Low Rank Approximation
Lecture 6

Daniel Kressner
Chair for Numerical Algorithms and HPC
Institute of Mathematics, EPFL
daniel.kressner@epfl.ch

Tensor Train (TT) decomposition

A tensor $\mathcal{X}$ is in TT decomposition if it can be written as

$$\mathcal{X}(i_1, \ldots, i_d) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_{d-1}=1}^{r_{d-1}} \mathcal{U}_1(1, i_1, k_1)\mathcal{U}_2(k_1, i_2, k_2) \cdots \mathcal{U}_d(k_{d-1}, i_d, 1).$$

- Smallest possible tuple $(r_1, \ldots, r_{d-1})$ is called TT rank of $\mathcal{X}$.
- $\mathcal{U}_\mu \in \mathbb{R}^{r_{\mu-1} \times n_\mu \times r_\mu}$ (formally set $r_0 = r_d = 1$) are called TT cores for $\mu = 1, \ldots, d$.
- If TT ranks are not large $\rightsquigarrow$ high compression ratio as d grows.
- TT decomposition multilinear wrt cores.
- TT decomposition connects to
  - matrix products $\rightsquigarrow$ Matrix Product States (MPS) in physics (see [Grasedyck/Kressner/Tobler'2013] for references)
  - simultaneous matrix factorizations $\rightsquigarrow$ SVD-based compression
  - contractions and tensor network diagrams $\rightsquigarrow$ design of efficient contraction-based algorithms

TT decomposition and matrix products

$$\mathcal{X}(i_1, \dots, i_d) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_{d-1}=1}^{r_{d-1}} \mathcal{U}_1(1, i_1, k_1)\mathcal{U}_2(k_1, i_2, k_2) \cdots \mathcal{U}_d(k_{d-1}, i_d, 1).$$

Let $U_\mu(i_\mu)$ be i_μth slice of μth core: $U_\mu(i_\mu) := \mathcal{U}_\mu(:, i_\mu, :) \in \mathbb{R}^{r_{\mu-1} \times r_\mu}$. Then

$$\mathcal{X}(i_1, i_2, \dots, i_d) = U_1(i_1)U_2(i_2) \cdots U_d(i_d).$$

EFY. Given two tensors $\mathcal{X}$ and $\mathcal{Y}$ in TT decomposition, derive a TT decomposition for $\mathcal{X} + \mathcal{Y}$.

EFY. Show that a tensor of tensor rank R has TT rank at most $(R, \dots, R)$ by converting its CP decomposition into a TT decomposition.

EFY. Given a tensor in TT decomposition, derive upper bounds for its multilinear rank.

Remark: Error analysis of matrix multiplication [Higham'2002] shows that TT decomposition may suffer from numerical instabilities if

$$\|U_1(i_1)\|_2 \|U_2(i_2)\|_2 \cdots \|U_d(i_d)\|_2 \gg |\mathcal{X}(i_1, i_2, \dots, i_d)|.$$

See [Bachmayr/Kazeev: arXiv:1802.09062] for more details.

TT decomposition and matrix factorizations

$$\mathcal{X}(i_1,\dots,i_d) = \sum_{k_1,k_2,\dots,k_{d-1}} \mathcal{U}_1(1,i_1,k_1)\mathcal{U}_2(k_1,i_2,k_2)\cdots\mathcal{U}_d(k_{d-1},i_d,1).$$

For any $1 \le \mu \le d-1$ group first μ factors and last $d-\mu$ factors together:

$$\begin{aligned} &\mathcal{X}(i_1,\dots,i_\mu,i_{\mu+1},\dots i_d) \\ = &\sum_{k_\mu=1}^{r_\mu} \Big(\sum_{k_1,\dots,k_{\mu-1}} \mathcal{U}_1(1,i_1,k_1)\cdots\mathcal{U}_\mu(k_{\mu-1},i_\mu,k_\mu)\Big) \\ &\cdot \Big(\sum_{k_{\mu+1},\dots,k_{d-1}} \mathcal{U}_{\mu+1}(k_\mu,i_{\mu+1},k_{\mu+1})\cdots\mathcal{U}_d(k_{d-1},i_d,1)\Big) \end{aligned}$$

This can be interpreted as a matrix-matrix product of two (large) matrices!

TT decomposition and matrix factorizations

The μth unfolding of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is obtained by arranging the entries in a matrix

$$X^{<\mu>} \in \mathbb{R}^{(n_1 n_2 \cdots n_\mu) \times (n_{\mu+1} \cdots n_d)}$$

where the corresponding index map is given by

$$\iota : \mathbb{R}^{n_1 \times \cdots \times n_d} \to \mathbb{R}^{n_1 \cdots n_\mu} \times \mathbb{R}^{n_{\mu+1} \cdots n_d}, \quad \iota(i_1, \ldots, i_d) = (i_{\text{row}}, i_{\text{col}}),$$

$$i_{\text{row}} = 1 + \sum_{\nu=1}^{\mu} (i_\nu - 1) \prod_{\tau=1}^{\nu-1} n_\tau, \quad i_{\text{col}} = 1 + \sum_{\nu=\mu+1}^{d} (i_\nu - 1) \prod_{\tau=\mu+1}^{\nu-1} n_\tau.$$

TT decomposition and matrix factorizations

Define interface matrices

$$X_{\leq \mu} \in \mathbb{R}^{n_1 n_2 \cdots n_\mu \times r_\mu}, \quad X_{\geq \mu+1} \in \mathbb{R}^{n_{\mu+1} n_{\mu+2} \cdots n_d \times r_\mu}$$

as

$$X_{\leq \mu}(i_{\text{row}}, j) = \sum_{k_1, \ldots, k_{\mu-1}} \mathcal{U}_1(1, i_1, k_1) \cdots \mathcal{U}_{\mu-1}(k_{\mu-2}, i_{\mu-1}, k_{\mu-1}) \mathcal{U}_\mu(k_{\mu-1}, i_\mu, j)$$

$$X_{\geq \mu+1}(i_{\text{col}}, j) = \sum_{k_{\mu+1}, \ldots, k_{d-1}} \mathcal{U}_{\mu+1}(j, i_{\mu+1}, k_{\mu+1}) \mathcal{U}_{\mu+2}(k_{\mu+1}, i_{\mu+2}, k_{\mu+2}) \cdots \mathcal{U}_d(k_{d-1}, i_d, 1)$$

Matrix factorizations

$$X^{<\mu>} = X_{\leq \mu} X_{\geq \mu+1}^T, \quad \mu = 1, \ldots, d-1.$$

TT decomposition and matrix factorizations

Important: These matrix factorizations are nested!

$$X_{\leq\mu} = (I_{n_\mu} \otimes X_{\leq\mu-1})U^{\mathsf{L}}_\mu, \quad \text{and} \quad X^T_{\geq\mu} = U^{\mathsf{R}}_\mu (X^T_{\geq\mu+1} \otimes I_{n_\mu}),$$

where

$$U^{\mathsf{L}}_\mu = U^{<2>}_\mu, \qquad U^{\mathsf{R}}_\mu = U^{(1)}_\mu = U^{<1>}_\mu.$$

The relations $X_{\leq 1} = U_1 \equiv \mathcal{U}_1$ and

$$X_{\leq\mu} = (I_{n_\mu} \otimes X_{\leq\mu-1})U^{\mathsf{L}}_\mu, \quad \mu = 2, \ldots, d,$$

fully characterize the TT decomposition:

$$\begin{aligned}
\mathrm{vec}(\mathcal{X}) &= X_{\leq d} \\
&= (I \otimes X_{\leq d-1})U^{\mathsf{L}}_d = (I \otimes X_{\leq d-1})\,\mathrm{vec}(\mathcal{U}_d) \\
&= (I \otimes I \otimes X_{\leq d-2})(I \otimes U^{\mathsf{L}}_{d-1})\,\mathrm{vec}(\mathcal{U}_d) \\
&\vdots \\
&= (I \otimes \cdots \otimes I \otimes U^{\mathsf{L}}_1) \cdots (I \otimes U^{\mathsf{L}}_{d-1})\,\mathrm{vec}(\mathcal{U}_d)
\end{aligned}$$

EFY. Perform an analogous calculation for $X^T_{\geq\mu}$, that is, resolve the recursion $X^T_{\geq\mu} = U^{\mathsf{R}}_\mu (X^T_{\geq\mu+1} \otimes I_{n_\mu})$.

TT decomposition and matrix factorizations

Lemma

The TT rank of a tensor is given by

$$\big(\operatorname{rank} X^{<1>}, \ldots, \operatorname{rank} X^{<d-1>}\big)$$

Proof. Because of the connection to matrix factorizations, the TT rank cannot be smaller than $\big(\operatorname{rank} X^{<1>}, \ldots, \operatorname{rank} X^{<d-1>}\big)$. We need to exclude that it can be larger. For this purpose, we construct a TT decomposition with

$$(r_1, \ldots, r_{d-1}) := \big(\operatorname{rank} X^{<1>}, \ldots, \operatorname{rank} X^{<d-1>}\big).$$

Step 1: Factorize

$$X^{<1>} = U_1 \tilde{X}^{<1>}, \quad U_1 \in \mathbb{R}^{n_1 \times r_1}, \quad \tilde{X}^{<1>} \in \mathbb{R}^{r_1 \times n_2 \cdots n_d},$$

and hence

$$\tilde{X}^{<1>} = U_1^{\dagger} X^{<1>}, \quad U_1^{\dagger} = (U_1^T U_1)^{-1} U_1^T$$

In terms of tensors: $\mathcal{X} = U_1 \circ_1 \tilde{\mathcal{X}}$.

$\mathcal{U}_1 \equiv U_1$ is the first TT core (and $X_{\geq 1}^T := \tilde{X}^{<1>}$).

Relation for second unfolding via Kronecker product:

$$X^{<2>} = (I_{n_2} \otimes U_1)\tilde{X}^{<2>}.$$

Together with full column rank of U_1, this implies

$$\operatorname{rank}(\tilde{X}^{<2>}) = \operatorname{rank}(X^{<2>}) = r_2.$$

Step 2: Factorize

$$\tilde{X}^{<2>} = U_2^{\mathsf{L}}\hat{X}^{<1>}, \quad U_2^{\mathsf{L}} \in \mathbb{R}^{r_1 n_2 \times r_2}, \quad \hat{X}^{<1>} \in \mathbb{R}^{r_2 \times n_3 \cdots n_d},$$

U_2^{L} gives the second TT core $\mathcal{U}_2 \in \mathbb{R}^{r_1 \times n_2 \times r_2}$ and $X_{\geq 2}^T := \hat{X}^{<1>}$.
Relation for third unfolding via Kronecker product:

$$X^{<3>} = (I_{n_3} \otimes I_{n_2} \otimes U_1)\tilde{X}^{<3>} = (I_{n_3} \otimes I_{n_2} \otimes U_1)(I_{n_3} \otimes U_2^{\mathsf{L}})\hat{X}^{<2>}$$

Together with full column ranks of U_1 and U_2^{L}, this implies

$$\operatorname{rank}(\hat{X}^{<2>}) = \operatorname{rank}(X^{<3>}) = r_3.$$

Continuing in this manner gives cores $\mathcal{U}_\mu \in \mathbb{R}^{r_{\mu-1} \times n_\mu \times r_\mu}$ such that

$$\operatorname{vec}(\mathcal{X}) = (I \otimes \cdots \otimes I \otimes U_1)\cdots(I \otimes U_{d-1}^{\mathsf{L}})\operatorname{vec}(\mathcal{U}_d)$$

This defines a TT decomposition.

Truncation in TT format

Proof of Lemma can be turned into practical algorithm (TT-SVD by [Oseledets'2011]) for approximating a given tensor $\mathcal{X}$ in TT format:

Input: $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, target TT rank $(r_1, \ldots, r_{d-1})$.
Output: TT cores $\mathcal{U}_\mu \in \mathbb{R}^{r_{\mu-1} \times n_\mu \times r_\mu}$ that define a TT decomposition approximating $\mathcal{X}$.

```
Set r_0 = r_d = 1. (and formally add leading singleton dimension to X ∈ R^{1×n_1×···×n_d}).
for μ = 1, ..., d − 1 do
    Reshape X into X^{<2>} ∈ R^{r_{μ−1} n_μ × n_{μ+1}···n_d}.
    Compute rank-r_μ approximation X^{<2>} ≈ UΣV^T (e.g., via SVD)
    Reshape U into U_μ ∈ R^{r_{μ−1}×n_μ×r_μ}.
    Update X via X^{<2>} ← U^T X^{<2>} = ΣV^T.
end for
Set U_d = X.
```

Truncation in TT format

Theorem

Let $\mathcal{X}_{\text{SVD}}$ denote the tensor in TT decomposition obtained from TT-SVD. Then

$$\|\mathcal{X} - \mathcal{X}_{\text{SVD}}\| \leq \sqrt{\varepsilon_1^2 + \cdots + \varepsilon_d^2},$$

where

$$\varepsilon_\mu^2 = \|X^{<\mu>} - \mathcal{T}_{r_\mu}(X^{<\mu>})\|_F^2 = \sigma_{r_\mu+1}(X^{<\mu>})^2 + \cdots .$$

Proof. After μ steps of the algorithm we have the following situation:

- Core tensors $\mathcal{U}_1, \ldots, \mathcal{U}_\mu$ have been computed, defining $X_{\leq\mu}$.
- Remaining tensor has size $r_\mu \times n_{\mu+1} \times \cdots \times n_d$.

Reshape remaining tensor into matrix $Y_{\geq\mu}^T \in \mathbb{R}^{r_\mu \times n_{\mu+1} \cdots n_d}$. Will prove relations

$$X_{\leq\mu}^T X_{\leq\mu} = I, \qquad Y_{\geq\mu+1}^T = X_{\leq\mu}^T X^{<\mu>},$$

and

$$\|X^{<\mu>} - X_{\leq\mu} Y_{\geq\mu+1}^T\|_F \leq \sqrt{\varepsilon_1^2 + \cdots + \varepsilon_\mu^2} \qquad (1)$$

for $\mu = 1, \ldots, d-1$ by induction. For $\mu = d-1$, this shows the theorem.

Line 3 in the μth step of the algorithm proceeds by reshaping the remaining tensor from step $\mu - 1$ (corresponding to $Y_{\geq\mu}^T$) into an array of size $Y^{<2>} \in \mathbb{R}^{r_{\mu-1}n_\mu \times n_{\mu+1}\cdots n_d}$. By induction assumption,

$$Y_{\geq\mu}^T = X_{\leq\mu-1}^T X^{<\mu-1>} \quad \Rightarrow \quad Y^{<2>} = (I_{n_\mu} \otimes X_{\leq\mu-1})^T X^{<\mu>}. \qquad (2)$$

The matrix $U_\mu^{\mathsf{L}} \equiv U$ computed in Line 4 contains left singular vectors of $Y^{<2>}$. In particular, $(U_\mu^{\mathsf{L}})^T U_\mu^{\mathsf{L}} = I$. Together with the induction assumption and the relation

$$X_{\leq\mu} = (I_{n_\mu} \otimes X_{\leq\mu-1}) U_\mu^{\mathsf{L}},$$

this implies $X_{\leq\mu}^T X_{\leq\mu} = I$ and $Y_{\geq\mu+1}^T = X_{\leq\mu}^T X^{<\mu>}$. Moreover,

$$\begin{aligned} \|(I - U_\mu^{\mathsf{L}}(U_\mu^{\mathsf{L}})^T) Y^{<2>}\|_F &= \|Y^{<2>} - \mathcal{T}_{r_\mu}(Y^{<2>})\|_F \\ &\leq \|X^{<\mu>} - \mathcal{T}_{r_\mu}(X^{<\mu>})\|_F = \varepsilon_\mu. \end{aligned}$$

Finally, we obtain:

$$
\begin{aligned}
& \|X^{<\mu>} - X_{\leq\mu} X_{\leq\mu}^T X^{<\mu>}\|_F^2 \\
= \;& \|X^{<\mu>} - (I \otimes X_{\leq\mu-1}) U_\mu^{\mathsf{L}} (U_\mu^{\mathsf{L}})^T (I \otimes X_{\leq\mu-1})^T X^{<\mu>}\|_F^2 \\
= \;& \|X^{<\mu>} - (I \otimes X_{\leq\mu-1})(I \otimes X_{\leq\mu-1})^T X^{<\mu>}\|_F^2 \\
+ \;& \|(I \otimes X_{\leq\mu-1})(I \otimes X_{\leq\mu-1})^T X^{<\mu>} \\
& -(I \otimes X_{\leq\mu-1}) U_\mu^{\mathsf{L}} (U_\mu^{\mathsf{L}})^T (I \otimes X_{\leq\mu-1})^T X^{<\mu>}\|_F^2 \\
= \;& \|X^{<\mu-1>} - X_{\leq\mu-1} X_{\leq\mu-1}^T X^{<\mu-1>}\|_F^2 + \|(I - U_\mu^{\mathsf{L}} (U_\mu^{\mathsf{L}})^T) Y^{<2>}\|_F^2 \\
\leq \;& \varepsilon_1^2 + \cdots + \varepsilon_{\mu-1}^2 + \varepsilon_\mu^2
\end{aligned}
$$

This completes the proof.

Truncation in TT format

Consequence of the theorem:

Corollary

Let $\mathcal{X}_{\text{best}}$ denote the best approximation of $\mathcal{X}$ with TT rank $(r_1, \ldots, r_{d-1})$. Then

$$\|\mathcal{X} - \mathcal{X}_{\text{SVD}}\| \leq \sqrt{d-1}\|\mathcal{X} - \mathcal{X}_{\text{best}}\|.$$

Proof.

Since $X_{\text{best}}^{<\mu>}$ has rank r_μ we have

$$\begin{aligned}
\|\mathcal{X} - \mathcal{X}_{\text{best}}\| &\geq \max_{\mu=1,\ldots,r-1} \|X^{<\mu>} - X_{\text{best}}^{<\mu>}\|_F \\
&= \max_{\mu=1,\ldots,r-1} \varepsilon_\mu \geq \frac{1}{\sqrt{d-1}}\sqrt{\varepsilon_1^2 + \cdots + \varepsilon_d^2} \\
&\geq \frac{1}{\sqrt{d-1}}\|\mathcal{X} - \mathcal{X}_{\text{SVD}}\|
\end{aligned}$$

□

EFY. Develop a variant of TT-SVD, which for a prescribed accuracy $\varepsilon > 0$ determines the TT ranks adaptively such that $\|\mathcal{X} - \mathcal{X}_{\text{SVD}}\| \leq \varepsilon$.

Orthogonal TT decompositions

The TT decomposition constructed by TT-SVD satisfies orthogonality relations

$$(U_\mu^{\mathrm{L}})^T U_\mu^{\mathrm{L}} = I_{r_\mu}, \quad X_{\le\mu}^T X_{\le\mu} = I_{r_\mu} \quad \text{for} \quad \mu = 1, \dots, d-1.$$

Such a TT decomposition is called left-orthogonal.

EFY. Show that $\|\mathcal{X}\| = \|\mathcal{U}_d\|$ holds for a left-orthogonal TT decomposition. What can you say about the singular values of $X^{<\mu-1>}$?

A TT decomposition is called right-orthogonal if

$$U_\mu^{\mathrm{R}} (U_\mu^{\mathrm{R}})^T = I_{r_\mu}, \quad X_{\ge\mu} X_{\ge\mu}^T = I_{r_\mu} \quad \text{for} \quad \mu = 2, \dots, d.$$

EFY. Show that $\|\mathcal{X}\| = \|\mathcal{U}_1\|$ holds for a right-orthogonal TT decomposition. What can you say about the singular values of $X^{<1>}$?

Benefits of orthogonal decompositions

- Numerical stability (see, e.g., Section 3.7 of [Higham'2002] on stability of matrix products).
- Acceleration of operations.

Orthogonal TT decompositions

Given TT decomposition with cores $\mathcal{U}_1, \ldots, \mathcal{U}_d$, can always compute left-/right-orthogonal TT decomposition for the same tensor.

Algorithm to obtain left-orthogonal TT decomposition:

for $\mu = 1, 2, \ldots, d-1$ **do**
  Compute QR decomposition $U_\mu^{\mathsf{L}} = QR$.
  Set $U_\mu^{\mathsf{L}} \leftarrow Q$.
  Update $U_{\mu+1} \leftarrow R \circ_1 U_{\mu+1}$.
end for

Algorithm to obtain right-orthogonal TT decomposition:

for $\mu = d, d-1, \ldots, 2$ **do**
  Compute QR decomposition $(U_\mu^{\mathsf{R}})^T = QR$.
  Set $U_\mu^{\mathsf{R}} \leftarrow Q^T$.
  Update $U_{\mu-1} \leftarrow R \circ_3 U_{\mu-1}$.
end for

Recompression of TT decomposition

Task: Approximate TT decomposition with TT cores $\mathcal{U}_1, \ldots, \mathcal{U}_d$ having relatively large TT ranks $(R_1, \ldots, R_d)$ by TT decomposition with TT cores $\mathcal{V}_1, \ldots, \mathcal{V}_d$ having smaller TT ranks $(r_1, \ldots, r_d)$.

TT-SVD starts with SVD of

$$X^{<1>} = U_1^L X_{\geq 2}^T.$$

Observation: If $\mathcal{X}$ is right-orthogonal then $X_{\geq 2}$ has orthonormal columns and only need to compute SVD of U_1^L!

Compute rank-r_1 approximation (e.g., via truncated SVD):

$$U_1^L \approx V_1^L \Sigma W^T$$

This defines the first TT core $\mathcal{V}_1$ of the compressed tensor. Move rest to second core:

$$\mathcal{U}_2 \leftarrow \Sigma W^T \circ_1 \mathcal{U}_2.$$

$X_{\geq 3}$ remains the same and thus still has orthonormal columns!

Recompression of TT decomposition

Input: $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ in TT decomposition with TT cores $\mathcal{U}_\mu \in \mathbb{R}^{R_{\mu-1} \times n_\mu \times R_\mu}$, smaller target TT rank $(r_1, \dots, r_{d-1})$.
Output: TT cores $\mathcal{V}_\mu \in \mathbb{R}^{r_{\mu-1} \times n_\mu \times r_\mu}$ that define a TT decomposition approximating $\mathcal{X}$.

Set $r_0 = r_d = 1$.
Right-orthogonalize TT cores $\mathcal{U}_\mu$.
for $\mu = 1, \dots, d-1$ **do**
    Compute rank-r_μ approximation $U_\mu^L \approx V \Sigma W^T$ (e.g., via SVD)
    Reshape V into $\mathcal{V}_\mu \in \mathbb{R}^{r_{\mu-1} \times n_\mu \times r_\mu}$.
    Update $\mathcal{U}_{\mu+1} \leftarrow \Sigma W^T \circ_1 \mathcal{U}_{\mu+1}$.
end for
Set $\mathcal{V}_d = \mathcal{U}_d$.

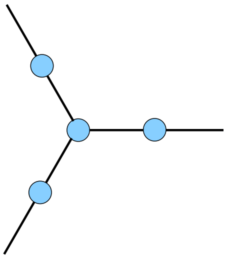
- Complexity: $\mathcal{O}(dnR^3)$.
- Mathematical equivalence to full TT-SVD $\rightsquigarrow$ error bounds carry over.

Tensor network diagrams

- Introduced by Roger Penrose.
- Heavily used in quantum mechanics (spin networks).
- Useful to gain intuition and guide design of algorithms.
- This is the matrix product $C = AB$:

$$C_{ij} = \sum_{k=1}^{r} A_{ik} B_{kj}$$

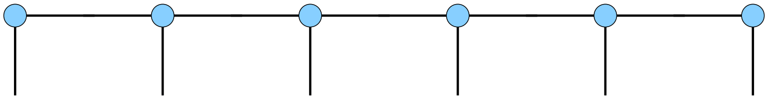
Tensor of order 3 in Tucker decomposition



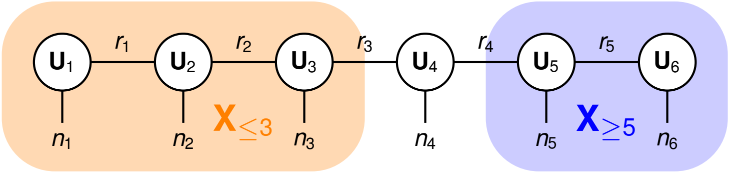
$$\mathcal{X}_{ijk} = \sum_{\ell_1=1}^{r_1} \sum_{\ell_2=1}^{r_2} \sum_{\ell_3=1}^{r_3} \mathcal{C}_{\ell_1 \ell_2 \ell_3} U_{i\ell_1} V_{j\ell_2} W_{k\ell_3}$$

- $r_1 \times r_2 \times r_3$ core tensor $\mathcal{C}$
- $n_1 \times r_1$ matrix U spans first mode
- $n_2 \times r_2$ matrix V spans second mode
- $n_3 \times r_3$ matrix W spans third mode.

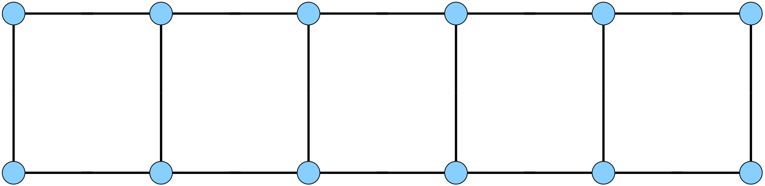
Tensor of order 6 in TT decomposition



- $\mathcal{X}$ implicitly represented by four $r \times n \times r$ tensors and two $n \times r$ matrices
- More detailed picture:



Inner product of two tensors in TT decomposition



- Carrying out contractions requires $O(dnr^4)$ instead of $O(n^d)$ operations for tensors of order d.

EFY. Develop an efficient method for computing the mean,

$$\overline{\mathcal{X}} = \frac{1}{n_1 \cdots n_d} \sum_{i_1,\ldots,i_d} \mathcal{X}(i_1,\ldots,i_d),$$

of a tensor in TT decomposition.

Linear operators in TT decomposition

Consider linear operator

$$\mathcal{L} : \mathbb{R}^{n_1 \times \cdots \times n_d} \to \mathbb{R}^{n_1 \times \cdots \times n_d}$$

and corresponding matrix representation $\mathcal{L}((i_1, \ldots, i_d), (j_1, \ldots, j_d))$.
$\mathcal{L}$ is TT operator if it takes the form

$$\mathcal{L}((i_1, \ldots, i_d), (j_1, \ldots, j_d)) =$$
$$\sum_{k_1=1}^{R_1} \cdots \sum_{k_{d-1}=1}^{R_{d-1}} \mathcal{A}_1(1, i_1, j_1, k_1)\mathcal{A}_2(k_1, i_2, j_2, k_2) \cdots \mathcal{A}_d(k_{d-1}, i_d, j_d, 1)$$

with $R_{\mu-1} \times n_\mu \times n_\mu \times R_\mu$ tensors $\mathcal{A}_\mu$.
Operator TT rank of $\mathcal{L} =$ Smallest possible values of $(R_1, \ldots, R_{\mu-1})$.
Matrix Product Operator (MPO):

$$\mathcal{L}((i_1, \ldots, i_d), (j_1, \ldots, j_d)) = A_1(i_1, j_1)A_2(i_2, j_2) \cdots A_d(i_d, j_d),$$

with $A_\mu(i_\mu, j_\mu) = \mathcal{A}_\mu(:, i_\mu, j_\mu, :)$.

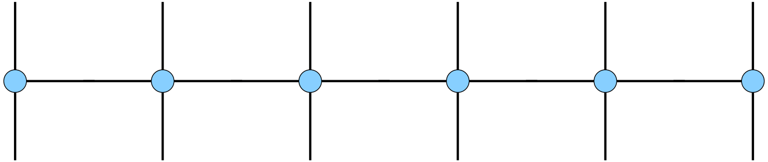
Linear operators in TT decomposition

Alternative view: Reinterpret matrix representation of $\mathcal{A}$ as tensor

$$\text{ten}(\mathcal{A}) \in \mathbb{R}^{n_1^2 \times \cdots \times n_d^2},$$

by merging indices (i_μ, j_μ) applying TT format, and separating indices (i_μ, j_μ).
In practice: Perfect shuffle permutation of modes.

Tensor network diagram for operator TT decomposition:



EFY. Show that the identity matrix has operator TT ranks 1.

Linear operators in TT decomposition

Example: Discrete Laplace-like operator takes the form

$$\mathcal{L} : \mathcal{X} \mapsto L_1 \circ_1 \mathcal{X} + \cdots + L_d \circ_d \mathcal{X}$$

for matrices $L_\mu \in \mathbb{R}^{n_\mu \times n_\mu}$. Matrix representation

$$\mathcal{L} = I_{n_d} \otimes \cdots \otimes I_{n_2} \otimes L_1 + \cdots + L_d \otimes I_{n_{d-1}} \otimes \cdots \otimes I_{n_1}$$

TT representation with cores

$$\begin{aligned} A_1(i_1, j_1) &= \begin{pmatrix} L_1(i_1, j_1) & I_{n_1}(i_1, j_1) \end{pmatrix} \\ A_\mu(i_\mu, j_\mu) &= \begin{pmatrix} I_{n_\mu}(i_\mu, j_\mu) & 0 \\ L_\mu(i_\mu, j_\mu) & I_{n_\mu}(i_\mu, j_\mu) \end{pmatrix} \\ A_d(i_d, j_d) &= \begin{pmatrix} I_{n_d}(i_d, j_d) \\ L_d(i_d, j_d) \end{pmatrix} \end{aligned}$$

Multiplication with linear operators in TT decomposition

Consider matrix-vector product $\tilde{x} = \mathcal{L}x$ or, equivalently, $\tilde{\mathcal{X}} = \mathcal{L}(\mathcal{X})$.

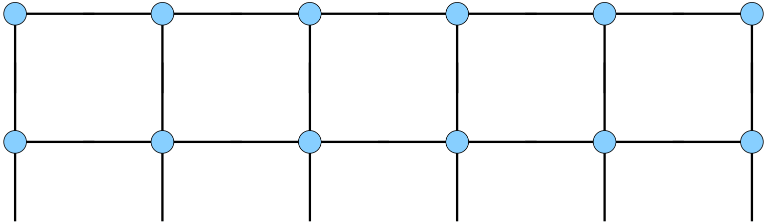
Assume that $\mathcal{L}$ and $\mathcal{U}$ are in TT decomposition $\rightsquigarrow$

$$
\begin{aligned}
& \tilde{\mathcal{X}}(i_1, \ldots, i_d) \\
= & \sum_{j_1, \ldots, j_d} \mathcal{L}((i_1, \ldots, i_d), (j_1, \ldots, j_d)) \mathcal{X}(j_1, \ldots, j_d) \\
= & \sum_{j_1, \ldots, j_d} A_1(i_1, j_1) \cdots A_d(i_d, j_d) U_1(j_1) \cdots U_d(j_d) \\
= & \sum_{j_1, \ldots, j_d} (A_1(i_1, j_1) \otimes U_1(j_1)) \cdots (A_d(i_d, j_d) \otimes U_d(j_d)) \\
= & \Big[\sum_{j_1} A_1(i_1, j_1) \otimes U_1(j_1) \Big] \cdots \Big[\sum_{j_d} A_d(i_d, j_d) \otimes U_d(j_d) \Big] \\
=: & \tilde{U}_1(i_1) \cdots \tilde{U}_d(i_d),
\end{aligned}
$$

that is, TT ranks of $\tilde{\mathcal{X}}$ are bounded by TT ranks of $\mathcal{X}$ multiplied with TT operator ranks of $\mathcal{L}$.

Multiplication with linear operators in TT decomposition

Illustration by tensor network diagrams:



- Carrying out contractions requires $O(dnr^4)$ operations for tensors of order d.

TT decomposition – Summary of operations

Easy:

- (partial) contractions
- multiplication with TT operators
- recompression/truncation

Medium:

- entrywise products

Hard:

- almost everything else

Software:

- TT toolox (Matlab, Python), TTeMPS (Matlab), ...