RESEARCH ARTICLE

Tensor train completion: Local recovery guarantees via Riemannian optimization

Stanislav Budzinskiy© | Nikolai Zamarashkin

Marchuk Institute of Numerical Mathematics RAS, Moscow, Russia

Correspondence
Stanislav Budzinskiy, Marchuk Institute of Numerical Mathematics RAS, Moscow, Russia. Email: stanislav.budzinskiy@protonmail.ch

Funding information
Russian Science Foundation, Grant/Award Number: 21-71-10072

Abstract
In this work, we estimate the number of randomly selected elements of a tensor that with high probability guarantees local convergence of Riemannian gradient descent for tensor train completion. We derive a new bound for the orthogonal projections onto the tangent spaces based on the harmonic mean of the unfoldings’ singular values and introduce a notion of core coherence for tensor trains. We also extend the results to tensor train completion with auxiliary subspace information and obtain the corresponding local convergence guarantees.

KEYWORDS
incoherence, Riemannian optimization, subspace information, tensor completion, tensor train, tensors

1 | INTRODUCTION

The problem of recovering algebraically structured data from scarce measurements has already become a classic one. The data under consideration are typically sparse vectors or low-rank matrices and tensors, while the measurements are obtained by applying a linear operator that satisfies a variant of the so-called restricted isometry property (RIP).

In this work, we focus on tensor completion, which consists in recovering a tensor in the tensor train (TT) format from a small subset of its entries. Specifically, we consider it as a Riemannian optimization problem on the smooth manifold of tensors with fixed TT ranks and derive sufficient conditions (essentially, the RIP) for local convergence of the Riemannian gradient descent. We further estimate the number of randomly selected entries of a tensor with low TT ranks that is sufficient for the RIP to hold with high probability and, as a consequence, for the Riemannian gradient descent to converge locally. We leave aside the question of producing a starting point that lies close enough to the solution and concentrate instead on reducing the required number of samples.

Before presenting our main contributions to tensor completion, we suggest to step back and take a look at a particular case of two-dimensional tensors with low TT ranks, which is low-rank matrices. The research into matrix completion, as opposed to multi-dimensional tensor completion, is more mature: not only are the properties of the matrix completion problem better understood, but also the key ideas that were produced in the process have been extended to the tensor case and greatly influenced its development. Therefore, starting with matrix completion is both historically motivated and allows one to better grasp the main concepts: the notion of coherence and the RIP. They lie at the heart of low-rank matrix completion, allowing one to prove that there are computationally feasible methods to solve the problem.

Let $A \in \mathbb{R}^{n_1 \times n_2}$ be a rank-$r$ matrix and let $\Omega \subseteq [n_1] \times [n_2]$ with $[k] = \{1, \ldots, k\}$ be a collection of indices. Assuming that $A(i_1, i_2)$ are known for $(i_1, i_2) \in \Omega$, we aim to find a matrix $X \in \mathbb{R}^{n_1 \times n_2}$ that solves the following rank minimization problem:

$$\text{rank}(X) \to \min \quad \text{s.t.} \quad X(i_1, i_2) = A(i_1, i_2), \quad (i_1, i_2) \in \Omega.$$  

(1)
Two important questions arise: what are the requirements for (1) to have a unique solution and whether the problem is computationally tractable.

Rank minimization problems with affine constraints are NP-hard in general,\(^6\) and Fazel\(^7\) developed a heuristic that consists in minimizing the nuclear norm, that is, the sum of the singular values

\[
\|X\|_* = \min_{k=1}^{\min(n_1,n_2)} \sigma_k(X).
\]

The matrix completion problem (1) then turns into a convex optimization problem

\[
\|X\|_* \rightarrow \min \text{ s.t. } X(i_1,i_2) = A(i_1,i_2), \ (i_1,i_2) \in \Omega,
\]

and can be solved as a semidefinite program. A breakthrough in understanding the properties of the nuclear norm minimization for matrix completion was achieved by Candès, Recht, and Tao\(^8\)-\(^10\) who established sufficient conditions under which \(A\) is the unique solution to (2). Their main contribution consists in showing that these sufficient conditions hold with high probability provided that sufficiently many indices \(\Omega\) are chosen uniformly at random. To this end, the authors introduced several key notions and assumptions that limit the class of matrices amenable for completion.

The coherence of a linear subspace is one of them. For an \(r\)-dimensional linear subspace \(T\) of \(\mathbb{R}^n\), its coherence \(\mu(T)\) is defined as

\[
\mu(T) = \frac{n}{r} \max_{i \in [n]} \|P_T e_i\|_2^2, \quad 1 \leq \mu(T) \leq \frac{n}{r},
\]

where \(e_i \in \mathbb{R}^n\) are canonical basis vectors and \(P_T : \mathbb{R}^n \rightarrow T\) is the orthogonal projection operator. With a slight abuse of notation, we will write \(\mu(U) = \mu(T)\) for any matrix \(U\) whose columns span \(T\). If \(U\) happens to have orthonormal columns, the coherence of its column space can be computed as

\[
\mu(U) = \frac{n}{r} \max_{i \in [n]} \|U^T e_i\|_2^2.
\]

The worst case for matrix completion is a rank-1 matrix of the form \(A = e_i e_j^T\): there is no hope for recovery unless we observe all of its entries. Similarly pessimistic are \(A = u e_i^T\) and \(A = e_i e_j^T\). For these examples, their column and/or row spaces have the maximum possible coherences. A reasonable assumption, then, is that both column and row spaces of \(A\) are incoherent, that is, their coherences are bounded by a small constant

\[
\mu(U) \leq \mu_0, \quad \mu(V) \leq \mu_0.
\]

Here, \(U \in \mathbb{R}^{n_1 \times r}\) and \(V \in \mathbb{R}^{n_2 \times r}\) are the left and right singular factors of \(A\).

Another object that plays an important role in References 8-10 is the following linear subspace of \(\mathbb{R}^{n_1 \times n_2}\), associated with \(A\),

\[
T_A = \{UM + NV^T : M \in \mathbb{R}^{r \times n_2}, N \in \mathbb{R}^{n_1 \times r}\} \subset \mathbb{R}^{n_1 \times n_2}
\]

together with the corresponding orthogonal projection operator

\[
P_{T_A} X = UU^T X + XVV^T - UU^T XVV^T \in T_A.
\]

In fact, \(T_A\) is exactly the tangent space at \(A\) to the smooth manifold of rank-\(r\) matrices.\(^{11}\) Let \(R_\Omega : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^{n_1 \times n_2}\) denote the sampling operator that sets to zero all elements of a matrix that do not lie in the index set \(\Omega\):

\[
R_\Omega X = \sum_{(i_1,i_2) \in \Omega} X(i_1,i_2) e_{i_1} e_{i_2}^T.
\]
One of the assumptions made in References 8-10 to prove that $A$ is the unique solution to (2) is that $R_\Omega$ satisfies a variant of the RIP with $\epsilon = 1/2$:

$$\|\rho^{-1}P_{T_\Omega}R_\Omega P_{T_\Omega} - P_{T_\Omega}\| < \epsilon, \quad \rho = \frac{|\Omega|}{n_1n_2}. \tag{8}$$

where $\| \cdot \|$ is the operator norm induced by the Frobenius norm. Calling (8) a RIP is justified, since its direct consequence is

$$(1 - \epsilon)\|X\|_F \leq \|\rho^{-1}P_{T_\Omega}R_\Omega X\|_F \leq (1 + \epsilon)\|X\|_F, \quad X \in T_A.$$ 

Candès and Recht proved estimates on the number of known elements $|\Omega|$ that guarantees the RIP (8).

**Theorem 1** (References 8, Theorem 4.1 and 10, Theorem 6). Let the matrix $A$ have incoherent column and row spaces (4) and assume that the index set $\Omega$ is chosen uniformly at random with

$$|\Omega| \geq \frac{1}{\epsilon^2} \mu_0 r n \log(n), \quad n = \max(n_1, n_2).$$

Then the RIP (8) holds with high probability.

The RIP (8) alone, however, is not sufficient for $A$ to be the unique minimizer of (2). The best (to date) estimate on the number of known elements that guarantees that nuclear norm minimization (2) solves the matrix completion problem was derived in Reference 12 with the help of leave-one-out analysis.

**Theorem 2** (Reference 12, Theorem 2). Let the matrix $A$ have incoherent column and row spaces (4) and assume that the index set $\Omega$ is chosen uniformly at random with

$$|\Omega| \geq \mu_0 r \log(\mu_0 r)n \log(n), \quad n = \max(n_1, n_2).$$

Then $A$ is the unique minimizer of (2).

This bound is almost optimal, considering that

$$|\Omega| \geq \mu_0 r n \log(n)$$

random samples are necessary to rule out the situation when several rank-$r$ matrices agree on the sample $\Omega$ (see Reference 9). It is also interesting to note that both necessary and sufficient conditions amount to only polylogarithmic oversampling as $r(n_1 + n_2 - r)$ parameters describe every rank-$r$ matrix of size $n_1 \times n_2$.

A different approach to matrix completion is to minimize the residual on the sampling set under the rank constraint:

$$\|R_\Omega X - R_\Omega A\|_F^2 \rightarrow \min \quad \text{s.t.} \quad \text{rank}(X) \leq r. \tag{9}$$

Unlike (2), this optimization problem is non-convex and, as a result, can have multiple local minima and saddle points. A closely related perspective builds upon a geometric fact that the set

$$M_r = \{X \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(X) = r\}$$

is a smooth embedded submanifold of $\mathbb{R}^{n_1 \times n_2}$ (see References 11 and 5). This means that the problem

$$\|R_\Omega X - R_\Omega A\|_F^2 \rightarrow \min \quad \text{s.t.} \quad X \in M_r \tag{10}$$

can be solved using Riemannian optimization methods. The Riemannian gradient descent (RGD) reads as

$$X_{t+1} = \text{SVD}_r \left( X_t - \alpha_t P_{T_\Omega, M_r} [R_\Omega X_t - R_\Omega A] \right), \tag{11}$$
where \( a_t > 0 \) is the step size, \( T_X, M_r \) is the tangent space to \( M_r \) at \( X_t \in M_r \) given by (5), and \( P_{T_X, M_r} \) is the corresponding orthogonal projection operator (6). The local linear convergence of the RGD (11) was studied in Reference 14, where it was proved that the RIP (8) is, essentially, the only sufficient condition.

**Theorem 3** (Reference 14, Theorem 2.2). Assume that the sampling operator \( R_{\Omega} \) is bounded \( \| R_{\Omega} \| \leq C \) and satisfies the RIP (8) with \( \varepsilon < 1/22 \). If the initial point \( X_0 \in M_r \) satisfies

\[
\frac{\| X_0 - A \|_F}{\sigma_{\text{min}}(A)} < \frac{\varepsilon \sqrt{\rho}}{2C(1 + \varepsilon)},
\]

where \( \sigma_{\text{min}}(A) \) is the smallest positive singular value of \( A \), then the RGD (11) converges linearly to \( A \) as

\[
\| X_t - A \|_F < \left( \frac{18\varepsilon}{1 - 4\varepsilon} \right)^t \| X_0 - A \|_F.
\]

Our intention with this paper is to look at the notion of coherence (3) and the RIP (8) in the multi-dimensional setting of tensors with low TT ranks, explore how they affect the properties of the RGD for TT completion, and relate them to estimate the required number of known elements. In pursuing our goal, we will follow a sequence of steps:

1. prove a new theorem on local linear convergence of the RGD for TT completion (an analog of Theorem 3);
2. introduce a new notion of core coherence for tensors in the TT format (an extension of the coherence (3));
3. formulate a new incoherence assumption (an analog of (4)) and derive a new estimate on the number of randomly selected elements of a tensor in the TT format that guarantees the RIP with high probability (an analog of Theorem 1).

We set up the notation and list the basic facts about the TT format in Section 2. The next Section 3 collects what we consider to be the main contributions of our paper: their formulations, the motivation behind them, and a detailed comparison with the existing literature. Section 4 is entirely devoted to the proofs of our main results. In Section 5, we attempt to evaluate our findings and outline directions for the future research. The paper also has two Appendices: Appendix A, where we provide a broader context of tensor completion and overview other important developments in the field, and Appendix B, where we adapt our results on TT completion to a modified problem with auxiliary subspace information.

## 2 | NOTATION AND PRELIMINARIES

We denote matrices by capital letters \( X, Y, Z \) and tensors by bold capital letters \( \mathbf{X}, \mathbf{Y}, \mathbf{Z} \). An element of a \( d \)-dimensional tensor \( \mathbf{X} \) at position \((i_1, \ldots, i_d)\) is marked as \( X(i_1, \ldots, i_d) \). The identity matrix of size \( n \) is written as \( I_n \). We denote its columns, the canonical basis vectors of \( \mathbb{R}^n \), by \( e_j \) for all \( j \in [n] = \{1, \ldots, n\} \), and the size of \( e_j \) will be clear from the context. Calligraphic letters such as \( P, R, S \) denote linear operators acting on matrices or tensors, \( Id \) is the identity operator. The Frobenius norm of a matrix or tensor is denoted by \( \| \cdot \|_F \). This is a Euclidean norm with the standard inner product

\[
\| \mathbf{X} \|_F = \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle_F}, \quad \langle \mathbf{X}, \mathbf{Y} \rangle_F = \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} X(i_1, \ldots, i_d) Y(i_1, \ldots, i_d).
\]

The operator norm induced by it is marked as \( \| \cdot \| \). We write \( \| \cdot \|_2 \) for the \( l_2 \) norm of a vector and the spectral norm of a matrix.

The Kronecker product is denoted by \( \otimes \), and \( \circ \) stands for the outer product. For instance, for every multi-index \( \omega = (i_1, \ldots, i_d) \in [n_1] \times \cdots \times [n_d] \), the corresponding canonical basis tensor \( E_\omega \) and its vectorization \( e_\omega \) can be represented as

\[
E_\omega = e_{i_1} \circ \cdots \circ e_{i_d}, \quad e_\omega = e_{i_1} \otimes \cdots \otimes e_{i_d}.
\]

A mode-\( k \) product of a tensor \( \mathbf{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d} \) with a matrix \( U \in \mathbb{R}^{m_k \times n_k} \) is denoted by \( \mathbf{X}_k \) so that

\[
\mathbf{Y} = \mathbf{X} \times_k U \in \mathbb{R}^{n_1 \times \cdots \times n_{k-1} \times m_k \times n_{k+1} \times \cdots \times n_d} \quad \text{and} \quad Y(i_1, \ldots, i_{k-1}, j_k, i_{k+1}, \ldots, i_d) = \sum_{i_k=1}^{n_k} X(i_1, \ldots, i_d) U(j_k, i_k).
\]
For a tensor $X \in \mathbb{R}^{n_1 \times \ldots \times n_d}$, its mode-$k$ flattening is a matrix of size $n_k \times \prod_{j \neq k} n_j$ denoted by $X^{(k)}$, the columns of $X^{(k)}$ are called mode-$k$ fibers. The $k$th unfolding of $X$ is a matrix of size $(n_1 \ldots n_k) \times (n_{k+1} \ldots n_d)$ denoted by $X^{(k)}$. A tensor is said to be in the tensor train (TT) format \footnote{BUDZINSKIY and ZAMARASHKIN} if each of its elements can be evaluated according to

$$X(i_1, \ldots, i_d) = \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_{d-1}=1}^{r_{d-1}} G_1(i_1, \alpha_1)G_2(\alpha_1, i_2, \alpha_2) \cdots G_{d-1}(\alpha_{d-2}, i_{d-1}, \alpha_{d-1})G_d(\alpha_{d-1}, i_d).$$

The matrices $G_1 \in \mathbb{R}^{n_1 \times n_k}$, $G_d \in \mathbb{R}^{n_{d-1} \times n_d}$ and the $3$-dimensional tensors $G_k \in \mathbb{R}^{n_{k-1} \times n_k \times n_k}$ are called TT cores. The upper limits of the summations, $r_k \in \mathbb{N}$, are conventionally combined into a tuple $r = (r_1, \ldots, r_{d-1})$ that is called the TT rank of the decomposition. To make the notation more consistent, we will write $G_1 \in \mathbb{R}^{n_1 \times n_k}$ and $G_d \in \mathbb{R}^{n_{d-1} \times n_d \times n_d}$ with $r_0 = r_d = 1$ for the first and last TT cores. We will also denote by $X = [G_1, G_2, \ldots, G_d]$ the TT representation itself.

Every tensor $X$ can be represented in the TT format. This can be achieved with the TT-SVD algorithm, \footnote{BUDZINSKIY and ZAMARASHKIN} and the TT ranks of the resulting representation are equal to the ranks of the unfolding matrices $X^{(k)}$. The unfolding matrices can be factorized as products of interface matrices $X^{(k)} = X_{2\leq k}X_{2\geq k+1}^T$, which can be defined recursively as

$$X_{2\leq 1} = G_1, \quad X_{2\geq k} = (U_{n_k} \otimes X_{2\leq k-1})G_k^L \in \mathbb{R}^{(n_1 \ldots n_{k-1}) \times n_k},$$

$$X_{2\leq d} = G_d^R, \quad X_{2\geq k+1} = (X_{2\leq k+2} \otimes I_{n_{k+1}})G_{k+1}^R \in \mathbb{R}^{(n_k \ldots n_d) \times n_{k+1}}.$$

The matrices $G_k^L \in \mathbb{R}^{n_{k-1} \times n_k \times n_k}$ and $G_k^R \in \mathbb{R}^{n_k \times n_k \times n_d}$ are the left and right unfoldings of the $k$th TT core $G_k$, respectively.

While a tensor can admit various TT representations with different TT ranks, under certain minimality conditions of the representation (satisfied by what TT-SVD outputs) the TT ranks are unique.\footnote{BUDZINSKIY and ZAMARASHKIN} Namely, for every TT core its left and right unfoldings must be full-rank. This justifies the notion of the TT rank of a tensor

$$\text{rank}_{TT}(X) = (\text{rank}(X^{(1)}), \ldots, \text{rank}(X^{(d-1)})).$$

The set of tensors of tensors with fixed TT rank will be denoted by

$$M_r = \{X \in \mathbb{R}^{n_1 \times \ldots \times n_d} : \text{rank}_{TT}(X) = r\},$$

and it is a smooth embedded submanifold \footnote{BUDZINSKIY and ZAMARASHKIN} of $\mathbb{R}^{n_1 \times \ldots \times n_d}$ of dimension

$$\dim M_r = \sum_{k=1}^d r_{k-1}n_kr_k - \sum_{k=1}^{d-1} r_k^2.$$  

Among all minimal representations specifically useful are $k$-orthogonal representations

$$X = [U_1, \ldots, U_{k-1}, G_k, V_{k+1}, \ldots, V_d]$$

such that every $U_i$ is left-orthogonal and every $V_j$ is right-orthogonal

$$(U_i^L)^T U_i^L = I_{r_i}, \quad i = 1, \ldots, k - 1, \quad V_j^R V_j^R = I_{r_{j+1}}, \quad j = k + 1, \ldots, d.$$

We call $1$-orthogonal and $d$-orthogonal representations right- and left-orthogonal, respectively. A minimal $k$-orthogonal representation of a tensor can be computed with TT-SVD followed by a partial sweep of QR (or RQ) orthogonalizations.

The truncated TT-SVD algorithm can be used to approximate $X$ with a tensor of given TT rank $r \in \mathbb{N}^{d-1}$. Unlike the truncated SVD for matrices, the resulting approximation is not optimal but is quasi-optimal nonetheless

$$\|\text{TT-SVD}_r(X) - X\|_F \leq \sqrt{d - 1}\|\text{opt}_r(X) - X\|_F,$$

where $\text{opt}_r(X)$ is the best rank-$r$ approximation of $X$ in the Frobenius norm.
3 | OUR CONTRIBUTIONS

3.1 Curvature bound

For the needs of the convergence analysis, we are interested in estimating how quickly the projection operator onto the tangent space $P_{T_XM_r}$ changes as we move around on the manifold $M_r$. Another concern is the following. Every $X \in M_r$ belongs to its own tangent space $X \in T_X M_r$ but it is also important to know how well $X$ can be approximated by other tangent spaces in its neighborhood, which essentially gives a bound on the curvature of the manifold.

Our first result (Lemma 1) is a new curvature bound for $M_r$. Denote by $\sigma_{\min}(\cdot)$ the smallest positive singular value of a matrix and, with some abuse of notation, the harmonic mean of the smallest positive singular values of the unfoldings of a tensor

$$\sigma_{\min}(X) = (d - 1)\left(\frac{1}{\sum_{k=1}^{d-1} \frac{1}{\sigma_{\min}(X^{(k)})}}\right)^{-1}.$$  

Lemma 1. For every pair of tensors $X, \tilde{X} \in M_r$ with the same TT ranks it holds that

$$\|(\text{Id} - P_{T_XM_r})X\|_F \leq (d - 1)\frac{\|X - \tilde{X}\|^2_F}{\sigma_{\min}(X)} \quad \text{and} \quad \|P_{T_XM_r} - P_{T_XM_r}\| \leq 2(d - 1)\frac{\|X - \tilde{X}\|_F}{\sigma_{\min}(X)}.$$  

Similar bounds were obtained in Reference 16 (Lemma 4.5) for tensors in the hierarchical Tucker format, of which TT is a particular case. Most importantly, the bounds in Lemma 1 remain valid for any tensor $\tilde{X} \in M_r$, while those in Reference 16 hold only in a neighborhood of $X$. Analogous global upper bounds were also derived in Reference 17 for $\|(\text{Id} - P_{T_XM_r})X\|_F$ (Lemma 27) and for $\|P_{T_XM_r} - P_{T_XM_r}\|$ (Equation 38). However, our bounds are tighter: (1) the constants are smaller; (2) we use the harmonic mean of the singular values, while References 16 and 17 work with the minimum of the singular values, and $\sigma_{\min}(X) \geq \min_{k \in [d-1]} \sigma_{\min}(X^{(k)})$.

The two types of averaging coincide when all $\sigma_{\min}(X^{(k)})$ are the same. For instance, if we take Lemma 1 with $d = 2$, we recover the curvature bounds for the matrix manifold as in Reference 14 (Lemma 4.1). The situation is different when $d > 2$ and some of the unfoldings are ill-conditioned. This can occur when a full-rank tensor is approximated in the TT format with overestimated TT ranks. Assume that $\sigma_{\min}(X^{(k)})$ are equal to 1 for $d - 1$ and $s$ unfoldings and to $0 < \varepsilon < 1$ for the remaining $s$ unfoldings. We have

$$\min_{k \in [d-1]} \sigma_{\min}(X^{(k)}) = \varepsilon, \quad \sigma_{\min}(X) = 1 - \frac{s(1 - \varepsilon)}{s + (d - 1 - s)\varepsilon},$$

and $\sigma_{\min}(X)$ can be seen as a convex combination of 1 and $\varepsilon$ with a dimension-dependent coefficient $0 < \alpha_{s,d} < 1$:

$$\sigma_{\min}(X) = (1 - \alpha_{s,d}) + \alpha_{s,d}\varepsilon, \quad \alpha_{s,d} = \frac{s}{s + (d - 1 - s)\varepsilon}.$$  

Lemma 1 shows that the curvature is tolerant to ill-conditioned unfoldings for high-dimensional tensors, while the previous results overestimate it. For example, take $\varepsilon = 10^{-2}$ and $d = 100$. Such high-dimensional tensors appear when the quantized TT format is used to approximate differential operators and solve differential equations. We get $\sigma_{\min}(X) \approx 0.17$ for $s = 5$ and $\sigma_{\min}(X) \approx 0.09$ for $s = 10$, which are about $d/s$ times larger than $\varepsilon$. As a result, the previous curvature bounds $(d - 1)/\varepsilon$ go down to about $s/\varepsilon$. If we let $d$ grow with $s$ and $\varepsilon$ fixed, the asymptotics are

$$\sigma_{\min}(X) = 1 - \frac{s}{d}(\varepsilon^{-1} - 1) + O\left(\frac{1}{d^2}\right).$$

3.2 Local convergence of Riemannian gradient descent

3.2.1 Tensor recovery

The TT completion problem is a particular instance of a more general TT recovery problem with a linear measurement operator $R : \mathbb{R}^{n_1 \times \cdots \times n_d} \to \mathbb{R}^s$, where one needs to recover a tensor $A$ in the TT format given the measurements $RA$:

$$\|RX - RA\|^2_F \to \min \quad \text{s.t.} \quad X \in M_r.$$
This Riemannian optimization problem can be solved with the RGD, and to explicitly formulate the method, we need to choose the step size and the retraction mapping. The truncated TT-SVD is a valid retraction on $\mathcal{M}_r$ (see Reference 20), hence our RGD step is

$$X_{t+1} = \text{TT-SVD}_r(X_t - \alpha_t Y_t) \in \mathcal{M}_r, \quad Y_t = P_X R^\ast (RX_t - RA) \in T_X \mathcal{M}_r, \quad (13)$$

where we use $P_X$ as an alias for $P_{T_X \mathcal{M}_r}$ and the step size is chosen via exact line search in the tangent space $T_X \mathcal{M}_r$:

$$\alpha_t = \|Y_t\|^2_F / \|R Y_t\|^2_F.$$

The appeal of this step size is in its closed-form formula, which greatly simplifies the analysis of the RGD (13). From the numerical perspective, such $\alpha_t$ can be used as a good starting point for the backtracking scheme applied along the geodesic; typically, though, $\alpha_t$ itself is sufficient. In general tensor recovery problems, the measurement operator $R$ is assumed to exhibit a more standard (than (8)) variant of the RIP. We will say that $R$ satisfies the standard RIP of order $r$ if the following two-sided bound holds for all tensors $X$ of TT rank at most $r$ with a RIP constant $0 < \delta_r < 1$. An example of a measurement operator for which the standard RIP (14) holds with high probability are i.i.d. random Gaussian measurements. The sampling operator $R_\Omega$ of tensor completion, however, cannot fulfill the standard RIP (14) for all tensors with low TT ranks (consider a sparse tensor). Nonetheless, the RGD convergence analyses for TT recovery and TT completion are very similar, and the proof of the former (which is slightly easier) can be easily adapted to fit the latter. This brings us to the new Theorem 4, which establishes local linear convergence of the RGD (13) for the TT recovery problem.

**Theorem 4.** Let $A \in \mathcal{M}_r$ be a tensor of TT rank $r$. Suppose that the measurement operator $R$ satisfies the standard RIP (14) of order $2r$ with a RIP constant $0 < \delta_{2r} < 1$ and is bounded $\|R^\ast R\| \leq C$. Then the error on the current step of the RGD (13) is estimated via the previous error

$$\|X_{t+1} - A\|_F \leq \beta_t \|X_t - A\|_F$$

with a constant

$$\beta_t = (1 + \sqrt{d - 1}) \left[ \frac{2 \delta_{2r}}{1 - \delta_{2r}} + \left(1 + \frac{C}{1 - \delta_{2r}}\right)(d - 1) \frac{\|X_t - A\|_F}{\sigma_{\min}(A)} \right].$$

If $\delta_{2r} < (3 + 2\sqrt{d - 1})^{-1}$ and the initial condition $X_0 \in \mathcal{M}_r$ satisfies

$$(d - 1) \frac{\|X_0 - A\|_F}{\sigma_{\min}(A)} < \frac{1}{1 + C - \delta_{2r}} \left(\frac{1 - \delta_{2r}}{1 + \sqrt{d - 1} - 2 \delta_{2r}}\right),$$

the iterations of RGD converge linearly to $A$ at a rate

$$\|X_{t+1} - A\|_F < \beta_0^{t+1} \|X_0 - A\|_F, \quad \beta_0 < 1.$$

If $R$ satisfies the standard RIP (14) of order $3r$, the above results remain valid when $C$ is replaced with $1 + \delta_{3r}$.

The novelty of our Theorem 4 is that we require the standard RIP (14) of order $2r$, while an analogous result from Reference 22(Theorem 3) uses order $3r$. We achieve this by leveraging the upper bound $\|R^\ast R\| \leq C$. In addition, we consider a varying step size (Reference 22 sets $\alpha_t = 1$ for the proof) and provide explicit expressions for the radius of convergence and the convergence rate. We have not seen results similar to Theorem 4 with the standard RIP (14) of order $2r$ in the matrix case, either.
3.2.2 | Tensor completion

Turning to the tensor completion problem, we introduce a collection of multi-indices \( \Omega \subset [n_1] \times \cdots \times [n_d] \) and denote by \( \rho = |\Omega|/(n_1 \cdots n_d) \) the density of known elements. We define the sampling operator \( R_\Omega : \mathbb{R}^{n_1 \times \cdots \times n_d} \rightarrow \mathbb{R}^{n_1 \times \cdots \times n_d} \) as

\[
R_\Omega X = \sum_{\omega \in \Omega} X(\omega) E_\omega, \quad \omega = (i_1, \ldots, i_d),
\]

where \( E_\omega = e_{i_1} \otimes \cdots \otimes e_{i_d} \) are canonical basis tensors. This definition allows \( \Omega \) to contain repeated elements so in general \( R_\Omega \) is not a projection operator. It is, however, self-adjoint and positive semi-definite. For the ease of presentation, we will use \( R = \sqrt{R_\Omega} \) as the measurement operator to reformulate the RGD (13) for the specific case of tensor completion:

\[
X_{t+1} = \text{TT-SVD}_r (X_t - \varrho Y_t) \in \mathcal{M}_r, \quad Y_t = P_X [R_\Omega X_t - R_\Omega A] \in T_X \mathcal{M}_r,
\]

with the step size

\[
\alpha_t = \frac{||Y_t||_F^2}{\langle R_\Omega Y_t, Y_t \rangle_F}.
\]

As we discussed previously, the sampling operator cannot satisfy the standard RIP (14), so we resort to a weaker (see Lemma 5) assumption, which is just a verbatim translation of the RIP (8) from matrix completion to the multi-dimensional setting:

\[
||P_{T_X \mathcal{M}_r} - \rho^{-1}P_{T_X \mathcal{M}_r} R_\Omega P_{T_X \mathcal{M}_r}|| < \varepsilon. \tag{16}
\]

Armed with this assumption, we can prove a new Theorem 5 on the convergence of the RGD (15) for TT completion.

**Theorem 5.** Let \( A \in \mathcal{M}_r \) be a tensor of TT rank \( r \). Suppose that the sampling operator \( R_\Omega \) satisfies the RIP (16) and is bounded \( ||R_\Omega|| \leq C \). Then the error on the current step of the RGD (15) is estimated via the previous error

\[
||X_{t+1} - A||_F \leq \beta_t ||X_t - A||_F
\]

with a constant

\[
\beta_t = (1 + \sqrt{d - 1}) \left[ \frac{2 \varepsilon_t}{1 - \varepsilon_t} + \left( 1 + \frac{C}{1 - \varepsilon_t} \right) (d - 1) \frac{||X_t - A||_F}{\sigma_{\min}(A)} \right] \cdot \varepsilon_t = \varepsilon + (2 + 4C \rho^{-1}) (d - 1) \frac{||X_t - A||_F}{\sigma_{\min}(A)}.
\]

If \( \varepsilon < (3 + 2 \sqrt{d - 1})^{-1} \) and the initial condition \( X_0 \in \mathcal{M}_r \) satisfies

\[
(d - 1) \frac{||X_0 - A||_F}{\sigma_{\min}(A)} < \min \left( \frac{1 - \varepsilon}{2 + 4C \rho^{-1}}, \frac{5 + 8C \rho^{-1} + 2 + 4C \rho^{-1}}{1 + \sqrt{d - 1} - \varepsilon} \right),
\]

the iterations of RGD converge linearly to \( A \) at a rate

\[
||X_{t+1} - A||_F \leq \beta_0^{t+1} ||X_0 - A||_F. \quad \beta_0 < 1.
\]

To the best of our knowledge, local linear convergence of the RGD (15) has not been established before. In Reference 17, Riemannian TT completion is addressed from the theoretical point of view as well, but the algorithm is different there: an additional trimming procedure is applied on every iteration before TT-SVD to ensure that all the elements of the tensor remain below a certain threshold. This algorithm is proved to locally linearly converge in Reference 17 (Lemmas 5 and 9), but the assumptions are stricter than in our Theorem 5: in addition to the RIP (16) and the bound \( ||R_\Omega|| \leq C \) (which are not present in the formulations of the Lemmas, but can be found in the proof of Lemma 9 under the names \( E_1 \) and \( E_2 \)), the initial condition \( X_0 \) is required to have low interface coherence (we will talk about this notion later on).
Meanwhile, our Theorem 5 guarantees local linear convergence for any initial condition as long as it is sufficiently close to $A$.

We can also compare the implications of Theorem 5 for matrix completion with Theorem 3. Our result guarantees convergence when the RIP ($\varepsilon$) holds with a larger $\varepsilon$ (1/5 against 1/22) and, as a consequence, when fewer elements of the matrix are known, owing to Theorem 1.

From the numerical perspective, the trimming step in Reference 17 renders the whole algorithm expensive both in terms of memory requirements and computational complexity, since a full tensor needs to be assembled from its TT representation and then TT-SVD is applied to a full tensor too. It is noted, however, that in numerical experiments the iterations with and without trimming behave in a nearly identical manner. While the algorithm without trimming (which is exactly our RGD (15)) is much more efficient, there is still a question of how to choose the initial point $X_0$: the sequential spectral initialization of Reference 17 (Alg. 3) and the possible multi-dimensional extensions of the initialization strategies in Reference 14 may be provable, but their computational complexities are likely to dwarf the resources needed to carry out the RGD iterations for large tensors. We leave aside the problem of choosing the initial point in this paper, but we believe that a more promising direction that can lead to a provably convergent computationally efficient method is random initialization.23,24

It should be noted that in Theorems 4 and 5, we implicitly assume that the sequences generated by the RGD always remain on the manifold $\mathcal{M}_r$; however, though virtually unseen in practice, the TT rank can become smaller. This phenomenon was studied in the matrix case for a projected linesearch method on the algebraic variety of matrices with rank not bigger (as opposed to equal) than a certain fixed value.25

### 3.3 Recovery guarantees

The main assumption in Theorem 5 that guarantees local linear convergence of the RGD (15) is that the sampling operator $R_{\Omega}$ satisfies the RIP (16). Assuming that the indices $\Omega$ are chosen uniformly at random with replacement, we want to derive probabilistic sufficient conditions which ensure that the RIP holds with high probability. In this setting, we can also obtain a new bound on the sampling operator.

**Lemma 2.** Let $\Omega \subset [n_1] \times \ldots \times [n_d]$ be a collection of indices sampled uniformly at random with replacement. Then the norm of the sampling operator is bounded by

$$\|R_{\Omega}\| \leq \frac{d\beta}{w(d)} \log(n), \quad n = \max(n_1, \ldots, n_d),$$

with probability at least $1 - n^{d(1 - \beta)}$ for $n \geq 16$ and $\beta > 1$. Here, $w(d)$ is the principal branch of the Lambert $W$ function, also known as product logarithm.

**Proof.** The norm $\|R_{\Omega}\|$ is nothing but the maximum number of repetitions in the sample. Consider $|\Omega|$ i.i.d. Bernoulli random variables $\xi_i$ with probability of success $1/(n_1 \ldots n_d)$ and let $\xi = \sum_j \xi_j$. Since all the indices in $\Omega$ are drawn with equal probability with replacement, $\xi$ describes how many times a single fixed entry is sampled. Then the probability of it being sampled more than $k$ times can be upper bounded with the help of the Chernoff bound

$$P\{\xi > k\} \leq \left(\frac{\rho}{\lambda}\right)^{x} \exp(x - \rho), \quad \rho = \frac{|\Omega|}{n_1 \ldots n_d}.$$  

The union bound over all the entries leads to

$$P\{\|R_{\Omega}\| > k\} \leq (n_1 \ldots n_d) \cdot P\{\xi > k\} \leq n^d \left(\frac{\rho}{\lambda}\right)^{x} \exp(x - \rho) < n^d \left(\frac{1}{\lambda}\right)^{x} \exp(x).$$

It remains to substitute $x = d\beta \log(n)/w(d)$ and note that for $n \geq 16 > \exp(e)$,

$$w(d) \exp(w(d)) = d \leq \frac{\log(n)}{e} d < \frac{\log(n)}{e} d\beta.$$
Lemma 2 is a direct multi-dimensional extension of Reference 10, Proposition 5. An analogous result appears in Reference 17(Lemma 33) with a $d \beta \log(n)$ bound. The bound we prove is tighter, especially for large $d$. Indeed, the Lambert W function behaves as $w(d) = \log(d) - \log(\log(d)) + o(1)$, so the bound grows as $d \log(n)/\log(d)$. A similar asymptotic was mentioned in Reference 10.

Going back to the RIP (16), we want to extend Theorem 1 from matrices to tensors in the TT format. To this end, we need to generalize the assumption of bounded coherence (4) and/or the notion of coherence (3) itself. Every matrix, seen as a tensor, coincides with its unfolding $A = A^{(1)}$ and has its column and row spaces spanned by the columns of the interface matrices $A_{\leq 1}$ and $A_{\geq 2}$, respectively. The incoherence assumption (4) can then be written in a way that is easily extended to the multi-dimensional case:

$$\mu(A_{\leq 1}) \leq \mu_0, \quad \mu(A_{\geq 2}) \leq \mu_0.$$

We define the interface coherence of a tensor $A$ as the maximum coherence of its left and right interface matrices:

$$\mu_I(A) = \max(\mu(A_{\leq 1}), \mu(A_{\geq 2}), \ldots, \mu(A_{\leq d-1}), \mu(A_{\geq d})).$$

Recalling the definition of the coherence (3), we get

$$\mu(A_{\leq k}) = \frac{n_1 \cdots n_k}{r_k} \max_{i_k \in [n_k], \ldots, i_1 \in [n_1]} \|P_{\leq k}(e_{i_k} \otimes \cdots \otimes e_{i_1})\|^2_2,$$

$$\mu(A_{\geq k+1}) = \frac{n_{k+1} \cdots n_d}{r_k} \max_{i_{k+1} \in [n_{k+1}], \ldots, i_d \in [n_d]} \|P_{\geq k+1}(e_{i_{k+1}} \otimes \cdots \otimes e_{i_d})\|^2_2.$$

As we replace the incoherence assumption (4) with the interface incoherence, we can prove an analog of Theorem 1 for tensors with low TT ranks.

**Theorem 6.** Let $A \in \mathcal{M}_r$ be a tensor of TT rank $r$ with bounded interface coherence $\mu_I(A) \leq \mu_0$ and let $\Omega \subset [n_1] \times \cdots \times [n_d]$ be a collection of indices sampled uniformly at random with replacement. Then the RIP (16)

$$\|P_{T_k\mathcal{M}_r} - \rho^{-1} P_{T_k\mathcal{M}_r} R_\Omega P_{T_k\mathcal{M}_r}\| < \epsilon, \quad \rho = \frac{|\Omega|}{n_1 \cdots n_d},$$

holds with probability at least $1 - 2n^{d(1-\beta)}$, $n = \max(n_1, \ldots, n_d)$, for all $\beta > 1$ provided that

$$|\Omega| \geq \frac{8}{3} \frac{\beta}{\epsilon^2} \mu_0 \left(n_1 r_1 + \mu_0 \sum_{k=2}^{d-1} r_{k-1} n_k r_k + r_{d-1} n_d\right) d \log(n).$$

The interface coherence (17) can also be found in Reference 17 under the name Incoh($A$) (to be precise, Incoh($A$) = $\sqrt{\mu_I(A)}$). A very similar result was proved independently there (Lemma 31). The differences are minor: we treat $\epsilon$ as a parameter and our estimate is more detailed. In the two-dimensional case, Theorem 6 becomes exactly Reference 10, Theorem 6.

Theorem 6, coupled with Theorem 5 and Lemma 2, shows that, with high probability, the RGD (15) converges locally to $A$ when the number of elements in the sample is of order

$$|\Omega| \gtrsim \mu_0^2 d^2 r^2 n \log(n),$$

where $n = \max(n_1, \ldots, n_d)$ and $r = \max(r_1, \ldots, r_{d-1})$. Every tensor of TT rank $r$ is described with $O(dnr^2)$ parameters, so the local recovery is highly probable with $d \log(n)$ oversampling, just as in the matrix case. To compare, the algorithm in Reference 17(Lemma 5) is proved to converge locally when

$$|\Omega| \gtrsim C_d \mu_0^d r^d n^d \log^d(n), \quad C_d = C_d(d).$$

The problem, however, is that for a tensor $A$ with minimal TT representation $A = [G_1, \ldots, G_d]$, the interface matrices are intimately interconnected (see Equation (12)).
\[ A_{\leq k} = (I_n \otimes A_{\leq k-1}) G_k^L, \]

and so their coherences are also far from being independent. Moreover, \( \mu(A_{\leq d-1}) \) and \( \mu(A_{\geq 2}) \) can become as high as \( nd^{-1}/r \) and hence the value of the interface coherence \( \mu_I(A) \) is a source of potential problems for the sample complexity.

In defining interface coherence, we were inspired by a particular way to express incoherence for matrices, via interface matrices. Here we draw a different analogy. Let \( A = [G_1, G_2] \) be a minimal representation of a matrix. Since their left and right unfoldings satisfy (see Equation (12))

\[ A = G_1^L(G_2^R)^T, \]

we can rewrite the incoherence assumption (4) as

\[ \mu(G_1^L) \leq \mu_0, \quad \mu((G_2^R)^T) \leq \mu_0. \]

We will try to extend the notion of coherence to tensors through the TT cores.

Let \( U \in \mathbb{R}^{r \times n \times s} \) be a three-dimensional left-orthogonal tensor. Denote by \( U^{(i)} \in \mathbb{R}^{r \times s} \) the \( i \)th subblock of its left unfolding:

\[ U^L = \begin{bmatrix} U^{(1)} \\ \vdots \\ U^{(n)} \end{bmatrix} \in \mathbb{R}^{rn \times s}. \]

We define the **left coherence of a three-dimensional left-orthogonal tensor** as

\[ \mu_L(U) = \frac{rn}{s} \max_{i \in [n]} \| U^{(i)} \|_2^2. \quad (18) \]

When \( r = 1 \), the tensor \( U \) becomes a matrix, the subblocks \( U^{(i)} \) become rows, their spectral norm equals to the Euclidean norm, and we recognize that the left coherence is just the coherence of a matrix with orthonormal columns.

Likewise, let \( V \in \mathbb{R}^{r \times n \times s} \) be a right-orthogonal tensor and let \( (V^{(i)})^T \in \mathbb{R}^{r \times s} \) be the \( i \)th subblock of the right unfolding:

\[ V^R = \begin{bmatrix} (V^{(1)})^T & \ldots & (V^{(n)})^T \end{bmatrix} \in \mathbb{R}^{r \times ns}. \]

We define the **right coherence of a three-dimensional right-orthogonal tensor** \( V \) as

\[ \mu_R(V) = \frac{sn}{r} \max_{i \in [n]} \| V^{(i)} \|_2^2. \quad (19) \]

In complete analogy, the right coherence of a three-dimensional right-orthogonal tensor becomes the coherence of a (transposed) matrix with orthonormal rows when \( s = 1 \). Note that while we defined the coherence (3) for arbitrary matrices, the notions of left and right coherences require orthogonality.

Now, consider a \( d \)-dimensional tensor \( X \) in a minimal left-orthogonal TT representation \( X = [U_1, \ldots, U_{d-1}, G_d] \). Since the first \( d-1 \) TT cores are left-orthogonal, we can compute their left coherences \( \{ \mu_L(U_k) \}_{k \in [d-1]} \). But do these values characterize the specific TT representation of \( X \) or the tensor itself? What happens to them when we choose a different minimal left-orthogonal TT representation? The following Lemma 3 shows that \( \{ \mu_L(U_k) \}_{k \in [d-1]} \) do not change.

**Lemma 3.** Let \( X = [U_1, \ldots, U_{d-1}, G_d] = [\tilde{U}_1, \ldots, \tilde{U}_{d-1}, \tilde{G}_d] \) be two minimal left-orthogonal TT representations. Then the left coherences of their TT cores coincide:

\[ \mu_L(U_k) = \mu_L(\tilde{U}_k), \quad k \in [d-1]. \]

The same is true for any two right-orthogonal TT representations and the right coherences of their TT cores.
Proof. We carry out the proof for the left coherences. Consider the column span of the first interface matrix $X_{\leq 1}$. It is spanned by two orthonormal bases $U^L_1$ and $U^L_1$, so there exists an orthogonal matrix $Q_1 \in \mathbb{R}^{n_1 \times r_1}$ such that $U^L_1 = U^L_1 Q_1$ and

$$
\mu_L(\tilde{U}_1) = \frac{r_0 r_1}{r_1} \max_{i \in [n_1]} \| \tilde{U}^{(i)} \|_2^2 = \frac{r_0 r_1}{r_1} \max_{i \in [n_1]} \| U^{(i)} Q_1 \|_2^2 = \frac{r_0 r_1}{r_1} \max_{i \in [n_1]} \| U^{(i)} \|_2^2 = \mu_L(U_1).
$$

By factoring $Q_1$ out of the first TT core and attaching it to the second TT core as

$$
\tilde{U}^L_1 = U^L_1, \quad \tilde{U}^L_2 = (I_{n_2} \otimes Q_1) \tilde{U}^L_2 = \begin{bmatrix} Q_1 \tilde{U}^{(1)}_2 \\ \vdots \\ Q_1 \tilde{U}^{(n_2)}_2 \end{bmatrix},
$$

we get a new minimal left-orthogonal TT representation with $\mu_L(\tilde{U}_1) = \mu_L(\tilde{U}_2)$:

$$
X = [U_1, \tilde{U}_2, \tilde{U}_3, \ldots, \tilde{U}_{d-1}, \tilde{G}_d].
$$

Now suppose we have a minimal left-orthogonal TT representation with $\mu_L(\tilde{U}_k) = \mu_L(\tilde{U}_k)$:

$$
X = [U_1, \ldots, U_{k-1}, \tilde{U}_k, \tilde{U}_{k+1}, \ldots, \tilde{U}_{d-1}, \tilde{G}_d].
$$

As the recursive formulas for the interface matrices (12) show, the column space of $X_{\leq k}$ is spanned by two orthonormal bases that are related via an orthogonal matrix $Q_k \in \mathbb{R}^{n_k \times r_k}$ so that

$$(I_{n_k} \otimes U_{\leq k-1}) U^L_k = (I_{n_k} \otimes U_{\leq k-1}) \tilde{U}^L_k Q_k.$$

Since $U^L_{\leq k-1} U_{\leq k-1} = I_{n_k}$, we get $U^L_k = \tilde{U}^L_k Q_k$ and $\mu_L(U_k) = \mu_L(\tilde{U}_k) = \mu_L(\tilde{U}_k)$. Attaching $Q_k$ to the next TT core gives a new minimal left-orthogonal TT representation

$$
X = [U_1, \ldots, U_k, \tilde{U}_{k+1}, \tilde{U}_{k+2}, \ldots, \tilde{U}_{d-1}, \tilde{G}_d]
$$

with $\mu_L(\tilde{U}_{k+1}) = \mu_L(\tilde{U}_{k+1})$ if $k \leq d - 2$ and $\tilde{G}_d = G_d$ if $k = d - 1$.

Since $\{\mu_L(U_k)\}_{k \in [d-1]}$ are a property of the tensor, rather than its TT representation, this motivates us to define a new notion of core coherence of a tensor.

**Definition 1.** Let $X \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ be a tensor with minimal left- and right-orthogonal TT representations

$$
X = [U_1, \ldots, U_{d-1}, G_d] = [G_1, V_2, \ldots, V_d].
$$

The $k$th left core coherence $\mu_L^{(k)}(X)$ of $X$ is defined as the left coherence (18) of the $k$th TT core of its minimal left-orthogonal TT representation:

$$
\mu_L^{(k)}(X) = \mu_L(U_k), \quad k \in [d - 1].
$$

The $(k + 1)$th right core coherence $\mu_R^{(k+1)}(X)$ of $X$ is defined as the right coherence (19) of the $(k + 1)$th TT core of its minimal right-orthogonal TT representation:

$$
\mu_R^{(k+1)}(X) = \mu_R(V_{k+1}), \quad k \in [d - 1].
$$

The core coherence $\mu_C(X)$ of $X$ is defined as the maximum of its left and right core coherences:

$$
\mu_C(X) = \max \left( \mu_L^{(1)}(X), \ldots, \mu_L^{(d-1)}(X), \mu_R^{(2)}(X), \ldots, \mu_R^{(d)}(X) \right).
$$

(20)
When $X$ is a matrix, $\mu^{(1)}(X)$ is the coherence (3) of its column space and $\mu^{(2)}(X)$ is the coherence of its row space. The core coherence $\mu_c(X)$ and the interface coherence $\mu_I(X)$ coincide for $d = 2$ as well, but are very different for $d > 2$. The following Lemma 4 shows the relationship between the two in the multi-dimensional case.

**Lemma 4.** Let $A \in \mathcal{M}_r$ be a tensor of TT rank $r$ with bounded core coherence $\mu_c(A) \leq \mu_1$ Then the coherences of its left and right interface matrices are estimated as

$$\mu(A_{\leq k}) \leq \mu_1^k, \quad \mu(A_{\geq k+1}) \leq \mu_1^{d-k}, \quad k \in [d-1].$$

**Proof.** Consider a minimal left-orthogonal TT representation $A = [U_1, \ldots, U_{d-1}, G_d]$. The projection onto the column space of an interface matrix $P_{\leq k} = U_{\leq k}U_{\leq k}^T$ can be written, with the help of the recursive formulas (12), as

$$U_{\leq 1} = U_1^L, \quad U_{\leq k} = (I_{n_k} \otimes U_{\leq k-1})U_k^L.$$

It follows that

$$U_{\leq k}^T(e_{i_k} \otimes \cdots \otimes e_{i_1}) = \left(U_{l_1}^{(i_1)} U_{l_2}^{(i_2)} \cdots U_k^{(i_k)}\right)^T \in \mathbb{R}^{n_k}$$

and

$$\|P_{\leq k}(e_{i_k} \otimes \cdots \otimes e_{i_1})\|_2^2 = \left\|\left(U_{l_1}^{(i_1)} U_{l_2}^{(i_2)} \cdots U_k^{(i_k)}\right)^T\right\|_2^2 \leq \|U_{l_1}^{(i_1)}\|_2^2 \cdots \|U_k^{(i_k)}\|_2^2 \leq \frac{r_1}{n_1} \frac{r_2}{n_2} \cdots \frac{r_k}{n_{k-1} n_k} \mu_1^k = \frac{r_k}{n_1 \cdots n_k} \mu_1^k.$$

The proof is the same for the right interface matrices. $\blacksquare$

We propose to replace the interface incoherence assumption in Theorem 6 with a new core incoherence one, which leads to Theorem 7.

**Theorem 7.** Let $A \in \mathcal{M}_r$ be a tensor of TT rank $r$ with bounded core coherence $\mu_c(A) \leq \mu_1$ and let $\Omega \subset [n_1] \times \cdots \times [n_d]$ be a collection of indices sampled uniformly at random with replacement. Then the RIP (16)

$$\|P_{\Omega A} - \rho^{-1}P_{\Omega A}R_{\Omega}P_{\Omega A}\| < \varepsilon, \quad \rho = \frac{|\Omega|}{n_1 \cdots n_d}$$

holds with probability at least $1 - 2n^{d(1-\beta)}$, $n = \max(n_1, \ldots, n_d)$, for all $\beta > 1$ provided that

$$|\Omega| \geq \frac{8}{3} \frac{\beta}{\varepsilon^2} \frac{\mu_1^{d-1}}{\mu_1} \left(\sum_{k=1}^{d} r_{k-1} n_k r_k\right) d \log(n).$$

Lemma 4 shows that, in the worst case, the interface coherence (17) can be bounded by $\mu_1^{d-1}$ and, consequently, Theorem 6 (and Reference 17, Lemma 31) gives the sample complexity of

$$|\Omega| \geq \mu_1^{2d-2} d^2 r^2 n \log(n),$$

where $n = \max(n_1, \ldots, n_d)$ and $r = \max(r_1, \ldots, r_{d-1})$. Once we use the core coherence (20) directly, Theorem 7 allows us to improve the estimate $\mu_1^{d-1}$ times:

$$|\Omega| \geq \mu_1^{d-1} d^2 r^2 n \log(n).$$

Given this many elements of a tensor, Theorem 5 ensures, with high probability, that the RGD (15) converges locally to $A$. When $d = 2$, Theorems 6 and 7 are equivalent and repeat Reference 10 (Theorem 6).
3.4 | Tensor train completion with auxiliary subspace information

We would also like to show how the core coherence (20) can be used in a different setting. As an example, we choose tensor completion with auxiliary subspace information. In this scenario, in addition to the elements of the tensor $R_k \Omega A$, we know that the mode-$k$ fiber spans of $A$ belong to particular low-dimensional subspaces. Such formulations appear, for example, in multi-label learning\(^{26}\) and bioinformatics.\(^{27,28}\)

Namely, let matrices $Q_k \in \mathbb{R}^{n_k \times m_k}$ have orthonormal columns that span the subspaces in question. If $m_k = n_k$, no extra information is given about the mode-$k$ fibers. The unknown tensor $A$ can then be represented as

$$A = B \times_1 Q_1 \times_2 \ldots \times_d Q_d.$$  \hspace{1cm} (21)

This is a Tucker decomposition of $A$ with the Tucker core $B \in \mathbb{R}^{m_1 \times \ldots \times m_d}$ and Tucker factors $Q_k$. Since $A \in \mathcal{M}_r$, we can also write its minimal TT representation $A = [G_1, G_2, \ldots, G_d]$, and it follows that $B$ admits a minimal TT representation

$$B = [S_1, S_2, \ldots, S_d]$$

with TT cores $S_k = G_k \times_2 Q_k^T$ and the same TT ranks $r$. Due to orthogonality, we also have $G_k = S_k \times_2 Q_k$.

There are then two ways to look at TT completion with subspace information. First, it is a usual TT completion problem for $A$, where we know some of its elements $\Omega A$ and impose an additional constraint $G_k = G_k \times_2 Q_k Q_k^T$ on the TT cores. Second, we can treat it as a TT recovery problem for the small tensor $B$ with a special measurement operator

$$RB = R_k \Omega (B \times_1 Q_1 \times_2 \ldots \times_d Q_d).$$

Whatever the preferred point of view, the number of parameters that describe $A$ is $O(dm r^2)$, where $m = \max(m_1, \ldots, m_d)$ and $r = \max(r_1, \ldots, r_{d-1})$. Therefore, it is reasonable to expect that the required number of known elements $|\Omega|$ should be reduced in the presence of auxiliary subspace information. And we prove the corresponding Theorem. Informally, it states that if $A$ has bounded core coherence $\mu_c(A) \leq \mu_1$ and the auxiliary subspaces have bounded coherences $\mu(Q_k) \leq \mu_2$, then, with high probability, a modified RGD converges locally to $A$ when

$$|\Omega| \gtrsim \frac{\mu_1^{d-1} \mu_2 d^2 r^2 m \log(m)}{\mu_1 \mu_2}.$$  

This is the first theoretical result on the sample complexity of TT completion with auxiliary subspace information. We leave the more detailed and rigorous discussion for Appendix B.

4 | PROOFS OF MAIN RESULTS

4.1 | Curvature bound

To describe the tangent spaces to $\mathcal{M}_r$, consider minimal left- and right-orthogonal TT representations of $X \in \mathcal{M}_r$ denoted by

$$X = [U_1, \ldots, U_{d-1}, G_d] = [G_1, V_2, \ldots, V_d].$$

Every tangent vector $Y \in T_X \mathcal{M}_r$ can be uniquely represented as a sum $Y = \sum_{k=1}^d Y_k$ with non-minimal TT representations\(^{20}\)

$$Y_k = [U_1, \ldots, U_{k-1}, Y_k, V_{k+1}, \ldots, V_d],$$

where for $k \in [d - 1]$ the TT cores $Y_k \in \mathbb{R}^{n_{k-1} \times n_k \times r_k}$ satisfy the gauge conditions for the left unfoldings

$$(U_k^L)^T Y_k^L = 0 \in \mathbb{R}^{n \times r_k}.$$
The last TT core $Y_d$ does not have a gauge condition. On introducing the subspaces

$$T_k = \left\{ [U_1, \ldots, U_{k-1}, Y_k, V_{k+1}, \ldots, V_d] : Y_k \in \mathbb{R}^{r_k \times n_k \times r_k}, \left( U_k^T \right)^T Y_k = 0 \right\},$$

$$T_d = \left\{ [U_1, \ldots, U_{d-1}, Y_d] : Y_d \in \mathbb{R}^{r_d \times n_d \times r_d} \right\}$$

we can decompose the tangent space $T_X M_r$ into a direct orthogonal sum

$$T_X M_r = T_1 \oplus \ldots \oplus T_d. \quad (22)$$

A useful fact that is derived by simple inspection is that all tensors in the tangent space $T_X M_r$ have TT ranks that are at most $2r$. It suffices to see that a tangent vector $Y = \sum_{k=1}^d Y_k$ admits the following, non-minimal, TT representation

$$Y = \sum_{k=1}^d Y_k = \begin{bmatrix} Y_1 & U_1 \end{bmatrix}, \begin{bmatrix} V_2 & 0 \\ Y_2 & U_2 \end{bmatrix}, \ldots, \begin{bmatrix} V_{d-1} & 0 \\ Y_{d-1} & U_{d-1} \end{bmatrix}, \begin{bmatrix} V_d \\ Y_d \end{bmatrix},$$

we use block notation for the TT cores

$$\begin{bmatrix} Y_1 & U_1 \end{bmatrix} \in \mathbb{R}^{r_1 \times n_1 \times 2r_1}, \begin{bmatrix} V_k & 0 \\ Y_k & U_k \end{bmatrix} \in \mathbb{R}^{2r_k \times n_k \times 2r_k}, \begin{bmatrix} V_d \\ Y_d \end{bmatrix} \in \mathbb{R}^{2r_d \times n_d \times r_d}. $$

The formula for the orthogonal projection onto the tangent space $T_X M_r$ was derived in Reference 29. To introduce it, we need to define the tensorization operation that reverts unfoldings to tensors

$$X = \text{ten}_k(X^{(k)}).$$

Consider the interface matrices $X_{sk}$ and $X_{2k+1}$ for $k \in [d-1]$. Let

$$P_{sk} = U_{sk}U_k^T \in \mathbb{R}^{(n_1 \ldots n_k) \times (n_1 \ldots n_k)}, \quad P_{2k+1} = V_{2k+1}V^T \in \mathbb{R}^{(n_{k+1} \ldots n_d) \times (n_{k+1} \ldots n_d)}$$

be the orthogonal projection onto their column spans. Owing to (12), we can write them down recursively as

$$U_{sk} = U_k^T, \quad U_{sk} = (I_{n_k} \otimes U_{sk-1})U_k^T \in \mathbb{R}^{(n_1 \ldots n_k) \times n_k},$$

$$V_{2k+1} = (V_{2k+1}^R)^T, \quad V_{2k+1} = (V_{2k+2} \otimes I_{n_{k+1}})(V_{k+1}^R)^T \in \mathbb{R}^{(n_{k+1} \ldots n_d) \times n_k}.$$

The orthogonal projection operator onto the tangent space $P_{T_X M_r} : \mathbb{R}^{n_1 \times \ldots \times n_d} \rightarrow T_X M_r$ is then given by

$$P_{T_X M_r} = \sum_{k=1}^{d-1} (P_{2k-1} - P_{sk})P_{2k+1} + P_{2d-1}, \quad (23)$$

where

$$P_{sk} : Z \mapsto \text{ten}_k(P_{sk}Z^{(k)}), \quad P_{2k+1} : Z \mapsto \text{ten}_k(Z^{(k)}P_{2k+1}), \quad P_{2d} = \text{Id}. $$

Let us try to better understand the roles played by each individual projection operator $P_{sk}$ and $P_{2k+1}$ (23). Denote by

$$X = [U_1, \ldots, U_{d-1}, G_d] = [G_1, V_2, \ldots, V_d].$$

are minimal left- and right-orthogonal TT representations of $X$. Consider a tensor $Z$ of TT rank $r'$ with minimal TT representation $Z = [C_1, \ldots, C_d]$. The projection $P_{sk}$ onto the column span of the left interface matrix results in a tensor with a non-minimal TT representation

$$P_{sk}Z = [U_1, \ldots, U_{k-1}, \overline{U}_k, C_{k+1}, \ldots, C_d]$$
by replacing the $k-1$ leftmost TT cores of $Z$ with the left-orthogonal TT cores of $X$, keeping the $d-k$ rightmost TT cores of $Z$, and computing a new TT core $\overline{U}_k$ such that $U_k = U_k^R W_k$ for a square matrix $W_k \in \mathbb{R}^{r \times r}$. In the same vein, $P_{\geq k+1}$ produces

$$P_{\geq k+1} Z = [C_1, \ldots, C_k, \overline{V}_{k+1}, V_{k+2}, \ldots, V_d]$$

with $\overline{V}_{k+1} = H_{k+1} V_{k+1}^R$, $H_{k+1} \in \mathbb{R}^{r \times r}$. It is important to note that $W_k$ and $H_{k+1}$ can be taken out of $U_k$ and $\overline{V}_{k+1}$ and multiplied onto $C_{k+1}$ and $C_k$, respectively, instead:

$$P_{\leq k} Z = [U_1, \ldots, U_{k-1}, U_k, \overline{C}_{k+1}, \ldots, C_d], \quad \overline{C}_{k+1} = W_k C_{k+1}^R,$$

$$P_{\geq k+1} Z = [C_1, \ldots, \overline{C}_k, V_{k+1}, V_{k+2}, \ldots, V_d], \quad \overline{C}_k = C_k^T H_{k+1}.$$

We can also deduce from these formulations that $P_{\leq j}$ and $P_{\geq k}$ commute when $j < k$ (even when they are connected with different tangent spaces).

Going further, we see that $P_{\leq k-1} Z - P_{\leq k}$ is a projection operator as well. Indeed, it multiplies $Z^{(k)}$ by an orthogonal projection on the left:

$$(P_{\leq k-1} - P_{\leq k}) Z = \text{ten}_k \left( [(I_{n_k} \otimes P_{\leq k-1}) - P_{\leq k}] Z^{(k)} \right)$$

$$= \text{ten}_k \left( (I_{n_k} \otimes U_{\leq k-1})(I_{n_k} - U_k^T U_k^R)(I_{n_k} \otimes U_k^T Z^{(k)}) \right).$$

It is clear that for $k \in [d-1]$ every $P_{\leq k-1} Z - P_{\leq k}$ acts by imposing the orthogonal gauge condition onto the $k$th TT core:

$$(P_{\leq k-1} Z - P_{\leq k}) Z = [U_1, \ldots, U_{k-1}, Y_k, C_{k+1}, \ldots, C_d], \quad (U_k^T)^T Y_k^T = 0.$$

And by analogy for $k \in [d-1]$ (denote $P_{\geq d+1} = \text{Id}$) we have

$$(P_{\leq k+2} Z - P_{\geq k+1}) Z = [C_1, \ldots, C_k, \Xi_{k+1}, V_{k+2}, \ldots, V_d], \quad \Xi_{k+1}^T V_{k+1}^R = 0.$$

We can now align the decomposition of the tangent space (22) with the definition of the orthogonal projection operator (23) since

$$(P_{\leq k-1} - P_{\leq k}) P_{\geq k+1} : \mathbb{R}^{n_1 \times \cdots \times n_d} \to T_k, \quad k \in [d-1],$$

$$P_{\leq d-1} : \mathbb{R}^{n_1 \times \cdots \times n_d} \to T_d.$$

The complementary orthogonal projection operator admits a simple expression too

$$\text{Id} - P = \sum_{k=1}^{d-1} (P_{\leq k-1} - P_{\leq k})(\text{Id} - P_{\leq k+1}),$$

where we can represent each $\text{Id} - P_{\geq k+1}$ as a sum of projections that we already understand:

$$\text{Id} - P_{\geq k+1} = \sum_{j=k}^{d-1} (P_{\leq j+2} - P_{\leq j+1}).$$

Now, we are in position to prove Lemma 1. At first, let us show that

$$\|P_{\leq k} Z - P_{\leq k}\| \leq \frac{\|X - \hat{X}\|}{\sigma_{\min}(X^{(k)})}, \quad \|P_{\geq k+1} Z - P_{\geq k+1}\| \leq \frac{\|X - \hat{X}\|}{\sigma_{\min}(X^{(k)})}.$$
Let $X^{(k)} = UΣVT$ be the truncated SVD of rank $r_k$. Then we have

$$
\|P_{≤k} - P_{≤k}\| = \|(I - P_{≤k})P_{≤k}\| = \|(I - P_{≤k})UUT\| = \|(I - P_{≤k})X^{(k)}VΣ^{-1}U^T\|
$$

$$
= \|(I - P_{≤k})(X^{(k)} - X^{(k)})VΣ^{-1}U^T\|
$$

$$
≤ \|(I - P_{≤k})\|X^{(k)}\|V\|Σ^{-1}\|U^T\|
$$

$$
= \|X^{(k)} - X^{(k)}\|/\sigma_{\text{min}}(X^{(k)}) \leq \|X - \bar{X}\|_F/\sigma_{\text{min}}(X^{(k)}).
$$

An analogous argument works for the right interface matrix.

For brevity, denote $P_{T_xM_r}$ as $P$ and $P_{T_xM_r}$ as $\bar{P}$. Using the decomposition of $\text{Id} - \bar{P}$ we prove the first part of the Lemma:

$$
\|(\text{Id} - \bar{P})X\|_F = \left\| \sum_{k=1}^{d-1} (\bar{P}_{2k} - \bar{P}_{≤k})(\text{Id} - \bar{P}_{2k+1})X \right\|_F \leq \sum_{k=1}^{d-1} \left\| (\bar{P}_{2k+1} - \bar{P}_{≤k})(\text{Id} - \bar{P}_{2k+1})X \right\|_F
$$

$$
= \sum_{k=1}^{d-1} \left\| \bar{P}_{2k+1} - \bar{P}_{≤k} \right\| \left\| \bar{P}_{2k+1} - \bar{P}_{≤k} \right\| \|X - \bar{X}\|_F = \sum_{k=1}^{d-1} \left\| \bar{P}_{2k+1} - \bar{P}_{≤k} \right\| \|X - \bar{X}\|_F
$$

$$
\leq \|X - \bar{X}\|_F \sum_{k=1}^{d-1} \frac{1}{\sigma_{\text{min}}(X^{(k)})}.
$$

Above, we also use the identity $X = P_{≥k}X$, the commutativity of $\bar{P}_{≤k}$ and $P_{≥k+1}$; the identity $\bar{P}_{≤k}X = \bar{P}_{≤k}X$; and the fact that $\bar{P}_{≤k} - \bar{P}_{≤k}$ is a projector. Straightforward calculation shows that

$$
P - \bar{P} = \sum_{k=1}^{d-1} \left[ (P_{≤k} - \bar{P}_{≤k})(P_{≥k+2} - P_{≥k+1}) + (\bar{P}_{≤k+1} - \bar{P}_{≤k})(P_{≥k+1} - \bar{P}_{≥k+1}) \right].
$$

Then the second assertion follows from

$$
\|P - \bar{P}\| \leq \sum_{k=1}^{d-1} \left[ \|P_{≤k} - \bar{P}_{≤k}\|\|P_{≥k+2} - P_{≥k+1}\| + \|\bar{P}_{≤k+1} - \bar{P}_{≤k}\|\|P_{≥k+1} - \bar{P}_{≥k+1}\| \right]
$$

$$
= \sum_{k=1}^{d-1} \left[ \|P_{≤k} - \bar{P}_{≤k}\| + \|P_{≥k+1} - \bar{P}_{≥k+1}\| \right] \leq 2\|X - \bar{X}\|_F \sum_{k=1}^{d-1} \frac{1}{\sigma_{\text{min}}(X^{(k)})}.
$$

### 4.2 Local convergence of Riemannian gradient descent: Tensor train recovery

To prove Theorem 4, we need to establish a technical Lemma 5. It shows that the standard RIP (14) implies that the RIP (16) holds globally on $M_r$, that is, for all of its tangent spaces.

**Lemma 5.** Let the linear operator $R$ satisfy the standard RIP (14) of order $2r$ with a RIP constant $0 < \delta_{2r} < 1$. Then for an arbitrary tensor $X \in M_r$ of TT rank $r$ the following RIP holds with the same constant:

$$
\|P_{T_xM_r} - P_{T_xM_r} R^* R P_{T_xM_r}\| < \delta_{2r}.
$$
Proof. Observe that $P_{T_rM_r} - P_{T_rM_r}^* R P_{T_rM_r}$ is a self-adjoint operator so its norm can be characterized as
\[ \| P_{T_rM_r} - P_{T_rM_r}^* R P_{T_rM_r} \| = \max_{Z: \|Z\|_F = 1} \langle (P_{T_rM_r} - P_{T_rM_r}^* R P_{T_rM_r})Z, Z \rangle_F. \]

It follows that
\[ \| P_{T_rM_r} - P_{T_rM_r}^* R P_{T_rM_r} \| = \max_{Z: \|Z\|_F = 1} \left( \| P_{T_rM_r} Z \|_F^2 - \| R P_{T_rM_r} Z \|_F^2 \right) \leq \max_{Z: \|Z\|_F = 1} (\delta_{2r} \| P_{T_rM_r} Z \|_F^2) \leq \delta_{2r}, \]
because the elements of every tangent space to $M_r$ have ranks equal to at most $2r$.

In the proof of Theorem 4, we are only going to use the result of Lemma 5 and not the standard RIP (14) itself. This explains why the proof can be adapted to the TT completion case.

Proof of Theorem 4. The new iterate is given by (13) so by using the quasi-optimality of TT-SVD projection we get
\[ \| x_{t+1} - A \|_F = \| \text{TT-SVD}_r(x_t - a_t y_t) - A \|_F \]
\[ \leq \| \text{TT-SVD}_r(x_t - a_t y_t) - (x_t - a_t y_t) \|_F + \| (x_t - a_t y_t) - A \|_F \]
\[ \leq \sqrt{d - 1} \max_{r} \| (x_t - a_t y_t - (x_t - a_t y_t)) \|_F + \| (x_t - a_t y_t) - A \|_F \]
\[ \leq (1 + \sqrt{d - 1}) \| (x_t - a_t y_t) - A \|_F. \]

We then separate this Frobenius norm into a sum of several components that we will bound one by one
\[ \| (x_t - a_t y_t) - A \|_F = \| x_t - a_t P_X^* R (x_t - A) - A \|_F = \| (1 - a_t P_X^* R) (x_t - A) \|_F \]
\[ \leq \| (1 - P_X) (x_t - A) \|_F + \| P_X^* R P_X (x_t - A) \|_F \]
\[ + |1 - a_t| \| P_X^* R P_X (x_t - A) \|_F + |a_t| \| P_X^* R (1 - P_X) (x_t - A) \|_F. \]

For the first term we use the curvature bound Lemma 1 to get
\[ \| (1 - P_X) (x_t - A) \|_F \leq (d - 1) \max_{r} \| x_t - A \|_F^2. \]

The bound for the second term follows from Lemma 5:
\[ \| (P_X^* R P_X (x_t - A) \|_F \leq \delta_{2r} \| x_t - A \|_F. \]

To estimate the third term, we note that the step size $a_t = \| y_t \|_F^2/\| R y_t \|_F^2$ is close to one. Indeed, $y_t$ has TT ranks at most $2r$ since it belongs to the tangent space and so
\[ \frac{1}{1 + \delta_{2r}} \leq a_t \leq \frac{1}{1 - \delta_{2r}}. \]

We then use the variational characterization of the Frobenius norm
\[ \| P_X^* R P_X (x_t - A) \|_F = \max_{Z: \|Z\|_F = 1} \langle P_X^* R P_X (x_t - A), Z \rangle_F = \max_{Z: \|Z\|_F = 1} \langle R P_X (x_t - A), R P_X Z \rangle_F \]
\[ \leq \max_{Z: \|Z\|_F = 1} \| R P_X (x_t - A) \|_F \| R P_X Z \|_F \leq (1 + \delta_{2r}) \| P_X (x_t - A) \|_F \| P_X Z \|_F \]
\[ \leq (1 + \delta_{2r}) \| x_t - A \|_F. \]
Thus the third term is bounded by
\[ |1 - \alpha_i| \|P_XR^*RP_X(X_i - A)\|_F \leq \delta_2r \frac{1 + \delta_2r}{1 - \delta_2r} \|X_i - A\|_F. \]

For the fourth term, we use the operator norm bound \( \|R^*R\| \leq C \):
\[ |\alpha_i| \|P_XR^*(Id - P_X)(X_i - A)\|_F \leq \frac{C}{1 - \delta_2r} (d - 1) \frac{\|X_i - A\|^2_F}{\sigma_{\min}(A)}. \]

Finally, collecting the terms, we get
\[ \|X_{i+1} - A\|_F \leq (1 + \sqrt{d - 1}) \left[ \frac{2\delta_2r}{1 - \delta_2r} + \left(1 + \frac{C}{1 - \delta_2r}(d - 1) \frac{\|X_i - A\|_F^2}{\sigma_{\min}(A)} \right) \right] \|X_i - A\|_F. \]

If the initial condition \( X_0 \in M_r \) is close enough
\[ (d - 1) \frac{\|X_0 - A\|_F}{\sigma_{\min}(A)} < \frac{1}{1 + C - \delta_2r} \left( \frac{1 - \delta_2r}{1 + \sqrt{d - 1}} - 2\delta_2r \right), \]
the rate \( \beta_0 \) becomes smaller than one and as a consequence \( \beta_i < \beta_0 < 1 \).

To prove the final assertion we note that the TT rank of \((Id - P_X)(X_i - A)\) is at most \( 3r \) and so the standard RIP can be used to estimate the fourth term:
\[ |\alpha_i| \|P_XR^*(Id - P_X)(X_i - A)\|_F \leq \frac{1 + \delta_3r}{1 - \delta_2r} (d - 1) \frac{\|X_i - A\|^2_F}{\sigma_{\min}(A)}. \]

where we used the variational form of the Frobenius norm and the fact that \( \delta_2r \leq \delta_3r \).

### 4.3 Local convergence of Riemannian gradient descent: Tensor train completion

For TT completion, we need an analog of Lemma 5. While the standard RIP (14) guarantees that the RIP (16) holds globally, Lemma 6 shows that if the RIP (16) holds on a particular tangent space, so does it locally on the neighboring tangent spaces, though with a degrading constant.

**Lemma 6.** Let \( A \in M_r \) be a tensor of TT rank \( r \) and suppose that \( R_{\Omega} \) satisfies the RIP (16) and is bounded
\[ \|P_{T_XM_r} - \rho^{-1}P_{T_XM_r}R_{\Omega}P_{T_XM_r}\| < \epsilon, \quad \|R_{\Omega}\| \leq C. \]

Then for every tensor \( X \in M_r \), with a sufficiently close tangent space \( \|P_{T_XM_r} - P_{T_XM_r}\| < \delta \), the sampling operator \( R_{\Omega} \) satisfies the RIP (16) on it as well
\[ \|P_{T_XM_r} - \rho^{-1}P_{T_XM_r}R_{\Omega}P_{T_XM_r}\| < E(\delta) \equiv \epsilon + \delta \left(1 + 2C\rho^{-1}\right). \]

**Proof.** Denote by \( P_A \) the projection \( P_{T_XM_r} \) and similarly for \( P_X \). Then
\[ \|P_X - \rho^{-1}P_XR_{\Omega}P_X\| \leq \|P_A - \rho^{-1}PA\| + \|P_X - P_A\| + \rho^{-1}\|P_XR_{\Omega}P_X - PA\|| \]
\[ \leq \epsilon + \delta + \rho^{-1}\|P_XR_{\Omega}P_X - PA\| + \rho^{-1}\|P_XR_{\Omega}P_X - PA\| \leq \epsilon + \delta + \rho^{-1}\|P_X - P_A\| \]
\[ \leq \epsilon + \delta \left(1 + 2C\rho^{-1}\right). \]
A tighter bound can be derived if we estimate \( \|R_{\Omega}P_A\| \) with more care using the RIP, see Reference 14. 

\[ \Box \]
Knowing how the RIP (16) behaves in a neighborhood of $A$, we can adapt the proof of Theorem 4. Note that neither Lemma 6 nor Theorem 5 exploits the actual nature of $R_{\Omega}$, so their proofs apply to any other measurement operator with the RIP (16).

**Proof of Theorem 5.** We basically repeat the proof of Theorem 4 with certain modifications related to the RIP (16). We immediately get that

$$\|X_{t+1} - A\|_F \leq (1 + \sqrt{d-1})\|X_t - a_t Y_t - A\|_F$$

and

$$\|X_t - a_t Y_t - A\|_F \leq \|(\text{Id} - P_{X_t})X_t - A\|_F + \|P_{X_t} - \rho^{-1}P_{X_t} R_{\Omega X_t} (X_t - A)\|_F + \|\rho^{-1} - a_t\| \|P_{X_t} R_{\Omega X_t} (X_t - A)\|_F.$$ 

Each term is then estimated using Lemmas 1 and 6. We only need to bound $a_t = \|Y_t\|_F^2 / (R_{\Omega} Y_t, Y_t)_F$. Denote

$$\delta_t = 2(d - 1) \|X_t - A\|_F^2 / \sigma_{\min}(A).$$

The operator $P_{X_t} - \rho^{-1}P_{X_t} R_{\Omega X_t}$ being self-adjoint, we have

$$-E(\delta_t)(Y_t, Y_t)_F < (\rho^{-1} P_{\Omega X_t} - P_t) Y_t, Y_t)_F < E(\delta_t)(Y_t, Y_t)_F.$$

As a consequence,

$$\frac{\rho^{-1}}{1 + E(\delta_t)} \leq a_t \leq \frac{\rho^{-1}}{1 - E(\delta_t)}$$

and the Theorem follows. □

### 4.4 Recovery guarantees

The interface incoherence assumption that we make in Theorem 6 allows us to estimate the norm of the projection of a canonical basis tensor $E_{\omega} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ onto the tangent space $T_A \mathcal{M}_r$, that is, estimate the coherence of the tangent space.

**Lemma 7.** Let $A \in \mathcal{M}_r$ be a tensor of TT rank $r$ with bounded interface coherence $\mu_t(A) \leq \mu_0$. Then for every canonical basis tensor $E_{\omega}$, $\omega \in [n_1] \times \cdots \times [n_d]$, its projection onto the tangent space $T_A \mathcal{M}_r$ can be bounded from above as

$$\|P_{T_A \mathcal{M}_r} E_{\omega}\|_F^2 \leq C_0 \equiv \frac{\mu_0}{n_1 \cdots n_d} \left( n_1 r_1 + \mu_0 \sum_{k=2}^{d-1} r_k r_k + r_{d-1} n_d \right).$$

**Proof.** Every canonical basis tensor $E_{\omega}$ can be represented as an outer product of canonical basis vectors $E_{\omega} = e_{i_1} \cdots e_{i_d}$ with $e_{i_k} \in \mathbb{R}^{n_k}$. Then using the definition of the projection onto the tangent space (23) we get

$$\|P_{T_A \mathcal{M}_r} E_{\omega}\|_F^2 = \sum_{k=1}^{d-1} \left[ \|P_{\Sigma k-1} P_{\Sigma k+1} E_{\omega}\|_F^2 - \|P_{\Sigma k} P_{\Sigma k+1} E_{\omega}\|_F^2 \right] + \|P_{\Sigma d-1} E_{\omega}\|_F^2$$

$$\leq \|P_{\Sigma 2} E_{\omega}\|_F^2 + \sum_{k=2}^{d-1} \|P_{\Sigma k-1} P_{\Sigma k+1} E_{\omega}\|_F^2 + \|P_{\Sigma d-1} E_{\omega}\|_F^2.$$ 

The first and last terms are bounded directly using the interface incoherence property because

$$\|P_{\Sigma 2} E_{\omega}\|_F^2 = \|e_{i_1} (e_{i_2} \otimes \cdots \otimes e_{i_d})^T P_{\Sigma 2}\|_F^2 = \|P_{\Sigma 2} (e_{i_1} \otimes \cdots \otimes e_{i_d})\|_2^2 \leq \frac{r_1}{n_2 \cdots n_d} \mu_0$$

and

$$\|P_{\Sigma d-1} E_{\omega}\|_F^2 \leq \|P_{\Sigma d-1} (e_{i_1} \otimes \cdots \otimes e_{i_d})\|_2^2 \leq \frac{r_{d-1}}{n_1 \cdots n_{d-2}} \mu_0.$$
and

\[ \| P_{\leq d-1} E_\omega \|_F^2 = \| P_{\leq d-1} (e_{i_{d-1}} \otimes \ldots \otimes e_{i_1}) e_{i_d}^T \|_F^2 = \| P_{\leq d-1} (e_{i_{d-1}} \otimes \ldots \otimes e_{i_1}) \|_F^2 \leq \frac{r_{d-1}}{n_1 \ldots n_{d-1}} \mu_0. \]

We then estimate every summand \( \| P_{\leq k-1} P_{\geq k+1} E_\omega \|_F^2 \) as follows

\[ \| P_{\leq k-1} P_{\geq k+1} E_\omega \|_F^2 = \| P_{\leq k-1} (e_{i_{k-1}} \otimes \ldots \otimes e_{i_1}) \circ e_{i_k} \circ P_{\geq k+1} (e_{i_{k+1}} \otimes \ldots \otimes e_{i_1}) \|_F^2 \]
\[ = \| P_{\leq k-1} (e_{i_{k-1}} \otimes \ldots \otimes e_{i_1}) \|_2^2 \cdot \| P_{\geq k+1} (e_{i_{k+1}} \otimes \ldots \otimes e_{i_1}) \|_2 \]
\[ \leq \frac{r_{k-1}}{n_1 \ldots n_{k-1}} \mu_0 \cdot \frac{r_k}{n_k \ldots n_d} \mu_0. \]

It remains to add the estimates together.

The main probabilistic tool we need in order to prove Theorem 6 is the noncommutative Bernstein inequality (Theorem 8), which is used in analyzing large deviation bounds.

**Theorem 8** (Reference 10, Theorem 4). Let \( X_1, \ldots, X_K \in \mathbb{R}^{n_1 \times n_2} \) be independent zero-mean random matrices. Suppose that

\[ \sigma_k^2 = \max \left( \| \mathbb{E} [X_k X_k^T] \|_2, \| \mathbb{E} [X_k^T X_k] \|_2 \right) \]

and \( \| X_k \|_2 \leq R \) almost surely for every \( k \). Then for any \( \tau > 0 \),

\[ \mathbb{P} \left\{ \left\| \sum_{k=1}^K X_k \right\|_2 > \tau \right\} \leq (s_1 + s_2) \exp \left( \frac{-\tau^2/2}{\sum_{k=1}^K \sigma_k^2 + R \tau / 3} \right). \]

If in addition \( \tau \leq \sum_{k=1}^K \sigma_k^2 / R \),

\[ \mathbb{P} \left\{ \left\| \sum_{k=1}^K X_k \right\|_2 > \tau \right\} \leq (s_1 + s_2) \exp \left( \frac{-3 \tau^2 / 8}{\sum_{k=1}^K \sigma_k^2} \right). \]

**Proof of Theorem 6.** Since any tensor \( Z \) can be represented as a linear combination of canonical basis tensors

\[ Z = \sum_{\omega \in [n_1] \times \ldots \times [n_d]} Z(i_1, \ldots, i_d) E_\omega = \sum_{\omega \in [n_1] \times \ldots \times [n_d]} \langle Z, E_\omega \rangle_F E_\omega, \]

the application of the operator \( P_A R_\Omega P_A \)—where we once again write \( P_A \) as a shorthand for \( P_{\Omega_1 \cdots \Omega_d} \)—can be computed as

\[ P_A R_\Omega P_A Z = P_A \left( \sum_{\omega \in [n_1] \times \ldots \times [n_d]} \langle P_A Z, E_\omega \rangle_F E_\omega \right) = \sum_{\omega \in [n_1] \times \ldots \times [n_d]} \langle Z, P_A E_\omega \rangle_F P_A E_\omega. \]

Every \( \omega \in \Omega \) is a uniformly distributed random variable so \( P_A R_\Omega P_A \) is a sum of \( |\Omega| \) i.i.d. random operators

\[ P_A R_\Omega P_A = \sum_{\omega \in [n_1] \times \ldots \times [n_d]} S_\omega, \quad S_\omega Z = \langle Z, P_A E_\omega \rangle_F P_A E_\omega. \]

The expected value of \( S_\omega \) is \( \frac{1}{n_1 \ldots n_d} P_A \) and we can estimate the norm of the deviation as

\[ \| S_\omega - \frac{1}{n_1 \ldots n_d} P_A \| \leq \max \left( \| S_\omega \|, \frac{1}{n_1 \ldots n_d} \| P_A \| \right) = \max \left( \| P_A E_\omega \|_F^2, \frac{1}{n_1 \ldots n_d} \right) = C_0. \]
The first inequality holds since both $S_\omega$ and $\frac{1}{n_1 \ldots n_d} P_A$ are positive semidefinite. To apply the noncommutative Bernstein inequality we also need a bound for the variance of $S_\omega$:

$$\left\| \mathbb{E}(S_\omega - \frac{1}{n_1 \ldots n_d} P_A)^2 \right\| = \left\| \mathbb{E}(\|P_A E_\omega\|_F^2 S_\omega) - \frac{1}{(n_1 \ldots n_d)^2} P_A \right\| \leq \max \left( \mathbb{E}(\|P_A E_\omega\|_F^2 S_\omega), \frac{1}{(n_1 \ldots n_d)^2} \right)$$

$$\leq \max \left( \frac{C_0}{n_1 \ldots n_d}, \frac{1}{(n_1 \ldots n_d)^2} \right) = \frac{C_0}{n_1 \ldots n_d}.$$

We then apply the second part of Theorem 8 to $S_\omega = \frac{1}{n_1 \ldots n_d} P_A$ for $\omega \in \Omega$. When $\tau / \rho = \epsilon < 1$, we have

$$\mathbb{P} \left( \|P_A - \rho^{-1} P_A R_\omega P_A\| > \tau / \rho = \epsilon \right) \leq 2(n_1 \ldots n_d) \exp \left( -\frac{3}{8} \epsilon \rho C_0 \right) \leq 2n^d \exp \left( -\frac{3}{8} \rho C_0 \right) \leq 2n^{d(1-\beta)}$$

provided that $\rho \geq \frac{8 C_0}{3 \epsilon^2} d \beta \log(n)$. $\blacksquare$

Theorem 7 is proved in exactly the same way as Theorem 6, except we use a refined estimate of $\|P_{T_A M_r} E_\omega\|_F$ that is given in the following Lemma 8.

**Lemma 8.** Let $A \in M_r$ be a tensor of TT rank $r$ with bounded core coherence $\mu_c(A) \leq \mu_1$. Then for every canonical basis tensor $E_\omega$, $\omega \in [n_1] \times \ldots \times [n_d]$, its projection onto the tangent space $T_A M_r$ can be bounded from above as

$$\|P_{T_A M_r} E_\omega\|_F^2 \leq \frac{\mu_1^{d-1}}{n_1 \ldots n_d} \sum_{k=1}^d r_{k-1} n_k r_k.$$

**Proof.** The proof goes along the same line as Lemma 7 and uses Lemma 4 to obtain the bounds. $\blacksquare$

## 5 | DISCUSSION

The sample complexities that we obtained for TT completion (Theorem 7) and TT completion with auxiliary subspace information (Theorem 9) depend on the core coherence (20) as $\mu_c(A)^{d-1}$. It is, thus, important to have a qualitative estimate of how large the core coherence can be. In the matrix case, Reference 8, $\mu_c(A)$ was proved to be of order $\max(r, \log(n))$ for matrices, whose left and right singular factors are chosen uniformly at random from the set of orthonormal columns $S_{[n]}$. To sample such factors, one can take a random $n \times r$ matrix with i.i.d standard Gaussian entries and apply Gram-Schmidt orthogonalization.

Consider now a minimal left-orthogonal TT representation of a tensor $A = [U_1, \ldots, U_{d-1}, G_d]$, whose TT cores $U_k$ are sampled uniformly from $S_{(r_{k-1} n_k, r_k)}$. What can be said about the distribution of their subblocks $U_k^{(l)}$ that are used in the definition of the left coherence of a TT core (18)? It is known that if we take a random orthogonal matrix $Q_{p,q}^{(n)} \in \mathbb{R}^{p \times q}$, pick any of its subblocks $Q_{p,q}^{(n)} \in \mathbb{R}^{p \times q}$, and let $n \to \infty$, then the matrix $\sqrt{n} Q_{p,q}^{(n)}$ converges in distribution to a matrix with i.i.d. standard Gaussian entries. As a consequence, we can, informally, treat the blocks $\sqrt{n_k} U_k^{(l)}$ as random matrices sampled from the standard Gaussian distribution. Random matrix theory provides probabilistic estimates on the spectral norm of a standard Gaussian random matrix. With probability at least $1 - \epsilon$ $\exp(-t^2/2)$, we have

$$\|\sqrt{n_k} U_k^{(l)}\|_2 \leq \sqrt{r_{k-1}} + \sqrt{r_k} + t.$$

It follows that, with high probability,

$$\frac{r_{k-1} n_k}{r_k} \|U_k^{(l)}\|_2 \leq \frac{r_{k-1}}{r_k} (\sqrt{r_{k-1}} + \sqrt{r_k} + t)^2$$

and so $\mu_c(A)$ should be of order $\max(r, \log(n))$ as well, if we set $t = c \sqrt{\log(n)}$. 


The exponential dependence \( \mu(A)^{d-1} \) originates in Lemma 4, where we bound the spectral norms of the row vectors

\[
U_1^{(i_1)} U_2^{(i_2)} \ldots U_k^{(i_k)}
\]

using the submultiplicative property. Assume, once again, that we can informally treat the subblocks \( \sqrt{n_k} U_k^{(i_k)} \) as standard Gaussian random matrices. The product of a Gaussian random matrix and a Gaussian random vector has a known distribution. In our case, the first product \((U_1^{(i_1)} U_2^{(i_2)})^T \in \mathbb{R}^{r_2}\) is distributed as

\[
\sqrt{n_1 n_2} (U_1^{(i_1)} U_2^{(i_2)})^T \sim \sqrt{s_1(r_1)} z,
\]

where \( s_1(r_1) \sim \chi^2(r_1) \) is a chi-squared random variable with \( r_1 \) degrees of freedom and \( z \in \mathbb{R}^{r_2} \) is a standard Gaussian random vector independent of \( s_1 \). Multiplying further, we find that

\[
\sqrt{n_1 \ldots n_k} (U_1^{(i_1)} U_2^{(i_2)} \ldots U_k^{(i_k)})^T \sim \sqrt{s_1(r_1) \ldots s_k(r_k)} z
\]

with a standard Gaussian random vector \( z \in \mathbb{R}^{r_k} \). The squared Euclidean norm of this vector is distributed as a product of \( k \) independent chi-squared random variables with the number of degrees of freedom equal to the corresponding TT rank:

\[
\begin{align*}
(n_1 \ldots n_k) \| U_1^{(i_1)} U_2^{(i_2)} \ldots U_k^{(i_k)} \|^2_2 &\sim s_1(r_1) \ldots s_k(r_k).
\end{align*}
\]

Its expectation is a good reference value to compare \( \mu(A_{\omega_k}) \) against:

\[
\mathbb{E} \left\{ \frac{n_1 \ldots n_k}{r_k} \| U_1^{(i_1)} U_2^{(i_2)} \ldots U_k^{(i_k)} \|^2_2 \right\} = r_1 \ldots r_{k-1} \leq r^{k-1}.
\]

The exponential dependence on \( k \) leads to the exponential dependence on \( d \) in the sample complexity via Lemma 8.

It is possible, however, that the distribution is not concentrated around the expected value but is spread out, that is, the majority of random row-vectors \( U_1^{(i_1)} U_2^{(i_2)} \ldots U_k^{(i_k)} \) has very small norms. In other words, for a significant subset of multi-indices \( \omega \) the projections \( \| \mathcal{P}_\omega \mathbf{E}_\omega \|_F \) might be small. In this case, the Bernstein inequality, which is the crux of Theorem 7, can produce crude estimates—as it requires a uniform upper bound of the random variable that holds almost surely—and a different tail bound such as in Reference 35 could lead to finer results. To check this hypothesis, we can estimate the median of \( s_1(r_1) \ldots s_k(r_k) \) in a numerical simulation. Unfortunately, the results in Figure 1 show that the median grows exponentially too, and so the squared norm is of order \( r^k \) for many row-vectors.

Still, we hope that the 'true' estimate of \( |\Omega| \) should not depend exponentially on the number of dimensions \( d \), the phase plot in Figure 2 supports our hopes. We applied the RGD (15) to TT completion with \( n = 50, r = (3, \ldots, 3) \), and varying number of dimensions \( d \) and sample size \( |\Omega| \); for every combination of \( d \) and \( |\Omega| \) we carried out 5 random experiments.

**FIGURE 1** Numerically computed median of \( \prod_{j=1}^{k} \chi^2(5) \).
Figure 2 Phase plot of the Riemannian gradient descent for \( n = 50, r = 3 \), and varying number of dimensions \( d \). The values between 0 and 1 are the frequencies of successful recovery for the given parameters. The orange (solid) and blue (dashed) curves correspond to \(|\Omega| = d^2 r^2 n \log(n)/10 \) and \(|\Omega| = d^2 r^2 n \log(n)/10 \), respectively.

In each of them, we 1) generated a random tensor \( A \) and a random initial approximation \( X_0 \), both of TT rank \( r \), with i.i.d. standard Gaussian TT cores; 2) generated a uniformly distributed sampling set \( \Omega_1 \) and a uniformly distributed test set \( \Omega_2 \), both of size \(|\Omega|\); 3) ran 500 iterations of the RGD with data \( \mathbf{R}_{\Omega_1} A \) starting from \( X_0 \); 4) and called the iterations successful if the relative error on the test set \( \Omega_2 \) was below \( 10^{-4} \):

\[
\| \mathbf{R}_{\Omega_2} A - \mathbf{R}_{\Omega_2} X_{500} \|_F < 10^{-4} \| \mathbf{R}_{\Omega_2} A \|_F.
\]

The implementation of the RGD was taken from the TTeMPS Toolbox (the RTTC method, see https://www.epfl.ch/labs/anchp/index-html/software/ttemps/). The phase plot in Figure 2 shows the frequency of success for every combination of \( d \) and \(|\Omega|\). We see that the phase transition curve between the “never successful” (black) and “always successful” (white) regions seems to exhibit polynomial growth.

The practical implications of the polynomial dependence are best seen on the following example. Consider a dataset of \( 2^d \) real numbers, of which we know only \(|\Omega|\). We can rearrange it into a \( d \)-dimensional tensor and, if \( r \) is its largest TT rank, recover all of the data from \(|\Omega| \geq d^2 r^2 \) elements, as Figure 2 suggests. If, instead, we ignore the tensor structure and treat the data as a \( 2^{d-1} \times 2^{d-1} \) matrix, the same bound evaluates to \(|\Omega| \geq d^{2d-2} \). Therefore, using the tensor structure (if it exists) of a large dataset, we shall manage to recover it from significantly fewer elements.

A different kind of reasoning might be needed to bridge the gap between the theoretical exponential bound (7) and the numerical polynomial bound from Figure 2. One possible direction can lie in relaxing the RIP (16). Currently, it states that the sampling operator \( \mathbf{R}_\Omega \) is well-conditioned on the whole tangent space \( T_A M_r \). Recent research into non-convex optimization for matrix completion and phase retrieval, however, shows that the gradients are far from being arbitrary and enjoy good entrywise bounds\(^{12,23}\) when the initial condition is incoherent with respect to the measurement operator. So, by adapting the RIP (16) to gradients with entrywise bounds, it might be possible to reduce the sample complexity in Theorem 7.

Acknowledgments
We are grateful to the referees, whose thoughtful comments helped us present our ideas in a clearer and more structured way. This work was supported by Russian Science Foundation (project 21-71-10072).

Conflict of Interest Statement
The authors declare no potential conflict of interests.

Data Availability Statement
Data sharing is not applicable to this article as no new data were created or analyzed in this study.
REFERENCES

1. Candès EJ, Tao T. Decoding by linear programming. IEEE Trans Inf Theory. 2005 Dec;51(12):4203–15.
2. Oseledets IV, Tyrtyshnikov EE. Breaking the curse of dimensionality, or how to use SVD in many dimensions. SIAM J Sci Comput. 2009;31(5):3744–59.
3. Oseledets IV. Tensor-train decomposition. SIAM J Sci Comput. 2011;33(5):2295–317.
4. Absil PA, Mahony R, Sepulchre R. Optimization algorithms on matrix manifolds. Princeton, NJ: Princeton University Press; 2008.
5. Uschmajew A, Vandereycken B. Geometric methods on low-rank matrix and tensor manifolds. In: Grohs P, Holler M, Weinmann A, editors. Handbook of variational methods for nonlinear geometric data. Cham: Springer; 2020. p. 261–313.
6. Recht B, Fazel M, Parrilo PA. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. 2010;52(3):471–501.
7. Fazel M. Matrix rank minimization with applications, Ph.D. thesis. Stanford, CA: Stanford University; 2002.
8. Candès EJ, Recht B. Exact matrix completion via convex optimization. Found Comput Math. 2009;9(6):717.
9. Candès EJ, Tao T. The power of convex relaxation: near-optimal matrix completion. IEEE Trans Inf Theory. 2010;56(5):2053–80.
10. Recht B. A simpler approach to matrix completion. J Mach Learn Res. 2011;12:3413–30.
11. Lee J. Introduction to smooth manifolds. Graduate texts in mathematics. 2nd ed. New York: Springer-Verlag; 2012.
12. Ding L, Chen Y. Leave-one-out approach for matrix completion: primal and dual analysis. IEEE Trans Inf Theory. 2020;66(11):7274–301.
13. Vandereycken B. Low-rank matrix completion by Riemannian optimization. SIAM J Optim. 2013 Jan;23(2):1214–36.
14. Wei K, Cai JF, Chan TF, Leung S. Guarantees of riemannian optimization for low rank matrix completion. arXiv:160306610 [math]. 2016.
15. Holtz S, Rohwedder T, Schneider R. On manifolds of tensors of fixed TT-rank. Numer Math. 2012;120(4):701–31.
16. Lubich C, Rohwedder T, Schneider R, Vandereycken B. Dynamical approximation by hierarchical Tucker and Tensor-train tensors. SIAM J Matrix Anal Appl. 2013;34(2):470–94.
17. Cai JF, Li J, and Xia D. Provable tensor-train format tensor completion by Riemannian optimization. arXiv:210812163 [cs, math, stat]. 2021.
18. Kazeer VA, Khoromskij BN. Low-rank explicit QTT representation of the Laplace operator and its inverse. SIAM J Matrix Anal Appl. 2012;33(3):742–58.
19. Dolgov SV, Khoromskij BN, Oseledets IV. Fast solution of parabolic problems in the tensor train/quantized tensor train format with initial application to the Fokker–Planck equation. SIAM J Sci Comput. 2012;34(6):A3016–38.
20. Steinlechner M. Riemannian optimization for high-dimensional tensor completion. SIAM J Sci Comput. 2016 Jan;38(5):S461–84.
21. Rauhut H, Schneider R, Stojanac Ž. Low rank tensor recovery via iterative hard thresholding. Linear Algebra Appl. 2017;523:220–62.
22. Rauhut H, Schneider R, Stojanac Ž. Tensor completion in hierarchical tensor representations. In: Boche H, Calderbank R, Kutyniok G, Vybiral J, editors. Compressed sensing and its applications: MATHION workshop 2013. Applied and numerical harmonic analysis. Cham: Springer International Publishing; 2015. p. 419–50.
23. Chen Y, Chi Y, Fan J, Ma C. Gradient descent with random initialization: fast global convergence for nonconvex phase retrieval. Math Program. 2019 Jul;176(1):5–37.
24. Ma C, Wang K, Chi Y, Chen Y. Implicit regularization in nonconvex statistical estimation: gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. Found Comput Math. 2020;20(3):451–632.
25. Schneider R, Uschmajew A. Convergence results for projected line-search methods on varieties of low-rank matrices via Łojasiewicz inequality. SIAM J Optim. 2015;25(1):622–46.
26. Xu M, Jin R, Zhou ZH. Speedup matrix completion with side information: application to multi-label learning. In: Burges CIC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, editors. Advances in neural information processing systems 26. Curran Associates: Inc; 2013. p. 2301–9.
27. Natarajan N, Dhillon IS. Inductive matrix completion for predicting gene–disease associations. Bioinformatics. 2014 Jun;30(12):i60–8.
28. Chen X, Wang L, Qu J, Guan NN, Li JQ. Predicting miRNA–disease association based on inductive matrix completion. Bioinformatics. 2018;34(24):4256–65.
29. Lubich C, Oseledets IV, Vandereycken B. Time integration of tensor trains. SIAM J Numer Anal. 2015;53(2):917–41.
30. Eaton ML. Multivariate statistics: a vector space Approach. first printing ed. New York: John Wiley & Sons Inc; 1983.
31. Eaton ML. Group invariance applications in statistics. Reg Conf Ser Probab Stat. 1989;1:i–133.
32. Diaconis PW, Eaton ML, Lauritzen SL. Finite De Finetti theorems in linear models and multivariate analysis. Scand J Stat. 1992;19(4):289–315.
33. Vershynin R. Introduction to the non-asymptotic analysis of random matrices. arXiv:10113027 [cs, math]. 2011.
34. Mattei PA. Multiplying a Gaussian matrix by a Gaussian vector. Stat Probab Lett. 2017;128:67–70.
35. Maurer A. A bound on the deviation probability for sums of non-negative random variables. J Inequalities Pure Appl Math. 2003;4(1):15.
36. Gandy S, Recht B, Yamada I. Tensor completion and low-n-rank tensor recovery via convex optimization. Inverse Problems. 2011 Jan;27(2):025010.
37. Signoretto M, Lathauwer LD, Suykens JAK. Nuclear norms for tensors and their use for convex multilinear estimation. Linear Algebra Appl. 2010;43.
APPENDIX A. OTHER APPROACHES TO MATRIX AND TENSOR COMPLETION

Given the success of nuclear norm minimization for matrices—in terms of both computational feasibility and sample complexity—the transition to the multi-dimensional case did not suffer from the lack of ideas. The nuclear norm heuristic was extended as a convex surrogate of Tucker (also known as multilinear) ranks\(^{36-38}\) and TT ranks\(^{39}\) by setting the cost function to the sum of the nuclear norms (SNN) of the tensor flattenings or unfoldings.

The Tucker/multilinear ranks of a \(d\)-dimensional tensor \(A\) are defined as a tuple of ranks of all the mode-\(j\) flattenings

\[
\text{rank}_{\text{Tucker}}(A) = (\text{rank}(A_{(1)}), \ldots, \text{rank}(A_{(d)})).
\]

Assume for simplicity that all the sizes are equal to \(n\) and all the Tucker/multilinear ranks are equal to \(r\). The sample complexity of SNN for Tucker recovery from random Gaussian measurements was studied in References 40 and 41. Tucker completion via SNN was treated in Reference 42 where the authors assumed the incoherence (3) of one of the mode-\(j\) flattenings \(A_{(j)}\); the RIP (8) was proved to hold with high probability if the sample \(\Omega \subseteq [n]^d\) contains

\[
|\Omega| \gtrsim \mu_0 d^r n^{d-1} \log(n)
\]

randomly chosen elements. With the help of an additional mutual incoherence property of the tensor, it was proved that SNN can recover \(A\) with high probability if

\[
|\Omega| \gtrsim \mu_0 d^4 n^{d-1} \log^2(n).
\]
A different view on the tensor nuclear norm and tensor completion consists in extending the spectral norm and taking its dual. This approach, however, is mostly of theoretical value: the norm in question is computationally intractable but leads to improved estimates of the sample size compared to SNN. In Reference 44, a special incoherent nuclear norm was constructed for the Tucker completion problem. An analog of the RIP (16) was proved to hold with high probability under the incoherence assumption (4) for all mode-\(j\) fiberspans provided that

\[|\Omega| \gtrsim \mu_0^{-1} d^{d-1} n \log(n)\]

samples are drawn uniformly at random. The minimization of the incoherent nuclear norm was proved to recover \(A\) when

\[|\Omega| \gtrsim C_d (\mu_0^{-d} d^{-1} n^2 \log^2(n), \quad C_d = C_d(d).\]

Another approach to tensor completion is to minimize the residual under the rank constraint (as in Equation (9)), without going into the geometric nuances on Riemannian optimization. In the matrix case, the singular value projection (SVP) algorithm (see also a closely related iterative hard thresholding algorithm) was developed as a projected gradient descent method

\[X_{t+1} = \text{SVD}_r \left( X_t - \frac{\rho^{-1}}{1 + \delta_2} [R_{\Omega} X_t - R_{\Omega} A] \right), \quad X_0 = 0. \tag{A1}\]

Here, \(0 < \delta_2 < 1\) is a RIP constant, where RIP is understood as

\[(1 - \delta_2)\|X\|^2_F \leq \rho^{-1}\|R_{\Omega} X\|^2_F \leq (1 + \delta_2)\|X\|^2_F\]

for all matrices of rank at most \(2r\) and with bounded coherence (4). This RIP holds with high probability when

\[|\Omega| \gtrsim \mu_0^2 r^2 n \log(n),\]

which exceeds what is required by Theorem 1 for the RIP (8). The convergence of the SVP is, however, only conjectured in Reference 45: the problem is that \(X_{t+1} - X_t\) and \(X_t - A\) need to have uniformly bounded coherences (3). Linear convergence in the entrywise norm was later proved, in the symmetric case, in Reference 12 when

\[|\Omega| \gtrsim \kappa^6 \mu_0^4 r^6 n \log(n),\]

where \(\kappa\) is the condition number (in the spectral norm) of \(A\).

The SVP framework has been extended to tensor recovery in Tucker and TT formats under the assumption that the measurement operator satisfies the standard RIP (14). In the multi-dimensional setting, the exact SVD-based matrix projection is replaced with HOSVD and TT-SVD, which are the standard generalizations of SVD to the Tucker and TT formats. The main difference between the matrix and tensor cases is that the truncated HOSVD and TT-SVD are quasi-optimal projections as opposed to the optimal truncated SVD. The theory of the SVP convergence for matrices has been extended to quasi-optimal projections. For HOSVD and TT-SVD, the quasi-optimality constant is rather large, \(\sqrt{d}\), a fact that poses problems for theoretical analysis (but less so for practical purposes since \(\sqrt{d}\) corresponds to the worst case). That is why a local optimality assumption accompanies the standard RIP of order \(3r\)—note that matrix SVP requires the standard RIP of order \(2r\) (see Reference 45)—in the proof of global SVP convergence for tensor recovery. We are not aware of any theoretical results about tensor completion using SVP.

An iteration of Riemannian gradient descent for Tucker recovery can be written with the help of notation we introduced above:

\[X_{t+1} = \text{HOSVD}_r \left( X_t - \alpha_t P_{\Omega^c} [RX_t - RA] \right).\]

Its local convergence was proved in Reference 22 for \(R\) satisfying RIP of order \(3r\), which was improved to \(2r\) in Reference 49. The authors of the latter also show that one step of Tucker-SVP with zero initial condition gives an estimate that is sufficiently close to \(A\) for local convergence to start working. Riemannian Tucker and TT completion were studied in References 50 and 20 but the number of samples was estimated only numerically.
The RGD for Tucker recovery was proposed in Reference 22 and proved to converge locally for the measurement operators satisfying the standard RIP (14) of order 3\( r \), which was improved to \( 2r \) in Reference 49. The authors of the latter also show that one step of Tucker-SVP with zero initial condition gives an estimate that is sufficiently close to the solution for the local convergence to start working. The algorithms for Riemannian Tucker and TT completion were studied in References 50 and 20 but the number of samples was estimated only numerically.

By comparing the current state of affairs in matrix and tensor completion, we can now see what principal difficulties are brought in by the multiple dimensions. For matrices, the nuclear norm formulation appeared to be a perfect object from the theoretical point of view. Indeed, it exhibits both polynomial computational complexity and nearly optimal sample complexity. Meanwhile, the computable SNN model leads to poor recovery guarantees for Tucker completion, and the tightest known sample complexity is achieved by the computationally intractable incoherent nuclear norm. Likewise, if we look at the development of SVP and Riemannian optimization for matrix and tensor completion in parallel, we will note that the RIP of the sampling operator and the recovery guarantees for tensor completion are only beginning to be explored in the literature.

### APPENDIX B. TENSOR TRAIN COMPLETION WITH AUXILIARY SUBSPACE INFORMATION

Typically, algorithms for matrix and tensor completion with subspace information are developed as generalizations of the methods used in usual matrix/tensor completion. The nuclear norm minimization approach was used in References 26, 51, 52 to recover a low-rank matrix \( A \) with additional subspace information (21):

\[
\| W \|_* \to \min \quad \text{s.t.} \quad R_{\Omega}(Q_1 W Q_2^T) = R_{\Omega} A. \tag{B1}
\]

It is claimed in Reference 26 that \( Q_1^T A Q_2 \in \mathbb{R}^{m_1 \times m_2} \) is the unique solution to (B1) when

\[
| \Omega | \gtrsim \mu^2 r m \log(m) \log(n), \quad n = \max(n_1, n_2), \quad m = \max(m_1, m_2),
\]

indices are chosen uniformly at random. Note that the estimate depends only logarithmically on the matrix size \( n \) (cf. Theorem 2). The coefficient \( \mu^2 \) depends on the coherences (3) of the column and row spaces of \( A \) and of the additionally known subspaces spanned by \( Q_1 \) and \( Q_2 \).

In Reference 53, a Riemannian algorithm for TT completion with subspace information was proposed and its sample complexity was studied numerically. Here, we want to address the question from the theoretical point of view. Tensor completion with subspace information for other tensor formats was treated in References 54 and 55.

#### B.1 Riemannian gradient descent

First of all, we need to establish the geometry of the problem. Denote by \( M_r^{(m)} \) and \( M_r^{(n)} \) the submanifolds of small \( m_1 \times \ldots \times m_d \) and large \( n_1 \times \ldots \times n_d \) tensors of TT rank \( r \), respectively. The following Lemma 9 shows that the size of the tensor can be increased by applying mode-\( k \) products without altering the rank.

**Lemma 9.** Let \( S_k \in \mathbb{R}^{n_k \times m_k} \) be a matrix of rank \( m_k \). Then for any tensor \( W \in \mathbb{R}^{m_1 \times \ldots \times m_d} \) the mode-\( k \) product with \( S_k \) does not change its TT rank:

\[
\text{rank}_{\text{TT}}(W) = \text{rank}_{\text{TT}}(W \times_k S_k).
\]

**Proof.** Let \( W = [C_1, \ldots, C_d] \) be a minimal TT representation of \( W \in \mathbb{R}^{m_1 \times \ldots \times m_d} \). By definition of the mode-\( k \) product,

\[
W \times_k S_k = [C_1, \ldots, C_{k-1}, D_k, C_{k+1}, C_d], \quad D_k = C_k \times_2 S_k \in \mathbb{R}^{r_k \times n_k \times x_k}.
\]

It suffices to show that this TT representation is also minimal, that is, that the left and right unfoldings of \( D_k \) are full-rank. It is easy to see that the left and right unfoldings satisfy

\[
D_k^L = (S_k \otimes I_{n_k}) C_k^L, \quad D_k^R = C_k^R (S_k^T \otimes I_{n_k})
\]

and are full-rank as products of full-rank matrices.
Thanks to Lemma 9, the linear operator $Q : \mathbb{R}^{m_1 \times \cdots \times m_d} \to \mathbb{R}^{n_1 \times \cdots \times n_d}$ defined by

$$QW = W \times_1 Q_1 \times_2 \cdots \times_d Q_d,$$

where $Q_k$ are matrices with orthonormal columns, can be restricted to the submanifold $\mathcal{M}_r^{(m)}$ as $Q : \mathcal{M}_r^{(m)} \to \mathcal{M}_r^{(n)}$. Its image $Q(\mathcal{M}_r^{(m)})$ is an embedded submanifold\(^{11}\) of $\mathcal{M}_r^{(n)}$ and the adjoint operator

$$Q^*X = X \times_1 Q_1^T \times_2 \cdots \times_d Q_d^T$$

acts as the left inverse $Q^*Q = \text{Id}$. The set $Q(\mathcal{M}_r^{(m)})$ contains precisely those tensors that satisfy the rank and subspace requirements (this explains Equation (21)).

**Lemma 10.** Let $A \in \mathcal{M}_r^{(n)}$ be a tensor of TT rank $r$. All of its mode-$k$ fiber spans belong to the subspaces spanned by the columns of $Q_k$ if and only if $A \in Q(\mathcal{M}_r^{(m)})$.

**Proof.** Let $A = QB$ with $B \in \mathcal{M}_r^{(m)}$. Then by the definition of the mode-$k$ product

$$A_{(k)} = Q_kB_{(k)}, \quad k \in [d],$$

and the inclusion of subspaces follows. Conversely, let the mode-$k$ fiber spans of $A$ belong to the column spans of $Q_k$. Then $Q_kQ_k^TA_{(k)} = A_{(k)}$ and $QQ^*A = A$. Then $B = Q^*A$ lies in $\mathcal{M}_r^{(m)}$ since if it had different TT ranks, so would $QB$ by Lemma 9.

This means that Riemannian optimization can be applied to TT completion with subspace information, and we only need to narrow down the manifold:

$$\|\sqrt{R}\Omega X - \sqrt{R}\Omega A\|_F^2 \to \min \quad \text{s.t.} \quad X \in Q(\mathcal{M}_r^{(m)}).$$

The projection onto the new tangent space can be easily computed as

$$P_{T_X Q(\mathcal{M}_r^{(m)})} = Q^*P_{T_X \mathcal{M}_r^{(m)}}$$

and the formulation of the RGD follows immediately. However, for the theoretical analysis we prefer to use an equivalent optimization problem that works on $\mathcal{M}_r^{(m)}$ rather than directly on $Q(\mathcal{M}_r^{(m)})$. Let $A = QB$ with $B \in \mathcal{M}_r^{(m)}$, then we consider

$$\|\sqrt{R}\Omega QW - \sqrt{R}\Omega QB\|_F^2 \to \min \quad \text{s.t.} \quad W \in \mathcal{M}_r^{(m)}.$$ 

The modified sampling operator is $Q^*R\Omega Q$ and a step of the RGD can be written as

$$W_{t+1} = \text{TT-SVD}_r(W_t - \alpha_t Y_t) \in \mathcal{M}_r^{(m)}, \quad Y_t = P_W[Q^*R\Omega QW_t - Q^*R\Omega QB] \in T_W, \mathcal{M}_r^{(m)}$$

with the step size

$$\alpha_t = \frac{\|Y_t\|_F^2}{\langle Q^*R\Omega QY_t, Y_t \rangle_F}.$$

The RIP (16), Lemma 6, and Theorem 5 for TT completion undergo simple modifications according to

$$\left\|P_{T_X \mathcal{M}_r^{(m)}} - \rho^{-1}P_{T_X \mathcal{M}_r^{(m)}}Q^*R\Omega QP_{T_X \mathcal{M}_r^{(m)}}\right\| \leq \varepsilon, \quad \|Q^*R\Omega Q\| \leq C. \quad (B2)$$

Other than this, the formulations and proofs transfer verbatim to the current scenario. The convergence rate and the estimate of the local convergence basin that are then given in terms of $W_t$ and $B$ hold identically for $QW_t$ and $A$ since $\|W - B\|_F = \|QW - A\|_F$ and $\sigma_{\min}(B) = \sigma_{\min}(A)$. 
B.2 Recovery guarantees
Let us show that these assumptions hold with high probability. First of all, note that \( \|Q^* R_\Omega \| \leq \| R_\Omega \| \), hence Lemma 2 applies. To derive probabilistic sufficient conditions for the modified RIP (B2), we need to prove analogs of Lemma 4, Lemma 8, and Theorem 7.

**Lemma 11.** Let \( A = Q B \in M_{t_i}^{(n)} \) be a tensor of TT rank \( r \) with bounded core coherence \( \mu_C(A) \leq \mu_1 \). Then for every \( k \in [d - 1] \) and for all multi-indices \( (i_1, \ldots, i_d) \in [n_1] \times \ldots \times [n_d] \) we have

\[
\frac{n_1 \ldots n_k}{r_k} P_{\leq k}(Q^T_i e_{i_k} \otimes \ldots \otimes Q^T_{i_d} e_{i_d}) \leq \mu_1^k, \quad \frac{n_{k+1} \ldots n_d}{r_k} P_{\geq k+1}(Q^T_{i_k} e_{i_{k+1}} \otimes \ldots \otimes Q^T_{i_d} e_{i_d}) \leq \mu_1^{d-k},
\]

where \( P_{\leq k} \) and \( P_{\geq k+1} \) are the orthogonal projections onto the column spans of the interface matrices \( B_{\leq k} \) and \( B_{\geq k+1} \).

**Proof.** Let \( B = [U_1, \ldots, U_{d-1}, G_d] \) be a minimal left-orthogonal TT representation. Then \( S_k \in \mathbb{R}^{n_1 \times \ldots \times n_{k-1} \times n_k \times \ldots \times n_d} \) defined as

\[
S^L_k = (Q_k \otimes I_{n_{k+1}}) U^L_k
\]
give a minimal left-orthogonal TT representation of \( A \). Denote by \( \xi_k \) the \( r_k \)-dimensional row-vector whose norm we need to estimate

\[
\xi_k = U^T_{\leq k}(Q^T e_{i_k} \otimes \ldots \otimes Q^T_{i_d} e_{i_d}).
\]

Given the recursive formula (4.1) we establish that

\[
\xi_k = (U^L_k)^T [Q_k e_{i_k} \otimes \ldots \otimes Q_k e_{i_d}], \quad \xi_0 = 1.
\]

The incoherence assumption for \( A \) tells us that

\[
\max_{i \in [n_i]} \| S^{(i)}_k \|_2^2 \leq \frac{r_k}{r_{k-1} n_k} \mu_1
\]

and so since

\[
S^{(i_k)}_k = (e_{i_k}^T \otimes I_{n_{k+1}}) S^L_k = (e_{i_k}^T Q_k \otimes I_{n_{k+1}}) U^L_k,
\]

we obtain

\[
\| \xi_k \|_2^2 = \| (U^L_k)^T [Q_k e_{i_k} \otimes \xi_{k-1}] \|_2^2 = \| (S^{(i_k)}_k \otimes \xi_{k-1}) \|_2^2 \leq \frac{r_k}{r_{k-1} n_k} \mu_1 \| \xi_{k-1} \|_2^2 \leq \frac{r_k}{n_1 \ldots n_{k-1}} \mu_1^k.
\]

The argument is the same for the right unfoldings. ■

**Lemma 12.** Let \( A = Q B \in M_{t_i}^{(n)} \) be a tensor of TT rank \( r \) with bounded core coherence \( \mu_C(A) \leq \mu_1 \). Assume that the coherences of the auxiliary subspaces are bounded as well \( \mu(Q_k) \leq \mu_2 \). Then for every canonical basis tensor \( E_\omega \in \mathbb{R}^{n_1 \times \ldots \times n_d} \), \( \omega \in [n_1] \times \ldots \times [n_d] \), its projection onto the tangent space \( T_B M_{t_i}^{(m)} \) can be bounded from above as

\[
\| P_{T_B M_{t_i}^{(m)}} Q^* E_\omega \|_F \leq \frac{\mu_1^{d-1} \mu_2}{n_1 \ldots n_d} \sum_{k=1}^d r_{k-1} m_k r_k.
\]

**Proof.** We have

\[
\| P_{T_B M_{t_i}^{(m)}} Q^* E_\omega \|_F^2 \leq \| P_{\geq 2} Q^* E_\omega \|_F^2 + \sum_{k=1}^{d-1} \| P_{\leq k} P_{\geq k+1} Q^* E_\omega \|_F^2 + \| P_{\leq d-1} Q^* E_\omega \|_F^2.
\]
For the first and last terms we obtain
\[
\| P_{\leq 1}^{(1)} Q^* E_{\omega} \|_F^2 = \| Q^T e_i \circ P_{\leq 1} (Q^T e_i) \|_F^2 \leq \frac{m_1 \mu_2^{d-1} r_1}{n_1 n_2 \ldots n_d} \mu_2^{d-1}
\]
and
\[
\| P_{\leq d-1}^{(1)} Q^* E_{\omega} \|_F^2 = \| P_{\leq d-1} (Q^T e_i) \circ Q^T e_i \|_F^2 \leq \frac{r_{d-1}}{n_1 \ldots n_d} \mu_1^{d-1} \mu_2^{d-1}.
\]

The summands \( \| P_{\leq k-1}^{(1)} P_{\leq k+1}^{(1)} Q^* E_{\omega} \|_F^2 \) in the middle are equal to
\[
\| P_{\leq k-1}^{(1)} (Q^T e_i) \circ Q^T e_i \circ P_{\leq k+1}^{(1)} (Q^T e_i) \|_F^2 \leq \frac{r_{k-1}}{n_1 \ldots n_{k-1}} \mu_1^{k-1} \mu_2^{k-1} r_k \mu_1^{d-k}.
\]

It remains to combine the estimates.

**Theorem 9.** Let \( A = QB \in \mathcal{M}_r^{(n)} \) be a tensor of TT rank \( r \) with bounded core coherence \( \mu_c(A) \leq \mu_1 \). Assume that the coherences of the auxiliary subspaces are bounded as well \( \mu(Q_k) \leq \mu_2 \) and let \( \Omega \subset [n_1] \times \ldots \times [n_d] \) be a collection of indices sampled uniformly at random with replacement. Then the modified RIP (B2)

\[
\left\| P_{\tau_{\Omega} \mathcal{M}_r^{(n)}} - \rho^{-1} P_{\tau_{\Omega} \mathcal{M}_r^{(n)}} Q^* R_{\Omega} Q P_{\tau_{\Omega} \mathcal{M}_r^{(n)}} \right\| < \varepsilon, \quad \rho = \frac{|\Omega|}{n_1 \ldots n_d},
\]

holds with probability at least \( 1 - 2m^{d(1-\beta)} \), \( m = \max(m_1, \ldots, m_d) \), for all \( \beta > 1 \) provided that

\[
|\Omega| \geq \frac{8 \beta}{3} \mu_1^{d-1} \mu_2^d \left( \sum_{k=1}^d r_{k-1} m_k r_k \right) d \log (m).
\]

**Proof.** For an arbitrary tensor \( Z \in \mathbb{R}^{m_1 \times \ldots \times m_d} \) we can represent \( QZ \) as

\[
QZ = \sum_{\omega \in [n_1] \times \ldots \times [n_d]} \langle QZ, E_{\omega} \rangle_{F} E_{\omega}.
\]

Denote by \( P_B \) the projection \( P_{\tau_{\Omega} \mathcal{M}_r^{(n)}} \). It follows that

\[
P_B Z = \sum_{\omega \in [n_1] \times \ldots \times [n_d]} \langle Z, P_B Q^* E_{\omega} \rangle_{F} P_B Q^* E_{\omega}
\]

and

\[
P_B Q^* R_{\Omega} Q P_B Z = \sum_{\omega \in \Omega} \langle Z, P_B Q^* E_{\omega} \rangle_{F} P_B Q^* E_{\omega}.
\]

As we introduce operators \( S_{\omega} : \mathbb{R}^{m_1 \times \ldots \times m_d} \to \mathbb{R}^{m_1 \times \ldots \times m_d} \) defined by

\[
S_{\omega} Z = \langle Z, P_B Q^* E_{\omega} \rangle_{F} P_B Q^* E_{\omega}
\]

the proof follows the proof of Theorem 6.

We believe that Theorem 9 is the first theoretical estimate for the sample complexity of TT completion with subspace information. Previous results on matrix completion with subspace information contained a \( \log(n) \) factor in the sample complexity;²⁶ our bound, which guarantees the local convergence of the RGD, depends only on the dimensions of the auxiliary subspaces and not on the dimensions of the tensor:

\[
|\Omega| \geq \mu_1^{d-1} \mu_2 d^2 r^2 m \log (m).
\]

This behavior is further well-aligned with the numerical experiments carried out in Reference 53.