From Matrix to Tensor:
The Transition to Numerical Multilinear Algebra

Lecture 5. The CP Representation and Tensor Rank

Charles F. Van Loan

Cornell University

The Gene Golub SIAM Summer School 2010
Selva di Fasano, Brindisi, Italy

Where We Are

What is this Lecture About?

Sums of Rank-1 Tensors

The SVD of a matrix A expresses A as a very special sum of rank-1 matrices.

Let us do the same thing as much as possible with tensor $\mathcal{A}$.

This requires an understanding of (a) rank-1 tensors and their unfoldings, (b) the Kruskal tensor format, (c) the alternating least squares framework for multilinear sum-of-squares optimization, and (d) the notion of tensor rank.

We use the order-3 case to motivate the main ideas.

What is this Lecture About?

A Note on Terminology

The central decomposition in this lecture is the CP Decomposition.

It also goes by the name of the CANDECOMP/PARAFAC Decomposition.

CANDECOMP	=	Canonical Decomposition
PARAFAC	=	Parallel Factors Decomposition

Rank-1 Tensors (Order-3)

Definition

If $f \in \mathbb{R}^{n_1}$, $g \in \mathbb{R}^{n_2}$, and $h \in \mathbb{R}^{n_3}$, then

$$\mathcal{B} \;=\; f \circ g \circ h$$

is defined by

$$\mathcal{B}(i_1, i_2, i_3) \;=\; f(i_1)g(i_2)h(i_3).$$

The tensor $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a rank-1 tensor.

The Kronecker Product Connection...

$$\mathcal{B} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \circ \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} \circ \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \quad \Leftrightarrow \quad \begin{bmatrix} b_{111} \\ b_{211} \\ b_{121} \\ b_{221} \\ b_{131} \\ b_{231} \\ b_{112} \\ b_{212} \\ b_{122} \\ b_{222} \\ b_{132} \\ b_{232} \end{bmatrix} = \begin{bmatrix} f_1 g_1 h_1 \\ f_2 g_1 h_1 \\ f_1 g_2 h_1 \\ f_2 g_2 h_1 \\ f_1 g_3 h_1 \\ f_2 g_3 h_1 \\ f_1 g_1 h_2 \\ f_2 g_1 h_2 \\ f_1 g_2 h_2 \\ f_2 g_2 h_2 \\ f_1 g_3 h_2 \\ f_2 g_3 h_2 \end{bmatrix} = h \otimes g \otimes f$$

Rank-1 Tensors (Order-3)

The Modal Unfoldings...

If

$$\mathcal{B} = f \circ g \circ h = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \circ \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} \circ \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

then

$$\begin{aligned} \mathcal{B}_{(1)} &= \begin{bmatrix} f_1 g_1 h_1 & f_1 g_2 h_1 & f_1 g_3 h_1 & f_1 g_1 h_2 & f_1 g_2 h_2 & f_1 g_3 h_2 \\ f_2 g_1 h_1 & f_2 g_2 h_1 & f_2 g_3 h_1 & f_2 g_1 h_2 & f_2 g_2 h_2 & f_2 g_3 h_2 \end{bmatrix} \\ &= \begin{bmatrix} f_1 \cdot (h \otimes g)^T \\ f_2 \cdot (h \otimes g)^T \end{bmatrix} \\ &= f \otimes (h \otimes g)^T \end{aligned}$$

Rank-1 Tensors (Order-3)

The Modal Unfoldings...

If

$$\mathcal{B} = f \circ g \circ h = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \circ \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} \circ \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

then

$$\mathcal{B}_{(2)} = \begin{bmatrix} f_1 g_1 h_1 & f_2 g_1 h_1 & f_1 g_1 h_2 & f_2 g_1 h_2 \\ f_1 g_2 h_1 & f_2 g_2 h_1 & f_1 g_2 h_2 & f_2 g_2 h_2 \\ f_1 g_3 h_1 & f_2 g_3 h_1 & f_1 g_3 h_2 & f_2 g_3 h_2 \end{bmatrix}$$

$$= \begin{bmatrix} g_1 \cdot (h \otimes f)^T \\ g_2 \cdot (h \otimes f)^T \\ g_3 \cdot (h \otimes f)^T \end{bmatrix}$$

$$= g \otimes (h \otimes f)^T$$

Rank-1 Tensors (Order-3)

The Modal Unfoldings...

If

$$\mathcal{B} = f \circ g \circ h = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \circ \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} \circ \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

then

$$\begin{aligned} \mathcal{B}_{(3)} &= \begin{bmatrix} f_1 g_1 h_1 & f_2 g_1 h_1 & f_1 g_2 h_1 & f_2 g_2 h_1 & f_1 g_3 h_1 & f_2 g_3 h_1 \\ f_1 g_1 h_2 & f_2 g_1 h_2 & f_1 g_2 h_2 & f_2 g_2 h_2 & f_1 g_3 h_2 & f_2 g_3 h_2 \end{bmatrix} \\ &= \begin{bmatrix} h_1 \cdot (g \otimes f)^T \\ h_2 \cdot (g \otimes f)^T \end{bmatrix} \\ &= h \otimes (g \otimes f)^T \end{aligned}$$

Problem 5.1. Suppose $a \in \mathbb{R}^{n_1 n_2 n_3}$. Show how to compute $f \in \mathbb{R}^{n_1}$ and $g \in \mathbb{R}^{n_2}$ so that $\| a - h \otimes g \otimes f \|_2$ is minimized where $h \in \mathbb{R}^{n_3}$ is given. Hint. It's an SVD problem.

Problem 5.2. Given $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with positive entries, how would you choose $\mathcal{B} = f \circ g \circ h \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ so that

$$\phi(f, g, h) \;=\; \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \left| \log(\mathcal{A}(\mathbf{i})) \;-\; \log(\mathcal{B}(\mathbf{i})) \right|^2$$

is minimized.

Notation

Given $\lambda \in \mathbb{R}^r$, $F \in \mathbb{R}^{n_1 \times r}$, $G \in \mathbb{R}^{n_2 \times r}$, and $H \in \mathbb{R}^{n_3 \times r}$, we define $[[\,\lambda;\ F,\ G,\ H\,]] \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ by

$$[[\,\lambda,\ F,\ G,\ H\,]] \;=\; \sum_{j=1}^{r} \lambda_j \cdot F(:,j) \circ G(:,j) \circ H(:.j)$$

A weighted sum of rank-1 tensors where the vectors that specify the rank-1's are columns of the matrices F, G, and H.

Kruskal Form

We say that $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is in Kruskal form if

$$\mathcal{X} \;=\; [[\, \lambda;\; F,\; G,\; H \,]].$$

where $\lambda \in \mathbb{R}^r$, $F \in \mathbb{R}^{n_1 \times r}$, $G \in \mathbb{R}^{n_2 \times r}$, and $H \in \mathbb{R}^{n_3 \times r}$.

Can we write a given tensor $\mathcal{A}$ as an illuminating sum of rank-1 tensors? I.e., Given $\mathcal{A}$, can we find $\mathcal{X} = [[\, \lambda;\; F,\; G,\; H \,]]$ so that $\mathcal{A} \approx \mathcal{X}$ in some meaningful way?

The CP Representation (Order-3)

Equivalent Formulations...

If $\mathcal{X} = [[\,\lambda,\ F,\ G,\ H\,]] \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then

$$\mathcal{X}(i_1, i_2, i_3) = \sum_{j=1}^{r} \lambda_j \cdot F(i_1, j) \cdot G(i_2, j) \cdot H(i_3, j))$$

$$\mathrm{vec}(\mathcal{X}) = \sum_{j=1}^{r} \lambda_j \cdot H(:, j) \otimes G(:, j) \otimes F(:, j)$$

MATLAB Tensor Toolbox: **ktensor Set-Up**

```
n = [5 8 3]; r = 4;
% Set up a random, length-r ktensor...
F = randn(n(1),r); G = randn(n(2),r);
H = randn(n(3),r); lambda = ones(r,1);
X = ktensor(lambda,{F,G,H});
Fsize = size(X.U{1}); Gsize = size(X.U{2});
Hsize = size(X.U{3});
L = length(X.lambda); s = size(X);
```

A ktensor is a structure with two fields that is used to represent a tensor in Kruskal form. In the above, `X.lambda` houses the vector of weights while `X.U` is a cell array of the matrices that define the tensor `X`.

Variable	Value
`Fsize`	[5,4]
`Gsize`	[8,4]
`Hsize`	[3,4]
`L`	4
`s`	[5 8 3]

MATLAB Tensor Toolbox: **Norm of a ktensor**

```
function alfa = normKruskal(X)
% X is a ktensor and alfa is the Frobenius norm
% of the tensor it represents.
N = prod(size(X));
% Create a multidimensional array that houses
% the Kruskal tensor...
Xarray = double(X);
% Reshape as a vector and compute its 2-norm...
alfa = norm(reshape(Xarray,N,1));
```

Problem 5.3. Write a MATLAB function `Y = normalize(X)` that takes a takes a ktensor `X` and returns a ktensor `Y` with the property that (a)

```
Y.U{j}(:,k) = X.U{j}(:,k)/norm(X.U{j}(:,k))
```

for all appropriate values of `k` and `j` and (b) `double(X) = double(Y)`.

The CP Representation (Order-3)

The CP Approximation Problem

Given $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and r, determine $\lambda \in \mathbb{R}^r$, $F \in \mathbb{R}^{n_1 \times r}$, $G \in \mathbb{R}^{n_2 \times r}$, and $H \in \mathbb{R}^{n_3 \times r}$ so that

$$\mathcal{A} \approx [[\,\lambda;\ F,\ G,\ H\,]] = \mathcal{X}.$$

Using Least Squares...

Choose λ, F, G, and H so that

$$\|\,\mathcal{A} - \mathcal{X}\,\|_F^2 = \left\| \mathrm{vec}(\mathcal{A}) - \sum_{j=1}^{r} \lambda_j \cdot H(:,j) \otimes G(:,j) \otimes F(:,j) \right\|_2^2$$

is minimized.

A multilinear optimization problem. Reshape using the Khatri-Rao Product...

The Khatri-Rao Product

Definition

If

$$\begin{aligned} B &= \left[\ b_1 \mid \cdots \mid b_r\ \right] \in \mathbb{R}^{n_1 \times r} \\ C &= \left[\ c_1 \mid \cdots \mid c_r\ \right] \in \mathbb{R}^{n_2 \times r} \end{aligned}$$

then the Khatri-Rao product of B and C is given by

$$B \odot C = \left[\ b_1 \otimes c_1 \mid \cdots \mid b_r \otimes c_r\ \right].$$

Note that $B \odot C \in \mathbb{R}^{n_1 n_2 \times r}$.

The Khatri-Rao Product

"Fast" Property 1.

If $B \in \mathbb{R}^{n_1 \times r}$ and $C \in \mathbb{R}^{n_2 \times r}$, then

$$(B \odot C)^T (B \odot C) \; = \; (B^T B) .* (C^T C)$$

where ".$*$" denotes pointwise multiplication.

Problem 5.4. Prove this property using the Kronecker product facts (i) $(W \otimes X)(Y \otimes Z) = WY \otimes XZ$ and (ii) $(W \otimes X)^T = W^T \otimes X^T$. How many flops are required?

The Khatri-Rao Product

"Fast" Property 2.

If

$$B = \left[\, b_1 \mid \cdots \mid b_r \,\right] \in \mathbb{R}^{n_1 \times r}$$

$$C = \left[\, c_1 \mid \cdots \mid c_r \,\right] \in \mathbb{R}^{n_2 \times r}$$

$x \in \mathbb{R}^{n_1 n_2}$, and $y = (B \odot C)^T x$, then

$$y = \begin{bmatrix} c_1^T X b_1 \\ \vdots \\ c_r^T X b_r \end{bmatrix} \qquad X = \text{reshape}(x, n_2, n_1)$$

Problem 5.5. Prove this property using $\text{vec}(YXW^T) = (W \otimes Y) \cdot \text{vec}(X)$. How many flops are required?

Problem 5.6. Complete the following function so that it performs as specified.

```
function x = KRLS(B,C,d)
% B is n1-by-1, C is n2-by-1, and d is n1*n2-by-1
% x minimizes norm(Ax - d,2) where A is the Khatri-Rao
% product of B and C.
```

Use the method of normal equations and assume that A has full column rank. Is there an equally efficient way to solve the problem via the QR factorization of A.

Unfolding Tensors in the Kruskal Form

Given $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ our goal is to minimize

$$\| \mathcal{A} - \mathcal{X} \|_F = \| \mathcal{A}_{(k)} - \mathcal{X}_{(k)} \|_F$$

where

$$\mathcal{X} = [[\, \lambda;\ F,\ G,\ H \,]] = \sum_{j=1}^{r} \lambda_j \cdot f_j \circ g_j \circ h_j$$

with

$$F = [\ f_1 \mid \cdots \mid f_r\ ] \quad G = [\ g_1 \mid \cdots \mid g_r\ ] \quad H = [\ h_1 \mid \cdots \mid h_r\ ]$$

So what do the modal unfoldings of $\mathcal{X}$ look like?

It will be a sum of rank-1 tensor unfoldings...

The CP Representation (Order-3)

Unfolding Tensors in the Kruskal Form

Since $\mathcal{B} = f \circ g \circ h$ implies

$$\begin{aligned}
\mathcal{B}_{(1)} &= f \otimes (h \otimes g)^T \\
\mathcal{B}_{(2)} &= g \otimes (h \otimes f)^T \\
\mathcal{B}_{(3)} &= h \otimes (g \otimes f)^T
\end{aligned}$$

we have

$$\begin{aligned}
\mathcal{X}_{(1)} &= \sum_{j=1}^{r} \lambda_j \cdot f_j \otimes (h_j \otimes g_j)^T = F \cdot \mathrm{diag}(\lambda_j) \cdot (H \odot G)^T \\
\mathcal{X}_{(2)} &= \sum_{j=1}^{r} \lambda_j \cdot g_j \otimes (h_j \otimes f_j)^T = G \cdot \mathrm{diag}(\lambda_j) \cdot (H \odot F)^T \\
\mathcal{X}_{(3)} &= \sum_{j=1}^{r} \lambda_j \cdot h_j \otimes (g_k \otimes f_j)^T = H \cdot \mathrm{diag}(\lambda_j) \cdot (G \odot F)^T
\end{aligned}$$

The CP Representation (Order-3)

The Alternating LS Solution Framework...

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

$$=$$

$$\| \mathcal{A}_{(1)} - F \cdot \mathrm{diag}(\lambda_j) \cdot (H \odot G)^T \|_F \quad \Leftarrow$$

1. Fix G and H and improve λ and F.

$$=$$

$$\| \mathcal{A}_{(2)} - G \cdot \mathrm{diag}(\lambda_j) \cdot (H \odot F)^T \|_F \quad \Leftarrow$$

2. Fix F and H and improve λ and G.

$$=$$

$$\| \mathcal{A}_{(3)} - H \cdot \mathrm{diag}(\lambda_j) \cdot (G \odot F)^T \|_F \quad \Leftarrow$$

3. Fix F and G and improve λ and H.

The Alternating LS Solution Framework

Repeat:

1. Let $\tilde{F}$ minimize $\| \mathcal{A}_{(1)} - \tilde{F} \cdot (H \odot G)^T \|_F$ and for $j = 1{:}r$ set

$$\lambda_j = \| \tilde{F}(:,j) \|_2 \quad \text{and} \quad F(:,j) \; = \; \tilde{F}(:,j)/\lambda_j.$$

2. Let $\tilde{G}$ minimize $\| \mathcal{A}_{(2)} - \tilde{G} \cdot (H \odot F)^T \|_F$ and for $j = 1{:}r$ set

$$\lambda_j = \| \tilde{G}(:,j) \|_2 \quad \text{and} \quad G(:,j) \; = \; \tilde{G}(:,j)/\lambda_j.$$

3. Let $\tilde{H}$ minimize $\| \mathcal{A}_{(3)} - \tilde{H} \cdot (G \odot F)^T \|_F$ and for $j = 1{:}r$ set

$$\lambda_j = \| \tilde{H}(:,j) \|_2 \quad \text{and} \quad H(:,j) \; = \; \tilde{H}(:,j)/\lambda_j.$$

These are linear least squares problems. The columns of F, G, and H are normalized.

Solving the LS Problems

The solution to

$$\min_{\tilde{F}} \| \mathcal{A}_{(1)} - \tilde{F} \cdot (H \odot G)^T \|_F \;=\; \min_{\tilde{F}} \| \mathcal{A}_{(1)}^T - (H \odot G)\tilde{F}^T \|_F$$

can be obtained by solving the normal equation system

$$(H \odot G)^T (H \odot G)\tilde{F}^T \;=\; (H \odot G)^T \mathcal{A}_{(1)}^T$$

Can be solved efficiently by exploiting the ideas in Problem 5.6.

Problem 5.7. Write a MATLAB function `X = MyKruskal(A,r,itMax)` that takes an order-3 tensor `A` and returns a ktensor `X` with the property that $\mathcal{A} \approx \mathcal{X}$. $\mathcal{X} = [[\, \lambda;\ F,\ G,\ H \,]]$ should be obtained by applying the following improvement steps `itMax` times:

1. Solve $(H \odot G)^T (H \odot G)\tilde{F}^T = (H \odot G)^T \mathcal{A}_{(1)}^T$ and for $j = 1{:}r$ set

$$\lambda_j = \| \tilde{F}(:,j) \|_2 \quad \text{and} \quad F(:,j) = \tilde{F}(:,j)/\lambda_j.$$

2. Solve $(H \odot F)^T (H \odot F)\tilde{G}^T = (H \odot F)^T \mathcal{A}_{(2)}^T$ and for $j = 1{:}r$ set

$$\lambda_j = \| \tilde{G}(:,j) \|_2 \quad \text{and} \quad G(:,j) = \tilde{G}(:,j)/\lambda_j.$$

3. Solve $(G \odot F)^T (G \odot F)\tilde{H}^T = (G \odot F)^T \mathcal{A}_{(3)}^T$ and for $j = 1{:}r$ set

$$\lambda_j = \| \tilde{H}(:,j) \|_2 \quad \text{and} \quad H(:,j) = \tilde{H}(:,j)/\lambda_j.$$

Choose the initial F, G, and H randomly unless you can think of something more clever.

MATLAB Tensor Toolbox: **The Function** cp_als

```
n = [ 5 6 7 ]; rmax = 35;
% Generate a random tensor...
A = tenrand(n);
for r = 1:rmax
   % Find the closest length-r ktensor...
   X = cp_als(A,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 cp_als returns a ktensor. Default values for the number of iterations and the termination criteria can be modified:

```
X = cp_als(A,r,'maxiters',20,'tol',.001)
```

Problem 5.8. Compare the efficiency of `MyKruskal` and `cp_als`.

The CP Representation: General Order

Rank-1 Tensors: Definition

If $u_k \in \mathbb{R}^{n_k}$ for $k = 1{:}d$, then

$$\mathcal{B} \;=\; u_1 \circ u_2 \circ \cdots \circ u_d,$$

is defined by

$$\mathcal{B}(i_1, \ldots, i_d) \;=\; u_1(i_1) \cdot u_2(i_2) \cdots u_d(i_d)$$

is a rank-1 tensor. Note that $\mathcal{B} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$.

Rank-1 Tensors: Modal Unfoldings

If

$$\mathcal{B} \;=\; u_1 \circ u_2 \circ \cdots \circ u_d\,,$$

then

$$\text{vec}(\mathcal{B}) \;=\; u_d \otimes \cdots \otimes u_2 \otimes u_1$$

and

$$\mathcal{B}_{(k)} \;=\; u_k \otimes \left(u_d \otimes \cdots \otimes u_{k+1} \otimes u_{k-1} \otimes \cdots \otimes u_1\right)^T.$$

Problem 5.9. Suppose $\mathcal{B} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is the rank-1 tensor defined above. Characterize the unfolding $M = \texttt{tenmat}(B, [1{:}p], [p+1{:}d])$ where $1 \leq p < d$.

The CP Representation: General Order

Notation

Given $\lambda \in \mathbb{R}^r$ and matrices $U_1, \ldots, U_d$ with unit column norms, define

$$[[\, \lambda; U_1, \ldots, U_d \,]] \;=\; \sum_{j=1}^{r} \lambda_j \cdot U_1(:,j) \circ \cdots \circ U_d(:,j)$$

Assume that $U_k \in \mathbb{R}^{n_k \times r}$.

A weighted sum of rank-1 tensors where the vectors that specify the rank-1's are columns of the matrices $U_1, \ldots, U_d$.

The Kruskal Form

We say that $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is in Kruskal form if

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

where $\lambda \in \mathbb{R}^r$ and $U_k \in \mathbb{R}^{n_k \times r}$ for $k = 1{:}d$.

Can we write a given tensor $\mathcal{A}$ as an illuminating sum of rank-1 tensors? I.e., Given $\mathcal{A}$, can we find a tensor $\mathcal{X}$ in Kruskal form so that $\mathcal{A} \approx \mathcal{X}$ in some meaningful way?

The CP Representation: General Order

Equivalent Formulations...

If $\mathcal{X} = [[\, \lambda;\ U_1, \ldots, U_d \,]] \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, then

$$\mathcal{X}(i_1, \ldots, i_d) = \sum_{j=1}^{r} \lambda_j \cdot U_1(i_1, j) \cdots U_d(i_d, j)$$

$$\mathrm{vec}(\mathcal{X}) = \sum_{j=1}^{r} \lambda_j \cdot U_d(:, j) \otimes \cdots \otimes U_1(:, j)$$

MATLAB Tensor Toolbox: **ktensor Operations**

```
  function X = KruskalRandn(n,r)
% Creates a random order-d, length-r
% ktensor having size determined
% by the length-d integer vector n. The
% columns of X.U{1},...,X.U{r} have unit
% 2-norm.
  U = cell(r,1);
  lambda = ones(r,1);
  for k=1:d
     U{k} = randn(n(k),r);
  end
  X0 = ktensor(lambda,U);
  X = arrange(X0);
```

The function `arrange` normalizes the columns of the matrices that define `X0` so that they have unit length. The weight vector is adjusted so that `double(X)` and `double(X0)` have the same value.

Problem 5.10. Implement a function `Y = MyArrange(X)` that normalizes a ktensor `X` in the same way as `arrange`.

The CP Representation: General Order

Unfolding Tensors in Kruskal Form

If

$$\mathcal{X} \;=\; [[\,\lambda;\; U_1, \ldots, U_d]] \;=\; \sum_{j=1}^{r} \lambda_j \cdot U_1(:,j) \circ \cdots \circ U_d(:,j)$$

then

$$\mathcal{X}_{(k)} \;=\; U_k \cdot \mathrm{diag}(\lambda_i) \cdot (U_d \odot \cdots \odot U_{k+1} \odot U_{k-1} \odot \cdots \odot U_1)^T$$

Note that Khatri-Rao products can be sequenced:

$$F \odot G \odot H \;=\; (F \odot G) \odot H \;=\; F \odot (G \odot H).$$

The CP Representation: General Case

The CP Approximation Problem

Given $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and r, determine

$$\mathcal{X} \; = \; [[\, \lambda;\; U_1, \ldots, U_d \,]] \in \mathbb{R}^{n_1 \times \cdots \times n_d}$$

so that

$$\| \, \mathcal{A} \; - \; \mathcal{X} \, \|_F \; = \; \| \, \mathcal{A}_{(k)} \; - \; \mathcal{X}_{(k)} \, \|_F$$

is minimized where

$$\mathcal{X}_{(k)} \; = \; U_k \cdot \mathrm{diag}(\lambda_i) \cdot (U_d \odot \cdots \odot U_{k+1} \odot U_{k-1} \odot \cdots \odot U_1)^T$$

The CP Representation: General Case

The Alternating Least Squares Framework

for $k = 1{:}d$

Fix $U_1, \ldots, U_{k-1}, U_{k+1}, \ldots, U_d$.

Improve λ and U_k by minimizing

$$\| \mathcal{A}_k - \tilde{U}_k \left(U_d \odot \cdots \odot U_{k+1} \odot U_{k-1} \odot \cdots \odot U_1 \right) \|$$

for $j = 1{:}r$

$$\lambda_j = \| \tilde{U}_k(:,j) \|$$

$$U_k(:,j) = \tilde{U}_k(:,j)/\lambda_j$$

end

end

Problem 5.11. Assume that $B_k \in \mathbb{R}^{n_k \times r}$ for $k = 1{:}d$. (a) Show how to compute

$$\tilde{d} = (B_1 \odot \cdots \odot B_d)^T d$$

efficiently where $d \in \mathbb{R}^N$ with $N = n_1 \cdots n_d$. (b) Show how to compute efficiently

$$C = (B_1 \odot \cdots \odot B_d)^T (B_1 \odot \cdots \odot B_d).$$

(c) Write a MATLAB function `x = KRLS(B,d)` that solves the least squares problem

$$\min \| \, (B_1 \odot \cdots \odot B_d)x - d \, \|$$

Assume that `B` is a cell array that houses the matrices $B_1, \ldots, B_d$. (See Problem 5.6.)

Problem 5.12. Refer to Problem 5.7 and develop a general order version of `X = MyKruskal(A,r,itMax)` based on the preceding alternating least squares framework. Take advantage of the ideas in Problem 5.11. Compare your implementation with `cp_als`.

What About r?

In the CP approximation problem we have assumed that r, the length of the approximating ktensor, is given:

$$\mathcal{A} \approx \mathcal{X} = \sum_{j=1}^{r} \lambda_j U_1(:,j) \circ \cdots \circ U_d(:,j)$$

We can think of $\mathcal{X}$ as a rank-r approximation to $\mathcal{A}$.

Tensor Rank

Departure from Matrix Case...

Suppose

$$\mathcal{X}_r \;=\; \sum_{j=1}^{r} \lambda_j U_1(:,j) \circ \cdots \circ U_d(:,j)$$

is the best rank-r approximation of $\mathcal{A}$ and

$$\mathcal{X}_{r+1} \;=\; \sum_{j=1}^{r+1} \lambda_j \hat{U}_1(:,j) \circ \cdots \circ \hat{U}_d(:,j)$$

is the best rank-$(r+1)$ approximation of $\mathcal{A}$.

IT DOES NOT FOLLOW THAT $\mathcal{X}_{r+1}$ is $\mathcal{X}_r$ plus a rank-1.

In this regard, the best Kruskal approximation is not SVD-like.

Tensor Rank

Definition

The rank of a tensor $\mathcal{A}$ is the smallest number of rank-1 tensors that sum to $\mathcal{A}$.

This agrees with the definition for matrices. But there are some differences that make tensor rank a more complicated issue...

Anomaly 1

The largest rank attainable for an n_1-by-...-n_d tensor is called the maximum rank. It is *not* a simple formula that depends on the dimensions $n_1, \ldots, n_d$. Indeed, its precise value is only known for small examples.

Maximum rank does not equal $\min\{n_1, \ldots, n_d\}$ *unless* $d \leq 2$.

Tensor Rank

Anomaly 2

If the set of rank-k tensors in $\mathbb{R}^{n_1 \times \cdots \times n_d}$ has positive Lebesgue measure, then k is a typical rank.

Size	Typical Ranks
$2 \times 2 \times 2$	2,3
$3 \times 3 \times 3$	4
$3 \times 3 \times 4$	4,5
$3 \times 3 \times 5$	5,6

For n_1-by-n_2 matrices, typical rank and maximal rank are both equal to the small of n_1 and n_2.

Tensor Rank

Anomaly 3

The rank of a particular tensor over the real field may be different than its rank over the complex field.

Anomaly 4

A tensor with a given rank may be approximated with arbitrary precision by a tensor of lower rank. Such a tensor is said to be degenerate.

Problem 5.13. For various small choices for $\mathbf{n} = [n_1, \ldots, n_d]$, see if you can discover the typical rank possibilities by using `cp_als`. In particular, for a randomly generated $\mathcal{A}$, compute the smallest r such that

$$\| \mathcal{A} - \texttt{cp\_als(A,r)} \|_F \leq 10^{-6}$$

By running a sufficient number of examples, see what you can deduce about the typical rank of tensors in $\mathbb{R}^{n_1 \times \cdots \times n_d}$.

Summary of Lecture 5.

Key Words

- An order-d rank-1 tensor is the outer product of d vectors.
- A tensor in Kruskal form has length r, if it is the sum of r rank-1 tensors. For an order-d tensor, the vectors that make up the rank-1's are specified as columns from d matrices.
- The CP approximation problem for a given tensor $\mathcal{A}$ and a given integer r involves finding the nearest length-r ktensor to $\mathcal{A}$ in the Frobenius norm.
- The alternating least squares framework is used by `cp_als` to solve the CP approximation problem. It proceeds by solving a sequence of structured linear least squares problems.