A Scalable Randomized SVD with Multiple Sketches for Big Data Analysics

Weichung Wang
Institute of Applied Mathematical Sciences
National Taiwan University

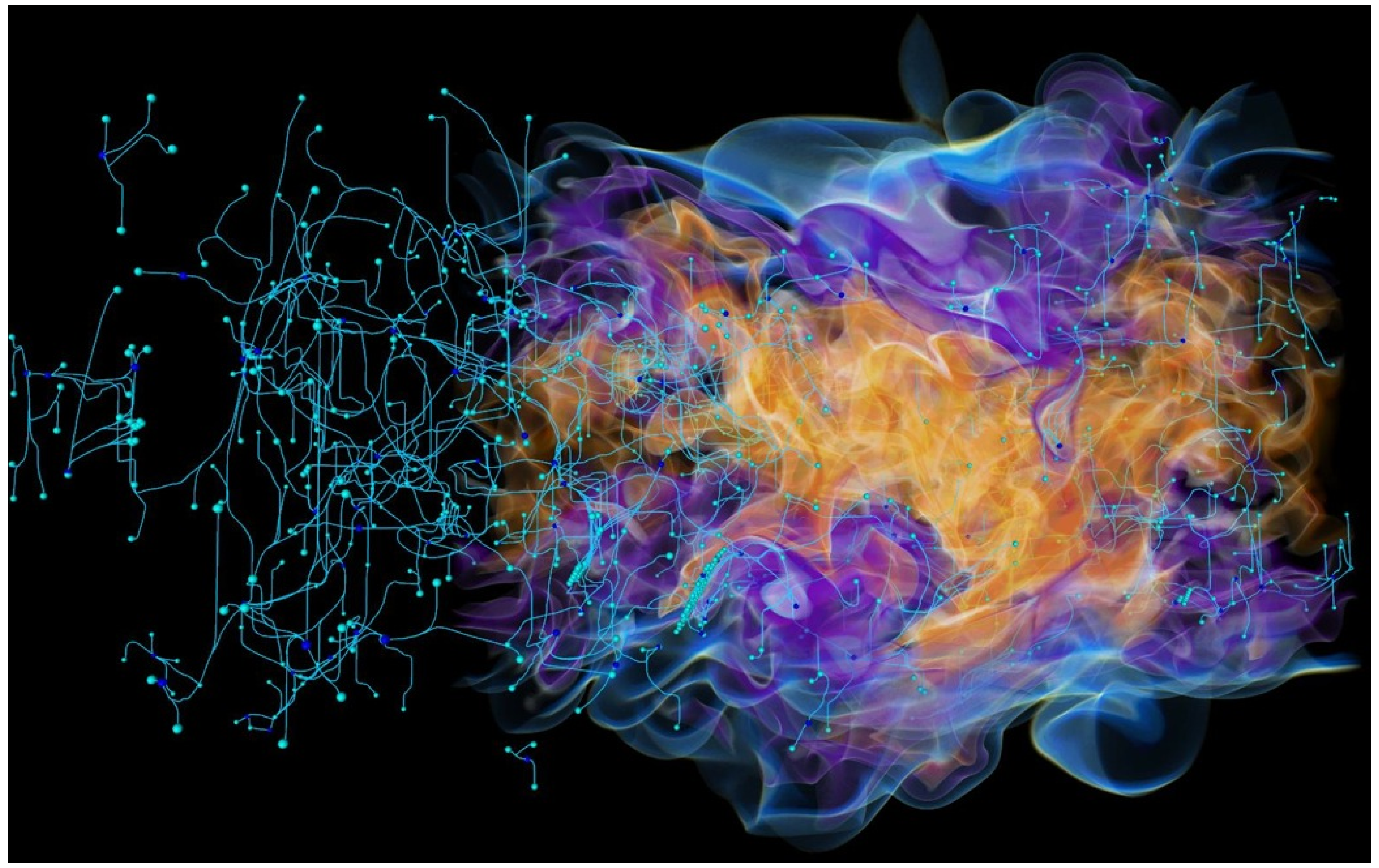
SIAM CSE, Spokane Convention Center, Spokane, Washington, USA, 2019/2/27

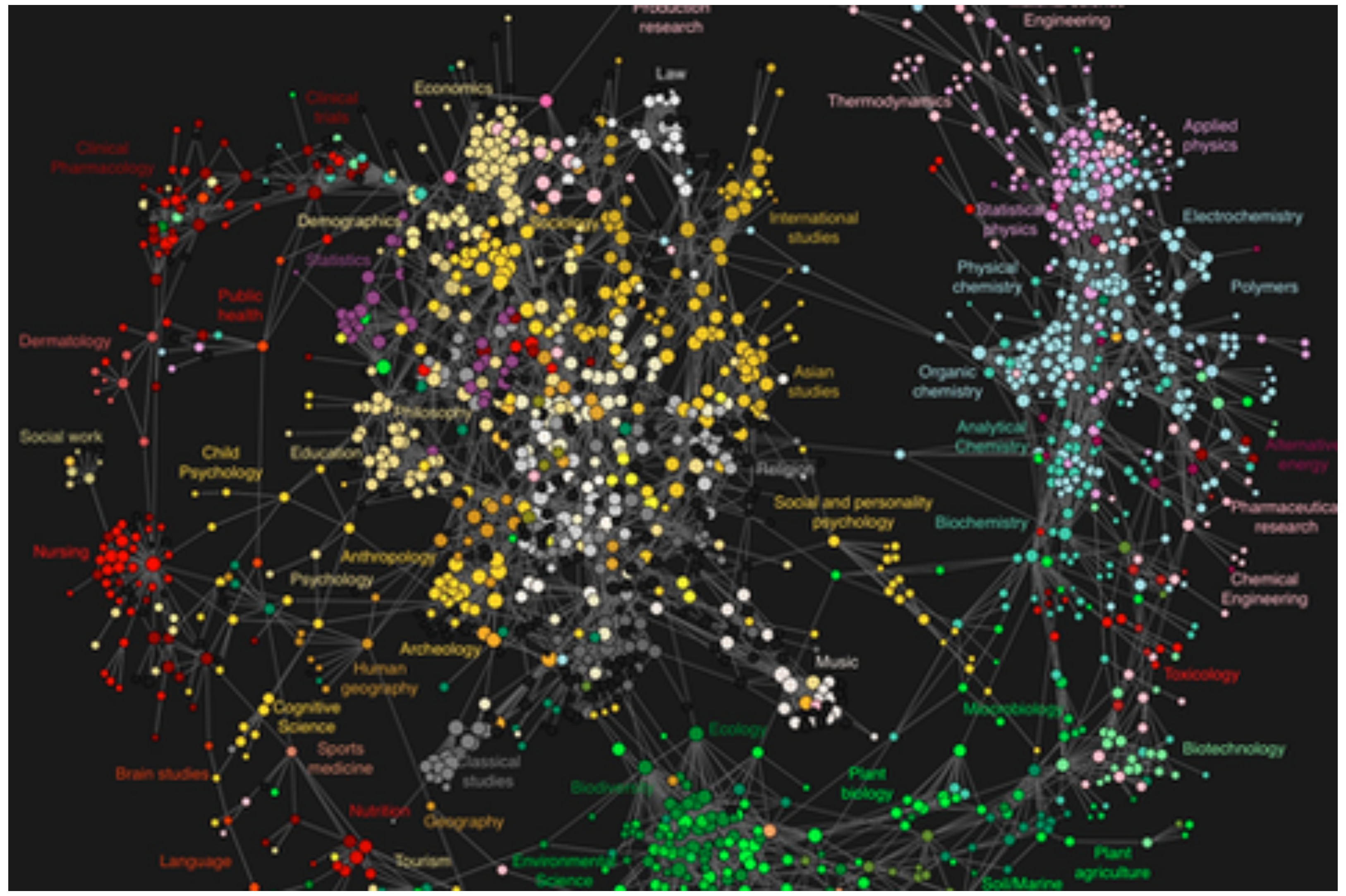
Collaborators

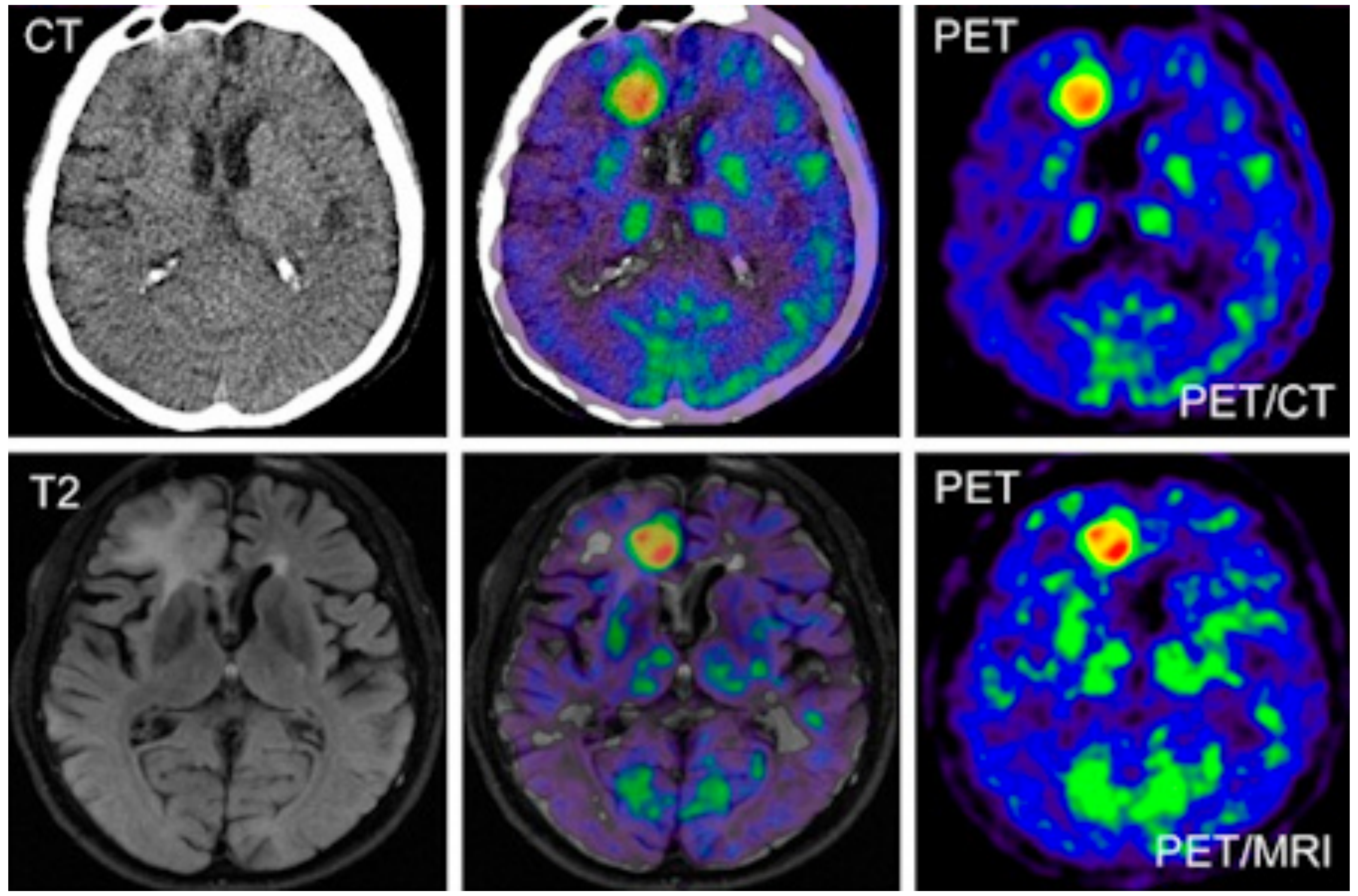
Su-Yun Huang, Ting-Li Chen
Institute of Statistical Science, Academia Sinica

Hung Chen, Dawei D. Chang, Mu Yang, Chen-Yao Lin
Institute of Applied Mathematical Sciences, National Taiwan University

Data Driven Discoveries







- Large scientific/internet/experimental datasets are censored, measured, collected, computed

- **Data driven** analysis
  - Information of m objects with n features
  - Correlations between all pairs of m objects
  - Connectivity between all pairs of m nodes in a network

iSVD

integrated singular value decomposition

Leading-k SVD

$$\boldsymbol{A} \approx \boldsymbol{U}_k \boldsymbol{\Sigma}_k \boldsymbol{V}_k^\top$$

$\boldsymbol{U}_k$ is an $m \times k$ orthonormal matrix that $k < m$, $\boldsymbol{\Sigma}_k$ is a $k \times k$ diagonal matrix, and $\boldsymbol{V}_k$ is an $n \times k$ orthonormal matrix. The columns of $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$ are the leading left singular vectors and right singular vectors of $\boldsymbol{A}$, respectively. The diagonal entries of $\boldsymbol{\Sigma}_k$ are the k largest singular values of $\boldsymbol{A}$.

- Well-studied

  numerical linear algebra, applied mathematics, statistics, computer sciences, data analytics, physical sciences, and engineering,…

- Many applications

  imaging, medicine, social networks, signal processing, machine learning, information compression, principal component analysis, finance,…

Random Sketches

- In many cases, one random sketch for a subspace is sufficient*

- Our idea
    - **multiple sketches** and then **integrate** the multiple subspaces
    - higher accuracy and higher stability
    - suitable for parallel computers and big matrices

*Halko, Nathan, Per-Gunnar Martinsson, and Joel A. Tropp. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions." *SIAM review* 53.2 (2011): 217-288.
*Rokhlin, Vladimir, Arthur Szlam, and Mark Tygert. "A randomized algorithm for principal component analysis." *SIAM Journal on Matrix Analysis and Applications* 31.3 (2009): 1100-1124.

Integrated SVD (iSVD)

- **Multiple** random sketching

$$\boldsymbol{Y}_{[i]} \leftarrow \boldsymbol{A\Omega}_{[i]}, \quad i = 1, \ldots, N$$

- Orthonormal basis of each sketched subspace

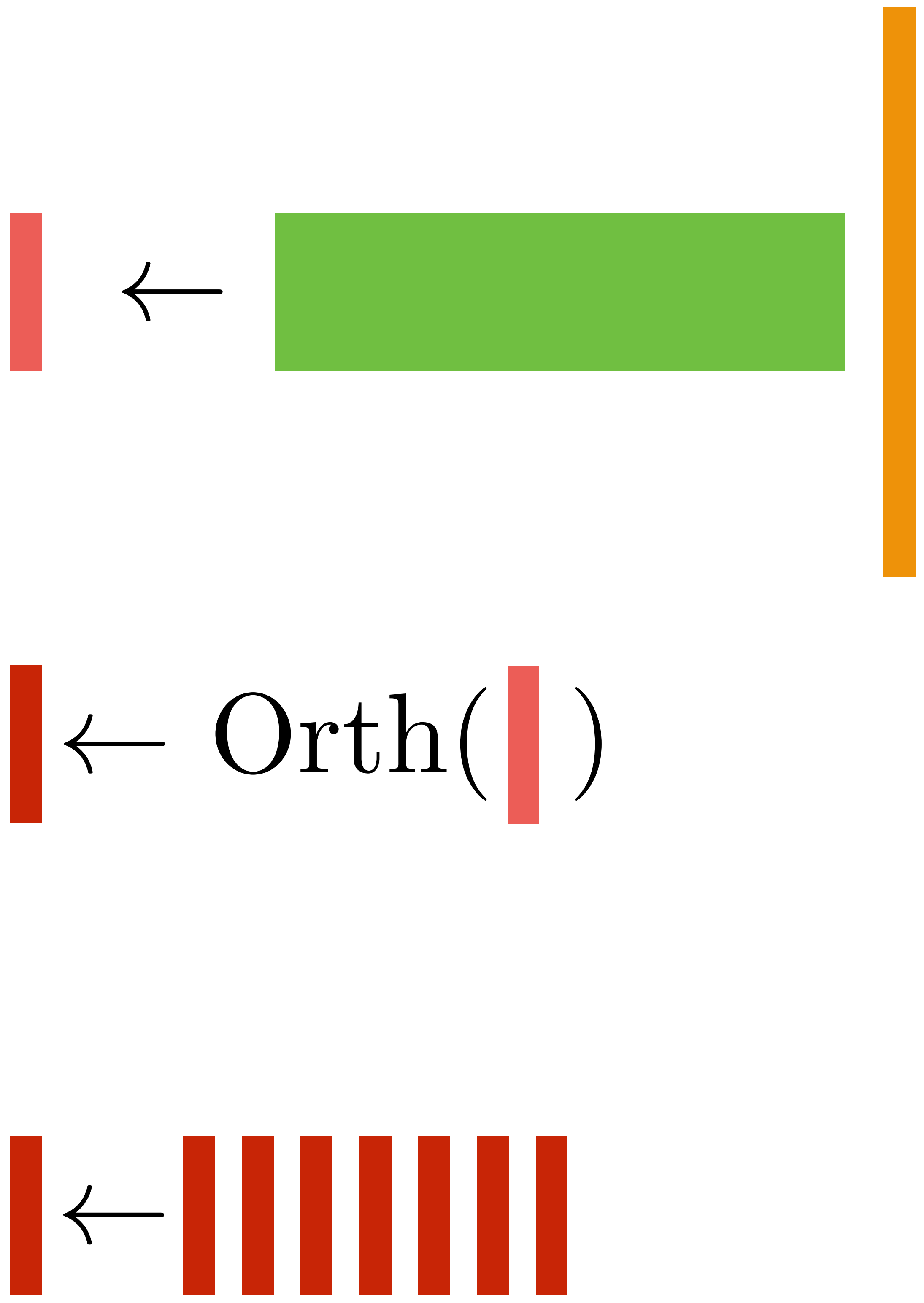
$$\boldsymbol{Q}_{[i]} \leftarrow \mathrm{Orth}(\boldsymbol{Y}_{[i]})$$

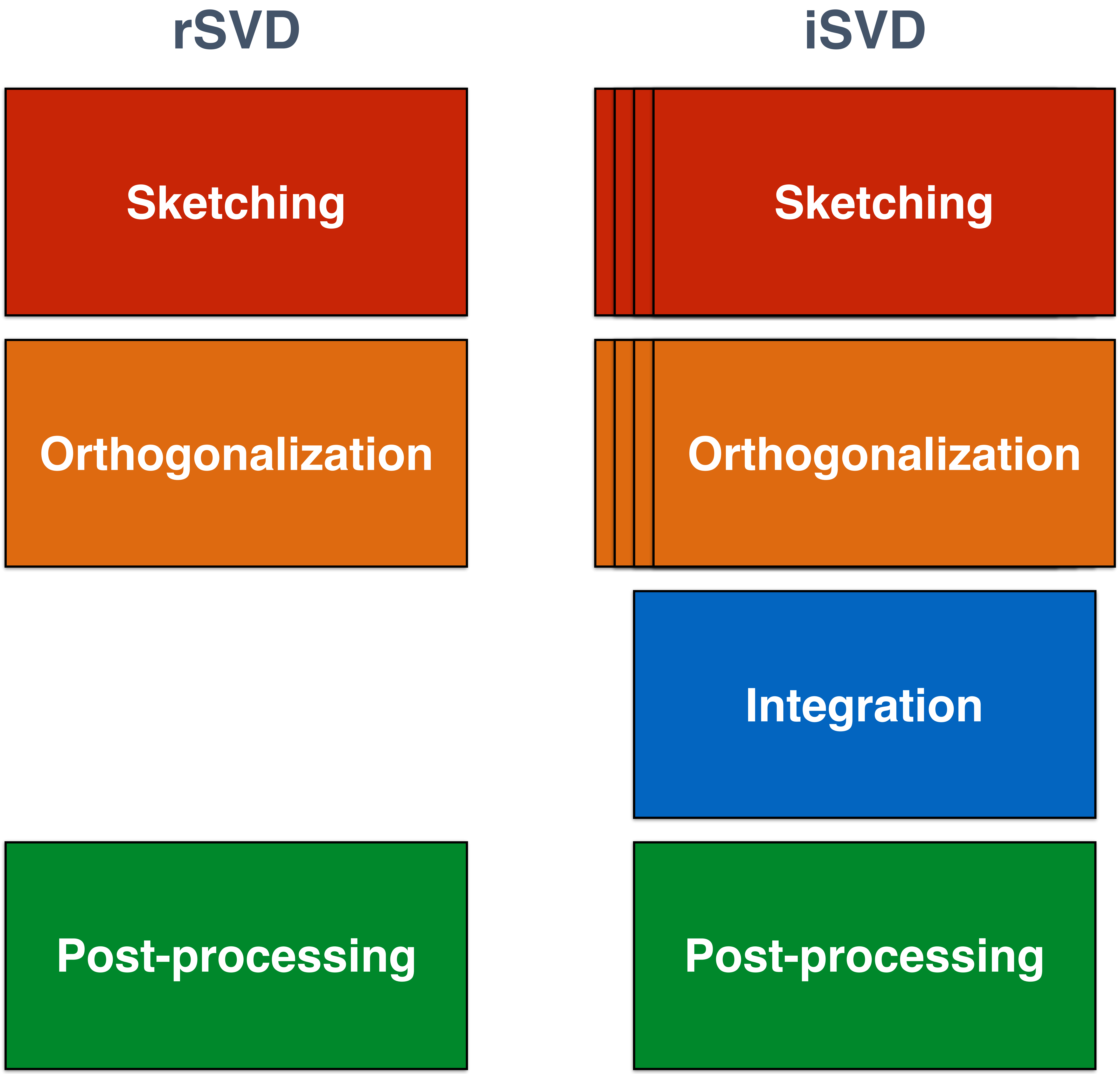
- **Integration of the basis matrices**

$$\overline{\boldsymbol{Q}} \leftarrow \{\boldsymbol{Q}_{[i]}\}_{i=1}^{N}$$

- Post-processing: SVD on the $QQ^\top$-projected subspace

$$\overline{\boldsymbol{Q}}(\overline{\boldsymbol{Q}}^\top \boldsymbol{A}) = \overline{\boldsymbol{Q}}(\widehat{\boldsymbol{W}}_\ell \, \widehat{\boldsymbol{\Sigma}}_\ell \, \widehat{\boldsymbol{V}}_\ell^\top) = \widehat{\boldsymbol{U}}_\ell \, \widehat{\boldsymbol{\Sigma}}_\ell \, \widehat{\boldsymbol{V}}_\ell^\top, \quad \ell = k + p$$




rSVD
iSVD
Sketching
Sketching
Orthogonalization
Orthogonalization
Integration
Post-processing
Post-processing

piSVD (2018)

Naïve Parallelism | Row-Block Parallelism | Column-Block Parallelism

Sketching

- Gaussian Projection
- Column Sampling

Orthogonalization

- Canonical
- Gramian
- Tall-Skinny QR

Integration

- Kolmogorov-Nagumo
- Wen-Yin
- Multilevel Pairwise

Post-processing

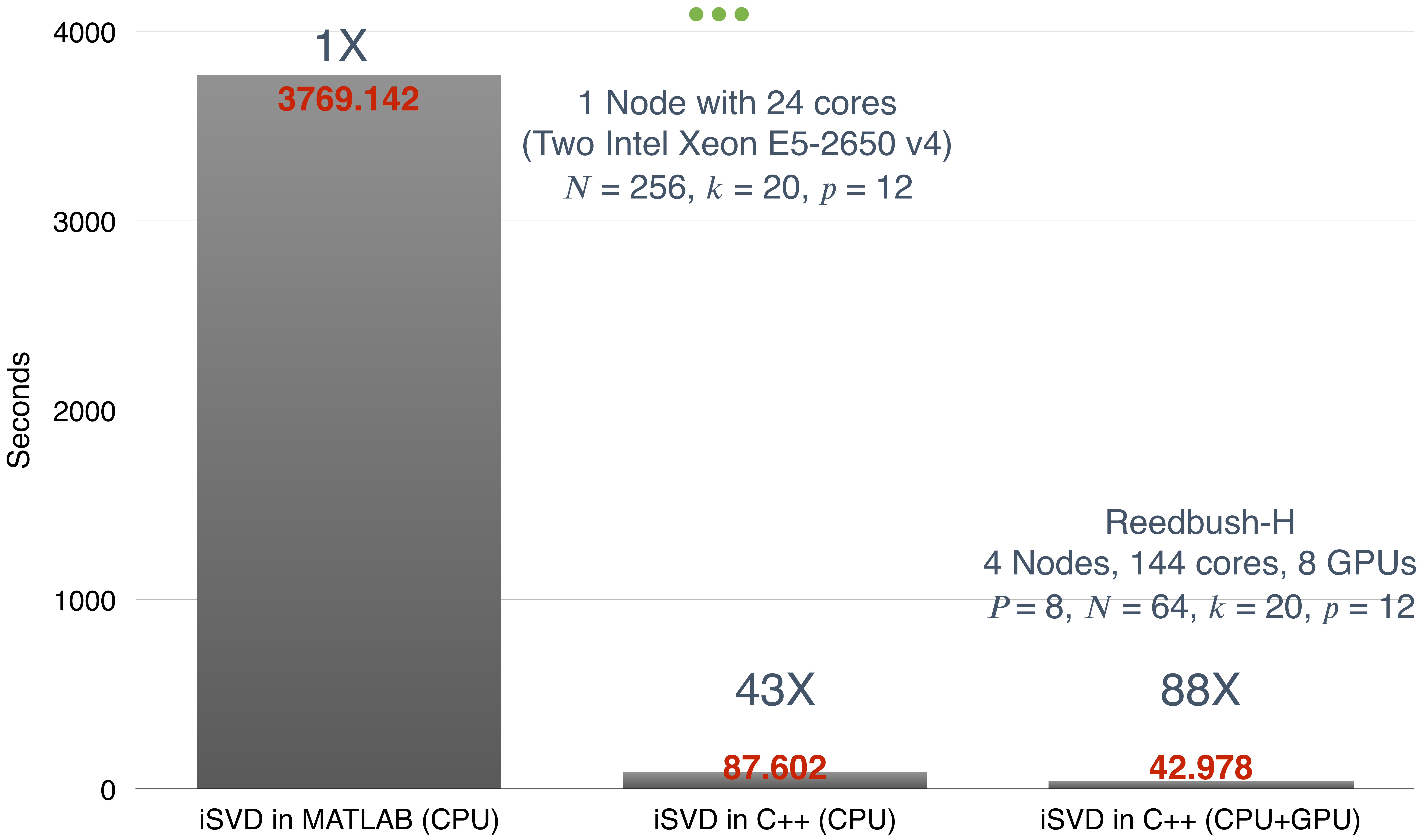
- Canonical
- Gramian
- Tall-Skinny QR
- Symmetric

1000 Genomes Project Phase 1

- 1,092 × 36,781,560 matrix $\boldsymbol{A}$

- iSVD
  - Column-block Gaussian projection sketching (CPU/GPU)
  - Row-block Gramian orthogonalization
  - Row-block Wen-Yin integration
  - Column-block Gramian former

1000 Genomes Project (1,092 × 36,781,560)



Integration

Integrated SVD (iSVD)

- **Multiple** random sketching

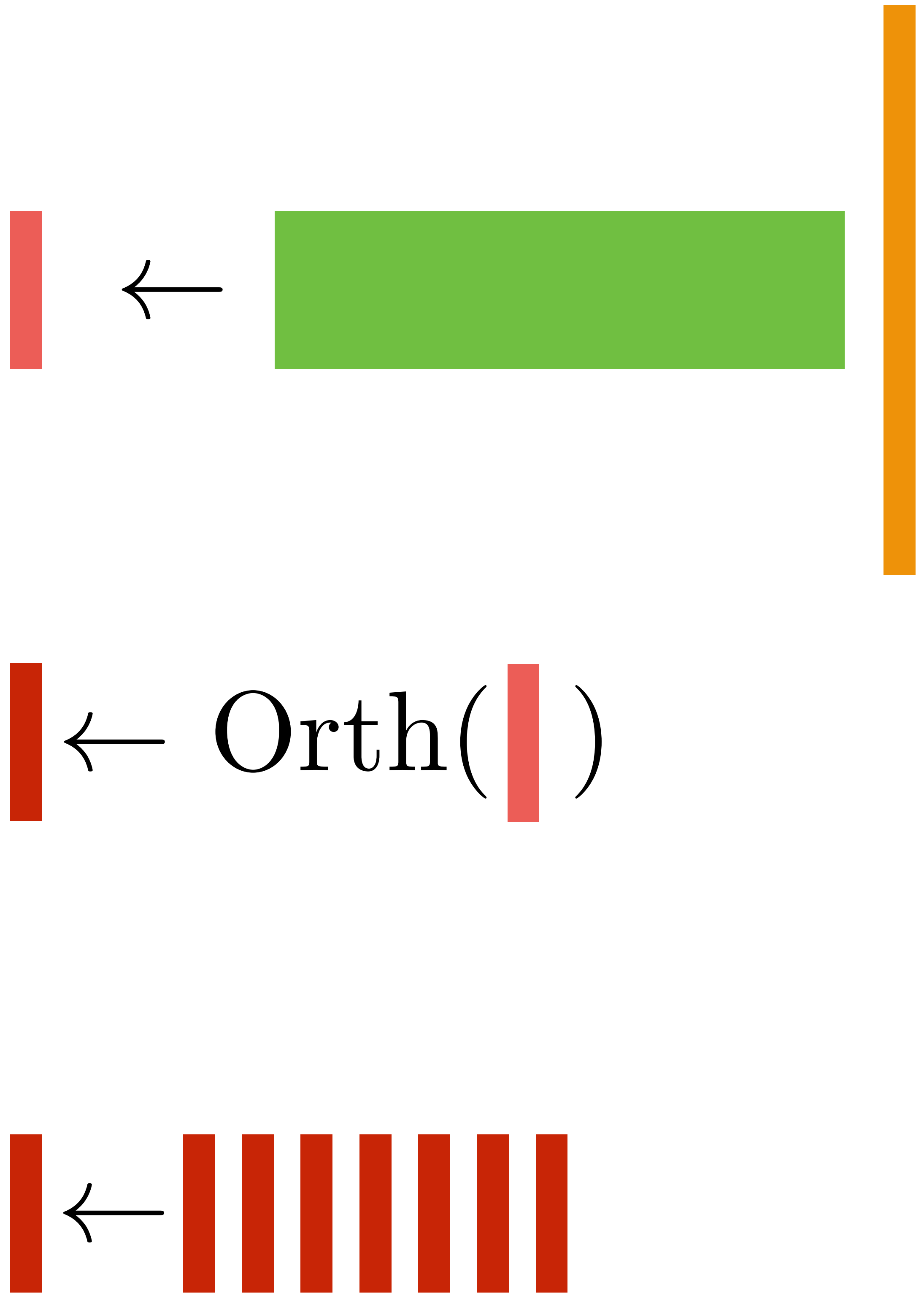
$$\boldsymbol{Y}_{[i]} \leftarrow \boldsymbol{A}\boldsymbol{\Omega}_{[i]}, \quad i = 1, \ldots, N$$

- Orthonormal basis of each sketched subspace

$$\boldsymbol{Q}_{[i]} \leftarrow \mathrm{Orth}(\boldsymbol{Y}_{[i]})$$

- **Integration of the basis matrices**

$$\overline{\boldsymbol{Q}} \leftarrow \{\boldsymbol{Q}_{[i]}\}_{i=1}^{N}$$



- Post-processing: SVD on the QQ$^\top$-projected subspace

$$\overline{\boldsymbol{Q}}(\overline{\boldsymbol{Q}}^\top \boldsymbol{A}) = \overline{\boldsymbol{Q}}(\widehat{\boldsymbol{W}}_\ell \, \widehat{\boldsymbol{\Sigma}}_\ell \, \widehat{\boldsymbol{V}}_\ell^\top) = \widehat{\boldsymbol{U}}_\ell \, \widehat{\boldsymbol{\Sigma}}_\ell \, \widehat{\boldsymbol{V}}_\ell^\top, \quad \ell = k + p$$

Variations of iSVD

Sketching	
Gaussian Projection	Column Sampling

Orthogonalization		
Canonical	Gramian	Tall-Skinny QR

Integration		
Kolmogorov-Nagumo	**Wen-Yin**	**Multi-level Pairwise**

Post-processing			
Canonical	Gramian	Tall-Skinny QR	Symmetric

Target Optimization Problem

Optimal Representation

- Best representation of the projections

$$\overline{\boldsymbol{Q}} = \operatorname{argmin}_{\boldsymbol{Q}\in\mathcal{S}_{m,\ell}} \sum_{i=1}^{N} \left\| \boldsymbol{Q}_{[i]}\boldsymbol{Q}_{[i]}^{\intercal} - \boldsymbol{Q}\boldsymbol{Q}^{\intercal} \right\|_F^2$$

  - Invariant of rotations: $(\boldsymbol{Q}_{[i]}\boldsymbol{R}_\theta)(\boldsymbol{Q}_{[i]}\boldsymbol{R}_\theta)^{\intercal} = \boldsymbol{Q}_{[i]}\boldsymbol{Q}_{[i]}^{\intercal}$
  - l_2-discrepancy

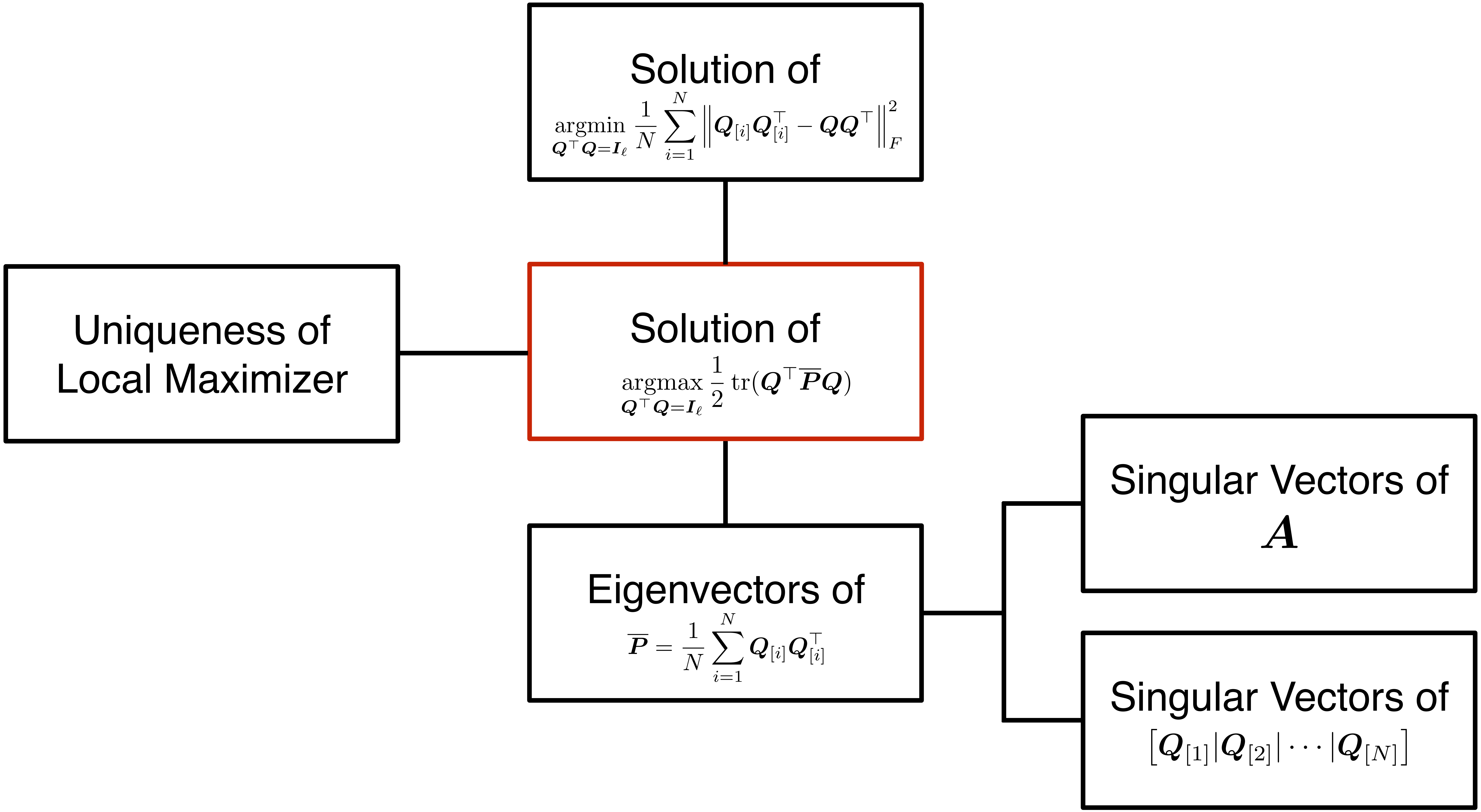
- Stiefel Manifold

$$\mathcal{S}_{m,\ell} = \{\boldsymbol{Q} \in \mathbb{R}^{m\times\ell} : \boldsymbol{Q}^{\intercal}\boldsymbol{Q} = \boldsymbol{I} \text{ and } m \geq \ell\}$$

Integration Methods

- Canonical SVD (The SVD routine in MATLAB or LAPACK…)
- Statistical average by Kolmogorov-Nagumo average on Stiefel Manifold*
- Optimization by line search method proposed by Wen and Yin+
- **Multi-level pairwise integration**

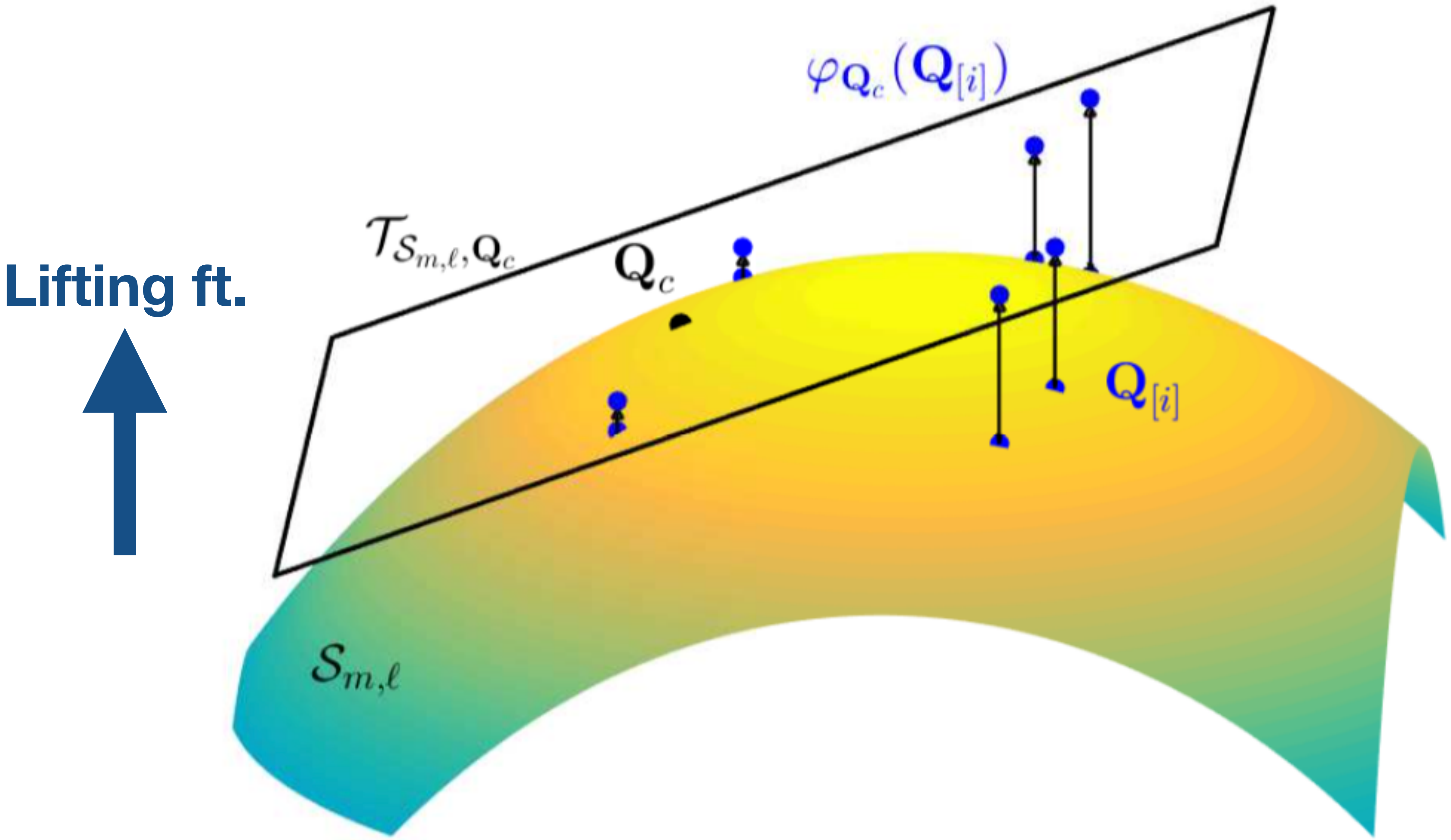
*Fiori, Simone, Tetsuya Kaneko, and Toshihisa Tanaka. "Mixed maps for learning a Kolmogorov-Nagumo-type average element on the compact Stiefel manifold." *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014.
*Kaneko, Tetsuya, Simone Fiori, and Toshihisa Tanaka. "Empirical arithmetic averaging over the compact Stiefel manifold." *IEEE Transactions on Signal Processing* 61.4 (2013): 883-894.
+ Wen, Zaiwen, and Wotao Yin. "A feasible method for optimization with orthogonality constraints." *Mathematical Programming* (2013): 1-38.


Solution of
$\underset{Q^\top Q = I_\ell}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \left\| Q_{[i]} Q_{[i]}^\top - Q Q^\top \right\|_F^2$
Uniqueness of Local Maximizer
Solution of
$\underset{Q^\top Q = I_\ell}{\operatorname{argmax}} \frac{1}{2} \operatorname{tr}(Q^\top \overline{P} Q)$
Eigenvectors of
$\overline{P} = \frac{1}{N} \sum_{i=1}^{N} Q_{[i]} Q_{[i]}^\top$
Singular Vectors of
A
Singular Vectors of
$[Q_{[1]} | Q_{[2]} | \cdots | Q_{[N]}]$

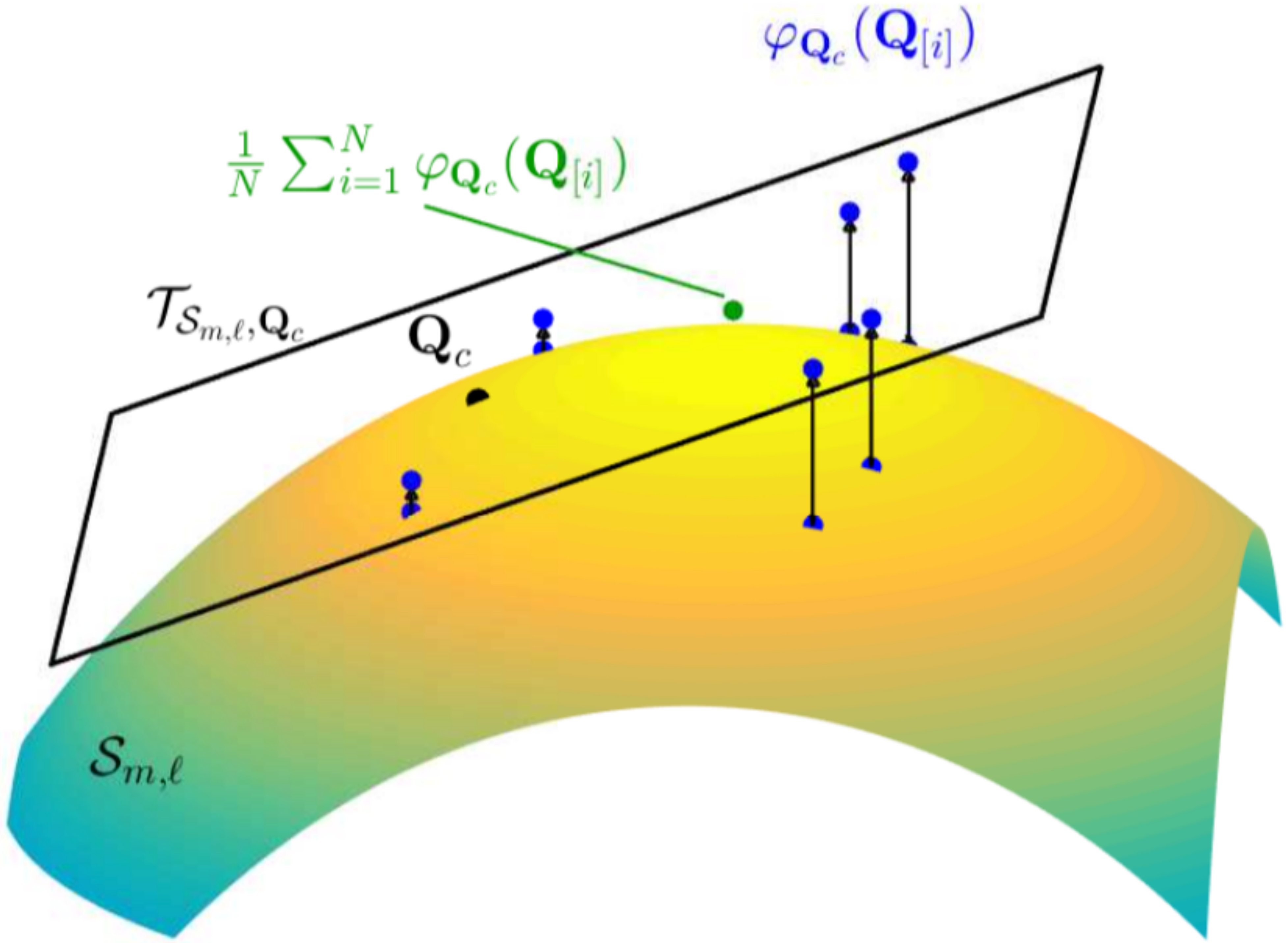
Kolmogorov-Nagumo Average

A Statistical View

One Step Moving in KN Average

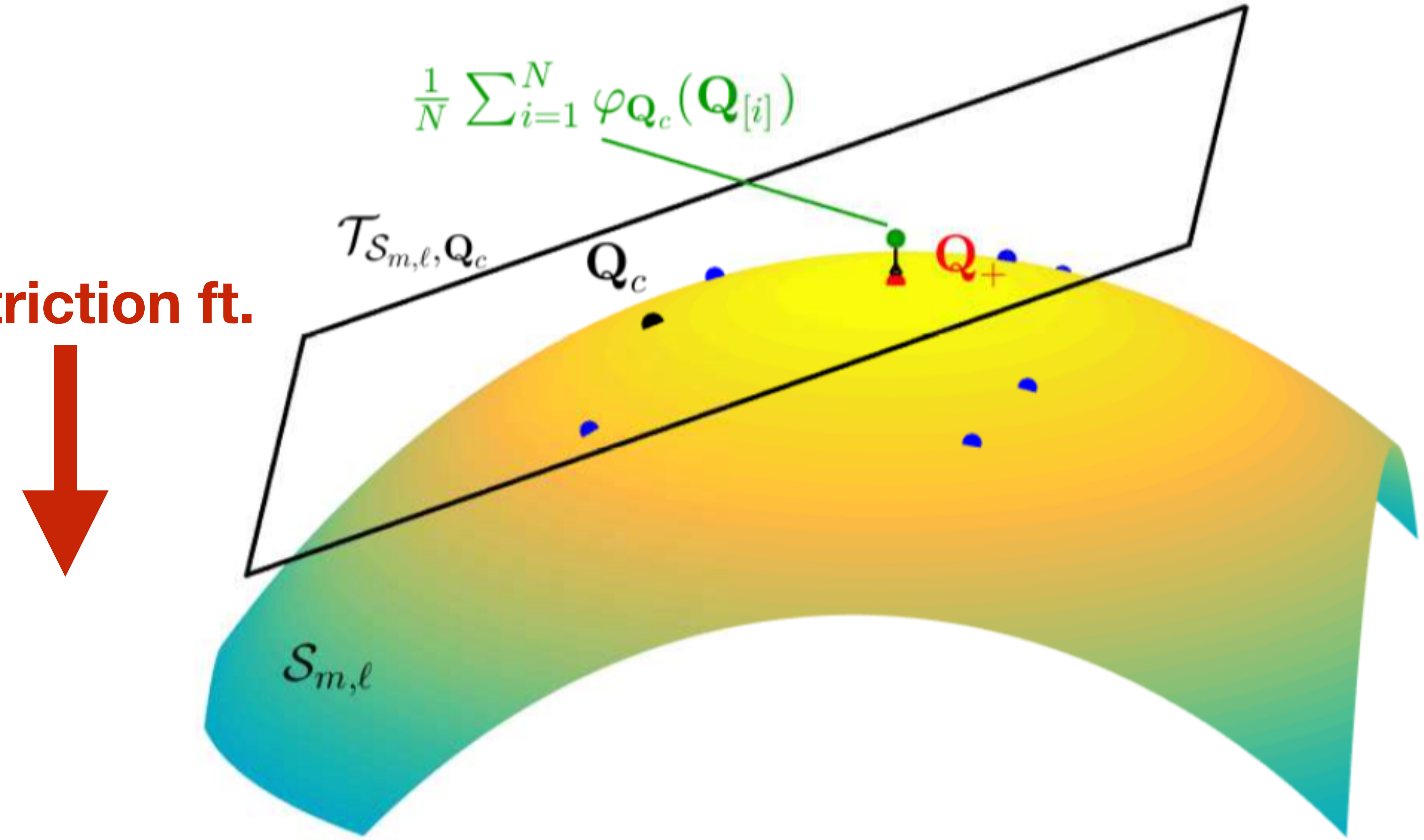


Average

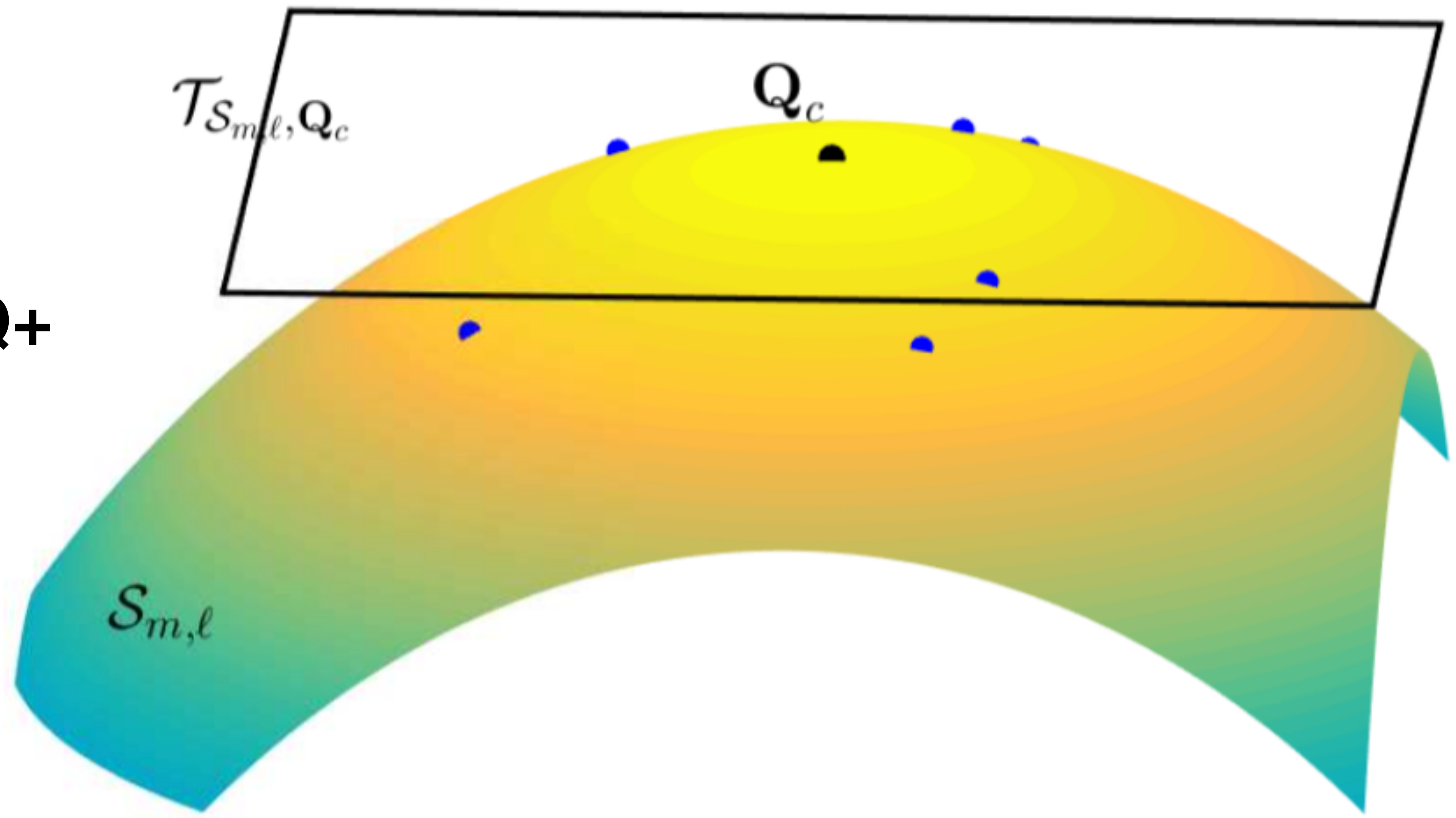


One Step Moving in KN Average

Restriction ft.



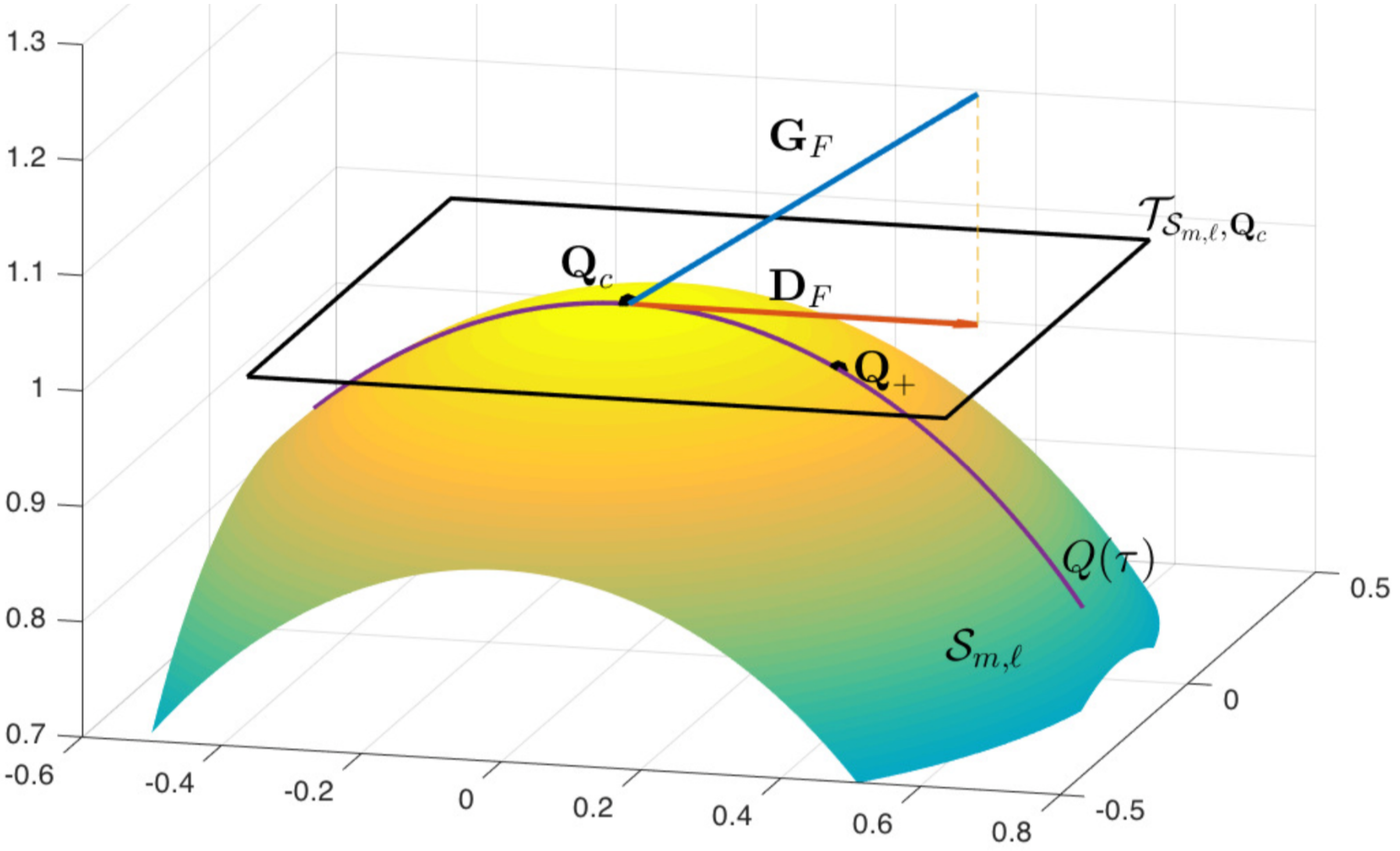
Qc := Q+



Wen-Yin Optimization

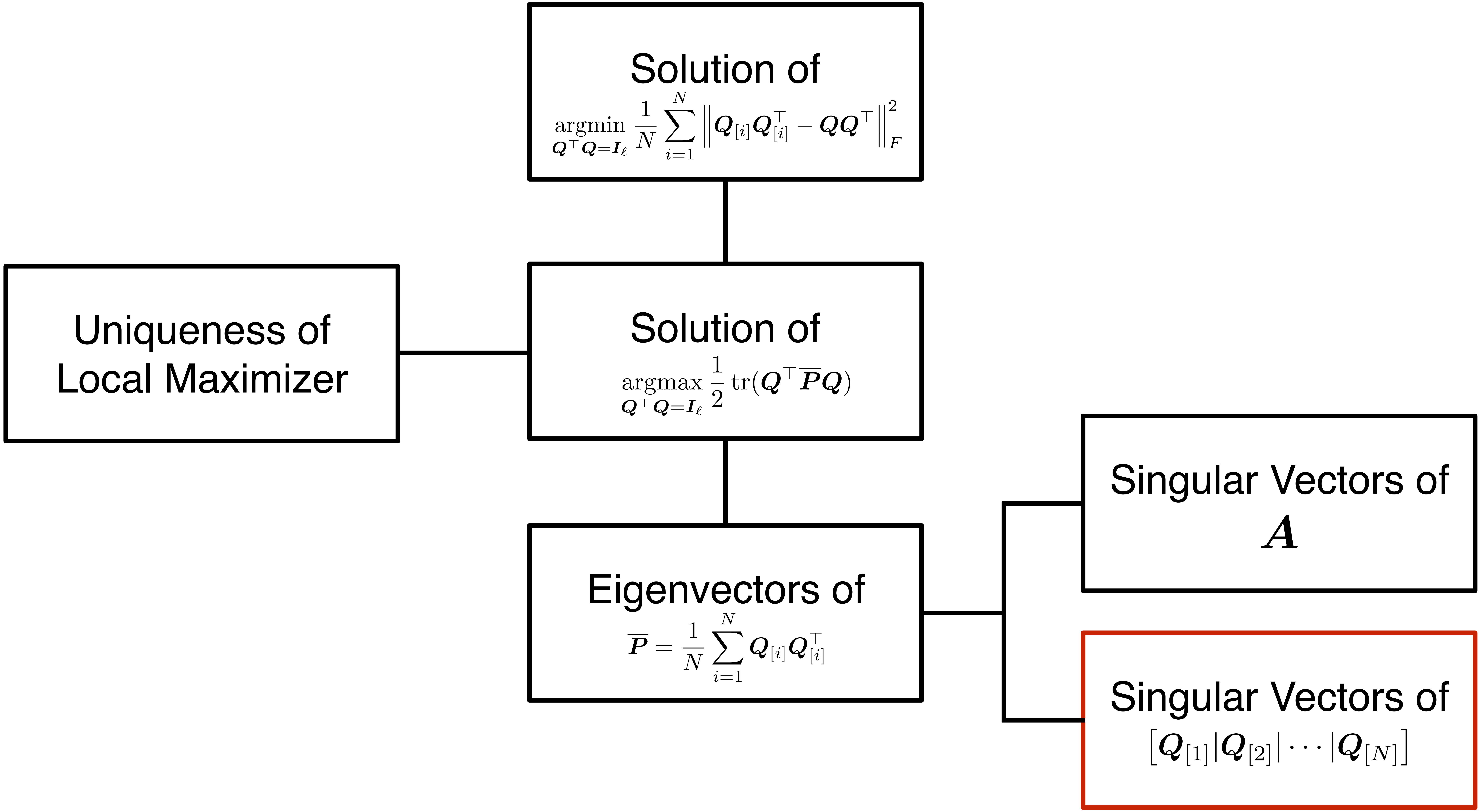
An Optimization View

A Gradient Ascent Method with Line Search



Multilevel Pairwise Integration

A Fast and Parallel Approach


Solution of
$\underset{Q^\top Q = I_\ell}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \left\| Q_{[i]} Q_{[i]}^\top - Q Q^\top \right\|_F^2$
Uniqueness of
Local Maximizer
Solution of
$\underset{Q^\top Q = I_\ell}{\operatorname{argmax}} \frac{1}{2} \operatorname{tr}(Q^\top \overline{P} Q)$
Eigenvectors of
$\overline{P} = \frac{1}{N} \sum_{i=1}^{N} Q_{[i]} Q_{[i]}^\top$
Singular Vectors of
A
Singular Vectors of
$[Q_{[1]} | Q_{[2]} | \cdots | Q_{[N]}]$

Integrated Subspace of Two Sketched Subspaces

- The integrated subspace of a pair of sketched subspaces (N=2) is

$$\overline{\boldsymbol{Q}} = \underset{\boldsymbol{Q} \in \mathcal{S}_{m,\ell}}{\operatorname{argmin}} \left\| \boldsymbol{Q}_{[1]} \boldsymbol{Q}_{[1]}^{\top} - \boldsymbol{Q}\boldsymbol{Q}^{\top} \right\|_F^2 + \left\| \boldsymbol{Q}_{[2]} \boldsymbol{Q}_{[2]}^{\top} - \boldsymbol{Q}\boldsymbol{Q}^{\top} \right\|_F^2$$

- The optimal solution of the above optimization problem is
  - the leading ℓ eigenvectors of $\boldsymbol{Q}_{[1]}\boldsymbol{Q}_{[1]}^{\top} + \boldsymbol{Q}_{[2]}\boldsymbol{Q}_{[2]}^{\top}$
  - or equivalently, the leading ℓ singular vectors of $[\boldsymbol{Q}_{[1]} | \boldsymbol{Q}_{[2]}]$

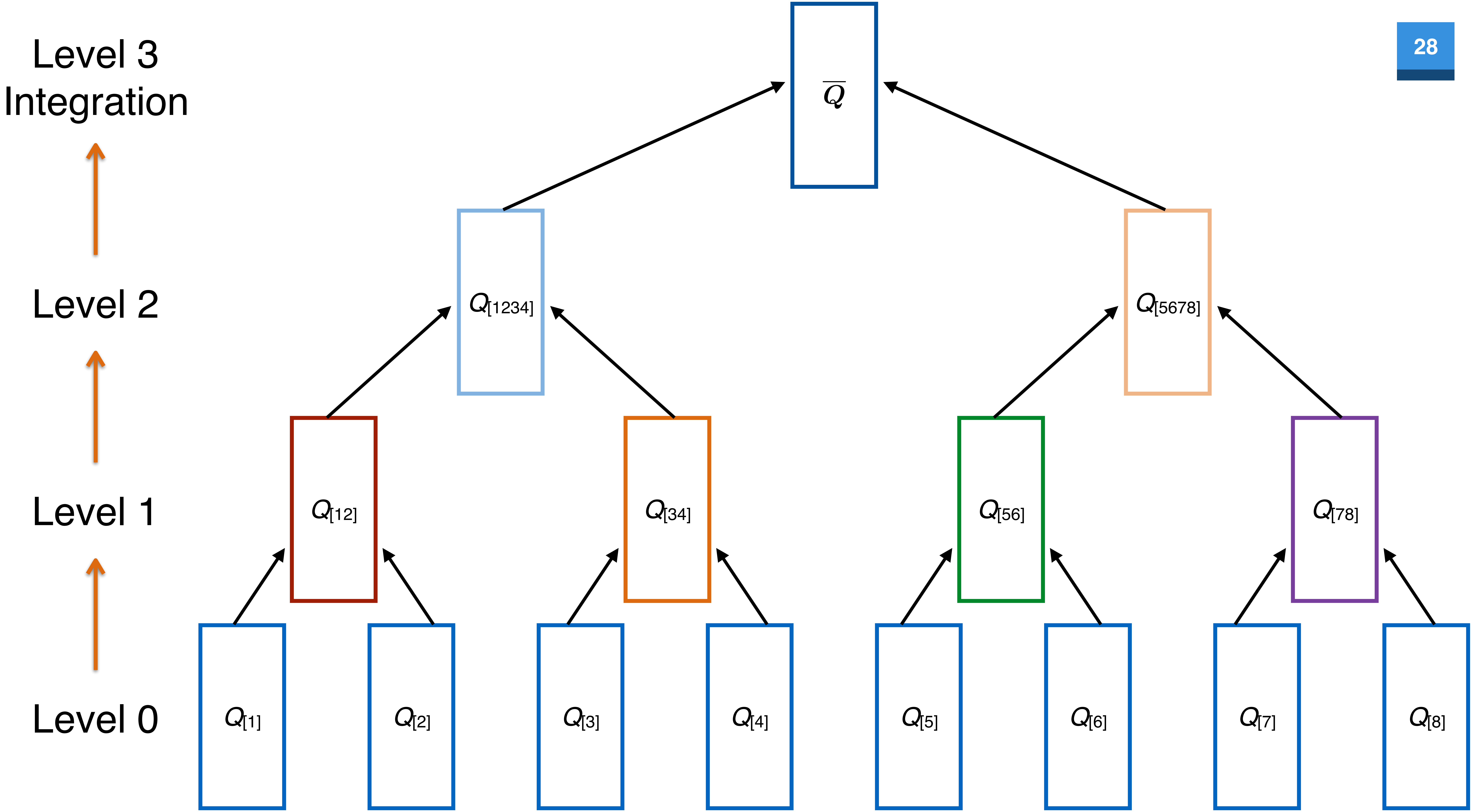
Integration by A Fast Pairwise Sketched Subspace Average

- Let $M = [Q_1 \,|\, Q_2] = L\Sigma R^T \approx L_\ell \Sigma_\ell R_\ell^T$ and $Q_1^T Q_2 = USV^T$. We have

$$\boldsymbol{M}^\top \boldsymbol{M} = \begin{bmatrix} \boldsymbol{I}_\ell & \boldsymbol{Q}_1^\top \boldsymbol{Q}_2 \\ \boldsymbol{Q}_2^\top \boldsymbol{Q}_1 & \boldsymbol{I}_\ell \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_\ell & \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top \\ \boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^\top & \boldsymbol{I}_\ell \end{bmatrix} = \left(\frac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{U} & \boldsymbol{U} \\ \boldsymbol{V} & -\boldsymbol{V} \end{bmatrix}\right) \begin{bmatrix} \boldsymbol{I}+\boldsymbol{S} & \\ & \boldsymbol{I}-\boldsymbol{S} \end{bmatrix} \left(\frac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{U} & \boldsymbol{U} \\ \boldsymbol{V} & -\boldsymbol{V} \end{bmatrix}\right)^\top .$$

$$M^T M = R\Sigma^2 R^T$$

$$\boldsymbol{\Sigma}_\ell^2 = \boldsymbol{I}+\boldsymbol{S},\ \boldsymbol{R}_\ell = \tfrac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix} \text{ and } \boldsymbol{L}_\ell = \boldsymbol{M}\frac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix}(\boldsymbol{I}+\boldsymbol{S})^{-\frac{1}{2}} = (\boldsymbol{Q}_1\boldsymbol{U}+\boldsymbol{Q}_2\boldsymbol{V})(2(\boldsymbol{I}+\boldsymbol{S}))^{-\frac{1}{2}}.$$


Level 3
Integration
Level 2
Level 1
Level 0
$\overline{Q}$
$Q_{[1234]}$
$Q_{[5678]}$
$Q_{[12]}$
$Q_{[34]}$
$Q_{[56]}$
$Q_{[78]}$
$Q_{[1]}$
$Q_{[2]}$
$Q_{[3]}$
$Q_{[4]}$
$Q_{[5]}$
$Q_{[6]}$
$Q_{[7]}$
$Q_{[8]}$

Algorithm and Complexity

Algorithm 2-2 Hierarchical Reduction

Require: The orthogonal matrices to be integrated $\boldsymbol{Q}_{[1]}, \boldsymbol{Q}_{[2]}, \ldots, \boldsymbol{Q}_{[N]}$.

Ensure: The average $\overline{\boldsymbol{Q}}$.

Set $n = N$.
while $n > 1$ **do** ← Loop for levels
  Set $m = \lfloor \frac{n}{2} \rfloor$
  for $i = 1, 2, \ldots, m$ **do** ← Loop for pairs
    Find SVD of $\boldsymbol{Q}_{[2i-1]}^{\top} \boldsymbol{Q}_{[2i]}$ as $\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\top}$. $O(\ell^3)$
    $\boldsymbol{Q}_{[i]} \leftarrow (\boldsymbol{Q}_{[2i-1]}\boldsymbol{U} + \boldsymbol{Q}_{[2i]}\boldsymbol{V})(2(\boldsymbol{I} + \boldsymbol{S}))^{-\frac{1}{2}}$. $O(m\ell^2)$
  end for
  $n \leftarrow \lceil \frac{n}{2} \rceil$
end while
$\overline{\boldsymbol{Q}} = \boldsymbol{Q}_{[1]}$.

Total: $O(Nm\ell^2 + N\ell^3)$

Comparison of Integration Methods

	Canonical SVD	KN Average WL Optimization	Multilevel Pairwise Integration
Complexity	$O(N^2 m\ell^2)$	$O(Nm\ell^2 \#Iter)$	$O(Nm\ell^2)$
Theoretical Accuracy of Result	Exactly the integrated subspaces defined previously.	Close to integrated subspace for few interaction steps. Exactly the integrated subspace while converged.	Approximation of integrated subspace.

Numerical Experiments

Setting and Environment

- The desired rank in all tests is $k = 10$.
- The oversampling number is $p = 12$.
- The test codes are implemented in MATLAB without optimization on speed.
- The tests are done in different machine due to the issue of memory size.
- All the timing tests are done in MacBook Pro (Mid. 2014). (Processor: 2.6 GHz Intel Core i5. 2 cores. 4 threads. Memory: 8 GB 1600 MHz DDR3)

Test Matrices

- The test matrices in the following tests are generated by

$$\boldsymbol{A} = \boldsymbol{H}_m \boldsymbol{\Sigma} \boldsymbol{H}_n^\top$$

where $\boldsymbol{H}_m, \boldsymbol{H}_n$ denote the Hadamard matrix with size $m = 2^d$, $n = 2^{d+1}$. The diagonal matrix $\boldsymbol{\Sigma}$ is given by different entries in different test matrices for $k = 10$.

$$A_H(10^{-1}) : \sigma_{i,i} = \begin{cases} (10^{-1})^{\frac{i-1}{k}} & \text{if } i \leq k \\ \dfrac{10^{-1}(m-i)}{m-k-1} & \text{otherwise} \end{cases} \qquad A_H(10^{-3}) : \sigma_{i,i} = \begin{cases} (10^{-3})^{\frac{i-1}{k}} & \text{if } i \leq k \\ \dfrac{10^{-3}(m-i)}{m-k-1} & \text{otherwise} \end{cases}$$

- Some matrices from SuiteSparse matrix collection

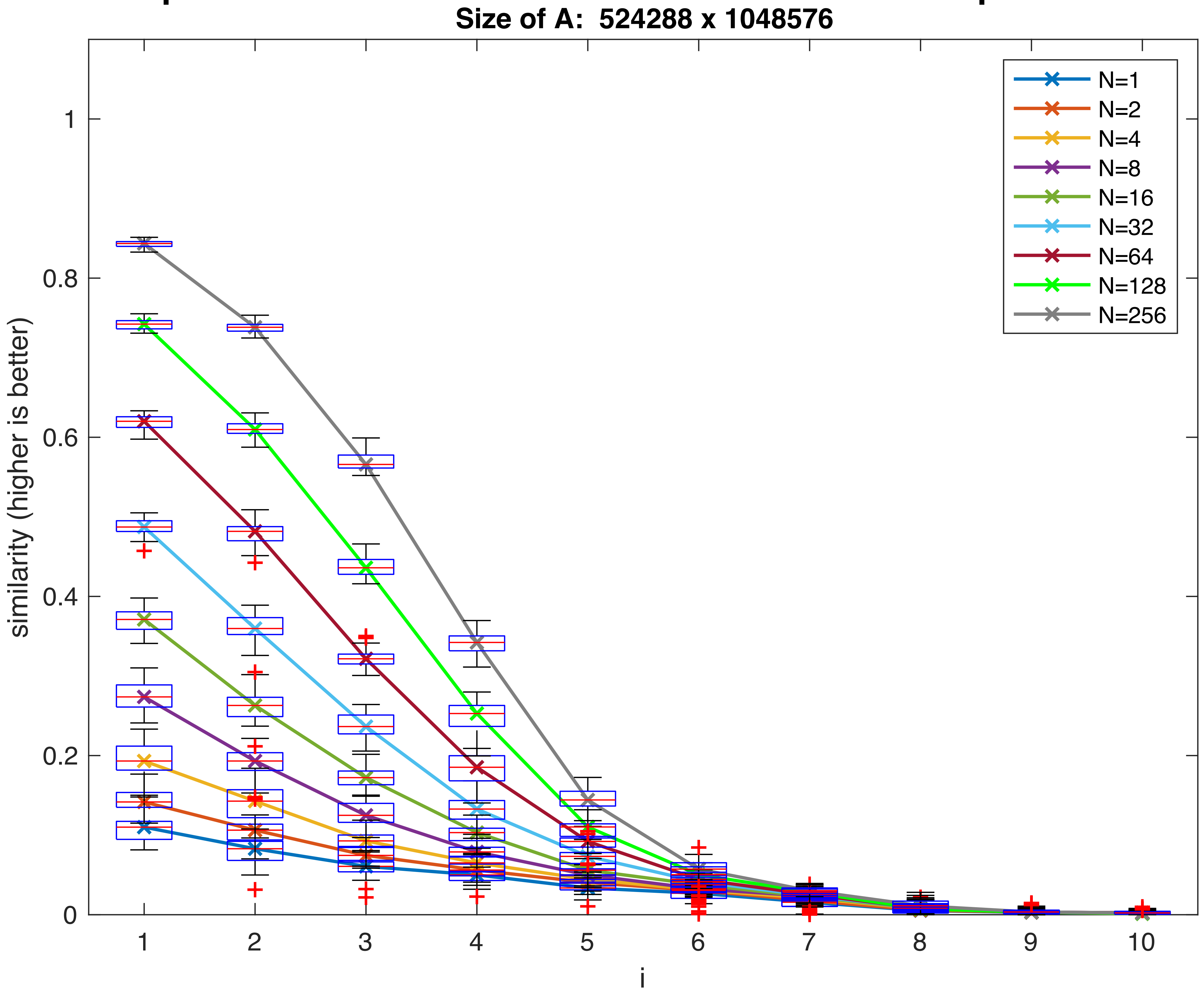
Error Measurement

- Singular vector similarity
  - Inner product of each columns between Q_{test} and Q_{true}
  - The angle of each singular vectors
  - The values are close to 1 if the approximation is good

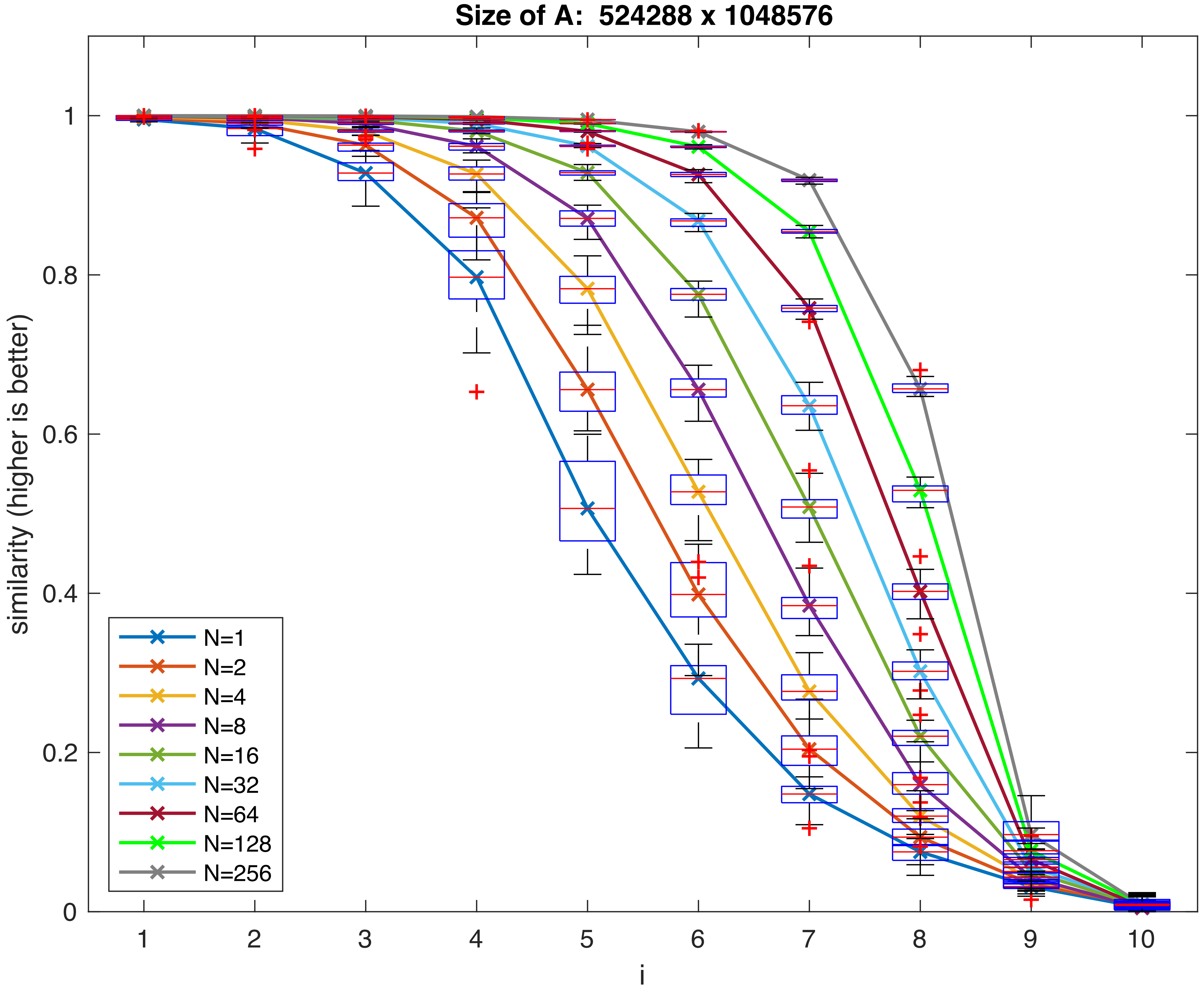
- Canonical angles
  - Singular values of the matrix $Q_{test}^\top Q_{true}$
  - Distance of two subspaces
  - The values are close to 1 if the approximation is good

Comparison of Different N

- Repeat each case for 30 times and plot in box plot.



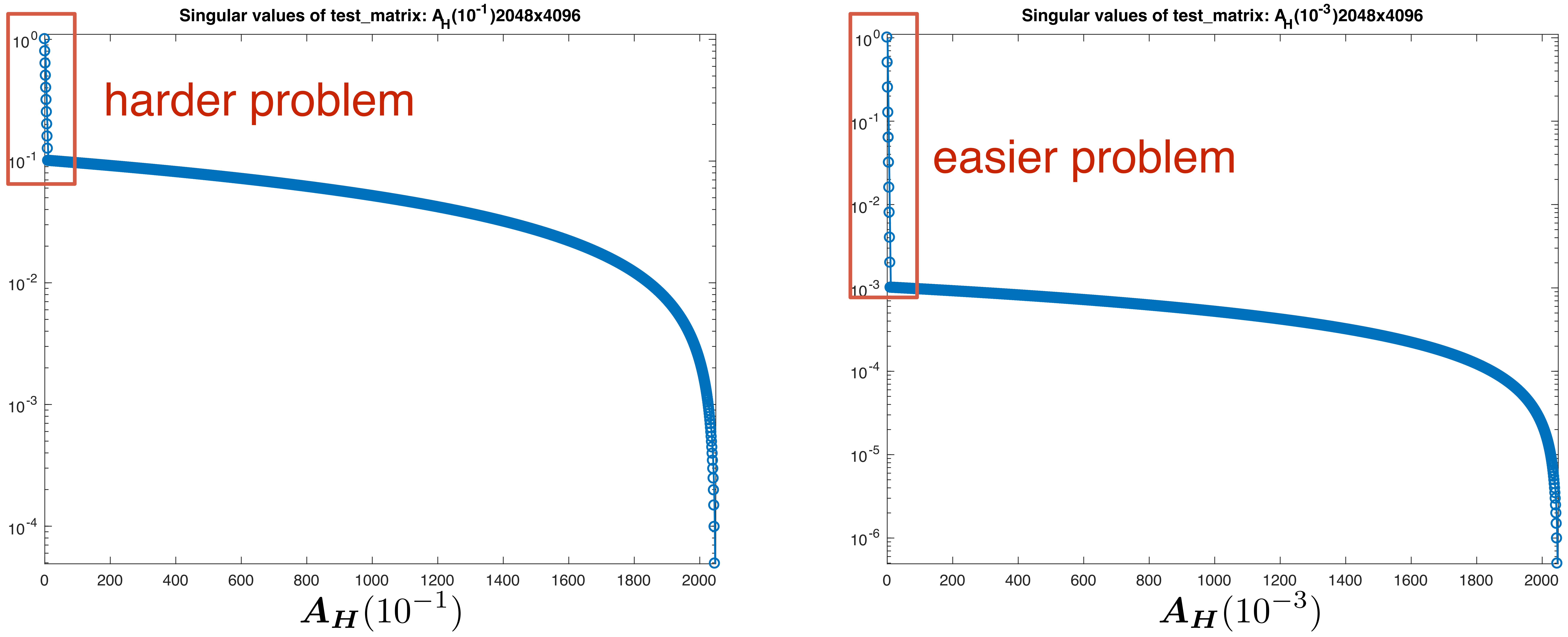
$\boldsymbol{A_H}(10^{-1})$



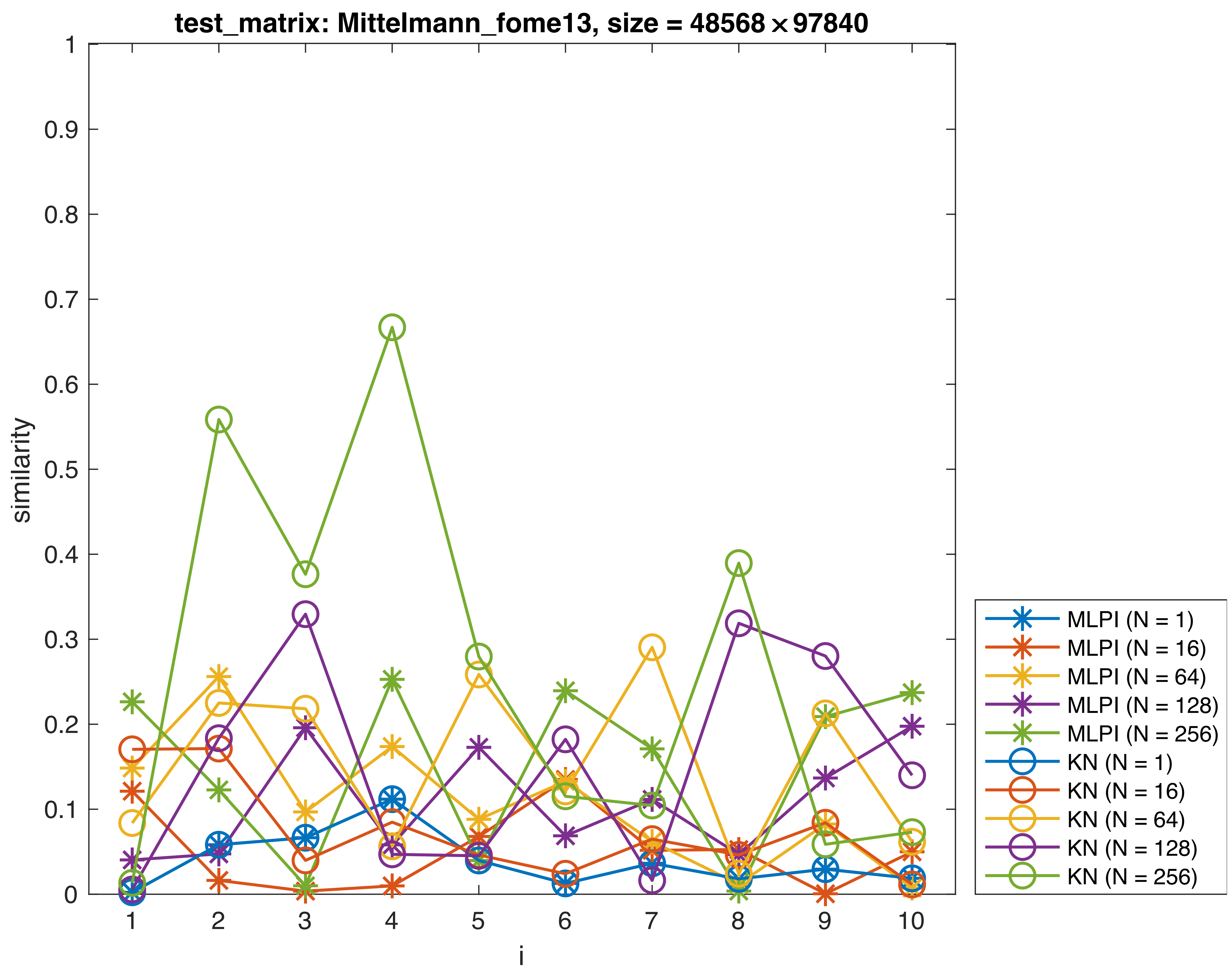
$\boldsymbol{A_H}(10^{-3})$

Singular Values of Test Matrices

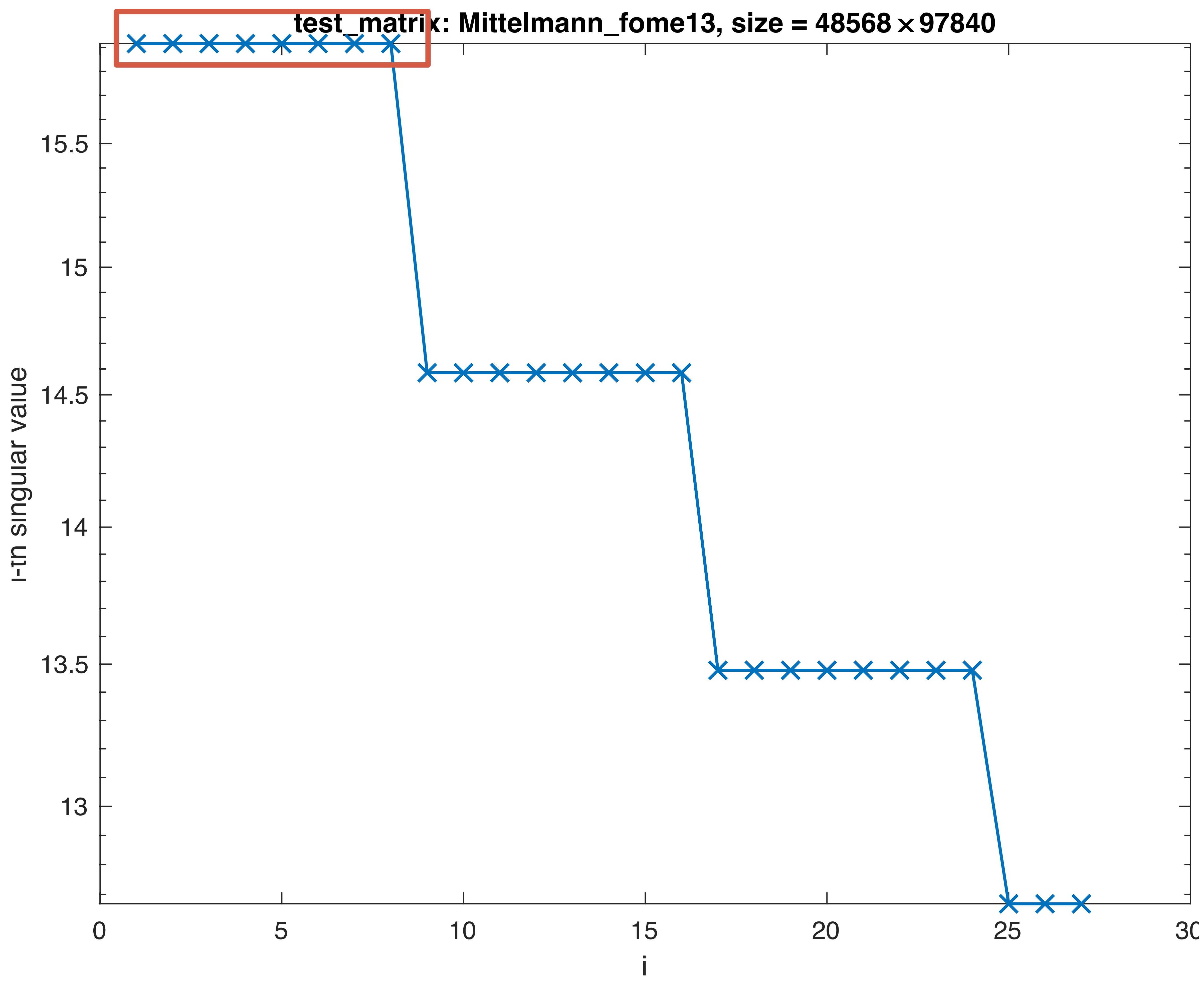
- The larger the difference between each singular values, the easier to capture the leading singular vectors by Gaussian projection.



SuiteSparse Test Matrix: Mittelmann_fome13

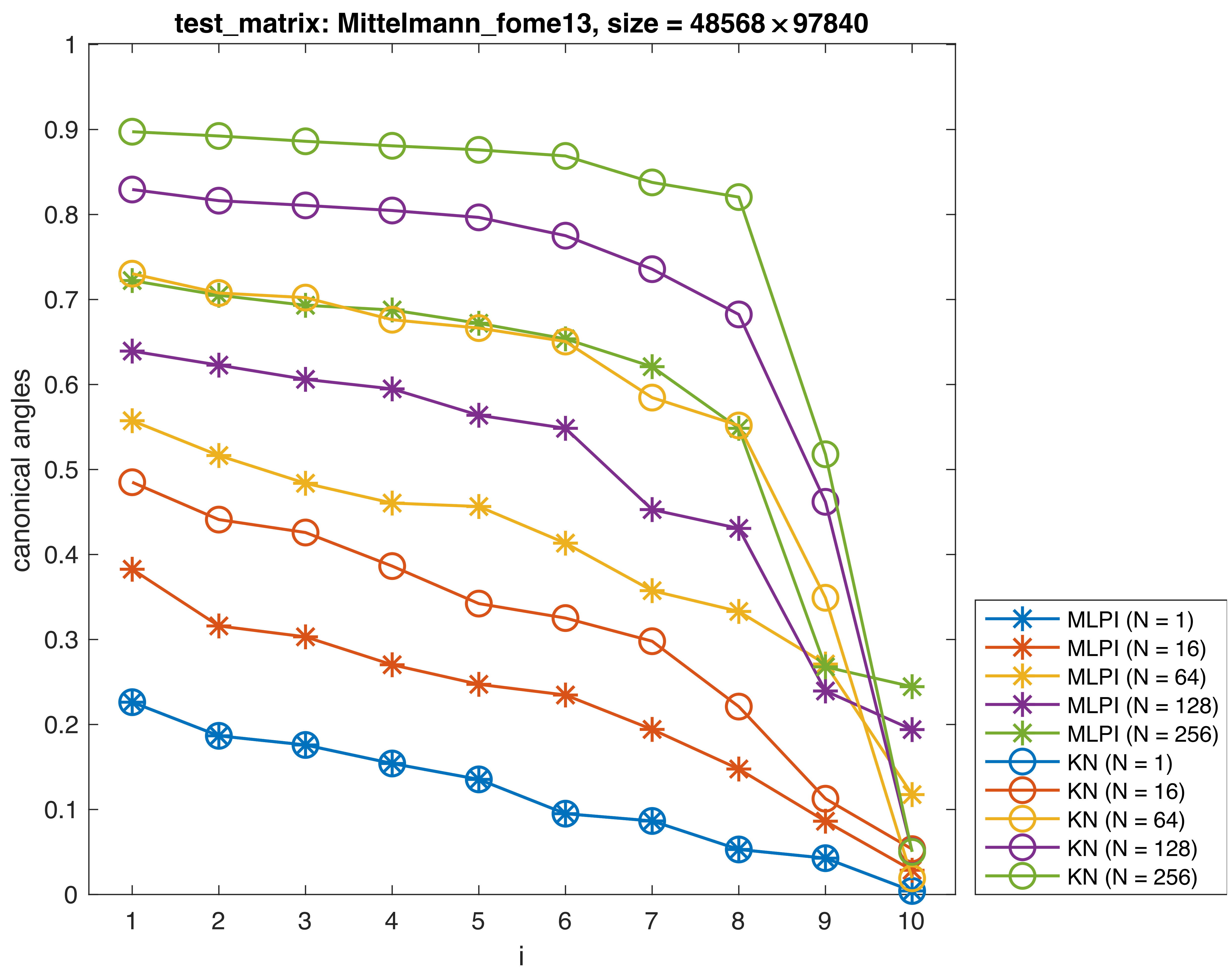


Similarities

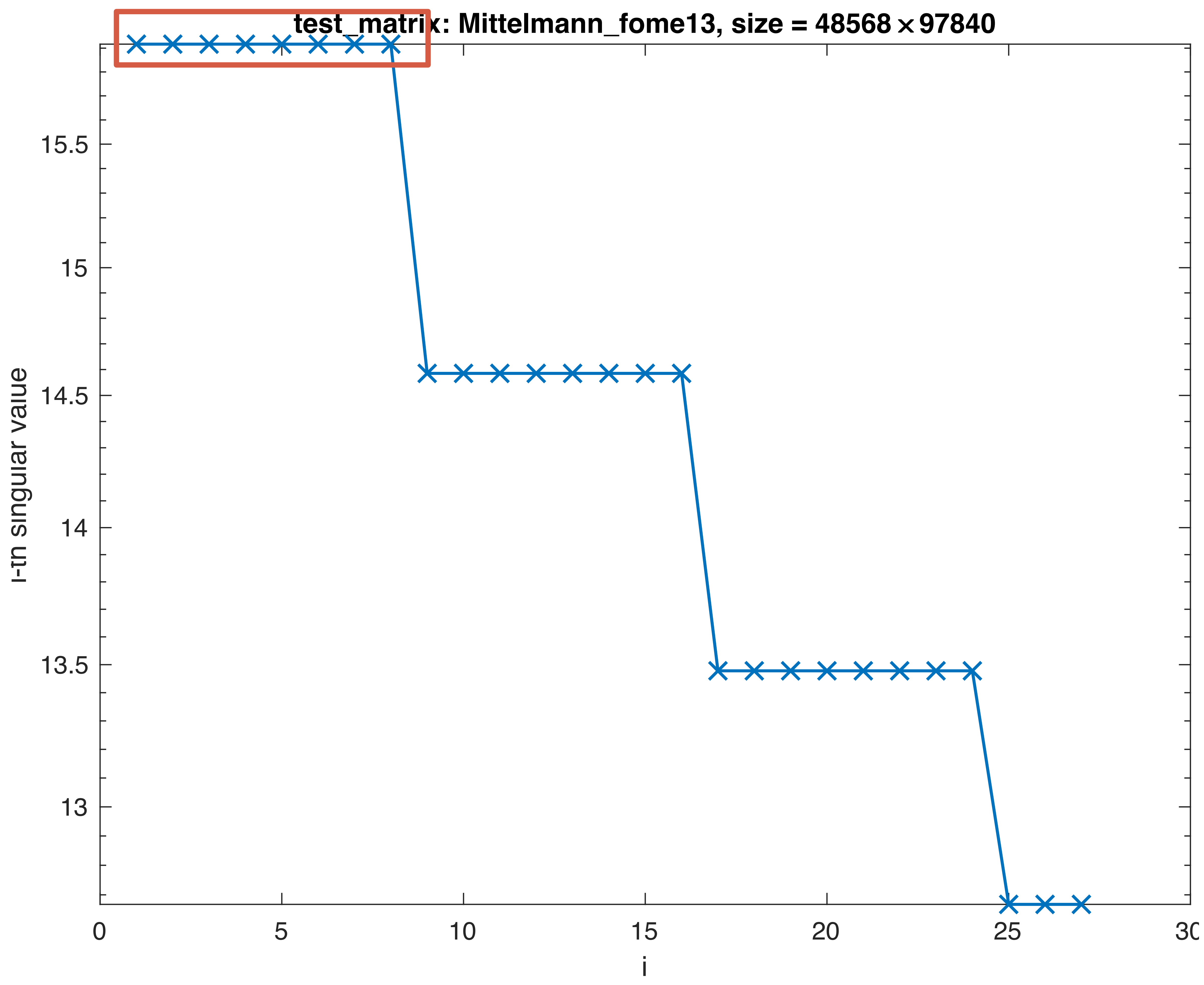


First 27 singular values of test matrix

SuiteSparse Test Matrix: Mittelmann_fome13

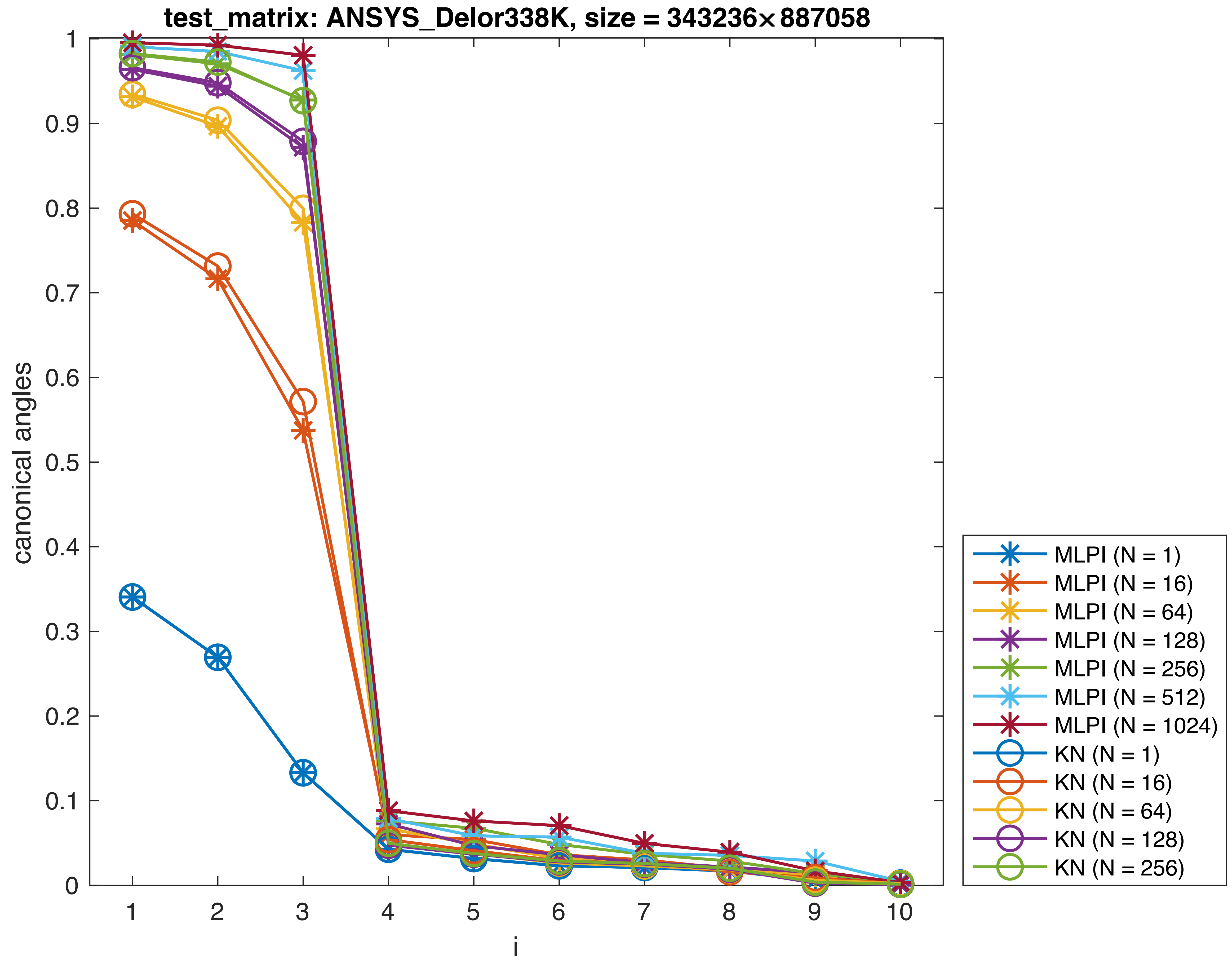


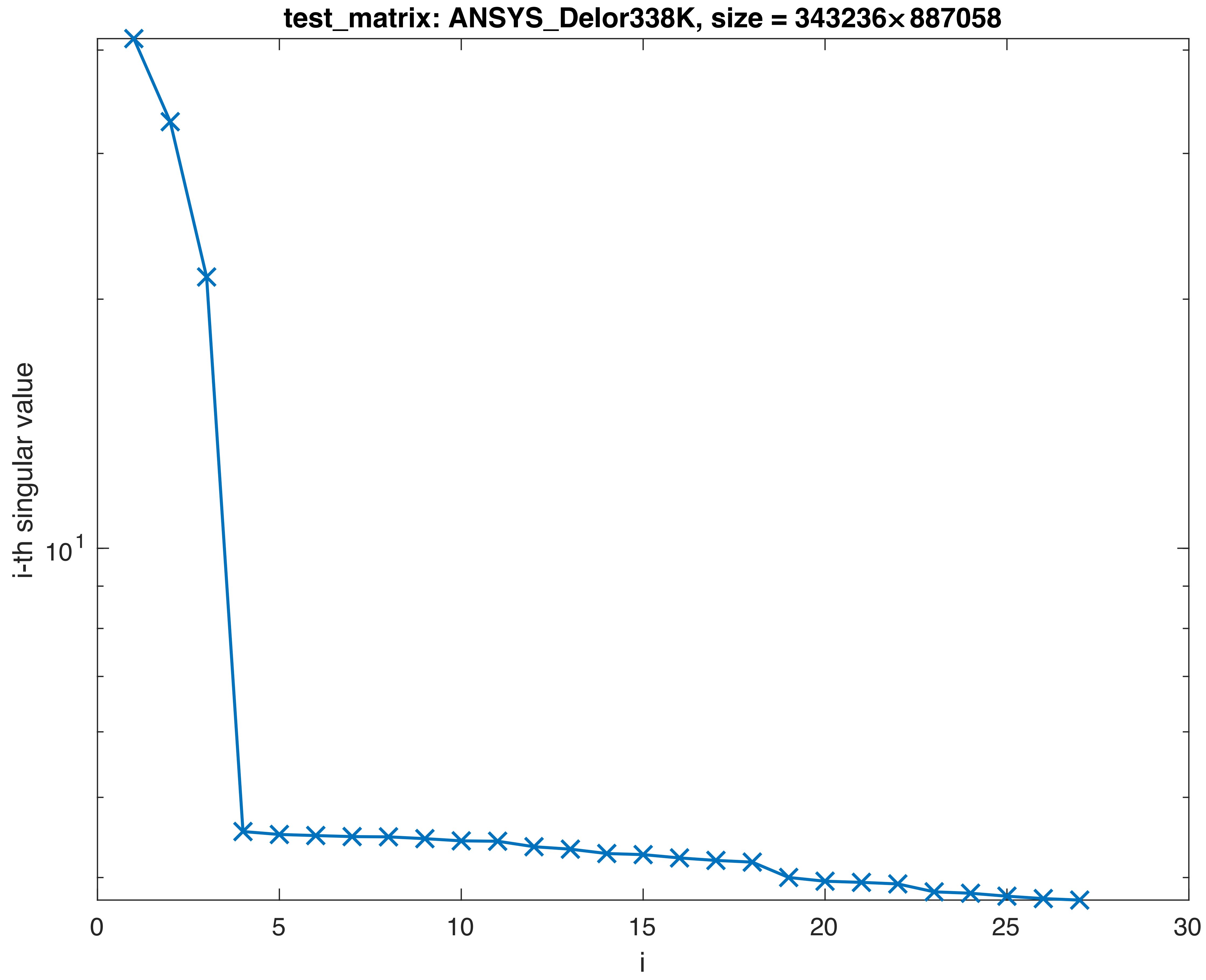
Canonical Angles



First 27 singular values of test matrix

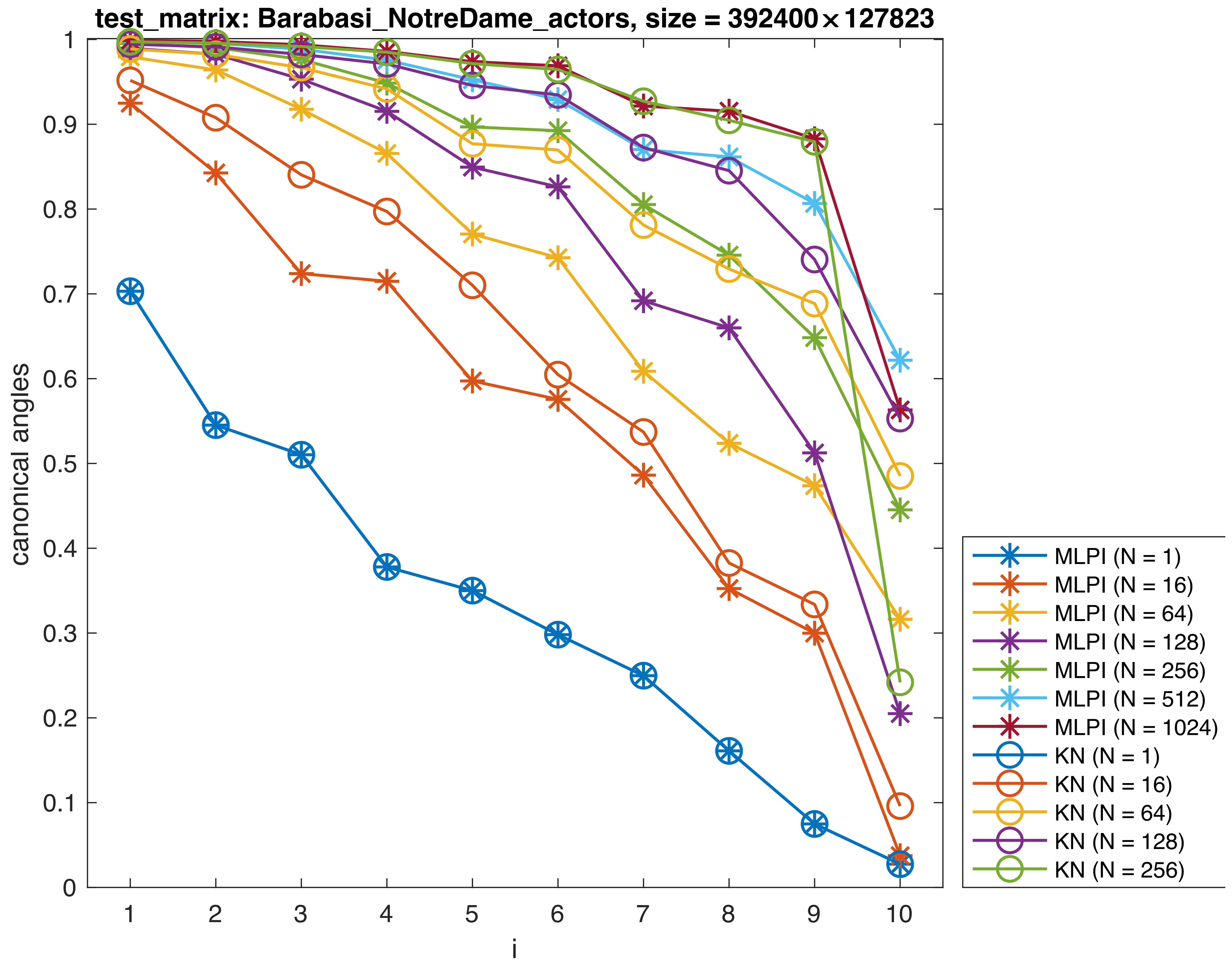
SuiteSparse Test Matrix: ANSYS_Delor338K

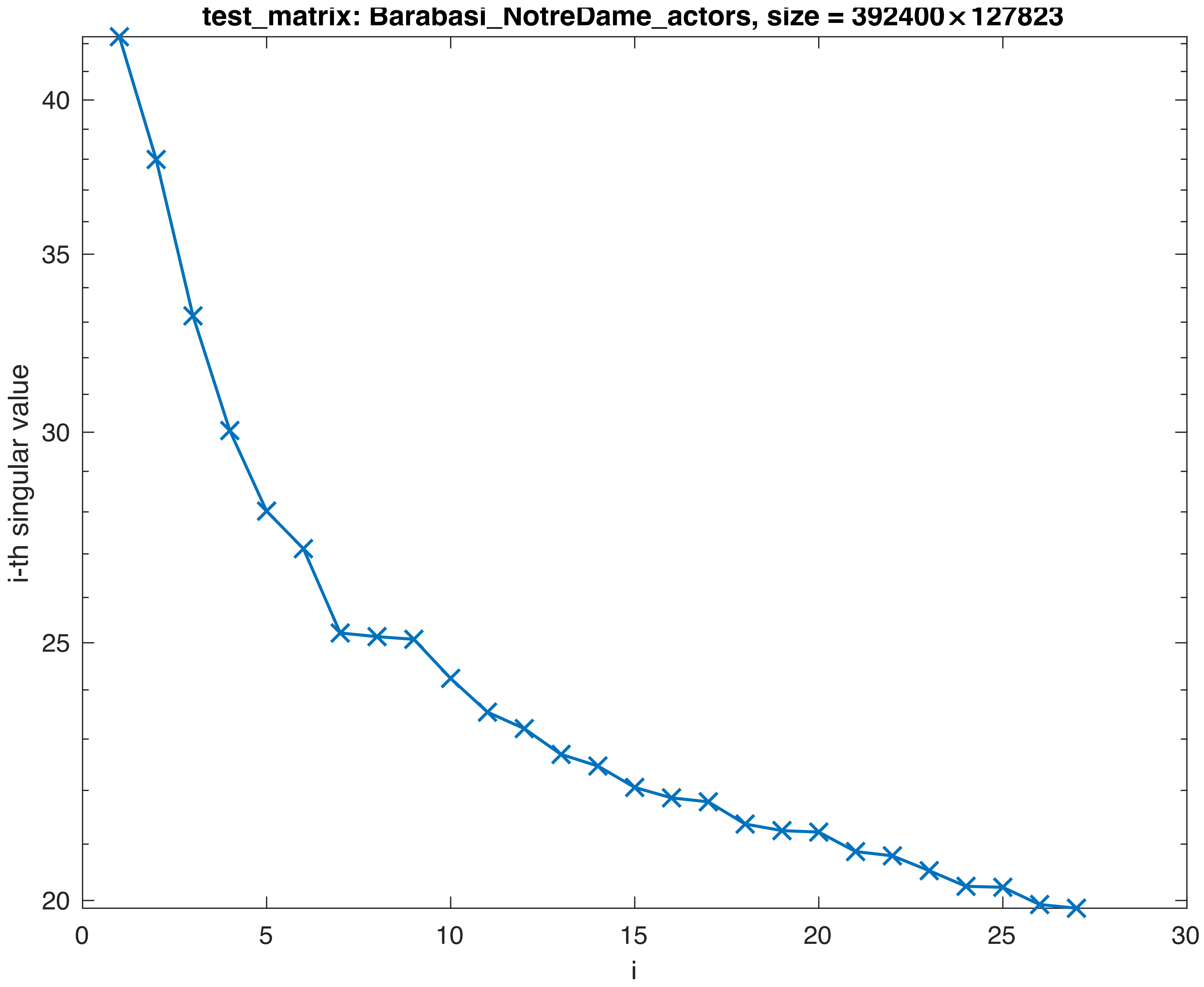




First 27 singular values of test matrix

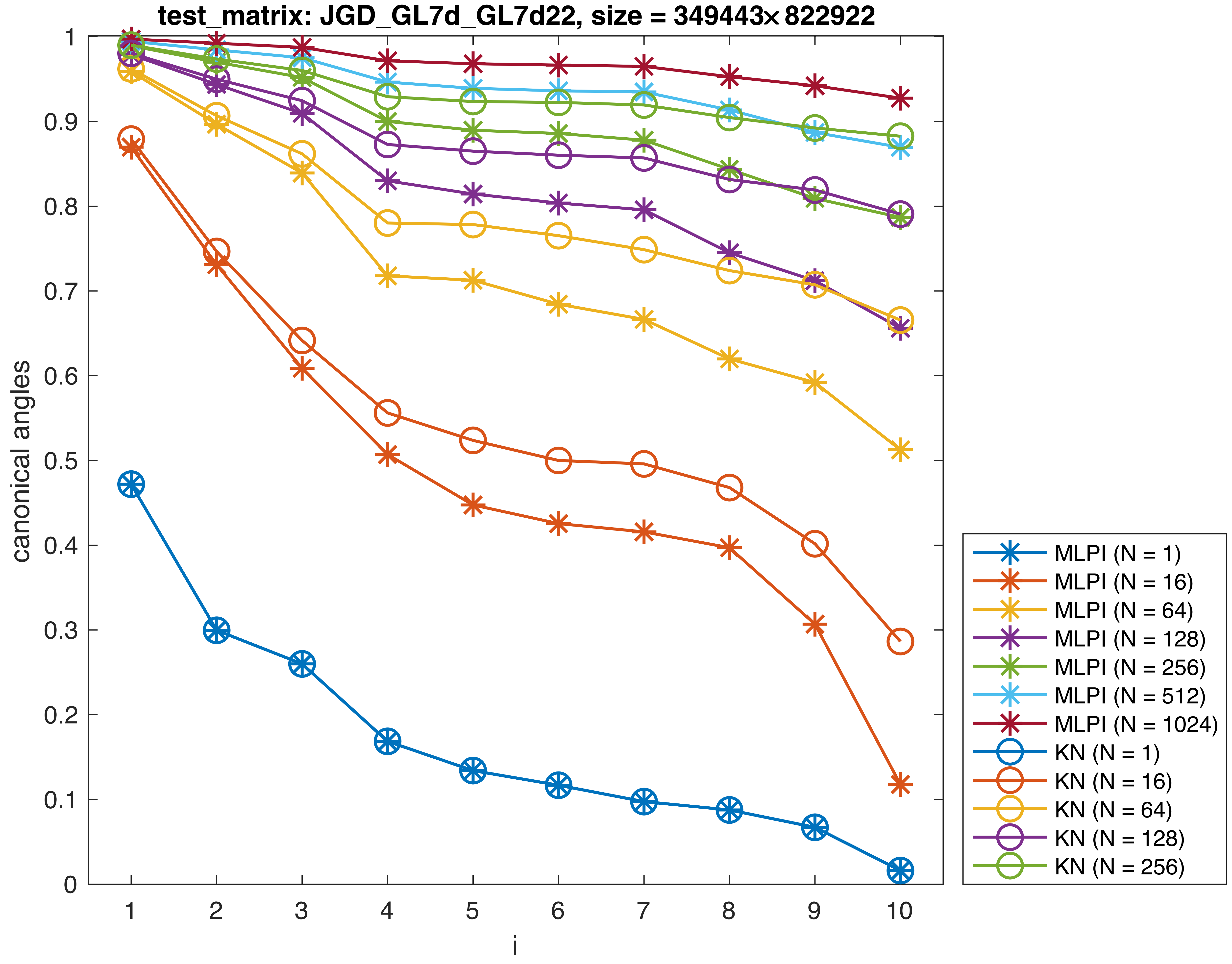
SuiteSparse Test Matrix: Barbasi_NotreDame_actors

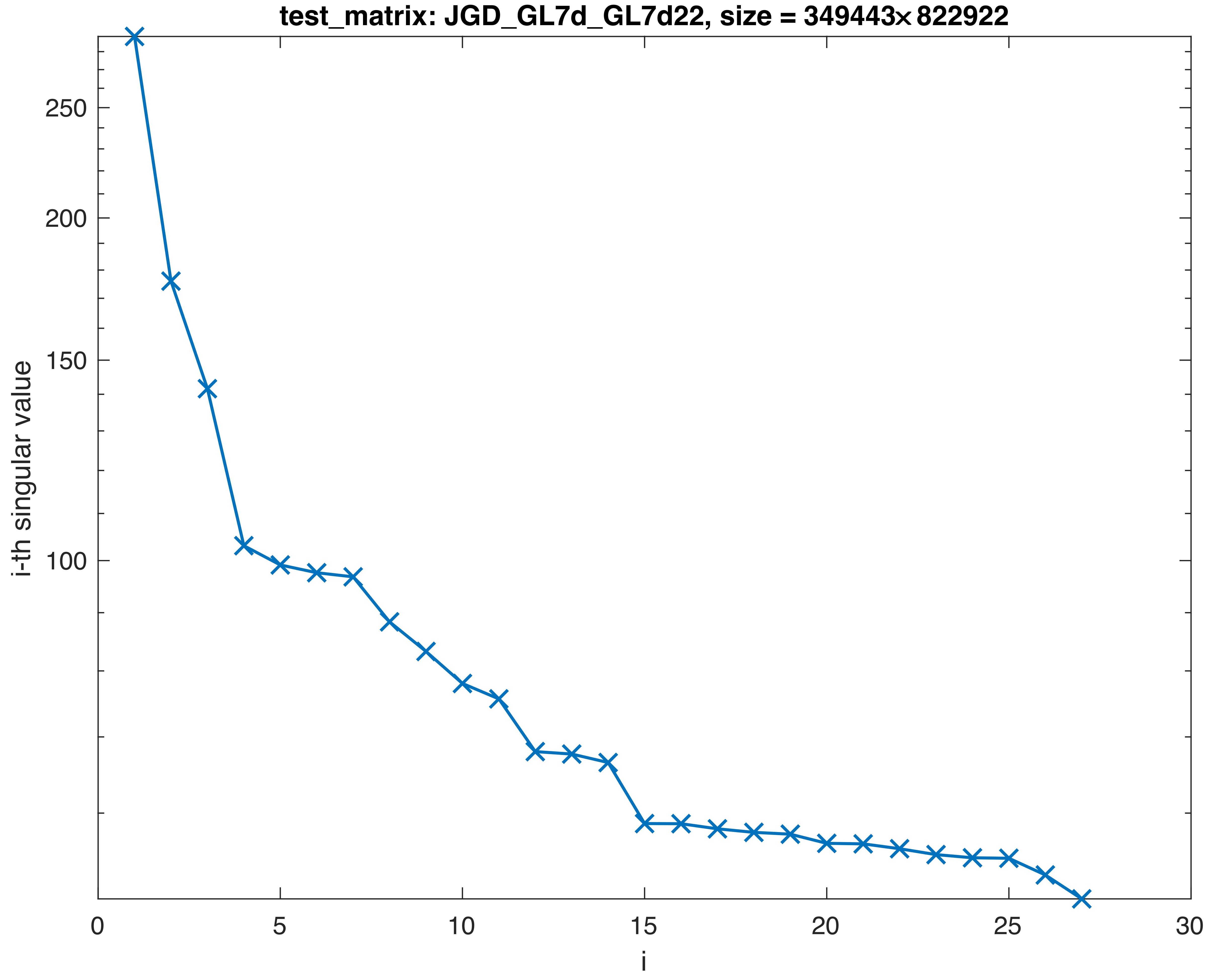




First 27 singular values of test matrix

SuiteSparse Test Matrix: JGD_GL7d_GL7d22

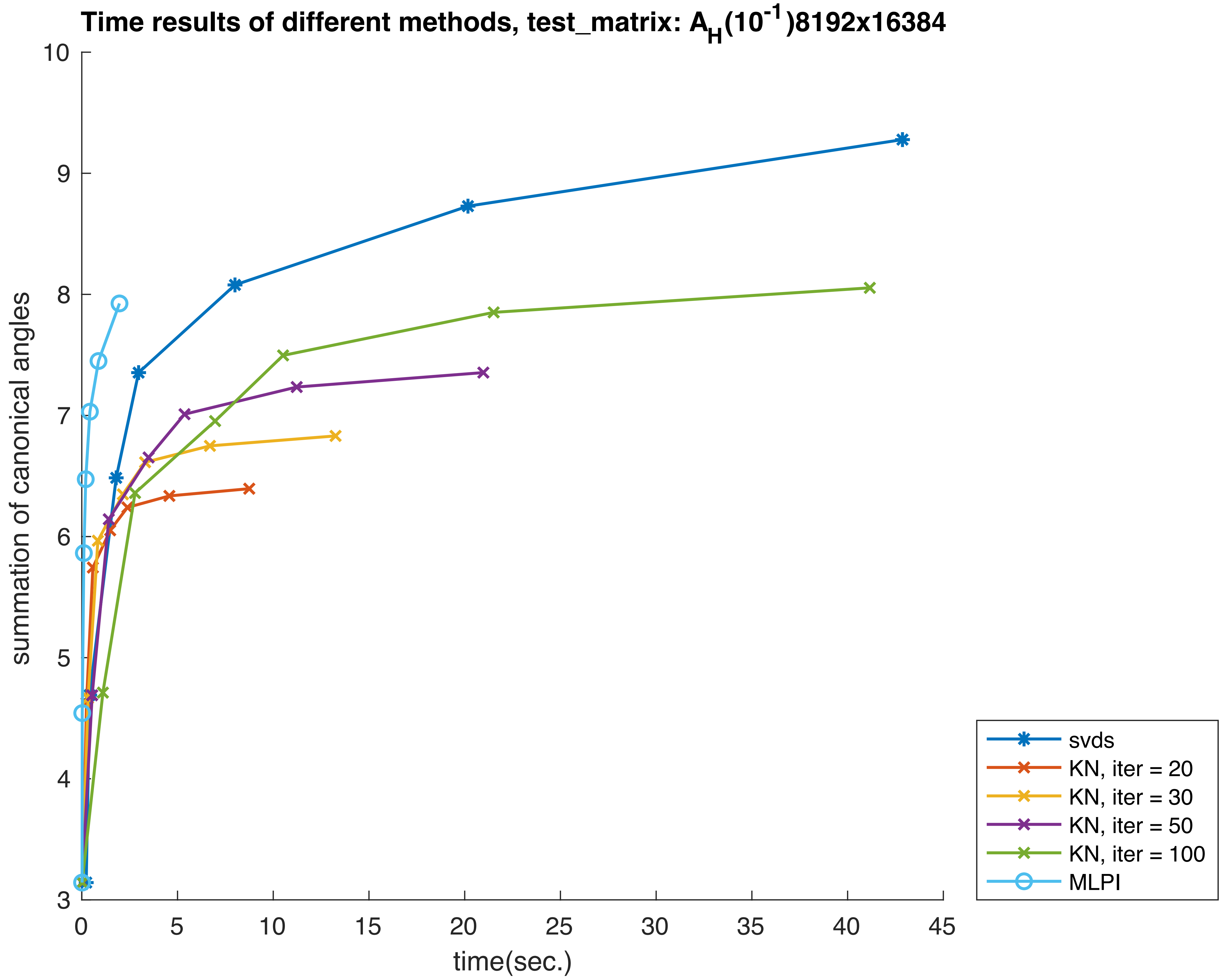




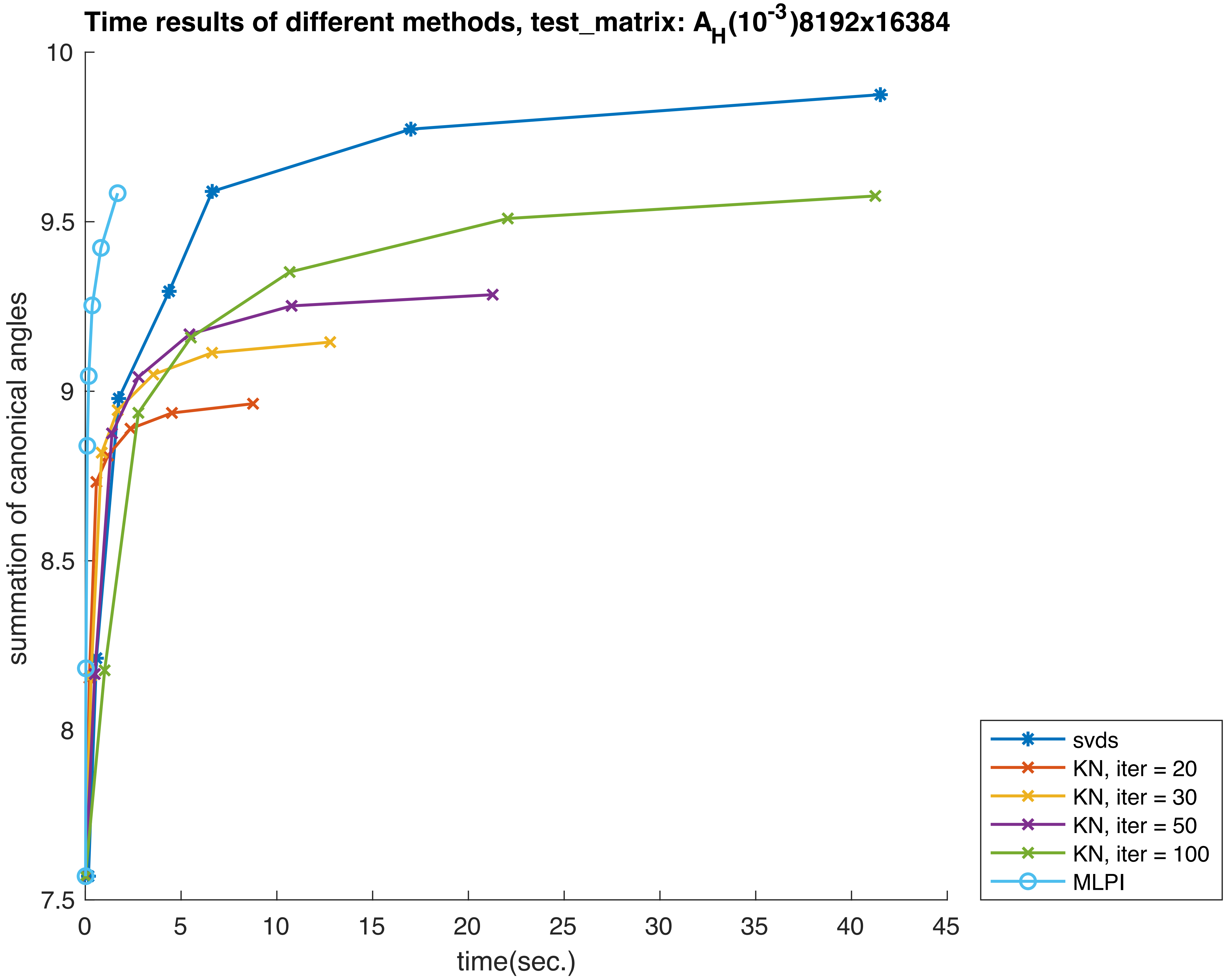
First 27 singular values of test matrix

Timing Results

- Each points represent a test case with N = 1, 4, 16, 32, 64, 128, 256



$\boldsymbol{A_H}(10^{-1})$



$\boldsymbol{A_H}(10^{-3})$

Summary

- Multiple random sketches based SVD
- Multilevel pairwise integration is a fast approximate method in iSVD
- Can be easily paralleled
- Can be used as an initial guess for KN average or WL optimization

Thank you.

Questions/comments/collaborations are welcome!