Structured Matrix Computations from Structured Tensors

Lecture 3. The Tucker and Tensor Train Decompositions

Charles F. Van Loan

Cornell University

CIME-EMS Summer School
June 22-26, 2015
Cetraro, Italy

The Setting

Good News/Bad News

The singular values of a general matrix and the eigenvalues of a symmetric matrix have variational definitions and these ideas can be extended to tensors.

However, these ideas are not strong enough to put together a tensor decomposition like the SVD:

$$A = U\Sigma V^T = \sum_{k=1}^{\text{rank}(A)} \sigma_k u_k v_k^T$$

Why Do We Like Matrix Factorizations?

The Factorization Paradigm in Matrix Computations

Typical...

Convert the given problem into an equivalent easy-to-solve problem by using the "right" matrix decomposition.

$$PA = LU, \quad Ly = Pb, \quad Ux = y \quad \Longrightarrow \qquad Ax = b$$

Also Typical...

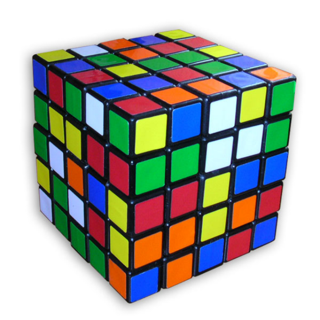
Uncover hidden relationships by computing the "right" decomposition of the data matrix.

$$A \ = \ U\Sigma V^T \quad \Longrightarrow \qquad A \ \approx \sum_{i=1}^{\tilde{r}} \sigma_i u_i v_i^T$$

$A = U\Sigma V^T \quad PA = LU \quad A = QR \quad A = GG^T \quad PAP^T = LDL^T \quad Q^TAQ = D$
$X^{-1}AX = J \quad U^TAU = T \quad AP = QR \quad A = ULV^T \quad PAQ^T = LU \quad A = U\Sigma V^T$
$PA = LU \quad A = QR \quad A = GG^T \quad PAP^T = LDL^T \quad Q^TAQ = D \quad X^{-1}AX = J$
$U^TAU = T \quad AP = QR \quad A = ULV^T \quad PAQ^T = LU \quad A = U\Sigma V^T \quad PA = LU$
$A = QR \quad A = GG^T \quad PAP^T = LDL^T \quad Q^TAQ = D \quad X^{-1}AX = J \quad U^TAU = T$
$AP = QR \quad A = ULV^T \quad PAQ^T = LU \quad A = U\Sigma V^T \quad PA = LU \quad A = QR$
$A = GG^T \quad PAP^T = LDL^T \quad Q^TAQ = D \quad X^{-1}AX = J \quad U^TAU = T$
$A = ULV^T \quad PAQ^T = LU \quad A = U\Sigma V^T \quad PA = LU \quad A = QR \quad A = GG^T$
$PAP^T = LDL^T \quad Q^TAQ = D \quad X^{-1}AX = J \quad U^TAU = T \quad AP = QR$
$A = ULV^T \quad PAQ^T = LU \quad A = U\Sigma V^T \quad PA = LU \quad A = QR \quad A = GG^T$
$PAP^T = LDL^T \quad Q^TAQ = D \quad X^{-1}AX = J \quad AP = QR \quad A = ULV^T$
$PAQ^T = LU \quad A = U\Sigma V^T \quad PA = LU \quad A = QR \quad A = GG^T \quad PAP^T = LDL^T$
$Q^TAQ = D \quad X^{-1}AX = J \quad U^TAU = T \quad AP = QR \quad A = ULV^T \quad PAQ^T = LU$
$A = U\Sigma V^T \quad PA = LU \quad A = QR \quad A = GG^T \quad PAP^T = LDL^T \quad Q^TAQ = D$
$X^{-1}AX = J \quad U^TAU = T \quad AP = QR \quad A = ULV^T \quad PAQ^T = LU \quad A = U\Sigma V^T$
$PA = LU \quad A = QR \quad PAP^T = LDL^T \quad Q^TAQ = D \quad X^{-1}AX = J \quad U^TAU = T$
$AP = QR \quad A = ULV^T \quad PAQ^T = LU \quad A = U\Sigma V^T \quad PA = LU \quad A = QR$

It's a Language

Anticipating the Same Thing for Tensors



$$= \sigma_1\, w_1 \circ v_1 \circ u_1 \;+\; \sigma_2\, w_2 \circ v_2 \circ u_2 + \;\ldots$$

Anticipating the Same Thing for Tensors

Question 1

Can we solve tensor problems by converting them to (approximately) equivalent easy-to-solve problems using a tensor decomposition?

Question 2

Can we uncover hidden patterns in tensor data by computing an appropriate tensor decomposition?

These questions will be addessed in this lecture and the next.

What is this Lecture About?

Outline

- The Tucker Product Representation and Its Properties
- The Mode-k Product and the Tucker Product
- The Higher-Order SVD of a tensor
- An Alternating Least Squares Framework for Reduced-Rank Tucker Approximation
- The Tensor Train Representation

The Tucker Product Representation

Tucker Product: The Matrix Case

Definition

The Tucker product between a matrix

$$S : r_1 \times r_2$$

and matrices

$$\begin{aligned} U_1 &: n_1 \times r_1 \\ U_2 &: n_2 \times r_2 \end{aligned}$$

is the $n_1 \times n_2$ matrix defined by

$$A(i_1, i_2) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \mathcal{S}(j_1, j_2) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2)$$

Tucker Product: The Matrix Case

It is Actually Just the Product of Three Matrices

$$
\begin{aligned}
A(i_1, i_2) &= \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \mathcal{S}(j_1, j_2) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2) \\
A &= U_1 S U_2^T
\end{aligned}
$$

Tucker Product: The Matrix Case

It is Actually the Sum of Rank-1 Matrices

$$A(i_1, i_2) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \mathcal{S}(j_1, j_2) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2)$$

$$A = U_1 S U_2^T$$

$$A = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \mathcal{S}(j_1, j_2) \cdot U_1(:, j_1) \cdot U_2(:, j_2)^T$$

Tucker Product: The Matrix Case

It is Actually the Sum of Kronecker Products Between Vectors

$$A(i_1, i_2) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \mathcal{S}(j_1, j_2) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2)$$

$$A = U_1 S U_2^T$$

$$A = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \mathcal{S}(j_1, j_2) \cdot U_1(:, j_1) \cdot U_2(:, j_2)^T$$

$$\text{vec}(A) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} S(j_1, j_2) \cdot U_2(:, j_2) \otimes U_1(:, j_2)$$

Tucker Product: The Matrix Case

It is Actually a Giant Matrix-Vector Product

$$A(i_1, i_2) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \mathcal{S}(j_1, j_2) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2)$$

$$A = U_1 S U_2^T$$

$$A = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \mathcal{S}(j_1, j_2) \cdot U_1(:, j_1) \cdot U_2(:, j_2)^T$$

$$\text{vec}(A) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} S(j_1, j_2) U_2(:, j_2) \otimes U_1(:, j_2)$$

$$\text{vec}(A) = (U_2 \otimes U_1) \cdot \text{vec}(S)$$

Tucker Product: The Tensor Case

Definition (Order-3)

The Tucker product between a tensor

$$\mathcal{S} : r_1 \times r_2 \times r_3$$

and matrices

$$\begin{aligned} U_1 &: n_1 \times r_1 \\ U_2 &: n_2 \times r_2 \\ U_3 &: n_3 \times r_3 \end{aligned}$$

is the $n_1 \times n_2 \times n_3$ tensor defined by

$$\mathcal{A}(i_1, i_2, i_3) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2) \cdot U_3(i_3, j_3)$$

Tucker Product: The Tensor Case

It is Actually the Sum of Rank-1 Tensors...

$$\mathcal{A}(i_1, i_2, i_3) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2) \cdot U_3(i_3, j_3)$$

$$\mathcal{A} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)$$

Tucker Product: The Tensor Case

It is Actually the Sum of Kronecker Products Between Vectors

$$\mathcal{A}(i_1, i_2, i_3) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2) \cdot U_3(i_3, j_3)$$

$$\mathcal{A} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)$$

$$\text{vec}(\mathcal{A}) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_3(:, j_3) \otimes U_2(:, j_2) \otimes U_1(:, j_1)$$

Tucker Product: The Tensor Case

It is Actually a Giant Matrix-Vector Product

$$\mathcal{A}(i_1, i_2, i_3) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2) \cdot U_3(i_3, j_3)$$

$$\mathcal{A} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)$$

$$\text{vec}(\mathcal{A}) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_3(:, j_3) \otimes U_2(:, j_2) \otimes U_1(:, j_1)$$

$$\text{vec}(\mathcal{A}) = (U_3 \otimes U_2 \otimes U_1) \cdot \text{vec}(\mathcal{S})$$

The Tucker Product

It is a "Representation"

$$\mathcal{A}(i_1, i_2, i_3) \;=\; \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2) \cdot U_3(i_3, j_3)$$

We are **representing** the tensor $\mathcal{A}$ in terms of the tensor $\mathcal{S}$ and the matrices U_1, U_2, and U_3.

Can we compute a Tucker Product representation that is especially illuminating or useful?

Computing the SVD of a Matrix

Have:

$$A = U_1 S U_2^T \qquad U_1,\ U_2 \text{ Orthogonal}$$

Improve:

$$A = (U_1 \Delta_1)({\Delta_1}^T S \Delta_2)(U_2 \Delta_2)^T$$

E.g., make S more diagonal by choosing clever orthogonal Δ_1 and Δ_2

Update:

$$S \leftarrow {\Delta_1}^T S \Delta_2 \qquad U_1 \leftarrow U_1 \Delta_1 \qquad U_2 \leftarrow U_2 \Delta_2$$

We would like to do the same thing for tensors, but what are the "update operations"?

The Mode-k Product

The Mode-k Product

Main Idea

Given $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, a mode k , and a matrix M, we apply M to every mode-k fiber.

Recall that

$$\mathcal{A}_{(2)} = \begin{bmatrix} a_{111} & a_{211} & a_{311} & a_{411} & a_{112} & a_{212} & a_{312} & a_{412} \\ a_{121} & a_{221} & a_{321} & a_{421} & a_{122} & a_{222} & a_{322} & a_{422} \\ a_{131} & a_{231} & a_{331} & a_{431} & a_{132} & a_{232} & a_{332} & a_{432} \end{bmatrix}$$

is the mode-2 unfolding of $\mathcal{A} \in \mathbb{R}^{4 \times 3 \times 2}$ and its columns are its mode-2 fibers

The Mode-k Product

A Mode-2 Example When $A \in \mathbb{R}^{4\times 3\times 2}$

$$\begin{bmatrix} b_{111} & b_{211} & b_{311} & b_{411} & b_{112} & b_{212} & b_{312} & b_{412} \\ b_{121} & b_{221} & b_{321} & b_{421} & b_{122} & b_{222} & b_{322} & b_{422} \\ b_{131} & b_{231} & b_{331} & b_{431} & b_{132} & b_{232} & b_{332} & b_{432} \\ b_{141} & b_{241} & b_{341} & b_{441} & b_{142} & b_{242} & b_{342} & b_{442} \\ b_{151} & b_{251} & b_{351} & b_{451} & b_{152} & b_{252} & b_{352} & b_{452} \end{bmatrix}$$

$$=$$

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \\ m_{41} & m_{42} & m_{43} \\ m_{51} & m_{52} & m_{53} \end{bmatrix} \begin{bmatrix} a_{111} & a_{211} & a_{311} & a_{411} & a_{112} & a_{212} & a_{312} & a_{412} \\ a_{121} & a_{221} & a_{321} & a_{421} & a_{122} & a_{222} & a_{322} & a_{422} \\ a_{131} & a_{231} & a_{331} & a_{431} & a_{132} & a_{232} & a_{332} & a_{432} \end{bmatrix}$$

Note: (1) $B \in \mathbb{R}^{4\times 5\times 2}$ and (2) $\mathcal{B}_{(2)} = M \cdot \mathcal{A}_{(2)}$.

The Mode-k Product: Definition

Mode-1

If $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $M \in \mathbb{R}^{n_1 \times n_1}$, then the mode-1 product

$$\mathcal{B} = \mathcal{A} \times_1 M \in \mathbb{R}^{n_1 \times n_2 \times n_3}$$

is defined by

$$\mathcal{B}(i_1, i_2, i_3) \;=\; \sum_{k=1}^{n_1} M(i_1, k)\mathcal{A}(k, i_2, i_3)$$

Two Equivalent Formulations...

$$\mathcal{B}_{(1)} \;=\; M \cdot \mathcal{A}_{(1)}$$

$$\text{vec}(\mathcal{B}) \;=\; (I_{n_3} \otimes I_{n_2} \otimes M)\text{vec}(\mathcal{A})$$

For now, assume M is square. Not necessary in general.

The Mode-k Product: Definition

Mode-2

If $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $M \in \mathbb{R}^{n_2 \times n_2}$, then the mode-2 product

$$\mathcal{B} = \mathcal{A} \times_2 M \in \mathbb{R}^{n_1 \times n_2 \times n_3}$$

is defined by

$$\mathcal{B}(i_1, i_2, i_3) \;=\; \sum_{k=1}^{n_2} M(i_2, k)\mathcal{A}(i_1, k, i_3)$$

Two Equivalent Formulations...

$$\mathcal{B}_{(2)} \;=\; M \cdot \mathcal{A}_{(2)}$$

$$\text{vec}(\mathcal{B}) \;=\; (I_{n_3} \otimes M \otimes I_{n_1})\text{vec}(\mathcal{A})$$

The Mode-k Product: Definition

Mode-3

If $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $M \in \mathbb{R}^{n_3 \times n_3}$, then the mode-3 product

$$\mathcal{B} = \mathcal{A} \times_3 M \in \mathbb{R}^{n_1 \times n_2 \times m_3}$$

is defined by

$$\mathcal{B}(i_1, i_2, i_3) \;=\; \sum_{k=1}^{n_3} M(i_3, k)\mathcal{A}(i_1, i_2, k)$$

Two Equivalent Formulations...

$$\mathcal{B}_{(3)} \;=\; M \cdot \mathcal{A}_{(3)}$$

$$\text{vec}(\mathcal{B}) \;=\; (M \otimes I_{n_2} \otimes I_{n_1})\text{vec}(\mathcal{A})$$

The Mode-k Product: Properties

Successive Products in the Same Mode

If $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $M_1, M_2 \in \mathbb{R}^{n_k \times n_k}$, then

$$(\mathcal{A} \times_k M_1) \times_k M_2 \;=\; \mathcal{A} \times_k (M_1 M_2).$$

Successive Products in Different Modes

If $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $M_k \in \mathbb{R}^{n_k \times n_k}$, $M_j \in \mathbb{R}^{n_j \times n_j}$, and $k \neq j$, then

$$(\mathcal{A} \times_k M_k) \times_j M_j \;=\; (\mathcal{A} \times_j M_j) \times_k M_k$$

The order is not important so we just write $\mathcal{A} \times_j M_j \times_k M_k$.

The Tucker Product

It is a Collection of Modal Products

The Tucker Product of the tensor

$$\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$$

with the matrices $U_1 \in \mathbb{R}^{n_1 \times r_1}$, $U_2 \in \mathbb{R}^{n_2 \times r_2}$, and $U_3 \in \mathbb{R}^{n_3 \times r_3}$ is given by

$$\begin{aligned} \mathcal{A}(i_1, i_2, i_3) &= \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2) \cdot U_3(i_3, j_3) \\ &= \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3 \end{aligned}$$

The Tucker Product Representation

A Simple but Important Result

If $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $U_1 \in \mathbb{R}^{n_1 \times n_1}$, $U_2 \in \mathbb{R}^{n_2 \times n_2}$, and $U_3 \in \mathbb{R}^{n_3 \times n_3}$ are nonsingular, then

$$\mathcal{A} \;=\; \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$$

where

$$\mathcal{S} \;=\; \mathcal{A} \times_1 U_1^{-1} \times_2 U_2^{-1} \times_3 U_3^{-1}.$$

We will refer to the U_k as the inverse factors and $\mathcal{S}$ as the core tensor.

The matrix version: $A \;=\; U_1(U_1^{-1} A U_2^{-1}) U_2 \;=\; U_1 S U_2$

Proof.

$$\begin{aligned}
\mathcal{A} &= \mathcal{A} \times_1 (U_1^{-1}U_1) \times_2 (U_2^{-1}U_2) \times_3 (U_3^{-1}U_3) \\
&= \left(\mathcal{A} \times_1 U_1^{-1} \times_2 U_2^{-1} \times_3 U_3^{-1}\right) \times_1 U_1 \times_2 U_2 \times_3 U_3 \\
&= \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3
\end{aligned}$$

An Orthogonal Tucker Product Representation

If the U's are Orthogonal

If $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $U_1 \in \mathbb{R}^{n_1 \times n_1}$, $U_2 \in \mathbb{R}^{n_2 \times n_2}$, and $U_3 \in \mathbb{R}^{n_3 \times n_3}$ are orthogonal, then

$$\mathcal{A} \;=\; \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$$

where

$$\mathcal{S} \;=\; \mathcal{A} \times_1 U_1^T \times_2 U_2^T \times_3 U_3^T .$$

We are representing $\mathcal{A}$ as Tucker product of a "core tensor" $\mathcal{S}$ and three orthogonal matrices.

The Higher-Order SVD

The Tucker Product Representation

The Challenge

Given $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, compute

$$\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$$

and

$$U_1 \in \mathbb{R}^{n_1 \times r_1}, \; U_2 \in \mathbb{R}^{n_2 \times r_2}, \; U_3 \in \mathbb{R}^{n_3 \times r_3}$$

such that

$$\mathcal{A} \; = \; \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$$

is an "illuminating" Tucker product representation of $\mathcal{A}$.

The Higher Order SVD (HOSVD)

If the U's are from the Modal Unfolding SVDs...

Suppose $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is given. If

$$\begin{aligned} \mathcal{A}_{(1)} &= U_1 \Sigma_1 V_1^T \\ \mathcal{A}_{(2)} &= U_2 \Sigma_2 V_2^T \\ \mathcal{A}_{(3)} &= U_3 \Sigma_3 V_3^T \end{aligned}$$

are SVDs and

$$\mathcal{S} = \mathcal{A} \times_1 U_1^T \times_2 U_2^T \times_3 U_3^T,$$

then

$$\mathcal{A} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3,$$

is the higher-order SVD of $\mathcal{A}$.

The Higher-Order SVD (HOSVD)

The HOSVD of a Matrix IS the SVD of that Matrix

If $d = 2$ then $\mathcal{A}$ is a matrix and the HOSVD is the SVD. Indeed, if

$$\begin{aligned} A &= A_{(1)} = U_1 \Sigma_1 V_1^T \\ A^T &= A_{(2)} = U_2 \Sigma_2 V_2^T \end{aligned}$$

then we can set $U = U_1 = V_2$ and $V = U_2 = V_1$. Note that

$$\mathcal{S} = (\mathcal{A} \times_1 U_1^T) \times_2 U_2^T = (U_1^T A) \times_2 U_2 = U_1^T A V_1 = \Sigma_1.$$

The HOSVD

Core Tensor Properties

If

$$\mathcal{A}_{(1)} = U_1 \Sigma_1 V_1^T \qquad \mathcal{A}_{(2)} = U_2 \Sigma_2 V_2^T \qquad \mathcal{A}_{(3)} = U_3 \Sigma_3 V_3^T$$

are SVDs and

$$\mathcal{A} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$$

then

$$\mathcal{A}_{(1)} = U_1 \mathcal{S}_{(1)} (U_3 \otimes U_2)^T \quad \text{and} \quad \mathcal{S}_{(1)} = \Sigma_1 V_1 (U_3 \otimes U_2)$$

It follows that the rows of $S_{(1)}$ are mutually orthogonal and that the singular values of $\mathcal{A}_{(1)}$ are the 2-norms of these rows.

The HOSVD

Core Tensor Properties

If

$$\mathcal{A}_{(1)} = U_1 \Sigma_1 V_1^T \qquad \mathcal{A}_{(2)} = U_2 \Sigma_2 V_2^T \qquad \mathcal{A}_{(3)} = U_3 \Sigma_3 V_3^T$$

are SVDs and

$$\mathcal{A} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$$

then

$$\mathcal{A}_{(2)} = U_2 \mathcal{S}_{(2)} (U_3 \otimes U_1)^T \quad \text{and} \quad \mathcal{S}_{(2)} = \Sigma_2 V_2 (U_3 \otimes U_1)$$

It follows that the rows of $S_{(2)}$ are mutually orthogonal and that the singular values of $\mathcal{A}_{(2)}$ are the 2-norms of these rows.

The HOSVD

Core Tensor Properties

If

$$\mathcal{A}_{(1)} = U_1 \Sigma_1 V_1^T \qquad \mathcal{A}_{(2)} = U_2 \Sigma_2 V_2^T \qquad \mathcal{A}_{(3)} = U_3 \Sigma_3 V_3^T$$

are SVDs and

$$\mathcal{A} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$$

then

$$\mathcal{A}_{(3)} = U_3 \mathcal{S}_{(3)} (U_2 \otimes U_1)^T \quad \text{and} \quad \mathcal{S}_{(3)} = \Sigma_3 V_3 (U_2 \otimes U_1)$$

It follows that the rows of $S_{(3)}$ are mutually orthogonal and that the singular values of $\mathcal{A}_{(3)}$ are the 2-norms of these rows.

The Core Tensor $\mathcal{S}$ is Graded

$$\mathcal{S}_{(1)} \;=\; \Sigma_1 V_1 (U_3 \otimes U_2) \;\Rightarrow\; \| \, \mathcal{S}(j,:,:) \, \|_F \;=\; \sigma_j(\mathcal{A}_{(1)}) \qquad j = 1{:}n_1$$

$$\mathcal{S}_{(2)} \;=\; \Sigma_2 V_2 (U_3 \otimes U_1) \;\Rightarrow\; \| \, \mathcal{S}(:,j,:) \, \|_F \;=\; \sigma_j(\mathcal{A}_{(2)}) \qquad j = 1{:}n_2$$

$$\mathcal{S}_{(3)} \;=\; \Sigma_3 V_3 (U_2 \otimes U_1) \;\Rightarrow\; \| \, \mathcal{S}(:,:,j) \, \|_F \;=\; \sigma_j(\mathcal{A}_{(3)}) \qquad j = 1{:}n_3$$

The norms of slices are getting smaller as you move away from $\mathcal{A}(1,1,1)$

Notation: $\sigma_j(\mathrm{C})$ is the jth largest singular value of the matrix C.

Thinking About the HOSVD

It is a Graded Sum of Rank-1 Tensors...

If $\mathcal{A} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$ is the HOSVD of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then

$$\mathcal{A} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)$$

where $r_1 = \text{rank}(A_{(1)})$, $r_2 = \text{rank}(A_{(2)})$, and $r_3 = \text{rank}(A_{(3)})$

And It Can Be Truncated...

If $\mathcal{A} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3$ is the HOSVD of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then

$$\mathcal{A} \approx \sum_{j_1=1}^{\tilde{r}_1} \sum_{j_2=1}^{\tilde{r}_2} \sum_{j_3=1}^{\tilde{r}_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)$$

where $\tilde{r}_1 \leq r_1$, $\tilde{r}_2 \leq r_2$, and $\tilde{r}_3 \leq r_3$.

The Truncated HOSVD

Just "Shorten" the Summations

$$\mathcal{A} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)$$

$$\mathcal{A}_{\mathbf{r}} = \sum_{j_1=1}^{\tilde{r}_1} \sum_{j_2=1}^{\tilde{r}_2} \sum_{j_3=1}^{\tilde{r}_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)$$

What can we say about the "thrown away" terms?

The Truncated HOSVD

Just "Shorten" the Summations

$$\mathcal{A} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)$$

$$\mathcal{A}_{\mathbf{r}} = \sum_{j_1=1}^{\tilde{r}_1} \sum_{j_2=1}^{\tilde{r}_2} \sum_{j_3=1}^{\tilde{r}_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)$$

Use these results...

$$\begin{aligned} \| \mathcal{S}(j,:,:) \|_F &= \sigma_j(\mathcal{A}_{(1)}) \qquad j = 1{:}n_1 \\ \| \mathcal{S}(:,j,:) \|_F &= \sigma_j(\mathcal{A}_{(2)}) \qquad j = 1{:}n_2 \\ \| \mathcal{S}(:,:,j) \|_F &= \sigma_j(\mathcal{A}_{(3)}) \qquad j = 1{:}n_3 \end{aligned}$$

Optional "Fun" Problem

Problem E3. What can you say about $\| \mathcal{A} - \mathcal{A}_r \|_F$ assuming that $\sigma_{\tilde{r}_1}(A_{(1)}) \leq \delta$, $\sigma_{\tilde{r}_2}(A_{(2)}) \leq \delta$, and $\sigma_{\tilde{r}_3}(A_{(3)}) \leq \delta$?

Problem A3. In the QR with column pivoting (QRP) decomposition $AP = QR$ the upper triangular matrix $R \in \mathbb{R}^{n \times n}$ is graded in the sense that

$$r_{jj}^2 \geq \sum_{i=j}^{k} r_{ik}^2 \qquad k = j{:}n$$

Formulate an HOQRP factorization for a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ that is based on the QR-with-column-pivoting factorizations

$$\mathcal{A}_{(k)} P_k \;=\; Q_k R_k$$

for $k = 1{:}3$. Does the core tensor have any special "grading" properties?

The Tucker Nearness Problem

Modal Rank

Definition

We say that

$$\mathcal{A} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)$$

has modal rank (r_1, r_2, r_3) if $r_1 = \text{rank}(A_{(1)})$, $r_2 = \text{rank}(A_{(2)})$, and $r_3 = \text{rank}(A_{(3)})$,

The Tucker Nearness Problem

Approximation With a "Shorter" Tucker Product

Assume that $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ has modal rank (r_1, r_2, r_3). Given integers $\tilde{r}_1$, $\tilde{r}_2$ and $\tilde{r}_3$ that satisfy $\tilde{r}_1 \leq r_1$, $\tilde{r}_2 \leq r_2$, and $\tilde{r}_3 \leq r_3$, compute

$$
\begin{array}{lll}
U_1: & n_1 \times \tilde{r}_1, & \text{orthonormal columns} \\
U_2: & n_2 \times \tilde{r}_2, & \text{orthonormal columns} \\
U_3: & n_3 \times \tilde{r}_3, & \text{orthonormal columns}
\end{array}
$$

and tensor $\mathcal{S} \in \mathbb{R}^{\tilde{r}_1 \times \tilde{r}_2 \times \tilde{r}_3}$ so that

$$
\left\| \mathcal{A} - \sum_{j_1=1}^{\tilde{r}_1} \sum_{j_2=1}^{\tilde{r}_2} \sum_{j_3=1}^{\tilde{r}_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3) \right\|_F
$$

is minimized.

The Tucker Nearness Problem

The Plan...

Develop a component-wise optimization framework for minimizing

$$\left\| \mathcal{A} - \sum_{j_1=1}^{\tilde{r}_1} \sum_{j_2=1}^{\tilde{r}_2} \sum_{j_3=1}^{\tilde{r}_3} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3) \right\|_F$$

Equivalent to finding U_1, U_2, and U_3 (all with orthonormal columns) and core tensor $\mathcal{S} \in \mathbb{R}^{\tilde{r}_1 \times \tilde{r}_2 \times \tilde{r}_3}$ so that

$$\| \, \text{vec}(\mathcal{A}) \; - \; (U_3 \otimes U_2 \otimes U_1)\text{vec}(\mathcal{S}) \, \|_F$$

is minimized.

The Tucker Nearness Problem

The "Removal" of $\mathcal{S}$

Since $\mathcal{S}$ must minimize

$$\| \operatorname{vec}(\mathcal{A}) \; - \; (U_3 \otimes U_2 \otimes U_1) \cdot \operatorname{vec}(\mathcal{S}) \; \|$$

and $U_3 \otimes U_2 \otimes U_1$ has orthonormal columns, we see that

$$\mathcal{S} = \left(U_3^T \otimes U_2^T \otimes U_1^T \right) \cdot \operatorname{vec}(\mathcal{A}).$$

Thus, the goal is to choose the U_i so that

$$\| \; \left(I \; - \; (U_3 \otimes U_2 \otimes U_1) \left(U_3^T \otimes U_2^T \otimes U_1^T\right)\right) \operatorname{vec}(\mathcal{A}) \; \|$$

is minimized.

The Tucker Nearness Problem

Reformulation...

Since $U_3 \otimes U_2 \otimes U_1$ has orthonormal columns, it follows that minimizing

$$\| \left(I \;-\; (U_3 \otimes U_2 \otimes U_1)\left(U_3^T \otimes U_2^T \otimes U_1^T\right)\right) \mathrm{vec}(\mathcal{A}) \,\|$$

is the same as maximizing

$$\| \left(U_3^T \otimes U_2^T \otimes U_1^T\right) \cdot \mathrm{vec}(\mathcal{A}) \,\|$$

If Q has orthonormal columns then $\| \,(I - QQ^T)a \,\|_2^2 \;=\; \| \,a\, \|^2 \;-\; \| \,Q^T a\, \|_2^2$.

The Tucker Nearness Problem

Three Reshapings of the Objective Function...

$$\| \left(U_3^T \otimes U_2^T \otimes U_1^T \right) \cdot \text{vec}(\mathcal{A}) \|$$

$$=$$

$$\| \; U_1^T \cdot A_{(1)} \cdot (U_3 \otimes U_2) \; \|_F$$

$$=$$

$$\| \; U_2^T \cdot A_{(2)} \cdot (U_3 \otimes U_1) \; \|_F$$

$$=$$

$$\| \; U_3^T \cdot A_{(3)} \cdot (U_2 \otimes U_1) \; \|_F$$

Sets the stage for a componentwise optimization solution approach...

Componentwise Optimization Framework

A Sequence of Three Linear Problems...

$$\| \left(U_3^T \otimes U_2^T \otimes U_1^T\right) \cdot \text{vec}(\mathcal{A}) \|$$

$$=$$

$$\| \; U_1^T \cdot A_{(1)} \cdot (U_3 \otimes U_2) \; \|_F \quad \Leftarrow$$

1. Fix U_2 and U_3 and maximize with U_1.

$$=$$

$$\| \; U_2^T \cdot A_{(2)} \cdot (U_3 \otimes U_1) \; \|_F \quad \Leftarrow$$

2. Fix U_1 and U_3 and maximize with U_2.

$$=$$

$$\| \; U_3^T \cdot A_{(3)} \cdot (U_2 \otimes U_1) \; \|_F \quad \Leftarrow$$

3. Fix U_1 and U_2 and maximize with U_3.

These max problems are SVD problems...

How do you maximize $\| Q^T M \|_F$ where $Q \in \mathbb{R}^{m\times r}$ has orthonormal columns, $M \in \mathbb{R}^{m\times n}$, and $r \leq n$?

If

$$M = U\Sigma V^T$$

is the SVD of M, then

$$\begin{aligned} \| Q^T M \|_F^2 &= \| Q^T U\Sigma V^T \|_F^2 = \| Q^T U\Sigma \|_F^2 \\ &= \sum_{k=1}^{n} \sigma_k^2 \| Q^T U(:,k) \|_2^2. \end{aligned}$$

The best you can do is to set $Q = U(:,1{:}r)$.

Solution Framework

A Sequence of Three Linear Problems...

Repeat:

1. Compute the SVD $\mathcal{A}_{(1)} \cdot (U_3 \otimes U_2) \;=\; \tilde{U}_1 \Sigma_1 V_1^T$

   and set $U_1 \;=\; \tilde{U}_1(:, 1{:}\tilde{r}_1)$.

2. Compute the SVD $\mathcal{A}_{(2)} \cdot (U_3 \otimes U_1) \;=\; \tilde{U}_2 \Sigma_2 V_2^T$

   and set $U_2 \;=\; \tilde{U}_2(:, 1{:}\tilde{r}_2)$.

3. Compute the SVD $\mathcal{A}_{(3)} \cdot (U_2 \otimes U_1) \;=\; \tilde{U}_3 \Sigma_3 V_3^T$

   and set $U_3 \;=\; \tilde{U}_3(:, 1{:}\tilde{r}_3)$.

Initial guess via the HOSVD. The highlighted matrix-matrix products are structured and ecomomies can be realized.

A Jacobi Variant

A Jacobi Procedure

Maximizing Mass on the Diagonal

Assume that $\mathcal{A}$ is m-by-m-by-m and define

$$\phi(\mathcal{A}) \;=\; \sum_{i=1}^{n} a_{iii}$$

Our goal is to compute orthogonal U, V, and W so that if the tensor tensor $\mathcal{S}$ is defined by

$$\text{vec}(\mathcal{S}) \;=\; (W \otimes V \otimes U)\text{vec}(\mathcal{A})$$

then $\phi(\mathcal{S})$ is maximized.

The Jacobi SVD procedure for matrices can be derived with a trace max objective function.

A Jacobi Procedure

Updating: Make $\mathcal{S}$ More Diagonal

Currrent: $\text{vec}(\mathcal{A}) \; = (W \otimes V \otimes U) \cdot \text{vec}(\mathcal{S})$

Determine: Orthogonal $\tilde{U}$, $\tilde{V}$, and $\tilde{W}$ so that if

$$\text{vec}(\tilde{\mathcal{S}}) \;=\; (\tilde{W} \otimes \tilde{V} \otimes \tilde{U})^T \cdot \text{vec}(\mathcal{S})$$

then $\phi(\tilde{\mathcal{S}}) > \phi(\mathcal{S})$.

Update:

$$\begin{aligned} \text{vec}(\mathcal{A}) &= (W \otimes V \otimes U) \cdot \text{vec}(\mathcal{S}) \\ &= (W \otimes V \otimes U) \cdot \left(\tilde{W} \otimes \tilde{V} \otimes \tilde{U}\right) \cdot \text{vec}(\tilde{\mathcal{S}}) \\ &= \left(W \cdot \tilde{W} \otimes V \cdot \tilde{V} \otimes U \cdot \tilde{U}\right) \cdot \text{vec}(\tilde{\mathcal{S}}) \end{aligned}$$

A Jacobi Procedure

Simple, Tractable Choices...

$$\tilde{W} \otimes \tilde{V} \otimes \tilde{U} = \begin{cases} I_n \otimes J_{pq}(\beta) \otimes J_{pq}(\alpha) \\ J_{pq}(\beta) \otimes I_n \otimes J_{pq}(\alpha) \\ J_{pq}(\beta) \otimes J_{pq}(\alpha) \otimes I_n \end{cases}$$

where $J_{pq}(\theta)$ is a Jacobi rotation in planes p and q.

These updates modify only two diagonal entries: (p,p,p) and (q,q,q).
Sweep through all possible (p,q) and all three types of updates.

A Jacobi Procedure

A Sample 2-by-2-by-2 Subproblem

Choose $c_\alpha = \cos(\alpha)$, $s_\alpha) = \sin(\alpha)$, $c_\beta = \cos(\beta)$, and $s_\beta) = \sin(\beta)$, so that if

$$\begin{bmatrix} \sigma_{111} & \sigma_{121} \\ \sigma_{211} & \sigma_{221} \end{bmatrix} = \begin{bmatrix} c_\alpha & s_\alpha \\ -s_\alpha & c_\alpha \end{bmatrix}^T \begin{bmatrix} s_{111} & s_{121} \\ s_{211} & s_{221} \end{bmatrix} \begin{bmatrix} c_\beta & s_\beta \\ -s_\beta & c_\beta \end{bmatrix}$$

and

$$\begin{bmatrix} \sigma_{112} & \sigma_{122} \\ \sigma_{212} & \sigma_{222} \end{bmatrix} = \begin{bmatrix} c_\alpha & s_\alpha \\ -s_\alpha & c_\alpha \end{bmatrix}^T \begin{bmatrix} s_{112} & s_{122} \\ s_{212} & s_{222} \end{bmatrix} \begin{bmatrix} c_\beta & s_\beta \\ -s_\beta & c_\beta \end{bmatrix}$$

then $\sigma_{111} + \sigma_{222}$ is maximized.

The Tensor Train Representation

The Tensor Train Idea

A Data Sparse Representation

Approximate a high-order tensor with a collection of order-3 tensors.

Each order-3 tensor is connected to its left and right "neighbor" through a simple summation.

Tensor Train: An Example

Given the "carriages"...

$$\begin{aligned} \mathcal{G}_1: &\quad n_1 \times r_1 \\ \mathcal{G}_2: &\quad r_1 \times n_2 \times r_2 \\ \mathcal{G}_3: &\quad r_2 \times n_3 \times r_3 \\ \mathcal{G}_4: &\quad r_3 \times n_4 \times r_4 \\ \mathcal{G}_5: &\quad r_4 \times n_5 \end{aligned}$$

We define the train" $\mathcal{A}(1{:}n_1, 1{:}n_2, 1{:}n_3, 1{:}n_4, 1{:}n_5)$...

$$\mathcal{A}(i_1, i_2, i_3, i_4, i_5)$$
$$=$$
$$\sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \sum_{k_3=1}^{r_3} \sum_{k_4=1}^{r_4} \mathcal{G}_1(i_1, k_1) \cdot \mathcal{G}_2(k_1, i_2, k_2) \cdot \mathcal{G}_3(k_2, i_3, k_3) \cdot \mathcal{G}_4(k_3, i_4, k_4) \cdot \mathcal{G}_5(k_4, i_5)$$

Think of a graph where the nodes are low-order tensors and the edges are the summations.

Tensor Train: An Example

Given the "carriages"...

$$\begin{aligned} \mathcal{G}_1: &\quad n_1 \times r_1 \\ \mathcal{G}_2: &\quad r_1 \times n_2 \times r_2 \\ \mathcal{G}_3: &\quad r_2 \times n_3 \times r_3 \\ \mathcal{G}_4: &\quad r_3 \times n_4 \times r_4 \\ \mathcal{G}_5: &\quad r_4 \times n_5 \end{aligned}$$

We define the train" $\mathcal{A}(1{:}n_1, 1{:}n_2, 1{:}n_3, 1{:}n_4, 1{:}n_5)$...

$$\mathcal{A}(i_1, i_2, i_3, i_4, i_5)$$
$$=$$
$$\sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \sum_{k_3=1}^{r_3} \sum_{k_4=1}^{r_4} \mathcal{G}_1(i_1, k_1) \cdot \mathcal{G}_2(k_1, i_2, k_2) \cdot \mathcal{G}_3(k_2, i_3, k_3) \cdot \mathcal{G}_4(k_3, i_4, k_4) \cdot \mathcal{G}_5(k_4, i_5)$$

$O(nr^2)$ vs $O(n^5)$

Tensor Train: An Example

Given the "carriages"...

$$\begin{array}{ll} \mathcal{G}_1: & n_1 \times r_1 \\ \mathcal{G}_2: & r_1 \times n_2 \times r_2 \\ \mathcal{G}_3: & r_2 \times n_3 \times r_3 \\ \mathcal{G}_4: & r_3 \times n_4 \times r_4 \\ \mathcal{G}_5: & r_4 \times n_5 \end{array}$$

We define the train" $\mathcal{A}(1{:}n_1, 1{:}n_2, 1{:}n_3, 1{:}n_4, 1{:}n_5)$...

$$\mathcal{A}(i_1, i_2, i_3, i_4, i_5)$$

$$=$$

$$\sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \sum_{k_3=1}^{r_3} \sum_{k_4=1}^{r_4} \mathcal{G}_1(i_1, k_1) \cdot \mathcal{G}_2(k_1, i_2, k_2) \cdot \mathcal{G}_3(k_2, i_3, k_3) \cdot \mathcal{G}_4(k_3, i_4, k_4) \cdot \mathcal{G}_5(k_4, i_5)$$

$O(nr^2)$ vs $O(n^5)$

Tensor Train: An Example

Given the "carriages"...

$$\mathcal{G}_1: \quad n_1 \times r_1$$

$$\mathcal{G}_2: \quad r_1 \times n_2 \times r_2$$

$$\mathcal{G}_3: \quad r_2 \times n_3 \times r_3$$

$$\mathcal{G}_4: \quad r_3 \times n_4 \times r_4$$

$$\mathcal{G}_5: \quad r_4 \times n_5$$

We define the train" $\mathcal{A}(1{:}n_1, 1{:}n_2, 1{:}n_3, 1{:}n_4, 1{:}n_5)$...

$$\mathcal{A}(i_1, i_2, i_3, i_4, i_5)$$

$$=$$

$$\sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \sum_{k_3=1}^{r_3} \sum_{k_4=1}^{r_4} \mathcal{G}_1(i_1, k_1) \cdot \mathcal{G}_2(k_1, i_2, k_2) \cdot \mathcal{G}_3(k_2, i_3, k_3) \cdot \mathcal{G}_4(k_3, i_4, k_4) \cdot \mathcal{G}_5(k_4, i_5)$$

$O(nr^2)$ vs $O(n^5)$

Tensor Train: An Example

Given the "carriages"...

$$\mathcal{G}_1: \quad n_1 \times r_1$$

$$\mathcal{G}_2: \quad r_1 \times n_2 \times r_2$$

$$\mathcal{G}_3: \quad r_2 \times n_3 \times r_3$$

$$\mathcal{G}_4: \quad r_3 \times n_4 \times r_4$$

$$\mathcal{G}_5: \quad r_4 \times n_5$$

We define the train" $\mathcal{A}(1{:}n_1, 1{:}n_2, 1{:}n_3, 1{:}n_4, 1{:}n_5)$...

$$\mathcal{A}(i_1, i_2, i_3, i_4, i_5)$$
$$=$$
$$\sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \sum_{k_3=1}^{r_3} \sum_{k_4=1}^{r_4} \mathcal{G}_1(i_1, k_1)\cdot\mathcal{G}_2(k_1, i_2, k_2)\cdot\mathcal{G}_3(k_2, i_3, k_3)\cdot\mathcal{G}_4(k_3, i_4, k_4)\cdot\mathcal{G}_5(k_4, i_5)$$

$O(nr^2)$ vs $O(n^5)$

Computing a Tensor Train Representation

Main Idea

A sequence of unfoldings is produced.

The unfoldings get narrower and narrower.

A rank-revealing SVD $U(\Sigma V^T) = UZ$ is computed each time.

The "carriages" are reshaped U-matrices.

Computing a Tensor Train Representation

1(a) Rank-revealing SVD: $\texttt{reshape}(A, [n_1\ ,\ n_2 n_3 n_4 n_5])\ =\ U_1 Z_1.$

$$\mathcal{G}_1 = \texttt{reshape}(U_1, [n_1, r_1]).$$

Computing a Tensor Train Representation

1(a) Rank-revealing SVD: $\texttt{reshape}(A, [n_1\ ,\ n_2 n_3 n_4 n_5]) \ = \ U_1 Z_1.$

$\mathcal{G}_1 = \texttt{reshape}(U_1, [n_1, r_1]).$

2(a) Rank-revealing SVD: $\texttt{reshape}(Z_1, [r_1 n_2, n_3 n_4 n_5]) \ = \ U_2 Z_2.$

$\mathcal{G}_2 = \texttt{reshape}(U_2, [r_1, n_2, r_2]).$

Computing a Tensor Train Representation

1(a) Rank-revealing SVD: $\texttt{reshape}(A, [n_1\ ,\ n_2 n_3 n_4 n_5]) \ = \ U_1 Z_1.$

$\mathcal{G}_1 = \texttt{reshape}(U_1, [n_1, r_1]).$

2(a) Rank-revealing SVD: $\texttt{reshape}(Z_1, [r_1 n_2, n_3 n_4 n_5]) \ = \ U_2 Z_2.$

$\mathcal{G}_2 = \texttt{reshape}(U_2, [r_1, n_2, r_2]).$

3(a) Rank-revealing SVD: $\texttt{reshape}(Z_2, [r_2 n_3, n_4 n_5]) \ = \ U_3 Z_3.$

$\mathcal{G}_3 = \texttt{reshape}(U_3, [r_2, n_3, r_3]).$

Computing a Tensor Train Representation

1(a) Rank-revealing SVD: $\texttt{reshape}(A, [n_1\ ,\ n_2 n_3 n_4 n_5]) \ = \ U_1 Z_1.$

$\mathcal{G}_1 = \texttt{reshape}(U_1, [n_1, r_1]).$

2(a) Rank-revealing SVD: $\texttt{reshape}(Z_1, [r_1 n_2, n_3 n_4 n_5]) \ = \ U_2 Z_2.$

$\mathcal{G}_2 = \texttt{reshape}(U_2, [r_1, n_2, r_2]).$

3(a) Rank-revealing SVD: $\texttt{reshape}(Z_2, [r_2 n_3, n_4 n_5]) \ = \ U_3 Z_3.$

$\mathcal{G}_3 = \texttt{reshape}(U_3, [r_2, n_3, r_3]).$

4(a) Rank-revealing SVD: $\texttt{reshape}(Z_3, [r_3 n_4, n_5]) \ = \ U_4 Z_4.$

$\mathcal{G}_4 = \texttt{reshape}(U_4, [r_3, n_4, r_4]).$

$\mathcal{G}_5 = \texttt{reshape}(Z_4, [r_4, n_5]).$