Four Talks on Tensor Computations

2. SVD-Based Tensor Decompositions

Charles F. Van Loan
Cornell University

SCAN Seminar

November 3, 2014
What is this Lecture About?

Pushing the SVD Envelope

The SVD is a powerful tool for exposing the structure of a matrix and for capturing its essence through optimal, data-sparse representations:

\[
A = \sum_{k=1}^{\text{rank}(A)} \sigma_k u_k v_k^T \approx \sum_{k=1}^{\hat{r}} \sigma_k u_k v_k^T
\]

For a tensor \( \mathcal{A} \), let’s try for something similar...

\[
\mathcal{A} = \sum \text{whatever!}
\]
What We will Need...

A Mechanism for Updating

\[ U^T AV = \Sigma \]

We will need operations that can transform the given tensor into something simple.

The Notion of an Abbreviated Decomposition

\[ \text{vec}(A) = (V \otimes U) \text{vec}(\Sigma) \]

\[ = \text{a structured sum of the } V(:,i) \otimes U(:,j) \]

\[ \approx \text{a SHORTER structured sum of the } V(:,i) \otimes U(:,j) \]

We’ll need a way to “pull out” the essential part of a decomposition.
What is this Lecture About?

Outline

- The Mode-k Product and the Tucker Product.
- The Tucker Representation and Its Properties
- The Higher-Order SVD of a tensor.
- An ALS Framework for Reduced-Rank Tucker Approximation
- Approximation via the Kronecker Product SVD
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

$$\mathcal{A}_{(2)} = \begin{bmatrix}
    a_{11} & a_{21} & a_{31} & a_{41} & a_{11} & a_{21} & a_{31} & a_{41} \\
    a_{12} & a_{22} & a_{32} & a_{42} & a_{12} & a_{22} & a_{32} & a_{42} \\
    a_{13} & a_{23} & a_{33} & a_{43} & a_{13} & a_{23} & a_{33} & a_{43}
\end{bmatrix}$$

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-\(k\) Matrix Product: An Example

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 that (1) \(B \in \mathbb{R}^{4 \times 5 \times 2}\) and (2) \(B_{(2)} = M \cdot A_{(2)}\).
A mode-1 example when the Tensor $A$ is Second Order...

\[
\begin{bmatrix}
  b_{11} & b_{12} & b_{13} \\
  b_{21} & b_{22} & b_{23}
\end{bmatrix}
= \begin{bmatrix}
  m_{11} & m_{12} \\
  m_{21} & m_{22}
\end{bmatrix}
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23}
\end{bmatrix}
\]

(The fibers of $A$ are its columns.)

A mode-2 example when the Tensor $A$ is Second Order...

\[
\begin{bmatrix}
  b_{11} & b_{12} \\
  b_{21} & b_{22} \\
  b_{31} & b_{32} \\
  b_{41} & b_{42}
\end{bmatrix}
= \begin{bmatrix}
  m_{11} & m_{12} & m_{13} \\
  m_{21} & m_{22} & m_{23} \\
  m_{31} & m_{32} & m_{33} \\
  m_{41} & m_{42} & m_{43}
\end{bmatrix}
\begin{bmatrix}
  a_{11} & a_{21} \\
  a_{12} & a_{22} \\
  a_{13} & a_{23}
\end{bmatrix}
\]

(The fibers of $A$ are its rows.)
The Mode-\( k \) Product: Definitions

### Mode-1

If \( A \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) and \( M \in \mathbb{R}^{m_1 \times n_1} \), then the mode-1 product

\[
B = A \times_1 M \in \mathbb{R}^{m_1 \times n_2 \times n_3}
\]

is defined by

\[
B(i_1, i_2, i_3) = \sum_{k=1}^{n_1} M(i_1, k)A(k, i_2, i_3)
\]

### Two Equivalent Formulations...

\[
B_{(1)} = M \cdot A_{(1)}
\]

\[
\text{vec}(B) = (M \otimes I_{n_2} \otimes I_{n_3})\text{vec}(B)
\]
Mode-2

If \( A \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) and \( M \in \mathbb{R}^{m_2 \times n_2} \), then the mode-1 product

\[
B = A \times_2 M \in \mathbb{R}^{n_1 \times m_2 \times n_3}
\]

is defined by

\[
B(i_1, i_2, i_3) = \sum_{k=1}^{n_2} M(i_2, k)A(i_1, k, i_3)
\]

Two Equivalent Formulations...

\[
B(2) = M \cdot A(2)
\]

\[
\text{vec}(B) = (I_{n_1} \otimes M \otimes I_{n_3})\text{vec}(B)
\]
The Mode-$k$ Product: Definitions

**Mode-3**

If $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $M \in \mathbb{R}^{m_3 \times n_3}$, then the mode-1 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...**

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

$$\text{vec}(\mathcal{B}) = (I_{n_1} \otimes I_{n_2} \otimes M)\text{vec}(\mathcal{B})$$
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 $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $M_k \in \mathbb{R}^{m_k \times n_k}$, $M_j \in \mathbb{R}^{m_j \times n_j}$, 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.
Matlab Tensor Toolbox: Mode-$k$ Matrix Product Using $\ttm$

\begin{verbatim}
n = [2 5 4 7];
A = tenrand(n);
M = randn(5,5);
B = ttm(A,M,k);
\end{verbatim}
The Tucker Product

Definition

The Tucker Product $\mathcal{X}$ of the tensor

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

with the matrices

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

is given by

$$\mathcal{X} = \mathcal{S} \times_1 U_1 \times_2 U_2 \times_3 U_3 = ((\mathcal{S} \times_1 U_1) \times_2 U_2) \times_3 U_3$$

It is a succession of mode-$k$ products.
The Tucker Product

Notation

\[
[[ S ; U_1, U_2, U_3 ]] \\
\equiv \\
S \times_1 U_1 \times_2 U_2 \times_3 U_3
\]
As a Scalar Summation...

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

$$X = [[S ; U_1, U_2, U_3]]$$

then

$$X(i_1, i_2, i_3) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} S(j_1, j_2, j_3) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2) \cdot U_3(i_3, j_3)$$
As a Scalar Summation and as a Sum of Rank-1 Tensors...

If \( S \in \mathbb{R}^{r_1 \times r_2 \times r_3} \), \( U_1 \in \mathbb{R}^{n_1 \times r_1} \), \( U_2 \in \mathbb{R}^{n_2 \times r_2} \), \( U_3 \in \mathbb{R}^{n_3 \times r_3} \), and

\[
X = \left[ [S; U_1, U_2, U_3] \right]
\]

then

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

\[
X = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} S(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)
\]
As a Giant Matrix-Vector Product...

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

$$\mathcal{X} = \left[ [ S ; U_1, U_2, U_3 ] \right]$$

then

$$\text{vec}(\mathcal{X}) = (U_3 \otimes U_2 \otimes U_1) \cdot \text{vec}(S)$$
function B = TuckerProd(A,M)
% A is a n(1)-by-...-n(d) tensor.
% M is a length-d cell array with
%   M{k} an m(k)-by-n(k) matrix.
% B is an m(1)-by-...-by-m(d) tensor given by
%   B = A x1 M{1} x2 M{2} ... xd M{d}
% where "xk" denotes mode-k matrix product.
    B = A;
    for k=1:length(A.size)
        B = ttm(B,M{k},k);
    end
MATLAB Tensor Toolbox: Tucker Tensor Set-Up

n = [5 8 3]; m = [4 6 2];
F = randn(n(1),m(1)); G = randn(n(2),m(2));
H = randn(n(3),m(3));
S = tenrand(m);
X = ttensor(S,{F,G,H});

A ttensor is a structure with two fields that is used to represent a tensor in Tucker form. In the above, X.core houses the the core tensor S while X.U is a cell array of the matrices F, G, and H that define the tensor X.
The Challenge

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

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

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

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

is an “illuminating” Tucker product representation of $\mathcal{A}$.
A Simple but Important Result

If $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 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

$$A = S \times_1 U_1 \times_2 U_2 \times_3 U_3$$

where

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

We will refer to the $U_k$ as the inverse factors and $S$ as the core tensor.

$$A = U_1(U_1^{-1}AU_2^{-1})U_2 = U_1SU_2$$
Proof.

\[ \mathcal{A} = \mathcal{A} \times_1 (U_1^{-1} U_1) \times_2 (U_2^{-1} U_2) \times_3 (U_3^{-1} U_3) \]

\[ = \left( \mathcal{A} \times_1 U_1^{-1} \times_2 U_2^{-1} \times_3 U_3^{-1} \right) \times_1 U_1 \times_2 U_2 \times_3 U_3 \]

\[ = S \times_1 U_1 \times_2 U_2 \times_3 U_3 \]
If the $U$'s are Orthogonal

If $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 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

$$A = S \times_1 U_1 \times_2 U_2 \times_3 U_3$$

with

$$S = A \times_1 U_1^T \times_2 U_2^T \times_3 U_3^T.$$
The Tucker Product Representation

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

\[
\begin{align*}
A^{(1)} &= U_1 \Sigma_1 V_1^T \\
A^{(2)} &= U_2 \Sigma_2 V_2^T \\
A^{(3)} &= U_3 \Sigma_3 V_3^T
\end{align*}
\]

are SVDs and

\[
S = A \times_1 U_1^T \times_2 U_2^T \times_3 U_3^T,
\]

then

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

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

function [S,U] = HOSVD(A)
% A is an n(1)-by-...-by-n(d) tensor.
% U is a length-d cell array with the
% property that U\{k\} is the left singular
% vector matrix of A's mode-k unfolding.
% S is an n(1)-by-...-by-n(d) tensor given by
% A x1 U\{1\} x2 U\{2\} ... xd U\{d\}

S = A;
for k=1:length(A.size)
    C = tenmat(A,k);
    [U{k},Sigma,V] = svd(C.data);
    S = ttm(S,U{k}',k);
end
The Higher-Order SVD (HOSVD)

The HOSVD of a Matrix

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

$$ A = A_{(1)} = U_1 \Sigma_1 V_1^T $$

$$ A^T = A_{(2)} = U_2 \Sigma_2 V_2^T $$

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

$$ S = (\mathcal{A} \times_1 U_1^T) \times_2 U_2^T = (U_1^T A) \times_2 U_2 = U_1^T AV_1 = \Sigma_1. $$
The HOSVD

Core Tensor Properties

If

\[
\mathcal{A}(1) = U_1 \Sigma_1 V_1^T \quad \mathcal{A}(2) = U_2 \Sigma_2 V_2^T \quad \mathcal{A}(3) = U_3 \Sigma_3 V_3^T
\]

are SVDs and

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

then

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

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.
Core Tensor Properties

If

\[ A_{(1)} = U_1 \Sigma_1 V_1^T \quad A_{(2)} = U_2 \Sigma_2 V_2^T \quad A_{(3)} = U_3 \Sigma_3 V_3^T \]

are SVDs and

\[ A = S \times_1 U_1 \times_2 U_2 \times_3 U_3 \]

then

\[ A_{(2)} = U_2 S_{(2)} (U_3 \otimes U_1)^T \quad \text{and} \quad S_{(2)} = \Sigma_2 V_2 (U_3 \otimes U_1) \]

It follows that the rows of \( S_{(2)} \) are mutually orthogonal and that the singular values of \( A_{(2)} \) are the 2-norms of these rows.
The HOSVD

Core Tensor Properties

If

\[ A^{(1)} = U_1 \Sigma_1 V_1^T \quad A^{(2)} = U_2 \Sigma_2 V_2^T \quad A^{(3)} = U_3 \Sigma_3 V_3^T \]

are SVDs and

\[ A = S \times_1 U_1 \times_2 U_2 \times_3 U_3 \]

then

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

It follows that the rows of \( S^{(3)} \) are mutually orthogonal and that the singular values of \( A^{(3)} \) are the 2-norms of these rows.
The Core Tensor $S$ is Graded

$$S_{(1)} = \Sigma_1 V_1(U_3 \otimes U_2) \Rightarrow \| S(j, :, :) \|_F = \sigma_j(A_{(1)}) \quad j = 1:n_1$$

$$S_{(2)} = \Sigma_2 V_2(U_3 \otimes U_1) \Rightarrow \| S(:, j, :) \|_F = \sigma_j(A_{(2)}) \quad j = 1:n_2$$

$$S_{(3)} = \Sigma_3 V_3(U_2 \otimes U_1) \Rightarrow \| S(:, :, j) \|_F = \sigma_j(A_{(3)}) \quad j = 1:n_3$$

Entries are getting smaller as you move away from $A(1, 1, 1)$
The Higher-Order SVD (HOSVD)

The HOSVD as a Multilinear Sum

If \( \mathbf{A} = \mathbf{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 \) is the HOSVD of \( \mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3} \), then

\[
\mathbf{A}(i_1, i_2, i_3) = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \sum_{j_3=1}^{n_3} \mathbf{U}_1(i_1, j_1) \mathbf{U}_2(i_2, j_2) \mathbf{U}_3(i_3, j_3)
\]

The HOSVD as a Matrix-Vector Product

If \( \mathbf{A} = \mathbf{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 \) is the HOSVD of \( \mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3} \), then

\[
\text{vec}(\mathbf{A}) = (\mathbf{U}_3 \otimes \mathbf{U}_2 \otimes \mathbf{U}_1) \cdot \text{vec}(\mathbf{S})
\]

Note that \( \mathbf{U}_3 \otimes \mathbf{U}_2 \otimes \mathbf{U}_1 \) is orthogonal.
Problem 1. Formulate an HOQRP factorization for a tensor $\mathbf{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ that is based on the QR-with-column-pivoting factorizations

$$\mathbf{A}(k) P_k = Q_k R_k$$

for $k = 1:d$. Does the core tensor have any special properties?

Problem 2. Does this inequality hold?

$$\| \mathbf{A} - \mathbf{A}_r \|_F^2 \leq \sum_{j=r_1+1}^{n_1} \sigma_j(\mathbf{A}(1))^2 + \sum_{j=r_2+1}^{n_2} \sigma_j(\mathbf{A}(2))^2 + \sum_{j=r_3+1}^{n_3} \sigma_j(\mathbf{A}(3))^2$$

Can you do better?

Problem 3. Show that

$$|\mathbf{A}(i_1, i_2, i_3) - \mathbf{X}_r(i_1, i_2, i_3)| \leq \min\{\sigma_{r_1+1}(\mathbf{A}(1)), \sigma_{r_2+1}(\mathbf{A}(2)), \sigma_{r_3+1}(\mathbf{A}(3))\}$$
Abbreviate the HOSVD Expansion...

\[ A(i_1, i_2, i_3) = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \sum_{j_3=1}^{n_3} S(j_1, j_2, j_3) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2) \cdot U_3(i_3, j_3) \]

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

What can we say about the “thrown away” terms?
The Truncated HOSVD

Abbreviate the HOSVD Expansion...

\[ A(i_1, i_2, i_3) = \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \sum_{j_3=1}^{n_3} S(j_1, j_2, j_3) \cdot U_1(i_1, j_1) \cdot U_2(i_2, j_2) \cdot U_3(i_3, j_3) \]

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

\[ \| S(j, :, :) \|_F = \sigma_j(A_{(1)}) \quad j = 1:n_1 \]

\[ \| S(:, j, :) \|_F = \sigma_j(A_{(2)}) \quad j = 1:n_2 \]

\[ \| S(:, :, j) \|_F = \sigma_j(A_{(3)}) \quad j = 1:n_3 \]

Entries are getting smaller as you move away from \( A(1, 1, 1) \)
Modal Rank

Definition

We say that

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

has modal rank \((r_1, r_2, r_3)\)
The Tucker Nearness Problem

A Tensor Optimization Problem

Given $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and integers $r_1 \leq n_1$, $r_2 \leq n_2$, and $r_3 \leq n_3$
minimize

$$\| \mathcal{A} - \mathcal{B} \|_F$$

over all tensors $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ that have modal rank $(r_1, r_2, r_3)$
A Tensor Optimization Problem

Given $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and integers $r_1 \leq n_1$, $r_2 \leq n_2$, and $r_3 \leq n_3$ minimize

$$\| A - B \|_F$$

over all tensors $B \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ that have modal rank $(r_1, r_2, r_3)$

A Matrix Optimization Problem

Given $A \in \mathbb{R}^{n_1 \times n_2}$ and integer $r \leq \min\{n_1, n_2\}$, minimize

$$\| A - B \|_F$$

over all matrices $B \in \mathbb{R}^{n_1 \times n_2}$ that have rank $r$.

The matrix problem has a happy solution via the SVD $A = U\Sigma V^T$:

$$B_{opt} = \sigma_1 U(:, 1)V(:, 1)^T + \cdots + \sigma_r U(:, r)V(:, r)^T$$
The Tucker Nearness Problem

The Plan...

Develop an Alternating Least Squares framework for minimizing

\[ \| A - \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} S(j_1, j_2, j_3) \cdot U_1(:, j_1) \odot U_2(:, j_2) \odot U_3(:, j_3) \|_F \]

Equivalent to finding \( U_1, U_2, \) and \( U_3 \) (all with orthonormal columns) and core tensor \( S \in \mathbb{R}^{r_1 \times r_2 \times r_3} \) so that

\[ \| \text{vec}(A) - (U_3 \otimes U_2 \otimes U_1)\text{vec}(S) \|_F \]

is minimized.
The Tucker Nearness Problem

The “Removal” of $S$

Since $S$ must minimize

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

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

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

and so our goal is to choose the $U_i$ so that

$$\| (I - (U_3 \otimes U_2 \otimes U_1) (U_3^T \otimes U_2^T \otimes U_1^T)) \cdot \text{vec}(A) \|$$

is minimized.
Reformulation...

Since $U_3 \otimes U_2 \otimes U_1$ has orthonormal columns, it follows that our goal is to choose orthonormal $U_i$ so that

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

is maximized.

If $Q$ has orthonormal columns then

$$\| (I - QQ^T)a \|^2 = \| a \|^2 - \| Q^T a \|^2$$
The Tucker Nearness Problem

Three Reshapings of the Objective Function...

\[ \| (U^T_3 \otimes U^T_2 \otimes U^T_1) \cdot \text{vec}(\mathcal{A}) \| = \]

\[ = \| U^T_1 \cdot A_1(1) \cdot (U_3 \otimes U_2) \|_F = \]

\[ = \| U^T_2 \cdot A_2(2) \cdot (U_3 \otimes U_1) \|_F = \]

\[ = \| U^T_3 \cdot A_3(3) \cdot (U_2 \otimes U_1) \|_F \]

Sets the stage for an alternating least squares solution approach...
Alternating Least Squares Framework

A Sequence of Three Linear Problems...

\[ \| (U_3^T \otimes U_2^T \otimes U_1^T) \cdot \text{vec}(A) \| = \| U_1^T \cdot A(1) \cdot (U_3 \otimes U_2) \|_F \leftarrow 1. \text{ Fix } U_2 \text{ and } U_3 \text{ and maximize with } U_1. \]

\[ = \| U_2^T \cdot A(2) \cdot (U_3 \otimes U_1) \|_F \leftarrow 2. \text{ Fix } U_1 \text{ and } U_3 \text{ and maximize with } U_2. \]

\[ = \| U_3^T \cdot A(3) \cdot (U_2 \otimes U_1) \|_F \leftarrow 3. \text{ Fix } U_1 \text{ and } U_2 \text{ and maximize with } U_3. \]

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

$$\| Q^T M \|_F^2 = \| Q^T U \Sigma V^T \|_F^2 = \| Q^T U \Sigma \|_F^2$$

$$= \sum_{k=1}^{n} \sigma_k^2 \| Q^T U(\cdot, k) \|_2^2.$$ 

The best you can do is to set $Q = U(\cdot, 1:r)$. 
Alternating Least Squares 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 $U_1 = \tilde{U}_1(:, 1:r_1)$.

2. Compute the SVD $\mathcal{A}(2) \cdot (U_3 \otimes U_1) = \tilde{U}_2 \Sigma_2 V_2^T$
   and set $U_2 = \tilde{U}_2(:, 1:r_2)$.

3. Compute the SVD $\mathcal{A}(3) \cdot (U_2 \otimes U_1) = \tilde{U}_3 \Sigma_3 V_3^T$
   and set $U_3 = \tilde{U}_3(:, 1:r_3)$.

Initial guess via the HOSVD
MATLAB Tensor Toolbox: The Function `tucker_als`

```matlab
n = [ 5 6 7 ];
% Generate a random tensor...
A = tenrand(n);
for r = 1:min(n)
    % Find the closest length-[r r r] ttensor...
    X = tucker_als(A, [r r r]);
    % Display the fit...
    E = double(X) - double(A);
    fit = norm(reshape(E, prod(n), 1));
    fprintf('r = %1d, fit = %5.3e
', r, fit);
end
```

The function `Tucker_als` returns a ttensor. Default values for the number of iterations and the termination criteria can be modified:

```matlab
X = Tucker_als(A, r, 'maxiters', 20, 'tol', .001)
```
Motivation

Unfold $\mathcal{A} \in \mathbb{R}^{n \times n \times n \times n}$ into an $n^2$-by-$n^2$ matrix $A$.

Express $A$ as a sum of Kronecker products:

$$A = \sum_{k=1}^{r} \sigma_k B_k \otimes C_k \quad B_k, C_k \in \mathbb{R}^{n \times n}$$

Back to tensor:

$$\mathcal{A} = \sum_{k=1}^{r} \sigma_k C_k \circ B_k$$

i.e.,

$$\mathcal{A}(i_1, i_2, j_1, j_2) = \sum_{k=1}^{r} \sigma_k C_k(i_1, i_2)B_k(j_1, j_2)$$

There is an optimal way of doing this.
The Nearest Kronecker Product Problem

Reshaping the Objective Function (3-by-2 case)

\[
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} & a_{14} \\
  a_{21} & a_{22} & a_{23} & a_{24} \\
  a_{31} & a_{32} & a_{33} & a_{34} \\
  a_{41} & a_{42} & a_{43} & a_{44} \\
  a_{51} & a_{52} & a_{53} & a_{54} \\
  a_{61} & a_{62} & a_{63} & a_{64}
\end{bmatrix}
- \begin{bmatrix}
  b_{11} & b_{12} \\
  b_{21} & b_{22} \\
  b_{31} & b_{32}
\end{bmatrix}
\otimes \begin{bmatrix}
  c_{11} & c_{12} \\
  c_{21} & c_{22}
\end{bmatrix}
= \begin{bmatrix}
  b_{11} \\
  b_{21} \\
  b_{31} \\
  b_{12} \\
  b_{22} \\
  b_{32}
\end{bmatrix}
\otimes \begin{bmatrix}
  c_{11} & c_{21} & c_{12} & c_{22}
\end{bmatrix}
\]
Minimizing the Objective Function (3-by-2 case)

It is a nearest rank-1 problem,

\[ \phi_A(B, C) = \| \tilde{A} - \text{vec}(B) \text{vec}(C)^T \|_F \]

with SVD solution:

\[ \tilde{A} = U \Sigma V^T \]

\[ \text{vec}(B) = \sqrt{\sigma_1} U(:,1) \]

\[ \text{vec}(C) = \sqrt{\sigma_1} V(:,1) \]
### The Nearest Kronecker Product Problem

#### The “Tilde Matrix”

Let \( A \) be a matrix as follows:

\[
A = \begin{bmatrix}
  a_{11} & a_{12} & a_{13} & a_{14} \\
  a_{21} & a_{22} & a_{23} & a_{24} \\
  a_{31} & a_{32} & a_{33} & a_{34} \\
  a_{41} & a_{42} & a_{43} & a_{44} \\
  a_{51} & a_{52} & a_{53} & a_{54} \\
  a_{61} & a_{62} & a_{63} & a_{64}
\end{bmatrix}
\]

Then, \( \tilde{A} \) is defined as:

\[
\tilde{A} = \begin{bmatrix}
  \text{vec}(A_{11})^T \\
  \text{vec}(A_{21})^T \\
  \text{vec}(A_{31})^T \\
  \text{vec}(A_{12})^T \\
  \text{vec}(A_{22})^T \\
  \text{vec}(A_{32})^T
\end{bmatrix}
\]

The Kronecker product implies:

\[
\tilde{A} = \begin{bmatrix}
  a_{11} & a_{21} & a_{12} & a_{22} \\
  a_{31} & a_{41} & a_{32} & a_{42} \\
  a_{51} & a_{61} & a_{52} & a_{62} \\
  a_{13} & a_{23} & a_{14} & a_{24} \\
  a_{33} & a_{43} & a_{34} & a_{44} \\
  a_{53} & a_{63} & a_{54} & a_{64}
\end{bmatrix}
\]

implies

\[
\tilde{A} = \begin{bmatrix}
  \text{vec}(A_{11})^T \\
  \text{vec}(A_{21})^T \\
  \text{vec}(A_{31})^T \\
  \text{vec}(A_{12})^T \\
  \text{vec}(A_{22})^T \\
  \text{vec}(A_{32})^T
\end{bmatrix}
\]
The Kronecker Product SVD (KPSVD)

**Theorem**

If

\[ A = \begin{bmatrix}
A_{11} & \cdots & A_{1,c_2} \\
\vdots & \ddots & \vdots \\
A_{r_2,1} & \cdots & A_{r_2,c_2}
\end{bmatrix} \quad A_{i_2,j_2} \in \mathbb{R}^{r_1 \times c_1} \]

then there exist \( U_1, \ldots, U_{r_{KP}} \in \mathbb{R}^{r_2 \times c_2}, \ V_1, \ldots, V_{r_{KP}} \in \mathbb{R}^{r_1 \times c_1} \), and scalars \( \sigma_1 \geq \cdots \geq \sigma_{r_{KP}} > 0 \) such that

\[ A = \sum_{k=1}^{r_{KP}} \sigma_k U_k \otimes V_k. \]

The sets \( \{\text{vec}(U_k)\} \) and \( \{\text{vec}(V_k)\} \) are orthonormal and \( r_{KP} \) is the Kronecker rank of \( A \) with respect to the chosen blocking.
Constructive Proof

Compute the SVD of $\tilde{A}$:

$$\tilde{A} = U\Sigma V^T = \sum_{k=1}^{r_{KP}} \sigma_k u_k v_k^T$$

and define the $U_k$ and $V_k$ by

$$\text{vec}(U_k) = u_k$$
$$\text{vec}(V_k) = v_k$$

for $k = 1: r_{KP}$.

$$U_k = \text{reshape}(u_k, r_2, c_2), \ V_k = \text{reshape}(v_k, r_1, c_1)$$
The Kronecker Product SVD (KPSVD)

Nearest rank-$r$

If $r \leq r_{KP}$, then

$$ A_r = \sum_{k=1}^{r} \sigma_k U_k \otimes V_k $$

is the nearest matrix to $A$ (in the Frobenius norm) that has Kronecker rank $r$. 
The Nearest Kronecker Product Problem

The Objective Function in Tensor Terms

\[ \phi_A(B, C) = \| A - B \otimes C \|_F = \sqrt{\sum_{i_1=1}^{r_1} \sum_{j_1=1}^{c_1} \sum_{i_2=1}^{r_2} \sum_{j_2=1}^{c_2} A(i_1, j_1, i_2, j_2) - B(i_2, j_2) C(i_1, j_1)} \]

We are trying to approximate an order-4 tensor with a pair of order-2 tensors.

Analogous to approximating a matrix (an order-2 tensor) with a rank-1 matrix (a pair of order-1 tensors.)
The Nearest Kronecker Product Problem

\[ \phi_A(B, C, D) \]
\[ = \|
A - B \otimes C \otimes D \|_F \]
\[ = \sqrt{\sum_{i_1=1}^{r_1} \sum_{j_1=1}^{c_1} \sum_{i_2=1}^{r_2} \sum_{j_2=1}^{c_2} \sum_{i_3=1}^{r_3} \sum_{j_3=1}^{c_3} A(i_1, j_1, i_2, j_2, i_3, j_3) - B(i_3, j_3)C(i_2, j_2)D(i_1, j_1)} \]

We are trying to approximate an order-6 tensor with a triplet of order-2 tensors.


The **Mode-k Matrix Product** is a contraction between a tensor and a matrix that produces another tensor.

The **Modal-rank** of a tensor is the vector of mode-\(k\) unfolding ranks.

The **Higher Order SVD** of a tensor \(\mathcal{A}\) assembles the SVDs of \(\mathcal{A}\)'s modal unfoldings.

The **Tucker Nearness Problem** for a given tensor \(\mathcal{A}\) involves finding the nearest tensor that has a given modal rank. Solved via alternating LS problems that involve SVDs.

The **Kronecker Product SVD** characterizes a block matrix as a sum of Kronecker products. By applying it to an unfolding of a tensor \(\mathcal{A}\), an outer product expansion for \(\mathcal{A}\) is obtained.