Rational Krylov methods
for the approximation of matrix functions

Michael Eiermann, Oliver G. Ernst, and Stefan Güttel

Dresden Seminar
14 July 2009

Outline

Matrix functions

Given

- a large, sparse matrix $A \in \mathbb{C}^{N \times N}$,
- a vector $\boldsymbol{b} \in \mathbb{C}^N$,
- a sufficiently smooth scalar function f,

we are interested in approximating

$$f(A)\boldsymbol{b} := p_{N-1}(A)\boldsymbol{b},$$

where p_{N-1} Hermite-interpolates f at $\Lambda(A)$.

Problems: N is large, $\Lambda(A)$ is not known, $f(A)$ is a full matrix.

Main idea of polynomial Krylov methods: Replace p_{N-1} implicitly by small degree polynomial p_{m-1} such that

$$f(A)\boldsymbol{b} \approx p_{m-1}(A)\boldsymbol{b}.$$

Clearly, p_m needs to be a "good" approximation to f at $\Lambda(A)$.

Main idea of rational Krylov methods: Replace p_{N-1} implicitly by small degree rational function $r_{m-1} := p_{m-1}/q_{m-1}$ such that

$$f(A)\boldsymbol{b} \approx r_{m-1}(A)\boldsymbol{b}.$$

Clearly, r_{m-1} needs to be a "good" approximation to f at $\Lambda(A)$.
The denominators q_{m-1} are chosen by the user.

Motivation

We consider the initial-boundary value problem

$$\begin{aligned} \partial_t u - \Delta u &= 0 && \text{in } \Omega = (0,1)^3, \quad t > 0, \\ u(x,t) &= 0 && \text{on } \Gamma = \partial\Omega, \quad t > 0, \\ u(x,0) &= u_0(x) && \text{in } \Omega. \end{aligned}$$

Seven-point stencil discretization on a uniform grid involving n interior grid points in each Cartesian direction yields IVP

$$\begin{aligned} \boldsymbol{u}'(t) &= A\boldsymbol{u}(t), \quad t > 0, \\ \boldsymbol{u}(0) &= \boldsymbol{u}_0, \end{aligned}$$

with $N \times N$ matrix A ($N = n^3$) and an initial vector $\boldsymbol{u}_0$ consisting of the values $u_0(x)$ at the grid points x, the solution of which is given by

$$\boldsymbol{u}(t) = f^t(A)\boldsymbol{u}_0, \quad f^t(z) = \exp(t\,z).$$

Approximation methods for $f(A)\boldsymbol{b}$ can be characterized by

- the m-dimensional approximation space $\mathscr{V}_m$,
- the extraction method (here: Rayleigh-Ritz, shift-invert).

Define *polynomial Krylov space* of order m

$$\mathscr{K}_m(A, \boldsymbol{b}) = \mathrm{span}\{\boldsymbol{b}, A\boldsymbol{b}, \ldots, A^{m-1}\boldsymbol{b}\}.$$

- There exists an invariance index L such that $\mathscr{K}_{L-1} \subset \mathscr{K}_L = \mathscr{K}_{L+1}$.
- $\mathscr{K}_m \cong \mathscr{P}_{m-1}$ for $m \leq L$.
- The exact solution $f(A)\boldsymbol{b}$ is contained in $\mathscr{K}_L$.

Polynomial vs. rational Krylov

Method 1 (Lanczos)

Space: Approximate $f(A)\boldsymbol{b}$ in $\mathscr{V}_m = \mathscr{K}_m(A, \boldsymbol{b})$.

Extraction: Compute ONB V_m of $\mathscr{V}_m$ and set

$$\boldsymbol{f}_m = V_m f(V_m^* A V_m) V_m^* \boldsymbol{b}.$$

We refer to this as *Rayleigh-Ritz extraction*.

Note: $V_m^* A V_m$ is a *Rayleigh quotient* of A.

Method 2 (shift-invert Lanczos)

Space: Approximate $f(A)\boldsymbol{b}$ in $\mathscr{V}_m = \mathscr{K}_m((A-\xi I)^{-1}, \boldsymbol{b})$.

Extraction: Compute ONB V_m of $\mathscr{V}_m$ satisfying

$$(A-\xi I)^{-1} V_m = V_m T_m + \boldsymbol{v}_{m+1} t_{m+1,m} \boldsymbol{e}_m^T,$$

and back-transform T_m:

$$\boldsymbol{f}_m = V_m f([T_m^{-1} + \xi I]) V_m^* \boldsymbol{b}.$$

[Moret & Novati, 2004], [van den Eshof & Hochbruck, 2005]

Note: The Rayleigh quotient of A would be

$$V_m^* A V_m = [T_m^{-1} + \xi I] - V_m^* A \boldsymbol{v}_{m+1} t_{m+1,m} \boldsymbol{e}_m^T T_m^{-1}.$$

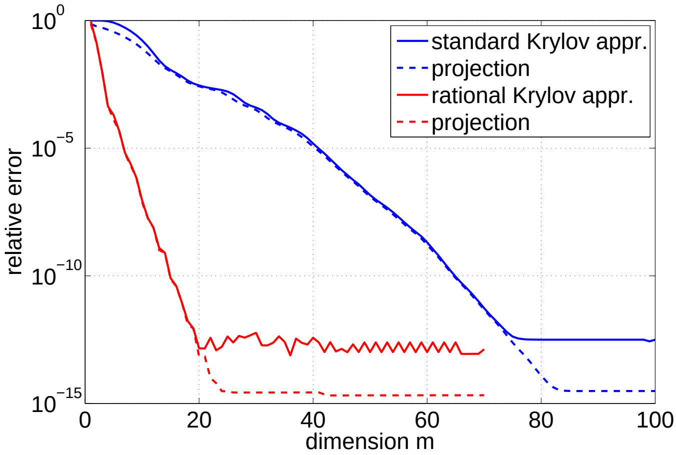


Figure: Heat equation $N = 15^3 = 3,375,\ t = 0.1,\ \xi = 1.$

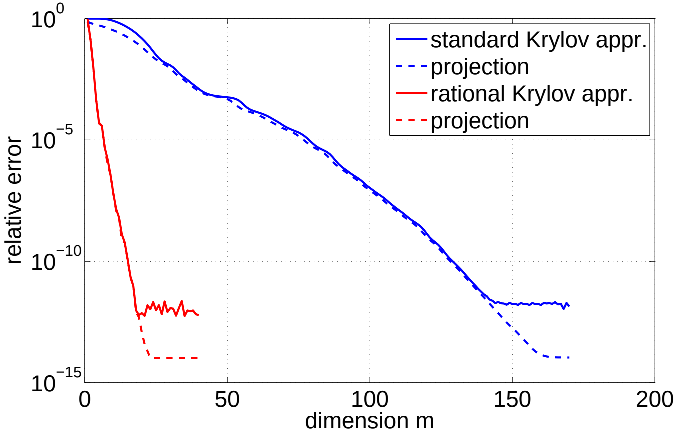


Figure: Heat equation $N = 31^3 = 29,791,\ t = 0.1,\ \xi = 1.$

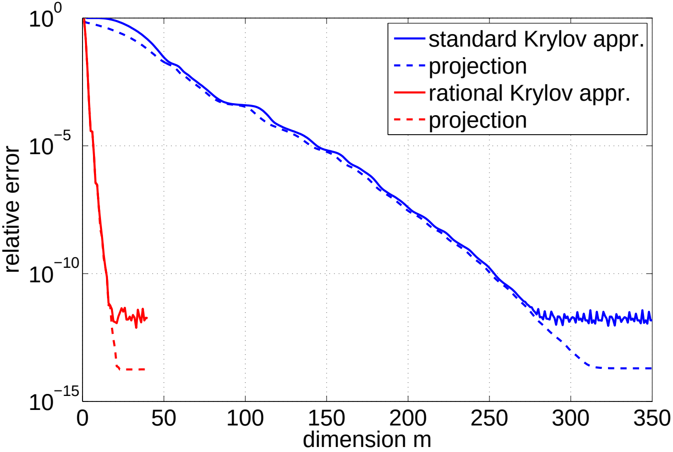


Figure: Heat equation $N = 63^3 = 250,047$, $t = 0.1$, $\xi = 1$.

Operation count

Recall: $N = n^3$

	Lanczos	**SI-Lanczos**
Iterations	$m \sim \sqrt{\|tA\|} = O(n)$ *	$m = \text{const.}$
Operator/It.	$O(N)$	$O(N)$ **
Orthogonalize/It.	$O(N)$	$O(N)$
Total	$O(nN)$	$O(N)$

Additionally, one evaluation of $f(T_m)$ typically involves $O(m^2)$ operations.

---

*) see [Hochbruck & Lubich, 1997]

**) for an ideal multigrid method

FEM discretization

... typically yields problems of the form $\exp(t\, M^{-1} K)\boldsymbol{b}$, where

- K is the stiffness matrix,
- M is the mass matrix,
- both matrices have (roughly) the same sparsity pattern.

Here rational Krylov methods seem even more natural since

$$(M^{-1}K - \xi I)^{-1} = (K - \xi M)^{-1} M.$$

Discussion

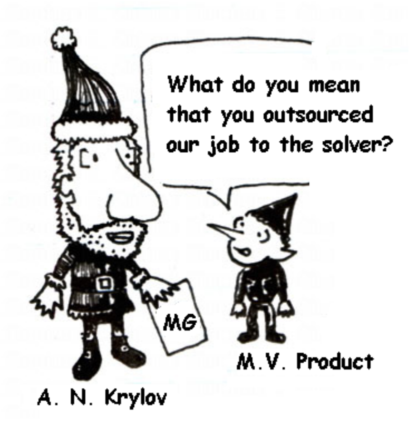
In polynomial Krylov methods the matrix A is incorporated only through matrix-vector products $A\boldsymbol{v}$.

In rational Krylov methods we need to solve linear systems $(A - \xi I)^{-1}\boldsymbol{v}$.

But: Rational functions have much better approximation properties than polynomials (if "good" poles are available).

And: Rational functions allow for straightforward parallelization by partial fraction expansion.

$\Rightarrow$ Maybe we could benefit from progress made for the solution of linear systems by displacing iteration work to the linear system solver.


What do you mean that you outsourced our job to the solver?
MG
M.V. Product
A. N. Krylov

Rational Krylov spaces

Definition

Let q_{m-1} be a polynomial of degree $\leq m-1$ which is nonzero at all eigenvalues of A. Then

$$\mathscr{V}_m = q_{m-1}(A)^{-1}\mathscr{K}_m(A, \boldsymbol{b})$$

is the *rational Krylov space of order* m *associated with* A*,* $\boldsymbol{b}$*, and* q_{m-1}.

Basic facts:

- $\mathscr{V}_m = \mathscr{K}_m(A, q_{m-1}(A)^{-1}\boldsymbol{b})$,
- $\mathscr{V}_m \cong \mathscr{P}_{m-1}/q_{m-1}$ for $m \leq L$,
- $\dim \mathscr{V}_m = \min\{m, L\}$,
- $\mathscr{V}_m \subseteq \mathscr{V}_L = \mathscr{K}_L$,
- $f(A)\boldsymbol{b} \in \mathscr{V}_L$.

Computationally more practical are nested spaces, which result when

$$q_{m-1}(z) := \prod_{\substack{j=1 \\ \xi_j \neq \infty}}^{m-1} (z - \xi_j)$$

for a given fixed pole sequence $(\xi_1, \xi_2, \ldots) \subset \overline{\mathbb{C}}$.

We then have

- $\mathscr{V}_1 \subset \mathscr{V}_2 \subset \cdots \subset \mathscr{V}_L = \mathscr{K}_L$.

Note: If all $\xi_j = \infty$, then $\mathscr{V}_m = q_{m-1}(A)^{-1}\mathscr{K}_m = \mathscr{K}_m$.

Algorithm 1: Rational Arnoldi algorithm [Ruhe, 1984]

Given: $(A, \boldsymbol{b}, q_m)$

$\boldsymbol{v}_1 := \boldsymbol{b}/\|\boldsymbol{b}\|$

for $j = 1, 2, \ldots, m$ **do**

  $\boldsymbol{x} := (I - A/\xi_j)^{-1} A\boldsymbol{v}_j$

  for $i = 1, 2, \ldots, j$ **do**

    $h_{i,j} := \langle \boldsymbol{x}, \boldsymbol{v}_i \rangle$

    $\boldsymbol{x} := \boldsymbol{x} - \boldsymbol{v}_i h_{i,j}$

  end

  $h_{j+1,j} := \|\boldsymbol{w}\|$

  $\boldsymbol{v}_{j+1} := \boldsymbol{w}/h_{j+1,j}$

end

(Problems with $\xi_j = 0$ and possible breakdown can be fixed.)

Rational Arnoldi decompositions

With $D_m = \mathrm{diag}(\xi_1^{-1}, \ldots, \xi_m^{-1})$ there holds

$$AV_m(H_m D_m + I_m) + A\boldsymbol{v}_{m+1} h_{m+1,m} \xi_m^{-1} \boldsymbol{e}_m^T = V_m H_m + \boldsymbol{v}_{m+1} h_{m+1,m} \boldsymbol{e}_m^T.$$

More shortly,

$$\boxed{AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}}$$

where $\underline{K_m}$ and $\underline{H_m}$ are upper Hessenberg matrices of size $(m+1) \times m$.

Note: If $\xi_m = \infty$, then $V_m^* A V_m = H_m(H_m D_m + I_m)^{-1}$
[Deckers & Bultheel, 2008], [Beckermann & Reichel, 2008].

Remarks:

- $V_m^* A V_m$ can also be computed recursively from $V_{m-1}^* A V_{m-1}$ without the artificial pole $\xi_m = \infty$.
- In SI-Lanczos the compression $H_m(H_m D_m + I_m)^{-1}$ is used without setting $\xi_m = \infty$.
- Generalizing the polynomial case, the RADs satisfy

$$(\tau A - \sigma I)\, V_{m+1} \underline{K_m} = V_{m+1}(\tau \underline{H_m} - \sigma \underline{K_m}).$$

  This makes rational Krylov methods suitable for functions $f^{\tau,\sigma}(z) = g(\tau z - \sigma)$.

Rayleigh-Ritz extraction

Lemma (Exactness)

Let $q_{m-1}(A)^{-1}\mathscr{K}_m(A,\boldsymbol{b})$ *be a rational Krylov space of dimension* m *with ONB* V_m*. Then for* $A_m := V_m^* A V_m$ *and every rational function* r *of the form* $r = p/q_{m-1}$ *(*$p \in \mathscr{P}_{m-1}$*) there holds*

$$r(A)\boldsymbol{b} = V_m r(A_m) V_m^* \boldsymbol{b}.$$

Proof: Use $q_{m-1}(A)^{-1}\mathscr{K}_m(A,\boldsymbol{b}) = \mathscr{K}_m(A, q_{m-1}(A)^{-1}\boldsymbol{b})$ and apply arguments of polynomial Krylov spaces.

Lemma (Interpolation)

Let $\boldsymbol{c} := q_{m-1}(A)^{-1}\boldsymbol{b}$ *and* V_m *be an ONB basis of* $q_{m-1}(A)^{-1}\mathscr{K}_m(A,\boldsymbol{b}) = \mathscr{K}_m(A,\boldsymbol{c})$, $A_m := V_m^* A V_m$. *Then*

$$V_m f(A_m)\, V_m^* \boldsymbol{b} = V_m \widetilde{f}(A_m)\, V_m^* \boldsymbol{c} = \widetilde{p}_{m-1}(A)\boldsymbol{c},$$

where $\widetilde{f} := q_{m-1} \cdot f$ *and* $\widetilde{p}_{m-1}$ *is a polynomial of degree* $m-1$ *that Hermite-interpolates* $\widetilde{f}$ *at* $\Lambda(A_m)$.

Proof: To prove the first equality, note that $\deg(q_{m-1}) \leq m-1$ and $q_{m-1}(A)\boldsymbol{c} = \boldsymbol{b}$. By Lemma (Exactness) there holds $\boldsymbol{b} = V_m q_{m-1}(A_m)\, V_m^* \boldsymbol{c}$. The second equality results from interpolation properties of polynomial Krylov methods.

Let w_m denote the characteristic polynomial of A_m and let Γ be a contour containing $\Lambda(A_m)$ in $\mathrm{int}(\Gamma)$. Furthermore f is assumed to be analytic in $\mathrm{int}(\Gamma)$ (then so is $\widetilde{f}$). The polynomial $\widetilde{p}_{m-1}$ can be expressed using Hermite's formula

$$\widetilde{p}_{m-1}(z) = \frac{1}{2\pi i} \int_\Gamma \frac{w_m(\zeta) - w_m(z)}{w_m(\zeta)(\zeta - z)} \widetilde{f}(\zeta)\, d\zeta,$$

and for the interpolation error

$$\widetilde{f}(z) - \widetilde{p}_{m-1}(z) = \frac{1}{2\pi i} \int_\Gamma \frac{w_m(z)}{w_m(\zeta)(\zeta - z)} \widetilde{f}(\zeta)\, d\zeta.$$

Dividing this equation by $q_{m-1}(z)$ and setting $r_m(z) := w_m(z)/q_{m-1}(z)$ we obtain

$$f(z) - \widetilde{p}_{m-1}(z)/q_{m-1}(z) = \frac{1}{2\pi i}\int_\Gamma \frac{r_m(z)}{r_m(\zeta)(\zeta - z)} f(\zeta)\, d\zeta.$$

Using Lemma (Interpolation) we have the following representation for the error of a Rayleigh-Ritz approximation

$$f(A)\boldsymbol{b} - V_m f(A_m) V_m^* \boldsymbol{b} = \left(\frac{1}{2\pi i}\int_\Gamma (\zeta I - A)^{-1} f(\zeta)/r_m(\zeta)\, d\zeta\right) r_m(A)\boldsymbol{b}$$

Example: Choose Γ to wind around $\Sigma = F(A)$. Then choose q_{m-1} such that $r_m = w_m/q_{m-1}$ is small on Σ and large on Γ.

A related problem is

Theorem

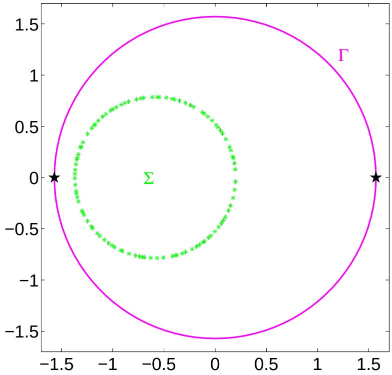
Given two compact disjoint sets Σ *and* Γ*. Denote by* $\mathscr{R}_{m-1}^{m}$ *the set of rational functions of type* $(m, m-1)$ *with zeros in* Σ *and poles in* Γ*. There holds*

$$\min_{\left(r_m \in \mathscr{R}_{m-1}^{m}\right)_{m\geq 1}} \limsup_{m\to\infty} \left(\frac{\max_{z\in\Sigma} r_m(z)}{\min_{z\in\Gamma} r_m(z)} \right)^{1/m} = c(\Sigma, \Gamma) < 1,$$

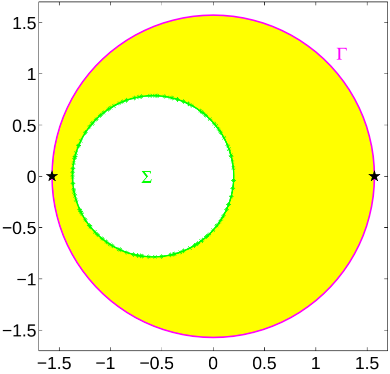
where $c(\Sigma, \Gamma)$ *is the capacity of the plane condenser with plates* Σ *and* Γ*.*

Idea: Use the poles of optimal rational functions r_m^* as poles for a rational Krylov space.

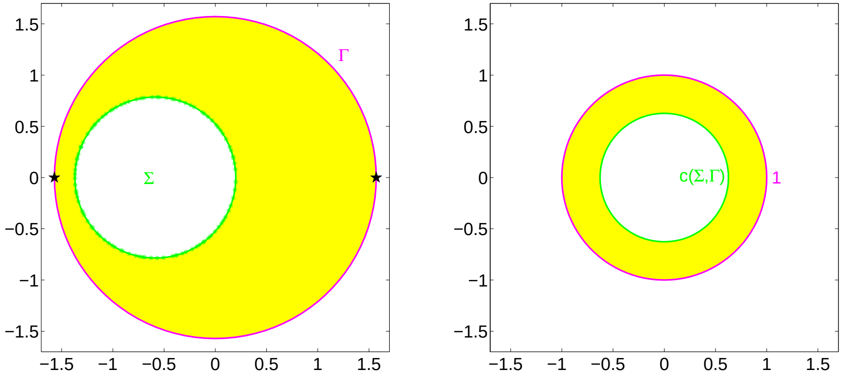
Example: (inspired by C. Beattie) We approximate $\tan(A)\boldsymbol{b}$ for $A = 0.2I - \frac{\pi}{4}U$, where U is a unitary random matrix.
We choose $\Gamma = \{z : |z| = \frac{\pi}{2}\}$.



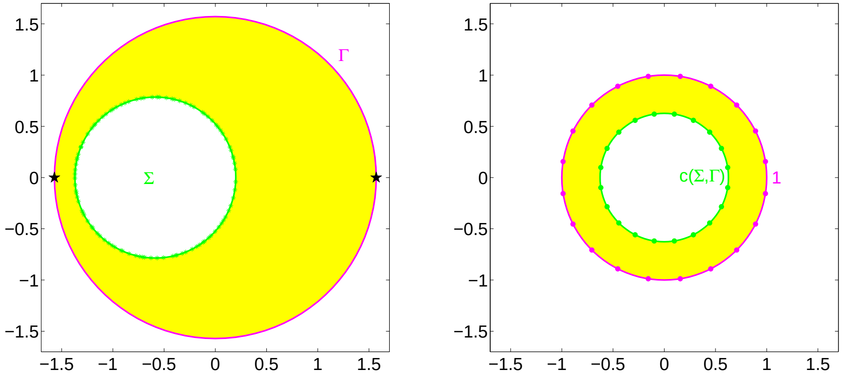
Example: (inspired by C. Beattie) We approximate $\tan(A)\boldsymbol{b}$ for $A = 0.2I - \frac{\pi}{4}U$, where U is a unitary random matrix.
We choose $\Gamma = \{z : |z| = \frac{\pi}{2}\}$.



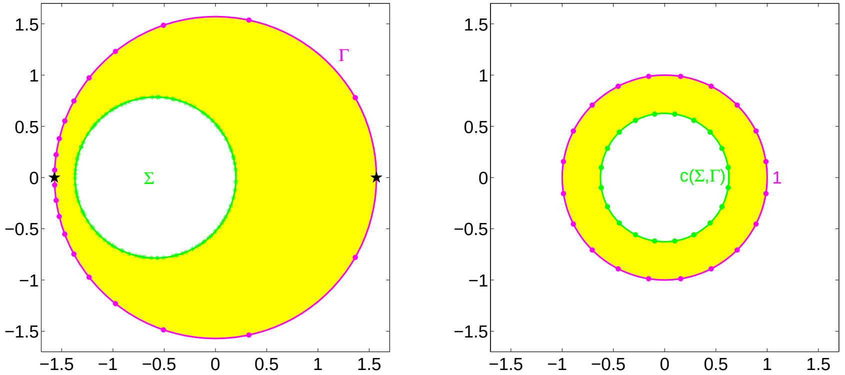
Example: (inspired by C. Beattie) We approximate $\tan(A)\boldsymbol{b}$ for $A = 0.2I - \frac{\pi}{4}U$, where U is a unitary random matrix.
We choose $\Gamma = \{z : |z| = \frac{\pi}{2}\}$.



Example: (inspired by C. Beattie) We approximate $\tan(A)\boldsymbol{b}$ for $A = 0.2I - \frac{\pi}{4}U$, where U is a unitary random matrix.
We choose $\Gamma = \{z : |z| = \frac{\pi}{2}\}$.



Example: (inspired by C. Beattie) We approximate $\tan(A)\boldsymbol{b}$ for $A = 0.2I - \frac{\pi}{4}U$, where U is a unitary random matrix.
We choose $\Gamma = \{z : |z| = \frac{\pi}{2}\}$.



Example: (inspired by C. Beattie) We approximate $\tan(A)\boldsymbol{b}$ for $A = 0.2I - \frac{\pi}{4}U$, where U is a unitary random matrix.
We choose $\Gamma = \{z : |z| = \frac{\pi}{2}\}$.

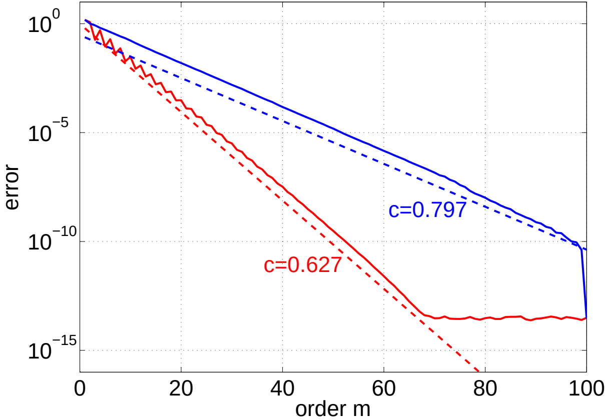


Figure: Rational (red) vs. polynomial (blue) Krylov approximation

Note: We have not chosen the σ_j given in the last theorem. These interpolation points are chosen implicitly by Rayleigh-Ritz extraction.

However, we observe that supplying optimal poles to the rational Krylov method yields an optimal error reduction in the sense, that the error is reduced with the optimal convergence rate $c(\Sigma, \Gamma)$ from the rational approximation problem.

Hence, the Rayleigh-Ritz extraction has implicitly chosen near-optimal interpolation points.

Lemma (Optimality)

Let $q_{m-1}(A)^{-1}\mathscr{K}_m(A,\boldsymbol{b})$ *be a rational Krylov space of dimension* m *with ONB* V_m*. Let* $A_m := V_m^* A V_m$ *and* $\Sigma = F(A)$*. Then*

$$\|f(A)\boldsymbol{b} - V_m f(A_m) V_m^* \boldsymbol{b}\| \le 2C\,\|\boldsymbol{b}\| \min_{p\in\mathscr{P}_{m-1}} \|f - p/q_{m-1}\|_\Sigma$$

with a constant $C \le 11.08$ *[Crouzeix, 2007].*
If A *is self-adjoint the result holds for* $C = 1$*.*

Aim: Choose poles q_{m-1} such that there exists a rational function p/q_{m-1} which is close to f on Σ.

- Lemma (Optimality) implies: Solving the linear systems in the RKM with unpreconditioned polynomial Krylov methods will require about the same number of MVPs as a purely polynomial Krylov method.
- More elaborate a-priori estimates are available, e.g., for Markov functions by using the Faber transform to consider approximation problems on the unit disk instead of Σ (see, e.g., [Druskin & Knizhnerman, 1989], [Knizhnerman, 1991], [Beckermann & Reichel, 2008]).
- It can be shown that SI-Lanczos extraction is also quasi-optimal. In some cases the optimal pole can be found by Remez' algorithm [Hochbruck & Lubich 1997].

Inexact solves

In each iteration of rational Arnoldi algorithm a linear system is solved.

Algorithm 2: Rational Arnoldi algorithm [Ruhe, 1984]

Given: $(A, \boldsymbol{b}, q_m)$

$\boldsymbol{v}_1 := \boldsymbol{b}/\|\boldsymbol{b}\|$

for $j = 1, 2, \ldots, m$ **do**

  $\boldsymbol{x}_j := (I - A/\xi_j)^{-1} A\boldsymbol{v}_j$

  for $i = 1, 2, \ldots, j$ **do**

    $h_{i,j} := \langle \boldsymbol{x}_j, \boldsymbol{v}_i \rangle$

    $\boldsymbol{x}_j := \boldsymbol{x}_j - \boldsymbol{v}_i h_{i,j}$

  end

  $h_{j+1,j} := \|\boldsymbol{x}_j\|$

  $\boldsymbol{v}_{j+1} := \boldsymbol{x}_j / h_{j+1,j}$

end

Inexact solves

In each iteration of rational Arnoldi algorithm a linear system is solved.

Algorithm 3: Rational Arnoldi algorithm [Ruhe, 1984]

Given: $(A, \boldsymbol{b}, q_m)$

$\boldsymbol{v}_1 := \boldsymbol{b}/\|\boldsymbol{b}\|$

for $j = 1, 2, \dots, m$ **do**

  $\widetilde{\boldsymbol{x}}_j \approx (I - A/\xi_j)^{-1} A\boldsymbol{v}_j$

  for $i = 1, 2, \dots, j$ **do**

    $h_{i,j} := \langle \widetilde{\boldsymbol{x}}_j, \boldsymbol{v}_i \rangle$

    $\widetilde{\boldsymbol{x}}_j := \widetilde{\boldsymbol{x}}_j - \boldsymbol{v}_i h_{i,j}$

  end

  $h_{j+1,j} := \|\widetilde{\boldsymbol{x}}_j\|$

  $\boldsymbol{v}_{j+1} := \widetilde{\boldsymbol{x}}_j / h_{j+1,j}$

end

The residual associated with $\widetilde{\boldsymbol{x}}_j \approx (I - A/\xi_j)^{-1} A\boldsymbol{v}_j$ is

$$\boldsymbol{r}_j := A\boldsymbol{v}_j - (I - A/\xi_j)\widetilde{\boldsymbol{x}}_j = (I - A/\xi_j)(\boldsymbol{x}_j - \widetilde{\boldsymbol{x}}_j).$$

Defining $R_m := [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m]$ we obtain a decomposition

$$AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m} + R_m.$$

Finally, with $E_m := -R_m\underline{K_m}^{\dagger} V_{m+1}^*$ we have

$$\boxed{(A + E_m)\, V_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}}$$

- One can show that this is a rational Arnoldi decomposition for $(A + E_m, \boldsymbol{b}, q_m)$ if q_m is nonzero at all eigenvalues of $A + E_m$.

- By $\widetilde{A}_m$ we denote the *inexact* Rayleigh quotient obtained from the inexact data $\underline{K_m}$ and $\underline{H_m}$ (it is the *exact* Rayleigh quotient for $A + E_m$).
- By $\widehat{A}_m$ we denote the *corrected* Rayleigh quotient obtained by explicit projection, i.e., $\widehat{A}_m = V_m^* A V_m$.

To quantify the error coming from the inexact solves we decompose the overall error

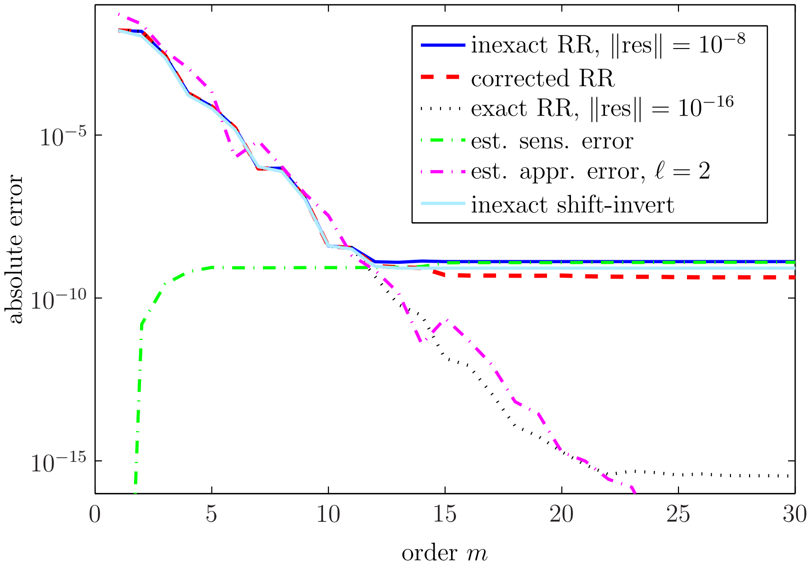
$$\| f(A)\boldsymbol{b} - V_m f(\widetilde{A}_m) V_m^* \boldsymbol{b} \|$$

$$\le \underbrace{\| f(A)\boldsymbol{b} - f(A + E_m)\boldsymbol{b} \|}_{\text{sensitivity error}} + \underbrace{\| f(A + E_m)\boldsymbol{b} - V_m f(\widetilde{A}_m) V_m^* \boldsymbol{b} \|}_{\text{approximation error}},$$

and estimate

$$\text{sensitivity error} \approx \| f(\widehat{A}_m) V_m^* \boldsymbol{b} - f(\widetilde{A}_m) V_m^* \boldsymbol{b} \|.$$

Error estimators, $f = \exp$, $N = 100$, A random neg. definite.



Remarks:

- Effective error estimators for the approximation error are also available.
- The level of the sensitivity error is usually well estimated after a few iterations already.
- One should stop iterating if the approximation error falls below the sensitivity error, because otherwise one only improves approximations to a series of “wrong” problems $f(A + E_m)\boldsymbol{b}$.

Orthogonal rational functions

Given ONB $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ of $q_{m-1}(A)^{-1}\mathscr{K}_m(A, \boldsymbol{b})$.

There exist polynomials $p_1, \ldots, p_m$ of (not nec. ascending) degree $m-1$ such that

$$\boldsymbol{v}_j = p_j(A) q_{m-1}(A)^{-1}\boldsymbol{b} =: r_j(A)\boldsymbol{b}.$$

Let $A = U\Lambda U^*$ be normal, $U^*U = I$. Then

$$\delta_{j,k} = \langle r_j(A)\boldsymbol{b}, r_k(A)\boldsymbol{b}\rangle_2 = \langle r_j(\Lambda)U^*\boldsymbol{b}, r_k(\Lambda)U^*\boldsymbol{b}\rangle_2 = \langle r_j(z), r_k(z)\rangle_\mu$$

where

$$\langle r_j(z), r_k(z)\rangle_\mu := \sum_{i=1}^{N} r_j(\lambda_i)\overline{r_k(\lambda_i)}\, w(\lambda_i)^2,$$

$w(\lambda_i) = |\boldsymbol{e}_i^T U^*\boldsymbol{b}|$. Hence, the r_j are orthonormal rational functions with respect to the inner product $\langle\,\cdot\,,\,\cdot\,\rangle_\mu$.

- It is well known that orthogonal rational functions satisfy short recurrences if A is self-adjoint. One can modify the rational Arnoldi algorithm to make it "Lanczos-like" (at the cost of 2 additional MVPs per iteration). [Bultheel et al, 1999].

There is much more to say

- How to solve/precondition the linear systems? What about multigrid methods for shifted sytems?
- How do the eigenvalues of the Rayleigh quotient A_m behave? [Beckermann, G. & Vandebril, 2009]
- Explain superlinear convergence of RKMs for $f(A)\boldsymbol{b}$!
- Possible spectral adaption of optimal poles?
- Develop (general) a-posteriori bounds for the error.
- Efficient implementation, parallelization?
- Application to real-world problems (e.g., from geophysical prospecting).