EXTRACTING ACCURATE SINGULAR VALUES FROM APPROXIMATE SUBSPACES

LORENZO LAZZARINO

Mathematical Institute - University of Oxford

Computational Mathematics Theme - STFC UKRI

PYSANUM, 10th October 2024

UNIVERSITY OF OXFORD
Mathematical Institute

Oxford Mathematics

EXTRACTING ACCURATE SINGULAR VALUES FROM APPROXIMATE SUBSPACES

INTRODUCTION

1

NUMERICAL LINEAR ALGEBRA

We devise and analyse methods for:

- Linear System: $$A \, x = b$$
- Eigenvalue Problem: $$A \, x = \lambda \, x$$

(•) Singular Value Decomposition:
  - Find (approximate) singular subspaces
  - Find (approximate) singular values
  - Low-rank approximations

NUMERICAL LINEAR ALGEBRA

We devise and analyse methods for:

- Linear System: $$\boxed{A}\;\boxed{x} = \boxed{b}$$
- Eigenvalue Problem: $$\boxed{A}\;\boxed{x} = \lambda\;\boxed{x}$$

(•) Singular Value Decomposition:
  - Find (approximate) singular subspaces
  - Find (approximate) singular values
  - Low-rank approximations

NUMERICAL LINEAR ALGEBRA

We devise and analyse methods for:

- Linear System: $\boxed{\quad A \quad}\,\boxed{x} = \boxed{b}$

- Eigenvalue Problem: $\boxed{\quad A \quad}\,\boxed{x} = \lambda\,\boxed{x}$

(•) Singular Value Decomposition:
- Find (approximate) singular subspaces
- Find (approximate) singular values
- Low-rank approximations

How?

- Complexity
- Accuracy
- Stability
- Use of inputs (e.g. Numeber of passes)

NUMERICAL LINEAR ALGEBRA

We devise and analyse methods for:

- Linear System: $\boxed{A}\,\boxed{x} = \boxed{b}$
- Eigenvalue Problem: $\boxed{A}\,\boxed{x} = \lambda\,\boxed{x}$

(•) Singular Value Decomposition:
- Find (approximate) singular subspaces
- Find (approximate) singular values
- Low-rank approximations

How?
- Complexity
- Accuracy
- Stability
- Use of inputs (e.g. Numeber of passes)

NLA TOOLS

- $\|A\|_F = \sqrt{\sum_i \sum_j |a_{ij}|^2}$, $\quad \|A\|_2 = \sup_x \frac{\|Ax\|_2}{\|x\|_2}$, with $\|x\|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2}$

- Orthogonal matrix: $\underset{m \times m}{Q^*}\, Q = I_m = Q\, Q^*$

- Orthonormal matrix: $\underset{n \times m}{Q^*}\, Q = I_n$

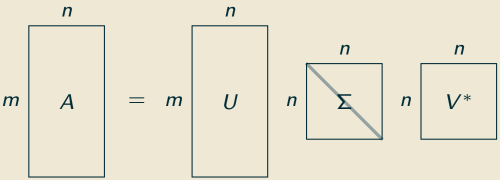
- QR factorization: For any $A \in \mathbb{R}^{m \times n}$ there exists a factorization $\underset{m \times n}{A} = \underset{m \times n}{Q}\, \underset{n \times n}{R}$

  where Q is orthonormal and R is upper triangular.

SINGULAR VALUE DECOMPOSITION

Singular Value Decomposition (SVD)

Any matrix A has the decomposition (assume $m \geq n$):



$$= \sum_{i=1}^{n} \sigma_i u_i v_i^*$$

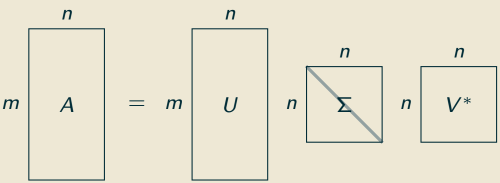
where $\Sigma = diag(\sigma_1, \dots, \sigma_n)$, with $(\sigma_{max} :=)\sigma_1 \geq \cdots \geq \sigma_n \geq 0$, and U, V are orthonormal matrices, that is, $U^*U = V^*V = I_n$.

Sec. 2.4 (Golub, Van Loan)
Lect. 4 (Trefethen, Bau, 2022)

SINGULAR VALUE DECOMPOSITION

Singular Value Decomposition (SVD)

Any matrix A has the decomposition (assume $m \geq n$):



$$= \sum_{i=1}^{n} \sigma_i u_i v_i^*$$

where $\Sigma = diag(\sigma_1, \dots, \sigma_n)$, with $(\sigma_{max} :=)\sigma_1 \geq \cdots \geq \sigma_n \geq 0$, and U, V are orthonormal matrices, that is, $U^*U = V^*V = I_n$.

Sec. 2.4 (Golub, Van Loan)
Lect. 4 (Trefethen, Bau, 2022)

Existance:
Always, from eigenvalues of A^*A

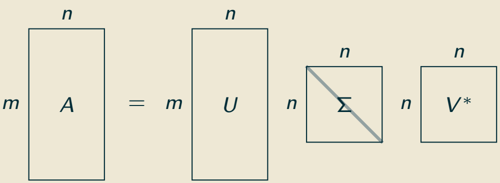
Uniqueness:

- Singular vectors
  - Can be fliped by signs
  - Multiple singular values
- Singular values
  - Always unique

SINGULAR VALUE DECOMPOSITION

Singular Value Decomposition (SVD)

Any matrix A has the decomposition (assume $m \geq n$):



$$= \sum_{i=1}^{n} \sigma_i u_i v_i^*$$

where $\Sigma = diag(\sigma_1, \ldots, \sigma_n)$, with $(\sigma_{max} :=)\sigma_1 \geq \cdots \geq \sigma_n \geq 0$, and U, V are orthonormal matrices, that is, $U^*U = V^*V = I_n$.

Sec. 2.4 (Golub, Van Loan)
Lect. 4 (Trefethen, Bau, 2022)

- $\sigma_i = \sqrt{\lambda_i(A^*A)}$, for $i = 1, \ldots n$
- $\|A\|_2 = \sigma_{max}$ and $\|A\|_F^2 = \sum_{i=1}^{n} \sigma_i^2$
- "full" SVD: $A = \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*$
- $\sigma_i(A) = \sigma_i(Q_1 A Q_2)$ for any Q_1, Q_2 orthogonal
- Can be computed by, e.g., Golub-Kahan bi-diagonalization cost $\mathcal{O}(mn^2)$

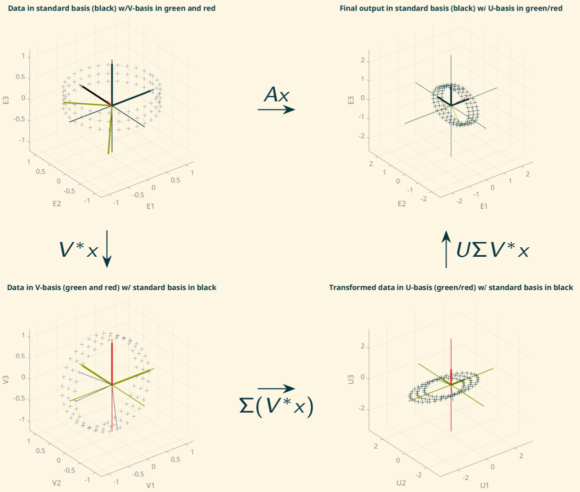
SINGULAR VALUE DECOMPOSITION › *Why do we care?*

It's beautiful!

Theoretical Beauty

- Existance
- Info about: norms, rank, subspaces
- Low-rank optimality
- reduce difficulties of problems:
  Linear system, eigenvalue problem, inverse problem
- Pseudoinverse

> Example by Eric Thomson, definetly worth having a look at http://neurochannels.blogspot.com/2008/02/visualizing-svd.html



Mathematical Institute

It's beautiful!

Applied Beauty

- Quantum information
- Immunology
- Molecular dynamics
- Information retrieval
- Pattern Recognition
- Weather forecast
- Astrodynamics
- Small-angle scattering

Mathematical Institute

It's beautiful!

Applied Beauty

- Gene expression data
- Quantum information
- Immunology
- Molecular dynamics
- Information retrieval
- Pattern Recognition
- Weather forecast
- Astrodynamics
- Small-angle scattering

It's beautiful!

Applied Beauty

- Signal Processing
- Gene expression data
- Quantum information
- Immunology
- Molecular dynamics
- Information retrieval
- Pattern Recognition
- Weather forecast
- Astrodynamics

SINGULAR VALUE DECOMPOSITION › *Why do we care?*

It's beautiful!

Applied Beauty

- Imaging processing and compression
- Signal Processing
- Gene expression data
- Quantum information
- Immunology
- Molecular dynamics
- Information retrieval
- Pattern Recognition
- Weather forecast

SINGULAR VALUE DECOMPOSITION › *Why do we care?*

It's beautiful!

Applied Beauty

- Choosing a Pizzeria

300 samples measuring 7 features of Pizze from 10 different Pizzerie!

> Brilliant example by Joachim Schork, see https://statisticsglobe.com/principal-component-analysis-pca

Pizzeria	water	protein	fat	ash	sodium	carbohydrates	calories
A	30.49	21.28	41.65	4.82	1.64	1.76	4.67
A	32.20	19.25	43.42	4.62	1.50	0.51	4.70
⋮							
B	50.33	13.96	29.25	3.42	0.96	3.04	3.31
⋮							
C	49.10	24.53	21.08	2.84	0.34	2.45	2.98
⋮							
D	47.45	22.37	20.97	4.06	0.70	5.15	2.99
⋮							
J	44.91	11.07	17.00	2.49	0.66	25.36	2.91

Mathematical Institute

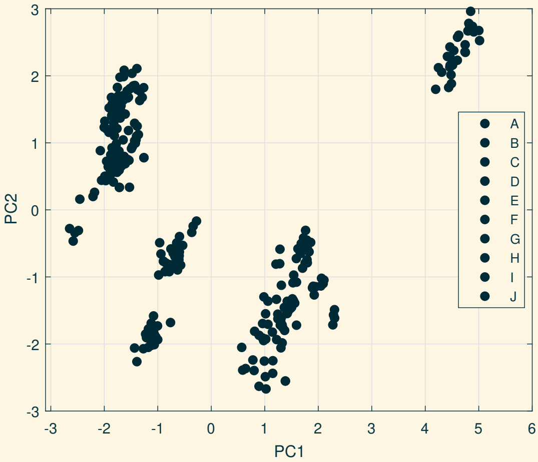
It's beautiful!

Applied Beauty

- Choosing a Pizzeria

300 samples measuring 7 features of Pizze from 10 different Pizzerie!

Brilliant example by Joachim Schork, see https://statisticsglobe.com/principal-component-analysis-pca



SINGULAR VALUE DECOMPOSITION › *Why do we care?*

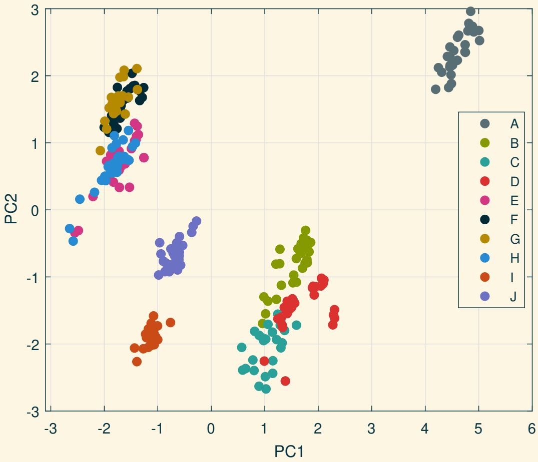
It's beautiful!

Applied Beauty

- Choosing a Pizzeria

300 samples measuring 7 features of Pizze from 10 different Pizzerie!

Brilliant example by Joachim Schork, see https://statisticsglobe.com/principal-component-analysis-pca



SINGULAR VALUE DECOMPOSITION › *Why do we care?*

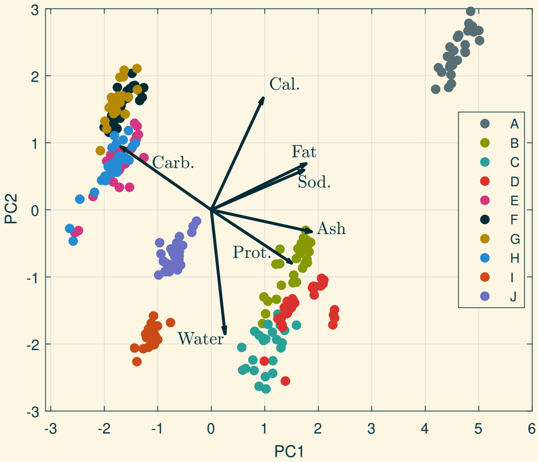
It's beautiful!

Applied Beauty

- Choosing a Pizzeria

300 samples measuring 7 features of Pizze from 10 different Pizzerie!

Brilliant example by Joachim Schork, see https://statisticsglobe.com/principal-component-analysis-pca



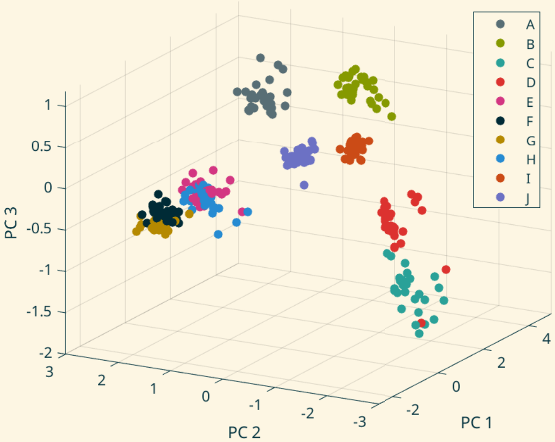
It's beautiful!

Applied Beauty

- Choosing a Pizzeria

300 samples measuring 7 features of Pizze from 10 different Pizzerie!

Brilliant example by Joachim Schork, see https://statisticsglobe.com/principal-component-analysis-pca



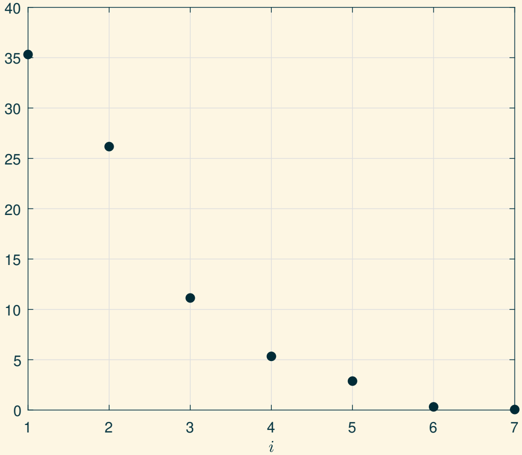
SINGULAR VALUE DECOMPOSITION › *Why do we care?*

It's beautiful!

Applied Beauty

- Choosing a Pizzeria

300 samples measuring 7 features of Pizze from 10 different Pizzerie!



> Brilliant example by Joachim Schork, see https://statisticsglobe.com/principal-component-analysis-pca

PROBLEM SETTING AND CLASSICAL APPROACHES

2

PROBLEM SETTING

Given $\tilde{U}$ and/or $\tilde{V}$ approximations of the leading singular subspaces of A

$$\begin{array}{cc} & r \\ n & \left[\ \tilde{V}\ \right] \end{array}, \quad \begin{array}{cc} & r+\ell \\ m & \left[\ \ \tilde{U}\ \ \right] \end{array}$$

AIM: Approximate the leading singular values $\{\sigma_i(A)\}_{i=1}^r$

PROBLEM SETTING

Given $\tilde{U}$ and/or $\tilde{V}$ approximations of the leading singular subspaces of A

$$\begin{array}{cc} & r \\ n & \begin{bmatrix} \\ \tilde{V} \\ \\ \end{bmatrix} \end{array}, \quad \begin{array}{cc} & r+\ell \\ m & \begin{bmatrix} \\ \\ \tilde{U} \\ \\ \\ \end{bmatrix} \end{array}$$

AIM: Approximate the leading singular values $\{\sigma_i(A)\}_{i=1}^{r}$

$\tilde{U}$ and $\tilde{V}$ could be obtained by:

- Subspace iteration
- Randomized techniques
- ...

PROBLEM SETTING

Given $\tilde{U}$ and/or $\tilde{V}$ approximations of the leading singular subspaces of A

$$\begin{array}{cc} & r \\ n & \begin{bmatrix} \\ \tilde{V} \\ \\ \end{bmatrix} \end{array}, \quad \begin{array}{cc} & r+\ell \\ m & \begin{bmatrix} \\ \\ \tilde{U} \\ \\ \\ \end{bmatrix} \end{array}$$

AIM: Approximate the leading singular values $\{\sigma_i(A)\}_{i=1}^{r}$

$\tilde{U}$ and $\tilde{V}$ could be obtained by:

- Subspace iteration
- Randomized techniques
- ...

Main message:

- $\tilde{U}$ or $\tilde{V}$ → (one-sided) SVD
- $\tilde{U}$ and $\tilde{V}$ → generalized Nyström
- Multiple passes with A → HMT

CLASSICAL APPROACHES > *Rayleigh Ritz and (one-sided) SVD approximations*

Rayleigh Ritz (RR)

$$\sigma_i(A) \approx \sigma_i(\tilde{U}^* A \tilde{V}) =: \sigma_i(A_{RR,\tilde{V},\tilde{U}})$$

(Dax, 2012)
(Saad, 2011)
(Xin-guo, 1992)

CLASSICAL APPROACHES › *Rayleigh Ritz and (one-sided) SVD approximations*

Mathematical Institute

Rayleigh Ritz (RR)

$$\sigma_i(A) \approx \sigma_i(\tilde{U}^* A \tilde{V}) =: \sigma_i(A_{RR,\tilde{V},\tilde{U}})$$

(Dax, 2012)
(Saad, 2011)
(Xin-guo, 1992)

- $N_r + \mathcal{O}(mr^2) + \mathcal{O}(r^3)$
- Single-pass
- 1 multiplication by A

CLASSICAL APPROACHES › *Rayleigh Ritz and (one-sided) SVD approximations*

Rayleigh Ritz (RR)

$$\sigma_i(A) \approx \sigma_i(\tilde{U}^* A \tilde{V}) =: \sigma_i(A_{RR,\tilde{V},\tilde{U}})$$

(Dax, 2012)
(Saad, 2011)
(Xin-guo, 1992)

- $N_r + \mathcal{O}(mr^2) + \mathcal{O}(r^3)$
- Single-pass
- 1 multiplication by A

$$Q_1 = \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \tilde{V} & \tilde{V}_\perp \end{bmatrix}$$

CLASSICAL APPROACHES › *Rayleigh Ritz and (one-sided) SVD approximations*

Rayleigh Ritz (RR)

$$\sigma_i(A) \approx \sigma_i(\tilde{U}^* A \tilde{V}) =: \sigma_i(A_{RR,\tilde{V},\tilde{U}})$$

(Dax, 2012)
(Saad, 2011)
(Xin-guo, 1992)

- $N_r + \mathcal{O}(mr^2) + \mathcal{O}(r^3)$
- Single-pass
- 1 multiplication by A

$$Q_1 = \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \tilde{V} & \tilde{V}_\perp \end{bmatrix}$$

$$\bar{A} = Q_1^* A Q_2$$

$$\sigma_i(A_{RR,\tilde{V},\tilde{U}}) = \sigma_i(\bar{A}_{RR,\begin{bmatrix} I_r \\ 0 \end{bmatrix},\begin{bmatrix} I_{r+\ell} \\ 0 \end{bmatrix}})|$$

$$= \sigma_i(\bar{A}_{11}) = \sigma_i\left(\begin{bmatrix} \bar{A}_{11} & 0 \\ 0 & 0 \end{bmatrix}\right)$$

CLASSICAL APPROACHES > *Rayleigh Ritz and (one-sided) SVD approximations*

Rayleigh Ritz (RR)

$$\sigma_i(A) \approx \sigma_i(\tilde{U}^* A \tilde{V}) =: \sigma_i(A_{RR,\tilde{V},\tilde{U}})$$

(Dax, 2012)
(Saad, 2011)
(Xin-guo, 1992)

(one-sided) SVD approximations

$$\sigma_i(A) \approx \sigma_i(A\tilde{V}) =: \sigma_i(A_{SVD,\tilde{V}})$$

- $N_r + \mathcal{O}(mr^2) + \mathcal{O}(r^3)$
- Single-pass
- 1 multiplication by A

$$Q_1 = \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \tilde{V} & \tilde{V}_\perp \end{bmatrix}$$

$$\bar{A} = Q_1^* A Q_2$$

$$\sigma_i(A_{RR,\tilde{V},\tilde{U}}) = \sigma_i(\bar{A}_{RR,\left[\begin{smallmatrix} I_r \\ 0 \end{smallmatrix}\right],\left[\begin{smallmatrix} I_{r+\ell} \\ 0 \end{smallmatrix}\right]})|$$

$$= \sigma_i(\bar{A}_{11}) = \sigma_i\left(\begin{bmatrix} \bar{A}_{11} & 0 \\ 0 & 0 \end{bmatrix}\right)$$

CLASSICAL APPROACHES › *Rayleigh Ritz and (one-sided) SVD approximations*

Rayleigh Ritz (RR)

$$\sigma_i(A) \approx \sigma_i(\tilde{U}^* A \tilde{V}) =: \sigma_i(A_{RR,\tilde{V},\tilde{U}})$$

- $N_r + \mathcal{O}(mr^2) + \mathcal{O}(r^3)$
- Single-pass
- 1 multiplication by A

(Dax, 2012)
(Saad, 2011)
(Xin-guo, 1992)

(one-sided) SVD approximations

$$\sigma_i(A) \approx \sigma_i(A\tilde{V}) =: \sigma_i(A_{SVD,\tilde{V}})$$

- $N_r + \mathcal{O}(mr^2)$
- Single-pass
- 1 multiplication by A

$$Q_1 = \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \tilde{V} & \tilde{V}_\perp \end{bmatrix}$$

$$\bar{A} = Q_1^* A Q_2$$

$$\sigma_i(A_{RR,\tilde{V},\tilde{U}}) = \sigma_i(\bar{A}_{RR,\left[\begin{smallmatrix} I_r \\ 0 \end{smallmatrix}\right],\left[\begin{smallmatrix} I_{r+\ell} \\ 0 \end{smallmatrix}\right]})|$$

$$= \sigma_i(\bar{A}_{11}) = \sigma_i\left(\begin{bmatrix} \bar{A}_{11} & 0 \\ 0 & 0 \end{bmatrix}\right)$$

CLASSICAL APPROACHES › *Rayleigh Ritz and (one-sided) SVD approximations*

Mathematical Institute

Rayleigh Ritz (RR)

$$\sigma_i(A) \approx \sigma_i(\tilde{U}^* A \tilde{V}) =: \sigma_i(A_{RR,\tilde{V},\tilde{U}})$$

- $N_r + \mathcal{O}(mr^2) + \mathcal{O}(r^3)$
- Single-pass
- 1 multiplication by A

(Dax, 2012)
(Saad, 2011)
(Xin-guo, 1992)

(one-sided) SVD approximations

$$\sigma_i(A) \approx \sigma_i(A\tilde{V}) =: \sigma_i(A_{SVD,\tilde{V}})$$

- $N_r + \mathcal{O}(mr^2)$
- Single-pass
- 1 multiplication by A

$$Q_1 = \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \tilde{V} & \tilde{V}_\perp \end{bmatrix}$$

$$\bar{A} = Q_1^* A Q_2$$

$$\sigma_i(A_{RR,\tilde{V},\tilde{U}}) = \sigma_i(\bar{A}_{RR, \begin{bmatrix} I_r \\ 0 \end{bmatrix}, \begin{bmatrix} I_{r+\ell} \\ 0 \end{bmatrix}})|$$

$$= \sigma_i(\bar{A}_{11}) = \sigma_i\left(\begin{bmatrix} \bar{A}_{11} & 0 \\ 0 & 0 \end{bmatrix}\right)$$

$$\tilde{A} = AQ_2 = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 \end{bmatrix}$$

$$\sigma_i(A_{SVD,\tilde{V}}) = \sigma_i(\tilde{A}_{SVD, \begin{bmatrix} I_r \\ 0 \end{bmatrix}})$$

$$= \sigma_i(\begin{bmatrix} \tilde{A}_1 & 0 \end{bmatrix})$$

CLASSICAL APPROACHES > *Rayleigh Ritz and (one-sided) SVD approximations* > *Accuracy*

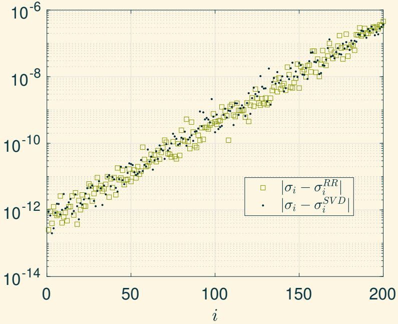
Rayleigh Ritz (RR)

$$\sigma_i(A) \approx \sigma_i(\tilde{U}^* A \tilde{V}) =: \sigma_i(A_{RR,\tilde{V},\tilde{U}})$$

(Dax, 2012)
(Saad, 2011)
(Xin-guo, 1992)

(one-sided) SVD approximations

$$\sigma_i(A) \approx \sigma_i(A\tilde{V}) =: \sigma_i(A_{SVD,\tilde{V}})$$



CLASSICAL APPROACHES › *Rayleigh Ritz and (one-sided) SVD approximations* › *Accuracy*

Rayleigh Ritz (RR)

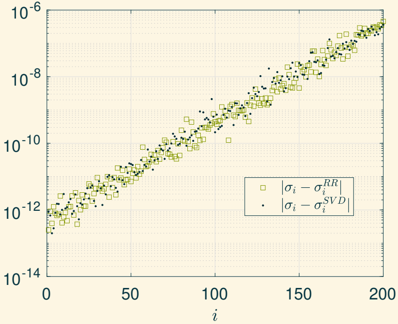
$$\sigma_i(A) \approx \sigma_i(\tilde{U}^* A \tilde{V}) =: \sigma_i(A_{RR,\tilde{V},\tilde{U}})$$

(Dax, 2012)
(Saad, 2011)
(Xin-guo, 1992)

(one-sided) SVD approximations

$$\sigma_i(A) \approx \sigma_i(A\tilde{V}) =: \sigma_i(A_{SVD,\tilde{V}})$$

Not bad...



CLASSICAL APPROACHES > *Rayleigh Ritz and (one-sided) SVD approximations* > *Accuracy*

Mathematical Institute

Rayleigh Ritz (RR)

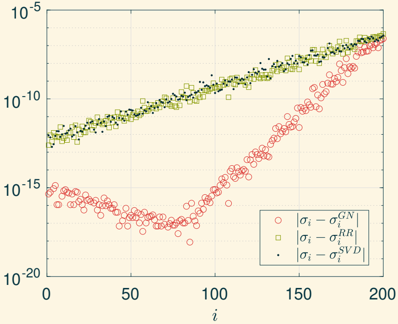
$$\sigma_i(A) \approx \sigma_i(\tilde{U}^* A \tilde{V}) =: \sigma_i(A_{RR,\tilde{V},\tilde{U}})$$

(Dax, 2012)
(Saad, 2011)
(Xin-guo, 1992)

(one-sided) SVD approximations

$$\sigma_i(A) \approx \sigma_i(A\tilde{V}) =: \sigma_i(A_{SVD,\tilde{V}})$$

Not bad...



BUT,
what if we could have this?

TECHNIQUES FROM: (RANDOMIZED) LOW-RANK APPROXIMATIONS

3

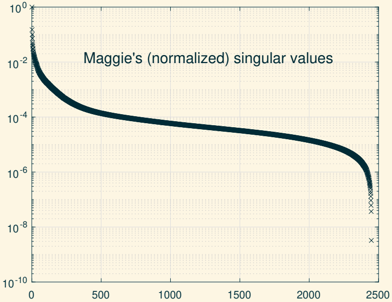
(NUMERICAL) RANK

- A has **rank** k if there exists E and F such that: $$m\,\overset{n}{\boxed{A}} = m\,\overset{k}{\boxed{E}}\; k\,\overset{n}{\boxed{F^*}}$$
  - rank = number of non-zero singular values
    $A^\dagger := V\ diag(\sigma_1^{-1}, \ldots, \sigma_k^{-1}, 0, \ldots, 0)U^*$
- A has ϵ-**rank** k if there exists E and F such that: $\|A - EF^*\| \leq \epsilon$
  - ϵ-rank = number of singular values greater than ϵ

(NUMERICAL) RANK

- A has **rank** k if there exists E and F such that: $$m \overset{n}{\boxed{A}} = m \overset{k}{\boxed{E}} \; k \overset{n}{\boxed{F^*}}$$
  - rank = number of non-zero singular values
    $A^\dagger := V\ diag(\sigma_1^{-1}, \dots, \sigma_k^{-1}, 0, \dots, 0) U^*$
- A has ϵ-**rank** k if there exists E and F such that: $\|A - EF^*\| \leq \epsilon$
  - ϵ-rank = number of singular values greater than ϵ

Maggie - 2448×2448

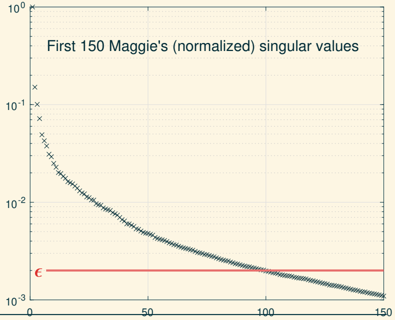


(NUMERICAL) RANK

- A has **rank** k if there exists E and F such that: $$m\ \overset{n}{\boxed{A}} = m\ \overset{k}{\boxed{E}}\ k\ \overset{n}{\boxed{F^*}}$$
  - rank = number of non-zero singular values
    $A^\dagger := V\ diag(\sigma_1^{-1}, \dots, \sigma_k^{-1}, 0, \dots, 0) U^*$
- A has ϵ-**rank** k if there exists E and F such that: $\|A - EF^*\| \leq \epsilon$
  - ϵ-rank = number of singular values greater than ϵ

Maggie - 2448×2448

rank 100



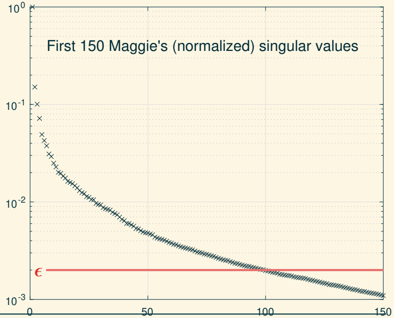
(NUMERICAL) RANK

- A has **rank** k if there exists E and F such that: $\quad m\underset{}{\overset{n}{\boxed{A}}} = m\overset{k}{\boxed{E}}\; k\overset{n}{\boxed{F^*}}$
  - rank = number of non-zero singular values
    $A^\dagger := V\ diag(\sigma_1^{-1}, \ldots, \sigma_k^{-1}, 0, \ldots, 0)U^*$
- A has ϵ-**rank** k if there exists E and F such that: $\|A - EF^*\| \leq \epsilon$
  - ϵ-rank = number of singular values greater than ϵ

Maggie - 2448 $\times$ 2448

rank 100

rank 5



(RANDOMIZED) LOW-RANK APPROXIMATIONS

Given a fix rank r, find $E \in \mathbb{R}^{m \times r}$ and $F \in \mathbb{R}^{n \times r}$ such that $A \approx EF^*$

$$A_r = \sum_{i=1}^{r} \sigma_i u_i v_i^*$$

is the best rank-r approximation of A in both 2-norm and F-norm

- $\|A - A_r\|_2 = \sigma_{r+1}$
- $\|A - A_r\|_F = \sqrt{\sigma_{r+1}^2 + \cdots + \sigma_{rank(A)}^2}$

(RANDOMIZED) LOW-RANK APPROXIMATIONS

Given a fix rank r, find $E \in \mathbb{R}^{m\times r}$ and $F \in \mathbb{R}^{n\times r}$ such that $A \approx EF^*$

$$A_r = \sum_{i=1}^{r} \sigma_i u_i v_i^*$$

is the best rank-r approximation of A in both 2-norm and F-norm

- $\|A - A_r\|_2 = \sigma_{r+1}$
- $\|A - A_r\|_F = \sqrt{\sigma_{r+1}^2 + \cdots + \sigma_{rank(A)}^2}$

Classical Approach

$$\|A - A_r\| = \|A - U_r U_r^* A\| = \inf_{P = r\text{-dim orth. proj.}} \|A - PA\|$$

→ Find cheaper (but not optimal) orthogonal projections: e.g.

- Gram-Schmidt on the columns/rows of A
  - cost $\mathcal{O}(mnr)$

(RANDOMIZED) LOW-RANK APPROXIMATIONS

Given a fix rank r, find $E \in \mathbb{R}^{m \times r}$ and $F \in \mathbb{R}^{n \times r}$ such that $A \approx EF^*$

$$A_r = \sum_{i=1}^{r} \sigma_i u_i v_i^*$$

is the best rank-r approximation of A in both 2-norm and F-norm

- $\|A - A_r\|_2 = \sigma_{r+1}$
- $\|A - A_r\|_F = \sqrt{\sigma_{r+1}^2 + \cdots + \sigma_{rank(A)}^2}$

Classical Approach

$$\|A - A_r\| = \|A - U_k U_K^* A\| = \inf_{P = k\text{-dim orth. proj.}} \|A - PA\|$$

→ Find cheaper (but not optimal) orthogonal projections: e.g.

- Gram-Schmidt on the columns/rows of A - cost $\mathcal{O}(mnr)$

Randomized Approach

Use randomization for a model reduction while (approximately) preserving properties of the big problem

Sketching → Random Embedding

- ☺ Reduced costs
- ☺ (often) near-optimal solutions
- 😐 Different outputs
- 😐 Can fail (with small probability)

RANDOMIZED SVD (HMT)

Randomized SVD

$$A \approx (A\Omega)(A\Omega)^{\dagger}A =: A_{HMT,\Omega}$$

(Clarkson, Woodruff, 2017)
(Halko, Martinsson, Tropp, 2011)
(Rokhlin, Szlam, Tygert, 2009)

1. Choose $\Omega \in \mathbb{R}^{n\times r}$ 2. Sketch: $X = A\Omega$ 3. $[Q,\sim] = \mathrm{qr}(X,0)$ 4. $A_{HMT,\Omega} = Q(Q^*A)$

RANDOMIZED SVD (HMT)

Randomized SVD

$$A \approx (A\Omega)(A\Omega)^{\dagger} A =: A_{HMT,\Omega}$$

(Clarkson, Woodruff, 2017)
(Halko, Martinsson, Tropp, 2011)
(Rokhlin, Szlam, Tygert, 2009)

1. Choose $\Omega \in \mathbb{R}^{n \times r}$ 2. Sketch: $X = A\Omega$ 3. $[Q,\sim] = \mathrm{qr}(X,0)$ 4. $A_{HMT,\Omega} = Q(Q^* A)$

- $N_r + \mathcal{O}(mr^2) + \tilde{N}_r$
- Double-pass
- 2 multiplications by A

RANDOMIZED SVD (HMT)

Randomized SVD

$$A \approx (A\Omega)(A\Omega)^{\dagger}A =: A_{HMT,\Omega}$$

(Clarkson, Woodruff, 2017)
(Halko, Martinsson, Tropp, 2011)
(Rokhlin, Szlam, Tygert, 2009)

1. Choose $\Omega \in \mathbb{R}^{n \times r}$ 2. Sketch: $X = A\Omega$ 3. $[\mathrm{Q},\sim] = \mathrm{qr}(X,0)$ 4. $A_{HMT,\Omega} = Q(Q^*A)$

- $N_r + \mathcal{O}(mr^2) + \tilde{N}_r$
- Double-pass
- 2 multiplications by A

Accuracy

$\hat{r} \leq r - 2$

$$\mathbb{E}\|A - A_{HMT,\Omega}\|_F \leq \sqrt{1 + \frac{r}{r - \hat{r} - 1}}\|A - A_{best,\hat{r}}\|_F$$

(Halko, Martinsson, Tropp, 2011)

RANDOMIZED SVD (HMT)

Randomized SVD

$$A \approx (A\Omega)(A\Omega)^{\dagger} A =: A_{HMT,\Omega}$$

(Clarkson, Woodruff, 2017)
(Halko, Martinsson, Tropp, 2011)
(Rokhlin, Szlam, Tygert, 2009)

1. Choose $\Omega \in \mathbb{R}^{n\times r}$ 2. Sketch: $X = A\Omega$ 3. $[\mathrm{Q},\sim] = \mathrm{qr}(X,0)$ 4. $A_{HMT,\Omega} = Q(Q^* A)$

- $N_r + \mathcal{O}(mr^2) + \tilde{N}_r$
- Double-pass
- 2 multiplications by A

Accuracy

$\hat{r} \leq r - 2$

$$\mathbb{E}\|A - A_{HMT,\Omega}\|_F \leq \sqrt{1 + \frac{r}{r - \hat{r} - 1}}\|A - A_{best,\hat{r}}\|_F$$

(Halko, Martinsson, Tropp, 2011)

Stability

Stable under rounding errors if computed with Householder QR

(Connolly, Higham, Pranesh, 2022)

GENERALIZED NYSTRÖM APPROXIMATION

Generalized Nyström

$$A \approx A\Omega_1(\Omega_2^* A\Omega_1)^\dagger \Omega_2^* A =: A_{GN,\Omega_1,\Omega_2}$$

(Clarkson, Woodruff, 2009)
(Nakatsukasa, 2020)
(Woolfe, Liberty, Rokhlin, Tygert, 2008)

1. Choose $\Omega_1 \in \mathbb{R}^{n\times r}, \Omega_2 \in \mathbb{R}^{m\times(r+\ell)}$
2. Two-side Sketch: $X = A\Omega_1$ and $Y = \Omega_2^* A$
3. $[\mathrm{Q,R}] = \mathrm{qr}(Y\Omega_1, 0)$
4. $A_{GN,\Omega_1,\Omega_2} = (XR^{-1})(Q^* Y)$

GENERALIZED NYSTRÖM APPROXIMATION

Mathematical Institute, University of Oxford

Generalized Nyström

$$A \approx A\Omega_1(\Omega_2^* A\Omega_1)^\dagger \Omega_2^* A =: A_{GN,\Omega_1,\Omega_2}$$

(Clarkson, Woodruff, 2009)
(Nakatsukasa, 2020)
(Woolfe, Liberty, Rokhlin, Tygert, 2008)

1. Choose $\Omega_1 \in \mathbb{R}^{n\times r}, \Omega_2 \in \mathbb{R}^{m\times (r+\ell)}$
2. Two-side Sketch: $X = A\Omega_1$ and $Y = \Omega_2^* A$
3. $[\mathrm{Q},\mathrm{R}] = \mathrm{qr}(Y\Omega_1, 0)$
4. $A_{GN,\Omega_1,\Omega_2} = (XR^{-1})(Q^* Y)$

- $N_{2r+\ell} + \mathcal{O}(r^3 + (m+n)r^2)$
- Single-pass
- 2 multiplications by A

GENERALIZED NYSTRÖM APPROXIMATION

Generalized Nyström

$$A \approx A\Omega_1(\Omega_2^* A\Omega_1)^\dagger \Omega_2^* A =: A_{GN,\Omega_1,\Omega_2}$$

(Clarkson, Woodruff, 2009)
(Nakatsukasa, 2020)
(Woolfe, Liberty, Rokhlin, Tygert, 2008)

1. Choose $\Omega_1 \in \mathbb{R}^{n\times r}, \Omega_2 \in \mathbb{R}^{m\times (r+\ell)}$
2. Two-side Sketch: $X = A\Omega_1$ and $Y = \Omega_2^* A$
3. $[Q,R] = \mathrm{qr}(Y\Omega_1,0)$
4. $A_{GN,\Omega_1,\Omega_2} = (XR^{-1})(Q^*Y)$

- $N_{2r+\ell} + \mathcal{O}(r^3 + (m+n)r^2)$
- Single-pass
- 2 multiplications by A

Accuracy

$\hat{r} \leq r - 2$

$$\mathbb{E}\|A - A_{GN,\Omega_1,\Omega_2}\|_F \leq \sqrt{1 + \frac{r+\ell}{\ell - 1}}\sqrt{1 + \frac{r}{r - \hat{r} - 1}}\|A - A_{best,\hat{r}}\|_F$$

(Tropp et al., 2017),(Nakatsukasa, 2020)

GENERALIZED NYSTRÖM APPROXIMATION

Generalized Nyström

$$A \approx A\Omega_1(\Omega_2^* A\Omega_1)^\dagger \Omega_2^* A =: A_{GN,\Omega_1,\Omega_2}$$

(Clarkson, Woodruff, 2009)
(Nakatsukasa, 2020)
(Woolfe, Liberty, Rokhlin, Tygert, 2008)

1. Choose $\Omega_1 \in \mathbb{R}^{n\times r}, \Omega_2 \in \mathbb{R}^{m\times(r+\ell)}$
2. Two-side Sketch: $X = A\Omega_1$ and $Y = \Omega_2^* A$
3. $[\mathrm{Q,R}] = \mathrm{qr}(Y\Omega_1, 0)$
4. $A_{GN,\Omega_1,\Omega_2} = (XR^{-1})(Q^*Y)$

- $N_{2r+\ell} + \mathcal{O}(r^3 + (m+n)r^2)$
- Single-pass
- 2 multiplications by A

Accuracy

$\hat{r} \leq r - 2$

$$\mathbb{E}\|A - A_{GN,\Omega_1,\Omega_2}\|_F \leq \sqrt{1 + \frac{r+\ell}{\ell-1}}\sqrt{1 + \frac{r}{r-\hat{r}-1}}\|A - A_{best,\hat{r}}\|_F$$

(Tropp et al., 2017),(Nakatsukasa, 2020)

Stability

$$(A\Omega_1)(\Omega_2^* A\Omega_1)_\epsilon^\dagger \Omega_2^* A$$

(Nakatsukasa, 2020)

ANALYSIS AND COMPARISON

4

GN APPROXIMATION AND EXTRACTING SINGULAR VALUES

Generalized Nyström

Given approximations $\tilde{U}$ and $\tilde{V}$ to the leading singular subspaces,

$$\sigma_i(A) \approx \sigma_i\left(A\tilde{V}(\tilde{U}^*A\tilde{V})^\dagger \tilde{U}^*A\right) =: \sigma_i^{GN}$$

$$\sigma_i\left(\boxed{A\tilde{V}} \quad \boxed{\tilde{U}^*A\tilde{V}}^{\dagger} \quad \boxed{\tilde{U}^*A} \right)$$

$$N_{2r+\ell}$$

GN APPROXIMATION AND EXTRACTING SINGULAR VALUES

Mathematical Institute

Generalized Nyström

Given approximations $\tilde{U}$ and $\tilde{V}$ to the leading singular subspaces,

$$\sigma_i(A) \approx \sigma_i \left(A\tilde{V}(\tilde{U}^* A\tilde{V})^\dagger \tilde{U}^* A \right) =: \sigma_i^{GN}$$

$$\sigma_i\Big(\boxed{Q_L} \; \boxed{R_L} \; \boxed{\tilde{U}^* A \tilde{V}}^{\dagger} \; \boxed{R_R^*} \; \boxed{Q_R^*} \Big)$$

$$N_{2r+\ell} + \mathcal{O}((m+n)r^2)$$

GN APPROXIMATION AND EXTRACTING SINGULAR VALUES

Generalized Nyström

Given approximations $\tilde{U}$ and $\tilde{V}$ to the leading singular subspaces,

$$\sigma_i(A) \approx \sigma_i\left(A\tilde{V}(\tilde{U}^*A\tilde{V})^\dagger \tilde{U}^*A\right) =: \sigma_i^{GN}$$

$$\sigma_i\left(\boxed{R_L} \; \boxed{\tilde{U}^*A\tilde{V}}^{\dagger} \; \boxed{R_R^*} \right)$$

$$N_{2r+\ell} + \mathcal{O}((m+n)r^2)$$

GN APPROXIMATION AND EXTRACTING SINGULAR VALUES

Generalized Nyström

Given approximations $\tilde{U}$ and $\tilde{V}$ to the leading singular subspaces,

$$\sigma_i(A) \approx \sigma_i\left(A\tilde{V}(\tilde{U}^* A\tilde{V})^\dagger \tilde{U}^* A\right) =: \sigma_i^{GN}$$

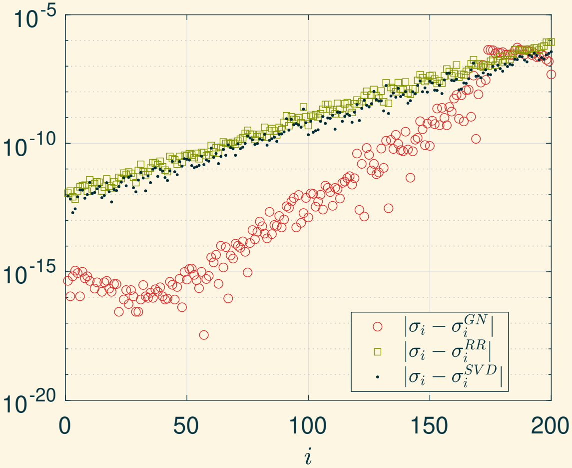
$$\sigma_i\left(\boxed{R_L} \quad \boxed{R_p^\dagger} \quad \boxed{Q_p^*} \quad \boxed{R_R^*} \right)$$

$$N_{2r+\ell} + \mathcal{O}((m+n)r^2)$$

MOTIVATIONAL COMPARISON

Single-pass methods

- $\sigma_i^{SVD} = \sigma_i(A\tilde{V})$
- $\sigma_i^{RR} = \sigma_i(\tilde{U}^* A\tilde{V})$
- $\sigma_i^{GN} = \sigma_i\left(A\tilde{V}(\tilde{U}^* A\tilde{V})^\dagger \tilde{U}^* A\right)$



GN AND MATRIX PERTURBATION THEORY

GN and Orthogonal Transformations

Consider T_1 and T_2 orthogonal matrices, then

$$T_1^*(M_{GN,\tilde{V},\tilde{U}})T_2 = (T_1^* M T_2)_{GN,T_2^*\tilde{V},T_1^*\tilde{U}}$$

For any orthonormal $\tilde{V}$ and $\tilde{U}$, we can:

1. Define $Q_1 = \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix}$ $\quad Q_2 = \begin{bmatrix} \tilde{V} & \tilde{V}_\perp \end{bmatrix}$;
2. Consider the transformed matrix: $Q_1^* A Q_2$;
3. Consider the transformed GN approximation:

$$Q_1^* A_{GN,\tilde{V},\tilde{U}} Q_2 = (Q_1^* A Q_2)_{GN,Q_2^*\tilde{V},Q_1^*\tilde{U}} = (Q_1^* A Q_2)_{GN,\begin{bmatrix} I_r \\ 0 \end{bmatrix},\begin{bmatrix} I_{r+\ell} \\ 0 \end{bmatrix}}.$$

GN AND MATRIX PERTURBATION THEORY

Mathematical Institute

GN and Orthogonal Transformations

Consider T_1 and T_2 orthogonal matrices, then

$$T_1^*(M_{GN,\tilde{V},\tilde{U}})T_2 = (T_1^* M T_2)_{GN,T_2^*\tilde{V},T_1^*\tilde{U}}$$

For any orthonormal $\tilde{V}$ and $\tilde{U}$, we can:

1. Define $Q_1 = \begin{bmatrix}\tilde{U} & \tilde{U}_\perp\end{bmatrix}$ $\quad Q_2 = \begin{bmatrix}\tilde{V} & \tilde{V}_\perp\end{bmatrix}$;
2. Consider the transformed matrix: $Q_1^* A Q_2$;
3. Consider the transformed GN approximation:

$$Q_1^* A_{GN,\tilde{V},\tilde{U}} Q_2 = (Q_1^* A Q_2)_{GN,Q_2^*\tilde{V},Q_1^*\tilde{U}} = (Q_1^* A Q_2)_{GN,\begin{bmatrix} I_r \\ 0\end{bmatrix},\begin{bmatrix} I_{r+\ell} \\ 0\end{bmatrix}}.$$

$$\rightarrow \quad |\sigma_i(A) - \sigma_i(A_{GN,\tilde{V},\tilde{U}})| = |\sigma_i(Q_1^* A Q_2) - \sigma_i((Q_1^* A Q_2)_{GN,\begin{bmatrix} I_r \\ 0\end{bmatrix},\begin{bmatrix} I_{r+\ell} \\ 0\end{bmatrix}})|$$

GN AND MATRIX PERTURBATION THEORY › *Express A_{GN} as a perturbation of the original matrix A*

$$\tilde{V} := \begin{array}{r c} & r \\ \begin{array}{r} r \\ n-r \end{array} & \left[\begin{array}{c} I_r \\ \hline 0 \end{array}\right] \end{array}, \quad \tilde{U} := \begin{array}{r c} & r+\ell \\ \begin{array}{r} r+\ell \\ m-(r+\ell) \end{array} & \left[\begin{array}{c} I_{r+\ell} \\ \hline 0 \end{array}\right] \end{array}, \quad A := \begin{array}{r c} & \begin{array}{cc} r & n-r \end{array} \\ \begin{array}{r} r+\ell \\ m-(r+\ell) \end{array} & \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] \end{array}$$

(Tropp, Webber, 2023)

$$A_{GN,\tilde{V},\tilde{U}} = A\tilde{V}\,(\tilde{U}^* A\tilde{V})^{\dagger}\,\tilde{U}^* A$$

GN AND MATRIX PERTURBATION THEORY › *Express A_{GN} as a perturbation of the original matrix A*

$$\tilde{V} := \begin{array}{c c} & r \\ \begin{array}{r} r \\ n-r \end{array} & \begin{bmatrix} I_r \\ - \\ 0 \end{bmatrix} \end{array}, \quad \tilde{U} := \begin{array}{c c} & r+\ell \\ \begin{array}{r} r+\ell \\ m-(r+\ell) \end{array} & \begin{bmatrix} I_{r+\ell} \\ - \\ 0 \end{bmatrix} \end{array}, \quad A := \begin{array}{c c} & \begin{array}{cc} r & n-r \end{array} \\ \begin{array}{r} r+\ell \\ m-(r+\ell) \end{array} & \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] \end{array}$$

$$A_{GN,\tilde{V},\tilde{U}} = \begin{bmatrix} A_{11} \\ - \\ A_{21} \end{bmatrix} (\tilde{U}^* A \tilde{V})^{\dagger} \, \tilde{U}^* A$$

GN AND MATRIX PERTURBATION THEORY > *Express A_{GN} as a perturbation of the original matrix A*

Mathematical Institute

$$\tilde{V} := \begin{array}{c c} & r \\ \begin{array}{r} r \\ n-r \end{array} & \left[\begin{array}{c} I_r \\ \hline 0 \end{array}\right] \end{array}, \quad \tilde{U} := \begin{array}{c c} & r+\ell \\ \begin{array}{r} r+\ell \\ m-(r+\ell) \end{array} & \left[\begin{array}{c} I_{r+\ell} \\ \hline 0 \end{array}\right] \end{array}, \quad A := \begin{array}{c c} & \begin{array}{cc} r & n-r \end{array} \\ \begin{array}{r} r+\ell \\ m-(r+\ell) \end{array} & \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] \end{array}$$

$$A_{GN,\tilde{V},\tilde{U}} = \left[\begin{array}{c} A_{11} \\ \hline A_{21} \end{array}\right] (\tilde{U}^* A \tilde{V})^\dagger \left[\begin{array}{c|c} A_{11} & A_{12} \end{array}\right]$$

GN AND MATRIX PERTURBATION THEORY > *Express A_{GN} as a perturbation of the original matrix A*

$$\tilde{V} := \begin{array}{c} \\ r \\ n-r \end{array}\!\!\begin{array}{c} r \\ \left[\begin{array}{c} I_r \\ \hline 0 \end{array}\right] \end{array}, \quad \tilde{U} := \begin{array}{c} \\ r+\ell \\ m-(r+\ell) \end{array}\!\!\begin{array}{c} r+\ell \\ \left[\begin{array}{c} I_{r+\ell} \\ \hline 0 \end{array}\right] \end{array}, \quad A := \begin{array}{c} \\ r+\ell \\ m-(r+\ell) \end{array}\!\!\begin{array}{c} \begin{array}{cc} r & n-r \end{array} \\ \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] \end{array}$$

$$A_{GN,\tilde{V},\tilde{U}} = \left[\begin{array}{c} A_{11} \\ \hline A_{21} \end{array}\right] (A_{11})^\dagger \left[\begin{array}{c|c} A_{11} & A_{12} \end{array}\right]$$

GN AND MATRIX PERTURBATION THEORY › *Express A_{GN} as a perturbation of the original matrix A*

$$\tilde{V} := \begin{array}{c c} & r \\ \begin{array}{r} r \\ n-r \end{array} & \left[\begin{array}{c} I_r \\ \hline 0 \end{array}\right] \end{array}, \quad \tilde{U} := \begin{array}{c c} & r+\ell \\ \begin{array}{r} r+\ell \\ m-(r+\ell) \end{array} & \left[\begin{array}{c} I_{r+\ell} \\ \hline 0 \end{array}\right] \end{array}, \quad A := \begin{array}{c c} & \begin{array}{cc} r & n-r \end{array} \\ \begin{array}{r} r+\ell \\ m-(r+\ell) \end{array} & \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] \end{array}$$

$$\boxed{MM^\dagger M = M}$$

$$A_{GN,\tilde{V},\tilde{U}} = \left[\begin{array}{c} A_{11} \\ \hline A_{21} \end{array}\right] (A_{11})^\dagger \left[\begin{array}{c|c} A_{11} & A_{12} \end{array}\right] = \left[\begin{array}{c|c} A_{11}A_{11}^\dagger A_{11} & A_{11}A_{11}^\dagger A_{12} \\ \hline A_{21}A_{11}^\dagger A_{11} & A_{21}A_{11}^\dagger A_{12} \end{array}\right]$$

GN AND MATRIX PERTURBATION THEORY › *Express A_{GN} as a perturbation of the original matrix A*

$$\tilde{V} := \begin{array}{c} \\ r \\ n-r \end{array}\!\!\begin{array}{c} r \\ \left[\begin{array}{c} I_r \\ \hline 0 \end{array}\right] \end{array}, \quad \tilde{U} := \begin{array}{c} \\ r+\ell \\ m-(r+\ell) \end{array}\!\!\begin{array}{c} r+\ell \\ \left[\begin{array}{c} I_{r+\ell} \\ \hline 0 \end{array}\right] \end{array}, \quad A := \begin{array}{c} \\ r+\ell \\ m-(r+\ell) \end{array}\!\!\begin{array}{cc} r & n-r \\ \multicolumn{2}{c}{\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]} \end{array}$$

M has linearly independent columns
$\implies M^{\dagger}M = M^{-1}M = M$

$$A_{GN,\tilde{V},\tilde{U}} = \left[\begin{array}{c} A_{11} \\ \hline A_{21} \end{array}\right](A_{11})^{\dagger}\left[\begin{array}{c|c} A_{11} & A_{12} \end{array}\right] = \left[\begin{array}{c|c} \overbrace{A_{11}A_{11}^{\dagger}A_{11}}^{=A_{11}} & A_{11}A_{11}^{\dagger}A_{12} \\ \hline A_{21}A_{11}^{\dagger}A_{11} & A_{21}A_{11}^{\dagger}A_{12} \end{array}\right]$$

GN AND MATRIX PERTURBATION THEORY › *Express A_{GN} as a perturbation of the original matrix A*

$$\tilde V := \begin{array}{c} \\ r \\ n-r \end{array}\!\!\begin{array}{c} r \\ \left[\begin{array}{c} I_r \\ \hline 0 \end{array}\right] \end{array}, \quad \tilde U := \begin{array}{c} \\ r+\ell \\ m-(r+\ell) \end{array}\!\!\begin{array}{c} r+\ell \\ \left[\begin{array}{c} I_{r+\ell} \\ \hline 0 \end{array}\right] \end{array}, \quad A := \begin{array}{c} \\ r+\ell \\ m-(r+\ell) \end{array}\!\!\begin{array}{cc} r & n-r \\ \multicolumn{2}{c}{\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]} \end{array}$$

$$A_{GN,\tilde V,\tilde U} = \left[\begin{array}{c} A_{11} \\ \hline A_{21} \end{array}\right] (A_{11})^\dagger \left[\begin{array}{c|c} A_{11} & A_{12} \end{array}\right] = \left[\begin{array}{c|c} A_{11} & A_{11}A_{11}^\dagger A_{12} \\ \hline \underbrace{A_{21}A_{11}^\dagger A_{11}}_{=A_{21}} & A_{21}A_{11}^\dagger A_{12} \end{array}\right]$$

GN AND MATRIX PERTURBATION THEORY › *Express A_{GN} as a perturbation of the original matrix A*

Mathematical Institute

$$\tilde{V} := \begin{array}{c} \\ r \\ n-r \end{array} \!\!\begin{array}{c} r \\ \left[\begin{array}{c} I_r \\ \hline 0 \end{array}\right] \end{array}, \quad \tilde{U} := \begin{array}{c} \\ r+\ell \\ m-(r+\ell) \end{array} \!\!\begin{array}{c} r+\ell \\ \left[\begin{array}{c} I_{r+\ell} \\ \hline 0 \end{array}\right] \end{array}, \quad A := \begin{array}{c} \\ r+\ell \\ m-(r+\ell) \end{array} \!\!\begin{array}{cc} r & n-r \\ \multicolumn{2}{c}{\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]} \end{array}$$

$$A_{GN,\tilde{V},\tilde{U}} = A - \left[\begin{array}{c|c} 0 & A_{12} - A_{11}A_{11}^{\dagger}A_{12} \\ \hline 0 & A_{22} - A_{21}A_{11}^{\dagger}A_{12} \end{array}\right] =: A - E_{GN}$$

GN AND MATRIX PERTURBATION THEORY › *Express A_{GN} as a perturbation of the original matrix A*

Mathematical Institute

$$\tilde V := \begin{array}{c c} & r \\ r & \\ n-r & \end{array}\!\!\begin{bmatrix} I_r \\ \hline 0 \end{bmatrix}, \quad \tilde U := \begin{array}{c c} & r \\ r & \\ m-r & \end{array}\!\!\begin{bmatrix} I_r \\ \hline 0 \end{bmatrix}, \quad A := \begin{array}{c c c} & r & n-r \\ r & & \\ m-r & & \end{array}\!\!\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]$$

No-oversample ($\ell = 0$)
→ $A_{12} - A_{11}A_{11}^{\dagger}A_{12} = 0$, but change of block sizes!

$$A_{GN,\tilde V,\tilde U} = A - \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & A_{22} - A_{21}A_{11}^{\dagger}A_{12} \end{array}\right] =: A - E_{GN}$$

GN AND MATRIX PERTURBATION THEORY › *Weyl's bound*

Oxford Mathematical Institute

Weyl's Theorem

For any matrix M we have that

$$|\sigma_i(M) - \sigma_i(M+E)| \leq \|E\|_2$$

Cor. 7.3.5 (Horn, Johnson, 2012)
Cor. I.4.31 (Stewart, 1998)

GN AND MATRIX PERTURBATION THEORY › *Weyl's bound*

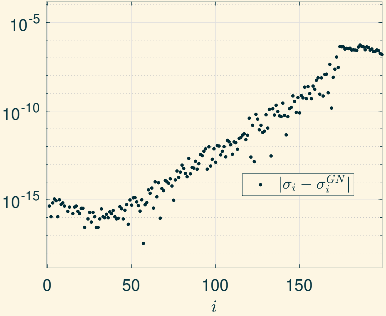
Weyl's Theorem

For any matrix M we have that

$$|\sigma_i(M) - \sigma_i(M+E)| \leq \|E\|_2$$

Cor. 7.3.5 (Horn, Johnson, 2012)
Cor. I.4.31 (Stewart, 1998)

$$|\sigma_i(A) - \sigma_i(A_{GN,\tilde{V},\tilde{U}})|$$



GN AND MATRIX PERTURBATION THEORY › *Weyl's bound*

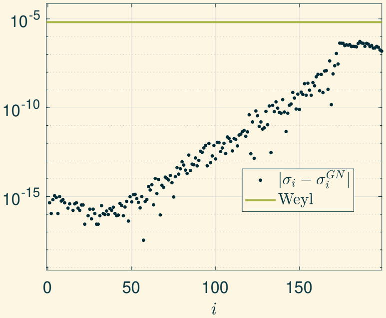
Weyl's Theorem

For any matrix M we have that

$$|\sigma_i(M) - \sigma_i(M+E)| \le \|E\|_2$$

Cor. 7.3.5 (Horn, Johnson, 2012)
Cor. I.4.31 (Stewart, 1998)

$$|\sigma_i(A) - \sigma_i(A_{GN,\tilde{V},\tilde{U}})| \le \|E_{GN}\|_2$$



RESULT ON SYMMETRIC MATRICES

Consider the $n \times n$ symmetric matrices

$$H := \begin{bmatrix} H_{11} & H_{21}^* \\ H_{21} & H_{22} \end{bmatrix}, \quad \hat{H} := H + \begin{bmatrix} E_{11} & E_{21}^* \\ E_{21} & E_{22} \end{bmatrix} =: H + E.$$

Theorem 3.2 (Nakatsukasa, 2012)

RESULT ON SYMMETRIC MATRICES

Consider the $n \times n$ symmetric matrices

$$H := \begin{bmatrix} H_{11} & H_{21}^* \\ H_{21} & H_{22} \end{bmatrix}, \quad \hat{H} := H + \begin{bmatrix} E_{11} & E_{21}^* \\ E_{21} & E_{22} \end{bmatrix} =: H + E.$$

Theorem 3.2 (Nakatsukasa, 2012)

Define

$$\tau_i = \left(\frac{\|H_{21}\|_2 + \|E_{21}\|_2}{\min_j |\lambda_i(H) - \lambda_j(H_{22})| - 2\|E\|_2} \right).$$

Then, for each i, if $\tau_i > 0$, then

$$|\lambda_i(H) - \lambda_i(\hat{H})| \leq \|E_{11}\|_2 + 2\|E_{21}\|_2 \tau_i + \|E_{22}\|_2 \tau_i^2,$$

RESULT ON SYMMETRIC MATRICES

Consider the $n \times n$ symmetric matrices

$$H := \begin{bmatrix} H_{11} & H_{21}^* \\ H_{21} & H_{22} \end{bmatrix}, \quad \hat{H} := H + \begin{bmatrix} E_{11} & E_{21}^* \\ E_{21} & E_{22} \end{bmatrix} =: H + E.$$

Theorem 3.2 (Nakatsukasa, 2012)

Define

$$\tau_i = \left(\frac{\|H_{21}\|_2 + \|E_{21}\|_2}{\min_j |\lambda_i(H) - \lambda_j(H_{22})| - 2\|E\|_2} \right).$$

Then, for each i, if $\tau_i > 0$, then

$$|\lambda_i(H) - \lambda_i(\hat{H})| \leq \|E_{11}\|_2 + 2\|E_{21}\|_2 \tau_i + \|E_{22}\|_2 \tau_i^2,$$

- $\tau_i < 1$ necessary to be better than Weyl
- If $\|E_{11}\|_2 \ll \|E\|_2$ and λ_i is far from the spectrum of H_{22} then $\tau_i \ll 1$
- If $E_{11} = E_{21} = 0$ and H_{21} is small, then λ_i is particularly insensitive to the perturbation E_{22} $\rightarrow$ bound proportional to $\|E_{22}\|_2 \|H_{21}\|_2^2$

FROM THE SYMMETRIC TO THE GENERAL RESULT

General case

$\downarrow$

Transform to symmetric

$\downarrow$

Obtain necessary structure

$\downarrow$

Apply symmetric Result

$\downarrow$

Transform back

$\downarrow$

General Result

Generalize (Nakatsukasa, 2012) to the 2×2 block matrix:

$$G := \begin{bmatrix} G_1 & B \\ C & G_2 \end{bmatrix},$$

and its perturbation:

$$\hat{G} := G + \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} =: G + F.$$

Strategy: Use a technique in (Li, Li, 2005)

FROM THE SYMMETRIC TO THE GENERAL RESULT

General case

↓

Transform to symmetric

↓

Obtain necessary structure

↓

Apply symmetric Result

↓

Transform back

↓

General Result

Thm. 7.3.3 (Horn, Johnson, 2012)
Thm. I.4.2 (Stewart, Sun, 1990)

Jordan-Wielandt (JW) Theorem

Let $\{\sigma_i(M)\}_{i=1}^n$ be the singular values of a matrix $M \in \mathbb{C}^{m\times n}$, with $m \geq n$. Then, the symmetric matrix

$$\begin{bmatrix} 0 & M \\ M^* & 0 \end{bmatrix} \tag{1}$$

has eigenvalues $\pm\sigma_1(M), \dots, \pm\sigma_n(M)$ and $m-n$ zeros eigenvalues.

FROM THE SYMMETRIC TO THE GENERAL RESULT

General case

↓

Transform to symmetric

↓

Obtain necessary structure

↓

Apply symmetric Result

↓

Transform back

↓

General Result

Thm. 7.3.3 (Horn, Johnson, 2012)
Thm. I.4.2 (Stewart, Sun, 1990)

Jordan-Wielandt (JW) Theorem

Let $\{\sigma_i(M)\}_{i=1}^n$ be the singular values of a matrix $M \in \mathbb{C}^{m\times n}$, with $m \geq n$. Then, the symmetric matrix

$$\begin{bmatrix} 0 & M \\ M^* & 0 \end{bmatrix} \tag{1}$$

has eigenvalues $\pm\sigma_1(M), \ldots, \pm\sigma_n(M)$ and $m-n$ zeros eigenvalues.

$$G \to G_{JW} := \left[\begin{array}{c|c} 0 & G \\ \hline G^* & 0 \end{array}\right] = \left[\begin{array}{cc|cc} 0 & 0 & G_1 & B \\ 0 & 0 & C & G_2 \\ \hline G_1^* & C^* & 0 & 0 \\ B^* & G_2^* & 0 & 0 \end{array}\right]$$

FROM THE SYMMETRIC TO THE GENERAL RESULT

General case
↓
Transform to symmetric
↓
Obtain necessary structure
↓
Apply symmetric Result
↓
Transform back
↓
General Result

Obtain a matrix similar to G_{JW} suitable for (Nakatsukasa, 2012) and with blocks reasonably related to the blocks of G

$$\left[\begin{array}{cc|cc} 0 & 0 & G_1 & B \\ 0 & 0 & C & G_2 \\ \hline G_1^* & C^* & 0 & 0 \\ B^* & G_2^* & 0 & 0 \end{array}\right]$$

FROM THE SYMMETRIC TO THE GENERAL RESULT

General case

↓

Transform to symmetric

↓

Obtain necessary structure

↓

Apply symmetric Result

↓

Transform back

↓

General Result

Obtain a matrix similar to G_{JW} suitable for (Nakatsukasa, 2012) and with blocks reasonably related to the blocks of G

$$\left[\begin{array}{cc|cc} 0 & 0 & G_1 & B \\ 0 & 0 & C & G_2 \\ \hline G_1^* & C^* & 0 & 0 \\ B^* & G_2^* & 0 & 0 \end{array}\right]$$

FROM THE SYMMETRIC TO THE GENERAL RESULT

General case

↓

Transform to symmetric

↓

Obtain necessary structure

↓

Apply symmetric Result

↓

Transform back

↓

General Result

Obtain a matrix similar to G_{JW} suitable for (Nakatsukasa, 2012) and with blocks reasonably related to the blocks of G

$$\left[\begin{array}{cc|cc} 0 & 0 & G_1 & B \\ G_1^* & C^* & 0 & 0 \\ \hline 0 & 0 & C & G_2 \\ B^* & G_2^* & 0 & 0 \end{array}\right]$$

FROM THE SYMMETRIC TO THE GENERAL RESULT

General case

↓

Transform to symmetric

↓

Obtain necessary structure

↓

Apply symmetric Result

↓

Transform back

↓

General Result

Obtain a matrix similar to G_{JW} suitable for (Nakatsukasa, 2012) and with blocks reasonably related to the blocks of G

$$\left[\begin{array}{cc|cc} 0 & 0 & G_1 & B \\ G_1^* & C^* & 0 & 0 \\ \hline 0 & 0 & C & G_2 \\ B^* & G_2^* & 0 & 0 \end{array}\right]$$

FROM THE SYMMETRIC TO THE GENERAL RESULT

General case

↓

Transform to symmetric

↓

Obtain necessary structure

↓

Apply symmetric Result

↓

Transform back

↓

General Result

Obtain a matrix similar to G_{JW} suitable for (Nakatsukasa, 2012) and with blocks reasonably related to the blocks of G

$$\left[\begin{array}{cc|cc} 0 & G_1 & 0 & B \\ G_1^* & 0 & C^* & 0 \\ \hline 0 & C & 0 & G_2 \\ B^* & 0 & G_2^* & 0 \end{array}\right] =: G_p$$

Note: $\lambda_i(G_p) = \lambda_i(G_{JW}) \overset{JW}{=} \pm\sigma_i(G)$

FROM THE SYMMETRIC TO THE GENERAL RESULT

General case

↓

Transform to symmetric

↓

Obtain necessary structure

↓

Apply symmetric Result

↓

Transform back

↓

General Result

Obtain a matrix similar to G_{JW} suitable for (Nakatsukasa, 2012) and with blocks reasonably related to the blocks of G

$$G_p = \left[\begin{array}{cc|cc} 0 & G_1 & 0 & B \\ G_1^* & 0 & C^* & 0 \\ \hline 0 & C & 0 & G_2 \\ B^* & 0 & G_2^* & 0 \end{array}\right]$$

$$\hat{G}_p = G_p + \left[\begin{array}{cc|cc} 0 & F_{11} & 0 & F_{12} \\ F_{11}^* & 0 & F_{21}^* & 0 \\ \hline 0 & F_{21} & 0 & F_{22} \\ F_{12}^* & 0 & F_{22}^* & 0 \end{array}\right] =: G_p + F_p.$$

FROM THE SYMMETRIC TO THE GENERAL RESULT

General case

↓

Transform to symmetric

↓

Obtain necessary structure

↓

Apply symmetric Result

↓

Transform back

↓

General Result

Define

$$\tau_i = \left(\frac{\left\| \begin{bmatrix} 0 & C \\ B^* & 0 \end{bmatrix} \right\|_2 + \left\| \begin{bmatrix} 0 & F_{21} \\ F_{12}^* & 0 \end{bmatrix} \right\|_2}{\min_j |\lambda_i - \lambda_j \left(\begin{bmatrix} 0 & G_2 \\ G_2^* & 0 \end{bmatrix} \right) | - 2 \|F_p\|_2} \right).$$

Then, for each i, if $\tau_i > 0$:

$$|\lambda_i(G_p) - \lambda_i(\hat{G}_p)| \leq \left\| \begin{bmatrix} 0 & F_{11} \\ F_{11}^* & 0 \end{bmatrix} \right\|_2 + 2 \left\| \begin{bmatrix} 0 & F_{21} \\ F_{12}^* & 0 \end{bmatrix} \right\|_2 \tau_i + \left\| \begin{bmatrix} 0 & F_{22} \\ F_{22}^* & 0 \end{bmatrix} \right\|_2 \tau_i^2,$$

FROM THE SYMMETRIC TO THE GENERAL RESULT

General case

↓

Transform to symmetric

↓

Obtain necessary structure

↓

Apply symmetric Result

↓

Transform back

↓

General Result

- $\left\| \begin{bmatrix} 0 & M_1 \\ M_2 & 0 \end{bmatrix} \right\|_2 = \max\{\|M_1\|_2, \|M_2\|_2\};$
- Jordan-Wielandt theorem
$$\implies \quad |\lambda_i(G_p) - \lambda_i(\hat{G}_p)| = |\sigma_i(G) - \sigma_i(\hat{G})|,$$
for $i = 1, \dots, n$;
- By Jordan-Wielandt theorem and by construction of F_p:
$$\|F_p\|_2 = \|F\|_2$$

FROM THE SYMMETRIC TO THE GENERAL RESULT › *Generalization of (Nakatsukasa, 2012)*

General case

↓

Transform to symmetric

↓

Obtain necessary structure

↓

Apply symmetric Result

↓

Transform back

↓

General Result

Theorem 4.1 (L.,Al Daas, Nakatsukasa,2024)

Consider the matrices

$$G := \begin{bmatrix} G_1 & B \\ C & G_2 \end{bmatrix}, \quad \hat{G} := G + \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} =: G + F,$$

and define

$$\tau_i = \left(\frac{\max\{\|B\|_2, \|C\|_2\} + \max\{\|F_{12}\|_2, \|F_{21}\|_2\}}{\min_j |\sigma_i(G) - \sigma_j(G_2)| - 2\|F\|_2} \right).$$

Then, for each i, if $\tau_i > 0$, then

$$|\sigma_i(G) - \sigma_i(\hat{G})| \leq \|F_{11}\|_2 + 2\max\{\|F_{12}\|_2, \|F_{21}\|_2\}\tau_i + \|F_{22}\|_2\, \tau_i^2,$$

FROM THE SYMMETRIC TO THE GENERAL RESULT › *Generalization of (Nakatsukasa, 2012)*

General case

↓

Transform to symmetric

↓

Obtain necessary structure

↓

Apply symmetric Result

↓

Transform back

↓

General Result

Theorem 4.1 (L.,Al Daas, Nakatsukasa,2024)

Consider the matrices

$$G := \begin{bmatrix} G_1 & B \\ C & G_2 \end{bmatrix}, \quad \hat{G} := G + \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} =: G + F,$$

and define

$$\tau_i = \left(\frac{\max\{\|B\|_2, \|C\|_2\} + \max\{\|F_{12}\|_2, \|F_{21}\|_2\}}{\min_j |\sigma_i(G) - \sigma_j(G_2)| - 2\|F\|_2} \right).$$

Then, for each i, if $\tau_i > 0$, then

$$|\sigma_i(G) - \sigma_i(\hat{G})| \leq \|F_{11}\|_2 + 2\max\{\|F_{12}\|_2, \|F_{21}\|_2\}\tau_i + \|F_{22}\|_2\, \tau_i^2,$$

- **Generalization to Block Tridiagonal**: A Singular Value is insensitive to blockwise perturbation if it is well-separated from the spectrum of the diagonal blocks near the perturbed blocks.

BOUND ON GN APPROXIMATION ERROR › *Derivation*

- $A, \tilde{V}, \tilde{U} \quad \to \quad A_{GN} = A\tilde{V}(\tilde{U}^* A\tilde{V})^\dagger \tilde{U}^* A$
- Define
  $\bar{A} = [\tilde{U}\ \tilde{U}_\perp]^* A[\tilde{V}\ \tilde{V}_\perp], \quad \bar{A}_{GN} = \Big([\tilde{U}\ \tilde{U}_\perp]^* A[\tilde{V}\ \tilde{V}_\perp]\Big)_{GN, \begin{bmatrix} I_r \\ 0 \end{bmatrix}, \begin{bmatrix} I_r \\ 0 \end{bmatrix}}$

$$\implies \bar{A}_{GN} = \bar{A} - \begin{bmatrix} 0 & 0 \\ 0 & \bar{A}_{22} - \bar{A}_{21}\bar{A}_{11}^\dagger \bar{A}_{12} \end{bmatrix} =: \bar{A} - E_{GN}$$

BOUND ON GN APPROXIMATION ERROR › *Derivation*

- $A, \tilde{V}, \tilde{U} \quad \to \quad A_{GN} = A\tilde{V}(\tilde{U}^* A\tilde{V})^\dagger \tilde{U}^* A$
- Define
  $\bar{A} = [\tilde{U}\ \tilde{U}_\perp]^* A [\tilde{V}\ \tilde{V}_\perp], \quad \bar{A}_{GN} = \left([\tilde{U}\ \tilde{U}_\perp]^* A [\tilde{V}\ \tilde{V}_\perp]\right)_{GN, \begin{bmatrix} I_r \\ 0 \end{bmatrix}, \begin{bmatrix} I_r \\ 0 \end{bmatrix}}$

Corollary 5.1
(L., Al Daas, Nakatsukasa, 2024)

$$\implies \bar{A}_{GN} = \bar{A} - \begin{bmatrix} 0 & 0 \\ 0 & \bar{A}_{22} - \bar{A}_{21}\bar{A}_{11}^\dagger \bar{A}_{12} \end{bmatrix} =: \bar{A} - E_{GN}$$

Define

$$\tau_i = \frac{\max\{\|\bar{A}_{12}\|_2, \|\bar{A}_{21}\|_2\}}{\min_j |\sigma_i(\bar{A}) - \sigma_j\left(\bar{A}_{22}\right)| - 2\,\|E_{GN}\|_2}.$$

Then, for each i, if $\tau_i > 0$

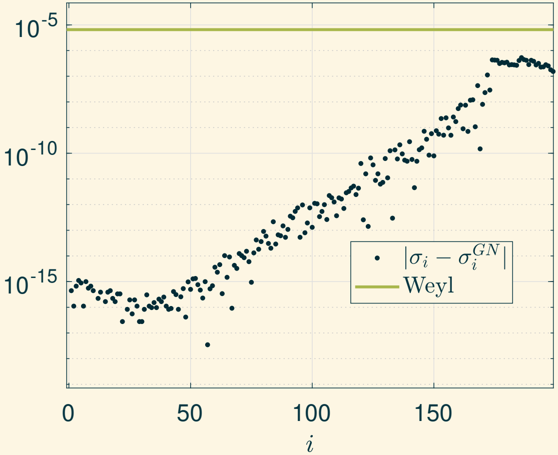
$$|\sigma_i(A) - \sigma_i(A_{GN})| = |\sigma_i(\bar{A}) - \sigma_i(\bar{A}_{GN})| \leq \left\|\bar{A}_{22} - \bar{A}_{21}\bar{A}_{11}^\dagger \bar{A}_{12}\right\|_2 \tau_i^2$$

- $\tau_i < 1$ necessary to be better than Weyl. If $\sigma_i(\bar{A})$ is far from the spectrum of $\bar{A}_{22}$ then $\tau_i \ll 1$

BOUND ON GN APPROXIMATION ERROR › *Numerical illustration*

- $\ell = 0$
- $A \in \mathbb{R}^{1000 \times 1000}$
- U_{ex}, V_{ex} Haar Matrices
- $\sigma_i(A)$ exponentially decaying
- $[\tilde{V}, \sim] = \mathtt{qr}(A^*\Omega, 0)$
- $[\tilde{U}, \sim] = \mathtt{qr}(A\Omega, 0)$
- $\tilde{V} \in \mathbb{R}^{1000 \times 200}$
- $\tilde{U} \in \mathbb{R}^{1000 \times 200}$
- Compute pseudoinverses by QR factorization

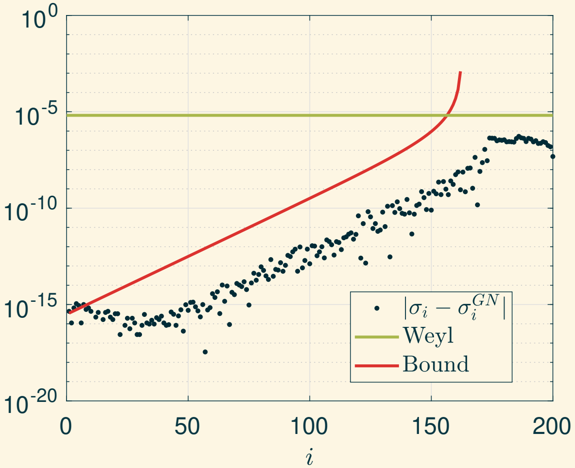
$$\sigma_i(A_{GN,\tilde{V},\tilde{U}}) = \sigma_i(A\tilde{V}(\tilde{U}^*A\tilde{V})^\dagger \tilde{U}^*A)$$



BOUND ON GN APPROXIMATION ERROR › *Numerical illustration*

- $\ell = 0$
- $A \in \mathbb{R}^{1000 \times 1000}$
- U_{ex}, V_{ex} Haar Matrices
- $\sigma_i(A)$ exponentially decaying
- $[\tilde{V}, \sim] = \mathtt{qr}(A^* \Omega, 0)$
- $[\tilde{U}, \sim] = \mathtt{qr}(A\Omega, 0)$
- $\tilde{V} \in \mathbb{R}^{1000 \times 200}$
- $\tilde{U} \in \mathbb{R}^{1000 \times 200}$
- Compute pseudoinverses by QR factorization

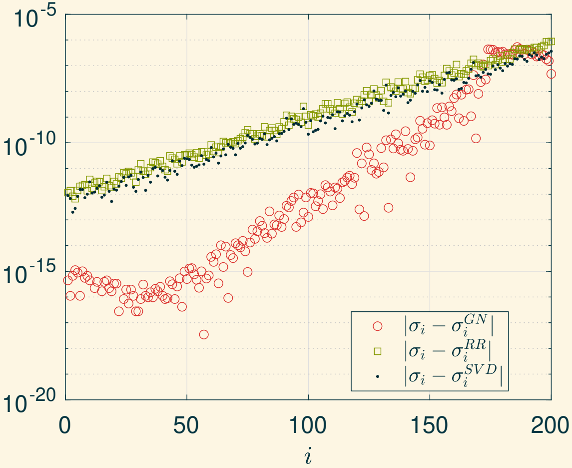
$$\sigma_i(A_{GN,\tilde{V},\tilde{U}}) = \sigma_i(A\tilde{V}(\tilde{U}^* A\tilde{V})^\dagger \tilde{U}^* A)$$



COMPARISON OF METHODS > *Idea*

Single-pass methods

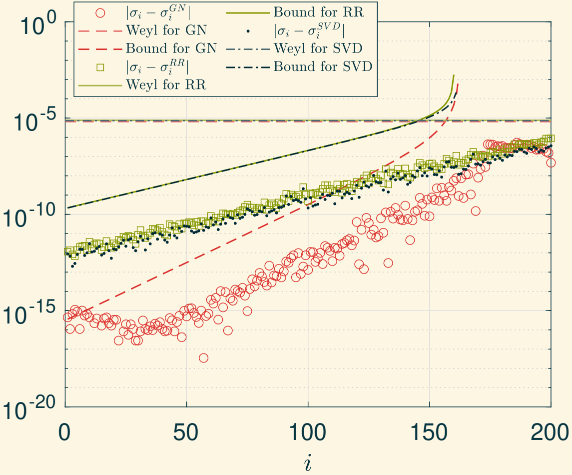
- $\sigma_i^{SVD} = \sigma_i(A\tilde{V})$
- $\sigma_i^{RR} = \sigma_i(\tilde{U}^* A\tilde{V})$
- $\sigma_i^{GN} = \sigma_i\left(A\tilde{V}(\tilde{U}^* A\tilde{V})^\dagger \tilde{U}^* A\right)$



COMPARISON OF METHODS > *Idea*

Single-pass methods

- $\sigma_i^{SVD} = \sigma_i(A\tilde{V})$
- $\sigma_i^{RR} = \sigma_i(\tilde{U}^* A\tilde{V})$
- $\sigma_i^{GN} = \sigma_i\left(A\tilde{V}(\tilde{U}^* A\tilde{V})^\dagger \tilde{U}^* A\right)$



THANK YOU!

😊

EXTRACTING ACCURATE SINGULAR VALUES FROM APPROXIMATE SUBSPACES

Lorenzo Lazzarino