Appendices

A Review of error analysis for matrix cross approximation

We first review the following result which establishes an element-wise approximation guarantee for a particular cross approximation.

Theorem 5 *[53] For any $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, if $\boldsymbol{U} = \boldsymbol{A}(I,J)$ is an $r\times r$ submatrix of maximal volume (maximum absolute value of the determinant), then*

$$\left\|\boldsymbol{A} - \boldsymbol{C}\boldsymbol{U}^\dagger\boldsymbol{R}\right\|_{\max} \le (r+1)\sigma_{r+1}(\boldsymbol{A}),$$

where $\|\boldsymbol{A}\|_{\max} = \max_{ij}|a_{ij}|$.

On one hand, similar to Theorem 1, Theorem 5 ensures no approximation error when $\boldsymbol{A}$ is low-rank. On the other hand, when $\boldsymbol{A}$ is approximately low-rank, Theorem 5 ensures that cross approximation via the maximal volume principle is stable as each entry is perturbed at most proportionally to $\sigma_{r+1}(\boldsymbol{A})$ which is expected to be small. However, if we are interested in an approximation guarantee for the entire matrix (say in the Frobenius norm) instead of each entry, then directly applying the above result leads to the loose bound

$$\left\|\boldsymbol{A} - \boldsymbol{C}\boldsymbol{U}^\dagger\boldsymbol{R}\right\|_F \le (r+1)\sqrt{mn}\cdot\sigma_{r+1}(\boldsymbol{A}),$$

which could be much worse than the best rank-r approximation. The recent work [58] provides a much tighter approximation guarantee in the Frobenius norm for cross approximation.

Theorem 6 *[58] For any $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, there exist indices $I \in [m], J \in [n]$ such that the cross approximation (1) satisfies*

$$\left\|\boldsymbol{A} - \boldsymbol{C}\boldsymbol{U}^\dagger\boldsymbol{R}\right\|_F \le (r+1)\left\|\boldsymbol{A} - \boldsymbol{A}_r\right\|_F,$$

where $\boldsymbol{A}_r$ denotes the best rank-r approximation of $\boldsymbol{A}$ measured by the Frobenius norm.

Theorem 6 shows that cross approximation could be stable and have approximation error comparable to the best rank-r approximation up to a factor of $(r+1)$. Theorem 6 is proved by viewing $I \in [m]$ and $J \in [n]$ with $|I| = |J| = r$ as random variables and studying the expectation of $\left\|\boldsymbol{A} - \boldsymbol{C}\boldsymbol{U}^\dagger\boldsymbol{R}\right\|_F$ over all (I,J). Note that Theorem 6 is not valid for cross approximations constructed using submatrices of maximum volume. In other words, one [58] could construct a counter-example $\boldsymbol{A}$ for which the cross approximations $\boldsymbol{C}\boldsymbol{U}^{-1}\boldsymbol{R}$ constructed using submatrices of maximum volume have approximation error larger than $\sqrt{\max(m,n)}\left\|\boldsymbol{A} - \boldsymbol{A}_r\right\|_F$. On the other hand, ignoring the worst-case examples, the work [59] establishes a similar guarantee for cross approximations constructed using submatrices of maximum projective volume for random matrices. Finally, we note that a derandomized algorithm is proposed in [60] that finds a cross approximation achieving the bound in Theorem 6. The work [62, Corollary 4.3] studies the approximation guarantee of cross approximation for the best rank-r $\boldsymbol{A}_r$ with any selected rows and columns. Below we extend this result for $\boldsymbol{A}$.

Theorem 7 *Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ be an approximately low-rank matrix that can be decomposed as $\boldsymbol{A} = \boldsymbol{A}_r + \boldsymbol{F}$. Then the cross approximation (1) with $\operatorname{rank}(\boldsymbol{A}_r(I,J)) = r$ satisfies*

$$\begin{aligned}\|\boldsymbol{A} - \boldsymbol{C}\boldsymbol{U}^\dagger\boldsymbol{R}\|_F \le{}& (\|\boldsymbol{W}^\dagger(I,:)\|_2 + \|\boldsymbol{V}^\dagger(J,:)\|_2 + 3\|\boldsymbol{W}^\dagger(I,:)\|_2\|\boldsymbol{V}^\dagger(J,:)\|_2 + 1)\|\boldsymbol{F}\|_F \\ &+ (\|\boldsymbol{W}^\dagger(I,:)\|_2 + \|\boldsymbol{V}^\dagger(J,:)\|_2 + \|\boldsymbol{W}^\dagger(I,:)\|_2\|\boldsymbol{V}^\dagger(J,:)\|_2 + 1)\|\boldsymbol{U}^\dagger\|_2\|\boldsymbol{F}\|_F^2, \quad (13)\end{aligned}$$

where $\boldsymbol{A}_r = \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ is the compact SVD of $\boldsymbol{A}_r$.

Note that (13) holds for any cross approximation as long as $\boldsymbol{A}_r(I,J)$ has rank r, but the quality of the cross approximation depends on the matrix $\boldsymbol{A}$ as well as the selected rows and columns as reflected by $\|\boldsymbol{W}^\dagger(I,:)\|_2$, $\|\boldsymbol{V}^\dagger(J,:)\|_2$, and $\|\boldsymbol{U}^\dagger\|_2$. For example, on one hand, $\|\boldsymbol{W}^\dagger(I,:)\|_2$ and $\|\boldsymbol{V}^\dagger(J,:)\|_2$ can achieve their smallest possible value 1 when the singular vectors $\boldsymbol{W}$ and $\boldsymbol{V}$ are the canonical basis. On the other hand, one may construct examples with large $\|\boldsymbol{W}^\dagger(I,:)\|_2$ and

$\|\boldsymbol{V}^\dagger(J,:)\|_2$. Nevertheless, these quantities can be upper bounded by selecting appropriate rows and columns [62, Proposition 5.1]. In particular, for any orthonormal matrix $\boldsymbol{W}$, if we select I such that $\boldsymbol{W}(I,:)$ has maximal volume among all $|I| \times r$ submatrices of $\boldsymbol{W}$, then we can always upper bound $\|\boldsymbol{W}^\dagger(I,:)\|_2$ by $\sqrt{1+\frac{r(m-|I|)}{|I|-r+1}}$. A similar result also holds for $\|\boldsymbol{V}^\dagger(J,:)\|_2$. Likewise, according to [62, Proposition 5.1], $\|\boldsymbol{U}^\dagger\|_2$ can be upper bounded by $\sqrt{1+\frac{r(m-|I|)}{|I|-r+1}}\sqrt{1+\frac{r(n-|J|)}{|J|-r+1}}\|\boldsymbol{A}^\dagger\|_2$.

Proof (of Theorem 7)

First note that

$$\|\boldsymbol{A}-\boldsymbol{C}\boldsymbol{U}^\dagger\boldsymbol{R}\|_F \le \|\boldsymbol{A}-\boldsymbol{A}_r\|_F + \|\boldsymbol{A}_r-\boldsymbol{C}\boldsymbol{U}^\dagger\boldsymbol{R}\|_F = \|\boldsymbol{F}\|_F + \|\boldsymbol{A}_r-\boldsymbol{C}\boldsymbol{U}^\dagger\boldsymbol{R}\|_F.$$

The proof is then completed by invoking [62, Corollary 4.3]:

$$\begin{aligned}\|\boldsymbol{A}_r-\boldsymbol{C}\boldsymbol{U}^\dagger\boldsymbol{R}\|_F \le &(\|\boldsymbol{W}^\dagger(I,:)\|_2 + \|\boldsymbol{V}^\dagger(J,:)\|_2 + 3\|\boldsymbol{W}^\dagger(I,:)\|_2\|\boldsymbol{V}^\dagger(J,:)\|_2)\|\boldsymbol{F}\|_F \\ &+ (\|\boldsymbol{W}^\dagger(I,:)\|_2 + \|\boldsymbol{V}^\dagger(J,:)\|_2 + \|\boldsymbol{W}^\dagger(I,:)\|_2\|\boldsymbol{V}^\dagger(J,:)\|_2 + 1)\|\boldsymbol{U}^\dagger\|_2\|\boldsymbol{F}\|_F^2.\end{aligned}$$

■

B Proof of Theorem 2

B.1 Overview of the Analysis

To bound the difference between $\mathcal{T}$ and $\widehat{\mathcal{T}}$, we use a similar approach as in [65] that exploits the approximate low-rankness in $\mathcal{T}$ and the same structures within $\mathcal{T}$ and $\widehat{\mathcal{T}}$. The point of departure is the fact that according to the expression for $\widehat{\mathcal{T}}$ in (4), if we let $\widehat{\boldsymbol{T}}^{\langle k\rangle}$ be the k-th unfolding of $\widehat{\mathcal{T}}$, then there exist $\widehat{\boldsymbol{C}}$ and $\widehat{\boldsymbol{R}}$ such that $\widehat{\boldsymbol{T}}^{\langle k\rangle} = \widehat{\boldsymbol{C}}[\boldsymbol{T}^{\langle k\rangle}(I^{\le k}, I^{>k})]_{\tau_k}^\dagger \widehat{\boldsymbol{R}}$. Note that $\widehat{\boldsymbol{C}}$ and $\widehat{\boldsymbol{R}}$ depend on k, but we omit such dependence to simplify the notation. On the other hand, $\boldsymbol{T}^{\langle k\rangle}$ (the k-th unfolding matrix of $\mathcal{T}$) is approximately low-rank, and thus can be approximated by the cross approximation in the form of $\boldsymbol{C}[\boldsymbol{T}^{\langle k\rangle}(I^{\le k}, I^{>k})]_{\tau_k}^\dagger \boldsymbol{R}$. Therefore, the difference between $\widehat{\boldsymbol{T}}^{\langle k\rangle}$ and $\boldsymbol{T}^{\langle k\rangle}$ is controlled by the differences between $\boldsymbol{C}$ and $\widehat{\boldsymbol{C}}$ and $\boldsymbol{R}$ and $\widehat{\boldsymbol{R}}$. We can then adopt the same approach to bound the difference between $\boldsymbol{C}$ and $\widehat{\boldsymbol{C}}$ by noting that $\boldsymbol{C} = \boldsymbol{T}^{\langle k\rangle}(:, I^{>k})$ contains selected columns of $\boldsymbol{T}^{\langle k\rangle}$, which if reshaped to another matrix (corresponding to another unfolding matrix $\widehat{\boldsymbol{T}}^{\langle k'\rangle}$ with $k' > k$) is also low-rank as shown in Figure 2. The difference between $\boldsymbol{R}$ and $\widehat{\boldsymbol{R}}$ can also be analyzed by the same approach. We can repeat the above step several times until the ground level where the associated matrices $\boldsymbol{C}$ and $\boldsymbol{R}$ are equal to $\widehat{\boldsymbol{C}}$ and $\widehat{\boldsymbol{R}}$, respectively.

Each step of the recursive procedure may amplify the approximation error. To reduce the total number of steps, we use the balanced canonical dimension tree [79, 65]. As an example, in Figure 4 (a modification of [65, Figure 1]), we illustrate the above mentioned interpolation steps with balanced canonical dimension tree. Since $\left\|\mathcal{T}-\widehat{\mathcal{T}}\right\|_F = \left\|\boldsymbol{T}^{\langle 4\rangle} - \widehat{\boldsymbol{T}}^{\langle 4\rangle}\right\|_F$, in the top level (i.e., 3-rd level) of the figure, we split the multi-index $\{d_1,\dots,d_8\}$ in two parts $\{d_1,\dots,d_4\}$ and $\{d_5,\dots,d_8\}$. Recall that $I^{\le k}$ and $\{d_1\cdots d_k\}$ (both have size r_4') denote the selected row indices from the multi-index and column indices from the multi-index $\{d_{k+1}\cdots d_N\}$, respectively. Since $\widehat{\boldsymbol{T}}^{\langle 4\rangle} = \widehat{\boldsymbol{C}}_{(4)}[\boldsymbol{T}^{\langle 4\rangle}(I^{\le k}, I^{>k})]_{\tau_k}^\dagger \widehat{\boldsymbol{R}}_{(4)}$ and $\boldsymbol{T}^{\langle 4\rangle} \approx \boldsymbol{C}_{(4)}[\boldsymbol{T}^{\langle 4\rangle}(I^{\le k}, I^{>k})]_{\tau_k}^\dagger \boldsymbol{R}_{(4)}$, where $\widehat{\boldsymbol{C}}_{(4)} = \widehat{\boldsymbol{T}}^{\langle 4\rangle}(:, I^{>4})$ and $\widehat{\boldsymbol{R}}_{(4)} = \widehat{\boldsymbol{T}}^{\langle 4\rangle}(I^{\le 4}, :)$ (similarly for $\boldsymbol{C}_{(4)}$ and $\boldsymbol{R}_{(4)}$), the task of analyzing $\left\|\boldsymbol{T}^{\langle 4\rangle} - \widehat{\boldsymbol{T}}^{\langle 4\rangle}\right\|_F$ can be reduced to analyzing $\|\widehat{\boldsymbol{C}}_{(4)} - \boldsymbol{C}_{(4)}\|_F$ and $\|\widehat{\boldsymbol{R}}_{(4)} - \boldsymbol{R}_{(4)}\|_F$. Taking $\|\widehat{\boldsymbol{C}}_{(4)} - \boldsymbol{C}_{(4)}\|_F$ as an example, we can represent $\boldsymbol{C}_{(4)}$ through the 2-nd unfolding matrix $\widehat{\boldsymbol{T}}^{\langle 2\rangle}$, i.e., $\boldsymbol{C}_{(4)} = \boldsymbol{T}^{\langle 4\rangle}(:, I^{>4})$ contains the same entries as $\boldsymbol{T}^{\langle 2\rangle}_{:,I>4}$, where $\boldsymbol{T}^{\langle 2\rangle}_{:,I>4}$ is a submatrix of $\boldsymbol{T}^{\langle 2\rangle}$ with the multi-index $d_5\cdots d_8$ restricted to $I^{>4}$. Here $\boldsymbol{T}^{\langle 2\rangle}_{:,I>4}$ has size $d_1d_2 \times d_3d_4r_4'$ but is low-rank, and thus can be approximated by $\boldsymbol{C}_{(2)}[\boldsymbol{T}^{\langle 2\rangle}_{:,I>4}(I^{\le 2}, I^{>2})]_{\tau_k}^\dagger \boldsymbol{R}_{(2)}$, where $\boldsymbol{C}_{(2)} = \boldsymbol{T}^{\langle 2\rangle}_{:,I>4}(:, I^{>2})$ and $\boldsymbol{R}_{(2)} = \boldsymbol{T}^{\langle 2\rangle}_{:,I>4}(I^{\le 2}, :)$. Therefore, $\|\widehat{\boldsymbol{C}}_{(4)} - \boldsymbol{C}_{(4)}\|_F$ can also be bounded through $\|\widehat{\boldsymbol{C}}_{(2)} - \boldsymbol{C}_{(2)}\|_F$ and $\|\widehat{\boldsymbol{R}}_{(2)} - \boldsymbol{R}_{(2)}\|_F$, where $\widehat{\boldsymbol{C}}_{(2)} = \widehat{\boldsymbol{T}}^{\langle 2\rangle}_{:,I>4}(:, I^{>2})$ and

$\widehat{\boldsymbol{R}}_{(2)} = \widehat{\boldsymbol{T}}^{\langle 2\rangle}_{:,I^{>4}}(I^{\leq 2},:)$. We can further analyze $\|\widehat{\boldsymbol{C}}_{(2)} - \boldsymbol{C}_{(2)}\|_F$ by connecting these matrices to $\boldsymbol{T}^{\langle 1\rangle}$ and $\widehat{\boldsymbol{T}}^{\langle 1\rangle}$, which have the same entries in the sampled locations. A similar approach can be applied for $\|\widehat{\boldsymbol{R}}_{(2)} - \boldsymbol{R}_{(2)}\|_F$ by connecting to the 2-nd unfolding of the tensors. Thus, as shown in Figure 4, the analysis for an 8-th order tensor involves a three-step decomposition until the ground level with no errors in the sampled locations. For a general N-th order tensor, the analysis will involve $\lceil \log_2 N\rceil$ steps of such a decomposition. The main task is to study how the approximation error depends on the previous layer.

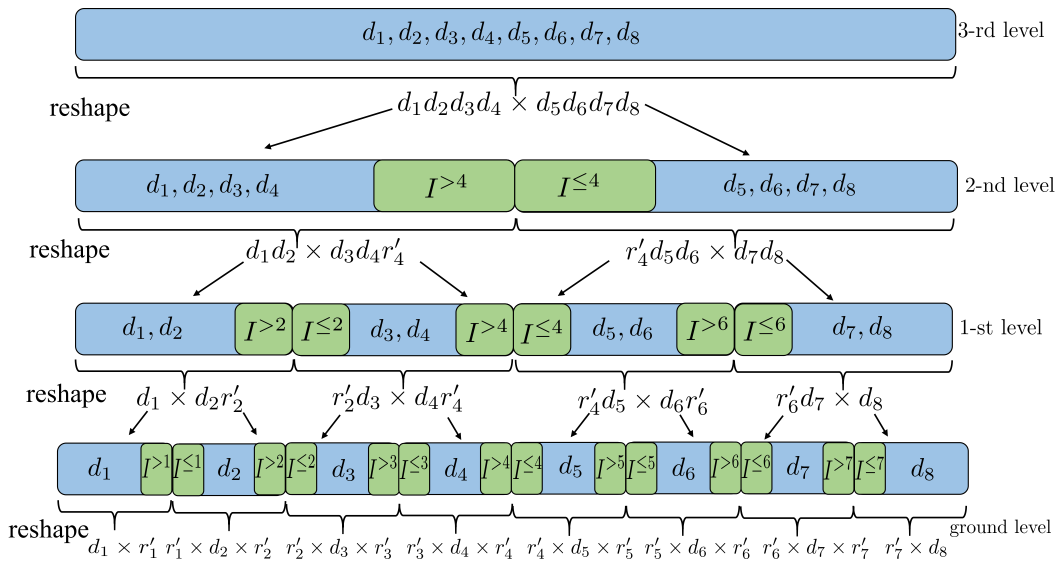

Figure 4: Interpolation steps on the balanced canonical dimension tree.

In a nutshell, the analysis mainly involves the following two procedures:

- **Error bound for how approximation error transfers to next level:** For any $1 \leq p < k < q \leq N$, we first define $\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}$ as a submatrix of $\boldsymbol{T}^{\langle k\rangle}$ when row and column indices $I^{\leq p-1}$ and $I^{>q}$ are respectively chosen from the multi-index $\{d_1 \cdots d_{p-1}\}$ and $\{d_{q+1} \cdots d_N\}$. Following the above discussion, $\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} \approx \boldsymbol{C}[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{R}$ since $\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}$ is low-rank and $\widehat{\boldsymbol{T}}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} = \widehat{\boldsymbol{C}}[\boldsymbol{U}]^{\dagger}_{\tau_k}\widehat{\boldsymbol{R}}$ according to (4), where $\boldsymbol{C} = \boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}(:,I^{>k})$, $\boldsymbol{R} = \boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}(I^{\leq k},:)$, $\boldsymbol{U} = \boldsymbol{T}^{\langle k\rangle}(I^{\leq k},I^{>k})$ as $\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},I^{>k})$ is the same as $\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}(I^{\leq k},I^{>k})$, and same notation holds for $\widehat{\boldsymbol{C}}$ and $\widehat{\boldsymbol{R}}$. We will bound $\|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widehat{\boldsymbol{T}}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F$ from above in the form of
$$\|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widehat{\boldsymbol{T}}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F \leq h_1 + h_2(\|\boldsymbol{E}_C\|_F, \|\boldsymbol{E}_R\|_F), \tag{14}$$
where
$$\boldsymbol{E}_C = \boldsymbol{C} - \widehat{\boldsymbol{C}}, \quad \boldsymbol{E}_R = \boldsymbol{R} - \widehat{\boldsymbol{R}}. \tag{15}$$
In (14), h_1 is independent to $\boldsymbol{E}_C$ and $\boldsymbol{E}_R$, while $h_2(\|\boldsymbol{E}_C\|_F, \|\boldsymbol{E}_R\|_F)$ highlights how the approximation error depends on the previous layer, or how the error transfers to next layer.
- **Error bound for the entire tensor:** We can then recursively apply (14) at most $\lceil \log_2 N\rceil$ times to get the bound for $\|\mathcal{T} - \widehat{\mathcal{T}}\|_F$. In particular, let e_ℓ denote the largest approximation error of $\|\boldsymbol{E}_C\|_F$ and $\|\boldsymbol{E}_R\|_F$ in the l-th layer as in Figure 4. Then (14) implies that
$$e_{l+1} \leq h_1 + h_2(e_l, e_l). \tag{16}$$
Using $e_0 = 0$, we recursively apply the above equation to get the bound for $e_{\lceil \log_2 N\rceil}$, which corresponds to $\left\|\mathcal{T} - \widehat{\mathcal{T}}\right\|_F$.

B.2 Main Proofs

Proof We now prove Theorem 2 by following the above two procedures.

Error bound for how approximation error transfers to next level: Our goal is to derive (14) with $\tau_k = 0$. Since $\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}}$ is a submatrix of $\boldsymbol{T}^{\langle k \rangle}$ and the latter satisfies $\epsilon_k = \|\boldsymbol{T}^{\langle k \rangle} - \boldsymbol{T}^{\langle k \rangle}_{r_k}\|_F$, it follows that $\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}}$ is also approximately low-rank with $\left\|\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}} - \boldsymbol{T}^{\langle k \rangle}_{r_k, I^{\leq p-1}, I^{>q}}\right\|_F \leq \epsilon_k$. Thus, Theorem 6 ensures that there exist indices $I^{\leq k}$ and $I^{>k}$ such that

$$\left\|\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}} - \boldsymbol{C}\boldsymbol{U}^{-1}\boldsymbol{R}\right\|_F \leq (r+1)\epsilon, \tag{17}$$

where $\epsilon = \max_{k=1,\dots,N-1} \epsilon_k$ and $r = \max_{k=1,\dots,N-1} r_k$, $\boldsymbol{C} = \boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}}(:, I^{>k})$.

We now bound the difference between $\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}}$ and $\widehat{\boldsymbol{T}}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}}$ by

$$\begin{aligned}
& \left\|\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}} - \widehat{\boldsymbol{T}}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}}\right\|_F \\
= \;& \left\|\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}} - (\boldsymbol{C} - \boldsymbol{E}_C)\,\boldsymbol{U}^{-1}\,(\boldsymbol{R} - \boldsymbol{E}_R)\right\|_F \\
\leq \;& \left\|\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}} - \boldsymbol{C}\boldsymbol{U}^{-1}\boldsymbol{R}\right\|_F + \left\|\boldsymbol{E}_C\boldsymbol{U}^{-1}\boldsymbol{R}\right\|_F + \left\|\boldsymbol{C}\boldsymbol{U}^{-1}\boldsymbol{E}_R\right\|_F + \left\|\boldsymbol{E}_C\boldsymbol{U}^{-1}\boldsymbol{E}_R\right\|_F \\
\leq \;& (r+1)\epsilon + \|\boldsymbol{E}_C\|_F\,\|\boldsymbol{U}^{-1}\boldsymbol{R}\|_2 + \|\boldsymbol{C}\boldsymbol{U}^{-1}\|_2\,\|\boldsymbol{E}_R\|_F + \|\boldsymbol{E}_C\boldsymbol{U}^{-1}\|_2\,\|\boldsymbol{E}_R\|_F \\
\leq \;& (r+1)\epsilon + \kappa\,\|\boldsymbol{E}_C\|_F + \kappa\,\|\boldsymbol{E}_R\|_F + \kappa\,\|\boldsymbol{E}_R\|_F \\
\leq \;& (r+1)\epsilon + 3\kappa\max\{\|\boldsymbol{E}_C\|_F, \|\boldsymbol{E}_R\|_F\},
\end{aligned} \tag{18}$$

where the third inequality follows because $\boldsymbol{C} = \boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}}(:, I^{>k})$ is a submatrix of $\boldsymbol{T}^{\langle k \rangle}(:, I^{>k})$ which implies that $\|\boldsymbol{C}\boldsymbol{U}^{-1}\|_2 \leq \left\|\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}}(:, I^{>k}) \cdot \boldsymbol{T}^{\langle k \rangle}(I^{\leq k}, I^{>k})^{-1}\right\|_2 \leq \kappa$ and by a similar argument we have $\left\|\boldsymbol{U}^{-1}\boldsymbol{R}\right\|_2 \leq \kappa$ and $\left\|\boldsymbol{E}_C\boldsymbol{U}^{-1}\right\|_2 \leq \kappa$.

Error bound for the entire tensor: Following the same notation as (16), (18) shows that

$$e_{l+1} \leq (r+1)\epsilon + 3\kappa e_l, \tag{19}$$

which together with $e_0 = 0$ implies that

$$\left\|\mathcal{T} - \widehat{\mathcal{T}}\right\|_F = e_{\lceil \log_2 N \rceil} \leq \frac{(3\kappa)^{\lceil \log_2 N \rceil} - 1}{3\kappa - 1}(r+1)\epsilon. \tag{20}$$

∎

C Proof of Theorem 3

Before proving Theorem 3, we provide several useful results.

Lemma 1 *[62, Proposition 6.4] For any rank-r $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with compact SVD $\boldsymbol{A} = \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ where $\boldsymbol{W} \in \mathbb{R}^{m \times r}, \boldsymbol{\Sigma} \in \mathbb{R}^{r \times r}$, and $\boldsymbol{V} \in \mathbb{R}^{n \times r}$, suppose $\boldsymbol{C}\boldsymbol{U}^\dagger\boldsymbol{R}$ is its CUR decomposition with selected row indices I and column indices J. Then*

$$\left\|\boldsymbol{C}\boldsymbol{U}^\dagger\right\|_2 = \left\|\boldsymbol{W}^\dagger(I,:)\right\|_2, \; \left\|\boldsymbol{U}^\dagger\boldsymbol{R}\right\|_2 = \left\|\boldsymbol{V}^\dagger(J,:)\right\|_2.$$

Noting that $\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}}$ is a submatrix of $\boldsymbol{T}^{\langle k \rangle}$ which is approximately low-rank, we can approximate $\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}}$ by cross approximation. In particular, using Theorem 7 and Lemma 1 for $\boldsymbol{T}^{\langle k \rangle}_{I^{\leq p-1}, I^{>q}}$, we can obtain the following result.

Lemma 2 *Suppose $\boldsymbol{T}^{\langle k \rangle}_{r_k}$ is the best rank r_k approximation of the k-th unfolding matrix $\boldsymbol{T}^{\langle k \rangle}$, such that $\boldsymbol{T}^{\langle k \rangle} = \boldsymbol{T}^{\langle k \rangle}_{r_k} + \boldsymbol{F}^{\langle k \rangle}$. For any $1 \leq p < k < q \leq N$ and indices $I^{\leq p-1}$ and $I^{>q}$, we have*

$\boldsymbol{T}^{\langle k\rangle}_{I^{\le p-1},I^{>q}} = \boldsymbol{T}^{\langle k\rangle}_{r_k,I^{\le p-1},I^{>q}} + \boldsymbol{F}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}$, where $\boldsymbol{T}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}$, $\boldsymbol{T}^{\langle k\rangle}_{r_k,I^{\le p-1},I^{>q}}$ and $\boldsymbol{F}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}$ are respectively submatrices of $\boldsymbol{T}^{\langle k\rangle}$, $\boldsymbol{T}^{\langle k\rangle}_{r_k}$ and $\boldsymbol{F}^{\langle k\rangle}$. Let $\boldsymbol{T}^{\langle k\rangle}_{r_k} = \boldsymbol{W}_{(k)}\boldsymbol{\Sigma}_{(k)}{\boldsymbol{V}_{(k)}}^T$ be the compact SVD of $\boldsymbol{T}^{\langle k\rangle}_{r_k}$. Then for any $(I^{\le k}, I^{>k})$ as long as $\operatorname{rank}(\boldsymbol{T}^{\langle k\rangle}_{r_k}(I^{\le k}, I^{>k})) = r_k$, we have

$$\begin{aligned}
&\left\|\boldsymbol{T}^{\langle k\rangle}_{I^{\le p-1},I^{>q}} - \boldsymbol{T}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}(:,I^{>k})[\boldsymbol{T}^{\langle k\rangle}(I^{\le k},I^{>k})]^\dagger \boldsymbol{T}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}(I^{\le k},:)\right\|_F \\
\le&\|\boldsymbol{F}^{\langle k\rangle}\|_F\Bigg(\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^\dagger\|_2+\|[\boldsymbol{V}_{(k)}(I^{>k},:)]^\dagger\|_2+3\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^\dagger\|_2\|[\boldsymbol{V}_{(k)}(I^{>k},:)]^\dagger\|_2+1 \\
&+d\Big(\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^\dagger\|_2+\|[\boldsymbol{V}_{(k)}(I^{>k},:)]^\dagger\|_2+\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^\dagger\|_2\|[\boldsymbol{V}_{(k)}(I^{>k},:)]^\dagger\|_2+1\Big)\Bigg),
\end{aligned}$$

where $d = \|[\boldsymbol{T}^{\langle k\rangle}(I^{\le k}, I^{>k})]^\dagger\|_2$.

The following result extends Lemma 1 to the case where the matrix is only approximately low-rank.

Lemma 3 *Suppose $\boldsymbol{A}$ is approximately low-rank of the form $\boldsymbol{A} = \boldsymbol{A}_r + \boldsymbol{F} = \boldsymbol{W\Sigma V}^\top + \boldsymbol{F}$, where $\boldsymbol{A}_r$ has rank r with compact SVD $\boldsymbol{A}_r = \boldsymbol{W\Sigma V}^\top$. Then for any selected row and column indices I, J, and $\tau \ge 0$, we have*

$$\left\|\boldsymbol{A}(:,J)[\boldsymbol{A}(I,J)]^\dagger_\tau\right\|_2 \le \left\|\boldsymbol{W}^\dagger(I,:)\right\|_2 + \left(1+\|\boldsymbol{W}^\dagger(I,:)\|_2\right)\left\|[\boldsymbol{A}(I,J)]^\dagger_\tau\right\|_2\|\boldsymbol{F}(:,J)\|_F, \tag{21}$$

and

$$\left\|[\boldsymbol{A}(I,J)]^\dagger_\tau\boldsymbol{A}(I,:)\right\|_2 \le \|\boldsymbol{V}^\dagger(J,:)\|_2 + \left(1+\|\boldsymbol{V}^\dagger(J,:)\|_2\right)\left\|[\boldsymbol{A}(I,J)]^\dagger_\tau\right\|_2\|\boldsymbol{F}(I,:)\|_F. \tag{22}$$

Proof Noting that $\boldsymbol{A}(:,J) = \boldsymbol{A}_r(:,J) + \boldsymbol{F}(;,J)$ and $\boldsymbol{A}(I,:) = \boldsymbol{A}_r(I,:) + \boldsymbol{F}(I,:)$, we have

$$\begin{aligned}
&\left\|\boldsymbol{A}(:,J)[\boldsymbol{A}(I,J)]^\dagger_\tau\right\|_2 \\
=\ & \|(\boldsymbol{A}_r(:,J)+\boldsymbol{F}(;,J))[\boldsymbol{A}(I,J)]^\dagger_\tau\|_2 \\
\le\ & \|\boldsymbol{A}_r(:,J)[\boldsymbol{A}(I,J)]^\dagger_\tau\|_2 + \|\boldsymbol{F}(;,J)[\boldsymbol{A}(I,J)]^\dagger_\tau\|_2 \\
\le\ & \|\boldsymbol{A}_r(:,J)\boldsymbol{A}_r^\dagger(I,J)\boldsymbol{A}_r(I,J)[\boldsymbol{A}(I,J)]^\dagger_\tau\|_2 + \|\boldsymbol{F}(;,J)\|_2\|[\boldsymbol{A}(I,J)]^\dagger_\tau\|_2 \\
\le\ & \|\boldsymbol{A}_r(:,J)\boldsymbol{A}_r^\dagger(I,J)\|_2\|(\boldsymbol{A}(I,J)-\boldsymbol{F}(I,J))[\boldsymbol{A}(I,J)]^\dagger_\tau\|_2 + \|\boldsymbol{F}(;,J)\|_2\|[\boldsymbol{A}(I,J)]^\dagger_\tau\|_2 \\
\le\ & \|\boldsymbol{A}_r(:,J)\boldsymbol{A}_r^\dagger(I,J)\|_2(1+\|\boldsymbol{F}(I,J)\|_2\|[\boldsymbol{A}(I,J)]^\dagger_\tau\|_2) + \|\boldsymbol{F}(;,J)\|_2\|[\boldsymbol{A}(I,J)]^\dagger_\tau\|_2 \\
\le\ & \left\|\boldsymbol{W}^\dagger(I,:)\right\|_2 + \left(1+\|\boldsymbol{W}^\dagger(I,:)\|_2\right)\left\|[\boldsymbol{A}(I,J)]^\dagger_\tau\right\|_2\|\boldsymbol{F}(:,J)\|_F,
\end{aligned} \tag{23}$$

where the penultimate line follows because $\|\boldsymbol{A}(I,J)[\boldsymbol{A}(I,J)]^\dagger_\tau\|_2 \le 1$ and in the last line we use Lemma 1. Likewise, we can obtain (22) with a similar argument.

■

We are now ready to prove Theorem 3.

Proof (of Theorem 3)

Recall the following definitions that will be used to simplify the presentation:

$$\begin{aligned}
&a = \max_{k=1,\dots,N-1}\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^\dagger\|_2, \quad b = \max_{k=1,\dots,N-1}\|[\boldsymbol{V}_{(k)}(I^{>k},:)]^\dagger\|_2, \\
&c = \max_{k=1,\dots,N-1}\|[\boldsymbol{T}^{\langle k\rangle}(I^{\le k},I^{>k})]^\dagger\|_2 \quad r = \max_{k=1,\dots,N-1} r_k, \quad \epsilon = \max_{k=1,\dots,N-1}\|\boldsymbol{F}^{\langle k\rangle}\|_F.
\end{aligned}$$

Error bound for how approximation error transfers to next level: To derive (14), we first exploit the approximate low-rankness of $\boldsymbol{T}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}$ to write it as

$$\boldsymbol{T}^{\langle k\rangle}_{I^{\le p-1},I^{>q}} = \boldsymbol{C}[\boldsymbol{U}]^\dagger_{\tau_k}\boldsymbol{R} + \boldsymbol{H}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}, \tag{24}$$

where $\boldsymbol{H}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}$ is the residual of the cross approximation.

Furthermore, by the construction of $\widehat{\boldsymbol{T}}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}$, we can also express it by the following cross approximation

$$\widehat{\boldsymbol{T}}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} = \widehat{\boldsymbol{C}}[\boldsymbol{U}]^{\dagger}_{\tau_k}\widehat{\boldsymbol{R}}. \tag{25}$$

We now quantify the difference between $\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}$ and $\widehat{\boldsymbol{T}}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}$ as

$$\begin{aligned}
&\|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widehat{\boldsymbol{T}}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F \\
=&\|\boldsymbol{C}[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{R} + \boldsymbol{H}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widehat{\boldsymbol{C}}[\boldsymbol{U}]^{\dagger}_{\tau_k}\widehat{\boldsymbol{R}}\|_F \\
\leq&\|\boldsymbol{H}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F + \|\boldsymbol{C}[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{R} - \widehat{\boldsymbol{C}}[\boldsymbol{U}]^{\dagger}_{\tau_k}\widehat{\boldsymbol{R}}\|_F \\
=&\|\boldsymbol{H}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F + \|\boldsymbol{C}[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{R} - (\boldsymbol{C}-\boldsymbol{E}_C)[\boldsymbol{U}]^{\dagger}_{\tau_k}(\boldsymbol{R}-\boldsymbol{E}_R)\|_F \\
\leq&\|\boldsymbol{H}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F + \|\boldsymbol{C}[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{E}_R\|_F + \|\boldsymbol{E}_C[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{E}_R\|_F + \|\boldsymbol{E}_C[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{R}\|_F.
\end{aligned} \tag{26}$$

Below we provide upper bounds for the four terms in the above equation. First, utilizing Lemma 3 gives

$$\begin{aligned}
&\|\boldsymbol{C}[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{E}_R\|_F \\
\leq&\ \|\boldsymbol{C}[\boldsymbol{U}]^{\dagger}_{\tau_k}\|_2\|\boldsymbol{E}_R\|_F \leq \|\boldsymbol{T}^{\langle k\rangle}(:,I^{>k})[\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}_{\tau_k}\|_2\|\boldsymbol{E}_R\|_F \\
\leq&\ (\|[\boldsymbol{W}_{(k)}(I^{\leq k},:)]^{\dagger}\|_2+(1+\|[\boldsymbol{W}_{(k)}(I^{\leq k},:)]^{\dagger}\|_2)\|[\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}_{\tau_k}\|_2\|\boldsymbol{F}^{\langle k\rangle}_{r_k}(:,I^{>k})\|_F)\|\boldsymbol{E}_R\|_F \\
\leq&\ (a+(1+a)c\epsilon)\|\boldsymbol{E}_R\|_F.
\end{aligned} \tag{27}$$

With the same argument, we have

$$\|\boldsymbol{E}_C[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{R}\|_F \leq (b+(1+b)c\epsilon)\|\boldsymbol{E}_C\|_F. \tag{28}$$

Also, noting that $\|[\boldsymbol{U}]^{\dagger}_{\tau_k}\|_2 \leq \frac{1}{\tau_k}$, we get

$$\|\boldsymbol{E}_C[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{E}_R\|_F \leq \|[\boldsymbol{U}]^{\dagger}_{\tau_k}\|_2\|\boldsymbol{E}_C\|_F\|\boldsymbol{E}_R\|_F \leq \frac{1}{\tau_k}\|\boldsymbol{E}_C\|_F\|\boldsymbol{E}_R\|_F. \tag{29}$$

The term $\|\boldsymbol{H}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F$ can be upper bounded as

$$\begin{aligned}
&\|\boldsymbol{H}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F \\
=&\ \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \boldsymbol{C}\boldsymbol{U}^{\dagger}\boldsymbol{R} + \boldsymbol{C}\boldsymbol{U}^{\dagger}\boldsymbol{R} - \boldsymbol{C}[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{R}\|_F \\
\leq&\ \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \boldsymbol{C}\boldsymbol{U}^{\dagger}\boldsymbol{R}\|_F + \|\boldsymbol{C}\boldsymbol{U}^{\dagger}\boldsymbol{R} - \boldsymbol{C}[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{R}\|_F \\
=&\ \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \boldsymbol{C}\boldsymbol{U}^{\dagger}\boldsymbol{R}\|_F + \|\boldsymbol{C}\boldsymbol{U}^{\dagger}\boldsymbol{U}(\boldsymbol{U}^{\dagger} - [\boldsymbol{U}]^{\dagger}_{\tau_k})\boldsymbol{U}\boldsymbol{U}^{\dagger}\boldsymbol{R}\|_F \\
=&\ \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \boldsymbol{C}\boldsymbol{U}^{\dagger}\boldsymbol{R}\|_F + \|\boldsymbol{C}\boldsymbol{U}^{\dagger}(\boldsymbol{U} - [\boldsymbol{U}]_{\tau_k})\boldsymbol{U}^{\dagger}\boldsymbol{R}\|_F \\
\leq&\ \|\boldsymbol{T}^{\langle k\rangle}(:,I^{>k})[\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\|_2\|\boldsymbol{U} - [\boldsymbol{U}]_{\tau_k}\|_F\|[\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},:)\|_2 \\
&\ + \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \boldsymbol{C}\boldsymbol{U}^{\dagger}\boldsymbol{R}\|_F \\
\leq&\ \sqrt{\|\boldsymbol{F}^{\langle k\rangle}(I^{\leq k},I^{>k})\|_F^2 + r_k^2\tau_k^2}\|\boldsymbol{T}^{\langle k\rangle}(:,I^{>k})[\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\|_2\|[\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},:)\|_2 \\
&\ + \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \boldsymbol{C}\boldsymbol{U}^{\dagger}\boldsymbol{R}\|_F \\
\leq&\ (\|\boldsymbol{F}^{\langle k\rangle}\|_F + r_k\tau_k)\|\boldsymbol{T}^{\langle k\rangle}(:,I^{>k})[\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\|_2\|[\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\boldsymbol{T}^{\langle k\rangle}(I^{\leq k},:)\|_2 \\
&\ + \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \boldsymbol{C}\boldsymbol{U}^{\dagger}\boldsymbol{R}\|_F \\
\leq&\ (a+b+3ab+1)\epsilon + (a+b+ab+1)c\epsilon^2 + (\epsilon + r\tau_k)(a+(1+a)c\epsilon)(b+(1+b)c\epsilon),
\end{aligned} \tag{30}$$

where $\boldsymbol{C} = \boldsymbol{C}\boldsymbol{U}^{\dagger}\boldsymbol{U}$, $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{U}^{\dagger}\boldsymbol{R}$ and $\boldsymbol{U}[\boldsymbol{U}]^{\dagger}_{\tau_k}\boldsymbol{U} = [\boldsymbol{U}]_{\tau_k}$ are respectively used in the second and third equality, the third inequality follows because $\|\boldsymbol{U} - [\boldsymbol{U}]_{\tau_k}\|_F \leq \sqrt{\|\boldsymbol{F}^{\langle k\rangle}(I^{\leq k},I^{>k})\|_F^2 + r_k^2\tau_k^2}$, and the last line uses Lemma 2 and Lemma 3.

Plugging (27), (28), (29), and (30) into (26) gives

$$\begin{aligned}
&\|\boldsymbol{T}^{\langle k\rangle}_{I_{\le p-1},I_{>q}} - \widehat{\boldsymbol{T}}^{\langle k\rangle}_{I_{\le p-1},I_{>q}}\|_F \\
\le\ & (a+b+3ab+1)\epsilon + (a+b+ab+1)c\epsilon^2 + (\epsilon + r\tau_k)(a+(1+a)c\epsilon)(b+(1+b)c\epsilon) \\
&+ (a+(1+a)c\epsilon)\|\boldsymbol{E}_R\|_F + (b+(1+b)c\epsilon)\|\boldsymbol{E}_C\|_F + \frac{1}{\tau_k}\|\boldsymbol{E}_C\|_F\|\boldsymbol{E}_R\|_F. \qquad (31)
\end{aligned}$$

Error bound for the entire tensor: Following the same notation as (16), (31) shows that (by setting $\tau_k = e_l$ for the l-th level)

$$\begin{aligned}
e_{l+1} \le\ & (a+b+3ab+1)\epsilon + (a+b+ab+1)c\epsilon^2 + \epsilon(a+(1+a)c\epsilon)(b+(1+b)c\epsilon) \\
&+ (1+a+b+(1+a)c\epsilon+(1+b)c\epsilon + r(a+(1+a)c\epsilon)(b+(1+b)c\epsilon))e_l \\
=\ & (a+b+4ab+1)\epsilon + (2a+2b+3ab+1)c\epsilon^2 + (1+a+b+ab)c^2\epsilon^3 \\
&+ \left(1+a+b+rab+(2+a+b+ar+br+2abr)c\epsilon + r(1+a+b+ab)c^2\epsilon^2\right)e_l, \qquad (32)
\end{aligned}$$

which together with $e_0 = 0$ implies that

$$\|\mathcal{T} - \widehat{\mathcal{T}}\|_F = e_{\lceil \log_2 N\rceil} \le \frac{\alpha_1(a,b,c,\epsilon,r)^{\lceil \log_2 N\rceil} - 1}{\alpha_1(a,b,c,\epsilon,r) - 1}\beta_1(a,b,c,\epsilon),$$

where

$$\begin{aligned}
\alpha_1(a,b,c,\epsilon,r) &= 1+a+b+rab+(2+a+b+ar+br+2abr)c\epsilon + r(1+a+b+ab)c^2\epsilon^2, \\
\beta_1(a,b,c,\epsilon) &= (a+b+4ab+1)\epsilon + (2a+2b+3ab+1)c\epsilon^2 + (1+a+b+ab)c^2\epsilon^3.
\end{aligned}$$

■

D Proof of Theorem 4

The proof is similar to the proof of Theorem 3. We include the proof for the sake of completeness.

Before deriving Theorem 4, we first consider matrix cross approximation with measurement error. In this case, we obtain noisy columns $\widetilde{\boldsymbol{C}}$, rows $\widetilde{\boldsymbol{R}}$, and intersection matrix $\widetilde{\boldsymbol{U}}$. To simplify the notation, we describe the measurement error in the selected rows I and columns J by $\boldsymbol{E} \in \mathbb{R}^{m\times n}$ such that $\boldsymbol{E}$ has non-zero elements only in the rows I or columns J corresponding to the noise. We can then rewrite

$$\widetilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{E},\ \widetilde{\boldsymbol{C}} = \widetilde{\boldsymbol{A}}(:,J), \widetilde{\boldsymbol{U}} = \widetilde{\boldsymbol{A}}(I,J), \widetilde{\boldsymbol{R}} = \widetilde{\boldsymbol{A}}(I,:) \qquad (33)$$

and $\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^\dagger\widetilde{\boldsymbol{R}}$ can be viewed as cross approximation for $\widetilde{\boldsymbol{A}}$. However, we note that here we want to ensure $\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^\dagger\widetilde{\boldsymbol{R}}$ is a stable approximation to $\boldsymbol{A}$ rather than $\widetilde{\boldsymbol{A}}$. Thus, we extend Theorem 7 to this case.

Theorem 8 *Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ be an approximately low-rank matrix that can be decomposed as $\boldsymbol{A} = \boldsymbol{A}_r + \boldsymbol{F}$, where $\boldsymbol{A}_r$ is rank-r. Let $\boldsymbol{A}_r = \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{V}^T$ be the compact SVD of $\boldsymbol{A}_r$. Then the noisy cross approximation defined in $\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^\dagger\widetilde{\boldsymbol{R}}$ (33) with $\operatorname{rank}(\boldsymbol{A}_r(I,J)) = r$ satisfies*

$$\begin{aligned}
&\|\boldsymbol{A} - \widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^\dagger\widetilde{\boldsymbol{R}}\|_F \\
\le\ & (\|\boldsymbol{W}^\dagger(I,:)\|_2 + \|\boldsymbol{V}^\dagger(J,:)\|_2 + 3\|\boldsymbol{W}^\dagger(I,:)\|_2\|\boldsymbol{V}^\dagger(J,:)\|_2)\|\boldsymbol{E}\|_F + \\
&(\|\boldsymbol{W}^\dagger(I,:)\|_2 + \|\boldsymbol{V}^\dagger(J,:)\|_2 + 3\|\boldsymbol{W}^\dagger(I,:)\|_2\|\boldsymbol{V}^\dagger(J,:)\|_2 + 1)\|\boldsymbol{F}\|_F + \\
&(\|\boldsymbol{W}^\dagger(I,:)\|_2 + \|\boldsymbol{V}^\dagger(J,:)\|_2 + \|\boldsymbol{W}^\dagger(I,:)\|_2\|\boldsymbol{V}^\dagger(J,:)\|_2 + 1)\|\widetilde{\boldsymbol{U}}^\dagger\|_2(\|\boldsymbol{E}\|_F + \|\boldsymbol{F}\|_F)^2. \qquad (34)
\end{aligned}$$

Proof In order to bound $\|\boldsymbol{A} - \widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^\dagger\widetilde{\boldsymbol{R}}\|_F$, similar to the derivation of Theorem 7, we first get

$$\begin{aligned}
&\|\boldsymbol{A}_r - \widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^\dagger\widetilde{\boldsymbol{R}}\|_F \\
\le\ & (\|\boldsymbol{W}^\dagger(I,:)\|_2 + \|\boldsymbol{V}^\dagger(J,:)\|_2 + 3\|\boldsymbol{W}^\dagger(I,:)\|_2\|\boldsymbol{V}^\dagger(J,:)\|_2)\|\boldsymbol{E}\|_F + \\
&(\|\boldsymbol{W}^\dagger(I,:)\|_2 + \|\boldsymbol{V}^\dagger(J,:)\|_2 + 3\|\boldsymbol{W}^\dagger(I,:)\|_2\|\boldsymbol{V}^\dagger(J,:)\|_2)\|\boldsymbol{F}\|_F + \\
&(\|\boldsymbol{W}^\dagger(I,:)\|_2 + \|\boldsymbol{V}^\dagger(J,:)\|_2 + \|\boldsymbol{W}^\dagger(I,:)\|_2\|\boldsymbol{V}^\dagger(J,:)\|_2 + 1)\|\widetilde{\boldsymbol{U}}^\dagger\|_2(\|\boldsymbol{E}\|_F + \|\boldsymbol{F}\|_F)^2. \qquad (35)
\end{aligned}$$

Combing $\|\boldsymbol{A}-\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{R}}\|_F \le \|\boldsymbol{F}\|_F + \|\boldsymbol{A}_r-\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{R}}\|_F$ and (35), we have (34).

$\blacksquare$

The following result extends Lemma 2 to the case with measurement error.

Lemma 4 *For any $1 \le p < k < q \le N$, a low-rank model with the measurement error $\widetilde{\boldsymbol{T}}^{\langle k\rangle}_{I^{\le p-1},I^{>q}} = \boldsymbol{T}^{\langle k\rangle}_{r_k,I^{\le p-1},I^{>q}} + \boldsymbol{E}^{\langle k\rangle}_{I^{\le p-1},I^{>q}} + \boldsymbol{F}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}$ is constructed where $\widetilde{\boldsymbol{T}}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}$ is a submatrix of $\widetilde{\boldsymbol{T}}^{\langle k\rangle}$. When $\operatorname{rank}(\boldsymbol{T}^{\langle k\rangle}_{r_k}(I^{\le k},I^{>k})) = r_k$ is satisfied, we have*

$$
\begin{aligned}
&\|\boldsymbol{T}^{\langle k\rangle}_{I^{\le p-1},I^{>q}} - \widetilde{\boldsymbol{T}}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}(:,I^{>k})[\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\le k},I^{>k})]^{\dagger}\widetilde{\boldsymbol{T}}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}(I^{\le k},:)\|_F\\
\le\ &(\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^{\dagger}\|_2 + \|[\boldsymbol{V}_{(k)}(I^{>k},:)]^{\dagger}\|_2 + 3\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^{\dagger}\|_2\|[\boldsymbol{V}_{(k)}(I^{>k},:)]^{\dagger}\|_2)\|\boldsymbol{E}^{\langle k\rangle}\|_F\\
&+(\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^{\dagger}\|_2+\|[\boldsymbol{V}_{(k)}(I^{>k},:)]^{\dagger}\|_2+3\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^{\dagger}\|_2\|[\boldsymbol{V}_{(k)}(I^{>k},:)]^{\dagger}\|_2+1)\|\boldsymbol{F}^{\langle k\rangle}_{r_k}\|_F\\
&+(\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^{\dagger}\|_2 + \|[\boldsymbol{V}_{(k)}(I^{>k},:)]^{\dagger}\|_2 + \|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^{\dagger}\|_2\|[\boldsymbol{V}_{(k)}(I^{>k},:)]^{\dagger}\|_2 + 1)\\
&\cdot\|[\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\le k},I^{>k})]^{\dagger}\|_2(\|\boldsymbol{E}^{\langle k\rangle}\|_F + \|\boldsymbol{F}^{\langle k\rangle}_{r_k}\|_F)^2,
\end{aligned}
\tag{36}
$$

where $\boldsymbol{T}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}$ is a submatrix of $\boldsymbol{T}^{\langle k\rangle}$ and $\boldsymbol{T}^{\langle k\rangle}_{r_k} = \boldsymbol{W}_{(k)}\boldsymbol{\Sigma}_{(k)}{\boldsymbol{V}_{(k)}}^T$ is the SVD of $\boldsymbol{T}^{\langle k\rangle}_{r_k}$.

Likewise, we extend Lemma 3 to the measurement noise case.

Lemma 5 *Given I and J, we set a low-rank model as $\widetilde{\boldsymbol{A}} = \boldsymbol{A}_r + \boldsymbol{E} + \boldsymbol{F} = \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{V}^T + \boldsymbol{E} + \boldsymbol{F}$. For $\tau \ge 0$, we have*

$$
\begin{aligned}
&\left\|\widetilde{\boldsymbol{A}}(:,J)[\widetilde{\boldsymbol{A}}(I,J)]^{\dagger}_{\tau}\right\|_2 \le \left\|\boldsymbol{W}^{\dagger}(I,:)\right\|_2 + \left(1+\|\boldsymbol{W}^{\dagger}(I,:)\|_2\right)\left\|[\widetilde{\boldsymbol{A}}(I,J)]^{\dagger}_{\tau}\right\|_2(\|\boldsymbol{E}(:,J)\|_F + \|\boldsymbol{F}(:,J)\|_F),\\
&\left\|[\widetilde{\boldsymbol{A}}(I,J)]^{\dagger}_{\tau}\widetilde{\boldsymbol{A}}(I,:)\right\|_2 \le \left\|\boldsymbol{V}^{\dagger}(J,:)\right\|_2 + (1+\left\|\boldsymbol{V}^{\dagger}(J,:)\right\|_2)\left\|[\widetilde{\boldsymbol{A}}(I,J)]^{\dagger}_{\tau}\right\|_2(\|\boldsymbol{E}(:,J)\|_F + \|\boldsymbol{F}(:,J)\|_F).
\end{aligned}
$$

We are now ready to prove Theorem 4.

Proof (of Theorem 4)

Error bound for how approximation error transfers to next level: To derive (14), we first exploit the approximate low-rankness of $\boldsymbol{T}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}$ to write it as

$$
\boldsymbol{T}^{\langle k\rangle}_{I^{\le p-1},I^{>q}} = \widetilde{\boldsymbol{C}}[\widetilde{\boldsymbol{U}}]^{\dagger}_{\tau_k}\widetilde{\boldsymbol{R}} + \boldsymbol{H}^{\langle k\rangle}_{I^{\le p-1},I^{>q}},
\tag{37}
$$

where $\widetilde{\boldsymbol{C}} = \widetilde{\boldsymbol{T}}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}(:,I^{>k})$, $\widetilde{\boldsymbol{R}} = \widetilde{\boldsymbol{T}}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}(I^{\le k},:)$ and $\widetilde{\boldsymbol{U}} = \widetilde{\boldsymbol{T}}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}(I^{\le k},I^{>k}) = \widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\le k},I^{>k})$. In addition, $\widehat{\boldsymbol{T}}^{\langle k\rangle}_{I^{\le p-1},I^{>q}}$ can be written as

$$
\widehat{\boldsymbol{T}}^{\langle k\rangle}_{I^{\le p-1},I^{>q}} = \widehat{\boldsymbol{C}}[\widetilde{\boldsymbol{U}}]^{\dagger}_{\tau_k}\widehat{\boldsymbol{R}}.
\tag{38}
$$

Denote by $\boldsymbol{E}_C = \widetilde{\boldsymbol{C}} - \widehat{\boldsymbol{C}}$ and $\boldsymbol{E}_R = \widetilde{\boldsymbol{R}} - \widehat{\boldsymbol{R}}$. Now using Lemma 5 and $\|[\widetilde{\boldsymbol{U}}]^{\dagger}_{\tau_k}\|_2 \le \frac{1}{\tau_k}$, we obtain

$$
\begin{aligned}
&\|\widetilde{\boldsymbol{C}}[\widetilde{\boldsymbol{U}}]^{\dagger}_{\tau_k}\boldsymbol{E}_R\|_F\\
\le\ &\|\widetilde{\boldsymbol{T}}^{\langle k\rangle}(:,I^{>k})[\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\le k},I^{>k})]^{\dagger}_{\tau_k}\|_2\|\boldsymbol{E}_R\|_F\\
\le\ &\Big(\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^{\dagger}\|_2 + (1+\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^{\dagger}\|_2)\|[\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\le k},I^{>k})]^{\dagger}\|_2\\
&\cdot(\|\boldsymbol{E}^{\langle k\rangle}(:,I^{>k})\|_2 + \|\boldsymbol{F}^{\langle k\rangle}(:,I^{>k})\|_2)\Big)\|\boldsymbol{E}_R\|_F\\
\le\ &(a+(1+a)c(\xi+\epsilon))\|\boldsymbol{E}_R\|_F,
\end{aligned}
\tag{39}
$$

where we restate that

$$
\begin{aligned}
&\xi = \|\boldsymbol{E}^{\langle k\rangle}\|_F,\ a = \max_{k=1,\ldots,N-1}\|[\boldsymbol{W}_{(k)}(I^{\le k},:)]^{\dagger}\|_2,\ b = \max_{k=1,\ldots,N-1}\|[\boldsymbol{V}_{(k)}(I^{>k},:)]^{\dagger}\|_2,\\
&c = \max_{k=1,\ldots,N-1}\|[\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\le k},I^{>k})]^{\dagger}\|_2,\ r = \max_{k=1,\ldots,N-1} r_k,\ \epsilon = \max_{k=1,\ldots,N-1}\|\boldsymbol{F}^{\langle k\rangle}\|_F.
\end{aligned}
$$

With the same argument, we can obtain

$$\|\boldsymbol{E}_C[\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger}\widetilde{\boldsymbol{R}}\|_F \leq (b+(1+b)c(\xi+\epsilon))\|\boldsymbol{E}_C\|_F, \tag{40}$$

and

$$\|\boldsymbol{E}_C[\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger}\boldsymbol{E}_R\|_F \leq \|[\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger}\|_2\|\boldsymbol{E}_C\|_F\|\boldsymbol{E}_R\|_F \leq \frac{1}{\tau_k}\|\boldsymbol{E}_C\|_F\|\boldsymbol{E}_R\|_F. \tag{41}$$

Furthermore, it follows from Lemma 4 and Lemma 5 that

$$\begin{aligned}
& \|\boldsymbol{H}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F \\
\leq\ & \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{R}}\|_F + \|\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{R}} - \widetilde{\boldsymbol{C}}[\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger}\widetilde{\boldsymbol{R}}\|_F \\
=\ & \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{R}}\|_F + \|\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{U}}(\widetilde{\boldsymbol{U}}^{\dagger} - [\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger})\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{R}}\|_F \\
=\ & \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{R}}\|_F + \|\widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^{\dagger}(\widetilde{\boldsymbol{U}} - [\widetilde{\boldsymbol{U}}]_{\tau_k})\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{R}}\|_F \\
\leq\ & \|\widetilde{\boldsymbol{T}}^{\langle k\rangle}(:,I^{>k})[\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\|_2\|\widetilde{\boldsymbol{U}} - [\widetilde{\boldsymbol{U}}]_{\tau_k}\|_F\|[\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\leq k},:)\|_2 \\
& + \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{R}}\|_F \\
\leq\ & \sqrt{\|\boldsymbol{E}^{\langle k\rangle}(I^{\leq k},I^{>k}) + \boldsymbol{F}^{\langle k\rangle}(I^{\leq k},I^{>k})\|_F^2 + r_k^2\tau_k^2}\,\|\widetilde{\boldsymbol{T}}^{\langle k\rangle}(:,I^{>k})[\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\|_2 \\
& \cdot\|[\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\leq k},:)\|_2 + \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{R}}\|_F \\
\leq\ & (\|\boldsymbol{E}^{\langle k\rangle}\|_F + \|\boldsymbol{F}^{\langle k\rangle}\|_F + r_k\tau_k)\|\widetilde{\boldsymbol{T}}^{\langle k\rangle}(:,I^{>k})[\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\|_2 \\
& \cdot\|[\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\leq k},I^{>k})]^{\dagger}\widetilde{\boldsymbol{T}}^{\langle k\rangle}(I^{\leq k},:)\|_2 + \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widetilde{\boldsymbol{C}}\widetilde{\boldsymbol{U}}^{\dagger}\widetilde{\boldsymbol{R}}\|_F \\
\leq\ & \epsilon + (\xi+\epsilon+r_k\tau_k)(a+(1+a)c(\xi+\epsilon))(b+(1+b)c(\xi+\epsilon)) \\
& + (a+b+3ab)(\xi+\epsilon) + (a+b+ab+1)c(\xi+\epsilon)^2.
\end{aligned} \tag{42}$$

Combining (39)-(42), we can obtain

$$\begin{aligned}
& \|\boldsymbol{T}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widehat{\boldsymbol{T}}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F = \|\widetilde{\boldsymbol{C}}[\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger}\widetilde{\boldsymbol{R}} + \boldsymbol{H}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}} - \widehat{\boldsymbol{C}}[\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger}\widehat{\boldsymbol{R}}\|_F \\
& \leq \|\boldsymbol{H}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F + \|\widetilde{\boldsymbol{C}}[\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger}\widetilde{\boldsymbol{R}} - (\widetilde{\boldsymbol{C}} - \boldsymbol{E}_C)[\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger}(\widetilde{\boldsymbol{R}} - \boldsymbol{E}_R)\|_F \\
& \leq \|\boldsymbol{H}^{\langle k\rangle}_{I^{\leq p-1},I^{>q}}\|_F + \|\widetilde{\boldsymbol{C}}[\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger}\boldsymbol{E}_R\|_F + \|\boldsymbol{E}_C[\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger}\boldsymbol{E}_R\|_F + \|\boldsymbol{E}_C[\widetilde{\boldsymbol{U}}]_{\tau_k}^{\dagger}\widetilde{\boldsymbol{R}}\|_F \\
& \leq \epsilon + (\xi+\epsilon+r_k\tau_k)(a+(1+a)c(\xi+\epsilon))(b+(1+b)c(\xi+\epsilon)) + (a+b+3ab)(\xi+\epsilon) \\
& \quad + (a+b+ab+1)c(\xi+\epsilon)^2 + (a+(1+a)c(\xi+\epsilon))\|\boldsymbol{E}_R\|_F + (b+(1+b)c(\xi+\epsilon))\|\boldsymbol{E}_C\|_F \\
& \quad + \frac{1}{\tau_k}\|\boldsymbol{E}_C\|_F\|\boldsymbol{E}_R\|_F.
\end{aligned} \tag{43}$$

Error bound for the entire tensor: Similar to the derivation of Theorem 3, (43) implies that (by setting $\tau_k = e_l$ in the l-th layer)

$$\begin{aligned}
e_{l+1} \leq\ & \epsilon + (a+b+4ab)(\xi+\epsilon) + (2a+2b+3ab+1)c(\xi+\epsilon)^2 + (1+a+b+ab)c^2(\xi+\epsilon)^3 \\
& + \Big(1+a+b+rab+(2+(a+b)(1+r)+2abr)c(\xi+\epsilon)+r(1+a+b+ab)c^2(\xi+\epsilon)^2\Big)e_l,
\end{aligned}$$

which together with $e_0 = 0$ implies that

$$\|\mathcal{T} - \widehat{\mathcal{T}}\|_F = \frac{\alpha_2(a,b,c,\epsilon,\xi,r)^{\lceil \log_2 N\rceil} - 1}{\alpha_2(a,b,c,\epsilon,\xi,r) - 1}\beta_2(a,b,c,\epsilon,\xi), \tag{44}$$

where

$$\begin{aligned}
\alpha_2(a,b,c,\epsilon,\xi,r) =\ & 1+a+b+rab+(2+a+b+ar+br+2abr)c(\xi+\epsilon) \\
& + r(1+a+b+ab)c^2(\xi+\epsilon)^2, \\
\beta_2(a,b,c,\epsilon,\xi) =\ & \epsilon + (a+b+4ab)(\xi+\epsilon) + (2a+2b+3ab+1)c(\xi+\epsilon)^2 \\
& + (1+a+b+ab)c^2(\xi+\epsilon)^3.
\end{aligned}$$

■