) \left(X_2''^{\mathrm{L}}\right)^\top = 0 \\ &\Leftrightarrow B_2 X_2{}^{\mathrm{L}} \left(X_2''^{\mathrm{L}}\right)^\top = 0 \\ &\Leftrightarrow B_2 R^{-1} X_2''^{\mathrm{L}} \left(X_2''^{\mathrm{L}}\right)^\top = 0 \Leftrightarrow B_2 = 0.\end{aligned}$$

Thus, it holds that $X_1' = A_1' B_1$ with $B_1 \in \mathrm{St}(r_1, r_1')$. From $\hat{W}_3 = 0$ in (14), it can similarly be derived that $B_4 = 0$. Hence, $X_3'' = B_3 A_3''$ and $B_3^\top \in \mathrm{St}(r_2, r_2')$. □

The two equalities in Proposition 2 enable to deduce the TT-rank of A—and thus a value of (k_1, k_2) for which the optimum of LRTAP (9) is zero—from any stationary point of $\min_{X \in \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1, r_2)}} f(X)$.

Proposition 2. *If the same conditions as in Lemma 1 hold, then* $\nabla f(X) \in T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (r_1', r_2')}$*, and*

$$\left(\mathrm{rank}\left(\left(\nabla f(X) \cdot X_3''^\top\right)^{\mathrm{L}}\right), \mathrm{rank}\left(\left(X_1'^\top \cdot \nabla f(X)\right)^R\right)\right) = \left(r_1' - r_1, r_2' - r_2\right).$$

Proof. By decomposing $\nabla f(X)$ as Y in (12) and by setting $\dot{W}_1$, $\tilde{W}_2$, and $\hat{W}_3$ in (13) to zero because $\mathcal{P}_{T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1, r_2)}} \nabla f(X) = 0$, we obtain

$$\begin{aligned}\nabla f(X) = {} & P^\perp_{X_1'} \cdot \nabla f(X) \cdot P^\perp_{X_3''^\top} \\ & + X_1' \cdot \left[P^\perp_{X_2'^{\mathrm{R}}} \left(X_1'^\top \cdot \nabla f(X)\right)^{\mathrm{R}}\right]^{r_1 \times n_2 \times n_3} \cdot P^\perp_{X_3''^\top} \\ & + P^\perp_{X_1'} \cdot \left[\left(\nabla f(X) \cdot X_3''^\top\right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}}\right)^\top}\right]^{n_1 \times n_2 \times r_2} \cdot X_3''.\end{aligned}$$

Multiplying both sides on the right by $X_3''^\top$, we obtain

$$\nabla f(X) \cdot X_3''^\top = P^\perp_{X_1'} \cdot \left[\left(\nabla f(X) \cdot X_3''^\top\right)^{\mathrm{L}} P^\perp_{\left(X_2''^{\mathrm{L}}\right)^\top}\right]^{n_1 \times n_2 \times r_2} = U_1 \cdot \dot{V}_2,$$

with U_1 and $\dot{V}_2$ as in (13) with Y replaced by $\nabla f(X)$. Furthermore, from Lemma 1, we know that $X_1' = A_1' B_1$, $X_3'' = B_3 A_3''$, and $X_2 = B_1^\top \cdot A_2 \cdot B_3^\top$, for some $B_1 \in \mathrm{St}(r_1, r_1')$ and $B_3^\top \in \mathrm{St}(r_2, r_2')$. Therefore,

$$\begin{aligned}&\mathrm{rank}\left(\left(\nabla f(X) \cdot X_3''^\top\right)^{\mathrm{L}}\right) \\ &= \mathrm{rank}\left(\left(X_1' \cdot X_2 - A_1' \cdot A_2 \cdot A_3'' \cdot X_3''^\top\right)^{\mathrm{L}}\right) \\ &= \mathrm{rank}\left(\left(A_1' B_1 \cdot B_1^\top \cdot A_2 \cdot B_3^\top - A_1' \cdot A_2 \cdot B_3^\top\right)^{\mathrm{L}}\right) \\ &= \mathrm{rank}\left(A_1' \left(B_1 B_1^\top - I_{r_1'}\right) \left(A_2 \cdot B_3^\top\right)^{\mathrm{L}}\right) \\ &= \mathrm{rank}\left(I_{r_1'} - B_1 B_1^\top\right) = r_1' - r_1\end{aligned}$$

because $\left(A_2 \cdot B_3^\top\right)^{\mathrm{L}}$ has full rank r_1', knowing that $X_2^{\mathrm{L}} = B_1^\top \left(A_2 \cdot B_3^\top\right)^{\mathrm{L}}$ has rank r_1 and using the Sylvester rank inequality. Thus, $\mathrm{rank}\left(U_1 \dot{V}_2^{\mathrm{L}}\right) = r_1' - r_1$. A similar derivation can be made for $\left(X_1'^\top \cdot \nabla f(X)\right)^{\mathrm{R}} = U_2^{\mathrm{R}} V_3$. Hence, $\nabla f(X)$ can be parameterized as in (10) with $s_1 = r_1' - r_1$, $s_2 = r_2' - r_2$), and thus by definition $\nabla f(X) \in T_X \mathbb{R}^{n_1 \times n_2 \times n_3}_{\leq (r_1', r_2')}$. □

We propose to exploit the two equalities from Proposition 2 in the context of LRTCP (1) by using the *estimated rank* of $B \in \mathbb{R}^{n \times m}$ which is inspired by [11] and defined as:

$$\tilde{r}_s(B) := \begin{cases} 0 & \text{if } B = 0, \\ \mathrm{argmax}_{j \leq s} \frac{\sigma_j(B) - \sigma_{j+1}(B)}{\sigma_j(B)} & \text{otherwise,} \end{cases} \tag{16}$$

where $\sigma_j(B)$, $j = 1 \ldots \mathrm{rank}(B)$, denote the singular values of B in decreasing order, i.e., $\sigma_i(B) \geq \sigma_j(B)$ for $i \leq j$, and $s < \mathrm{rank}(B)$. The upper bound s prevents the estimated rank from being too high and should be chosen by the user. Thus, we propose $\left(\tilde{r}_s\left(\left(\nabla f_\Omega(X) \cdot X_3''^\top\right)^{\mathrm{L}}\right), \tilde{r}_s\left(\left(X_1'^\top \cdot \nabla f_\Omega(X)\right)^R\right)\right)$ as an adequate value for (k_1, k_2) in (1), where $X = X_1' \cdot X_2 \cdot X_3''$ has been obtained by running a Riemannian optimization algorithm on

$$\min_{X \in \mathbb{R}^{n_1 \times n_2 \times n_3}_{(r_1, r_2)}} \frac{1}{2} \|X_\Omega - A_\Omega\|^2. \tag{17}$$

VI. Experiments

In this section, three experiments are generated where we have optimized (17) for $(r_1, r_2) := (2, 2)$, $n_1 := n_2 := n_3 := 100$, and $(r_1', r_2') := (6, 6)$, using a Riemannian conjugate gradient (CG) algorithm [5], [12]. The tensor A and the starting point X_0 given to the optimization algorithm are generated as follows:

$$A = \texttt{randn}(n_1, r_1') \cdot \texttt{randn}(r_1', n_2, r_2') \cdot \texttt{randn}(r_2', n_3),$$
$$X_0 = \texttt{randn}(n_1, r_1) \cdot \texttt{randn}(r_1, n_2, r_2) \cdot \texttt{randn}(r_2, n_3),$$

where `randn` is a built-in MATLAB function to generate pseudo-random numbers. It can be shown that the elements of A, generated in this way, have standard deviation $\sqrt{r_1' r_2'} = 6$. To obtain A_Ω, $4 \cdot 10^4$ random samples of this tensor were generated. An illustration of how the estimated rank of $\left(\nabla f_\Omega(X) \cdot X_3''^\top\right)^{\mathrm{L}}$ can be used to estimate a good value for (k_1, k_2) in (1) is given in Figure 1. The first 20 singular values of $\left(\nabla f_\Omega(X) \cdot X_3''^\top\right)^{\mathrm{L}}$ are shown in the left upper subfigure. The squared norm of the Riemannian gradient that is obtained

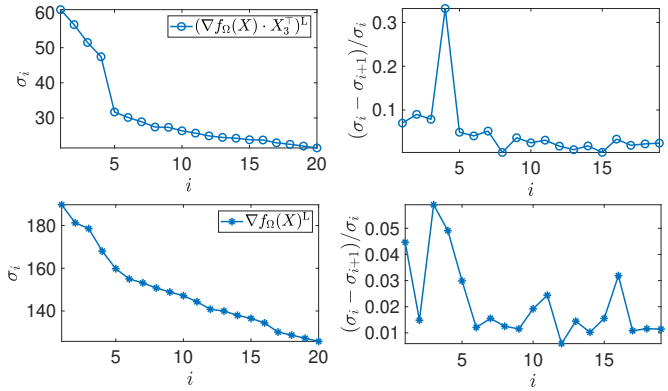

Figure 1. An illustration of the advantage of $\tilde{r}_{20}\left(\left(\nabla f_\Omega(X)\cdot X_3''^\top\right)^{\mathrm{L}}\right)$ compared to $\tilde{r}_{20}(\nabla f_\Omega(X)^{\mathrm{L}})$, to estimate the rank of A, for $\|\mathcal{P}_{T_X\mathbb{R}^{n_1\times n_2\times n_3}_{(r_1,r_2)}}\nabla f_\Omega(X)\|^2 = 10^{-8}$, obtained after 200 iterations.

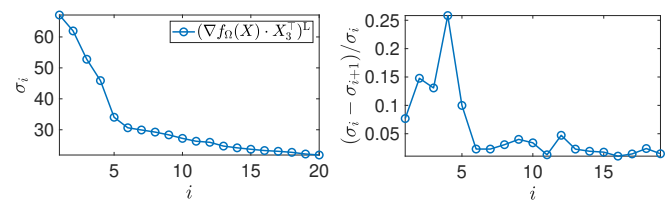

Figure 2. The first 20 singular values of $\left(\nabla f_\Omega(X)\cdot X_3''^\top\right)^{\mathrm{L}}$ (left) and their relative gap (right), for $\|\mathcal{P}_{T_X\mathbb{R}^{n_1\times n_2\times n_3}_{(r_1,r_2)}}\nabla f_\Omega(X)\|^2 = 504$, obtained after 10 iterations.

at X is approximately 10^{-8}. There were 200 iterations needed to obtain this accuracy. Based on the upper right subfigure, where the relative gap between the singular values is shown, it can be seen that the estimated rank equals $r_1' - r_1$. In the lower two subfigures, the first 20 singular values of $\nabla f_\Omega(X)^{\mathrm{L}}$ are shown. The estimated rank of $\nabla f_\Omega(X)^{\mathrm{L}}$ equals 3, and thus cannot be used to estimate the rank of A.

In Figure 2, it is shown that in practice the norm of the Riemannian gradient does not need to be very small for the estimated rank of $\left(\nabla f_\Omega(X)\cdot X_3''^\top\right)^{\mathrm{L}}$ to equal $r_1' - r_1$. In this experiment, only 10 iterations of the Riemannian CG algorithm where used, such that the squared norm of the Riemannian gradient was approximately 504. However the estimated rank still equals $r_1' - r_1$.

In a last experiment, another advantage of the proposed method is illustrated. For this experiment noise with $\eta = 10$ is added to the low-rank tensor as follows:

$$A_\eta = A + \eta \ \ \mathtt{randn}\,(n_1, n_2, n_3)\,. \tag{18}$$

This means that the noise has the same magnitude as A but the estimated rank of $\left(\nabla f_\Omega(X)\cdot X_3''^\top\right)^{\mathrm{L}}$ still equals $r_1' - r_1 = 4$ after 120 iterations of the CG algorithm, as shown in Figure 3. The squared norm of the Riemannian gradient equals 0.9.

VII. Conclusion

The two equalities given in Proposition 2 enable to compute the TT-rank of A based on a stationary point of LRTAP on the fixed-rank manifold, i.e., $\min_{X\in\mathbb{R}^{n_1\times n_2\times n_3}_{(r_1,r_2)}} \frac{1}{2}\|X - A\|^2$, which can be obtained using classic Riemannian optimization. Moreover, numerical experiments indicate that, for LRTCP (1), using these equalities with the rank replaced by the estimated rank (16) provides a plausible estimation of the TT-rank of A which can be used as an adequate value for (k_1, k_2).

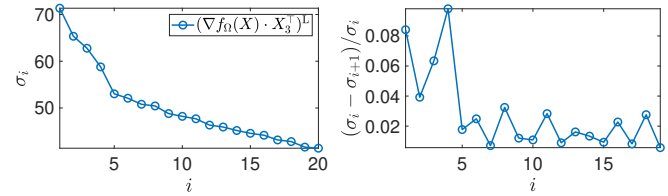

Figure 3. The first 20 singular values of $\left(\nabla f_\Omega(X)\cdot X_3''^\top\right)^{\mathrm{L}}$ (left) and their relative gap (right), for $\|\mathcal{P}_{T_X\mathbb{R}^{n_1\times n_2\times n_3}_{(r_1,r_2)}}\nabla f_\Omega(X)\|^2 = 0.9$, obtained after 120 iterations, and with noise added to the data as in (18) with $\eta = 10$.

We are working on a Riemannian rank-adaptive method using this rank estimation method on the LRTCP and additionally on an extension of this method to higher dimensions.

Acknowledgment

This work was supported by the Fonds de la Recherche Scientifique – FNRS and the Fonds Wetenschappelijk Onderzoek – Vlaanderen under EOS Project no 30468160, and by the Fonds de la Recherche Scientifique – FNRS under Grant no T.0001.23.

References

[1] B. Kutschan, "Tangent cones to tensor train varieties," *Linear Algebra Appl.*, vol. 544, pp. 370–390, 2018.
[2] I. V. Oseledets, "Tensor-train decompositions," *SIAM J. Sci. Comput.*, vol. 33, no. 5, pp. 2295–2317, 2011.
[3] S. Holtz, T. Rohwedder, and R. Schneider, "On manifolds of tensors of fixed TT-rank," *Numer. Math.*, vol. 120, no. 4, pp. 701–731, 2012.
[4] F. Harper and J. Konstan, "The MovieLens datasets: History and context," *ACM Trans. Interact. Intell. Syst.*, vol. 5, dec 2015.
[5] M. Steinlechner, "Riemannian optimization for high-dimensional tensor completion," *SIAM J. Sci. Comput.*, vol. 38, no. 5, pp. S461–S484, 2016.
[6] S. Budzinskiy and N. Zamarashkin, "Tensor train completion: local recovery guarantees via Riemannian optimization," *arXiv preprint arXiv:2110.03975*, 2021.
[7] D. Kressner, M. Steinlechner, and A. Uschmajew, "Low-rank tensor methods with subspace correction for symmetric eigenvalue problems," *SIAM J. Sci. Comput.*, vol. 36, no. 5, pp. A2346–A2368, 2014.
[8] M. Steinlechner, *Riemannian Optimization for Solving High-Dimensional Problems with Low-Rank Tensor Structure*. PhD thesis, EPFL, Lausanne, 2016.
[9] C. J. Hillar and L.-H. Lim, "Most tensor problems are NP-hard," *Journal of the ACM (JACM)*, vol. 60, no. 6, pp. 1–39, 2013.
[10] C. Lubich, I. V. Oseledets, and B. Vandereycken, "Time integration of tensor trains," *SIAM J. Numer. Anal.*, vol. 53, no. 2, pp. 917–941, 2015.
[11] B. Gao and P.-A. Absil, "A Riemannian rank-adaptive method for low-rank matrix completion," *Computational Optimization and Applications*, vol. 81, pp. 67–90, 2022.
[12] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *Journal of Machine Learning Research*, vol. 15, no. 42, pp. 1455–1459, 2014.