Structured Matrix Computations from Structured Tensors

Lecture 4. CP and KSVD Representations

Charles F. Van Loan

Cornell University

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

What is this Lecture About?

Two More "Tensor SVDs"

The CP Representation has "diagonal" aspect like the SVD but there is no orthogonality.

The Kronecker Product SVD can be used to write a given matrix as an "optimal" sum of Kronecker products. If the matrix is obtained via a tensor unfolding, then we obtain yet another SVD-like representation.

The CP Representation

The CP Representation

Definition

The CP representation for an $n_1 \times n_2 \times n_3$ tensor $\mathcal{A}$ has the form

$$\mathcal{A} \; = \; \sum_{k=1}^{r} \lambda_k F(:,k) \circ G(:,k) \circ H(:,k)$$

where λ's are real scalars and $F \in \mathbb{R}^{n_1 \times r}$, $G \in \mathbb{R}^{n_2 \times r}$, and $H \in \mathbb{R}^{n_3 \times r}$

Equivalent

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

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

Tucker Vs. CP

The Tucker Representation

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

The CP Representation

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

In Tucker the U's have orthonormal columns. In CP, the matrices F, G, and H do not have orthonormal columns.

In CP the core tensor is diagonal while in Tucker it is not.

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

A Little More About Tensor Rank

The CP Representation and Rank

Definition

If

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

is the shortset possible CP representation of $\mathcal{A}$, then

$$\text{rank}(\mathcal{A}) \;=\; r$$

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.

The Nearest CP Problem

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

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

then

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

is minimized.

A multilinear optimization problem.

The CP Approximation Problem

Equivalent Formulations

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

$$=$$

$$\left\| \mathcal{A}_{(1)} \; - \; \sum_{j=1}^{r} \lambda_j \cdot F(:,j) \; \otimes \; (\; H(:,j) \otimes G(:,j) \;)^T \right\|_F$$

$$=$$

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

$$=$$

$$\left\| \mathcal{A}_{(3)} \; - \; \sum_{j=1}^{r} \lambda_j \cdot H(:,j) \; \otimes \; (\; G(:,j) \otimes F(:,j) \;)^T \right\|_F$$

Introducing the Khatri-Rao Product

Definition

If

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

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

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

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

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

The CP Approximation Problem

Equivalent Formulations

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

$$=$$

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

$$=$$

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

$$=$$

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

The CP Approximation Problem

The Alternating LS Solution Framework...

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

$$=$$

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

1. Fix G and H and improve λ and F.

$$=$$

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

2. Fix F and H and improve λ and G.

$$=$$

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

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.

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

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

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

$z \in \mathbb{R}^{n_1 n_2}$, 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 $Mx = y$. $O(r^2)$.

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

The Kronecker Product SVD

The Nearest Kronecker Product Problem

Find B and C so that $\| A - B \otimes C \|_F = \min$

$$\left\| \left[\begin{array}{cc|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} \\ a_{61} & a_{62} & a_{63} & a_{64} \end{array} \right] - \left[\begin{array}{cc} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{array} \right] \otimes \left[\begin{array}{cc} c_{11} & c_{12} \\ c_{21} & c_{22} \end{array} \right] \right\|_F$$

$$=$$

$$\left\| \left[\begin{array}{cccc} a_{11} & a_{21} & a_{12} & a_{22} \\ \hline a_{31} & a_{41} & a_{32} & a_{42} \\ \hline a_{51} & a_{61} & a_{52} & a_{62} \\ \hline a_{13} & a_{23} & a_{14} & a_{24} \\ \hline a_{33} & a_{43} & a_{34} & a_{44} \\ \hline a_{53} & a_{63} & a_{54} & a_{64} \end{array} \right] - \left[\begin{array}{c} b_{11} \\ b_{21} \\ b_{31} \\ b_{12} \\ b_{22} \\ b_{32} \end{array} \right] \left[\begin{array}{cccc} c_{11} & c_{21} & c_{12} & c_{22} \end{array} \right] \right\|_F$$

The Nearest Kronecker Product Problem

Find B and C so that $\| A - B \otimes C \|_F = \min$

It is a nearest rank-1 problem,

$$\phi_A(B,C) = \left\| \begin{bmatrix} a_{11} & a_{21} & a_{12} & a_{22} \\ \hline a_{31} & a_{41} & a_{32} & a_{42} \\ \hline a_{51} & a_{61} & a_{52} & a_{62} \\ \hline a_{13} & a_{23} & a_{14} & a_{24} \\ \hline a_{33} & a_{43} & a_{34} & a_{44} \\ \hline a_{53} & a_{63} & a_{54} & a_{64} \end{bmatrix} - \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \\ b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} \begin{bmatrix} c_{11} & c_{21} & c_{12} & c_{22} \end{bmatrix} \right\|_F$$

$$= \| \tilde{A} - \text{vec}(B)\text{vec}(C)^T \|_F$$

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"

$$A = \left[\begin{array}{cc|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} \\ a_{61} & a_{62} & a_{63} & a_{64} \end{array}\right] = \left[\begin{array}{cc} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{array}\right]$$

implies

$$\tilde{A} = \left[\begin{array}{cccc} a_{11} & a_{21} & a_{12} & a_{22} \\ \hline a_{31} & a_{41} & a_{32} & a_{42} \\ \hline a_{51} & a_{61} & a_{52} & a_{62} \\ \hline a_{13} & a_{23} & a_{14} & a_{24} \\ \hline a_{33} & a_{43} & a_{34} & a_{44} \\ \hline a_{53} & a_{63} & a_{54} & a_{64} \end{array}\right] = \left[\begin{array}{c} \text{vec}(A_{11})^T \\ \text{vec}(A_{21})^T \\ \text{vec}(A_{31})^T \\ \text{vec}(A_{12})^T \\ \text{vec}(A_{22})^T \\ \text{vec}(A_{32})^T \end{array}\right].$$

The Kronecker Product SVD (KPSVD)

Theorem

If

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

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 $\{\mathrm{vec}(U_k)\}$ and $\{\mathrm{vec}(V_k)\}$ are orthonormal and r_{KP} is the Kronecker rank of A with respect to the chosen blocking.

The Kronecker Product SVD (KPSVD)

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

$$\begin{aligned} \text{vec}(U_k) &= u_k \\ \text{vec}(V_k) &= v_k \end{aligned}$$

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

$$U_k = \texttt{reshape}(u_k, r_2, c_2),\; V_k = \texttt{reshape}(v_k, r_1, c_1)$$

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.

Structured Kronecker Product Approximation

$\min_{B,C} \| A - B \otimes C \|_F$ Problems

If A is symmetric and positive definite, then so are B and C.

If A is a block Toeplitz with Toeplitz blocks, then B and C are Toeplitz.

If A is a block band matrix with banded blocks, the B and C are banded.

Can use Lanczos SVD if A is large and sparse.

A Tensor Approximation Idea

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:

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

Back to tensor:

$$\mathcal{A} = \sum_{k=1}^{r} \sigma_k \mathcal{C}_k \circ \mathcal{B}_k$$

i.e.,

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

Sums of tensor products of matrices instead of vectors.

The Nearest Kronecker Product Problem

Harder

$$\phi_A(B, C, D)$$

$$=$$

$$\| A - B \otimes C \otimes D \|_F$$

$$=$$

$$\sqrt{\sum_{i_1=1}^{r_1} \sum_{j_1=1}^{c_1} \sum_{i_2=1}^{r_2} \sum_{j_2=1}^{c_2} \sum_{i_3=1}^{r_3} \sum_{j_3=1}^{c_2} \mathcal{A}(i_1, j_1, i_2, j_2, i_3, j_3) - \mathcal{B}(i_3, j_3)\mathcal{C}(i_2, j_2)D(i_1, j_1)}$$

Trying to approximate an order-6 tensor with a triplet of order-2 tensors. Would have to apply componentwise optimization.

Concluding Remarks

Optional "Fun" Problems

Problem E4. Suppose

$$A = \begin{bmatrix} B_{11} \otimes C_{11} & B_{12} \otimes C_{12} \\ B_{21} \otimes C_{21} & B_{22} \otimes C_{22} \end{bmatrix}$$

and that the B_{ij} and C_{ij} are each m-by-m. (a) Assuming that structure is fully exploited, how many flops are required to compute $y = Ax$ where $x \in \mathbb{R}^{2m^2}$? (b) How many flops are required to explicitly form A? (c) How many flops are required to compute $y = Ax$ assuming that A has been explicitly formed?

Problem A4. Suppose A is n^2-by-n^2. How would you compute $X \in \mathbb{R}^{n \times n}$ so that $\| A - X \otimes X \|_F$ is minimized?