Properties of the Higher-Order GSVD

Charles Van Loan

Cornell University
Department of Computer Science

Workshop on Numerical Linear Algebra & Optimization

Vancouver

August 8–10, 2013

To: Previous Speakers

From: C. Van Loan

Date: August 10, 2013

Subject: Thanks for the tiny orthogonal complement.

How can I possibly add to the space?

$$\mathcal{S}_{\text{Overton}} \;=\; \text{span}\{v_1, \ldots, v_{26}\}$$

To: Previous Speakers

From: C. Van Loan

Date: August 10, 2013

Subject: Thanks for the tiny orthogonal complement.

How can I possibly add to the space?

$$\mathcal{S}_{\text{Overton}} = \text{span}\{v_1, \ldots, v_{26}\}$$

$$v_i = \text{Eigenvalues can be anti-social}$$

To: Previous Speakers

From: C. Van Loan

Date: August 10, 2013

Subject: Thanks for the tiny orthogonal complement.

How can I possibly add to the space?

$$\mathcal{S}_{\text{Overton}} = \text{span}\{v_1, \ldots, v_{26}\}$$

v_i = Eigenvalues frequently have other behavior problems

To: Previous Speakers

From: C. Van Loan

Date: August 10, 2013

Subject: Thanks for the tiny orthogonal complement.

How can I possibly add to the space?

$$\mathcal{S}_{\text{Overton}} = \text{span}\{v_1, \ldots, v_{26}\}$$

v_i = Eigenvalues like to shop at Whole Foods

To: Previous Speakers

From: C. Van Loan

Date: August 10, 2013

Subject: Thanks for the tiny orthogonal complement.

How can I possibly add to the space?

$$\mathcal{S}_{\text{Overton}} = \text{span}\{v_1, \ldots, v_{26}\}$$

v_i = Eigenvalues have been known to coalesce

To: Previous Speakers

From: C. Van Loan

Date: August 10, 2013

Subject: Thanks for the tiny orthogonal complement.

How can I possibly add to the space?

$$\mathcal{S}_{\text{Overton}} = \text{span}\{v_1, \ldots, v_{26}\}$$

$$v_i = \text{Eigenvalues sometimes travel in gangs}$$

To: Previous Speakers

From: C. Van Loan

Date: August 10, 2013

Subject: Thanks for the tiny orthogonal complement.

How can I possibly add to the space?

$$\mathcal{S}_{\text{Overton}} = \text{span}\{v_1, \ldots, v_{26}\}$$

$$v_i = \text{Eigenvalues require motivation to move}$$

Properties of the Higher-Order GSVD

Charles Van Loan

Cornell University
Department of Computer Science

Workshop on Numerical Linear Algebra & Optimization

Vancouver

August 8–10, 2013

The Setting

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.**

The Matrix V

It has something to do with this...

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

And it has something to do with this...

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

S_N is Diagonalizable

In General..

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

$$S_3 = \frac{\left(A_1^T A_1 + A_2^T A_2 + A_3^T A_3\right)\left((A_1^T A_1)^{-1} + (A_2^T A_2)^{-1} + (A_3^T A_3)^{-1}\right) - 3I}{6}$$

Product of two symmetric positive definite matrices

The HO-GSVD

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 Key Result

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

Common Features

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

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

See:

S. Priya Ponnapalli, Michael A. Saunders, Orly Alter, and CVL

A Higher Order Generalized Singular Value Decomposition for Comparison of Global mRNA Expression from Multiple Organisims, PLoS One, 6:12, 2011.

The 2-Matrix HO-GSVD

A.K.A. The Generalized Singular Value Decomposition

If $A \in \mathbb{R}^{m_1 \times n}$ and $A_2 \in \mathbb{R}^{m_2 \times n}$, there exist orthogonal U_1 and U_2 and nonsingular V so that

$$\begin{aligned} A_1 &= U_1 \Sigma_1 V^T \\ A_2 &= U_2 \Sigma_2 V^T \end{aligned}$$

where Σ_1 and Σ_2 are diagonal.

The Generalized Singular Value Problem

The Columns of $X = V^{-T}$ are the Generalized Singular Vectors

Since

$$A_1 = U_1 \Sigma_1 V^T = \mathrm{diag}(\sigma_k^{(1)}) \qquad A_2 = U_2 \Sigma_2 V^T = \mathrm{diag}(\sigma_k^{(2)})$$

it follows that

$$A_1^T A_1 \; - \; \mu^2 A_2^T A_2 \; = \; V \left(\Sigma_1^T \Sigma_1 - \mu^2 \Sigma_2^T \Sigma_2 \right) V^T.$$

Thus, if $V^{-T} \; = \; X \; = \left[\; x_1 \; \middle| \; \cdots \; \middle| \; x_n \; \right]$, then

$$A_1^T A_1 x_k \; = \; \mu_k^2 A_2^T A_2 x_k$$

where $\mu_k = \sigma_k^{(1)} / \sigma_k^{(2)}$ is a generalized singular value of $\{A_1, A_2\}$.

V the "Diagonalizer"

Look What V Does to S_2

Since

$$V^{-1}(A_1^T A_1)(A_2^T A_2)^{-1} V = (\Sigma_1^T \Sigma_1)(\Sigma_2^T \Sigma_2)^{-1}$$

$$V^{-1}(A_2^T A_2)(A_1^T A_1)^{-1} V = (\Sigma_2^T \Sigma_2)(\Sigma_1^T \Sigma_1)^{-1}$$

we have

$$V^{-1} S_2 V = \frac{1}{2} V^{-1}\left((A_1^T A_1)(A_2^T A_2)^{-1} + (A_2^T A_2)(A_1^T A_1)^{-1}\right) V$$

$$= \frac{1}{2}\left((\Sigma_1^T \Sigma_1)(\Sigma_2^T \Sigma_2)^{-1} + (\Sigma_2^T \Sigma_2)(\Sigma_1^T \Sigma_1)^{-1}\right)$$

The matrix S_2 is "symmetric" in A_1 and A_2.

$\lambda(S)$ and $\sigma(A_1, A_2)$

Here is the Connection

If $\mu_k = \sigma_k^{(1)}/\sigma_k^{(2)}$ is a generalized singular value of $\{A_1, A_2\}$, then

$$\lambda_k \;=\; \frac{1}{2}\left(\mu_k^2 \;+\; \frac{1}{\mu_k^2}\right)$$

is an eigenvalue of

$$V^{-1}S_2V \;=\; \frac{1}{2}\left((\Sigma_1^T\Sigma_1)(\Sigma_2^T\Sigma_2)^{-1} \;+\; (\Sigma_2^T\Sigma_2)(\Sigma_1^T\Sigma_1)^{-1}\right)$$

The function $f(z) = (z + 1/z)/2$ can never be smaller than one and that is why the eigenvalues of S_2 can never be smaller than one.

Computing the 2-Matrix GSVD

Three Simple Steps

1. Compute the QR factorization:

$$\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} R$$

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 $V^T = Z^T R$

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

Is there a higher-order CS decomposition?

Reminder About What We Need

We only Need Part of the "Complete" HO-GSVD

1. Diagonalize: $V^{-1}S_N V = \mathrm{diag}(\lambda_i)$ 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 $u_i^{(k)}$ and $\Sigma_k = \mathrm{diag}(\sigma_i^{(k)})$.

Just the v_i associated with the unit eigenvalues and the corresponding $u_i^{(k)}$ and $\sigma_i^{(k)}$. No inverses please!

Simplification of S_N via QR

A Thin QR Factorization...

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

Since $A_k = Q_k R$ and $Q_1^T Q_1 + \cdots + Q_N^T Q_N = I$ we can show...

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

where

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

Reminder:

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

Here is Why...

$$\sum_{i=1}^{2}\sum_{j=i+1}^{3}\left((Q_i^T Q_i)(Q_j^T Q_j)^{-1} + (Q_j^T Q_j)(Q_i^T Q_i)^{-1}\right)$$

$$=$$

$$\left(Q_1^T Q_1 + Q_2^T Q_2 + Q_3^T Q_3\right)\left((Q_1^T Q_1)^{-1} + (Q_2^T Q_2)^{-1} + (Q_3^T Q_3)^{-1}\right) - 3I$$

$$=$$

$$\left((Q_1^T Q_1)^{-1} + (Q_2^T Q_2)^{-1} + (Q_3^T Q_3)^{-1}\right) - 3I$$

The Higher-Order CS Decomposition

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 the (orthogonal) eigenvector matrix for

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

and $Q_k Z = U_k \Sigma_k =$ (Matrix with unit 2-norm columns)(Diagonal).

We won't need to compute all of this...

The Eigenvalues of T_N

The Connection Between S_N and T_N

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

where

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

Since we are interested in the eigenvalues of S_N that equal 1, we are interested in the eigenvalues of T_N that equal N.

The Eigenvalues of T_N

Key Result

Can show that if $T_N z = N \cdot z$ then

$$Q_k^T Q_k z = \frac{1}{N} z$$

for $k = 1{:}N$.

This says that z is a right singular vector for $Q_1, \ldots, Q_N$.

Further Properties

Let $Z^T T_N Z = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ be a Schur decomposition with

$$N = \lambda_1 = \cdots = \lambda_p < \lambda_{p+1} \leq \cdots \leq \lambda_n$$

and partition

$$Z = [Z^{(c)} \mid Z^{(u)}] \qquad U_k = [U_k^{(c)} \mid U_k^{(u)}] \qquad \Sigma_k = \begin{bmatrix} I_p/\sqrt{N} & 0 \\ 0 & \Sigma_k^{(u)} \end{bmatrix}$$

Then for $k = 1{:}N$

$$Q_k = U_k \Sigma_k Z^T = \frac{1}{\sqrt{N}} U_k^{(c)} Z^{(c)\,T} + U_k^{(u)} \Sigma_k^{(u)} Z^{(u)\,T}$$

and the columns of $U_k^{(c)}$ are orthonormal and

$$\mathrm{ran}(U_k^{(c)}) \perp \mathrm{ran}(U_k^{(u)})$$

Back to the HO-GSVD

- Thin QR: $A_k = Q_k R$.
- HO-CSD:

$$Q_k \;=\; U_k \Sigma_k Z^T \;=\; \frac{1}{\sqrt{N}} U_k^{(c)} Z^{(c)\,T} \;+\; U_k^{(u)} \Sigma_k^{(u)} Z^{(u)\,T}$$

and the columns of $U_k^{(c)}$ are orthonormal and

$$\mathrm{ran}(U_k^{(c)}) \;\perp\; \mathrm{ran}(U_k^{(u)})$$

- Setting $V^{(c)\,T} = Z^{(c)\,T} R$ and $V^{(u)\,T} = Z^{(c)\,T} R$ gives HO-GSVD:

$$A_k \;=\; U_k \Sigma_k Z^T R \;=\; \underbrace{\frac{1}{\sqrt{N}} U_k^{(c)}\, V^{(c)\,T}}_{\text{common part}} \;+\; \underbrace{U_k^{(u)}\, \Sigma_k^{(u)} V^{(u)\,T}}_{\text{uncommon part}}$$

Computing the Common HO-GSVD Subspace

Recall that if

$$\frac{(Q_1^T Q_1)^{-1} + \cdots + (Q_N^T Q_N)^{-1}}{N} z = Nz$$

then

$$Q_k^T Q_k z = \frac{1}{N} z$$

for $k = 1{:}N$.

This means that the common HO-GSVD subspace for $Q_1, \ldots, Q_N$ is the intersection of all H_{ij} where H_{ij} is the common HO-GSVD subspace associated with $\{Q_i, Q_j\}$.

Not the Only Show In Town

PARFAC2

Choose a parameter r that satisfies $r \leq n$ and a nonsingular $H \in \mathbb{R}^{r \times r}$ and then set out to minimize

$$\phi(U_1, \ldots, U_N, \Sigma_1, \ldots, \Sigma_N, V) = \sum_{k=1}^{N} \| A_k - U_k H \Sigma_k V^T \|_F^2$$

where

1. $V \in \mathbb{R}^{n \times r}$ has full column rank
2. each $U_k \in \mathbb{R}^{m_k \times r}$ has orthonormal columns
3. each $\Sigma_k \in \mathbb{R}^{r \times r}$ is diagonal

Is there a connection?

The All-Possible-Quotients Quadratic Form

Definition

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

Gradient

$$\nabla \phi(x) = c \cdot \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(\frac{\| A_i x \|^2}{\| A_j x \|^2} - \frac{\| A_j x \|^2}{\| A_i x \|^2} \right) \left(\frac{A_i^T A_i x}{\| A_i x \|^2} - \frac{A_j^T A_j x}{\| A_j x \|^2} \right)$$

The stationary vectors x for which $\phi(x) = 1$ relate to the common HO-GSVD subspace. This may point the way to interesting techniques for large and sparse $A_1, \ldots, A_N$.

Some Theorems are Missing

What if S_N has an minimum eigenvalue that is slightly bigger than 1?

Then we have an approximate HO-GSVD common subspace. And the associated u-vectors are approximate left singular vectors. HOW APPROXIMATE?

If everything is approximate, what are the ramifications when it comes to identifying common features in $A_1, \ldots, A_N$?

Summary

At the top level, the transformation matrices in the HO-GSVD are not orthogonal.

However, we only used a "subset" of the HO-GSVD and that subset has orthogonal features.

Those features made it possible to formulate a stable procedure that could identify common factors in the data matrix collection $\{A_1, \ldots, A_N\}$.

Now back to the

BIG

Picture...

To: Previous Speakers
From: C. Van Loan
Date: August 10, 2013
Subject: Thanks a lot for the tiny orthogononal complement!

How can I possibly add to the space?

$$\mathcal{S}_{\text{Overton}} = \text{span}\{v_1, \ldots, v_{26}, ?\}$$

To: Previous Speakers
From: C. Van Loan
Date: August 10, 2013
Subject: Thanks a lot for the tiny orthogononal complement!

How can I possibly add to the space?

$$\mathcal{S}_{\text{Overton}} \;=\; \text{span}\{v_1,\ldots,v_{26},v_{27}\}$$

$$\uparrow$$

$$v_{27} \;=\; \text{GVL4 Typo Space}$$

GVL4: Pages 401-402

$$C = \begin{bmatrix} \lambda & \times & \times & \times & \times & \times & \times \\ 0 & \lambda & \times & \times & \times & \times & \times \\ 0 & 0 & \lambda & \times & \times & \times & \times \\ 0 & 0 & 0 & \lambda & \times & \times & \times \\ 0 & 0 & 0 & 0 & \lambda & \times & \times \\ 0 & 0 & 0 & 0 & 0 & \lambda & \times \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

$$\Downarrow$$

$$V^T C V = \begin{bmatrix} \lambda & 0 & 0 & 0 & \times & \times & \times \\ 0 & \lambda & 0 & 0 & \times & \times & \times \\ 0 & 0 & \lambda & 0 & \times & \times & \times \\ 0 & 0 & 0 & \lambda & \times & \times & \times \\ 0 & 0 & 0 & 0 & \color{red}{\lambda} & \color{red}{\times} & \color{red}{a} \\ 0 & 0 & 0 & 0 & \color{red}{0} & \color{red}{\lambda} & \color{red}{0} \\ 0 & 0 & 0 & 0 & \color{red}{0} & \color{red}{0} & \color{red}{\lambda} \end{bmatrix}$$

} 4 blocks of order 1 or larger

} 2 blocks of order 2 or larger

} 1 block of order 3 or larger

How Michael says “You Screwed Up”

“Could there have been a shift in notation at some point? Or I am simply blind/idiotic?”

How Michael says "Look on the Bright Side"

"I suppose ... it can be material for your August talk!"

GVL4: Pages 401-402

$$C = \begin{bmatrix} \lambda & \times & \times & \times & \times & \times & \times \\ 0 & \lambda & \times & \times & \times & \times & \times \\ 0 & 0 & \lambda & \times & \times & \times & \times \\ 0 & 0 & 0 & \lambda & \times & \times & \times \\ 0 & 0 & 0 & 0 & \lambda & \times & \times \\ 0 & 0 & 0 & 0 & 0 & \lambda & \times \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

$$\Downarrow$$

$$V^T C V = \begin{bmatrix} \lambda & 0 & 0 & 0 & \times & \times & \times \\ 0 & \lambda & 0 & 0 & \times & \times & \times \\ 0 & 0 & \lambda & 0 & \times & \times & \times \\ 0 & 0 & 0 & \lambda & \times & \times & \times \\ 0 & 0 & 0 & 0 & \lambda & \times & a \\ 0 & 0 & 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix} \begin{matrix} \left.\right\} \text{4 blocks of order 1 or larger} \\ \left.\right\} \text{2 blocks of order 2 or larger} \\ \left.\right\} \text{1 block of order 3 or larger} \end{matrix}$$

GVL4: Pages 401-402 (Corrected)

$$C = \begin{bmatrix} \lambda & \times & \times & \times & \times & \times & \times \\ 0 & \lambda & \times & \times & \times & \times & \times \\ 0 & 0 & \lambda & \times & \times & \times & \times \\ 0 & 0 & 0 & \lambda & \times & \times & \times \\ 0 & 0 & 0 & 0 & \lambda & \times & \times \\ 0 & 0 & 0 & 0 & 0 & \lambda & \times \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

$$\Downarrow$$

$$V^T C V = \begin{bmatrix} \lambda & 0 & 0 & 0 & \times & \times & \times \\ 0 & \lambda & 0 & 0 & \times & \times & \times \\ 0 & 0 & \lambda & 0 & \times & \times & \times \\ 0 & 0 & 0 & \lambda & \times & \times & \times \\ 0 & 0 & 0 & 0 & \lambda & 0 & a \\ 0 & 0 & 0 & 0 & 0 & \lambda & b \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix}
\begin{matrix} \left.\vphantom{\begin{matrix}1\\1\\1\\1\end{matrix}}\right\} \text{4 blocks of order 1 or larger} \\ \left.\vphantom{\begin{matrix}1\\1\end{matrix}}\right\} \text{2 blocks of order 2 or larger} \\ \} \text{ 1 block of order 3 or larger} \end{matrix}$$

How I graciously said “Thanks For the Correction”

“You found a typo that has been out there for decades.”

“Perhaps that is why the Tacoma bridge collapsed!”

How Michael "Rubbed It In"

"Perhaps this is also why the Mt Vernon I-5 bridge collapsed, which is the one between Seattle airport and our new place in Bellingham."

How Michael "Wouldn't Let Go"!

"Too bad GVL4 wasn't fixed in time!"

How Michael said "What is In It For Me?!"

I guess your ill-conceived 5-dollar per typo program has expired!

How I demonstrated great flexibility!

“OK, it is now a one-cheap-brew-per-typo”

Yes, Michael Really Is Honest and Cerebral

"Sounds good!"