Is the Lanczos-Method for Matrix Functions Nearly Optimal?

Tyler Chen (joint with N. Amsel, A. Greenbaum, G. Meurant, C. Musco, C. Musco)

May 13, 2024

`chen.pw/slides`

What is a matrix function?

An $n \times n$ symmetric matrix $\mathbf{A}$ has real eigenvalues and orthonormal eigenvectors:

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top.$$

The matrix function $f(\mathbf{A})$ is defined as

$$f(\mathbf{A}) := \sum_{i=1}^{n} f(\lambda_i) \mathbf{u}_i \mathbf{u}_i^\top.$$

What are we doing with matrix functions?

Common matrix functions include:

- $f(x) = x^{-1}$
- $f(x) = \exp(-\beta x)$ for all β in some range
- $f(x) = \sqrt{x}$
- $f(x) = \text{sign}(x)$

Computational Task. Approximate $f(\mathbf{A})\mathbf{b}$

- Want to avoid forming $f(\mathbf{A})$
- w.l.o.g. assume $\|\mathbf{b}\| = 1$

Kryov subspace methods

Def. The k-th Krylov subspace (generated by $\mathbf{A}$ and $\mathbf{b}$) is

$$\mathcal{K}_k(\mathbf{A}, \mathbf{b}) := \operatorname{span}\{\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \ldots, \mathbf{A}^{k-1}\mathbf{b}\}.$$

The Lanczos algorithm can be used to obtain a (graded) orthonormal basis $\mathbf{q}_0, \ldots, \mathbf{q}_k$ for $\mathcal{K}_{k+1}(\mathbf{A}, \mathbf{b})$.

This basis satisfies a symmetric three-term recurrence

$$\mathbf{A}\mathbf{q}_n = \beta_{n-1}\mathbf{q}_{n-1} + \alpha_n\mathbf{q}_n + \beta_n\mathbf{q}_{n+1},$$

with initial conditions $\mathbf{q}_{-1} = \mathbf{0}$ and $\beta_{-1} = 0$.

Lanczos matrix relation

The coefficients $\{\alpha_n\}$ and $\{\beta_n\}$ defining the three term recurrence are also generated by the algorithm. This recurrence can be written in matrix form as

$$\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_k\mathbf{T}_k + \beta_{k-1}\mathbf{q}_k\mathbf{e}_k^\top$$

where

$$\mathbf{Q}_k := \begin{bmatrix} | & | & & | \\ \mathbf{q}_0 & \mathbf{q}_1 & \cdots & \mathbf{q}_{k-1} \\ | & | & & | \end{bmatrix}, \quad \mathbf{T}_k := \begin{bmatrix} \alpha_0 & \beta_0 & & \\ \beta_0 & \alpha_1 & \ddots & \\ & \ddots & \ddots & \beta_{k-2} \\ & & \beta_{k-2} & \alpha_{k-1} \end{bmatrix}.$$

The orthonormality of the Krylov basis implies that $\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k = \mathbf{T}_k$.

The Lanczos-method for matrix function approximation

Def. The k-th Lanczos-FA iterate is

$$\mathsf{lan\text{-}FA}_k(f) := \mathbf{Q}_k f(\mathbf{T}_k)\mathbf{e}_1.$$

Method introduced and studied in 1980s.[1]

Lots of competitors have better theoretical guarantees,[2] but Lanczos-FA typically works best in practice!

Ongoing Reserach Program: Explain why Lanczos-FA works so well.

[1] Nauts and Wyatt 1983; Park and Light 1986; Vorst 1987; Druskin and Knizhnerman 1988; Druskin and Knizhnerman 1989; Gallopoulos and Saad 1992; Saad 1992, etc.

[2] Eshof, Frommer, Lippert, Schilling, and Vorst 2002; Jin and Sidford 2019; Chen, Greenbaum, Musco, and Musco 2023.

The Lanczos-method for matrix function approximation

Def. The k-th Lanczos-FA iterate is

$$\mathsf{lan\text{-}FA}_k(f) := \mathbf{Q}_k f(\mathbf{T}_k)\mathbf{e}_1.$$

Method introduced and studied in 1980s.[1]

Lots of competitors have better theoretical guarantees,[2] but Lanczos-FA typically works best in practice!

Ongoing Reserach Program: Explain why Lanczos-FA works so well.

[1]Nauts and Wyatt 1983; Park and Light 1986; Vorst 1987; Druskin and Knizhnerman 1988; Druskin and Knizhnerman 1989; Gallopoulos and Saad 1992; Saad 1992, etc.

[2]Eshof, Frommer, Lippert, Schilling, and Vorst 2002; Jin and Sidford 2019; Chen, Greenbaum, Musco, and Musco 2023.

Exactness for low-degree polynomials

Lemma. Suppose $\deg(p) < k$. Then $\text{lan-FA}_k(p) = p(\mathbf{A})\mathbf{b}$.

Proof. By linearity, it suffices to verify for $\mathbf{A}^j\mathbf{b}$, $j < k$.

Note that $\mathbf{Q}_k\mathbf{Q}_k^\top$ is the orthogonal projector onto $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$. Hence,

$$
\begin{aligned}
\mathbf{A}^j\mathbf{b} &= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}^j\mathbf{b} \\
&= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}^{j-1}\mathbf{b} \\
&= \cdots \\
&= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k \cdots \mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}^\top\mathbf{b} \\
&= \mathbf{Q}_k\mathbf{T}_k^j\mathbf{e}_1.
\end{aligned}
$$

□

$\text{lan-FA}_k(f) = p(\mathbf{A})\mathbf{b}$ so that $f(\mathbf{T}_k) = p(\mathbf{T}_k)$; i.e. interpolating at the eigenvalues of $\mathbf{T}_k$.

Exactness for low-degree polynomials

Lemma. Suppose $\deg(p) < k$. Then $\text{lan-FA}_k(p) = p(\mathbf{A})\mathbf{b}$.

Proof. By linearity, it suffices to verify for $\mathbf{A}^j\mathbf{b}$, $j < k$.

Note that $\mathbf{Q}_k\mathbf{Q}_k^\top$ is the orthogonal projector onto $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$. Hence,

$$\begin{aligned}
\mathbf{A}^j\mathbf{b} &= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}^j\mathbf{b} \\
&= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}^{j-1}\mathbf{b} \\
&= \cdots \\
&= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k \cdots \mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}^\top\mathbf{b} \\
&= \mathbf{Q}_k\mathbf{T}_k^j\mathbf{e}_1.
\end{aligned}$$

□

$\text{lan-FA}_k(f) = p(\mathbf{A})\mathbf{b}$ so that $f(\mathbf{T}_k) = p(\mathbf{T}_k)$; i.e. interpolating at the eigenvalues of $\mathbf{T}_k$.

Exactness for low-degree polynomials

Lemma. Suppose $\deg(p) < k$. Then $\mathsf{lan\text{-}FA}_k(p) = p(\mathbf{A})\mathbf{b}$.

Proof. By linearity, it suffices to verify for $\mathbf{A}^j\mathbf{b}$, $j < k$.

Note that $\mathbf{Q}_k\mathbf{Q}_k^\top$ is the orthogonal projector onto $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$. Hence,

$$\begin{aligned}
\mathbf{A}^j\mathbf{b} &= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}^j\mathbf{b} \\
&= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}^{j-1}\mathbf{b} \\
&= \cdots \\
&= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\cdots\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}^\top\mathbf{b} \\
&= \mathbf{Q}_k\mathbf{T}_k^j\mathbf{e}_1.
\end{aligned}$$

□

$\mathsf{lan\text{-}FA}_k(f) = p(\mathbf{A})\mathbf{b}$ so that $f(\mathbf{T}_k) = p(\mathbf{T}_k)$; i.e. interpolating at the eigenvalues of $\mathbf{T}_k$.

Exactness for low-degree polynomials

Lemma. Suppose $\deg(p) < k$. Then $\mathsf{lan\text{-}FA}_k(p) = p(\mathbf{A})\mathbf{b}$.

Proof. By linearity, it suffices to verify for $\mathbf{A}^j\mathbf{b}$, $j < k$.

Note that $\mathbf{Q}_k\mathbf{Q}_k^\mathsf{T}$ is the orthogonal projector onto $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$. Hence,

$$\begin{aligned}\mathbf{A}^j\mathbf{b} &= \mathbf{Q}_k\mathbf{Q}_k^\mathsf{T}\mathbf{A}^j\mathbf{b} \\ &= \mathbf{Q}_k\mathbf{Q}_k^\mathsf{T}\mathbf{A}\mathbf{Q}_k\mathbf{Q}_k^\mathsf{T}\mathbf{A}^{j-1}\mathbf{b} \\ &= \cdots \\ &= \mathbf{Q}_k\mathbf{Q}_k^\mathsf{T}\mathbf{A}\mathbf{Q}_k\mathbf{Q}_k^\mathsf{T}\mathbf{A}\mathbf{Q}_k \cdots \mathbf{Q}_k^\mathsf{T}\mathbf{A}\mathbf{Q}_k\mathbf{Q}^\mathsf{T}\mathbf{b} \\ &= \mathbf{Q}_k\mathbf{T}_k^j\mathbf{e}_1.\end{aligned}$$

□

$\mathsf{lan\text{-}FA}_k(f) = p(\mathbf{A})\mathbf{b}$ so that $f(\mathbf{T}_k) = p(\mathbf{T}_k)$; i.e. interpolating at the eigenvalues of $\mathbf{T}_k$.

Exactness for low-degree polynomials

Lemma. Suppose $\deg(p) < k$. Then $\text{lan-FA}_k(p) = p(\mathbf{A})\mathbf{b}$.

Proof. By linearity, it suffices to verify for $\mathbf{A}^j\mathbf{b}$, $j < k$.

Note that $\mathbf{Q}_k\mathbf{Q}_k^\top$ is the orthogonal projector onto $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$. Hence,

$$\begin{aligned}\mathbf{A}^j\mathbf{b} &= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}^j\mathbf{b} \\ &= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}^{j-1}\mathbf{b} \\ &= \cdots \\ &= \mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\cdots\mathbf{Q}_k^\top\mathbf{A}\mathbf{Q}_k\mathbf{Q}^\top\mathbf{b} \\ &= \mathbf{Q}_k\mathbf{T}_k^j\mathbf{e}_1.\end{aligned}$$ □

$\text{lan-FA}_k(f) = p(\mathbf{A})\mathbf{b}$ so that $f(\mathbf{T}_k) = p(\mathbf{T}_k)$; i.e. interpolating at the eigenvalues of $\mathbf{T}_k$.

Optimality for the inverse: a connection to Conjugate Gradient

Theorem. If $f(x) = x^{-1}$ and $\mathbf{A}$ is positive definite, Lanczos-FA is $\mathbf{A}$-norm optimal.

Proof. Any vector $\mathbf{x} \in \mathcal{K}_k(\mathbf{A}, \mathbf{b})$ can be written as $\mathbf{x} = \mathbf{Q}_k\mathbf{c}$ for some $\mathbf{c} \in \mathbb{R}^k$.

Therefore

$$\begin{aligned}\operatorname*{argmin}_{\mathbf{x}\in\mathcal{K}_k(\mathbf{A},\mathbf{b})} \|\mathbf{A}^{-1}\mathbf{b} - \mathbf{x}\|_{\mathbf{A}} &= \mathbf{Q}_k \operatorname*{argmin}_{\mathbf{c}\in\mathbb{R}^k} \|\mathbf{A}^{-1}\mathbf{b} - \mathbf{Q}_k\mathbf{c}\|_{\mathbf{A}} \\ &= \mathbf{Q}_k \operatorname*{argmin}_{\mathbf{c}\in\mathbb{R}^k} \|\mathbf{A}^{-1/2}\mathbf{b} - \mathbf{A}^{1/2}\mathbf{Q}_k\mathbf{c}\|.\end{aligned}$$

The solution to this least squares problem is

$$\mathbf{Q}_k(\mathbf{Q}_k^\mathsf{T}\mathbf{A}^{1/2}\mathbf{A}^{1/2}\mathbf{Q}_k)^{-1}\mathbf{Q}_k^\mathsf{T}\mathbf{A}^{1/2}\mathbf{A}^{-1/2}\mathbf{b} = \mathbf{Q}_k\mathbf{T}_k^{-1}\mathbf{e}_1 = \mathsf{lan\text{-}FA}_k(x^{-1}).$$

□

Optimality for the inverse: a connection to Conjugate Gradient

Theorem. If $f(x) = x^{-1}$ and $\mathbf{A}$ is positive definite, Lanczos-FA is $\mathbf{A}$-norm optimal.

Proof. Any vector $\mathbf{x} \in \mathcal{K}_k(\mathbf{A}, \mathbf{b})$ can be written as $\mathbf{x} = \mathbf{Q}_k\mathbf{c}$ for some $\mathbf{c} \in \mathbb{R}^k$.

Therefore

$$\operatorname*{argmin}_{\mathbf{x}\in\mathcal{K}_k(\mathbf{A},\mathbf{b})} \|\mathbf{A}^{-1}\mathbf{b} - \mathbf{x}\|_{\mathbf{A}} = \mathbf{Q}_k \operatorname*{argmin}_{\mathbf{c}\in\mathbb{R}^k} \|\mathbf{A}^{-1}\mathbf{b} - \mathbf{Q}_k\mathbf{c}\|_{\mathbf{A}}$$
$$= \mathbf{Q}_k \operatorname*{argmin}_{\mathbf{c}\in\mathbb{R}^k} \|\mathbf{A}^{-1/2}\mathbf{b} - \mathbf{A}^{1/2}\mathbf{Q}_k\mathbf{c}\|.$$

The solution to this least squares problem is

$$\mathbf{Q}_k(\mathbf{Q}_k^\mathsf{T}\mathbf{A}^{1/2}\mathbf{A}^{1/2}\mathbf{Q}_k)^{-1}\mathbf{Q}_k^\mathsf{T}\mathbf{A}^{1/2}\mathbf{A}^{-1/2}\mathbf{b} = \mathbf{Q}_k\mathbf{T}_k^{-1}\mathbf{e}_1 = \mathsf{lan\text{-}FA}_k(x^{-1}).$$ □

Optimality for the inverse: a connection to Conjugate Gradient

Theorem. If $f(x) = x^{-1}$ and $\mathbf{A}$ is positive definite, Lanczos-FA is $\mathbf{A}$-norm optimal.

Proof. Any vector $\mathbf{x} \in \mathcal{K}_k(\mathbf{A}, \mathbf{b})$ can be written as $\mathbf{x} = \mathbf{Q}_k\mathbf{c}$ for some $\mathbf{c} \in \mathbb{R}^k$.

Therefore

$$\underset{\mathbf{x}\in\mathcal{K}_k(\mathbf{A},\mathbf{b})}{\operatorname{argmin}} \|\mathbf{A}^{-1}\mathbf{b} - \mathbf{x}\|_{\mathbf{A}} = \mathbf{Q}_k \underset{\mathbf{c}\in\mathbb{R}^k}{\operatorname{argmin}} \|\mathbf{A}^{-1}\mathbf{b} - \mathbf{Q}_k\mathbf{c}\|_{\mathbf{A}}$$

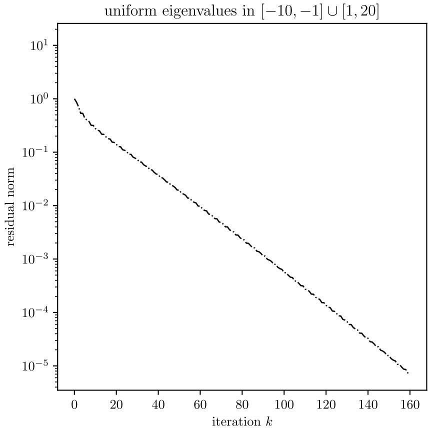
$$= \mathbf{Q}_k \underset{\mathbf{c}\in\mathbb{R}^k}{\operatorname{argmin}} \|\mathbf{A}^{-1/2}\mathbf{b} - \mathbf{A}^{1/2}\mathbf{Q}_k\mathbf{c}\|.$$

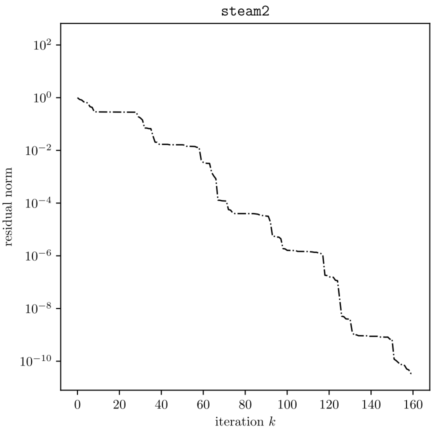
The solution to this least squares problem is

$$\mathbf{Q}_k(\mathbf{Q}_k^\mathsf{T}\mathbf{A}^{1/2}\mathbf{A}^{1/2}\mathbf{Q}_k)^{-1}\mathbf{Q}_k^\mathsf{T}\mathbf{A}^{1/2}\mathbf{A}^{-1/2}\mathbf{b} = \mathbf{Q}_k\mathbf{T}_k^{-1}\mathbf{e}_1 = \mathsf{lan\text{-}FA}_k(x^{-1}). \qquad \square$$

Warm up: CG on indefinite systems?

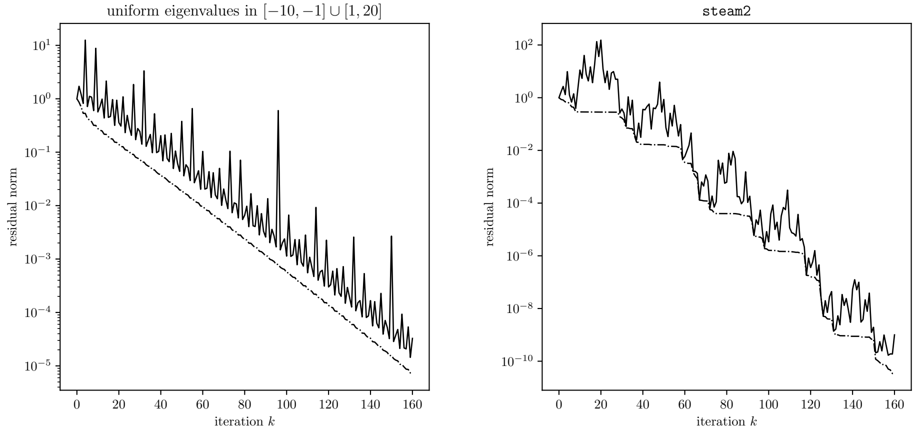
What happens to CG/Lanczos-FA if $\mathbf{A}$ is symmetric indefinite (or nonsymmetric)?





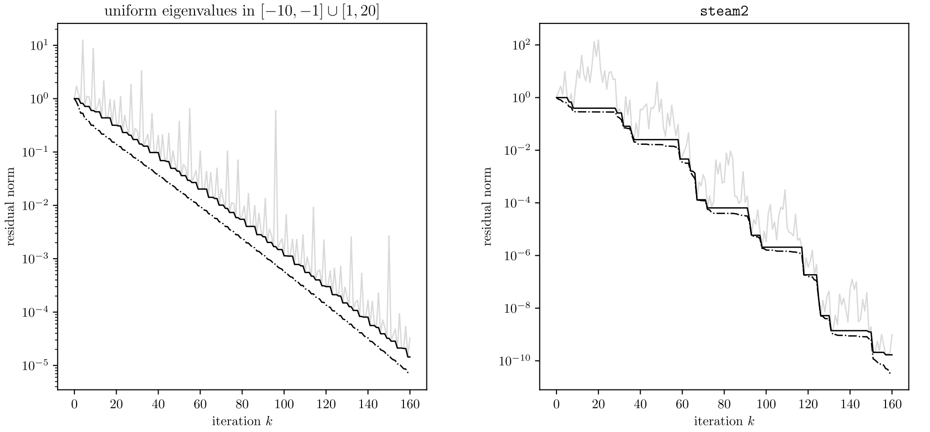
Warm up: CG on indefinite systems?

What happens to CG/Lanczos-FA if **A** is symmetric indefinite (or nonsymmetric)?



Warm up: CG on indefinite systems?

What happens to CG/Lanczos-FA if $\mathbf{A}$ is symmetric indefinite (or nonsymmetric)?



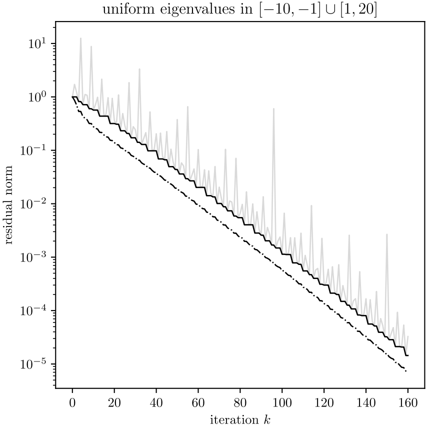
CG on indefinite systems[3]

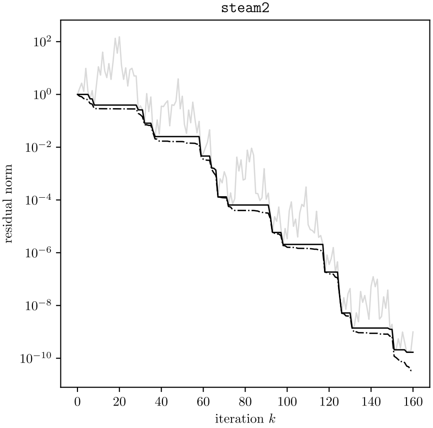
Theorem. Lanczos-FA satisfies $\min_{j\leq k} \|\mathbf{b} - \mathbf{A}\,\mathsf{lan\text{-}FA}_j(x^{-1})\| \leq \sqrt{k+1} \min_{\mathbf{x}\in\mathcal{K}_k(\mathbf{A},\mathbf{b})} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|$.

[3] Chen and Meurant 2024.

CG on indefinite systems[3]

Theorem. Lanczos-FA satisfies $\min_{j \le k} \|\mathbf{b} - \mathbf{A}\,\text{lan-FA}_j(x^{-1})\| \le \sqrt{k+1} \min_{\mathbf{x} \in \mathcal{K}_k(\mathbf{A},\mathbf{b})} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|$.

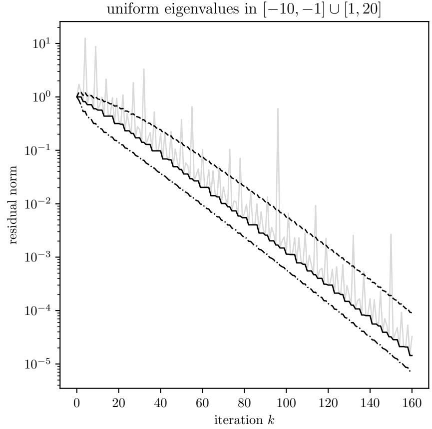


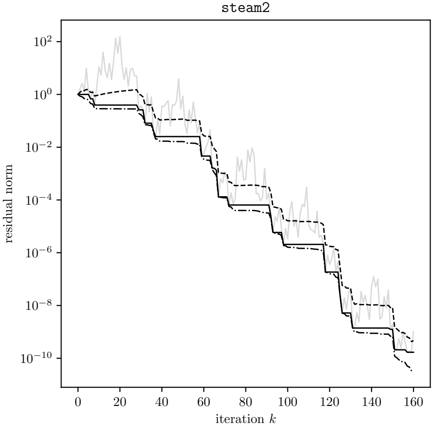


[3] Chen and Meurant 2024.

CG on indefinite systems[3]

Theorem. Lanczos-FA satisfies $\min_{j\leq k} \|\mathbf{b} - \mathbf{A}\,\mathrm{lan\text{-}FA}_j(x^{-1})\| \leq \sqrt{k+1} \min_{\mathbf{x}\in\mathcal{K}_k(\mathbf{A},\mathbf{b})} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|$.





[3] Chen and Meurant 2024.

What about the general case?

Theorem. Lanczos-FA satisfies $\|f(\mathbf{A})\mathbf{b} - \text{lan-FA}_k(f)\| \leq 2 \min_{\deg(p)<k} \|f - p\|_{[\lambda_{\min}, \lambda_{\max}]}$.

Proof. Fix a polynomial p with $\deg(p) < k$ and define $e(x) = f(x) - p(x)$. Then,

$$\begin{aligned}
\|f(\mathbf{A})\mathbf{b} - \text{lan-FA}_k(f)\| &= \|f(\mathbf{A})\mathbf{b} - p(\mathbf{A})\mathbf{b} + \text{lan-FA}_k(p) - \text{lan-FA}_k(f)\| \\
&\leq \|p(\mathbf{A})\mathbf{b} - f(\mathbf{A})\mathbf{b}\| + \|\text{lan-FA}_k(f) - \text{lan-FA}_k(p)\|. \\
&\leq \|e(\mathbf{A})\mathbf{b}\| + \|\mathbf{Q}_k e(\mathbf{T}_k)\mathbf{e}_1\| \\
&\leq \|e(\mathbf{A})\|_2 + \|\mathbf{Q}_k\|_2 \|e(\mathbf{T}_k)\|_2 \\
&= \|e\|_{\Lambda(\mathbf{A})} + \|e\|_{\Lambda(\mathbf{T}_k)} \leq 2\|e\|_{[\lambda_{\min}, \lambda_{\max}]}.
\end{aligned}$$

Optimizing over p gives the result. □

This bound is essentially sharp, but hard instances aren't real-world instances.

What about the general case?

Theorem. Lanczos-FA satisfies $\|f(\mathbf{A})\mathbf{b} - \text{lan-FA}_k(f)\| \leq 2 \min_{\deg(p)<k} \|f - p\|_{[\lambda_{\min},\lambda_{\max}]}$.

Proof. Fix a polynomial p with $\deg(p) < k$ and define $e(x) = f(x) - p(x)$. Then,

$$\begin{aligned}
\|f(\mathbf{A})\mathbf{b} - \text{lan-FA}_k(f)\| &= \|f(\mathbf{A})\mathbf{b} - p(\mathbf{A})\mathbf{b} + \text{lan-FA}_k(p) - \text{lan-FA}_k(f)\| \\
&\leq \|p(\mathbf{A})\mathbf{b} - f(\mathbf{A})\mathbf{b}\| + \|\text{lan-FA}_k(f) - \text{lan-FA}_k(p)\|. \\
&\leq \|e(\mathbf{A})\mathbf{b}\| + \|\mathbf{Q}_k e(\mathbf{T}_k)\mathbf{e}_1\| \\
&\leq \|e(\mathbf{A})\|_2 + \|\mathbf{Q}_k\|_2 \|e(\mathbf{T}_k)\|_2 \\
&= \|e\|_{\Lambda(\mathbf{A})} + \|e\|_{\Lambda(\mathbf{T}_k)} \leq 2\|e\|_{[\lambda_{\min},\lambda_{\max}]}.
\end{aligned}$$

Optimizing over p gives the result. □

This bound is essentially sharp, but hard instances aren't real-world instances.

What about the general case?

Theorem. Lanczos-FA satisfies $\|f(\mathbf{A})\mathbf{b} - \mathsf{lan\text{-}FA}_k(f)\| \leq 2 \min_{\deg(p)<k} \|f - p\|_{[\lambda_{\min},\lambda_{\max}]}$.

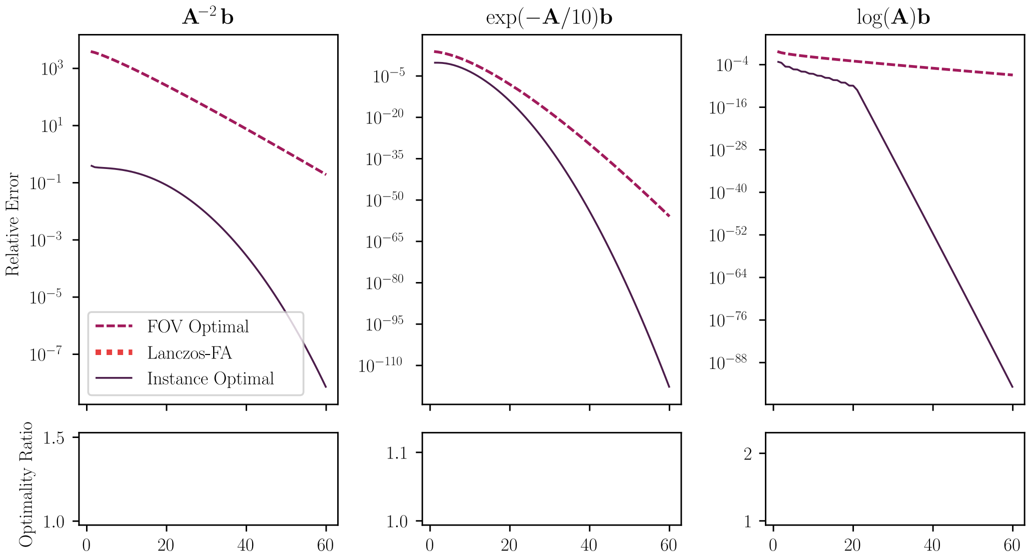
Proof. Fix a polynomial p with $\deg(p) < k$ and define $e(x) = f(x) - p(x)$. Then,

$$\begin{aligned} \|f(\mathbf{A})\mathbf{b} - \mathsf{lan\text{-}FA}_k(f)\| &= \|f(\mathbf{A})\mathbf{b} - p(\mathbf{A})\mathbf{b} + \mathsf{lan\text{-}FA}_k(p) - \mathsf{lan\text{-}FA}_k(f)\| \\ &\leq \|p(\mathbf{A})\mathbf{b} - f(\mathbf{A})\mathbf{b}\| + \|\mathsf{lan\text{-}FA}_k(f) - \mathsf{lan\text{-}FA}_k(p)\|. \\ &\leq \|e(\mathbf{A})\mathbf{b}\| + \|\mathbf{Q}_k e(\mathbf{T}_k)\mathbf{e}_1\| \\ &\leq \|e(\mathbf{A})\|_2 + \|\mathbf{Q}_k\|_2 \|e(\mathbf{T}_k)\|_2 \\ &= \|e\|_{\Lambda(\mathbf{A})} + \|e\|_{\Lambda(\mathbf{T}_k)} \leq 2\|e\|_{[\lambda_{\min},\lambda_{\max}]}. \end{aligned}$$

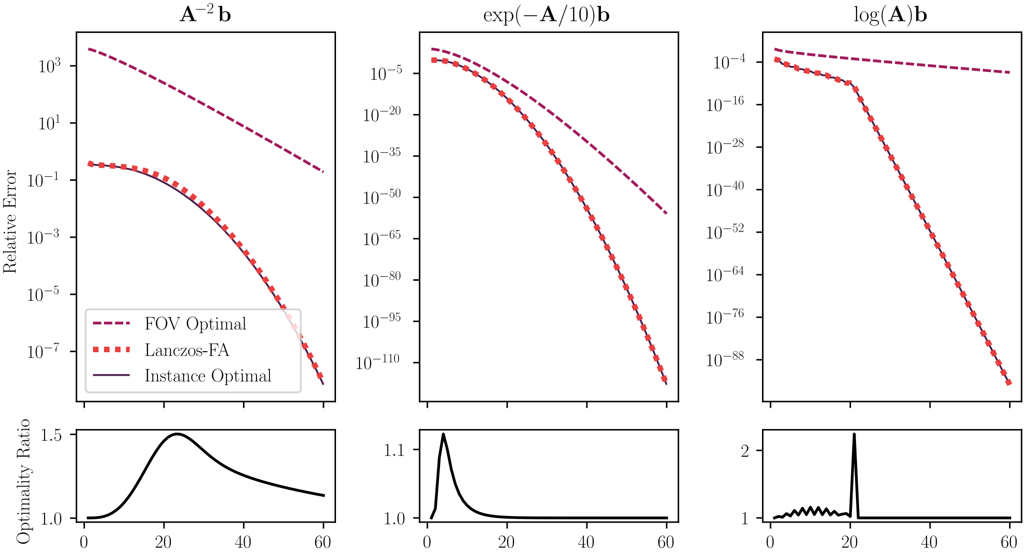
Optimizing over p gives the result. □

This bound is essentially sharp, but hard instances aren't real-world instances.

Some examples



Some examples



Is Lanczos-FA nearly optimal?

Interval optimality bound is clearly not good enough!

- For CG, this amounts to the $\sqrt{\kappa}$ bound, which is not usually indicative of the algorithm's actual behavior!

Research Question: Can we prove convergence guarantees for Lanczos-FA in terms of the best-possible Krylov Subspace Methods? I.e. does

$$\|f(\mathbf{A})\mathbf{b} - \text{lan-FA}_k(f)\| \leq C \min_{\mathbf{x} \in \mathcal{K}_k(\mathbf{A},\mathbf{b})} \|f(\mathbf{A})\mathbf{b} - \mathbf{x}\| ?$$

Is Lanczos-FA nearly optimal?

Interval optimality bound is clearly not good enough!

- For CG, this amounts to the $\sqrt{\kappa}$ bound, which is not usually indicative of the algorithm's actual behavior!

Research Question: Can we prove convergence guarantees for Lanczos-FA in terms of the best-possible Krylov Subspace Methods? I.e. does

$$\|f(\mathbf{A})\mathbf{b} - \text{lan-FA}_k(f)\| \leq C \min_{\mathbf{x} \in \mathcal{K}_k(\mathbf{A},\mathbf{b})} \|f(\mathbf{A})\mathbf{b} - \mathbf{x}\|?$$

An optimality bound[4]

Theorem. Let $r(x) = n(x)/m(x)$ be a degree (p, q) rational function, where $m(x) = (x - z_1)(x - z_2)\cdots(x - z_q)$ and $z_i \in \mathbb{R} \setminus [\lambda_{\min}, \lambda_{\max}]$. Then, provided $k > \max\{p, q-1\}$, the Lanczos-FA iterate satisfies the bound

$$\|\text{lan-FA}_k(r) - r(\mathbf{A})\mathbf{b}\| \leq C(r, \lambda_{\min}, \lambda_{\max}) \min_{\mathbf{x} \in \mathcal{K}_{k-q+1}(\mathbf{A},\mathbf{b})} \|r(\mathbf{A}) - \mathbf{x}\|.$$

This is a near-optimality guarantee for a very wide class of functions!

- Recovers CG bound when $n(x) = 1$ and $m(x) = (x - 0)$
- Often more indicative of true behavior of Lanczos-FA than interval bound

[4]Amsel, Chen, Greenbaum, Musco, and Musco 2023.

An optimality bound[4]

Theorem. Let $r(x) = n(x)/m(x)$ be a degree (p, q) rational function, where $m(x) = (x - z_1)(x - z_2) \cdots (x - z_q)$ and $z_i \in \mathbb{R} \setminus [\lambda_{\min}, \lambda_{\max}]$. Then, provided $k > \max\{p, q-1\}$, the Lanczos-FA iterate satisfies the bound

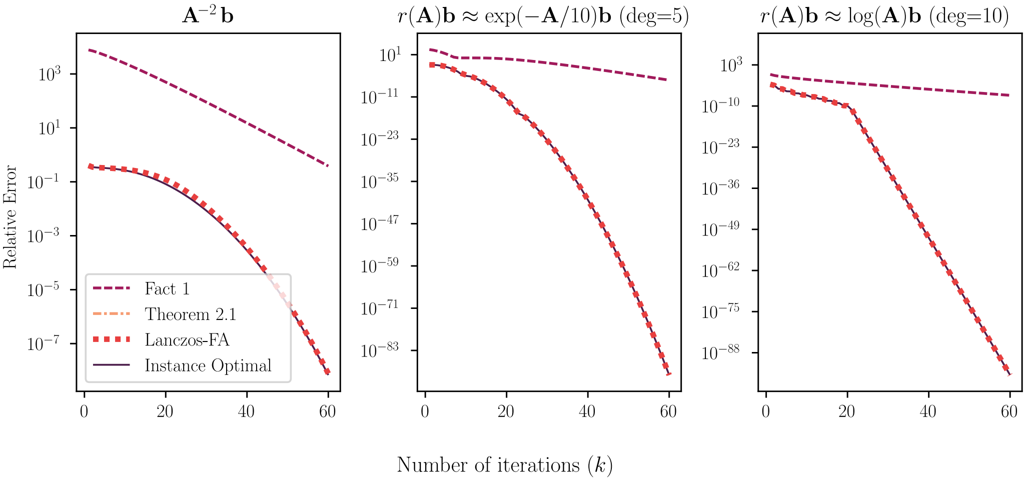
$$\|\mathsf{lan\text{-}FA}_k(r) - r(\mathbf{A})\mathbf{b}\| \leq C(r, \lambda_{\min}, \lambda_{\max}) \min_{\mathbf{x} \in \mathcal{K}_{k-q+1}(\mathbf{A},\mathbf{b})} \|r(\mathbf{A}) - \mathbf{x}\|.$$

This is a near-optimality guarantee for a very wide class of functions!

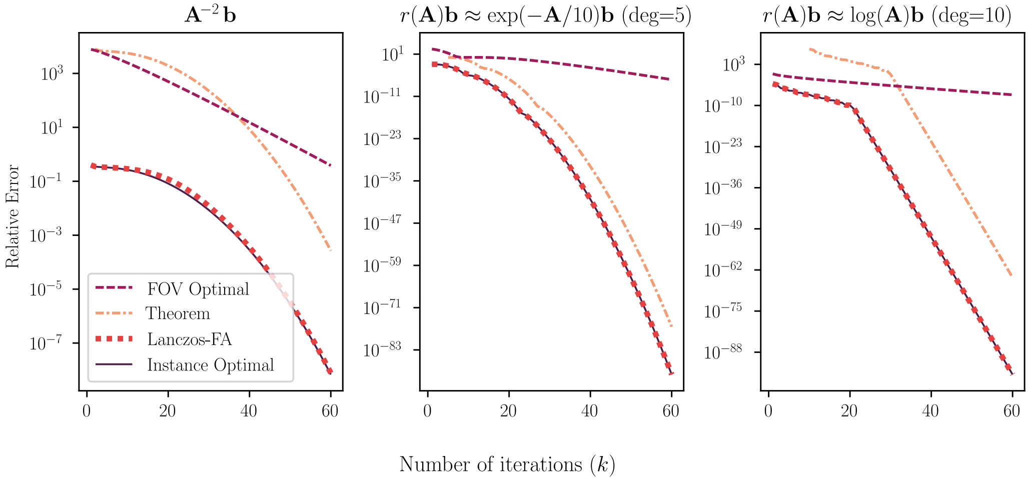
- Recovers CG bound when $n(x) = 1$ and $m(x) = (x - 0)$
- Often more indicative of true behavior of Lanczos-FA than interval bound

[4] Amsel, Chen, Greenbaum, Musco, and Musco 2023.

Some examples (revisited)



Some examples (revisited)



Proof sketch

Let's consider $r(x) = x^{-2}$ and PSD $\mathbf{A}$. By the triangle inequality,

$$\begin{aligned}\|\mathbf{A}^{-2}\mathbf{b} - \text{lan-FA}_k(x^{-2})\| &\leq \|\mathbf{A}^{-2}\mathbf{b} - \mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{A}^{-1}\mathbf{b}\| + \|\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{A}^{-1}\mathbf{b} - \mathbf{Q}\mathbf{T}^{-2}\mathbf{Q}^\mathsf{T}\mathbf{b}\| \\ &\leq \|\mathbf{A}^{-2}\mathbf{b} - \underbrace{\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{A}\mathbf{A}^{-2}\mathbf{b}}_{\mathbf{A}\text{-norm optimal}}\| + \underbrace{\|\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\|}_{\leq \lambda_{\min}^{-1}} \underbrace{\|\mathbf{A}^{-1}\mathbf{b} - \mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{b}\|}_{\text{CG error}}.\end{aligned}$$

Now note if $p(x) \approx x^{-2}$, then $xp(x) \approx x^{-1}$. So,

$$\begin{aligned}\min_{\deg(p)<k} \|\mathbf{A}^{-1}\mathbf{b} - p(\mathbf{A})\mathbf{b}\| &\leq \min_{\deg(p)<k-1} \|\mathbf{A}^{-1}\mathbf{b} - \mathbf{A}p(\mathbf{A})\mathbf{b}\| \\ &\leq \min_{\deg(p)<k-1} \|\mathbf{A}(\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b})\| \\ &\leq \min_{\deg(p)<k-1} \|\mathbf{A}\|\|\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b}\|.\end{aligned}$$

Together, (and using that the $\mathbf{A}$-norm and 2-norm are $\kappa^{1/2}$-equivalent) we get

$$\|\mathbf{A}^{-2}\mathbf{b} - \text{lan-FA}_k(x^{-2})\| \leq \kappa^{1/2} \min_{\deg(p)<k} \|\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b}\| + \kappa^{3/2} \min_{\deg(p)<k-1} \|\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b}\|.$$

Proof sketch

Let's consider $r(x) = x^{-2}$ and PSD $\mathbf{A}$. By the triangle inequality,

$$\begin{aligned}\|\mathbf{A}^{-2}\mathbf{b} - \text{lan-FA}_k(x^{-2})\| &\leq \|\mathbf{A}^{-2}\mathbf{b} - \mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{A}^{-1}\mathbf{b}\| + \|\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{A}^{-1}\mathbf{b} - \mathbf{Q}\mathbf{T}^{-2}\mathbf{Q}^\mathsf{T}\mathbf{b}\| \\ &\leq \|\mathbf{A}^{-2}\mathbf{b} - \underbrace{\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{A}\mathbf{A}^{-2}\mathbf{b}}_{\mathbf{A}\text{-norm optimal}}\| + \underbrace{\|\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\|}_{\leq \lambda_{\min}^{-1}} \underbrace{\|\mathbf{A}^{-1}\mathbf{b} - \mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{b}\|}_{\text{CG error}}.\end{aligned}$$

Now note if $p(x) \approx x^{-2}$, then $xp(x) \approx x^{-1}$. So,

$$\begin{aligned}\min_{\deg(p)<k} \|\mathbf{A}^{-1}\mathbf{b} - p(\mathbf{A})\mathbf{b}\| &\leq \min_{\deg(p)<k-1} \|\mathbf{A}^{-1}\mathbf{b} - \mathbf{A}p(\mathbf{A})\mathbf{b}\| \\ &\leq \min_{\deg(p)<k-1} \|\mathbf{A}(\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b})\| \\ &\leq \min_{\deg(p)<k-1} \|\mathbf{A}\|\|\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b}\|.\end{aligned}$$

Together, (and using that the $\mathbf{A}$-norm and 2-norm are $\kappa^{1/2}$-equivalent) we get

$$\|\mathbf{A}^{-2}\mathbf{b} - \text{lan-FA}_k(x^{-2})\| \leq \kappa^{1/2} \min_{\deg(p)<k} \|\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b}\| + \kappa^{3/2} \min_{\deg(p)<k-1} \|\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b}\|.$$

Proof sketch

Let's consider $r(x) = x^{-2}$ and PSD $\mathbf{A}$. By the triangle inequality,

$$\begin{aligned}\|\mathbf{A}^{-2}\mathbf{b} - \mathsf{lan\text{-}FA}_k(x^{-2})\| &\leq \|\mathbf{A}^{-2}\mathbf{b} - \mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{A}^{-1}\mathbf{b}\| + \|\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{A}^{-1}\mathbf{b} - \mathbf{Q}\mathbf{T}^{-2}\mathbf{Q}^\mathsf{T}\mathbf{b}\| \\ &\leq \|\mathbf{A}^{-2}\mathbf{b} - \underbrace{\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{A}\mathbf{A}^{-2}\mathbf{b}}_{\mathbf{A}\text{-norm optimal}}\| + \underbrace{\|\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\|}_{\leq \lambda_{\min}^{-1}} \underbrace{\|\mathbf{A}^{-1}\mathbf{b} - \mathbf{Q}\mathbf{T}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{b}\|}_{\text{CG error}}.\end{aligned}$$

Now note if $p(x) \approx x^{-2}$, then $xp(x) \approx x^{-1}$. So,

$$\begin{aligned}\min_{\deg(p)<k} \|\mathbf{A}^{-1}\mathbf{b} - p(\mathbf{A})\mathbf{b}\| &\leq \min_{\deg(p)<k-1} \|\mathbf{A}^{-1}\mathbf{b} - \mathbf{A}p(\mathbf{A})\mathbf{b}\| \\ &\leq \min_{\deg(p)<k-1} \|\mathbf{A}(\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b})\| \\ &\leq \min_{\deg(p)<k-1} \|\mathbf{A}\|\|\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b}\|.\end{aligned}$$

Together, (and using that the $\mathbf{A}$-norm and 2-norm are $\kappa^{1/2}$-equivalent) we get

$$\|\mathbf{A}^{-2}\mathbf{b} - \mathsf{lan\text{-}FA}_k(x^{-2})\| \leq \kappa^{1/2} \min_{\deg(p)<k} \|\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b}\| + \kappa^{3/2} \min_{\deg(p)<k-1} \|\mathbf{A}^{-2}\mathbf{b} - p(\mathbf{A})\mathbf{b}\|.$$

Caveats

The prefactor $C(r, \lambda_{\min}, \lambda_{\max})$ is constant for fixed r and matrices with bounded spectrum. But the value we obtain is very bad (proof artifact?).

- If $\mathbf{A}$ is positive definite and $z_i < 0$, then $C(r, \lambda_{\min}, \lambda_{\max}) \leq q\kappa(\mathbf{A})^q$.
- The worst dependence on κ and q we could find numerically is $\sqrt{q \cdot \kappa}$.

This bound does not account for finite precision arithmetic, but it can be connected[5]

[5] Greenbaum 1989.

Future work

- Improve the constant prefactor in the near-optimality bound
- Prove instance optimality guarantees for Markov/Stieltjes functions
  - $f(x) = \int w(z)(x - z)^{-1}$.
- Extend the result to more general rational functions
  - conjugate pairs of poles
  - poles in $[\lambda_{\min}, \lambda_{\max}]$.
- Prove instance optimality guarantees for the exponential[6] or other functions

[6]Druskin, Greenbaum, and Knizhnerman 1998.

Markov/Stieltjes functions

Consider functions of the form

$$f(x) = \int_{-\infty}^{0} w(z)(x-z)^{-1} \approx \sum_i w_i(x-z_i)^{-1}.$$

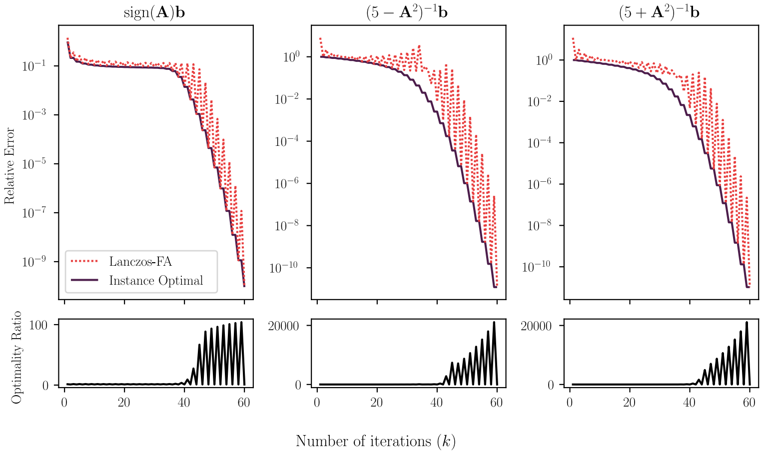
Then the Lanczos-FA iterate is

$$\textsf{lan-FA}_k(f) = \int_{-\infty}^{0} w(z)\textsf{lan-FA}_k((x-z)^{-1}) \approx \sum_i w_i\textsf{lan-FA}_k((x-z_i)^{-1}).$$

This is like CG on a bunch of shifted linear systems… (termwise optimal).

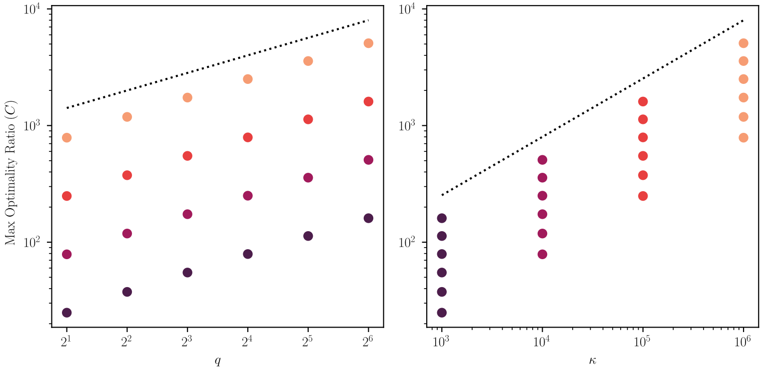
Indefinite problems

Even for other functions, Lanczos-FA seems nearly optimal in an overall sense.



Hard problems?

Different values of κ and q and the worst-case optimality ratio we could find.



References I

Amsel, Noah et al. (2023). *Near-Optimal Approximation of Matrix Functions by the Lanczos Method*.

Chen, Tyler and Gérard Meurant (2024). *Near-optimal convergence of the full orthogonalization method*.

Chen, Tyler et al. (May 2023). “Low-Memory Krylov Subspace Methods for Optimal Rational Matrix Function Approximation”. In: *SIAM Journal on Matrix Analysis and Applications* 44.2, pp. 670–692.

Druskin, Vladimir, Anne Greenbaum, and Leonid Knizhnerman (1998). “Using Nonorthogonal Lanczos Vectors in the Computation of Matrix Functions”. In: *SIAM Journal on Scientific Computing* 19.1, pp. 38–54.

Druskin, Vladimir and Leonid Knizhnerman (Jan. 1988). “Spectral Differential-Difference Method for Numeric Solution of Three-Dimensional Nonstationary Problems of Electric Prospecting”. In: *Physics of the Solid Earth* 24, pp. 641–648.

— (Jan. 1989). “Two polynomial methods of calculating functions of symmetric matrices”. In: *USSR Computational Mathematics and Mathematical Physics* 29.6, pp. 112–121.

Eshof, J. van den et al. (2002). “Numerical methods for the QCD overlap operator. I. Sign-function and error bounds”. In: *Computer Physics Communications* 146.2, pp. 203–224.

Gallopoulos, E. and Y. Saad (Sept. 1992). “Efficient Solution of Parabolic Equations by Krylov Approximation Methods”. In: *SIAM Journal on Scientific and Statistical Computing* 13.5, pp. 1236–1264.

Greenbaum, Anne (1989). “Behavior of slightly perturbed Lanczos and conjugate-gradient recurrences”. In: *Linear Algebra and its Applications* 113, pp. 7–63.

Jin, Yujia and Aaron Sidford (2019). “Principal Component Projection and Regression in Nearly Linear Time through Asymmetric SVRG”. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., pp. 3868–3878.

References II

Nauts, André and Robert E. Wyatt (Dec. 1983). “New Approach to Many-State Quantum Dynamics: The Recursive-Residue-Generation Method”. In: *Physical Review Letters* 51.25, pp. 2238–2241.

Park, Tae Jun and J. C. Light (Nov. 1986). “Unitary quantum time evolution by iterative Lanczos reduction”. In: *The Journal of Chemical Physics* 85.10, pp. 5870–5876.

Saad, Yousef (1992). “Analysis of Some Krylov Subspace Approximations to the Matrix Exponential Operator”. In: *SIAM Journal on Numerical Analysis* 29.1, pp. 209–228.

Vorst, H.A. Van Der (May 1987). “An iterative solution method for solving f(A)x = b, using Krylov subspace information obtained for the symmetric positive definite matrix A”. In: *Journal of Computational and Applied Mathematics* 18.2, pp. 249–263.