NYU

Near-Optimality Guarantees for Approximating Matrix Functions by the Lanczos Method

Noah Amsel, Tyler Chen, Anne Greenbaum†, Cameron Musco‡, Christopher Musco

†University of Washington ‡University of Massachusetts, Amherst

Problem

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

Setup

- $\mathbf{A} = \mathbf{A}^\top \in \mathbb{R}^{d\times d}$ and $\mathbf{b} \in \mathbb{R}^d$ are the problem instance
- $\Lambda \subset [\lambda_{\min}, \lambda_{\max}]$ are the eigenvalues of $\mathbf{A}$
- $\mathscr{K}_k(\mathbf{A},\mathbf{b}) = \operatorname{span}\{\mathbf{b}, \mathbf{A}\mathbf{b}, \ldots, \mathbf{A}^{k-1}\mathbf{b}\}$ is the kth Krylov subspace
- $f : \Lambda \to \mathbb{R}$ is a function, like $f(z) = 1/z,\ \sqrt{z},\ \exp(tz),$ or $\operatorname{sign}(z)$

Lanczos Method

1. Let $\mathbf{Q}$ be an orthonormal basis for the kth Krylov subspace
2. Approximate $\mathbf{A}$ by projecting it into the Krylov subspace:
   $\mathbf{A} \approx \mathbf{Q}\mathbf{Q}^\top\mathbf{A}\mathbf{Q}\mathbf{Q}^\top =: \mathbf{Q}\mathbf{T}\mathbf{Q}^\top$
   - where $\mathbf{T} := \mathbf{Q}^\top\mathbf{A}\mathbf{Q}$ is $k \times k$ tridiagonal
3. Output $\mathsf{lan}_k := \mathbf{Q} f(\mathbf{T})\mathbf{Q}^\top\mathbf{b} \approx f(\mathbf{A})\mathbf{b}$
   - Can compute $f(\mathbf{T})$ in $O(k^2)$ time by eigendecomposition
   - Fact: $\mathsf{lan}_k = p(\mathbf{A})\mathbf{b}$ for some degree $k-1$ polynomial p

Standard Analysis of Lanczos Method

Lanczos finds a degree $k-1$ approximation to f that is nearly optimal on the range of $\mathbf{A}$'s eigenvalues:

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

- Exponential convergence for smooth f
- We prove: for any f and Λ, this is tight for *some* $\mathbf{A}$ and $\mathbf{b}$
  - But it's loose for typical $\mathbf{A}$ and $\mathbf{b}$
- Weakness: we should not need to approximate f on all of $[\lambda_{\min}, \lambda_{\max}]$, just at the eigenvalues Λ

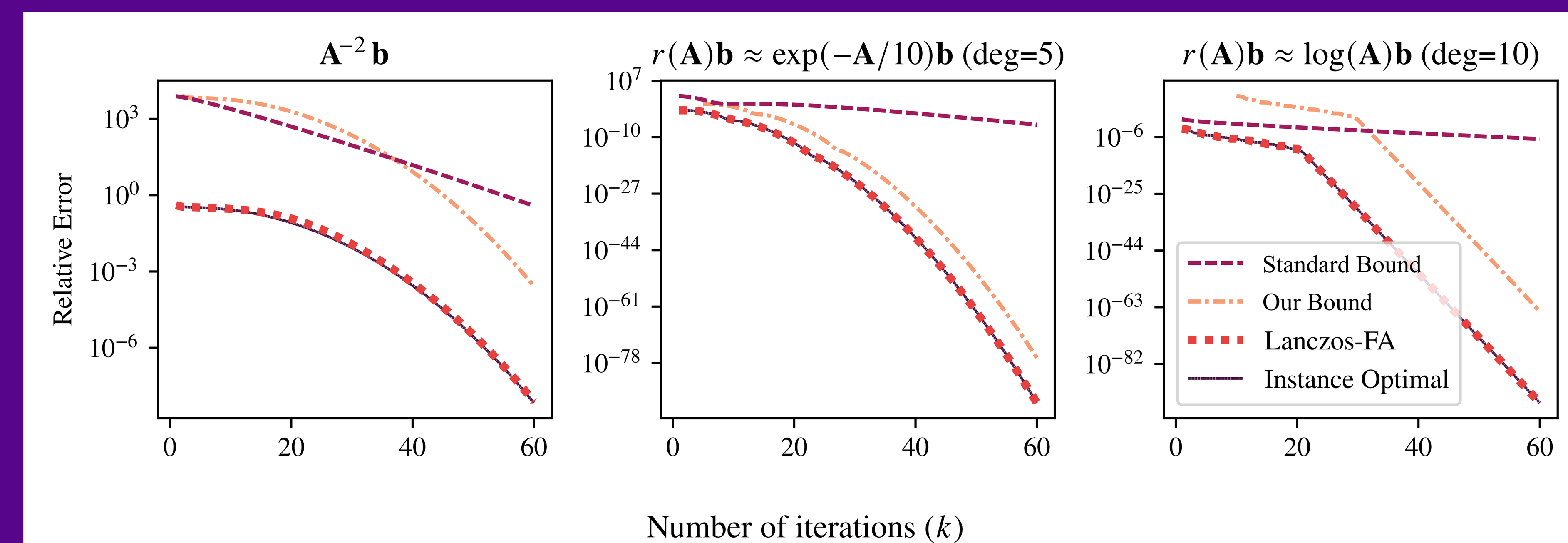
Prior Improved Analysis for $\mathbf{A}^{-1}\mathbf{b}$ and $\exp(\mathbf{A})\mathbf{b}$

- For $A \succeq 0$, Lanczos on $\mathbf{A}^{-1}\mathbf{b}$ is just conjugate gradients, so *super-exponential* convergence!

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

- For exp, there's a similar guarantee that adapts to $\mathbf{A}$ and $\mathbf{b}$
- Can we extend these guarantees to more functions, like $\mathbf{A}^{-2}\mathbf{b}$?

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



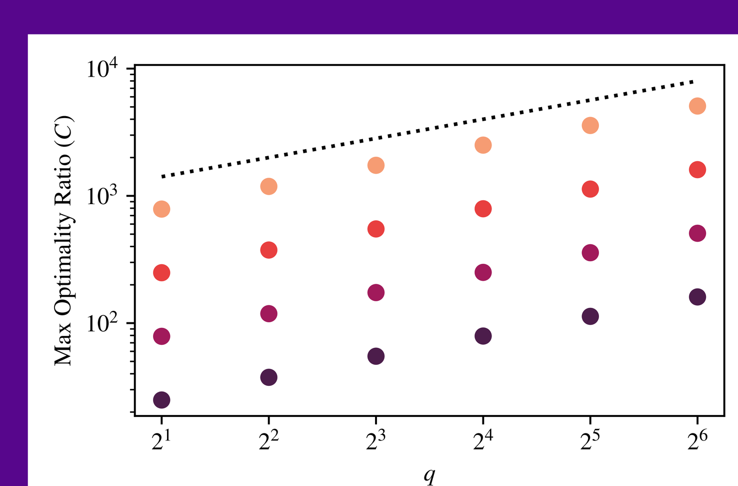
Main Theorem

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

Error of k-step Lanczos in exact arithmetic; Degree of r's denom; Condition number; Error of best degree $\approx k$ polynomial approximation

- Standard analysis does not depend on $\mathbf{A}$ and $\mathbf{b}$, just $\lambda_{\min}$ and $\lambda_{\max}$
- Our bound shows that Lanczos adapts to each specific $\mathbf{A}$ and $\mathbf{b}$. Much better at capturing the observed convergence behavior.

(Above is slightly simplified. In general, prefactor is this → where $z_1, \ldots, z_q \notin [\lambda_{\min}, \lambda_{\max}]$ are roots of r's denominator) $\quad q \cdot \prod_{j=1}^{q} \kappa\left(\mathbf{A} - z_j\mathbf{I}\right)$



Conjecture

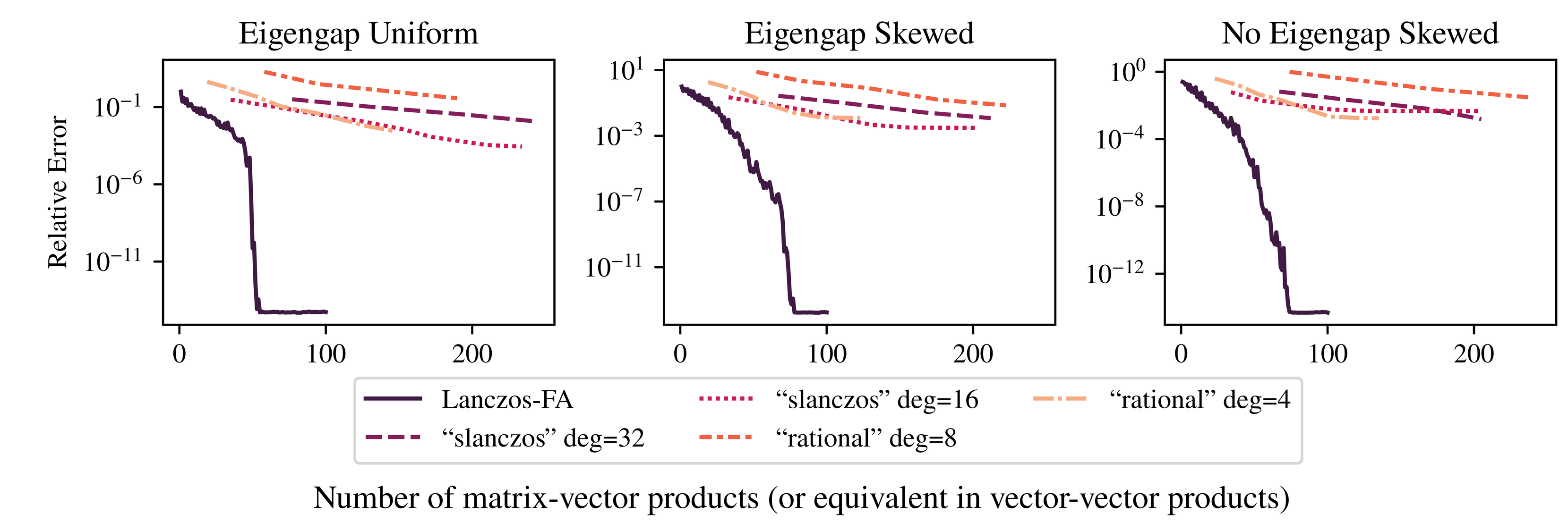
The prefactor can be improved to

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

That would match the hardest family of problems that we could find

Lanczos beats methods based on explicit rational approximations

- Many newer algorithms in the literature work as follows:
  1. Find a rational approximation $r(z) \approx f(z)$
  2. Compute $r(\mathbf{A})\mathbf{b}$ using a Krylov linear system solver
- But vanilla Lanczos is better in practice, e.g. for sign function:



Applying our bound to non-rational functions

- Our analysis automatically transfers to any f that is close to rational.
- If Lanczos is nearly optimal on rational r with up to a factor of C_r, then by triangle inequality

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

How well r approximates f on the range of eigenvalues; Error of best degree $\approx k$ polynomial approximation

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

Using different techniques, we prove a weaker, looser optimality guarantee for the matrix square root that adapts to $\mathbf{A}$ (but not $\mathbf{b}$):

$$\left\| \mathbf{A}^{-1/2}\mathbf{b} - \mathsf{lan}_k \right\|_2 \le \frac{3}{\sqrt{\pi k}} \sqrt{\kappa(\mathbf{A})} \cdot \min_{\deg(p)\le k/2} \max_{x\in\Lambda} \left| \frac{1}{\sqrt{x}} - p(x) \right|$$

Future Work

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