Four Talks on Tensor Computations

3. More Decompositions and Iterations

Charles F. Van Loan

Cornell University

SCAN Seminar

November 10, 2014

What is this Lecture About?

Still 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 alternating least squares framework for multilinear sum-of-squares optimization, and (c) the notion of tensor rank.

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

The Tucker Representation

Given...

$\mathcal{S} \in \mathbb{R}^{r \times r \times r}$

$U_1 \in \mathbb{R}^{n_1 \times r}$, $U_2 \in \mathbb{R}^{n_2 \times r}$, $U_3 \in \mathbb{R}^{n_3 \times r}$ (orthonormal cols)

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

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

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

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

The CP Representation

Given...

$\lambda_1, \cdots, \lambda_r$

$F \in \mathbb{R}^{n_1 \times r}$, $G \in \mathbb{R}^{n_2 \times r}$, $H \in \mathbb{R}^{n_3 \times r}$ (unit 2-norm columns)

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

$$
\begin{aligned}
\mathcal{X} &= \sum_{j=1}^{r} \lambda_j \cdot F(:,j) \circ G(:,j) \circ H(:,j) \\
\mathcal{X}(i_1, i_2, i_3) &= \sum_{j=1}^{r} \lambda_j \cdot F(i_1,j) \cdot G(i_2,j) \cdot H(i_3,j)) \\
\text{vec}(\mathcal{X}) &= \sum_{j=1}^{r} \lambda_j \cdot H(:,j) \otimes G(:,j) \otimes F(:,j)
\end{aligned}
$$

Tucker Vs. CP

Tucker:

$$\mathcal{X} = \sum_{j_1=1}^{r} \sum_{j_2=1}^{r} \sum_{j_3=1}^{r} \mathcal{S}(j_1, j_2, j_3) \cdot U_1(:, j_1) \circ U_2(:, j_2) \circ U_3(:, j_3)$$

CP:

$$\mathcal{X} = \sum_{j=1}^{r} \lambda_j$$

A Note on Terminology

The "CP" Decomposition

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

CANDECOMP = Canonical Decomposition

PARAFAC = Parallel Factors Decomposition

The CP Approximation Problem

Definition

Given: $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and r

Determine: $\lambda \in \mathbb{R}^r$ and $F \in \mathbb{R}^{n_1 \times r}$, $G \in \mathbb{R}^{n_2 \times r}$, and $H \in \mathbb{R}^{n_3 \times r}$ (with unit 2-norm columns) so that if

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

then

$$\| \, \mathcal{A} - \mathcal{X} \, \|_F^2$$

is minimized.

A multilinear optimization problem.

The CP Approximation Problem (Unfolded Form)

Definition

Given: $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and r

Determine: $\lambda \in \mathbb{R}^r$ and $F \in \mathbb{R}^{n_1 \times r}$, $G \in \mathbb{R}^{n_2 \times r}$, and $H \in \mathbb{R}^{n_3 \times r}$ (with unit 2-norm columns) so that if

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

then

$$\| \, \mathcal{A}_{(k)} - \mathcal{X}_{(k)} \, \|_F^2 \qquad k = 1, 2, 3$$

is minimized.

What do the modal unfoldings of the $F(:,j) \circ G(:,j) \circ F(:,j)$ look like?

Rank-1 Tensors

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.

Rank-1 Tensors

The Kronecker Product Connection...

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

The Modal Unfoldings of these things are Kronecker Products with unit rank

Unfolding a Rank-1 Tensor

The Mode-1 Unfolding...

If

$$\mathcal{B} \;=\; f \circ g \circ h \;=\; \left[\begin{array}{c} f_1 \\ f_2 \end{array}\right] \circ \left[\begin{array}{c} g_1 \\ g_2 \\ g_3 \end{array}\right] \circ \left[\begin{array}{c} h_1 \\ h_2 \end{array}\right]$$

then

$$\begin{aligned}
\mathcal{B}_{(1)} \;&=\; \left[\begin{array}{cccccc} f_1 g_1 h_1 & f_1 g_2 h_1 & f_1 g_3 h_1 & f_1 g_1 h_2 & f_1 g_2 h_2 & f_1 g_3 h_2 \\ f_2 g_1 h_1 & f_2 g_2 h_1 & f_2 g_3 h_1 & f_2 g_1 h_2 & f_2 g_2 h_2 & f_2 g_3 h_2 \end{array}\right] \\
&=\; \left[\begin{array}{c} f_1 \cdot (h \otimes g)^T \\ f_2 \cdot (h \otimes g)^T \end{array}\right] \\
&=\; f \otimes (h \otimes g)^T
\end{aligned}$$

Unfolding a Rank-1 Tensor

The Mode-2 Unfolding...

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

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

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

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

Unfolding a Rank-1 Tensor

The Mode-3 Unfolding...

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

$$\mathcal{B}_{(3)} \;=\; \begin{bmatrix} f_1g_1h_1 & f_2g_1h_1 & f_1g_2h_1 & f_2g_2h_1 & f_1g_3h_1 & f_2g_3h_1 \\ f_1g_1h_2 & f_2g_1h_2 & f_1g_2h_2 & f_2g_2h_2 & f_1g_3h_2 & f_2g_3h_2 \end{bmatrix}$$

$$= \begin{bmatrix} h_1 \cdot (g \otimes f)^T \\ h_2 \cdot (g \otimes f)^T \end{bmatrix}$$

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

Unfolding the CP Representation

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

$$\mathcal{B}_{(1)} \;=\; f \otimes (h \otimes g)^T \qquad \mathcal{B}_{(2)} \;=\; g \otimes (h \otimes f)^T \qquad \mathcal{B}_{(3)} \;=\; h \otimes (g \otimes f)^T$$

and

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

we have

$$\begin{aligned}
\mathcal{X}_{(1)} &= \sum_{j=1}^{r} \lambda_j \cdot F(:,j) \otimes \left(H(:,j) \otimes G(:,j) \right)^T \\
\mathcal{X}_{(2)} &= \sum_{j=1}^{r} \lambda_j \cdot G(:,j) \otimes \left(H(:,j) \otimes F(:,j) \right)^T \\
\mathcal{X}_{(3)} &= \sum_{j=1}^{r} \lambda_j \cdot H(:,j) \otimes \left(G(:,j) \otimes F(:,j) \right)^T
\end{aligned}$$

Use the Khatri-Rao Product to Define the $\mathcal{X}_{(k)}$

Definition

If

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

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

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

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

"Column-wise KPs". Note that $B \odot C \in \mathbb{R}^{n_1 n_2 \times r}$.

The Khatri-Rao Product

Definition

If

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

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

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

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

$$B \otimes C = \left[\ b_1 \mid b_2 \mid b_3\ \right] \otimes \left[\ c_1 \mid c_2 \mid c_3\ \right]$$

$$= \left[\ b_1 \otimes c_1 \mid b_1 \otimes c_2 \mid b_1 \otimes c_3 \mid b_2 \otimes c_1 \mid b_2 \otimes c_2 \mid b_2 \otimes c_3 \mid b_3 \otimes c_1 \mid b_3 \otimes c_2 \mid b_3 \otimes c_3\ \right]$$

Unfolding the CP Representation

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

$$\mathcal{B}_{(1)} = f \otimes (h \otimes g)^T \qquad \mathcal{B}_{(2)} = g \otimes (h \otimes f)^T \qquad \mathcal{B}_{(3)} = h \otimes (g \otimes f)^T$$

and

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

we have

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

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

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

The CP Approximation Problem

The Alternating LS Solution Framework...

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

$$=$$

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

1. Fix G and H and improve λ and F.

$$=$$

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

2. Fix F and H and improve λ and G.

$$=$$

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

3. Fix F and G and improve λ and H.

The CP Approximation Problem

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.

The CP Approximation Problem

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 two properties of the Khatri-Rao product.

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.

The Khatri-Rao Product

"Fast" Property 2.

If

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

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

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

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

Overall: The Khatri-Rao LS Problem

Structure

Given $B \in \mathbb{R}^{n_1 \times r}$, $C \in \mathbb{R}^{n_2 \times r}$, and $b \in \mathbb{R}^{n_1 n_2}$, minimize

$$\| \; B \odot C)x \; - \; z \, \|_2$$

Data Sparse: An $n_1 n_2$-by-r LS problem defined by $O((n_1 + n_2)r)$ data.

Solution Procedure

1. Form $M = (B^T B). * (C^T C)$. $O((n_1 + n_2)r^2)$.
2. Cholesky: $M = LL^T$. $O(r^3)$.
3. Form $y = (B \odot C)^T$ using Property 2. $O(n_1 n_2 r)$.
4. Solve $Mx = y$. $O(r^2)$.

$$O(n_1 n_2 r) \quad \text{vs} \quad O((n_1 n_2 r^2)$$

Problem . Is there a fast QR procedure for

$$\min \| \ B \odot C)x \ - \ z \ \|_2$$

Let's Think About Tensor Rank

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 CP 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...

Tensor Rank

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 smaller 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.

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)
```

Now Some Variants and Alternatives

Our Plan...

Three examples will be presented.

Each comes with a multilinear optimization problem.

Each tells a "story" about the matrix-tensor connection.

Some come with pointers to unsolved problems.

Variant 1. (Subspace Constrained CP Approximation)

Given...

Tensor:	$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$
Length:	r
Subspaces:	$S_1 \subseteq \mathbb{R}^{n_1}$, $S_2 \subseteq \mathbb{R}^{n_2}$, and $S_3 \subseteq \mathbb{R}^{n_3}$,

Compute...

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

$$\left\| \mathcal{A} - \sum_{k=1}^{r} \lambda_k \cdot F(:,k) \circ G(:,k) \circ H(:,k) \right\|_F = \min$$

subject to the constraint that

$$\text{range}(F) \subseteq S_1 \qquad \text{range}(G) \subseteq S_2 \qquad \text{range}(H) \subseteq S_3$$

Variant 1. (Subspace Constrained CP Approximation)

Comments

By forcing the factor matrices to sit in predefined spaces, the "post game" factor analysis can be more meaningful.

The technique illustrated goes by the name of CANDELINC (Canonical Decomposition with Linear Constraints).

Variant 1. (Subspace Constrained CP Approximation)

Intuition via a Matrix Analog

Solve:

$$\min_{\text{range}(U_i) \subseteq S_i} \| A - U_1 \Sigma U_2^T \|_F$$

If $\text{range}(Q_i) = S_i$ and the Q_i have orthonormal columns, then

$$\min_{T_1, T_2} \| A - Q_1 T_1 \Sigma T_2^T Q_2^T \|_F$$

is equivalent and we are led to a reduced-in-size, projected problem

$$\min_{T_1, T_2} \| (Q_1^T A Q_2) \; - \; T_1 \Sigma T_2^T \|_F \; = \; \min$$

Set $U_i^{opt} \; = \; Q_i T_i^{(opt)}$.

Variant 1. (Subspace Constrained CP Approximation)

The Projected Problem

If Q_i is orthonormal and range$(Q_i) = S_i$, then our goal is to determine $\tilde{F}$, $\tilde{G}$, and $\tilde{H}$ so that

$$\left\| \mathcal{A} - \sum_{k=1}^{r} \lambda_k \cdot Q_1 \tilde{F}(:,k) \circ Q_2 \tilde{G}(:,k) \circ Q_3 \tilde{H}(:,k) \right\|_F = \min$$

Reshape

$$\min_{\tilde{F},\tilde{G},\tilde{H}} \| \operatorname{vec}(\mathcal{A}) - (Q_1 \otimes Q_2 \otimes Q_3)(\tilde{F} \odot \tilde{G} \odot \tilde{H})\lambda \|$$

Variant 1. (Subspace Constrained CP Approximation)

Equivalent

$$\min_{\tilde{F},\tilde{G},\tilde{H}} \| (Q_1 \otimes Q_2 \otimes Q_3)^T \mathrm{vec}(\mathcal{A}) - (\tilde{F} \odot \tilde{G} \odot \tilde{H})\lambda \|$$

An Unconstrained CP Approximation Problem...

$$\min_{\tilde{F},\tilde{G},\tilde{H}} \left\| \tilde{\mathcal{A}} - \sum_{k=1}^{r} \lambda_k \cdot \tilde{F}(:,k) \circ \tilde{G}(:,k) \circ \tilde{H}(:,k) \right\|_F = \min$$

where $\tilde{\mathcal{A}} = \mathcal{A} \times_1 Q_1^T \times_2 Q_2^T \times_3 Q_3^T$.

$$F^{(opt)} = Q_1 \tilde{F}^{(opt)} \qquad G^{(opt)} = Q_2 \tilde{G}^{(opt)} \qquad H^{(opt)} = Q_3 \tilde{H}^{(opt)}$$

Variant 2. (Block Term Decompositions)

Given...

$$\text{Tensor:} \quad \mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$$
$$\text{Parameters:} \quad p, \ [r_1 \ r_2 \ r_3\,]$$

Compute...

For $i = 1{:}p$ compute $F_i \in \mathbb{R}^{n_1 \times r_1}$, $G_i \in \mathbb{R}^{n_2 \times r_2}$, $H_i \in \mathbb{R}^{n_3 \times r_3}$, and $\mathcal{S}_i \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ so that

$$\left\| \mathcal{A} - \sum_{i=1}^{p} [[\, \mathcal{S}_i \,;\, F_i,\ G_i,\ H_i \,]] \right\|_F \ = \ \min$$

$$[[\, \mathcal{S}_i \,;\, F_i,\ G_i,\ H_i \,]] \ = \ \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{r}} \mathcal{S}_i(\mathbf{j}) \cdot F_i(:, j_1) \circ G_i(:, j_2 \circ H_i(:, j_3)$$

Variant 2. (Block Term Decompositions)

Comments

The CP and Tucker models are special cases.

Proving to be useful in tensor-based signal processing.

Variant 2. (Block Term Decompositions)

Reshaping ($p = 2$)

$$\left\| \mathcal{A} - \sum_{i=1}^{2} [[\, \mathcal{S}_i \,;\, F_i,\ G_i,\ H_i \,]] \right\|_F$$

$$=$$

$$\left\| \mathrm{vec}(\mathcal{A}) - \left[\; H_1 \otimes G_1 \otimes F_1 \;\middle|\; H_2 \otimes G_2 \otimes F_2 \; \right] \begin{bmatrix} \mathrm{vec}(\mathcal{S}_1) \\ \mathrm{vec}(\mathcal{S}_2) \end{bmatrix} \right\|$$

How would you solve for $\mathcal{S}_1$ and $\mathcal{S}_2$ given that the F's, G's and H's are fixed with orthonormal columns?

Variant 3. (Tucker With Compression)

Given...

Tensor: $\mathcal{A} \in \mathbb{R}^{m \times m \times m}$

Compute...

$U, V, W \in \mathbb{R}^{m \times m}$ (all orthogonal) and $\mathcal{S} \in \mathbb{R}^{m \times m \times m}$ so that

$$\mathcal{A} = \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{m}} \mathcal{S}(\mathbf{j}) \cdot U(:, j_1) \circ V(:, j_2) \circ W(:, j_3)$$

and the core tensor $\mathcal{S}$ has as much mass on the diagonal as possible.

Trying to mimic the SVD's ability to compress data.

Variant 3. (Tucker With Compression)

Comments

Trying to mimic the SVD's ability to compress data.

Jacobi SVD for matrix minimizes $\text{off}(A) = \sum_{i \neq j} a_{ij}^2$.

Develop a tensor analog that maximizes what is on $\text{diag}(a_{111}, \ldots, a_{nnn})$.

Metrics (Both would yield SVD in the order-2 (Matrix) Case

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

$$\psi(\mathcal{A}) = \sum_{i=1}^{n} a_{iii}^2$$

Variant 3. (Tucker With Compression)

Reshaping

$$\mathcal{A} = \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{m}} \mathcal{S}(\mathbf{j}) \cdot U(:, j_1) \circ V(:, j_2) \circ W(:, j_3)$$

$$=$$

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

Variant 3. (Tucker With Compression)

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:

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

Variant 3. (Tucker With Compression)

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.*

Variant 3. (Tucker With Compression)

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.

Since

$$\sigma_{111} + \sigma_{222} = \begin{bmatrix} c_\beta \\ s_\beta \end{bmatrix}^T \begin{bmatrix} s_{111} + s_{222} & -s_{211} + s_{122} \\ -s_{121} + s_{212} & s_{221} + s_{112} \end{bmatrix} \begin{bmatrix} c_\alpha \\ s_\alpha \end{bmatrix}$$

the solution involves computing the svd of this 2x2. The sought after unit-vectors are the left and right singular vectors associated with the largest singular value

Tucker and CP as "Simultaneous SVDS" of Slices

Tucker as a Collection of Simultaneous SVDS

If

$$\mathcal{A} = \sum_{i_1=1}^{r}\sum_{i_2=1}^{r}\sum_{i_3=1}^{r} \mathcal{S}(i_1,i_2,i_3)U(:,i_1)\circ V(:,i_2)\circ W(:,i_3)$$

then

$$\mathcal{A}(:,:,k) = VS_kU^T \qquad S_k(i_1,i_2) = \sum_{i_3=1}^{r}\mathcal{S}(i_1,i_2,i_3)W(k,i_3)$$

CP as a Collection of Simultaneous SVDS

If

$$\mathcal{A} = \sum_{j=1}^{r}\lambda_j F(:,j)\circ G(:,j)\circ H(:,j)$$

then

$$\mathcal{A}(:,:,k) = G\cdot \mathrm{diag}(\lambda_j H(k,j))\cdot F^T$$

Collection-Driven Decompositions

Two Examples

PARAFAC2

Higher-Order GSVD

PARAFAC2

Given...

$$\begin{array}{rl} \text{Matrices:} & A_k \in \mathbb{R}^{n_k \times m},\ k = 1{:}N \\ \text{Length:} & r \end{array}$$

Compute...

$$\begin{array}{ll} Q_k \in \mathbb{R}^{n_k \times r} & \text{orthonormal columns, } k = 1{:}N \\ S_k \in \mathbb{R}^{r \times r} & \text{diagonal, } k = 1{:}N \\ H \in \mathbb{R}^{r \times r} & \\ V \in \mathbb{R}^{m \times r} & \end{array}$$

so that

$$\sum_{k=1}^{N} \| A_k - Q_k H S_k V^T \|_F^2 = \min$$

PARAFAC2

Comments

A kind of simultaneous diagonalization.

Because the A_k vary in size, it is not possible to reshape the computation as a tensor computation.

Nevertheless, ALS methods are applicable.

The ideas presented are similar to the PARFAC2 method.

PARAFAC2

Improving Q_k (Orthonormal)

$$\sum_{k=1}^{N} \| \; A_k \; - \; Q_k H S_k V^T \; \|_F^2$$

The problem of minimizing $\| \; Y - QZ \; \|_F$ where Q has orthonormal columns is solved by computing the SVD of YZ^T and building Q from the left singular vectors.

Do this for $k = 1{:}N$ with $Y = A_k$ and $Z = HS_k V^T$.

PARAFAC2

Improving S_k (Diagonal)

$$\sum_{k=1}^{N} \| \; A_k \; - \; Q_k H S_k V^T \; \|_F^2$$

The problem of minimizing $\| \; Y - WSZ^T \; \|_F$ with respect to $S = \text{diag}(s_i)$ is equivalent to minimizing

$$\| \; \text{vec}(Y) - (Z \odot W)\, s \; \|$$

This type of structured LS problem showed up in the ALS method for CP approximation.

HOGSVD

What We Are Given...

Data matrices $A_1, \ldots, A_N$ each with full column rank equal to n

What We Want...

Expose common features in $\{A_1, \ldots, A_N\}$ by computing a simultaneous diagonalization of the form

$$A_k = U_k \Sigma_k V^T \qquad k = 1{:}N$$

where the Σ_k are diagonal, the U_k have unit 2-norm columns, and **V is nonsingular and carefully chosen.**

HOGSVD

V is the Eigenvector Matrix for

$$S_N = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left((A_i^T A_i)(A_j^T A_j)^{-1} + (A_j^T A_j)(A_i^T A_i)^{-1} \right).$$

This matrix is diagonalizable.

The eigenvalues of this matrix are real and ≥ 1.

The eigenvectors associated with the $\lambda = 1$ eigenvalues relate to certain singular vectors of the A_k.

The HOGSVD

Input: $A_k \in \mathbb{R}^{m_k \times n} \quad k = 1{:}N$

The Computation...

1. $V^{-1} S_N V = \text{diag}(\lambda_i)$
2. For $k = 1{:}N$ compute

$$A_k V^{-T} \;=\; U_k \Sigma_k$$

where the U_k have unit 2-norm columns and the Σ_k are diagonal.

Output: $A_k = U_k \Sigma_k V^T \;=\; \sum_{i=1}^{n} \sigma_i^{(k)} u_i^{(k)} v_i^T$

The HOGSVD

The eigenvalues of S satisfy $\lambda \geq 1$ and the invariant subspace associated with $\lambda = 1$ is important.

Suppose $Sv_1 = v_1$ and $Sv_2 = v_2$. In the HO-GSVD expansion

$$A_k = \sigma_1 u_1^{(k)} v_1^T + \sigma_2 u_2^{(k)} v_2^T + \sum_{j=3}^{n} \sigma_j^{(k)} u_j^{(k)} v_j^T$$

it can be shown that

(1) the red vectors are orthogonal to the blue vectors.

(2) the red vectors are left singular vectors for A_k.

The subspace span$\{v_1, v_2\}$ is the **the common HO-GSVD subspace**.

HOGSVD

We were able to discover biological similarity among three organisms in how they regulate their cell-cycle programs via

$$A_k = \underbrace{\sigma_1 u_1^{(k)} v_1^T + \sigma_2 u_2^{(k)} v_2^T}_{\text{The critical part}} + \sum_{j=3}^{n} \sigma_j^{(k)} u_j^{(k)} v_j^T \qquad k = 1{:}3$$

Lecture Summary

Key Words

- The CP representation (like the Tucker representation) is a very basic way of representing a tensor.
- Both Tucker and CP approximation problems are approached via alternating least Squares
- Both representations have variants.
- They motivate various “simultaneous SVD” frameworks.

References

S.P. Ponnapalli, M.A. Saunders, C. Van Loan,and Orly Alter (2011). "A Higher Order Generalized Singular Value Decompoistion for Comparison of Global mRNA Expression from Multiple Organisms", *PLoS One 6,* Issue 12.

G.H. Golub and C.F. Van Loan, *Matrix Computations, Fourth Edition*, Johns Hopkins University Press, Baltimore, MD.

H.A.L. Kiers, J.M.F. Ten Berg, and R. Bro, *PARAFAC2–Part 1. A direct fitting algorithm for the PARAFAC2 model*, J. Chemometrics, 13 (1999), pp. 275–294.

L. de Lathauwer (1997). *Signal Processing Based on Multilinear Algebra*, PhD Thesis, K.U. Leuven.

A. Smilde, R. Bro, and P. Geladi (2004). *Multiway Analysis*, John Wiley, Chichester, England.

T.G. Kolda and B.W. Bader (2009). "Tensor Decompositions and Applications," *SIAM Review* 51, 455–500.

C. A. Andersson and R. Bro (2000). "The N-way Toolbox for MATLAB," *Chemometrics and Intelligent Laboratory Systems 52*, 1–4.