From Matrix to Tensor: 
The Transition to Numerical Multilinear Algebra

Lecture 6. The Tucker Representation

Charles F. Van Loan
Cornell University

The Gene Golub SIAM Summer School 2010
Selva di Fasano, Brindisi, Italy
Lecture 1. Introduction to Tensor Computations
Lecture 2. Tensor Unfoldings
Lecture 3. Transpositions, Kronecker Products, Contractions
Lecture 4. Tensor-Related Singular Value Decompositions
Lecture 5. The CP Representation and Tensor Rank
Lecture 6. The Tucker Representation
Lecture 7. Other Decompositions and Nearness Problems
Lecture 8. Multilinear Rayleigh Quotients
Lecture 9. The Curse of Dimensionality
Lecture 10. Special Topics
Approximate $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$

Find $\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ and orthonormal $U_1 \in \mathbb{R}^{n_1 \times r}$, $U_2 \in \mathbb{R}^{n_2 \times r}$ and $U_3 \in \mathbb{R}^{n_3 \times r}$ so that

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

Compare with the CP Representation from Last Lecture...

Find $\lambda \in \mathbb{R}^r$ $U_1 \in \mathbb{R}^{n_1 \times r}$, $U_2 \in \mathbb{R}^{n_2 \times r}$ and $U_3 \in \mathbb{R}^{n_3 \times r}$ so that

$$\mathcal{A} \approx \sum_{j=1}^{r} \lambda_j \cdot U_1(:, j) \circ U_2(:, j) \circ U_3(:, j)$$

The SVD Ambition: Illuminating Sums of Rank-1 Tensors
Approximate $A \in \mathbb{R}^{n_1 \times \cdots \times n_d}$

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

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

Compare with SVD of Matrix...

Find $S \in \mathbb{R}^{r_1 \times r_2}$ and orthonormal $U_1 \in \mathbb{R}^{n_1 \times r_1}$ and $U_2 \in \mathbb{R}^{n_3 \times r_2}$ so that

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

But we will not be able to make $S(1:r_1, 1:r_2, 1:r_3)$ diagonal.
What is This Lecture About?

The Plan...

Review the Tucker Representation and the HOSVD introduced in Lecture 4.

Develop an Alternating Least Squares framework for minimizing

$$\| A - \sum_{j=1}^{r} S(j) \cdot U_1(:,j_1) \circ U_2(:,j_2) \circ U_3(:,j_3) \|_F$$

Re-examine the tensor rank issue.

*Use Order-3 to Motivate Main Ideas.*
Mode-k Multiplication

Apply a matrix to all the mode-k fibers of a tensor.

For example, if $S \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ and $U_2 \in \mathbb{R}^{n_2 \times r_2}$, then

$$X = S \times_2 U_2 \iff X(2) = U_2 \cdot S(2)$$
The Tucker Product Representation (Brief Review)

The Tucker Product

A succession of mode-$k$ products.

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

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

$$= ((S \times_1 U_1) \times_2 U_2) \times_3 U_3$$

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

The tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is represented in Tucker form.
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

\[
X = \begin{bmatrix} S; U_1, U_2, U_3 \end{bmatrix}
\]

then

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

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

then

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

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

then

$$\text{vec}(\mathcal{X}) = (U_3 \otimes U_2 \otimes U_1) \cdot \text{vec}(S)$$
When the $U$’s are Orthogonal...

If $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is given 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 it is possible to determine

$$\chi = [[[S ; U_1, U_2, U_3]]]$$

so that $A = \chi$.

$$S = A \times_1 U_1^T \times_2 U_2^T \times_3 U_3^T$$
When the $U$'s are from the Modal Unfolding SVDs...

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

\[
A = [ [ S ; U_1, U_2, U_3 ] ]
\]

is the higher-order SVD of $A$. 

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});
Fsize = size(X.U{1}); Gsize = size(X.U{2});
Hsize = size(X.U{3});
Ssize = size(X.core); s = size(X);

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 $X$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fsize</td>
<td>[5,4]</td>
</tr>
<tr>
<td>Gsize</td>
<td>[8,6]</td>
</tr>
<tr>
<td>Hsize</td>
<td>[3,2]</td>
</tr>
<tr>
<td>Ssize</td>
<td>[4 6 2]</td>
</tr>
<tr>
<td>s</td>
<td>[5 8 3]</td>
</tr>
</tbody>
</table>
Problem 6.1. Suppose

\[ A = [[S; M_1, M_2, M_3]] \]

and that each \( M_i \) has more rows than columns. If \( M_i = Q_iR_i \) is a QR factorization, then

\[ A = [[S; Q_1, Q_2, Q_3]] \times_1 R_1 \times_2 R_2 \times_3 R_3 \]

can be regarded as a QR factorization of \( A \).

Write a \texttt{MATLAB} function \([Q,R] = \text{tensorQR}(A)\) that carries out this decomposition where \( A, Q \) and \( R \) are ttensors.
The Tucker Product Approximation Problem

**Definition**

Given \( A \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) and \( r \leq n \), determine

- \( S \in \mathbb{R}^{r_1 \times r_2 \times r_3} \) the "core tensor"
- \( U_1 \in \mathbb{R}^{n_1 \times r_1} \) orthonormal columns
- \( U_2 \in \mathbb{R}^{n_2 \times r_2} \) orthonormal columns
- \( U_3 \in \mathbb{R}^{n_3 \times r_3} \) orthonormal columns

such that \( \| A - \mathcal{X} \|_F \) is minimized where

\[
\mathcal{X} = [[S; U_1, U_2, U_3]] = \sum_{j=1}^{r} S(j) \cdot U_1(\cdot, j_1) \circ U_2(\cdot, j_2) \circ U_3(\cdot, j_3)
\]

We say that \( \mathcal{X} \) is a length-\( r \) Tucker tensor.

In the matrix case, we solve this problem by truncating \( A \)'s SVD.
The Truncated HOSVD

Definition

If

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

is the HOSVD of \( A \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) and \( r \leq n \), then

\[ A_r = \left[ [ S(1:r_1,1:r_2,1:r_3); U_1(:,1:r_1), U_2(:,1:r_2), U_3(:,1:r_3) ] \right] \]

is a truncated HOSVD of \( A \).

How good is \( \mathcal{X} = A_r \) as a minimizer of \( \| A - \mathcal{X} \|_F \)? Not optimal but...
Look at $\mathcal{A} \approx \mathcal{A}_r$ at the scalar level...

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

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

What can we say about the “thrown away” terms?
The Core Tensor $S$ is Graded

Recall from Lecture 4 that $\| S_k(j,:) \|$ is the $j$-th singular value of $A_k$. This means that

$$\| S(j,:) \|_F = \sigma_j(A_{(1)}) \quad j = 1:n_1$$
$$\| S(:,j,:) \|_F = \sigma_j(A_{(2)}) \quad j = 1:n_2$$
$$\| S(:,:,j) \|_F = \sigma_j(A_{(3)}) \quad j = 1:n_3$$

The entries in $S$ tend to get smaller as you move away from the $(1,1,1)$ entry.
**Problem 6.2.** Does this inequality hold?

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

Can you do better?

**Problem 6.3.** Show that

\[
|\mathcal{A}(i_1, i_2, i_3) - \mathcal{X}_r(i_1, i_2, i_3)| \leq \min\{\sigma_{r_1+1}(\mathcal{A}(1)), \sigma_{r_2+1}(\mathcal{A}(2)), \sigma_{r_3+1}(\mathcal{A}(3))\}
\]
Once again, the problem...

Given $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $r \leq n$, determine

$$S \in \mathbb{R}^{r_1 \times r_2 \times r_3} \quad \text{the “core tensor”}$$

$$U_1 \in \mathbb{R}^{n_1 \times r_1} \quad \text{orthonormal columns}$$

$$U_2 \in \mathbb{R}^{n_2 \times r_2} \quad \text{orthonormal columns}$$

$$U_3 \in \mathbb{R}^{n_3 \times r_3} \quad \text{orthonormal columns}$$

such that

$$\| A - [[S; U_1, U_2, U_3]] \|_F$$

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

is minimized.

Note that $\text{vec}(S)$ solves an ordinary least squares problem...
The Tucker Product Approximation Problem

Reformulation...

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

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

is minimized.
The Tucker Product Approximation Problem

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

$$\| (I - QQ^T)a \|_2^2 = \| a \|_2^2 - \| Q^T a \|_2^2$$
Reshaping...

\[
\| (U_3^T \otimes U_2^T \otimes U_1^T) \cdot \text{vec}(A) \| = \\
\| U_1^T \cdot A(1) \cdot (U_3 \otimes U_2) \|_F = \\
\| U_2^T \cdot A(2) \cdot (U_3 \otimes U_1) \|_F = \\
\| U_3^T \cdot A(3) \cdot (U_2 \otimes U_1) \|_F
\]

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

A Sequence of Three Linear Problems...

\[
\| (U_3^T \otimes U_2^T \otimes U_1^T) \cdot \text{vec}(A) \| = \\
\| U_1^T \cdot A(1) \cdot (U_3 \otimes U_2) \|_F \quad \Leftarrow \quad 1. \text{ Fix } U_2 \text{ and } U_3 \text{ and maximize with } U_1. \\
= \\
\| U_2^T \cdot A(2) \cdot (U_3 \otimes U_1) \|_F \quad \Leftarrow \quad 2. \text{ Fix } U_1 \text{ and } U_3 \text{ and maximize with } U_2. \\
= \\
\| U_3^T \cdot A(3) \cdot (U_2 \otimes U_1) \|_F \quad \Leftarrow \quad 3. \text{ Fix } U_1 \text{ and } U_2 \text{ and maximize with } U_3.
\]

These max problems are SVD problems...
Alternating Least Squares Framework

A Sequence of Three Linear Problems...

Repeat:

1. Compute the SVD $A(1) \cdot (U_3 \otimes U_2) = \tilde{U}_1 \Sigma_1 V_1^T$
   and set $U_1 = \tilde{U}_1(:,1:r_1)$.

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

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

Higher Order Orthogonal Iteration (HOOI)
Problem 6.4. Write a `MATLAB` function `X = MyTucker3(A,r,itmax)` that performs `itMax` steps of the HOOI algorithm to obtain a best length-`r` Tucker approximation to the order-3 tensor `A`. The output tensor `X` should be a ttensor. Use the truncated HOSVD to obtain an initial guess. Justify its use over a random starting guess. To make your implementation efficient, use the Kronecker product fact that if

\[ B = A \cdot (Y \otimes Z) \]

then `B(i,:)` is a reshaping of a matrix product that involves `Y`, `Z`, and a reshaping of `A(i,:)`.

Problem 6.5. How does `MyTucker3` behave if it is based on QR-with-column-pivoting instead of the SVD?
The Tucker Product Approximation Problem

Given $A \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and $r \leq n$, compute $S \in \mathbb{R}^{r_1 \times \cdots \times r_d}$ and orthonormal matrices $U_k \in \mathbb{R}^{n_k \times r_k}$ for $k = 1:d$ such that if

$$X = [[S; U_1, \ldots, U_d]] = \sum_{j=1}^{r} S(j) U_1(:,j_1) \circ \cdots \circ U_d(:,j_d)$$

then $\| A - X \|_F$ is minimized.
The Order-$d$ Case

The ALS Framework Involves a Sequence of $d$ SVD Problems...

Repeat:

for $k = 1:d$

Compute the SVD

$$A_{(k)} (U_d \otimes \cdots \otimes U_{k+1} \otimes U_{k-1} \otimes \cdots \otimes U_1) = \tilde{U}_k \Sigma_k V_k^T$$

and set $U_k = \tilde{U}_k(:,1:r_k)$

end

What about the choice of $r = [r_1, \ldots, r_d]$?
MATLAB Tensor Toolbox: The Function tucker_als

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\n',r,fit);
end

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

X = Tucker_als(A,r,’maxiter’,20,’tol’,.001)
Problem 6.6. Compare the efficiency of MyTucker3 and tucker_als.

Problem 6.7. Do cp_als(A,1) and tucker_als(A,1) return the same rank-1 tensor?
More on Tensor Rank

**k-Rank**

If $A \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and $1 \leq k \leq d$, then its $k$-rank is defined by

$$\text{rank}_k(A) = \text{rank}(A(e_k))$$

**Problem 6.8.** Show that if $A \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ then there exists

$$\mathcal{X} = [[S; U_1, \ldots, U_d]]$$

where for $k = 1:d$, $U_k \in \mathbb{R}^{n_k \times r_k}$ has orthonormal columns and $r_k = \text{rank}_k(A)$. This can be thought of as a “thin” HOSVD.
More on Tensor Rank

\[ \text{rank}(\text{Any Unfolding of } A) \leq \text{rank}(A) \]

Suppose \( U_k \in \mathbb{R}^{n_k \times r} \) and that

\[ A = \sum_{k=1}^{r} U_1(:, k) \circ \cdots \circ U_d(:, k). \]

If \( p \) is any permutation of \( 1:d \) and \( 1 \leq j < d \), then

\[
\text{tenmat}(A, p(1:j), p(j + 1:d)) = \\
\left[ U_{p(j)} \circ \cdots \circ U_{p(1)} \right] \left[ U_{p(d)} \circ \cdots \circ U_{p(j+1)} \right]^T \\
= \\
\sum_{k=1}^{r} (U_{p(j)}(:, k) \otimes \cdots \otimes U_{p(1)}(:, j)) (U_{p(d)}(:, k) \otimes \cdots \otimes U_{p(j+1)}(:, j))^T
\]
Problem 6.9. Does it follow that the “most square” unfolding of $A \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ has the highest rank?

Problem 6.10. Suppose $M$ is an unfolding of $A \in \mathbb{R}^{n_1 \times \cdots \times n_d}$. How might you construct a good ktensor or ttensor approximation to $A$ from the SVD of $M$?
The Tucker Approximation Problem for a given tensor $\mathcal{A}$ and a given integer vector $\mathbf{r}$, involves finding the nearest length-$r$ Tucker tensor to $\mathcal{A}$ in the Frobenius norm.

The alternating least squares framework is used by tucker_als to solve the Tucker approximation problem. It proceeds by solving a sequence of SVD problems.

The k-rank of a tensor is the matrix rank of its mode-$k$ unfolding.


