Structured Matrix Computations from Structured Tensors

Lecture 6. The Higher-Order Generalized Singular Value Decomposition

Charles F. Van Loan

Cornell University

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

A Proof that $3 \gg 2$

It is possible to reduce a pair of matrices to canonical form.

Generalized Schur Decomposition

Simultaneous upper triangularization:

$$Q^T A_1 Z = T_1 \qquad Q^T A_2 Z = T_2$$

The Generalized Singular Value Decomposition

Simultaneous diagonalization:

$$U_1^T A_1 V = \Sigma_1 \qquad U_2^T A_2 V = \Sigma_2$$

A Proof that $3 \gg 2$

It is possible to reduce a pair of matrices to canonical form.

Generalized Schur Decomposition

Simultaneous upper triangularization:

$$Q^T A_1 Z \;=\; T_1 \qquad Q^T A_2 Z \;=\; T_2$$

The Generalized Singular Value Decomposition

Simultaneous diagonalization:

$$U_1^T A_1 V \;=\; \Sigma_1 \qquad U_2^T A_2 V \;=\; \Sigma_2$$

But you can forget about this kind of simultaneous reduction when there are more than two matrices. Q.E.D.

Three is a Crowd

For example, there are **no** methods for the quadratic eigenvalue problem

$$(A_1 + \lambda A_2 + \lambda^2 A_3)x \;=\; 0$$

that work by simultaneously reducing all three matrices to a canonical form

$$Q^T A_1 Z = \tilde{A}_1 \qquad Q^T A_2 Z = \tilde{A}_2 \qquad Q^T A_3 Z = \tilde{A}_3$$

that "reveals" the solution

$$(\tilde{A}_1 + \lambda \tilde{A}_2 + \lambda^2 \tilde{A}_3)\tilde{x} \;=\; 0$$

Challenge

Given a collection of data matrices

$$\{A_1, \ldots, A_N\}$$

that each have the same number of columns, how can you discover features that they share in common?

Idea 1: Use a Tensor Decomposition

If each matrix in the collection $\{A_1, \ldots, A_N\}$ has the same number of rows, then "stack them up" into a tensor

$$\mathcal{A}(:,:,k) = A_k \qquad k = 1{:}N$$

and compute (say) a CP decomposition

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

Since

$$A(i,j,k) \;=\; \sum_{p=1}^{r} \lambda_p F(i,p) G(j,p) H(k,p)$$

this says

$$\mathcal{A}(:,:,k) \;=\; A_k \;=\; \sum_{p=1}^{r} (\lambda_p H(k,p)) F(:,p) G(:,p)^T \qquad k = 1{:}N$$

Idea 2. Approximate SVDs

Given $A_k \in \mathbb{R}^{m_k \times n}$ for $k = 1{:}N$ and an integer $r \leq n$, determine

$$
\begin{aligned}
&U_k \in \mathbb{R}^{m_k \times r} && k = 1{:}N, \text{Each with orthonormal columns} \\
&S_k \in \mathbb{R}^{r \times r} && k = 1{:}N, \text{Each diagonal} \\
&V \in \mathbb{R}^{n \times r} &&
\end{aligned}
$$

so that

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

is minimized. (We do not force V to have orthonormal columns.)

Idea 2. Approximate SVDs Using Alternating Least Squares

Improving the U_k (Orthonormal)

Fix the S_k and V and determine $U_1, \ldots, U_N$ so that

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

is minimized.

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

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

Idea 2. Approximate SVDs Using Alternating Least Squares

Improving the S_k (Diagonal)

Fix the U_k and V and determine the $S_1, \ldots, S_N$ so that

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

is minimized.

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

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

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

Idea 2. Approximate SVDs Using Alternating Least Squares

Improving V

Fix the U_k and the S_k and determine V so that

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

is minimized.

Hint: This is a least squares problem since

$$\sum_{k=1}^{N} \| A_k - U_k S_k V^T \|_F^2 = \left\| \begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix} - \begin{bmatrix} U_1 S_1 \\ \vdots \\ U_N S_N \end{bmatrix} V^T \right\|_F^2$$

Idea 2. Approximate SVDs Using Alternating Least Squares

This is the PARAFAC2 Framework

Repeat Until Happy

Improve $U_1, \ldots, U_N$

Improve $S_1, \ldots, S_N$

Improve V

But we are going to do something different...

Idea 3. Use the Higher-Order GSVD Framework

Assume that $A_1, \ldots, A_N$ each have full column rank.

1. Compute $V^{-1} S_N V = \mathrm{diag}(\lambda_i)$ where

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

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.

Upon completion we have $A_k \;=\; U_k \Sigma_k V^T, \quad k = 1{:}N$
The U-matrices in these expansions turns out to be connected in a very special way if S_N has an eigenvalue equal to one.

Idea 3. Use the Higher-Order GSVD Framework

The Common HO-GSVD Subspace: Definition

The eigenvectors associated with the unit eigenvalues of S_N define the common HO-GSVD subspace:

$$\text{HO-GSVD}(A_1, \ldots, A_N) \;=\; \{\, v : S_N v = v \,\}$$

Idea 3. Use the Higher-Order GSVD Framework

The Common HO-GSVD Subspace: Importance

In general, we have these rank-1 expansions

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

where $V = [v_1, \ldots, v_n]$.

But if (say) the HO-GSVD$(A_1, \ldots, A_N) = \text{span}\{v_1, v_2\}$, then

$$A_k \;=\; \sigma_1 u_1^{(k)} v_1^T \;+\; \sigma_2 u_2^{(k)} v_2^T \;+\; \sum_{i=3}^{n} \sigma_i^{(k)} u_i^{(k)} v_i^T \qquad k = 1{:}N$$

and $\{u_1^{(k)}, u_2^{(k)}\}$ is an orthonormal basis for $\text{span}\{u_3^{(k)}, \ldots, u_n^{(k)}\}^{\perp}$. Moreover, $u_1^{(k)}$ and $u_2^{(k)}$ are left singular vectors for A_k.

This expansion identifies features that are common across the datasets $A_1, \ldots, A_N$.

Idea 3. Use the Higher-Order GSVD Framework

The Common HO-GSVD Subspace: Importance

In general, we have these rank-1 expansions

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

where $V = [v_1, \ldots, v_n]$.

But if (say) the HO-GSVD$(A_1, \ldots, A_N) = \text{span}\{v_1, v_2\}$, then

$$A_k \;=\; \sigma_1 u_1^{(k)} v_1^T \;+\; \sigma_2 u_2^{(k)} v_2^T \;+\; \sum_{i=3}^{n} \sigma_i^{(k)} u_i^{(k)} v_i^T \qquad k = 1{:}N$$

and $\{u_1^{(k)}, u_2^{(k)}\}$ is an orthonormal basis for span$\{u_3^{(k)}, \ldots, u_n^{(k)}\}^\perp$. Moreover, $u_1^{(k)}$ and $u_2^{(k)}$ are left singular vectors for A_k.

Much to Explain!

The CS Decomposition

(The Two-Matrix Case)

The CS Decomposition

Definition

If

$$Q = \left[\begin{array}{ccc|cc} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \hline \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{array}\right]$$

has orthonormal columns, then there exist orthogonal U_1, U_2, Z_1 and Z_2 so that

$$\left[\begin{array}{c|c} U_1 & 0 \\ \hline 0 & U_2 \end{array}\right]^T Q \left[\begin{array}{c|c} Z_1 & 0 \\ \hline 0 & Z_2 \end{array}\right] = \left[\begin{array}{ccc|cc} c_1 & 0 & 0 & -s_1 & 0 \\ 0 & c_2 & 0 & 0 & -s_2 \\ 0 & 0 & c_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline s_1 & 0 & 0 & c_1 & 0 \\ 0 & s_2 & 0 & 0 & c_2 \\ 0 & 0 & s_3 & 0 & 0 \end{array}\right]$$

The SVDs of the blocks are related.

The CS Decomposition

Definition (Structured Special Case: Q)

If $Q \in \mathbb{R}^{2n \times 2n}$ is orthogonal and

$$J_{2n}^T Q J_{2n} = Q^{-1} \qquad J_{2n} = \left[\begin{array}{c|c} 0 & I_n \\ \hline -I_n & 0 \end{array}\right]$$

then

$$Q = \left[\begin{array}{c|c} Q_1 & -Q_2 \\ \hline Q_2 & Q_1 \end{array}\right]$$

and there exist orthogonal U and Z so that

$$\left[\begin{array}{c|c} U & 0 \\ \hline 0 & U \end{array}\right]^T Q \left[\begin{array}{c|c} Z & 0 \\ \hline 0 & Z \end{array}\right] = \left[\begin{array}{ccc|ccc} c_1 & 0 & 0 & -s_1 & 0 & 0 \\ 0 & c_2 & 0 & 0 & -s_2 & 0 \\ 0 & 0 & c_3 & 0 & 0 & -s_3 \\ \hline s_1 & 0 & 0 & c_1 & 0 & 0 \\ 0 & s_2 & 0 & 0 & c_2 & 0 \\ 0 & 0 & s_3 & 0 & 0 & c_3 \end{array}\right] = \left[\begin{array}{c|c} C & -S \\ \hline S & C \end{array}\right]$$

Q_2 nonsingular $\Rightarrow Q_1 Q_2^{-1} = U \cdot \mathrm{diag}(c_i/s_i) \cdot U^T$, a symmetric Schur Decomp.

The CS Decomposition

Definition (Thin Version)

If

$$\left[\begin{array}{c} Q_1 \\ \hline Q_2 \end{array}\right] = \left[\begin{array}{ccc} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \hline \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{array}\right]$$

has orthonormal columns, then there exist orthogonal U_1, U_2, and Z so that

$$\left[\begin{array}{c|c} U_1 & 0 \\ \hline 0 & U_2 \end{array}\right]^T \left[\begin{array}{c} Q_1 \\ \hline Q_2 \end{array}\right] Z = \left[\begin{array}{ccc} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \\ 0 & 0 & 0 \\ \hline s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{array}\right] = \left[\begin{array}{c} C \\ \hline S \end{array}\right]$$

Computation

Stable efficient methods exist.

Not straight forward.

You can't just compute the SVDs

$$U_1 Q_{11} V_1 = \Sigma_1 \qquad U_2 Q_{22} V_2 = \Sigma_2$$

and expect $U_1 Q_{12} V_2$ and $U_2 Q_2 V_1$ to be diagonal to within machine precision.

Rethinking the 2-Matrix Generalized Singular Value Decomposition

The 2-Matrix GSVD

Definition

If

$$A_1 = \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \qquad A_2 = \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix}$$

then there exist orthogonal U_1, orthogonal U_2 and **nonsingular** X so that

$$U_1^T A_1 X \;=\; \Sigma_1 \;=\; \begin{bmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad U_2^T A_2 X \;=\; \Sigma_2 \;=\; \begin{bmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \\ 0 & 0 & 0 \end{bmatrix}$$

The 2-Matrix GSVD

The Rank-1 Expansion Version

The GSVD basically says that there exist orthogonal U_1, orthogonal U_2 and **nonsingular** X so that

$$U_1^T A_1 X \;=\; \Sigma_1 \;=\; \text{diag}(c_k) \qquad U_2^T A_2 X \;=\; \Sigma_2 \;=\; \text{diag}(s_k)$$

are diagonal. Thus, if $U_1 = [u_1^{(1)}, \ldots, u_n^{(1)}]$, $U_2 = [u_1^{(2)}, \ldots, u_n^{(2)}]$, and

$$X^{-T} = V = [v_1, \ldots, v_n]$$

are column partitionings, then

$$A_1 \;=\; U_1 \Sigma V^T \;=\; \sum_{k=1}^{n} c_k u_k^{(1)} v_k^T \qquad A_2 \;=\; U_2 \Sigma V^T \;=\; \sum_{k=1}^{n} s_k u_k^{(2)} v_k^T$$

Moving X to the other side would be simpler if it was orthogonal for then $V = X^{-T} = X$.

The 2-Matrix GSVD

Applications

Many 2-matrix problems can be diagonalized via the GSVD. For example, in quadratically Constrained Least Squares we solve

$$\min \| A_1 x - b \|_2 \qquad \text{such that} \qquad \| A_2 x - d \|_2 \leq \alpha$$

By substituting the GSVD of A_1 and A_2 into this we get an easily solved equivalent problem with diagonal matrices:

$$\min \| \Sigma_1 \tilde{x} - \tilde{b} \|_2 \qquad \text{such that} \qquad \| \Sigma_2 \tilde{x} - \tilde{d} \|_2 \leq \alpha$$

The 2-Matrix GSVD

Computation

1. Compute the QR factorization:

$$\left[\begin{array}{c} A_1 \\ A_2 \end{array}\right] = \left[\begin{array}{c} Q_1 \\ Q_2 \end{array}\right] R$$

2. Compute the CS decomposition:

$$Q_1 \;=\; U_1 \cdot \text{diag}(c_i) \cdot Z^T \qquad Q_2 \;=\; U_2 \cdot \text{diag}(s_i) \cdot Z^T \qquad \text{(SVD's)}$$

3. Set $V^T = Z^T R$. Note: $X = V^{-T} = R^{-1} Z$

$$A_1 \;=\; Q_1 R \;=\; U_1 \cdot \text{diag}(c_i) \cdot (Z^T R) \;=\; U_1 \Sigma_1 V^T$$
$$A_2 \;=\; Q_2 R \;=\; U_2 \cdot \text{diag}(s_i) \cdot (Z^T R) \;=\; U_2 \Sigma_2 V^T$$

The 2-Matrix GSVD

Relevance to the Problem $A_1^T A_1 x = \tau^2 A_2^T A_2 x$

Since $U_1^T A_1 X = \Sigma_1$ and $U_2^T A X = \Sigma_2$, it follows that

$$X^T(A_1^T A_1 \; - \; \tau^2 A_2^T A_2)X \; = \; \Sigma_1^T \Sigma_1 - \tau^2 \Sigma_2^T \Sigma_2 \; = \; \mathrm{diag}(c_i^2 - \tau^2 s_i^2)$$

and so

$$A_1^T A_1 x_i \; = \; \left(\frac{c_i^2}{s_i^2}\right) A_2^T A_2 x_i$$

where $X = [\, x_1 \mid \cdots \mid x_n \,]$.

The c_i/s_i and x_i are the generalized singular values and vectors of $\{A_1, A_2\}$.

The 2-Matrix GSVD

Characterizing the V-Matrix

Since

$$A_1 = U_1 \Sigma_1 V^T \qquad A_2 = U_2 \Sigma_2 V^T$$

implies

$$A_1^T A_1 = V(\Sigma_1^T \Sigma_1) V^T \qquad A_2^T A_2 = V(\Sigma_2^T \Sigma_2) V^T$$

we see that

$$\begin{aligned}(A_2^T A_2)(A_1^T A_1)^{-1} &= V(\Sigma_2^T \Sigma_2)(\Sigma_1^T \Sigma_1)^{-1} V^{-1} = V \text{diag}((s_i^2/c_i^2) V^{-1} \\ (A_1^T A_1)(A_2^T A_2)^{-1} &= V(\Sigma_1^T \Sigma_1)(\Sigma_2^T \Sigma_2)^{-1} V^{-1} = V \text{diag}((c_i^2/s_i^2) V^{-1}\end{aligned}$$

The columns of V are eigenvectors for both

$(A_2^T A_2)(A_1^T A_1)^{-1}$ and $(A_1^T A_1)(A_2^T A_2)^{-1}$.

The 2-Matrix GSVD

Characterizing the V-Matrix

If

$$S = \frac{1}{2}\left((A_2^T A_2)(A_1^T A_1)^{-1} + (A_1^T A_1)(A_2^T A_2)^{-1}\right)$$

then since

$$\begin{aligned}(A_2^T A_2)(A_1^T A_1)^{-1} &= V(\Sigma_2^T \Sigma_2)(\Sigma_1^T \Sigma_1)^{-1} V^{-1} = V\text{diag}((s_i^2/c_i^2)V^{-1} \\ (A_1^T A_1)(A_2^T A_2)^{-1} &= V(\Sigma_1^T \Sigma_1)(\Sigma_2^T \Sigma_2)^{-1} V^{-1} = V\text{diag}((c_i^2/s_i^2)V^{-1}\end{aligned}$$

we have

$$S = V \cdot \text{diag}\left(\frac{1}{2}\left(\frac{s_i^2}{c_i^2} + \frac{c_i^2}{s_i^2}\right)\right) V^{-1}$$

The columns of V are eigenvectors for S and the eigenvalues are never smaller than 1 because the function $f(x) = (x + 1/x)/2$ is never smaller than 1.

The 2-Matrix GSVD

The Common Invariant Subspace Problem

Compute a matrix whose columns are an orthonormal basis for

$$\mathbf{C}_{\text{HOGSVD}}\{A_1, A_2\} \;=\; \{\, v : Sv = v\}$$

where $S = \left((A_1^T A_1)(A_2^T A_2)^{-1} + (A_2^T A_2)(A_1^T A_1)^{-1}\right)/2$.

Algorithm $\tilde{Q} = \texttt{Common}(A_1, A_2)$

1. Compute the GSVD: $A_1 = U_1 \text{diag}(c_i) V^T$, $A_2 = U_2 \text{diag}(s_i) V^T$.

2. Let $\tilde{V}$ consist of those columns of V associated with generalized singular values that equal 1 to within some tolerance, i.e., include $V(:, i)$ if $|c_i - s_i| \leq tol$.

3. Orthonormalize: $\tilde{V} = \tilde{Q}\tilde{R}$.

The Higher Order CS Decomposition

Higher-Order CSD: Motivation

If

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

and

$$\begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix} = \begin{bmatrix} Q_1 \\ \vdots \\ Q_N \end{bmatrix} R$$

is a thin QR factorization, then since $A_k = Q_k R$ we have

$$R^{-T} S R = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left((Q_i^T Q_i)(Q_j^T Q_j)^{-1} + (Q_j^T Q_j)(Q_i^T Q_i)^{-1} \right).$$

It follows that

$$R^{-T}SR^{T} \;=\; \frac{1}{N-1}(T-I)$$

where T is the symmetric matrix

$$T \;=\; \frac{1}{N}\left((Q_1^TQ_1)^{-1}+\cdots+(Q_N^TQ_N)^{-1}\right)$$

$$\begin{aligned}
R^{-T}SR^{T} \;&=\; \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left((Q_i^TQ_i)(Q_j^TQ_j)^{-1}+(Q_j^TQ_j)(Q_i^TQ_i)^{-1}\right)\\
&=\; \frac{1}{N(N-1)}\left(\left(Q_1^TQ_1+\cdots+Q_N^TQ_N\right)\left((Q_1^TQ_1)^{-1}+\cdots+(Q_N^TQ_N)^{-1}\right)-NI\right)\\
&=\; \frac{1}{N(N-1)}\left((Q_1^TQ_1)^{-1}+\cdots+(Q_N^TQ_N)^{-1}-NI\right)
\end{aligned}$$

The Higher-Order CS Decomposition (HO-CSD)

Definition

If

$$Q = \begin{bmatrix} Q_1 \\ \vdots \\ Q_N \end{bmatrix}$$

has orthonormal columns and each Q_k has full column rank, then its HO-CSD is given by

$$Q_k = U_k \Sigma_k Z^T \qquad k = 1{:}N$$

where Z is orthogonal such that

$$Z^T T Z = \text{diag}(\mu_k) \qquad T = \frac{1}{N}\left((Q_1^T Q_1)^{-1} + \cdots + (Q_N^T Q_N)^{-1}\right)$$

and for $k = 1{:}N$ we have

$$Q_k Z = U_k \Sigma_k = (\text{Matrix with unit 2-norm columns})\cdot(\text{Diagonal Matrix})$$

The Higher-Order CS Decomposition (HO-CSD)

Properties of T

The Cauchy inequality tells us that

$$y^T(Q_k^TQ_k)^{-1}y \;\geq\; \frac{1}{y^T(Q_k^TQ_k)y_k} \qquad k = 1{:}N$$

with equality iff y is an eigenvector for $Q_k^TQ_k$. Using this fact, it can be shown that if $\| \, y \, \|_2 = 1$, then

$$y^TTy \;\;=\;\; y^T\left(\frac{1}{N}\left((Q_1^TQ_1)^{-1} + \cdots + (Q_N^TQ_N)^{-1}\right)\right)y \;\geq N$$

with equality iff

$$Q_k^TQ_ky \;=\; \frac{1}{N}y \qquad k = 1{:}N$$

VERY BIG FACT: $Ty = N{\cdot}y \;\Leftrightarrow\;$ y is a right singular vector for each Q_k

The Higher-Order CS Decomposition (HO-CSD)

The Common HO-CSD Subspace

If the columns of

$$Q = \begin{bmatrix} Q_1 \\ \vdots \\ Q_N \end{bmatrix}$$

are orthonormal and if each block has full column rank, then the Common HO-CSD Subspace is defined by

$$\mathbf{C}_{\text{HOCSD}}\{Q_1, \ldots, Q_N\} = \{\, x \mid T_N x = N x \,\}.$$

The Higher-Order CS Decomposition (HO-CSD)

Canonical Form

Suppose the columns of

$$Q = \begin{bmatrix} Q_1 \\ \vdots \\ Q_N \end{bmatrix}$$

are orthonormal and each block has full column rank. Assume that

$$Z^T T_N Z = \text{diag}(\mu_i) \qquad Z = [z_1, \ldots, z_n]$$

is the Schur decomposition with

$$\text{span}\{z_1, \ldots, z_p\} = \mathbf{C}_{\text{HOCSD}}\{Q_1, \ldots, Q_N\}$$

Then...

The Higher-Order CS Decomposition (HO-CSD)

Canonical Form

$$Q_k \;=\; U_k \Sigma_k Z^T \qquad k = 1{:}N$$

where

$$U_k \;=\; \Big[\; \underbrace{U_k^{(c)}}_{p} \mid \underbrace{U_k^{(u)}}_{n-p} \;\Big] \qquad\qquad Z \;=\; \Big[\; \underbrace{Z^{(c)}}_{p} \mid \underbrace{Z^{(u)}}_{n-p} \;\Big]$$

and

$$\Sigma_k \;=\; \begin{bmatrix} I_p/\sqrt{N} & 0 \\ 0 & \Sigma_k^{(u)} \end{bmatrix}$$

is diagonal. Moreover, the columns of each $U_k^{(c)}$ are orthonormal and

$$[U_k^{(c)}]^T U_k^{(u)} = 0.$$

The Higher-Order CS Decomposition (HO-CSD)

Want to compute an Orthonormal Basis for $\mathbf{C}_{\text{HOCSD}}\{Q_1, \ldots, Q_N\}$

A Useful Characterization:

$$\begin{aligned} \mathbf{C}_{\text{HOCSD}}\{Q_1, \ldots, Q_N\} &= \bigcap_{1 \le i < j \le N} \mathbf{C}_{\text{HOGSVD}}\{Q_i, Q_j\} \\ &= \bigcap_{k=2}^{N} \mathbf{C}_{\text{HOGSVD}}\{Q_{k-1}, Q_k\} \end{aligned}$$

Algorithm (A Sequence of Ever-Thinner GSVD Problems)

$Z_c = \texttt{Common}(Q_1, Q_2)$

for $k = 3{:}N$

$\quad Z_k = \texttt{Common}(Q_{k-1} Z_c, Q_k Z_c)$

$\quad Z_c = Z_c Z_k$

The columns of Z_c span $\mathbf{C}_{\text{HOCSD}}\{Q_1, \ldots, Q_N\}$.

The Higher-Order GSVD

The Higher-Order GSVD Framework

Given: $A_i \in \mathbb{R}^{m_i \times n}$, $i = 1{:}N$ each with full column rank.

1. Assume $V^{-1} S_N V = \text{diag}(\lambda_i)$ where

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

2. For $k = 1{:}N$ set

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

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

What we have: $\quad A_k \;=\; U_k \Sigma_k V^T, \quad k = 1{:}N$

Properties of S_N

Use the Connection to T_N

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

$$T_N = \frac{1}{N} \left((Q_1^T Q_1)^{-1} + \cdots + (Q_N^T Q_N)^{-1} \right)$$

$$R^{-T} S_N R^T = \frac{1}{N-1} (T_N - I)$$

Here, $\begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix} = \begin{bmatrix} Q_1 \\ \vdots \\ Q_N \end{bmatrix} R$ is the thin QR factorization

Properties of S_N

Use the Connection to T_N

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

$$T_N = \frac{1}{N} \left((Q_1^T Q_1)^{-1} + \cdots + (Q_N^T Q_N)^{-1} \right)$$

$$R^{-T} S_N R^T = \frac{1}{N-1} (T_N - I)$$

1. S_N is similar to T_N, a symmetric matrix.
2. S_N where is diagonalizable with real eigenvalues.
3. If $Z^T T_N Z = \text{diag}(\mu_i)$, then $V^{-1} S_N V = \text{diag}(\lambda_i)$ where $V = R^T Z$ and $\lambda_i = (\mu_i - 1)/(N-1)$.

Properties of S_N

Use the Connection to T_N

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

$$T_N = \frac{1}{N} \left((Q_1^T Q_1)^{-1} + \cdots + (Q_N^T Q_N)^{-1} \right)$$

$$R^{-T} S_N R^T = \frac{1}{N-1} (T_N - I)$$

3. If $Z^T T_N Z = \text{diag}(\mu_i)$, then $V^{-1} S_N V = \text{diag}(\lambda_i)$ where $V = R^T Z$ and $\lambda_i = (\mu_i - 1)/(N-1)$.
4. Since the eigenvalues $\{\mu_i\}$ of T_N satisfy $\mu_i \geq N$, the eigenvalues $\{\lambda_i\}$ of S_N satisfy $\lambda_i \geq 1$.
5. $S_N x = x$ if and only if $y = R^{-1} x$ is a right singular vector for each Q_k.

Properties of S_N

The Common HO-GSVD Subspace: Definition

The eigenvectors associated with the unit eigenvalues of S_N define the common HO-GSVD subspace:

$$\mathbf{C}_{\text{HO-GSVD}}\{A_1, \ldots, A_N\} \;=\; \{\, v : S_N v = v \,\}$$

An Important Connection

Since

$$R^{-T} S_N R^T \;=\; \frac{1}{N-1}(T_N - I)$$

it follows that

$$\mathbf{C}_{\text{HO-GSVD}}\{A_1, \ldots, A_N\} \;=\; \{R^T z : z \in \mathbf{C}_{\text{HO-CSD}}\{Q_1, \ldots, Q_N\})$$

Properties of S_N

To Compute an Orthonormal Basis for $\mathbf{C}_{\text{HO-GSVD}}\{A_1, \ldots, A_N\}$

1. Compute the Thin QR factorization:

$$\begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix} = \begin{bmatrix} Q_1 \\ \vdots \\ Q_N \end{bmatrix} R$$

2. Compute a matrix Z_c with orthonormal columns that span $\mathbf{C}_{\text{HO-CSD}}\{Q_1, \ldots, Q_N\}$.
3. Compute the thin QR factorization $V_c R_c = (R^T Z_c)$.

The columns of V_c span $\mathbf{C}_{\text{HO-GSVD}}\{A_1, \ldots, A_N\}$

Properties of S_N

The Common HO-GSVD Subspace: Importance

In general, we have these rank-1 expansions

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

where $V = [v_1, \ldots, v_n]$.

But if (say) the HO-GSVD$(A_1, \ldots, A_N) = \text{span}\{v_1, v_2\}$, then

$$A_k \;=\; \sigma_1 u_1^{(k)} v_1^T \;+\; \sigma_2 u_2^{(k)} v_2^T \;+\; \sum_{i=3}^{n} \sigma_i^{(k)} u_i^{(k)} v_i^T \qquad k = 1{:}N$$

and $\{u_1^{(k)}, u_2^{(k)}\}$ is an orthonormal basis for span$\{u_3^{(k)}, \ldots, u_n^{(k)}\}^\perp$. Moreover, $u_1^{(k)}$ and $u_2^{(k)}$ are left singular vectors for A_k.

Identifies features that are common across the datasets $A_1, \ldots, A_N$.

A Partial GSVD

$N = 2$

$$A_k \;=\; \sigma_1^{(k)} u_1^{(k)} v_1^T \;+\; \sigma_2^{(k)} u_2^{(k)} v_2^T \;+ \cdots +\; \sigma_n^{(k)} u_n^{(k)} v_n^T \qquad k = 1, 2$$

and $\{u_1^{(k)}, \ldots, u_n^{(k)}\}$ an orthonormal basis for $\mathbb{R}^n$.

General N

$$A_k \;=\; \sigma_1^{(k)} u_1^{(k)} v_1^T \;+\; \sigma_2^{(k)} u_2^{(k)} v_2^T \;+\; \sum_{i=3}^{n} \sigma_i^{(k)} u_i^{(k)} v_i^T \qquad k = 1{:}N$$

and $\{u_1^{(k)}, u_2^{(k)}\}$ is an orthonormal basis for $\mathrm{span}\{u_3, \ldots, u_n^{(k)}\}^{\perp}$.

Not a simultaneous diagonalization, but good enough.

Open Problems

A Variational Approach

If $v \in \mathbf{C}_{\text{HO-GSVD}}\{A_1, \ldots, A_N\}$ then v is a stationary vector for

$$\phi(v) \;=\; \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{1}{2} \left(\frac{\| A_i v \|^2}{\| A_j v \|^2} + \frac{\| A_j v \|^2}{\| A_i v \|^2} \right) \;\geq\; 1$$

Does this open the door for sparse matrix friendly algorithm?

Rank-Deficient A_k

Everything revolves around

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

Is there a way to proceed in the event that one or more of the A_k is rank deficient? After all, the 2-matrix GSVD does not require the full rank assumption.

Summary

Tensor computations are prompting the development of new, structured matrix factorizations.

Tensor computations teach us to be relaxed about simultaneous diagonalization.