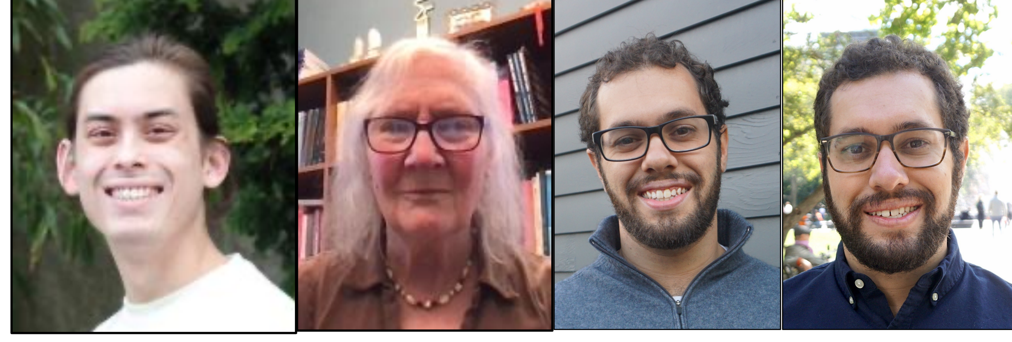
Noah Amsel, Tyler Chen, Anne Greenbaum[†]
Cameron Musco[‡], Christopher Musco
[†]University of Washington [‡]University of Massachusetts, Amherst



Goal: Approximate $f(\mathbf{A})\mathbf{b}$ using a few matvecs with $\mathbf{A}$

Set up

- d = ambient dimension
- k = num iterations = dim of Krylov subspace
- $\mathbf{A} = \mathbf{A}^\top \in \mathbb{R}^{d\times d}$ and $\mathbf{b} \in \mathbb{R}^d$ define the problem instance
- $\Lambda \in [\lambda_{\min}, \lambda_{\max}]$ are the eigenvalues of $\mathbf{A}$
- $\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \operatorname{span}\{\mathbf{b}, \mathbf{Ab}, \dots, \mathbf{A}^{k-1}\mathbf{b}\}$ is the Krylov subspace
- $f: \Lambda \to \mathbb{R}$ is a function. Often $f(z) = z^\alpha$, $\exp(tz)$, $\log(z)$, or $\operatorname{sgn}(z)$
- $r: \Lambda \to \mathbb{R}$ is a rational function of degree q with poles $z_1, \dots, z_q \notin [\lambda_{\min}, \lambda_{\max}]$
- lan_k is the output of the algorithm at k iterations and everything is in exact arithmetic

Lanczos-FA for Matrix Functions

1. $\mathbf{Q}$ is an orthonormal basis for the Krylov subspace
2. Approximate $\mathbf{A}$ by projecting it into the Krylov subspace: $\mathbf{A} \approx \mathbf{Q}\mathbf{Q}^\top\mathbf{A}\mathbf{Q}\mathbf{Q}^\top =: \mathbf{Q}\mathbf{T}\mathbf{Q}^\top$
   - $\mathbf{T}$ is $k \times k$ symmetric tridiagonal
3. $f(\mathbf{A})\mathbf{b} = \mathbf{Q}f(\mathbf{T})\mathbf{Q}^\top\mathbf{b} = p(\mathbf{A})\mathbf{b}$ for some degree $k-1$ polynomial
   - Compute $f(T)$ in $O(k^2)$ by eigendecomposition

Standard Error Bound for Lanczos-FA

"It converges as fast as polynomial approximation converges to f on the interval, measured in max norm" (i.e. uniform approximation)

$$\|f(\mathbf{A})\mathbf{b} - \mathsf{lan}_k\|_2 / \|\mathbf{b}\|_2 \le 2 \min_{\deg(p)<k} \left(\max_{x \in [\lambda_{\min}, \lambda_{\max}]} |f(x) - p(x)| \right)$$

- This bound only depends on $f, k, \lambda_{\min}, \lambda_{max}$
- Given those, it shows how Lanczos-FA does on the hardest choice of $\mathbf{A}, \mathbf{b}$
- Lanczos-FA is "Krylov-optimal" because no Krylov method can converge faster for worst-case $\mathbf{A}, \mathbf{b}$
- The optimality coefficient 2 is tight

But empirically, Lanczos-FA converges much faster than this bound for typical $\mathbf{A}, \mathbf{b}$ (see figure, center), which begs the question…

Our Question:
Is Lanczos-FA "instance optimal"?

- It seems to converge nearly as fast as the best polynomial approximation $p^*(\mathbf{A})\mathbf{b}$ does for *this* instance $\mathbf{A}, \mathbf{b}$

Prior Work: $\mathbf{A}^{-1}\mathbf{b}$ and $\exp(\mathbf{A})\mathbf{b}$

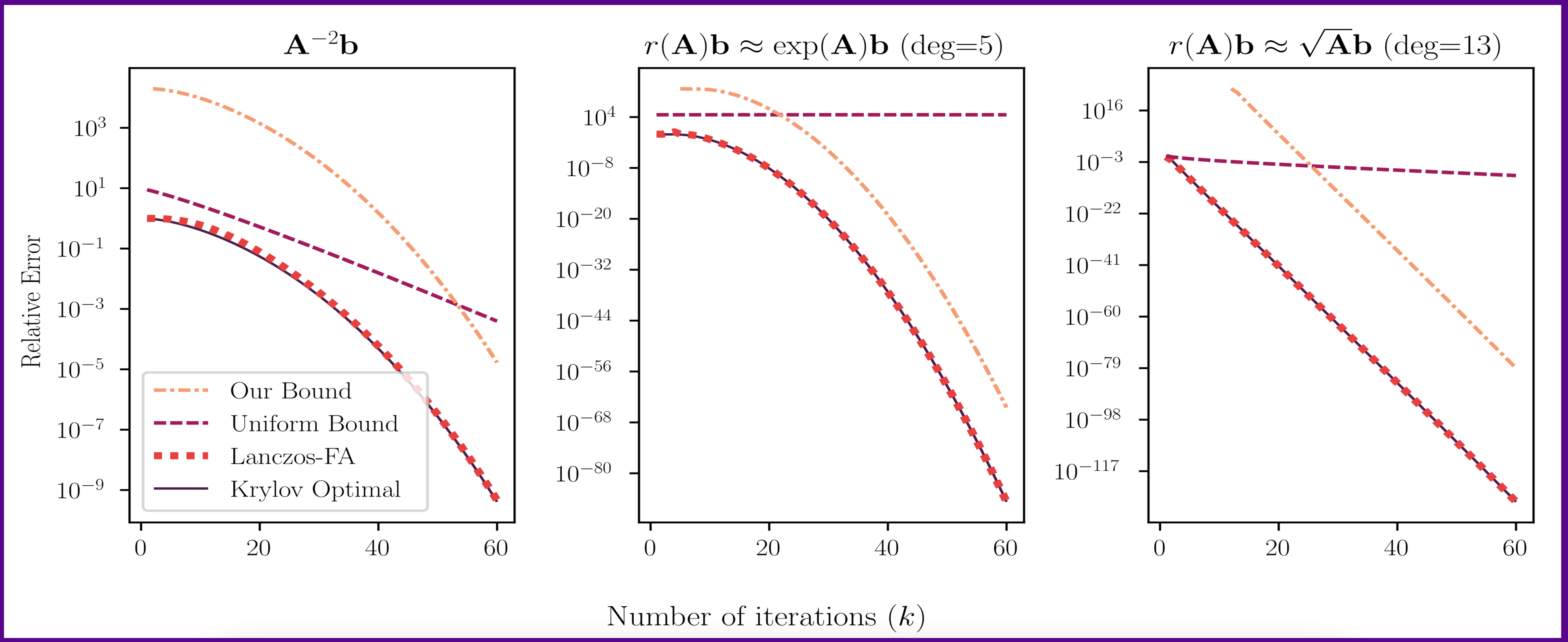
- For $A \succcurlyeq 0$, Lanczos-FA on $\mathbf{A}^{-1}\mathbf{b}$ is identical to Conjugate Gradients in exact arithmetic
- It's nearly instance optimal:

$$\|\mathbf{A}^{-1}\mathbf{b} - \mathsf{lan}_k\|_2 \le \sqrt{\kappa(\mathbf{A})} \min_{\deg(p)<k} \left\|\mathbf{A}^{-1}b - p(\mathbf{A})\mathbf{b}\right\|_2$$

- Much better! Super-exponential instead of exponential [5]
- For indefinite systems, doesn't converge monotonically but the best iteration is almost good as MINRES [3]
- For the exponential, we something similar to instance optimality but for the maximization[4]:

$$\|\exp(-t\mathbf{A})\mathbf{b} - \mathsf{lan}_k\| \le 3\|\mathbf{A}\|^2 t^2 \max_{0\le s\le t} \left(\min_{\deg(p)<k-2} \|\exp(-s\mathbf{A})\mathbf{b} - p(\mathbf{A})\mathbf{b}\| \right).$$

For rational functions r, the Lanczos method gives a nearly* optimal approximation to $r(\mathbf{A})\mathbf{b}$ from the Krylov subspace $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$



Our Bound:

"It converges as fast as if it gave the best approximation from the Krylov subspace, up to a prefactor and iteration delay depending on the degree of r"

$$\|r(\mathbf{A})\mathbf{b} - \mathsf{lan}_k\|_2 \le q \cdot \kappa(A)^q \cdot \min_{\deg(p)<k-q+1} \|r(\mathbf{A})\mathbf{b} - p(\mathbf{A})\mathbf{b}\|_2$$

(Error of k-step Lanczos-FA) (Error of best degree $\sim k$ polynomial approx)

- The standard bound depends only on $r, k, \lambda_{\min}, \lambda_{\max}$
- Like Lanczos-FA itself [1], our bound adapts to the specific $\mathbf{A}$ and $\mathbf{b}$ at hand so it captures the true shape of the convergence curve
- Above is for psd $\mathbf{A}$, roots of r are ≤ 0. In general, prefactor is $\prod_{j=1}^{q} \kappa(\pm(A - z_j I))$

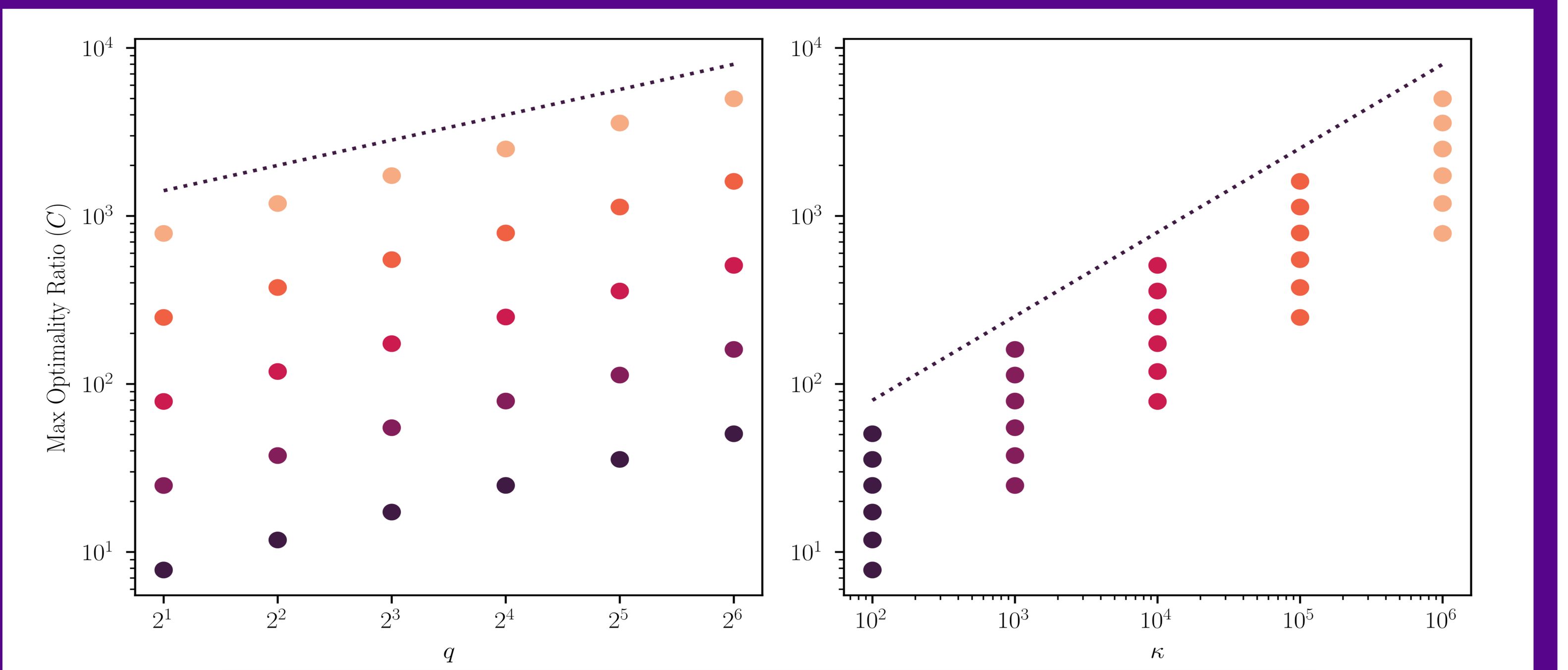
*Our Conjecture:

We can improve the prefactor to

$$O(\sqrt{q \cdot \kappa(\mathbf{A})})$$

- We've only found one family of problems that seems to need

$$\Omega(\sqrt{q \cdot \kappa(\mathbf{A})})$$



Applications to non-rational functions

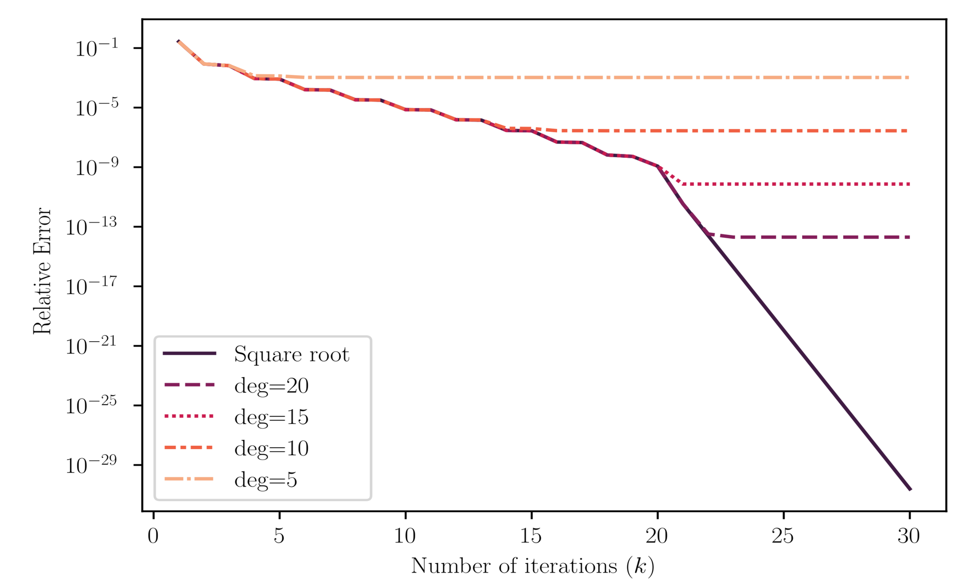
If f is uniformly well approximated by rational functions, then Lanczos-FA inherits optimality!

If Lanczos-FA is instance optimal for r with coeff C_r,

$$\|f(\mathbf{A})\mathbf{b} - \mathsf{lan}_k\|_2 \le \min_r \Bigg[(C_r + 2)\|\mathbf{b}\|_2 \max_{x\in[\lambda_{\min},\lambda_{\max}]} |r(x) - f(x)| + C_r \min_{\deg(p)<k-c_r} \|f(\mathbf{A})\mathbf{b} - p(\mathbf{A})\mathbf{b}\|_2 \Bigg]$$

- The extra (first) term captures the error of the rational uniform approximation
- Much stronger than the standard bound, which requires uniform approximation by *polynomials*
- No need to actually construct the approximant r. Run Lanczos-FA on f, and the bound picks an r that balances C_r against the approximation error

Lanczos-FA for (rational approximations of) square root:



Bonus: Pseudo-optimality for $\mathbf{A}^{-1/2}\mathbf{b}$

- "Pseudo-optimality" means the bound is specific to the spectrum of $\mathbf{A}$, but still only captures the hardest choice of $\mathbf{b}$
  - Better than standard bound, which depends only on $[\lambda_{\min}, \lambda_{\max}]$
- Using a different technique, we derive the following for inverse roots:

$$\|\mathbf{A}^{-1/2}\mathbf{b} - \mathsf{lan}_k\| \le \frac{3}{\sqrt{\pi k}}\sqrt{\kappa(\mathbf{A})} \cdot \min_{\deg(p)\le k/2} \max_{x\in\Lambda} \left| \frac{1}{\sqrt{x}} - p(x) \right| \le \frac{3}{\sqrt{\pi k}}\sqrt{\kappa(\mathbf{A})} \cdot \frac{1}{\min_j \langle \mathbf{v}_j, \mathbf{b}\rangle} \|p(\mathbf{A})\mathbf{b} - \mathbf{A}^{-1/2}\mathbf{b}\|$$

where v_j is the j^{th} eigenvector of $\mathbf{A}$

- When $\mathbf{b}$ is orthogonal to λ_j, you can get low error without approximating $1/\sqrt{\lambda_j}$ well, so our bound is loose

Extensions

1. Poles in the interval of the eigenvalues (cf. indefinite linear systems)
2. Finite precision arithmetic (already studied for exponential)
3. Optimality with respect to other norms that may be more natural (cf. Lanczos-OR [2])

References

1. Erin Carson, Jörg Liesen, and Zdenek Strakoš. 70 years of Krylov subspace methods: The journey continues, 2022.
2. Tyler Chen, Anne Greenbaum, Cameron Musco, and Christopher Musco. Low- memory Krylov subspace methods for optimal rational matrix function approxima- tion. 2023.
3. Jane Cullum and Anne Greenbaum. Relations between Galerkin and norm-minimizing iterative methods for solving linear systems. SIAM Journal on Matrix Analysis and Applications, 17(2):223–247, 1996.
4. Vladimir Druskin, Anne Greenbaum, and Leonid Knizhnerman. Using nonorthog- onal Lanczos vectors in the computation of matrix functions. SIAM Journal on Scientific Computing, 19(1):38–54, 1998.
5. Sadegh Jokar and Bahman Mehri. The best approximation of some rational func- tions in uniform norm. Applied Numerical Mathematics, 55(2):204–214, October 2005.