PDEs, Matrix Functions and Krylov Subspace Methods

Oliver Ernst

Institut für Numerische Mathematik und Optimierung
TU Bergakademie Freiberg, Germany

LMS Durham Symposium

Computational Linear Algebra for Partial Differential Equations
July 14–24, 2008

Deutsche Forschungsgemeinschaft DFG

BERGAKADEMIE TECHNISCHE UNIVERSITÄT FREIBERG

Collaborators

- Michael Eiermann, Martin Afanasjew, Stefan Güttel
  TU Bergakademie Freiberg
  Institute of Numerical Analysis and Optimization
- Ralph-Uwe Börner, Klaus Spitzer
  TU Bergakademie Freiberg
  Institute of Geophysics
- Bernhard Beckermann
  Labo Painlevé
  UST Lille

Outline

Outline

Initial Value Problems

By the variation-of-constants formula the solution of the IVP

$$\dot{\boldsymbol{u}} = \boldsymbol{A}\boldsymbol{u} + \boldsymbol{g}, \quad \boldsymbol{u}(t_0) = \boldsymbol{u}_0, \qquad \boldsymbol{A} \in \mathbb{C}^{N\times N}; \boldsymbol{g}, \boldsymbol{u}_0 \in \mathbb{C}^N,$$

is given by

$$\boldsymbol{u}(t) = e^{(t-t_0)\boldsymbol{A}}\boldsymbol{u}_0 + (t-t_0)\,\varphi_1\big((t-t_0)\boldsymbol{A}\big)\boldsymbol{g}, \qquad t > t_0,$$

with the "Phi-function"

$$\varphi_1(z) = \frac{e^z - 1}{z}.$$

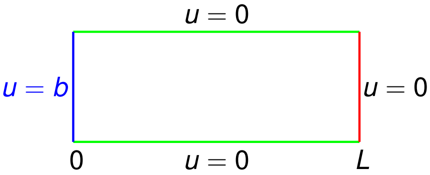
Such relations are the basis of exponential integrators, which address stiffness in ODE systems (in particular MOL semi-discretizations) by explicitly evaluating the action of $e^{\boldsymbol{A}}$ or $\varphi_1(\boldsymbol{A})$ on a vector.

[Hocbruck et al. (1998)], [Minchev & Wright (2005)], [Schmelzer (2007)].

Outline

Dirichlet-Neumann Maps (1)

$$
\begin{aligned}
Au(x) - u_{xx} &= 0, \quad x \in (0, L), \qquad L > 0\\
-u_x(0) &= b,\\
u(L) &= 0.
\end{aligned}
$$

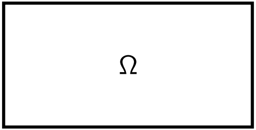


Mapping which assigns $b \mapsto u(0)$ (Neumann-Dirichlet map, impedance function) given by

$$
u(0) = f(A)b, \quad f(z) = \begin{cases} \frac{1}{\sqrt{z}}, & L = \infty, \\ \frac{\tanh(L\sqrt{z})}{\sqrt{z}}, & L < \infty. \end{cases}
$$

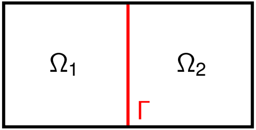
[Druskin & Knizhnerman (1999)]

Model problem



$$\begin{aligned} -\Delta u = f \quad &\text{on} \quad \Omega, \\ u = 0 \quad &\text{on } \partial\Omega \end{aligned}$$

may be reformulated as ($i = 1, 2$)



$$\begin{aligned} -\Delta u_i = f \quad &\text{on} \quad \Omega_i, \\ u_i = 0 \quad &\text{on } \partial\Omega_i \setminus \Gamma \\ \partial_n u_i = S u_i \quad &\text{on } \Gamma \end{aligned}$$

in terms of Dirichlet-Neumann mapping (Steklov-Poincaré operator) $S : H_{00}^{1/2}(\Gamma) \to H_{00}^{-1/2}(\Gamma)$. A spectrally equivalent preconditioner to S is given by $\boldsymbol{M}(\boldsymbol{M}^{-1}\boldsymbol{L})^{1/2}$, where $\boldsymbol{M}$ and $\boldsymbol{L}$ are Galerkin mass and stiffness matrices for basis functions restricted to Γ. [Arioli & Loghin (2008)].

Outline

Stochastic Differential Equations

Certain problems in population dynamics and neutron transport lead to Itô differential equations

$$d\boldsymbol{y}(t) = \boldsymbol{f}(t, \boldsymbol{y}(t))\, dt + \boldsymbol{A}^{1/2}(t, \boldsymbol{y}(t))\, d\boldsymbol{W}(t), \qquad \boldsymbol{y}(t_0) = \boldsymbol{y}_0,$$

with $\boldsymbol{f}$ and $\boldsymbol{A}$ known vector and matrix-valued functions and $\boldsymbol{W}(t)$ a (vector) Wiener process.
Approximation using the Euler-Maruyama method results in the iteration

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \Delta t\, \boldsymbol{f}(t_n, \boldsymbol{y}_n) + \sqrt{\Delta t}\boldsymbol{A}^{1/2}(t_n, \boldsymbol{y}_n)\omega_n$$

with ω_n sampled from a multivariate normal distribution.
[Allen, Baglama & Boyd (2000)]

Outline

1. Matrix Functions and Differential Equations
   - Initial Value Problems
   - Dirichlet-Neumann Maps
   - Stochastic Differential Equations
   - Frequency Domain Model Reduction

2. Krylov Subspace Approximation
   - Algorithm
   - Restarting
   - Convergence
   - A Posteriori Error Estimation

Frequency Domain Model Reduction

Time-dependent Maxwell's equations on a bounded domain Ω

$$\partial_t(\sigma \boldsymbol{E}) + \nabla\times(\mu\,\nabla\times \boldsymbol{E}) = -\partial_t \boldsymbol{J}^{(i)}, \quad \boldsymbol{n}\times\boldsymbol{E} = \boldsymbol{0} \text{ on } \partial\Omega, \quad \boldsymbol{E}(t_0) = \boldsymbol{E}_0.$$

Instead of MOL-discretization, switch to frequency domain

$$\nabla\times(\mu\,\nabla\times \boldsymbol{E}) + i\omega\sigma\boldsymbol{E} = \boldsymbol{q}, \qquad \boldsymbol{n}\times\boldsymbol{E} = 0 \text{ on } \partial\Omega$$

for $\omega \in [\omega_{\min}, \omega_{\max}]$. FE discretization in space gives

$$(\boldsymbol{K} + i\omega\boldsymbol{M})\boldsymbol{u} = \boldsymbol{q}, \qquad \omega \in [\omega_{\min}, \omega_{\max}].$$

If solution of interest only at p locations (receiver locations), introduce restriction matrix $\boldsymbol{R}$ and evaluate

$$\boldsymbol{f}(\omega) = \boldsymbol{R}^\top(\boldsymbol{K} + i\omega\boldsymbol{M})^{-1}\boldsymbol{q}, \qquad \omega \in [\omega_{\min}, \omega_{\max}].$$

[Börner, E. & Spitzer (2008)].

Outline

1. Matrix Functions and Differential Equations
   - Initial Value Problems
   - Dirichlet-Neumann Maps
   - Stochastic Differential Equations
   - Frequency Domain Model Reduction

2. Krylov Subspace Approximation
   - Algorithm
   - Restarting
   - Convergence
   - A Posteriori Error Estimation

Krylov Subspace Approximation of $f(\boldsymbol{A})\boldsymbol{b}$

Given	$\boldsymbol{A} \in \mathbb{C}^{n \times n}$,
	$f : D \to \mathbb{C}$ analytic, $W(\boldsymbol{A}) \subset D$,
	$\boldsymbol{b} \in \mathbb{C}^n$, $\|\boldsymbol{b}\| = 1$,
compute	$f(\boldsymbol{A})\boldsymbol{b}$.

Approximate in Kylov subspace

$$f(\boldsymbol{A})\boldsymbol{b} \approx \boldsymbol{f}_m \in \mathscr{K}_m(\boldsymbol{A}, \boldsymbol{b}) = \{\boldsymbol{v} = p(\boldsymbol{A})\boldsymbol{b} : p \in \mathscr{P}_{m-1}\}, \qquad m = 1, 2, \dots$$

Basic Algorithm

Arnoldi-like decomposition

$$\begin{aligned}\boldsymbol{A}\boldsymbol{V}_m = \boldsymbol{V}_m\boldsymbol{H}_m + h_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^\top \\ \operatorname{ran}(\boldsymbol{V}_m) = \mathscr{K}_m(\boldsymbol{A},\boldsymbol{b}),\ \ \boldsymbol{V}_m^H\boldsymbol{V}_m = \boldsymbol{I} \\ \boldsymbol{b} = \boldsymbol{V}_m\boldsymbol{e}_1, \\ \boldsymbol{H}_m \text{ unreduced upper Hessenberg}\end{aligned}$$

Approximant

$$\boldsymbol{f}_m := \boldsymbol{V}_m f(\boldsymbol{H}_m)\boldsymbol{e}_1 = \boldsymbol{V}_m f(\boldsymbol{H}_m)\boldsymbol{V}_m^H\boldsymbol{b}.$$

- Requires evaluation of (first column of) $f(\boldsymbol{H}_m)$ for small dense matrix $\boldsymbol{H}_m$.
- Simplification: $\boldsymbol{H}_m$ Hermitian tridiagonal for $\boldsymbol{A}$ Hermitian (Hermitian Lanczos process).

Three Interpretations

- **Subspace approximation.** $\boldsymbol{H}_m = \boldsymbol{V}_m^H \boldsymbol{A} \boldsymbol{V}_m$ represents $\boldsymbol{A}$ on $\mathscr{K}_m(\boldsymbol{A}, \boldsymbol{b})$ w.r.t. $\boldsymbol{V}_m$. Approximate $f(\boldsymbol{A})$ with $f(\boldsymbol{H}_m)$ there.

- **Cauchy integral.** For a contour Γ with $W(\boldsymbol{A}) \subset \operatorname{int} \Gamma$,

$$f(\boldsymbol{A})\boldsymbol{b} = \frac{1}{2\pi i} \int_\Gamma f(\lambda)(\lambda \boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{b} \, d\lambda$$
$$\approx \frac{1}{2\pi i} \int_\Gamma f(\lambda) \underbrace{\boldsymbol{V}_m(\lambda \boldsymbol{I} - \boldsymbol{H}_m)^{-1} \boldsymbol{V}_m^H \boldsymbol{b}}_{=:\boldsymbol{x}_m(\lambda)} \, d\lambda = \boldsymbol{V}_m f(\boldsymbol{H}_m) \boldsymbol{e}_1.$$

$\boldsymbol{x}_m(\lambda)$: Galerkin approx. of $\boldsymbol{x}(\lambda) := (\lambda \boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{b}$ w.r.t. $\mathscr{K}_m(\boldsymbol{A}, \boldsymbol{b})$.

- **Interpolation.** If $p \in \mathscr{P}_{m-1}$ Hermite-interpolates f at nodes $\Lambda(\boldsymbol{H}_m)$, then

$$f(\boldsymbol{A})\boldsymbol{b} \approx p(\boldsymbol{A})\boldsymbol{b} = \boldsymbol{V}_m p(\boldsymbol{H}_m) \boldsymbol{e}_1 = \boldsymbol{V}_m f(\boldsymbol{H}_m) \boldsymbol{e}_1.$$

Three Interpretations

- **Subspace approximation.** $\boldsymbol{H}_m = \boldsymbol{V}_m^H \boldsymbol{A} \boldsymbol{V}_m$ represents $\boldsymbol{A}$ on $\mathscr{K}_m(\boldsymbol{A}, \boldsymbol{b})$ w.r.t. $\boldsymbol{V}_m$. Approximate $f(\boldsymbol{A})$ with $f(\boldsymbol{H}_m)$ there.

- **Cauchy integral.** For a contour Γ with $W(\boldsymbol{A}) \subset \operatorname{int}\Gamma$,

$$\begin{aligned} f(\boldsymbol{A})\boldsymbol{b} &= \frac{1}{2\pi i} \int_\Gamma f(\lambda)(\lambda \boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{b} \, d\lambda \\ &\approx \frac{1}{2\pi i} \int_\Gamma f(\lambda) \underbrace{\boldsymbol{V}_m(\lambda \boldsymbol{I} - \boldsymbol{H}_m)^{-1} \boldsymbol{V}_m^H \boldsymbol{b}}_{=: \boldsymbol{x}_m(\lambda)} \, d\lambda = \boldsymbol{V}_m f(\boldsymbol{H}_m) \boldsymbol{e}_1. \end{aligned}$$

  $\boldsymbol{x}_m(\lambda)$: Galerkin approx. of $\boldsymbol{x}(\lambda) := (\lambda \boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{b}$ w.r.t. $\mathscr{K}_m(\boldsymbol{A}, \boldsymbol{b})$.

- **Interpolation.** If $p \in \mathscr{P}_{m-1}$ Hermite-interpolates f at nodes $\Lambda(\boldsymbol{H}_m)$, then

$$f(\boldsymbol{A})\boldsymbol{b} \approx p(\boldsymbol{A})\boldsymbol{b} = \boldsymbol{V}_m p(\boldsymbol{H}_m) \boldsymbol{e}_1 = \boldsymbol{V}_m f(\boldsymbol{H}_m) \boldsymbol{e}_1.$$

Three Interpretations

- **Subspace approximation.** $\boldsymbol{H}_m = \boldsymbol{V}_m^H \boldsymbol{A} \boldsymbol{V}_m$ represents $\boldsymbol{A}$ on $\mathscr{K}_m(\boldsymbol{A}, \boldsymbol{b})$ w.r.t. $\boldsymbol{V}_m$. Approximate $f(\boldsymbol{A})$ with $f(\boldsymbol{H}_m)$ there.

- **Cauchy integral.** For a contour Γ with $W(\boldsymbol{A}) \subset \operatorname{int}\Gamma$,

$$
\begin{aligned}
f(\boldsymbol{A})\boldsymbol{b} &= \frac{1}{2\pi i}\int_\Gamma f(\lambda)(\lambda \boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{b}\, d\lambda \\
&\approx \frac{1}{2\pi i}\int_\Gamma f(\lambda)\underbrace{\boldsymbol{V}_m(\lambda \boldsymbol{I} - \boldsymbol{H}_m)^{-1}\boldsymbol{V}_m^H\boldsymbol{b}}_{=:\boldsymbol{x}_m(\lambda)}\, d\lambda = \boldsymbol{V}_m f(\boldsymbol{H}_m)\boldsymbol{e}_1.
\end{aligned}
$$

  $\boldsymbol{x}_m(\lambda)$: Galerkin approx. of $\boldsymbol{x}(\lambda) := (\lambda \boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{b}$ w.r.t. $\mathscr{K}_m(\boldsymbol{A}, \boldsymbol{b})$.

- **Interpolation.** If $p \in \mathscr{P}_{m-1}$ Hermite-interpolates f at nodes $\Lambda(\boldsymbol{H}_m)$, then

$$
f(\boldsymbol{A})\boldsymbol{b} \approx p(\boldsymbol{A})\boldsymbol{b} = \boldsymbol{V}_m p(\boldsymbol{H}_m)\boldsymbol{e}_1 = \boldsymbol{V}_m f(\boldsymbol{H}_m)\boldsymbol{e}_1.
$$

A Key Relation

For Arnoldi(-like) decomposition of $\mathscr{K}_m(\boldsymbol{A}, \boldsymbol{b})$

$$\boldsymbol{A}\boldsymbol{V}_m = \boldsymbol{V}_m\boldsymbol{H}_m + h_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^\top,$$

denote $\gamma_m := \prod_{j=1}^m h_{j+1,j}$.

For any polynomial $p \in \mathscr{P}_{m-1}$ there holds

$$p(\boldsymbol{A})\boldsymbol{b} = \boldsymbol{V}_m\, p(\boldsymbol{H}_m)\boldsymbol{e}_1$$

and, for $p \in \mathscr{P}_m$ with leading coefficient α_m,

$$p(\boldsymbol{A})\boldsymbol{b} = \boldsymbol{V}_m\, p(\boldsymbol{H}_m)\boldsymbol{e}_1 + \alpha_m\gamma_m\boldsymbol{v}_{m+1}.$$

[Druskin & Knizhnerman (1989)], [Saad (1992)], [Paige & al. (1995)].

Outline

Restarting

- For large Krylov spaces storage and computation for Arnoldi process too expensive.
- Remedy: periodically restart Arnoldi process with new initial vector.
- Short recurrences for Arnoldi/Lanczos don't carry over to approximation; two-pass algorithm another option.
- Difficulties: no residual vector, recursive update of approximation.
- Restarting method based on divided differences.

Divided Differences

For function f, nodes $\vartheta_1, \dots, \vartheta_m \in \mathbb{C}$, denote by

$$w_m(z) := \prod_{j=1}^{m} (z - \vartheta_j) \quad \text{nodal polynomial,}$$

$$I_{w_m} f \in \mathscr{P}_{m-1} \quad \text{Hermite interpolant to } f \text{ at } \{\vartheta_j\}_{j=1}^m,$$

$$\Delta_{w_m} f := \frac{f - I_{w_m} f}{w_m} \quad m\text{-th order divided difference of } f \text{ w.r.t. } w_m.$$

Then $f = I_{w_m} f + \Delta_{w_m} f \cdot w_m,$

$$\begin{aligned} f(\boldsymbol{A})\boldsymbol{b} &= [I_{w_m} f](\boldsymbol{A})\boldsymbol{b} + [\Delta_{w_m} f](\boldsymbol{A})\, w_m(\boldsymbol{A})\boldsymbol{b} \\ &= \boldsymbol{V}_m [I_{w_m} f](\boldsymbol{H}_m)\boldsymbol{e}_1 + [\Delta_{w_m} f](\boldsymbol{A})(\boldsymbol{V}_m \underbrace{w_m(\boldsymbol{H}_m)}_{=O} \boldsymbol{e}_1 + \gamma_m \boldsymbol{v}_{m+1}) \\ &= \boldsymbol{f}_m + \gamma_m [\Delta_{w_m} f](\boldsymbol{A}) \boldsymbol{v}_{m+1}. \end{aligned}$$

General Error Representation

Theorem (Eiermann & E., 2006)

Given a function f, matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$, vector $\boldsymbol{b} \in \mathbb{C}^n$, and the Arnoldi decomposition $\boldsymbol{A}\boldsymbol{V}_m = \boldsymbol{V}_m\boldsymbol{H}_m + h_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^\top$, then the error of the Krylov subspace approximation $\boldsymbol{f}_m$ of $f(\boldsymbol{A})\boldsymbol{b}$ is given by

$$f(\boldsymbol{A})\boldsymbol{b} - \boldsymbol{f}_m = g(\boldsymbol{A})\boldsymbol{v}_{m+1}, \tag{1}$$

where $g(z) = \gamma_m[\Delta_{w_m} f](z)$ and $w_m \in \mathscr{P}_m$ denotes the (monic) nodal polynomial associated with $\Lambda(\boldsymbol{H}_m)$.

Naive approach: update $\boldsymbol{f}_m$ by explicit evaluation of divided differences (block Newton interpolation).

This is (severely) unstable.

Restart Algorithm 1 [Eiermann & E. (2006)]

k standard Arnoldi decompositions of $\boldsymbol{A}$

$$\boldsymbol{A}\boldsymbol{V}_j = \boldsymbol{V}_j\boldsymbol{H}_j + h_{j+1}\boldsymbol{v}_{jm+1}\boldsymbol{e}_m^T, \quad j = 1, 2, \dots, k,$$

of the m-dim. Krylov spaces $\mathscr{K}_m(\boldsymbol{A}, \boldsymbol{v}_{(j-1)m+1})$, glued together,

$$\boldsymbol{A}\hat{\boldsymbol{V}}_k = \hat{\boldsymbol{V}}_k\hat{\boldsymbol{H}}_k + h_{k+1}\boldsymbol{v}_{km+1}\boldsymbol{e}_{km}^T, \tag{2}$$

where $\hat{\boldsymbol{V}}_k := [\boldsymbol{V}_1 \ \boldsymbol{V}_2 \ \cdots \ \boldsymbol{V}_k] \in \mathbb{C}^{n\times km}$,

$$\hat{\boldsymbol{H}}_k := \begin{bmatrix} \boldsymbol{H}_1 & & & \\ \boldsymbol{E}_2 & \boldsymbol{H}_2 & & \\ & \ddots & \ddots & \\ & & \boldsymbol{E}_k & \boldsymbol{H}_k \end{bmatrix} \in \mathbb{C}^{km\times km}, \qquad \boldsymbol{E}_j := h_j\boldsymbol{e}_1\boldsymbol{e}_m^T \in \mathbb{R}^{m\times m}.$$

(2) is an Arnoldi-like decomposition of $\mathscr{K}_{km}(\boldsymbol{A}, \boldsymbol{b})$. Compute

$$\hat{\boldsymbol{f}}_k := \hat{\boldsymbol{V}}_k f(\hat{\boldsymbol{H}}_k)\boldsymbol{e}_1 = \hat{\boldsymbol{f}}_{k-1} + \boldsymbol{V}_k[f(\hat{\boldsymbol{H}}_k)\boldsymbol{e}_1]_{(k-1)m+1:km}.$$

Restart Algorithm 2 [Afanasjew, Eiermann, E. & Güttel (2008)]

Instead of $f(\boldsymbol{A})\boldsymbol{b}$, evaluate $r(\boldsymbol{A})\boldsymbol{b}$ where $f(\lambda) \approx r(\lambda) = \sum_{\ell=1}^{n_p} \frac{\alpha_\ell}{\omega_\ell - \lambda}$ is a suitably accurate rational approximation of f. Now

$$r(\hat{\boldsymbol{H}}_k)\boldsymbol{e}_1 = \sum_{\ell=1}^{n_p} \alpha_\ell(\omega_\ell \boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{e}_1 =: \sum_{\ell=1}^{n_p} \alpha_\ell \hat{\boldsymbol{r}}_\ell.$$

Due to block bidiagonal structure of $\hat{\boldsymbol{H}}_k$, each of the n_p systems $(\omega_\ell \boldsymbol{I} - \boldsymbol{A})\hat{\boldsymbol{r}}_\ell = \boldsymbol{e}_1$ can be solved recursively:

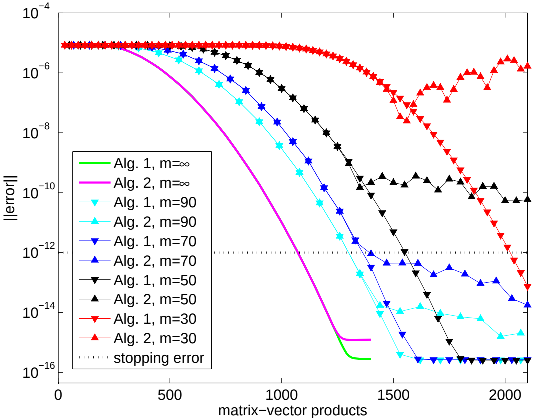
$$(\omega_\ell \boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{r}_{\ell,1} = \boldsymbol{e}_1, \qquad (\omega_\ell \boldsymbol{I} - \boldsymbol{H}_j)\boldsymbol{r}_{\ell,j} = \boldsymbol{E}_j \boldsymbol{r}_{\ell,j-1}, \quad j = 2, \ldots, k,$$

where $\hat{\boldsymbol{r}}_\ell = [\boldsymbol{r}_{\ell,1}^T, \boldsymbol{r}_{\ell,2}^T, \ldots, \boldsymbol{r}_{\ell,k}^T]^T$. Last block of $r(\hat{\boldsymbol{H}}_k)\boldsymbol{e}_1$ now obtained as

$$[O, \ldots, O, \boldsymbol{I}]\, r(\hat{\boldsymbol{H}}_k)\boldsymbol{e}_1 = \sum_{\ell=1}^{n_p} \alpha_\ell \boldsymbol{r}_{\ell,k}.$$

Numerical Example

$$\boldsymbol{f} = e^{t\boldsymbol{A}}\boldsymbol{b}$$
$$t = 10^{-3},$$
$$\boldsymbol{A} = [\nabla\times(\mu^{-1}\,\nabla\times\,\cdot)]_h$$
$$\dim \boldsymbol{A} = 565\,326$$
$$\Lambda(\boldsymbol{A}) \subset [-10^8, 0]$$



Deflated Restarting

- Compensate for deterioration of convergence due to restarting by augmenting the Krylov subspace with nearly invariant subspaces.
- Identify a subspace which slows convergence, approximate this space and eliminate its influence from the iteration process.
- In practice: Approximate eigenspaces associated with eigenvalues close to singularities of f (for $f = \exp$, approximate eigenspaces which belong to "large" eigenvalues).
- Well known for eigenproblems [Wu & Simon (2000)], [Stewart (2001)] and linear systems [Morgan (2002)].
  For matrix functions, first proposed by [Niehoff (2006)].

Numerical Example

$$\boldsymbol{f} = e^{t\boldsymbol{A}}\boldsymbol{b}$$

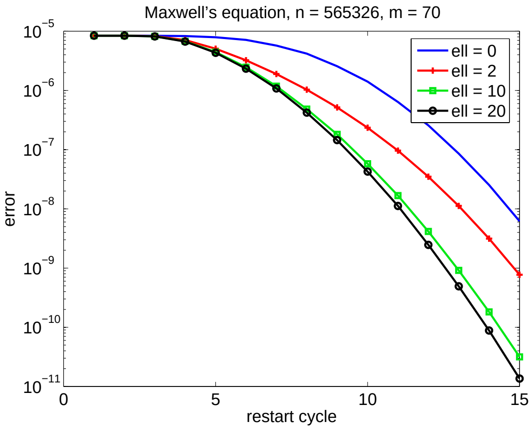
$$t = 10^{-3},$$

$$\boldsymbol{A} = [\nabla\times(\mu^{-1}\,\nabla\times\cdot)]_h$$

$$\dim \boldsymbol{A} = 565\,326$$

$$\Lambda(\boldsymbol{A}) \subset [-10^8, 0]$$

target: eigenvalues closest to origin.



Outline

Basic Error Bounds

For m-th (unrestarted) Krylov subspace approximation $\boldsymbol{f}_m \approx f(\boldsymbol{A})\boldsymbol{b}$ and any $p \in \mathscr{P}_{m-1}$, there holds

$$\begin{aligned} \|f(\boldsymbol{A})\boldsymbol{b} - \boldsymbol{f}_m\| &\leq \|f(\boldsymbol{A})\boldsymbol{b} - p(\boldsymbol{A})\boldsymbol{b}\| + \|\boldsymbol{f}_m - p(\boldsymbol{A})\boldsymbol{b}\| \\ &= \|(f-p)(\boldsymbol{A})\boldsymbol{b}\| + \|\boldsymbol{V}_m(f-p)(\boldsymbol{H}_m)\boldsymbol{e}_1\|. \end{aligned}$$

For $\boldsymbol{A} = \boldsymbol{A}^H$ we conclude

$$\|f(\boldsymbol{A})\boldsymbol{b} - \boldsymbol{f}_m\| \leq 2 \inf_{p \in \mathscr{P}_{m-1}} \|f - p\|_{\infty,[\lambda_{\min}(\boldsymbol{A}),\lambda_{\max}(\boldsymbol{A})]}.$$

For general $\boldsymbol{A}$:

$$\|f(\boldsymbol{A})\boldsymbol{b} - \boldsymbol{f}_m\| \leq C \inf_{p \in \mathscr{P}_{m-1}} \|f - p\|_{\infty,W(\boldsymbol{A})},$$

where $C \approx 13$ is Crouziex's universal constant.

Sequence of Krylov subspace approximations $\boldsymbol{f}_m \approx f(A)\boldsymbol{b}$ uniquely determined by (any) triangular scheme of interpolation nodes $\vartheta_j^{(m)} \in \mathbb{C}$ or their associated nodal polynomials $w_m \in \mathscr{P}_m$ ($w_0(z) \equiv 1$)

$$\begin{array}{llll} \vartheta_1^{(1)} & & & \\ \vartheta_1^{(2)} & \vartheta_2^{(2)} & & \\ \vartheta_1^{(3)} & \vartheta_2^{(3)} & \vartheta_3^{(3)} & \\ \vdots & \vdots & & \ddots \end{array} \qquad \begin{aligned} v_1(z) &= z - \vartheta_1^{(1)}, \\ v_2(z) &= (z - \vartheta_1^{(2)})(z - \vartheta_2^{(2)}), \\ v_3(z) &= (z - \vartheta_1^{(3)})(z - \vartheta_2^{(3)})(z - \vartheta_3^{(3)}), \\ &\vdots \end{aligned}$$

making up the vectors in associated Arnoldi-like decomposition, i.e., $\boldsymbol{v}_m = v_{m-1}(\boldsymbol{A})\boldsymbol{b}, m = 1, 2, \dots$.

Question: How quickly does $\boldsymbol{f}_m$ converge to $f(\boldsymbol{A})\boldsymbol{b}$ and how does this depend on $\boldsymbol{A}$, $\boldsymbol{b}$, f and $\{\vartheta_j^{(m)}\}$?

More Refined Bounds

Interpolation Theory: prescribed nodes

Under the assumption that the interpolation nodes are

- contained in a fixed compact set $\Omega \subset \mathbb{C}$,
- distributed asymptotically according to measure μ supported on Ω,

one can show

$$\begin{cases} \|f(\boldsymbol{A})\boldsymbol{b} - \boldsymbol{f}_m\|^{1/m} \leq C, & \text{if } f \text{ has finite singularities,} \\ (m\|f(\boldsymbol{A})\boldsymbol{b} - \boldsymbol{f}_m\|)^{1/m} \leq C, & \text{if } f \text{ is entire of order 1,} \end{cases}$$

where the constant C depends on

- the domain of analyticity and type of f,
- $\Lambda(A)$
- relative to the level curves of the logarithmic potential associated with μ.

Basic Quantities

Counting measure μ_m associated with m-th interpolation nodes:

$$\mu_m = \frac{1}{m}\sum_{j=1}^{m} \delta_{\vartheta_j^{(m)}} \qquad \text{where, for any } M \subset \mathbb{C},\ \delta_\vartheta(M) = \begin{cases} 1, & \vartheta \in M, \\ 0, & \text{otherwise.} \end{cases}$$

For every measure μ supported on the compact set $\Omega \subset \mathbb{C}$ we define the logarithmic potential $U^\mu : \mathbb{C} \to \mathbb{R}_0^+$ of μ by

$$U^\mu(z) = \int_\Omega \log \frac{1}{|z-t|}\, d\mu(t) = -\int_\Omega \log|z-t|\, d\mu(t).$$

For the counting measure we have

$$U^{\mu_m}(z) = -\frac{1}{m}\sum_{j=1}^{m} \log|z-\vartheta_j^{(m)}|,$$

and therefore, since $|v_m(z)|^{1/m}| = \left(\prod_{j=1}^{m} |z-\vartheta_j^{(m)}|\right)^{1/m}$,

$$\log|v_m(z)|^{1/m} = \frac{1}{m}\sum_{j=1}^{m} \log|z-\vartheta_j^{(m)}| = -U^{\mu_m}(z).$$

Example

Two well-known node sequences

Equidistant: $$\vartheta_j^{(m)} = -1 + 2\frac{j-1}{m-1}$$

$$\mu_m \xrightarrow{*} \mu, \; d\mu(t) = \tfrac{1}{2}\, dt$$

$$U^{\mu}(z) = 1 - \mathrm{Re}[(1-z)\log(1-z) + (1+z)\log(1+z)]$$

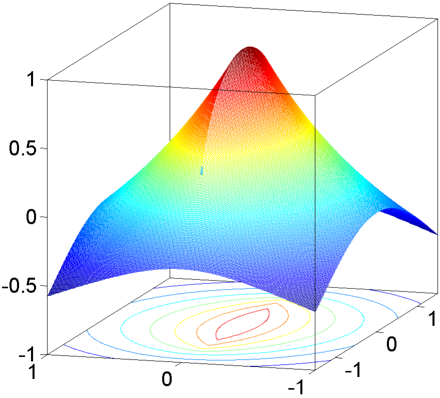
Chebyshev: $$\vartheta_j^{(m)} = \cos\frac{(j-1)\pi}{m-1}$$

$$\mu_m \xrightarrow{*} \mu, \; d\mu(t) = \frac{1}{\pi}\frac{dt}{\sqrt{1-t^2}}$$

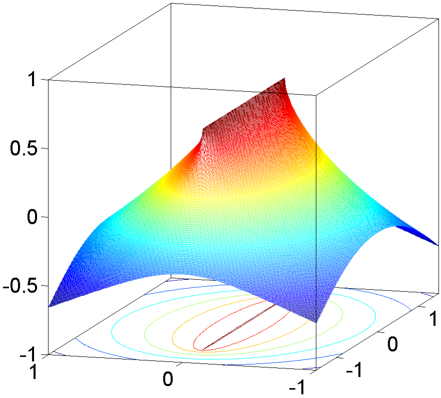
$$U^{\mu}(z) = e^{-1/2} - \log\left|z - \sqrt{z^2-1}\right|$$

Example

Their logarithmic potentials



Equidistant

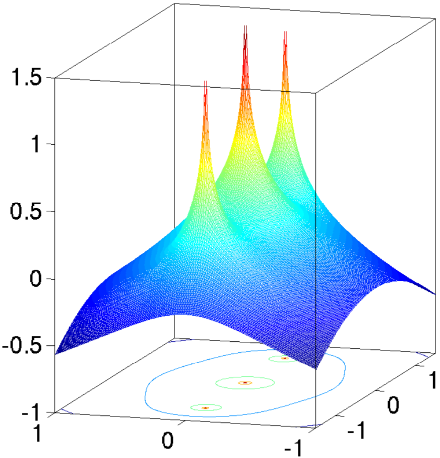


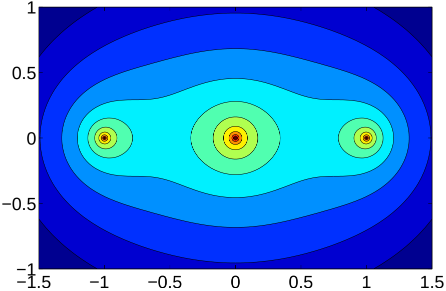
Chebyshev

Another Example

Typical for restarting

Repeat nodes $\vartheta = -1, 0, 1$ cyclically, $\mu_m \xrightarrow{*} \mu = \frac{1}{3}(\delta_{-1} + \delta_0 + \delta_1)$

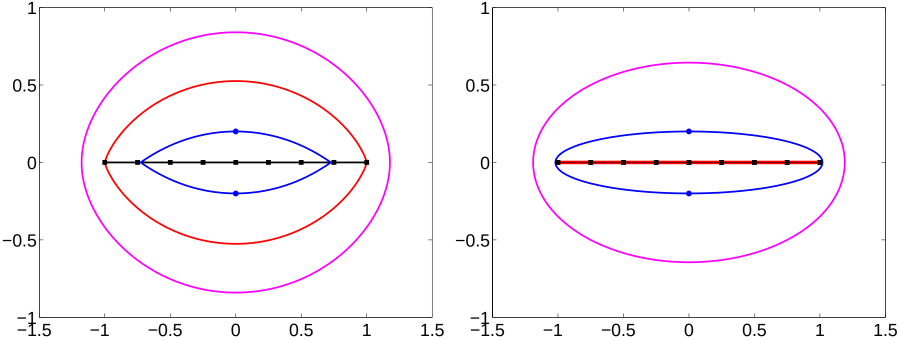




Potential Level Sets

For $\rho \geq 0$ define the level sets $\quad \Omega_\mu(\rho) := \{z : U^\mu(z) \geq -\log(\rho)\}$
and set

$$\rho_\mu(A) := \inf\{\rho : \Lambda(A) \subset \Omega_\mu(\rho)\},$$
$$\rho_\mu(f) := \inf\{\rho : f \text{ analytic in } \Omega_\mu(\rho)\}$$



$[\lambda_{\min}(A), \lambda_{\max}(A)] = [-1, 1], \qquad f(z) = 1/(1 + 25z^2)$

Can extend from linear systems to matrix functions the techniques of Kuijlaars and Beckermann to quantify the effect of interpolating at successively better approximations of parts of $\Lambda(\boldsymbol{A})$.

The asymptotic convergence factor of the Arnoldi approximation can be described using potentials of constrained equilibrium measures. The error is given by

- c_m^m if f has finite singularities, where $c_m < 1$ is a non-increasing function of m which depends on the eigenvalue distribution of A.
- $(c_m/m)^m$ if f is entire of order 1, where c_m is a non-increasing function of m.

Outline

1. Matrix Functions and Differential Equations
   - Initial Value Problems
   - Dirichlet-Neumann Maps
   - Stochastic Differential Equations
   - Frequency Domain Model Reduction

2. Krylov Subspace Approximation
   - Algorithm
   - Restarting
   - Convergence
   - A Posteriori Error Estimation

- For f rational, $\boldsymbol{A}$ Hermitian
  - Derive upper and lower bounds by exploiting collinearity of Galerkin residuals for shifted linear systems [Frommer & Simoncini (2008)]
  - Use CG-lower bounds of [Strakos & Tichy (2002)] for shifted systems and sum. [Frommer & Simoncini (2008)]
- For general f, Hermitian $\boldsymbol{A}$
  - Can derive upper and lower bounds based on error representation formula (divided differences) [Eiermann, E. & Güttel (2008)]
- For general f, general $\boldsymbol{A}$
  - Can use auxuliary nodes in error representation formula to obtain estimates, upper or lower bounds [Saad (1992)], [Philippe & Sidje (1993)], [Eiermann, E. & Güttel (2008)]

Summary

- Evaluation of $f(\boldsymbol{A})\boldsymbol{b}$ required for many PDE applications.
- (Restarted) Krylov subspace methods effective for large problems.
- Asymptotic convergence behavior well understood, at least in Hermitian case.
- Several estimators available for error of Krylov subspace approximation to $f(\boldsymbol{A})\boldsymbol{b}$.

Further Reading

Martin Afanasjew, Michael Eiermann, Oliver G. Ernst, and Stefan Güttel.
On a generalization of the steepest descent method for matrix functions.
Electron. Trans. Numer. Anal., 2008.
(to appear).

Martin Afanasjew, Michael Eiermann, Oliver G. Ernst, and Stefan Güttel.
Implementation of a restarted Krylov subspace method for the evaluation of matrix functions.
Linear Algebra and its Applications, 2008 (to appear).

Ralph-Uwe Börner, Oliver G. Ernst, and Klaus Spitzer.
Fast 3D simulation of transient electromagnetic fields by model reduction in the frequency domain using Krylov subspace projection.
Geophysical Journal International, 73(3):766–780, 2008.

Michael Eiermann and Oliver G. Ernst.
A restarted Krylov subspace method for the evaluation of matrix functions.
SIAM Journal on Numerical Analysis, 44:2481–2504, 2006.

Michael Eiermann, Oliver G. Ernst, and Stefan Güttel.
Asymptotic convergence analysis of Krylov subspace approximations to matrix funcions using potential theory.
(in preparation).

Michael Eiermann, Oliver G. Ernst, and Stefan Güttel.
A posteriori error estimation for Krylov subspace approximation of matrix functions.
(in preparation).