The Effectiveness of Single Vector Krylov Methods for Low-Rank Approximation

NYU Tandon School of Engineering, Christopher Musco

COLLABORATORS

Raphael Meyer
(NYU)

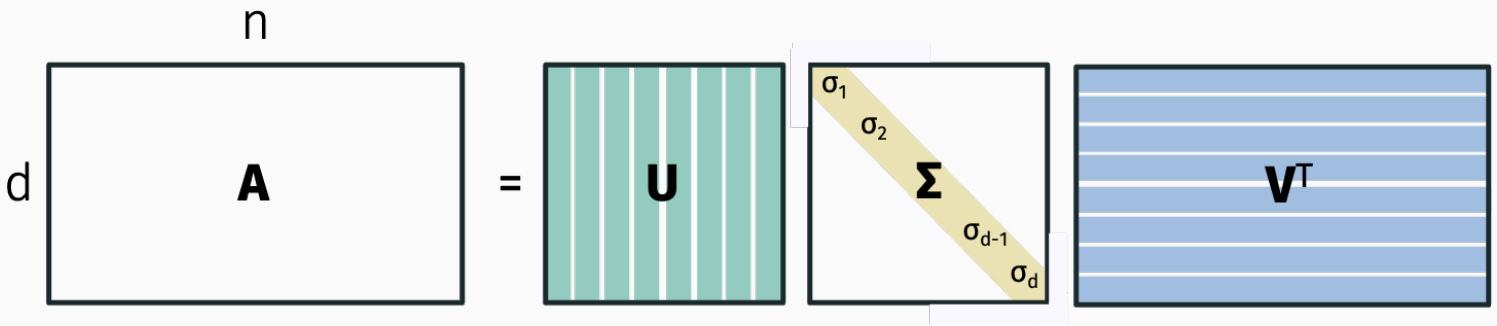
Cameron Musco
(UMass. Amherst)

Paper available at: https://arxiv.org/abs/2305.02535.

Raphael is graduating with his Ph.D. in CS in **Spring 2024** and will be looking for postdocs!

PROBLEM WE ARE STUDYING

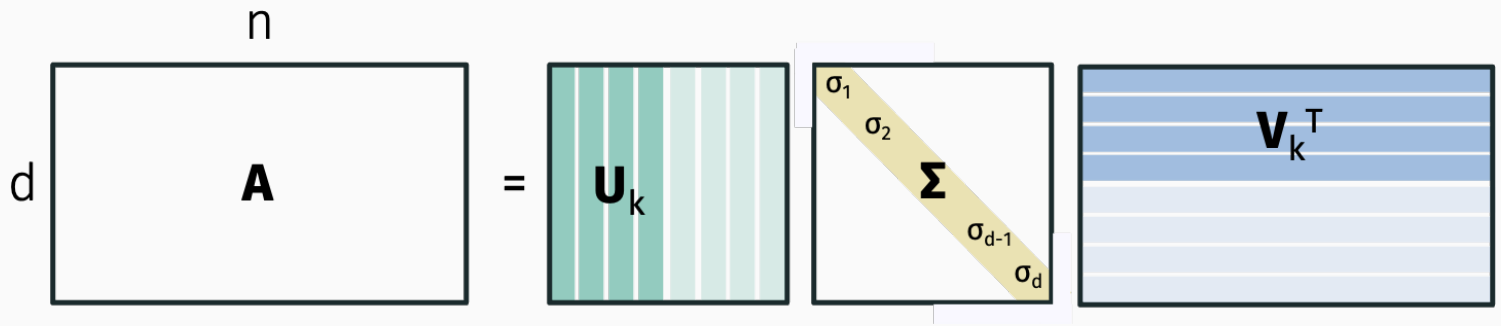
Have a matrix $\mathbf{A} \in \mathbb{R}^{d \times n}$ and want to quickly compute an optimal rank-k approximation.



Need to find the top k singular vectors of $\mathbf{A}$.

PROBLEM WE ARE STUDYING

Have a matrix $\mathbf{A} \in \mathbb{R}^{d \times n}$ and want to quickly compute an optimal rank-k approximation.



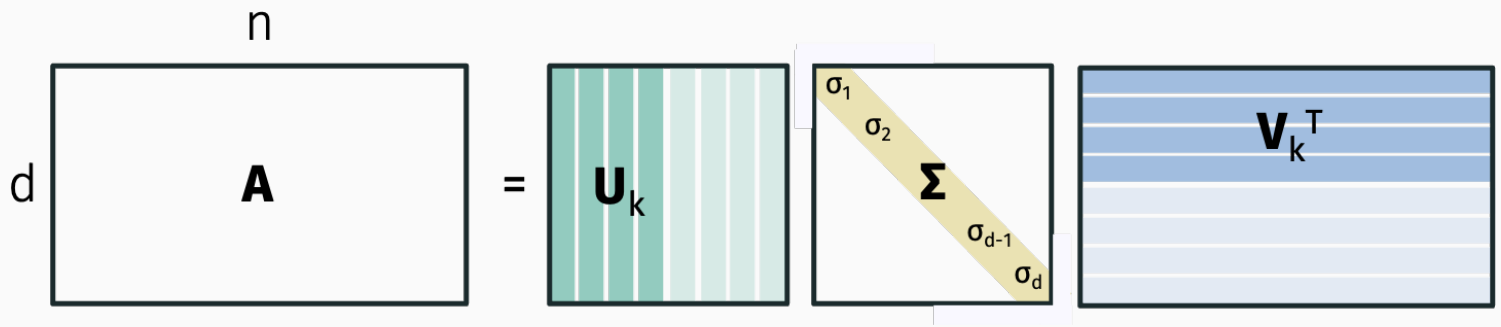
Need to find the top k singular vectors of $\mathbf{A}$. Call them $\mathbf{U}_k \in d \times k$. Then we can form the approximation:

$$\mathbf{A} \approx \mathbf{U}_k \mathbf{U}_k^T \mathbf{A} \stackrel{\text{def}}{=} \mathbf{A}_k$$

$\mathbf{A}_k$ is the optimal k-rank approximation to $\mathbf{A}$.

PROBLEM WE ARE STUDYING

Have a matrix $\mathbf{A} \in \mathbb{R}^{d \times n}$ and want to quickly compute an optimal rank-k approximation.



Need to find the top k singular vectors of $\mathbf{A}$. Call them $\mathbf{U}_k \in d \times k$. Then we can form the approximation:

$$\mathbf{A} \approx \mathbf{U}_k \mathbf{U}_k^T \mathbf{A} \stackrel{\text{def}}{=} \mathbf{A}_k$$

$\mathbf{A}_k$ is the optimal k-rank approximation to $\mathbf{A}$.

How do we find it quickly without computing a full SVD?

TYPICAL KRYLOV METHOD

Most common approach: Run a Krylov subspace method.

1. Choose random starting block $\mathsf{B} \in \mathbb{R}^{d \times b}$ ($b = 1$ or larger).
2. Compute orthonormal basis Z for the Krylov subspace:

$$\mathcal{K} = \begin{bmatrix} \mathsf{B} & \mathsf{A}\mathsf{A}^T\mathsf{B} & (\mathsf{A}\mathsf{A}^T)^2\mathsf{B} & \cdots & (\mathsf{A}\mathsf{A}^T)^t\mathsf{B} \end{bmatrix}$$

3. Return $\mathsf{Q} = \mathsf{Z}\tilde{\mathsf{U}}_k$ where $\tilde{\mathsf{U}}_k$ contains the top k eigenvectors of $\mathsf{Z}^T\mathsf{A}\mathsf{A}^T\mathsf{Z}$. This choice of Q minimizes $\|\mathsf{A} - \mathsf{Q}\mathsf{Q}^T\mathsf{A}\|_F$ amoung all rank k matrices in the span of $\mathcal{K}$.

Runtime of this method? Often dominated by the cost of multiplying vectors by A, so $O(T_{MV}(\mathsf{A}) \cdot b \cdot t)$.

TYPICAL KRYLOV METHOD

Two parameters to choose[1]**:**

- Number of iterations t.
- Size of starting block b.

Main question: For a choice of block-size b, how many matrix vector multiplications $b \cdot t$ do we need to ensure that:

$$\|A - QQ^T A\| \leq (1 + \epsilon)\|A - A_k\|?$$

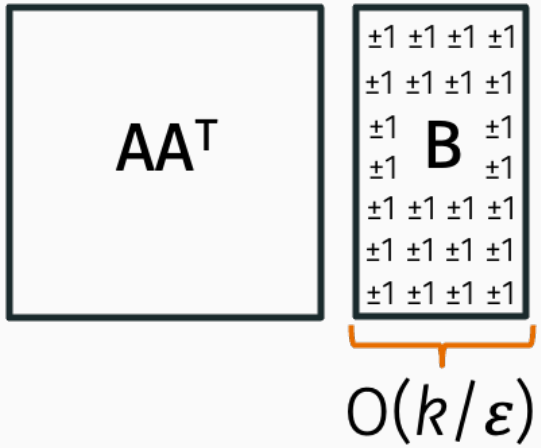
[1] Lots of other choices related to how to orthogonalize and post-process the Krylov subspace. These are not the focus of my talk, but very important!

LARGE BLOCK METHODS

Most prior theoretical work on k-rank approximation focuses on "large block" methods where $b \geq k$.

LARGE BLOCK METHODS

Most prior theoretical work on k-rank approximation focuses on "large block" methods where $b \geq k$. Two main regimes:

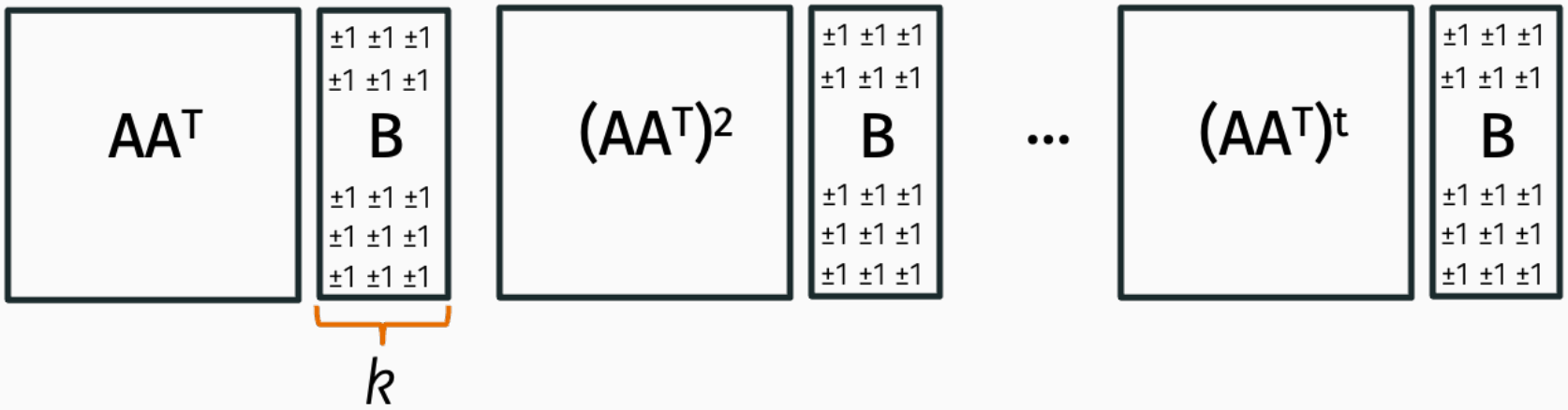


Sketching. Run for single iteration.

[Sarlós, 2006] [Martinsson, Rokhlin, Tygert, 2006] [Halko, Martinsson, Tropp 2011] [Clarkson, Woodruff 2009, 2013]

LARGE BLOCK METHODS

Most prior theoretical work on k-rank approximation focuses on "large block" methods where $b \geq k$. Two main regimes:

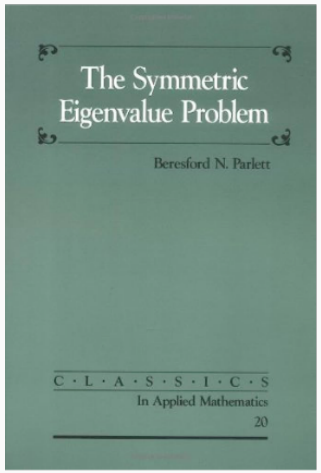


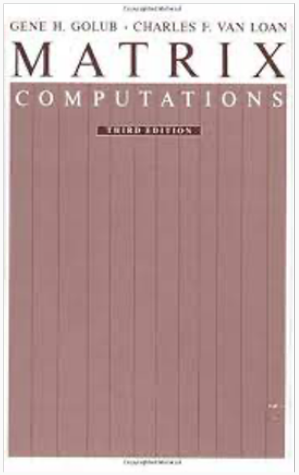
Block size k. Run for multiple iterations

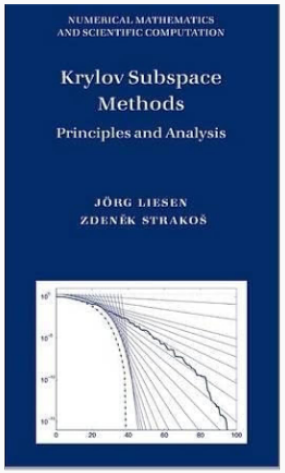
[Rokhlin, Szlam, and Tygert, 2009] [Halko, Martinsson, Tropp 2011] [Gu, 2015] [Musco, Musco 2015] [Drineas, Ipsen, Kontopoulous, Magdon-Ismail 2017]

SMALL BLOCK METHODS

Surely the case of $b = 1$ has also been studied? This is textbook single vector Krylov iteration.







Yes and no. Single vector Krylov methods have been studied extensively for the problem of eigenvector/eigenspace approximation. But this problem differs in subtle but important ways from the low-rank approximation problem.

THE VIRTUES OF LARGE BLOCK METHODS

1. **Effectively take advantage of parallelism.** Multiplying $\mathbf{A}\mathbf{A}^T$ by k vectors all at once might not be that much more expensive than multiplying by 1 vector.
2. **Do not require iteration for very large** b**.** Useful in streaming and distributed computing environments, for constrained low-rank approximation, and more.
3. **Enjoy "gap-independent" theoretical guarantees.**

THE VIRTUES OF LARGE BLOCK METHODS

1. **Effectively take advantage of parallelism.** Multiplying $\mathbf{A}\mathbf{A}^T$ by k vectors all at once might not be that much more expensive than multiplying by 1 vector.
2. **Do not require iteration for very large** b**.** Useful in streaming and distributed computing environments, for constrained low-rank approximation, and more.
3. Enjoy "gap-independent" theoretical guarantees.

EXAMPLE THEORETICAL BOUND

Theorem (Musco, Musco 2015)

When run for $O\left(\frac{\log d}{\sqrt{\epsilon}}\right)$ *iterations with a random starting block with* $b = k$ *columns, the Krylov method returns a rank* k *matrix* $\mathbf{Q}$ *satisfying:*

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T\mathbf{A}\|_{2,F} \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_{2,F}$$

$O\left(\frac{k \log d}{\sqrt{\epsilon}}\right)$ matrix-vector multiplications total.

EXAMPLE THEORETICAL BOUND

Theorem (Musco, Musco 2015)

When run for $O\left(\frac{\log d}{\sqrt{\epsilon}}\right)$ *iterations with a random starting block with* $b = k$ *columns, the Krylov method returns a rank* k *matrix* $\mathbf{Q}$ *satisfying:*

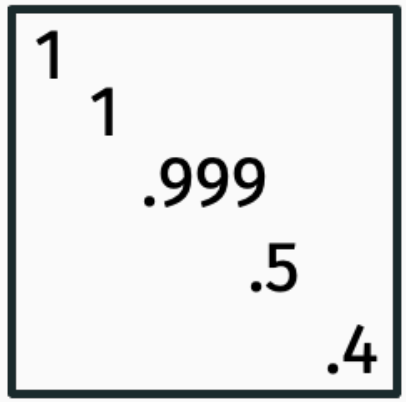
$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T\mathbf{A}\|_{2,F} \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_{2,F}$$

$O\left(\frac{k \log d}{\sqrt{\epsilon}}\right)$ matrix-vector multiplications total.

In contrast, guarantees for eigenvector approximation always depend on matrix dependent quantities like $\frac{\sigma_{i+1}}{\sigma_i - \sigma_{i+1}}$.

GAP DEPENDENT VS. GAP INDEPENDENT

The following matrix has top singular vectors $\mathbf{e}_1$ and $\mathbf{e}_2$:

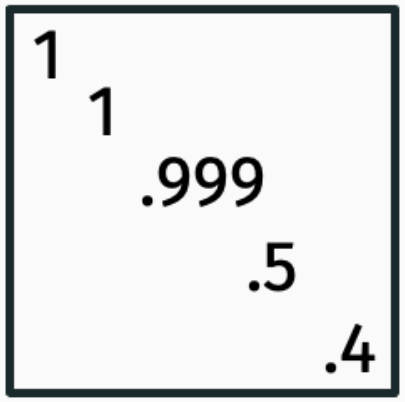


Convergence to the top subspace, $\mathbf{span}(\mathbf{e}_1, \mathbf{e}_2)$, inherently depends polynomially on the inverse gap $\frac{\sigma_3}{\sigma_2 - \sigma_3}$.

But for accurate rank 2 approximation, we can converge to any two vectors in the span of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. No gap dependence necessary.

GAP DEPENDENT VS. GAP INDEPENDENT

The following matrix has top singular vectors $\mathbf{e}_1$ and $\mathbf{e}_2$:

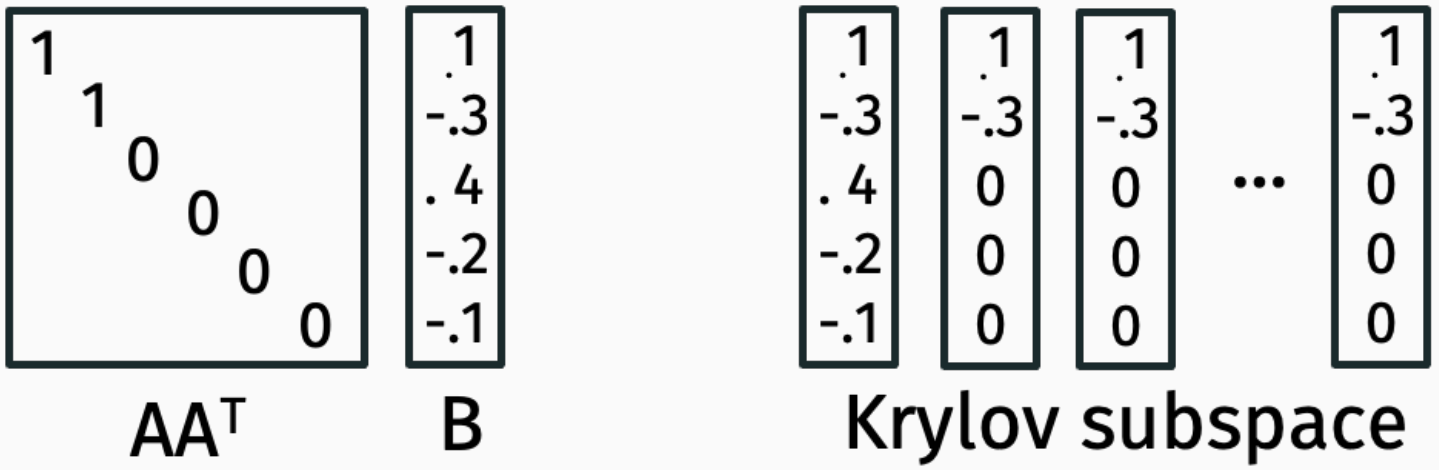


Convergence to the top subspace, $\mathbf{span}(\mathbf{e}_1, \mathbf{e}_2)$, inherently depends polynomially on the inverse gap $\frac{\sigma_3}{\sigma_2 - \sigma_3}$.

For more discussion, see e.g. Low-Rank Matrix Approximations Do Not Need a Singular Value Gap [Drineas, Ipsen, 2019].

SINGLE VECTOR METHODS AND SPECTRAL GAP DEPENDENCE

Without additional modifications, the convergence of single vector Krylov iteration must depend inversely on spectral gaps.
Even for low-rank approximation!



Krylov subspace never contains a good low-rank approximation.

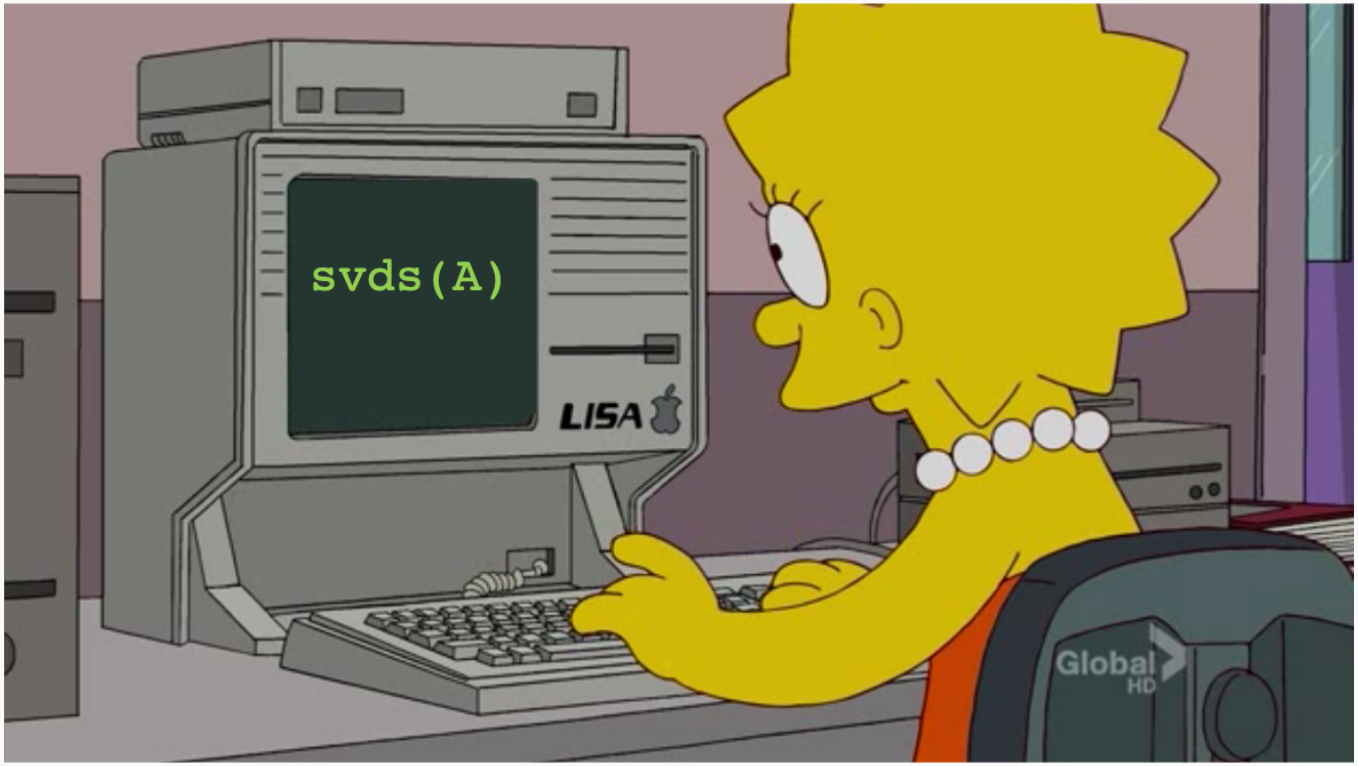
$\|A - A_2\|_F = 0$ but $\|A - QQ^T A\|_F$ is never less then 1, no matter how many iterations we take.

SINGLE VECTOR METHODS AND GAP DEPENDENCE

"Despite decades of research on [single vector] Lanczos methods, the theory for [large block] randomized methods is more complete and provides strong guarantees of excellent accuracy, whether or not there exist any gaps between the singular values." – Slzam, Kluger, Tygert, 2014

single vector Krylov methods for low-rank approximation

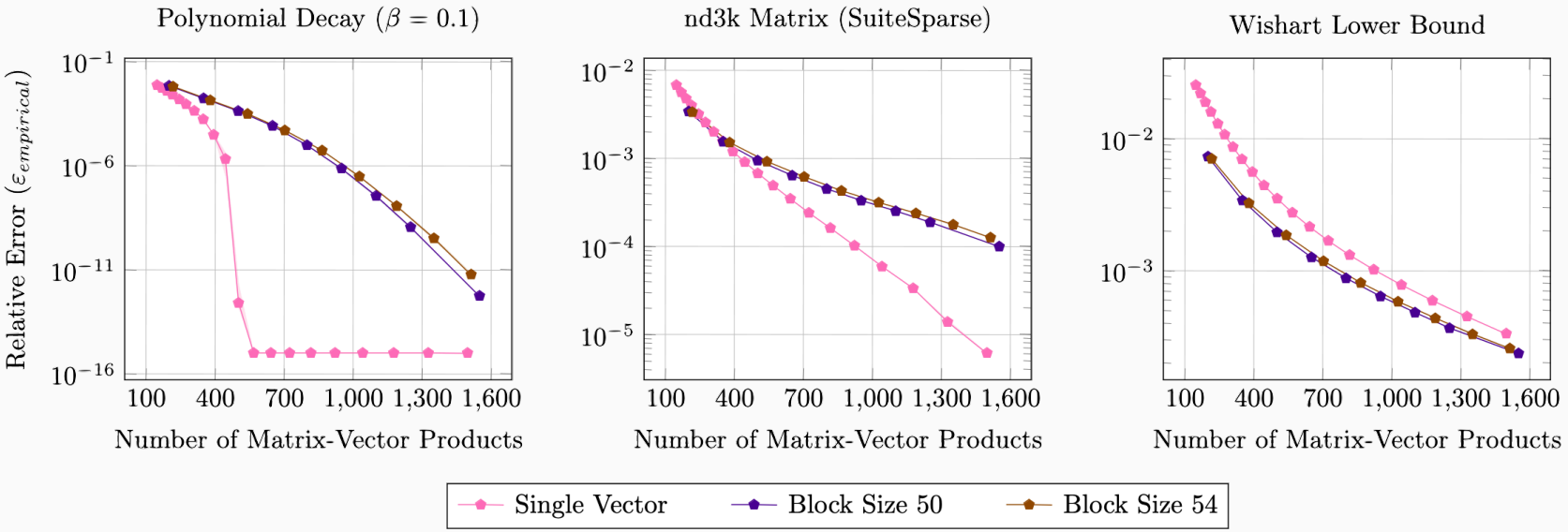
THEORY/PRACTICE GAP



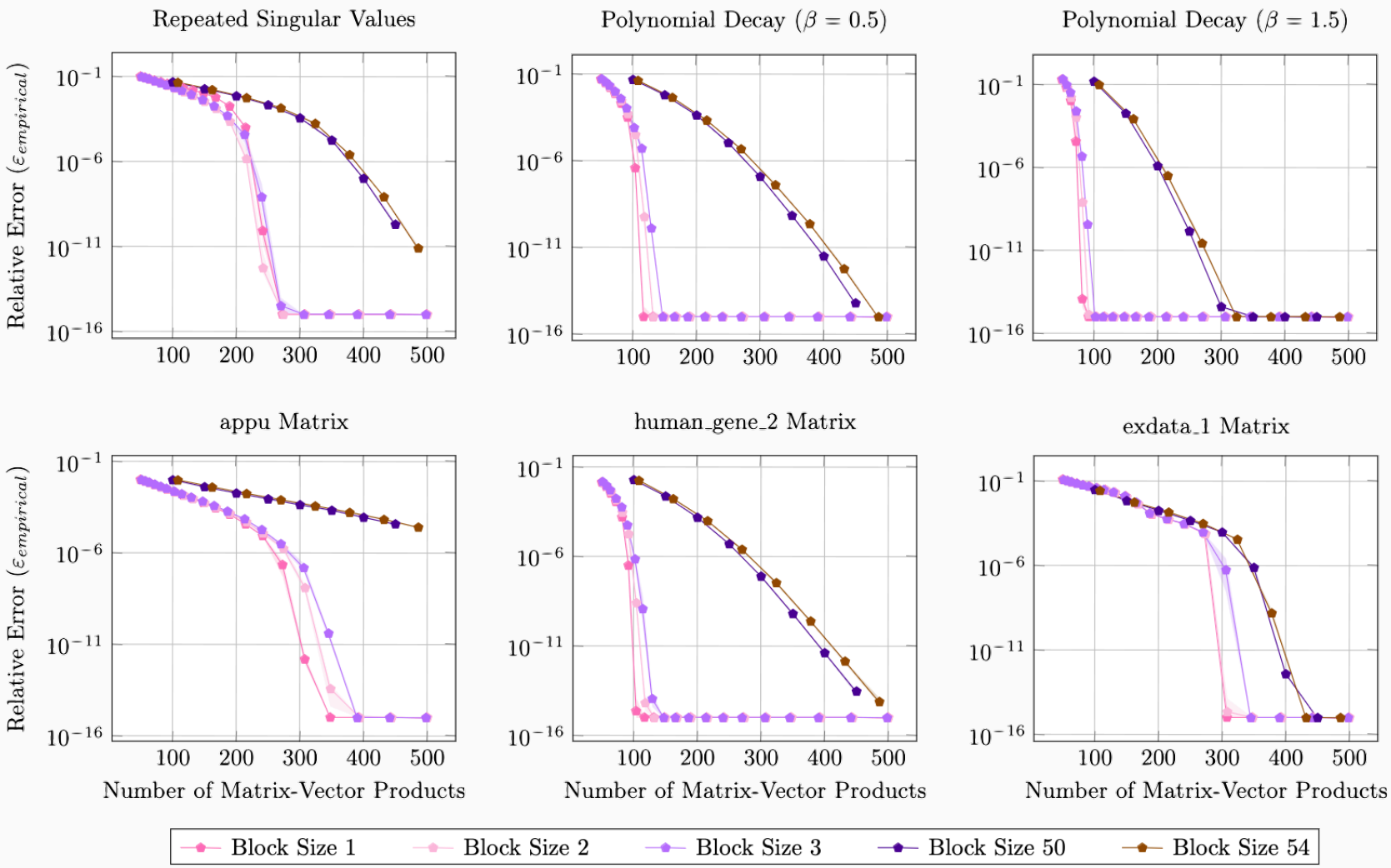
Most major numerical linear algebra libraries use a single random starting vector. And they usually work fine!

THEORY/PRACTICE GAP

If you are careful about numerical issues (e.g. build the Krylov subspace with sufficient reothogonalization) it is hard to find a problem where single vector Krylov methods lose to large block methods (in terms of number of matrix-vector products).



THEORY PRACTICE GAP



OUR MAIN RESULTS

Research question: For low-rank approximation, when and why do single vector Krylov methods require the same or fewer matrix-vector multiplications than large block methods?

OUR MAIN RESULTS

Research question: For low-rank approximation, when and why do single vector Krylov methods require the same or fewer matrix-vector multiplications than large block methods?

Answer: For low-rank approximation, single vector methods depend on gaps, but only in a very mild way!

Up to a logarithmic dependence on spectral gaps, single vector methods match (or beat) the performance of large block Krylov iteration run with any block size $b \geq k$.

EXAMPLE RESULTS

Theorem (Existing Large Block Result)

When run for $O\left(\frac{\log d}{\sqrt{\epsilon}}\right)$ *iterations with block size* $b = k$*, the Krylov method returns a rank* k *matrix* $\mathbf{Q}$ *satisfying:*

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T\mathbf{A}\|_{2,F} \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_{2,F}$$

Let $g_{min} = \min_{i=1,\ldots,k-1}\left(\frac{\sigma_i - \sigma_{i+1}}{\sigma_{i+1}}\right)$ be the minimum singular value gap in the top subspace.

Theorem (Small block result)

When run for $O\left(\frac{k\log(1/g_{min}) + \log d}{\sqrt{\epsilon}}\right)$ *iterations with block size* $b = 1$*, the Krylov method returns a rank* k *matrix* $\mathbf{Q}$ *satisfying:*

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T\mathbf{A}\|_{2,F} \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_{2,F}$$

EXAMPLE RESULTS

Block size k method: $O\left(\frac{k \log d}{\sqrt{\epsilon}}\right)$ matvecs for error ϵ.

Single vector method: $O\left(\frac{k \log(1/g_{min}) + \log d}{\sqrt{\epsilon}}\right)$ matvecs for error ϵ.

EXAMPLE RESULTS

Theorem (Existing Large Block Result)

When run for $O\left(\frac{\log(d/\epsilon)}{\sqrt{(\sigma_k-\sigma_{k+p})/\sigma_{k+p}}}\right)$ *iterations with block size* $b = k + p$*, Krylov iteration returns a rank* k $\mathbf{Q}$ *satisfying:*

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T\mathbf{A}\|_{2,F} \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_{2,F}$$

Let $g_{min} = \min_{i=1,\ldots,k+p-1}\left(\frac{\sigma_i-\sigma_{i+1}}{\sigma_{i+1}}\right)$.

Theorem (New small block result)

When run for $O\left(\frac{(k+p)\log(1/g_{min})+\log(d/\epsilon)}{\sqrt{(\sigma_k-\sigma_{k+p})/\sigma_{k+p}}}\right)$ *iterations with block size* $b = 1$*, Krylov iteration returns a rank* k $\mathbf{Q}$ *satisfying:*

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T\mathbf{A}\|_{2,F} \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_{2,F}$$

EXAMPLE RESULTS

Block size $k+p$: $O\left(\frac{(k+p)\log(d/\epsilon)}{\sqrt{(\sigma_k-\sigma_{k+p})/\sigma_{k+p}}}\right)$ matvecs for error ϵ.

Single vector: $O\left(\frac{(k+p)\log(1/g_{min})+\log(d/\epsilon)}{\sqrt{(\sigma_k-\sigma_{k+p})/\sigma_{k+p}}}\right)$ matvecs for error ϵ, where we can minimize over p.

Single-vector method obtains the "best of all worlds" without having to do any parameter selection.

EXAMPLE RESULTS

[Bakshi, Clarkson, Woodruff, 2022] shows how to compute $\mathbf{Q}$ satisfying:

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T\mathbf{A}\|_F \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_F$$

using $\tilde{O}\left(k/\epsilon^{1/3}\right)$. They run Krylov iteration with block sizes k and $k/\epsilon^{1/3}$ and take the best result in both subspaces.

Corollary of our result: Up to a logarithmic dependence on the minimum gap, we obtain the same bound with a simpler, parameter free method – simply run single vector Krylov.

PROOF APPROACH

All of these theorems can be proven using a single “gray box” reduction.

PROOF APPROACH

All of these theorems can be proven using a single “gray box” reduction. **Start with a naive observation:**

$$\text{span}\begin{bmatrix} b & AA^Tb & (AA^T)^2b & \cdot & (AA^T)^tb \end{bmatrix}$$

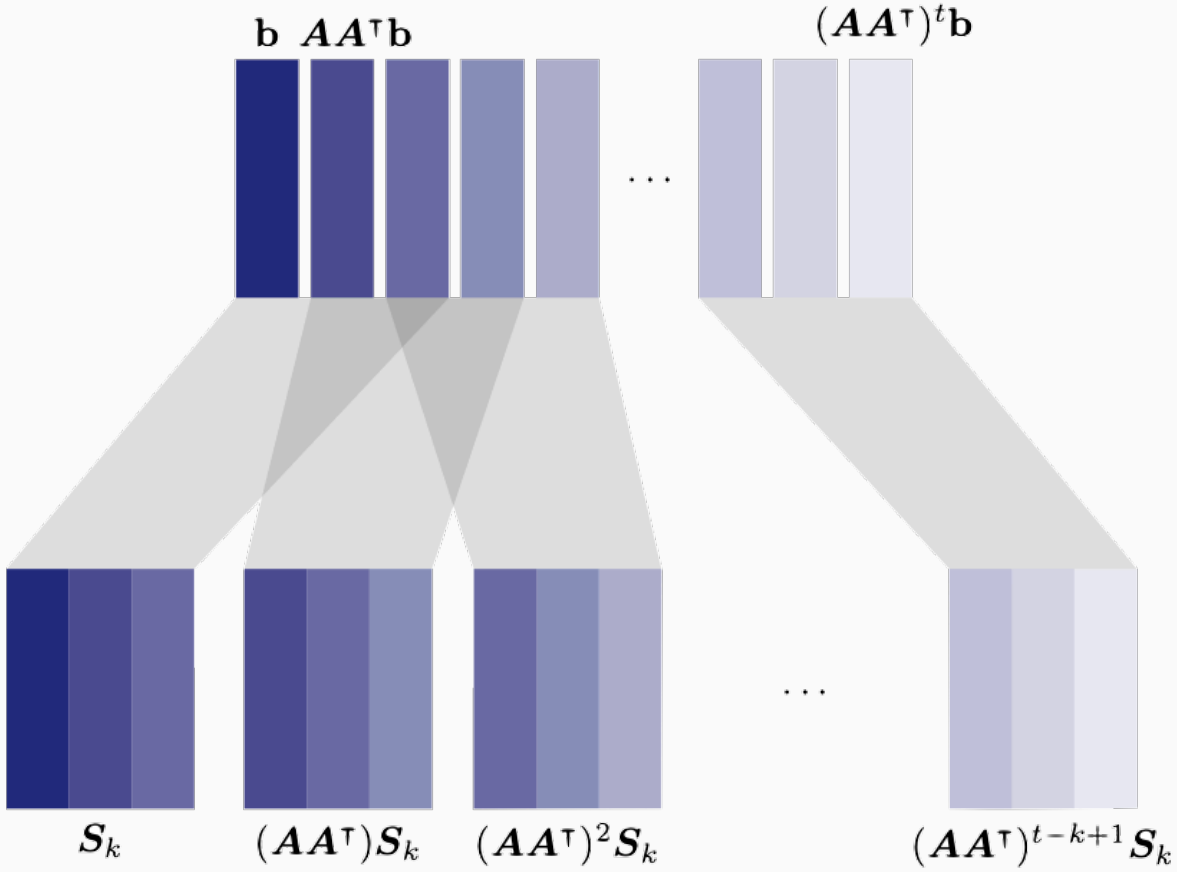
is equal to

$$\text{span}\begin{bmatrix} S_k & AA^TS_k & (AA^T)^2b & \cdot & (AA^T)^{t-k+1}S_k \end{bmatrix}$$

where $S_k = \begin{bmatrix} b & AA^Tb & \cdots & (AA^T)^{k-1}b \end{bmatrix}$ contains the first k columns of the Krylov subsoace.

PROOF APPROACH

Single vector iteration is equivalent to large block iteration with a particular choice of starting block.



IS THIS A REASONABLE IDEA?

- **Upside:** Every additional "iteration" of the block method only requires one matrix-vector multiply with $\mathbf{A}\mathbf{A}^T$. Typically would need kt matrix-vector multiplies to iterate a block of size k for t iterations. Only need $k + t$ if our starting block is $\mathbf{S}_k$.
- **Downside:** The starting block $\mathbf{S}_k$ looks very different from a random matrix. E.g. will typically be numerically low-rank.

$$\mathbf{S}_k = \begin{bmatrix} \mathbf{b} & \mathbf{A}\mathbf{A}^T\mathbf{b} & \cdots & (\mathbf{A}\mathbf{A}^T)^k\mathbf{b} \end{bmatrix}$$

This is huge upside and a huge downside.

WHAT MAKES A GOOD STARTING BLOCK?

Almost all prior analysis of block Krylov methods depend on showing that the random starting block $\mathbf{B}$ satisfies a condition similar to the following:

Definition ((k, L)-good starting matrix)

A starting matrix $\mathbf{B} \in \mathbb{R}^{d \times k}$ is a (k, L)-good starting matrix for $\mathbf{A}$ if, $\mathbf{Q} = \mathsf{span}(\mathbf{B})$ satisfies $\| \left(\mathbf{U}_k^T \mathbf{Q}\right)^{-1} \|_2^2 \leq L$.

We want L to be as small as possible. Requires that $\mathbf{Q}$ has good inner product with any vector in the span of $\mathbf{U}_k$.

Equivalent to requiring that all principal angles between subspaces $\mathbf{U}_k$ and $\mathbf{Q}$ have $\cos(\theta_i) \geq 1/\sqrt{L}$.

WHAT MAKES A GOOD STARTING BLOCK?

Almost all prior analysis of block Krylov methods depend on showing that the random starting block $\mathbf{B}$ satisfies a condition similar to the following:

Definition ((k, L)-good starting matrix)

A starting matrix $\mathbf{B} \in \mathbb{R}^{d \times k}$ is a (k, L)-good starting matrix for $\mathbf{A}$ if, $\mathbf{Q} = \mathbf{span}(\mathbf{B})$ satisfies $\| \left(\mathbf{U}_k^T \mathbf{Q}\right)^{-1} \|_2^2 \leq L$.

We want L to be as small as possible. Requires that $\mathbf{Q}$ has good inner product with any vector in the span of $\mathbf{U}_k$.

Now standard result [Rudelson, Vershynin 2010]: A random Gaussian matrix is (k, L) good with probability $1 - \delta$ for:

$$L = O(kd/\delta).$$

RESULT FOR GENERIC STARTING BLOCKS

Theorem (Musco, Musco 2015)

When run for $O\left(\frac{\log dL}{\sqrt{\epsilon}}\right)$ *iterations with a* (k, L)*-good starting block, the block Krylov method returns a rank* k *matrix* $\mathbf{Q}$ *satisfying:*

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T\mathbf{A}\|_{2,F} \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_{2,F}$$

Plugging in $L = O(kd/\delta)$ gives the desired bound for a random starting block.

Main question: Is the following matrix (k, L) good?

$$\mathsf{S}_k = \begin{bmatrix} \mathsf{b} & \mathsf{A}\mathsf{A}^T\mathsf{b} & \cdots & (\mathsf{A}\mathsf{A}^T)^k\mathsf{b} \end{bmatrix}$$

Main question: Is the following matrix (k, L) good?

$$\mathsf{S}_k = \begin{bmatrix} \mathsf{b} & \mathsf{A}\mathsf{A}^T\mathsf{b} & \cdots & (\mathsf{A}\mathsf{A}^T)^k\mathsf{b} \end{bmatrix}$$

Claim: With probability $(1-\delta)$, S_k is (k, L)-good for:

$$L = \mathsf{poly}(k, d, 1/\delta) \cdot \frac{1}{g_{min}^{4k}},$$

where $g_{min} = \mathsf{min}_{i=1,\ldots,k-1} \left(\frac{\sigma_i - \sigma_{i+1}}{\sigma_{i+1}} \right)$.

Main question: Is the following matrix (k, L) good?

$$\mathsf{S}_k = \begin{bmatrix} \mathsf{b} & \mathsf{A}\mathsf{A}^T\mathsf{b} & \cdots & (\mathsf{A}\mathsf{A}^T)^k\mathsf{b} \end{bmatrix}$$

Claim: With probability $(1 - \delta)$, S_k is (k, L)-good for:

$$L = \mathsf{poly}(k, d, 1/\delta) \cdot \frac{1}{g_{min}^{4k}},$$

where $g_{min} = \mathsf{min}_{i=1,\ldots,k-1} \left(\frac{\sigma_i - \sigma_{i+1}}{\sigma_{i+1}} \right)$.

This is exponentially worse than what we know for random matrices.

Which is great!!

Theorem (Musco, Musco 2015)

When run for $O\left(\frac{\log dL}{\sqrt{\epsilon}}\right)$ *iterations with a* (k, L)*-good starting block, the block Krylov method returns a rank* k *matrix* $\mathbf{Q}$ *satisfying:*

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T\mathbf{A}\|_{2,F} \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_{2,F}$$

Plugging in $L = \mathbf{poly}(k, d, 1/\delta) \cdot \frac{1}{g_{min}^{4k}}$ gives the desired bound for the single vector starting block $\mathbf{S}_k$ with:

$$t = O\left(\frac{k\log(1/g_{min}) + \log(d/\delta)}{\sqrt{\epsilon}}\right) \text{ iterations.}$$

Theorem (Musco, Musco 2015)

When run for $O\left(\frac{\log dL}{\sqrt{\epsilon}}\right)$ *iterations with a* (k, L)*-good starting block, the block Krylov method returns a rank* k *matrix* $\mathbf{Q}$ *satisfying:*

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T\mathbf{A}\|_{2,F} \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_{2,F}$$

Plugging in $L = \mathbf{poly}(k, d, 1/\delta) \cdot \frac{1}{g_{min}^{4k}}$ gives the desired bound for the single vector starting block $\mathbf{S}_k$ with:

$$t = O\left(\frac{k\log(1/g_{min}) + \log(d/\delta)}{\sqrt{\epsilon}}\right) \text{ iterations.}$$

And running this many iterations only requires $t + k$ matrix-vector multiplications with $\mathbf{A}\mathbf{A}^T$.

SINGLE VECTOR ANALYSIS

How to show that $\mathsf{S}_k = \begin{bmatrix} \mathsf{b} & \mathsf{A}\mathsf{A}^T\mathsf{b} & \cdots & (\mathsf{A}\mathsf{A}^T)^k\mathsf{b} \end{bmatrix}$ is (k, L)-good?

Reduce to a problem about polynomials, by taking advantage of the fact that:

$$\|(\mathsf{U}_k^T\mathsf{Q})^{-1}\|_2^2 = \|\mathsf{S}_k(\mathsf{U}_k^T\mathsf{S}_k)^{-1}\|_2^2 \max_{\mathsf{x}} \frac{\|\mathsf{S}_k\mathsf{x}\|_2^2}{\|\mathsf{U}_k^T\mathsf{S}_k\mathsf{x}\|_2^2}.$$

SINGLE VECTOR ANALYSIS

How to show that $\mathsf{S}_k = \begin{bmatrix} \mathsf{b} & \mathsf{A}\mathsf{A}^T\mathsf{b} & \cdots & (\mathsf{A}\mathsf{A}^T)^k\mathsf{b} \end{bmatrix}$ is (k, L)-good?

Reduce to a problem about polynomials, by taking advantage of the fact that:

$$\|(\mathsf{U}_k^T\mathsf{Q})^{-1}\|_2^2 = \|\mathsf{S}_k(\mathsf{U}_k^T\mathsf{S}_k)^{-1}\|_2^2 \max_{\mathsf{x}} \frac{\|\mathsf{S}_k\mathsf{x}\|_2^2}{\|\mathsf{U}_k^T\mathsf{S}_k\mathsf{x}\|_2^2}.$$

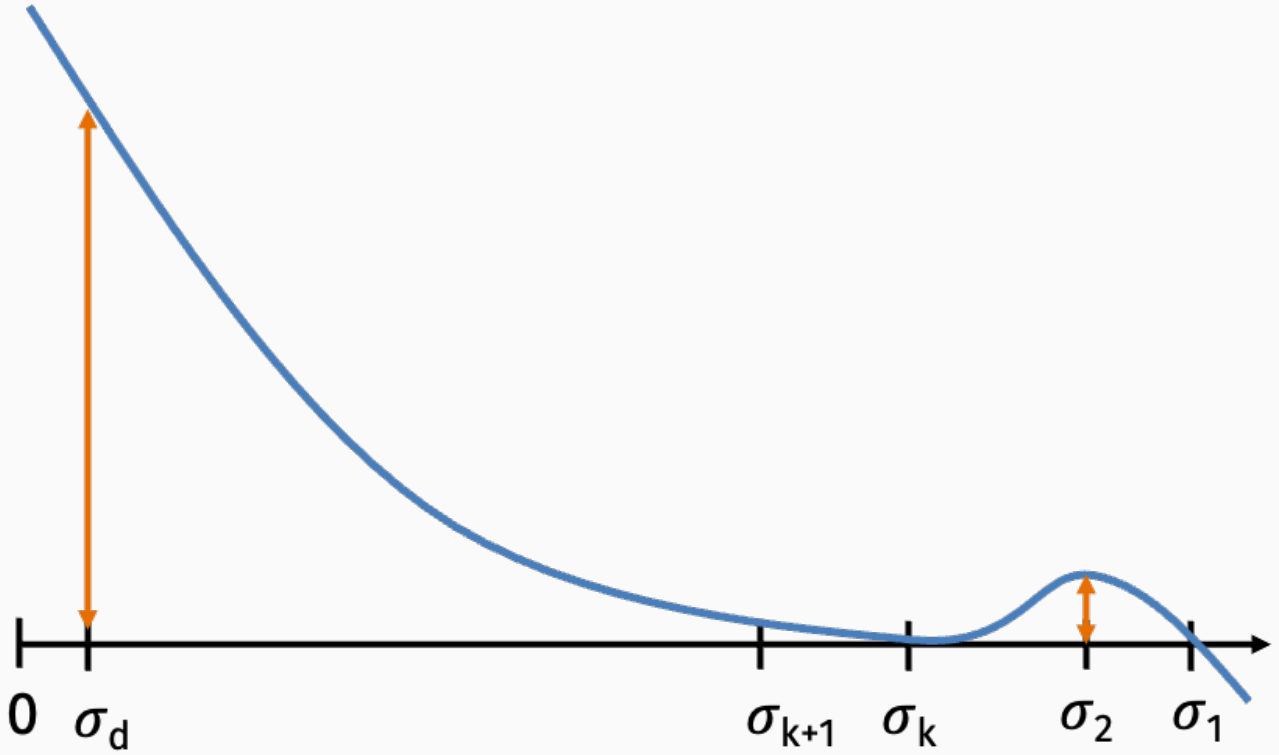
After applying some Gaussian anti-concentration, need to bound:

$$\max_{\text{degree } k-1 \text{ poly } p} \frac{\max_{i\in 1,\ldots,d} p(\sigma_i^2)}{\max_{i\in 1,\ldots,k} p(\sigma_i^2)}$$

SINGLE VECTOR ANALYSIS

Need to bound:

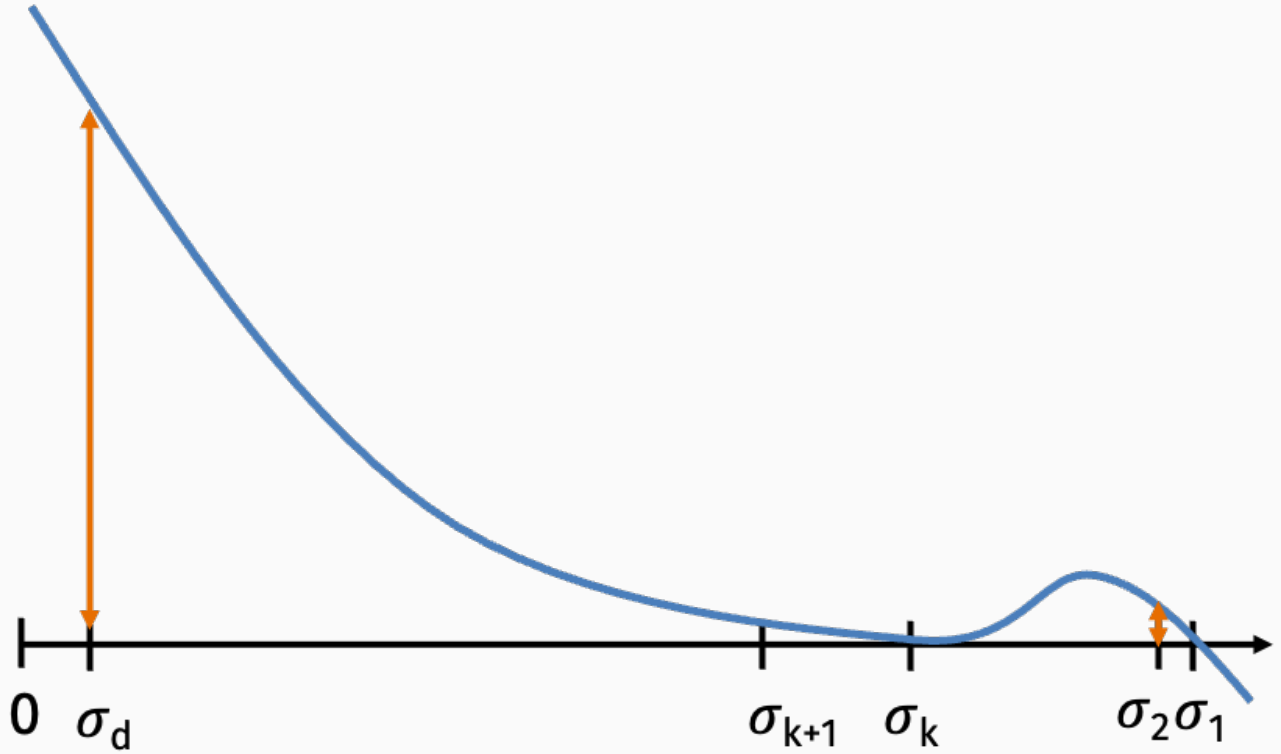
$$\max_{\text{degree } k-1 \text{ poly } p} \frac{\max_{i \in k+1,\ldots,d} p(\sigma_i^2)}{\max_{i \in 1,\ldots,k} p(\sigma_i^2)}$$



SINGLE VECTOR ANALYSIS

Need to bound:

$$\max_{\text{degree } k-1 \text{ poly } p} \frac{\max_{i \in k+1,\ldots,d} p(\sigma_i^2)}{\max_{i \in 1,\ldots,k} p(\sigma_i^2)}$$



Obtain a bound of $O(1/g_{min}^{4k})$ following an approach from [Saad, 1980].

CONCLUSION

Punchline: For low-rank approximation, single vector Krylov iteration matches large block methods (in terms of matrix-vector products) up to a logarithmic factor.

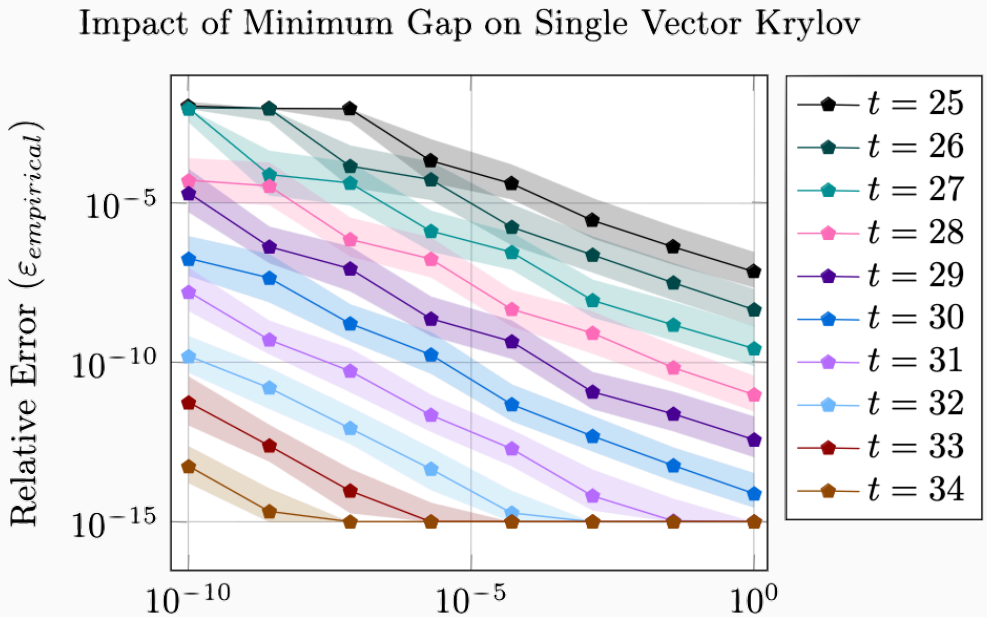
Theorem

When run for $O\left(\frac{k\log(1/g_{min})+\log d}{\sqrt{\epsilon}}\right)$ *iterations with block size* $b = 1$*, the Krylov method returns a rank* k *matrix* $\mathbf{Q}$ *satisfying:*

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T\mathbf{A}\|_{2,F} \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_{2,F}$$

Other results in the paper:

- Bounds for block size $1 < b < k$. Depend on minimum b^{th}-order gap.
- By adding a small random perturbation to $\mathbf{A}$, we can replace the $\log(1/g_{min})$ with a $\log(d)$ without hurting accuracy.
- More experiments.



CONCLUSION

Future work:

- Better understanding finite precision. Can we obtain similar results with partial orthogonalization?
- Improving bounds for block size $1 < b < k$. For $b = c \cdot k$, we require $O(k^2)$ matrix-vector multiplies. Not sure if tight.
- Are our current low-rank approximation algorithms optimal in terms of matrix-vector products? [Bakshi, Narayanan 2023].

questions?