Block Krylov methods for functions of matrices

Kathryn Lund-Nguyen, Andreas Frommer, and *Daniel B. Szyld*

Temple University, Philadelphia, USA; Wuppertal, Germany; *Temple Univ.*

NL2A, CIRM, Luminy
27 October 2016

Thanks to NSF DMS-1418882, DOE DE-SC0016578, and especially the organizers

Matrix functions, or Functions of Matrices

$f : \mathbb{C} \to \mathbb{C}$, $A \in \mathbb{C}^{n\times n}$, λ_i eigenvalues of A, and ψ the minimal polynomial of A

- If f is sufficiently differentiable on λ_i, then there exists a unique interpolating polynomial $p_{f,A}$ of degree less than $\deg \psi$ matching derivatives of f on λ_i. Then we define

$$f(A) := p_{f,A}(A)$$

- If f has an integral representation, e.g., f Stieltjes and $f(z) = \int_0^\infty \frac{1}{z+t}\,\mathrm{d}\mu(t)$, then we define

$$f(A) := \int_0^\infty (A + tI)^{-1}\,\mathrm{d}\mu(t)$$

Polynomial and integral definitions are equivalent when f is analytic

Matrix functions, or Functions of Matrices

$f : \mathbb{C} \to \mathbb{C}$, $A \in \mathbb{C}^{n\times n}$, λ_i eigenvalues of A, and ψ the minimal polynomial of A

- If f is sufficiently differentiable on λ_i, then there exists a unique interpolating polynomial $p_{f,A}$ of degree less than $\deg \psi$ matching derivatives of f on λ_i. Then we define

$$f(A) := p_{f,A}(A)$$

- If f has an integral representation, e.g., f Stieltjes and $f(z) = \int_0^\infty \frac{1}{z+t}\,\mathrm{d}\mu(t)$, then we define

$$f(A) := \int_0^\infty (A + tI)^{-1}\,\mathrm{d}\mu(t)$$

Polynomial and integral definitions are equivalent when f is analytic

We use the latter definition numerically with quadrature

Krylov subspace methods for $Ax = \boldsymbol{b}$ and for $f(A)\boldsymbol{b}$

Given $A \in \mathbb{C}^{n\times n}$ (not necessarily HPD) and $\boldsymbol{b} \in \mathbb{C}^{n\times 1}$

- Krylov subspace with respect to A and $\boldsymbol{b}$:

  $$\mathscr{K}_m(A, \boldsymbol{b}) = \operatorname{span}\{\boldsymbol{b}, A\boldsymbol{b}, A^2\boldsymbol{b}, \ldots, A^{m-1}\boldsymbol{b}\} = \{p(A)\boldsymbol{b} : p \in \mathbb{P}_{m-1}\}$$

- Arnoldi (Lanczos if A HPD) relation:

  $$AV_m = V_m H_m + h_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^*$$

- Full Orthogonalization Method (FOM) approximation for $A\boldsymbol{x} = \boldsymbol{b}$ [Saad, 1981]:

  $$\boldsymbol{x}_m := V_m H_m^{-1} V_m^* \boldsymbol{b} \in \mathscr{K}_m(A, \boldsymbol{b}),$$

  with $\boldsymbol{r}_m := \boldsymbol{b} - A\boldsymbol{x}_m \perp \mathscr{K}_m(A, \boldsymbol{b})$

Krylov subspace methods for $A\boldsymbol{x} = \boldsymbol{b}$ and for $f(A)\boldsymbol{b}$

Given $A \in \mathbb{C}^{n \times n}$ (not necessarily HPD) and $\boldsymbol{b} \in \mathbb{C}^{n \times 1}$

- Krylov subspace with respect to A and $\boldsymbol{b}$:

$$\mathscr{K}_m(A, \boldsymbol{b}) = \operatorname{span}\{\boldsymbol{b}, A\boldsymbol{b}, A^2\boldsymbol{b}, \dots, A^{m-1}\boldsymbol{b}\} = \{p(A)\boldsymbol{b} : p \in \mathbb{P}_{m-1}\}$$

- Arnoldi (Lanczos if A HPD) relation:

$$AV_m = V_m H_m + h_{m+1,m} \boldsymbol{v}_{m+1} \boldsymbol{e}_m^*$$

- Full Orthogonalization Method (FOM) approximation for $A\boldsymbol{x} = \boldsymbol{b}$ [Saad, 1981]:

$$\boldsymbol{x}_m := V_m H_m^{-1} V_m^* \boldsymbol{b} \in \mathscr{K}_m(A, \boldsymbol{b}),$$

  with $\boldsymbol{r}_m := \boldsymbol{b} - A\boldsymbol{x}_m \perp \mathscr{K}_m(A, \boldsymbol{b})$

- FOM approximation for $f(A)\boldsymbol{b}$

$$\boldsymbol{x}_m := V_m f(H_m) V_m^* \boldsymbol{b} \in \mathscr{K}_m(A, \boldsymbol{b}),$$

Krylov subspace methods for $Ax = \boldsymbol{b}$ and for $f(A)\boldsymbol{b}$

Given $A \in \mathbb{C}^{n \times n}$ (not necessarily HPD) and $\boldsymbol{b} \in \mathbb{C}^{n \times 1}$

- Krylov subspace with respect to A and $\boldsymbol{b}$:

$$\mathscr{K}_m(A, \boldsymbol{b}) = \operatorname{span}\{\boldsymbol{b}, A\boldsymbol{b}, A^2\boldsymbol{b}, \ldots, A^{m-1}\boldsymbol{b}\} = \{p(A)\boldsymbol{b} : p \in \mathbb{P}_{m-1}\}$$

- Arnoldi (Lanczos if A HPD) relation:

$$AV_m = V_m H_m + h_{m+1,m} \boldsymbol{v}_{m+1} \boldsymbol{e}_m^*$$

- Full Orthogonalization Method (FOM) approximation for $A\boldsymbol{x} = \boldsymbol{b}$ [Saad, 1981]:

$$\boldsymbol{x}_m := V_m H_m^{-1} V_m^* \boldsymbol{b} \in \mathscr{K}_m(A, \boldsymbol{b}),$$

  with $\boldsymbol{r}_m := \boldsymbol{b} - A\boldsymbol{x}_m \perp \mathscr{K}_m(A, \boldsymbol{b})$

- FOM approximation for $f(A)\boldsymbol{b}$

$$\boldsymbol{x}_m := V_m f(H_m) V_m^* \boldsymbol{b} \in \mathscr{K}_m(A, \boldsymbol{b}),$$

  We like to call this $(\text{FOM})^2$: FOM for functions of matrices

Multiple vectors

The problem we want to study is
$f(A)\boldsymbol{b}^{(i)}, i = 1, \ldots, s$, or

Multiple vectors

The problem we want to study is

$f(A)\boldsymbol{b}^{(i)}, i = 1, \dots, s$, or

$f(A)\boldsymbol{B}$, $\boldsymbol{B} \in \mathbb{C}^{n \times s}$

For example in QCD, s=12, f = sign

Multiple vectors

The problem we want to study is

$f(A)\boldsymbol{b}^{(i)}, i = 1, \dots, s$, or

$f(A)\boldsymbol{B}$, $\boldsymbol{B} \in \mathbb{C}^{n \times s}$

For example in QCD, s=12, $f = \text{sign}$

Of course we can use $(\text{FOM})^2$ for each $\boldsymbol{b}^{(i)}$, i.e.,
for $i = 1, \dots, s$, use $(\text{FOM})^2$ to approximate the solution of $f(A)\boldsymbol{b}^{(i)}$

Multiple vectors

The problem we want to study is

$f(A)\boldsymbol{b}^{(i)}, i = 1, \ldots, s$, or

$f(A)\boldsymbol{B}$, $\boldsymbol{B} \in \mathbb{C}^{n \times s}$

For example in QCD, s=12, f = sign

Of course we can use $(\text{FOM})^2$ for each $\boldsymbol{b}^{(i)}$, i.e.,
for $i = 1, \ldots, s$, use $(\text{FOM})^2$ to approximate the solution of $f(A)\boldsymbol{b}^{(i)}$

But we want to use blocks

Multiple vectors

The problem we want to study is

$f(A)\boldsymbol{b}^{(i)}, i = 1, \ldots, s$, or

$f(A)\boldsymbol{B}$, $\boldsymbol{B} \in \mathbb{C}^{n\times s}$

For example in QCD, s=12, f = sign

Of course we can use $(\text{FOM})^2$ for each $\boldsymbol{b}^{(i)}$, i.e.,
for $i = 1, \ldots, s$, use $(\text{FOM})^2$ to approximate the solution of $f(A)\boldsymbol{b}^{(i)}$

But we want to use blocks

Example: Loop interchange [Rashedi et al., 2016]
Do the $b^{(i)}$'s in parallel and interchange "loop over s" with "loop over k"
$\Rightarrow$ innermost loop exposes AV_k with $V_k = [v_k^{(1)} \mid \ldots \mid v_k^{(s)}]$

Why blocks?

$A\boldsymbol{B}$ vs. $[A\boldsymbol{b}^{(1)}| \dots A\boldsymbol{b}^{(s)}]$: only need to access A once

- Good if memory access is slow relative to other operations
- Good if A must be generated each time it is accessed
- In HPC, it is much cheaper to compute $A\boldsymbol{B}$ than s times $A\boldsymbol{b}^{(i)}$

Krylov spaces generated by $\boldsymbol{B}$ vs. a single $\boldsymbol{b}$ are “richer”

Examples of block Krylov spaces

$$
\begin{aligned}
\mathscr{K}_m^{\mathrm{Cl}}(A,\boldsymbol{B}) &= \left\{ \sum_{k=0}^{m-1} A^k \boldsymbol{B} C_k : C_k \in \mathbb{C}^{s\times s} \right\} \\
\mathscr{K}_m^{\mathrm{Gl}}(A,\boldsymbol{B}) &= \operatorname{span}\{\boldsymbol{B}, A\boldsymbol{B}, \ldots, A^{m-1}\boldsymbol{B}\} = \left\{ \sum_{k=0}^{m-1} A^k \boldsymbol{B} c_k : c_k \in \mathbb{C} \right\}, \\
&\ \text{where span is in the usual sense for the vector space } \mathbb{C}^{n\times s} \\
\mathscr{K}_m^{\mathrm{Li}}(A,\boldsymbol{B}) &= \mathscr{K}_m(A,\boldsymbol{b}^{(1)}) \times \cdots \times \mathscr{K}_m(A,\boldsymbol{b}^{(s)}) \\
&= \left\{ \sum_{k=0}^{m-1} A^k \boldsymbol{B} D_k : D_k \in \mathbb{C}^{s\times s} \text{ is diagonal} \right\}
\end{aligned}
$$

Some references on block methods

Classical block methods for linear systems:
block CG: [O'Leary, 1980], [Saad, 1987], ...
block GMRES: [Simoncini and Gallopoulos, 1995, 1996], [Simoncini, 1996], ..., [Gutknecht, 2007], ...
block MINRES: ..., [Soodhalter, 2015]
Also [Langou, thesis 2003] and some talks on blok GCR.

Global (block) methods:
global GMRES and global FOM [Jbilou, Messouadi and Sadok, 1999] **for matrix equations**,
[Heyouni and Essai, 2005], [Bouyouli, Jbilou, Sadaka, and Sadok, 2006], [Elbouyahyaoui, Messouadi, and Sadok, *ETNA*, 2009], ... **for linear systems**

For Functions of Matrices:
Block methods:[Lopez and Simoncini, 2006], [Al-Mohy and N. Higham, 2011] for $\exp(A)$, [Benner et al., 2015] for $\log(A)$.
We also mention [Arrigo, Benzi, and Fenu, *SIMAX*, 2016] for "generalized matrix functions" (yesterday's talk)

Our Contributions

- Block methods for general matrix functions
- Comparison of execution times of the three variants
  (classical, global, loop-interchange)
  plus apply function to each vector one at a time.
- Convergence theory: with and without restarts
  (mostly for A HPD and Stieltjes functions)
- Theory of bilinear forms encompassing all three examples of block Krylov subspaces

A "block inner product"

Let $\mathbb{S}$ be a subalgebra of $\mathbb{C}^{s \times s}$, i.e., a vector subspace closed under matrix multiplication such that $S^* \in \mathbb{S}$ for all $S \in \mathbb{S}$

Definition

A mapping $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$ from $\mathbb{C}^{n \times s} \times \mathbb{C}^{n \times s}$ to $\mathbb{S}$ is called a *block inner product onto* $\mathbb{S}$ if for all $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \in \mathbb{C}^{n \times s}$ and $S \in \mathbb{S}$ we have the following conditions:

(i) $\mathbb{S}$-*linearity*: $\langle\!\langle \boldsymbol{X} + \boldsymbol{Y} S, \boldsymbol{Z} \rangle\!\rangle_{\mathbb{S}} = \langle\!\langle \boldsymbol{X}, \boldsymbol{Z} \rangle\!\rangle_{\mathbb{S}} + \langle\!\langle \boldsymbol{Y}, \boldsymbol{Z} \rangle\!\rangle_{\mathbb{S}} S$,

(ii) *symmetry*: $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}} = \langle\!\langle \boldsymbol{Y}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}^*$,

(iii) *definiteness*: $\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle_{\mathbb{S}}$ is positive definite whenever $\boldsymbol{X}$ has full rank s.

A “block norm”

Definition

A mapping $N_{\mathbb{S}} : \mathbb{C}^{n\times s} \to \mathbb{S}$ is called a *normalizing quotient onto* $\mathbb{S}$ if for every $\boldsymbol{X} \in \mathbb{C}^{n\times s}$, $\boldsymbol{X} \neq 0$, there exists $\boldsymbol{Y} \in \mathbb{C}^{n\times s}$ such that $\boldsymbol{X} = \boldsymbol{Y} N_{\mathbb{S}}(\boldsymbol{X})$ and $\langle\langle \boldsymbol{Y}, \boldsymbol{Y} \rangle\rangle_{\mathbb{S}} = I_s$.

A “block norm”

Definition

A mapping $N_{\mathbb{S}} : \mathbb{C}^{n \times s} \to \mathbb{S}$ is called a *normalizing quotient onto* $\mathbb{S}$ if for every $\boldsymbol{X} \in \mathbb{C}^{n \times s}$, $\boldsymbol{X} \neq 0$, there exists $\boldsymbol{Y} \in \mathbb{C}^{n \times s}$ such that $\boldsymbol{X} = \boldsymbol{Y} N_{\mathbb{S}}(\boldsymbol{X})$ and $\langle\!\langle \boldsymbol{Y}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}} = I_s$.

Lemma

$N_{\mathbb{S}}(\boldsymbol{X})$ always exists via QR of $\boldsymbol{X}$.

Giving the natural choice: $N_{\mathbb{S}}(\boldsymbol{X}) = R$ where R is the (upper triangular) matrix of the Cholesky factorization $\langle\!\langle \boldsymbol{X}, \boldsymbol{X} \rangle\!\rangle = R^* R$

Some examples[1]

- **classical**: $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Cl}} = \boldsymbol{X}^* \boldsymbol{Y}$, and $\mathbb{S} = \mathbb{C}^{s \times s}$. $N_{\mathbb{S}}^{\mathrm{Cl}}(\boldsymbol{X}) = R$, where $\boldsymbol{Q}R = \boldsymbol{X}$ is an economical QR factorization
- **global**: $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Gl}} = \mathrm{trace}(\boldsymbol{X}^* \boldsymbol{Y}) I_s$, and $\mathbb{S}$ is set of diagonal matrices with constant entry on the diagonal.

$$N_{\mathbb{S}}^{\mathrm{Gl}}(\boldsymbol{X}) = \|\boldsymbol{X}\|_{\mathrm{F}}\, I_s.$$

- **loop-interchange**: $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}}^{\mathrm{Li}} = \mathrm{diag}(\boldsymbol{X}^* \boldsymbol{Y})$, and $\mathbb{S}$ is the set of diagonal matrices.

$$N_{\mathbb{S}}^{\mathrm{Li}}(\boldsymbol{X}) = \begin{bmatrix} \|\boldsymbol{x}_1\|_2 & & & \\ & \|\boldsymbol{x}_2\|_2 & & \\ & & \ddots & \\ & & & \|\boldsymbol{x}_s\|_2 \end{bmatrix}.$$

[1] First identified by Elbouyahyaoui, Messaoudi, and Sadok in their 2009 paper "Algebraic properties of the block GMRES and block Arnoldi methods"

Block orthogonality, block normalization, and block span

- $\boldsymbol{X}$ and $\boldsymbol{Y}$ are *block orthogonal* if $\langle\!\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\!\rangle_{\mathbb{S}} = 0_s$
- $\boldsymbol{X}$ is *block normalized* if $N_{\mathbb{S}}(\boldsymbol{X}) = I_s$
- A set of block vectors $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m\}$ is *block orthonormal* when $\langle\!\langle \boldsymbol{X}_i, \boldsymbol{X}_j \rangle\!\rangle_{\mathbb{S}} = \delta_{ij} I_s$, for all $i, j = 1, \ldots, m$
- *Block span*[2]:

$$\operatorname{span}^{\mathbb{S}}\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m\} := \left\{ \sum_{k=1}^{m} \boldsymbol{X}_k C_k : C_k \in \mathbb{S} \right\}$$

The mth block Krylov subspace with respect to A, $\boldsymbol{B}$, and the block inner-product Φ as

$$\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B}) := \operatorname{span}^{\mathbb{S}}\{\boldsymbol{B}, A\boldsymbol{B}, \ldots, A^{m-1}\boldsymbol{B}\} \leq \mathbb{C}^{n \times s}$$

[2] Inspired by Gutknecht's 2007 paper "Block Krylov space methods for linear systems with multiple right-hand sides: an introduction"

Matrix polynomials

$$\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B}) = \operatorname{span}^{\mathbb{S}}\{\boldsymbol{B}, A\boldsymbol{B}, \ldots, A^{m-1}\boldsymbol{B}\} = \left\{ \sum_{k=0}^{m-1} A^k \boldsymbol{B} C_k : C_k \in \mathbb{S} \right\}$$

Let $\mathbb{P}_{m-1}(\mathbb{S})$ denote the space of $m-1$ degree polynomials with coefficients in $\mathbb{S}$. Three possibilities:

- [Gohberg et al., 2009]: $P(\Lambda) = \sum_{k=0}^{m-1} \Lambda^k C_k$, with $\Lambda \in \mathbb{C}^{s \times s}$
- [Lancaster, 1966]: $P(\lambda) = \sum_{k=0}^{m-1} \lambda^k C_k$, with $\lambda \in \mathbb{C}$
- [Kent, 1989]: $P(A) \circ \boldsymbol{B} := \sum_{k=0}^{m-1} A^k \boldsymbol{B} C_k$

Matrix polynomials

$$\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B}) = \operatorname{span}^{\mathbb{S}}\{\boldsymbol{B}, A\boldsymbol{B}, \ldots, A^{m-1}\boldsymbol{B}\} = \left\{ \sum_{k=0}^{m-1} A^k \boldsymbol{B} C_k : C_k \in \mathbb{S} \right\}$$

Let $\mathbb{P}_{m-1}(\mathbb{S})$ denote the space of $m-1$ degree polynomials with coefficients in $\mathbb{S}$. Three possibilities:

- [Gohberg et al., 2009]: $P(\Lambda) = \sum_{k=0}^{m-1} \Lambda^k C_k$, with $\Lambda \in \mathbb{C}^{s \times s}$
- [Lancaster, 1966]: $P(\lambda) = \sum_{k=0}^{m-1} \lambda^k C_k$, with $\lambda \in \mathbb{C}$
- [Kent, 1989]: $P(A) \circ \boldsymbol{B} := \sum_{k=0}^{m-1} A^k \boldsymbol{B} C_k$

$$\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B}) = \{P(A) \circ \boldsymbol{B} : P \in \mathbb{P}_{m-1}(\mathbb{S})\}$$

The block Arnoldi algorithm

Given: A, $\boldsymbol{B}$, $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$, $N_{\mathbb{S}}$, m

Compute $B = N_{\mathbb{S}}(\boldsymbol{B})$ and $\boldsymbol{V}_1$ such that $\boldsymbol{V}_1 B = \boldsymbol{B}$ and $\langle\!\langle \boldsymbol{V}_1, \boldsymbol{V}_1 \rangle\!\rangle_{\mathbb{S}} = I_s$

for $k = 1, \dots, m$ **do**

  Compute $\boldsymbol{W} = A \boldsymbol{V}_k$

  for $j = 1, \dots, k$ **do**

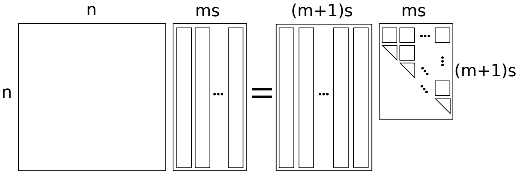
    $H_{j,k} = \langle\!\langle \boldsymbol{V}_j, \boldsymbol{W} \rangle\!\rangle_{\mathbb{S}}$

    $\boldsymbol{W} = \boldsymbol{W} - \boldsymbol{V}_j H_{j,k}$

  Compute $H_{k+1,k} = N_{\mathbb{S}}(\boldsymbol{W})$ and $\boldsymbol{V}_{k+1}$ such that $\boldsymbol{V}_{k+1} H_{k+1,k} = W$ and $\langle\!\langle \boldsymbol{V}_k, \boldsymbol{V}_k \rangle\!\rangle_{\mathbb{S}} = I_s$

Return B, $\boldsymbol{\mathcal{V}}_m = [\boldsymbol{V}_1 | \dots | \boldsymbol{V}_m]$, $\mathcal{H}_m = (H_{j,k})_{j,k=1}^m$, $\boldsymbol{V}_{m+1}$, and $H_{m+1,m}$

The block Arnoldi relation



$$A\boldsymbol{\mathcal{V}}_m = \boldsymbol{\mathcal{V}}_m \mathcal{H}_m + \boldsymbol{V}_{m+1} H_{m+1,m} \widehat{\boldsymbol{E}}_m^*,$$

where

- $\widehat{\boldsymbol{E}}_j = \widehat{\boldsymbol{e}}_j \otimes I_s$, and $\widehat{\boldsymbol{e}}_j$ is the jth standard unit vector
- $\boldsymbol{\mathcal{V}}_m = [\boldsymbol{V}_1 | \ldots | \boldsymbol{V}_m] \in \mathbb{C}^{n \times ms}$ and $\{\boldsymbol{V}_k\}_{k=1}^m$ is an $\mathbb{S}$-orthonormal basis spanning $\mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$
- $\mathcal{H}_m$ is block upper Hessenberg, with each $H_{j,k} \in \mathbb{S} \subseteq \mathbb{C}^{s \times s}$, and $H_{k+1,k}$ upper triangular

Block Full Orthogonalization Methods (BFOM)

The m-th BFOM approximation to $A\boldsymbol{X} = \boldsymbol{B}$ is the block vector $\boldsymbol{X}_m \in \mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$ is defined to satisfy the following Galerkin condition:

$$\boldsymbol{B} - A\boldsymbol{X}_m \perp_{\langle\langle \cdot,\cdot \rangle\rangle_{\mathbb{S}}} \mathscr{K}_m^{\mathbb{S}}(A, \boldsymbol{B})$$

Let A be HPD and $\boldsymbol{E}_m := \boldsymbol{X}_* - \boldsymbol{X}_m$, where $\boldsymbol{X}_*$ solves $A\boldsymbol{X} = \boldsymbol{B}$ exactly. Some new tools:

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_{\mathbb{S}} := \operatorname{trace}(\langle\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\rangle_{\mathbb{S}})$$

$$\|\boldsymbol{X}\|_{\mathbb{S}} := \langle \boldsymbol{X}, \boldsymbol{X} \rangle_{\mathbb{S}}^{\frac{1}{2}}$$

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_{\mathbb{S}\text{-}A} := \langle \boldsymbol{X}, A\boldsymbol{Y} \rangle_{\mathbb{S}}$$

Block Full Orthogonalization Methods (BFOM)

$$\kappa := \frac{\lambda_{\max}}{\lambda_{\min}} \quad c := \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \quad \xi_m := \frac{1}{\cosh(m \ln c)} \leq \frac{2}{c^m + c^{-m}}$$

Theorem

For a Hermitian positive definite matrix $A \in \mathbb{C}^{n \times n}$ and block right-hand-side vector $\boldsymbol{B} \in \mathbb{C}^{n \times s}$, the BFOM error $\boldsymbol{E}_m = \boldsymbol{X}_* - \boldsymbol{X}_m$ for $A\boldsymbol{X} = \boldsymbol{B}$ has the following bound:

$$\|\boldsymbol{E}_m\|_{\mathbb{S}\text{-}A} \leq \xi_m \|\boldsymbol{E}_0\|_{\mathbb{S}\text{-}A}.$$

Block Full Orthogonalization Methods (BFOM)

$$\kappa := \frac{\lambda_{\max}}{\lambda_{\min}} \quad c := \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \quad \xi_m := \frac{1}{\cosh(m \ln c)} \leq \frac{2}{c^m + c^{-m}}$$

Theorem

For a Hermitian positive definite matrix $A \in \mathbb{C}^{n \times n}$ and block right-hand-side vector $\boldsymbol{B} \in \mathbb{C}^{n \times s}$, the BFOM error $\boldsymbol{E}_m = \boldsymbol{X}_* - \boldsymbol{X}_m$ for $A\boldsymbol{X} = \boldsymbol{B}$ has the following bound:

$$\|\boldsymbol{E}_m\|_{\mathbb{S}\text{-}A} \leq \xi_m \|\boldsymbol{E}_0\|_{\mathbb{S}\text{-}A}.$$

For ClBFOM, GlBFOM, and LiBFOM, it is already well known that

$$\|\boldsymbol{E}_m\|_{A\text{-F}} \leq \xi_m \|\boldsymbol{E}_0\|_{A\text{-F}},$$

where $\|\boldsymbol{X}\|_{A\text{-F}} = \sqrt{\mathrm{trace}(\boldsymbol{X}^* A \boldsymbol{X})}$

The B(FOM)^2 approximation

Block full orthogonalization method for functions of matrices: B(FOM)^2

- For general f, we define the B(FOM)^2 approximation as follows:

$$\boldsymbol{F}_m := \boldsymbol{\mathcal{V}}_m f(\mathcal{H}_m) \widehat{\boldsymbol{E}}_1 B$$

The B(FOM)^2 approximation

Block full orthogonalization method for functions of matrices: B(FOM)^2

- For general f, we define the B(FOM)^2 approximation as follows:

$$\boldsymbol{F}_m := \boldsymbol{\mathcal{V}}_m f(\mathcal{H}_m) \widehat{\boldsymbol{E}}_1 B$$

- With restarts:

$$\boldsymbol{F}_m^{(k)} := \boldsymbol{F}_m^{(k-1)} + \boldsymbol{\mathcal{V}}_m^{(k)} \Delta_m^{(k-1)}(\mathcal{H}_m^{(k)}) \circ \widehat{\boldsymbol{E}}_1,$$

where the error function $\Delta_m^{(k-1)}(z)$ is such that

$$\Delta_m^{(k-1)}(A) \circ \boldsymbol{B} := f(A)\boldsymbol{B} - \boldsymbol{F}_m^{(k-1)}$$

B(FOM)^2(m) algorithm

Given: f, A, $\boldsymbol{B}$, $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$, $N_{\mathbb{S}}$, m, `tol`

1 Run Block Arnoldi with inputs A, $\boldsymbol{B}$, $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$, $N_{\mathbb{S}}$, and m to obtain $\boldsymbol{\mathcal{V}}_m^{(1)}$, $\mathcal{H}_m^{(1)}$, $H_{m+1,m}^{(1)}$, $\boldsymbol{V}_{m+1}^{(1)}$, and B

2 Compute $\boldsymbol{F}_m^{(1)} := \boldsymbol{\mathcal{V}}_m^{(1)} f\left(\mathcal{H}_m^{(1)}\right) \widehat{\boldsymbol{E}}_1 B$

3 **for** $k = 2, 3, \ldots,$ *until convergence* **do**

4 Determine $C_m^{(k-1)}(t)$ to define the new error function $\Delta_m^{(k-1)}(z)$

5 Run Block Arnoldi with inputs A, $\boldsymbol{V}_{m+1}^{(k-1)}$, $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{S}}$, $N_{\mathbb{S}}$, and m to obtain $\boldsymbol{\mathcal{V}}_m^{(k)}$, $\mathcal{H}_m^{(k)}$, $H_{m+1,m}^{(k)}$, and $\boldsymbol{V}_{m+1}^{(k)}$

6 Compute $\widetilde{\boldsymbol{D}}_m^{(k-1)} := \boldsymbol{\mathcal{V}}_m^{(k)} \Delta_m^{(k-1)}\left(\mathcal{H}_m^{(k)}\right) \circ \widehat{\boldsymbol{E}}_1$, where $\Delta_m^{(k-1)}(z)$ is evaluated via quadrature

7 Compute $\boldsymbol{F}_m^{(k)} := \boldsymbol{F}_m^{(k-1)} + \widetilde{\boldsymbol{D}}_m^{(k-1)}$

Convergence of $\mathrm{B(FOM)}^2(m)$

Recall f is a Stieltjes function: $f(z) = \int_0^\infty \frac{1}{z+t} \,\mathrm{d}\mu(t)$

$\kappa(t) := \frac{\lambda_{\max}(A+tI)}{\lambda_{\min}(A+tI)}, \quad c(t) := \frac{\sqrt{\kappa(t)}-1}{\sqrt{\kappa(t)}+1}, \quad \xi_m(t) := \frac{1}{\cosh(m \ln c(t))} \leq \frac{2}{c^m + c^{-m}}$

Theorem

Let A be an HPD matrix, $\boldsymbol{B} \in \mathbb{C}^{n \times s}$, and f a Stieltjes function. Let $t_0 \geq 0$ denote the minimum of the support of μ. Then the $\mathbb{S}$-A norm of the error of $\boldsymbol{F}_m^{(k)}$ has the following bound:

$$\left\| f(A)\boldsymbol{B} - \boldsymbol{F}_m^{(k)} \right\|_{\mathbb{S}\text{-}A} \leq \gamma \xi_m(t_0)^k,$$

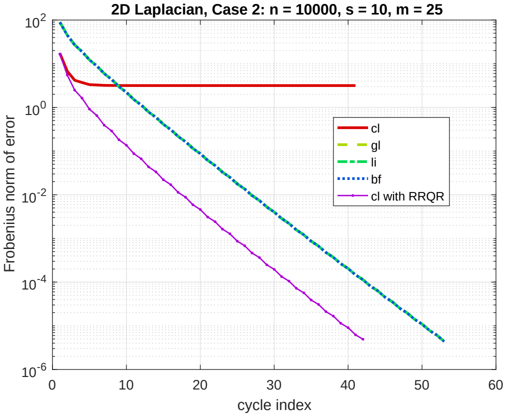
where $\gamma = \|\boldsymbol{B}\|_{\mathbb{S}} \sqrt{\lambda_{\max}} f(\sqrt{\lambda_{\min}\lambda_{\max}})$. In particular, B(FOM)^2(m) converges for all cycle lengths m as the restart index $k \to \infty$.

Comments on implementation

- Block Arnoldi algorithm
  - One should implement block inner product in a "smart" way
  - One must account for different kinds of break-downs for each block inner product
- Error function
  - An analytic form of the error function $\Delta_m^{(k-1)}$ must be written a priori coming from the integral representation of f.
  - The quadrature rule (used to evaluate $\Delta_m^{(k-1)}$) and tolerance must be chosen carefully
- Code written in Matlab and executed on a Dell desktop with a Linux 64-bit operating system, an Intel®Core™ i7-4770 CPU @ 3.40 GHz, and 32 GB of RAM

A discretized 2D Laplacian, $f(z) = z^{-1/2}$

Convergence history for $A^{-1/2}\boldsymbol{B}$, where $A \in \mathbb{C}^{10{,}000\times 10{,}000}$ is the discretized two-dimensional Laplacian, $\boldsymbol{B} \in \mathbb{C}^{10{,}000\times 10}$ is rank deficient, the cycle length is $m = 25$, and the error tolerance is 5e-6



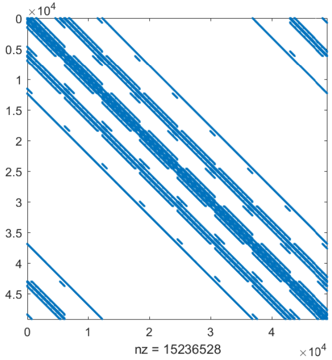
A discretized 2D Laplacian, $f(z) = z^{-1/2}$

Results for $A^{-1/2}\boldsymbol{B}$, where $A \in \mathbb{C}^{10{,}000 \times 10{,}000}$ is the discretized two-dimensional Laplacian, $\boldsymbol{B} \in \mathbb{C}^{10{,}000 \times 10}$ is rank deficient, the cycle length is $m = 25$, and the error tolerance is 5e-6

	wall time (s)	**number of cycles**	**true error ($\lVert\cdot\rVert_F$)**
CIB(FOM)$^2(m)$	–	–	–
GIB(FOM)$^2(m)$	10.31	53	4.41E-06
LiB(FOM)$^2(m)$	93.53	53	4.37E-06
(FOM)$^2(m)$	87.39	53	4.37E-06
CIB(FOM)$^2(m)$ with deflation	353	42	4.92E-06

Lattice Quantum Chromodynamics (QCD), $f(z) = \text{sign}(z) = (z^2)^{-1/2} z$

$Q \in \mathbb{C}^{49,152 \times 49,152}$ represents a periodic nearest-neighbor coupling on the lattice. Overlap Dirac operator requires $\text{sign}(Q)\boldsymbol{B}$. $\boldsymbol{B}$ random, $s = 12$, error tolerance set to 5e-6. Sparsity pattern of Q^2 given below.



Lattice Quantum Chromodynamics (QCD), $f(z) = \text{sign}(z) = (z^2)^{-1/2} z$

$Q \in \mathbb{C}^{49{,}152 \times 49{,}152}$ represents a periodic nearest-neighbor coupling on the lattice. Overlap Dirac operator requires $\text{sign}(Q)\boldsymbol{B}$. $\boldsymbol{B}$ random, error tolerance set to 5e-6.

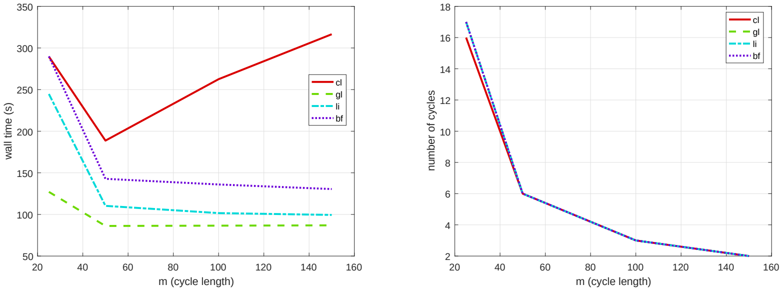


Figure: (left) Total time versus the cycle length. (right) Number of cycles versus the cycle length.

Conclusions

- New theory of bilinear form amenable to build block Arnoldi/Lanczos bases
- Formulation and convergence results for (restarted) BFOM for functions of matrices (for A HPD and f Stieltjes)
- Multiple experiments showing that block global is either faster with Block classical (or of comparable performance), and both much faster than loop-interchange

Conclusions

- New theory of bilinear form amenable to build block Arnoldi/Lanczos bases
- Formulation and convergence results for (restarted) BFOM for functions of matrices (for A HPD and f Stieltjes)
- Multiple experiments showing that block global is either faster with Block classical (or of comparable performance), and both much faster than loop-interchange

Look for research report in a couple of weeks at `math.temple.edu/szyld`