UNIVERSITÉ DE GENÈVE

Hybrid rational Krylov methods for matrix functions

Stefan Güttel
Swiss Numerics Colloquium, 16.04.2010

Outline

1. Introduction: Matrix functions
2. Rational Krylov decompositions
3. The rational Arnoldi method
4. A nonorthogonal rational Krylov method
5. Hybrid rational Krylov methods

1. Introduction: Matrix functions

Let $A \in \mathbb{C}^{N\times N}$ be a large sparse matrix.

Let $f(z)$ be a function which is analytic on $\Lambda(A)$.

Definition

The *matrix function* $f(A)$ is defined as $f(A) := p_A(A)$,
where $p_A \in \mathscr{P}_{N-1}$ interpolates f at $\Lambda(A)$.

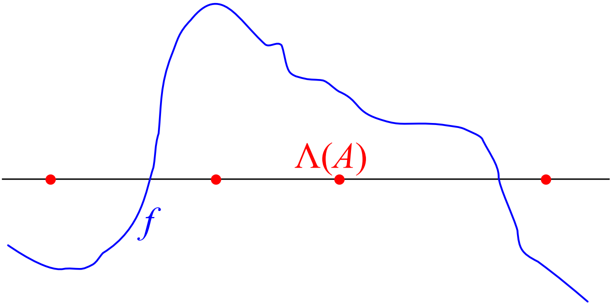
1. Introduction: Matrix functions

Let $A \in \mathbb{C}^{N \times N}$ be a large sparse matrix.

Let $f(z)$ be a function which is analytic on $\Lambda(A)$.

Definition

The *matrix function* $f(A)$ is defined as $f(A) := p_A(A)$,
where $p_A \in \mathscr{P}_{N-1}$ interpolates f at $\Lambda(A)$.



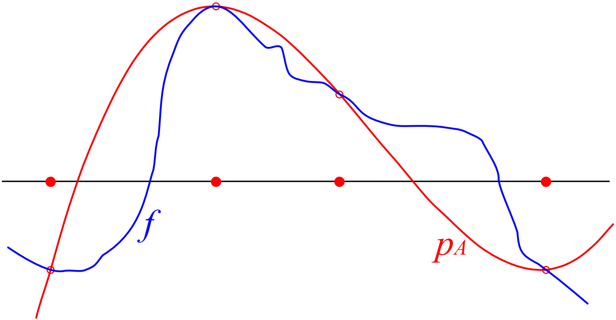
1. Introduction: Matrix functions

Let $A \in \mathbb{C}^{N \times N}$ be a large sparse matrix.

Let $f(z)$ be a function which is analytic on $\Lambda(A)$.

Definition

The *matrix function* $f(A)$ is defined as $f(A) := p_A(A)$,
where $p_A \in \mathscr{P}_{N-1}$ interpolates f at $\Lambda(A)$.



Example

Consider the initial value problem

$$\boldsymbol{u}'(t) = A\boldsymbol{u}(t) + \boldsymbol{g}(\boldsymbol{u}, t), \quad t \geq 0, \quad \boldsymbol{u}(0) = \boldsymbol{u}_0.$$

The *exponential forward Euler method* is given by

$$\boldsymbol{u}_{j+1} = e^{\Delta t\, A}\boldsymbol{u}_j + \Delta t\, \varphi_1(\Delta t\, A)\boldsymbol{g}(\boldsymbol{u}_j, t_j), \quad j = 0, 1, \ldots,$$

where $\varphi_1(z) = (e^z - 1)/z$.

Problem: Compute $f(A)\boldsymbol{b}_j$ for *many* vectors $\boldsymbol{b}_j \in \mathbb{C}^N$.

Must avoid computation of $f(A)$.

Principle of rational Krylov methods: Replace f (implicitly) by rational function $r_m = p_{m-1}/q_{m-1}$ of type $(m-1, m-1)$. Hopefully,

$$r_m(A)\boldsymbol{b} \approx f(A)\boldsymbol{b}.$$

The fixed denominator $q_{m-1} \in \mathscr{P}_{m-1}$ is chosen by the user.
It must be nonzero on $\Lambda(A)$.

Definition

The space of all such vectors $r_m(A)\boldsymbol{b}$ is called a *rational Krylov space*

$$\mathscr{Q}_m := \left\{ \frac{p_{m-1}}{q_{m-1}}(A)\boldsymbol{b} \,:\, p_{m-1} \in \mathscr{P}_{m-1} \right\}.$$

Note: With the space $\mathscr{Q}_m$ is always associated a denominator q_{m-1}.

We will consider sequences $\mathscr{Q}_1, \mathscr{Q}_2, \ldots$ of rational Krylov spaces.

It is useful to have nested spaces $\mathscr{Q}_1 \subset \mathscr{Q}_2 \subset \cdots$,
which is the case if the denominators $q_0, q_1, \ldots$ are defined as

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

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

2. How to extract approximations to $f(A)\boldsymbol{b}$ from $\mathcal{Q}_m$?

General approach: **Rational Krylov decomposition**

$$A V_m K_m = V_{m+1} \underline{H_m}, \qquad (*)$$

where

- $V_m \in \mathbb{C}^{N\times m}$, $\mathrm{range}(V_m) = \mathcal{Q}_m$,
- $V_{m+1} = [V_m, \boldsymbol{v}_{m+1}]$, $\mathrm{range}(V_{m+1}) = \mathcal{Q}_{m+1}$,
- $K_m \in \mathbb{C}^{m\times m}$,
- $\underline{H_m} \in \mathbb{C}^{(m+1)\times m}$ is of full rank m.

2. How to extract approximations to $f(A)\boldsymbol{b}$ from $\mathscr{Q}_m$?

General approach: **Rational Krylov decomposition**

$$AV_mK_m = V_{m+1}\underline{H_m}, \qquad (*)$$

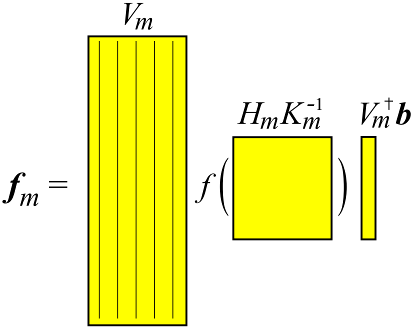
where

- $V_m \in \mathbb{C}^{N\times m}$, $\mathrm{range}(V_m) = \mathscr{Q}_m$,
- $V_{m+1} = [V_m, \boldsymbol{v}_{m+1}]$, $\mathrm{range}(V_{m+1}) = \mathscr{Q}_{m+1}$,
- $K_m \in \mathbb{C}^{m\times m}$,
- $\underline{H_m} \in \mathbb{C}^{(m+1)\times m}$ is of full rank m.

We associate with $(*)$ a **rational Krylov approximation**

$$\boldsymbol{f}_m := V_m f(H_m K_m^{-1}) V_m^{\dagger}\boldsymbol{b}.$$

Rational Krylov approximation $\boldsymbol{f}_m$ is easy to compute if decomposition $AV_mK_m = V_{m+1}\underline{H_m}$ is known:



Theorem

The rational Krylov approximation $\boldsymbol{f}_m = V_m f(H_m K_m^{-1}) V_m^\dagger \boldsymbol{b}$ satisfies

$$\boldsymbol{f}_m = r_m(A)\boldsymbol{b},$$

where $r_m = p_{m-1}/q_{m-1}$ interpolates f at $\Lambda(H_m K_m^{-1})$.

In particular: If $f \in \mathscr{P}_{m-1}/q_{m-1}$, then $\boldsymbol{f}_m = f(A)\boldsymbol{b}$.

Hence, there is an immediate connection to **rational interpolation**.

How good are rational Krylov approximations?

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m\|$$

How good are rational Krylov approximations?

$$
\begin{aligned}
& \|f(A)\boldsymbol{b} - \boldsymbol{f}_m\| \\
= \; & \|f(A)\boldsymbol{b} - V_m f(H_m K_m^{-1})\, V_m^{\dagger} \boldsymbol{b} - r_m(A)\boldsymbol{b} + V_m r_m(H_m K_m^{-1})\, V_m^{\dagger} \boldsymbol{b}\| \\
\leq \; & \|(f - r_m)(A)\boldsymbol{b}\| + \|V_m (f - r_m)(H_m K_m^{-1})\, V_m^{\dagger} \boldsymbol{b}\| \\
\leq \; & \|\boldsymbol{b}\| \left(\|(f - r_m)(A)\| + \|V_m\| \, \|(f - r_m)(H_m K_m^{-1})\| \right)
\end{aligned}
$$

How good are rational Krylov approximations?

$$
\begin{aligned}
& \|f(A)\boldsymbol{b} - \boldsymbol{f}_m\| \\
= \;& \|f(A)\boldsymbol{b} - V_m f(H_m K_m^{-1}) V_m^\dagger \boldsymbol{b} - r_m(A)\boldsymbol{b} + V_m r_m(H_m K_m^{-1}) V_m^\dagger \boldsymbol{b}\| \\
\leq \;& \|(f - r_m)(A)\boldsymbol{b}\| + \| V_m (f - r_m)(H_m K_m^{-1}) V_m^\dagger \boldsymbol{b}\| \\
\leq \;& \|\boldsymbol{b}\| \left(\|(f - r_m)(A)\| + \| V_m\| \, \|(f - r_m)(H_m K_m^{-1})\|\right) \\
\leq \;& C \, \|\boldsymbol{b}\| \left(\|f - r_m\|_{\mathbb{W}(A)} + \| V_m\| \, \|f - r_m\|_{\mathbb{W}(H_m K_m^{-1})}\right),
\end{aligned}
$$

where we have used

$$\|g(A)\| \leq C\|g\|_{\mathbb{W}(A)} = C \max_{z \in \mathbb{W}(A)} |g(z)| \quad \text{with a constant } C \leq 11.08$$

for every continuous function g on the numerical range $\mathbb{W}(A)$ [Crouzeix 07].

This bound holds for every $r_m \in \mathscr{P}_{m-1}/q_{m-1}$, hence

$$\|f(A)\boldsymbol{b}-\boldsymbol{f}_m\| \leq C\,\|\boldsymbol{b}\| \min_{r_m} \Big(\|f - r_m\|_{\mathbb{W}(A)} + \|\,V_m\|\, \|f - r_m\|_{\mathbb{W}(H_m K_m^{-1})} \Big)\,.$$

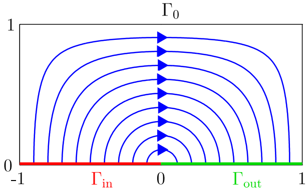
If V_m is *orthonormal*, the associated *rational Arnoldi approximation* $\boldsymbol{f}_m^*$ is **near-optimal**

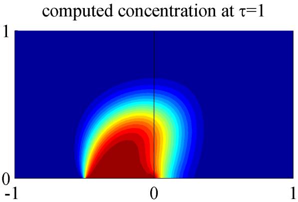
$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^*\| \leq 2\,C\,\|\boldsymbol{b}\| \min_{r_m} \|f - r_m\|_{\mathbb{W}(A)}.$$

[Druskin & Knizhnerman 98] [Beckermann & Reichel 09].

Example: Advection–Diffusion

$$
\begin{aligned}
\partial_t u &= \frac{1}{\mathrm{Pe}}\Delta u - \boldsymbol{a}\cdot\nabla u && \text{in } \Omega = (-1,1)\times(0,1),\\
u &= 1 - \tanh(\mathrm{Pe}) && \text{on } \Gamma_0,\\
u &= 1 + \tanh((2x+1)\,\mathrm{Pe}) && \text{on } \Gamma_{\mathsf{in}},\\
\frac{\partial u}{\partial n} &= 0 && \text{on } \Gamma_{\mathsf{out}},\\
u(\boldsymbol{x},0) &= u_0(\boldsymbol{x}) && \text{in } \Omega,\\
\boldsymbol{a}(x,y) &= [2y(1-x^2), -2x(1-y^2)] && \text{in } \Omega.
\end{aligned}
$$





Discretization by finite elements yields IVP

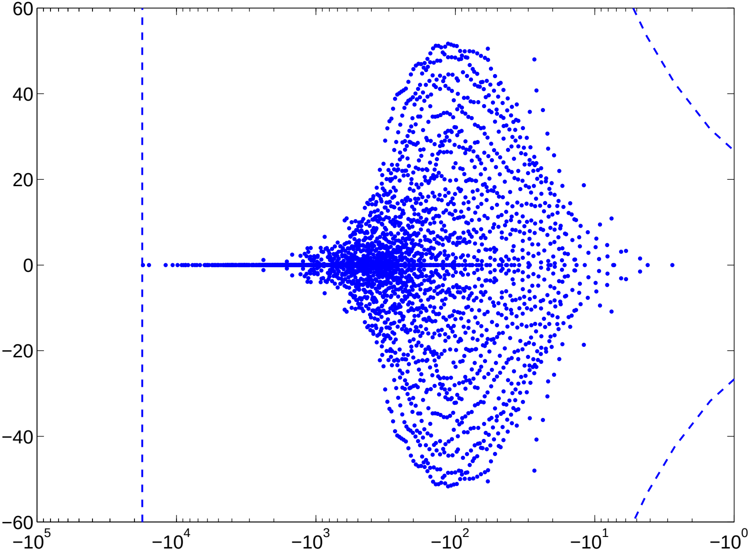
$$M\boldsymbol{u}'(t) = K\boldsymbol{u}(t) + \boldsymbol{g}, \quad \boldsymbol{u}(0) = \boldsymbol{u}_0,$$

with nonsymmetric K, $M \in \mathbb{R}^{N\times N}$ and $\boldsymbol{g} \in \mathbb{R}^N$ ($N = 2912$).

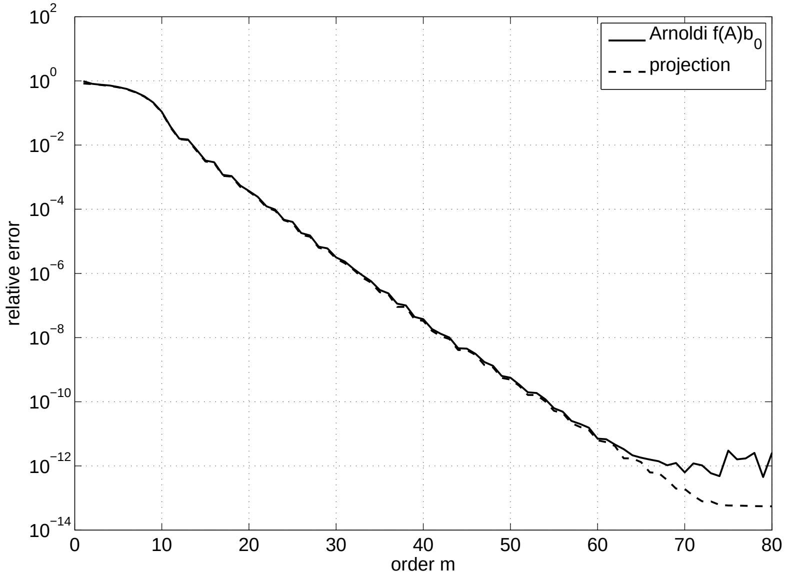
The exact solution is given by

$$\boldsymbol{u}(t) = e^{tA}\boldsymbol{u}_0 + t\,\varphi_1(tA)M^{-1}\boldsymbol{g}, \quad A = M^{-1}K.$$

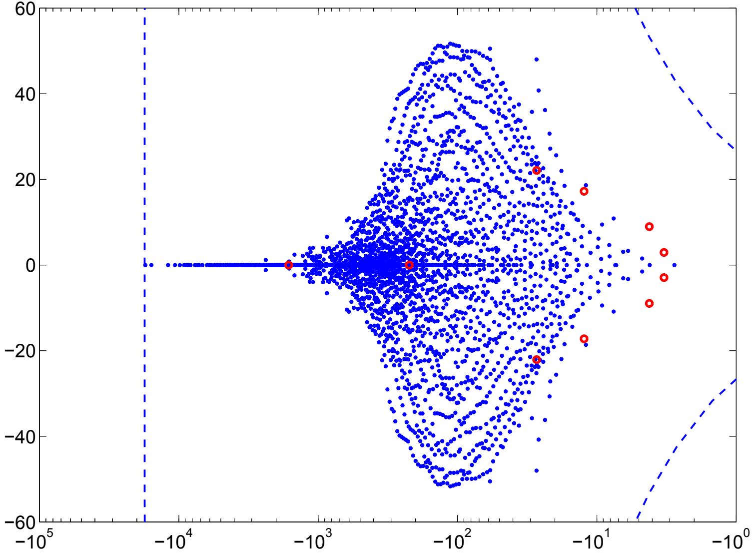
Spectrum and numerical range of A ($\mathrm{Pe} = 100$)



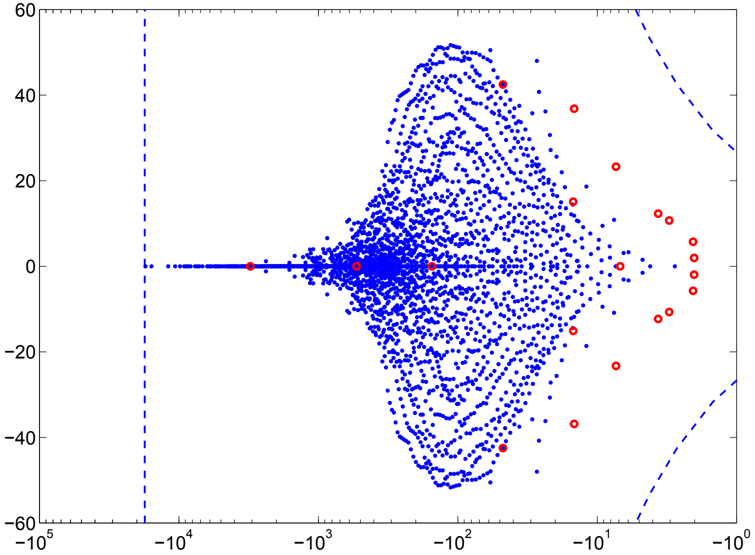
Convergence of the rational Arnoldi method (all poles at $\xi_j = 20$)



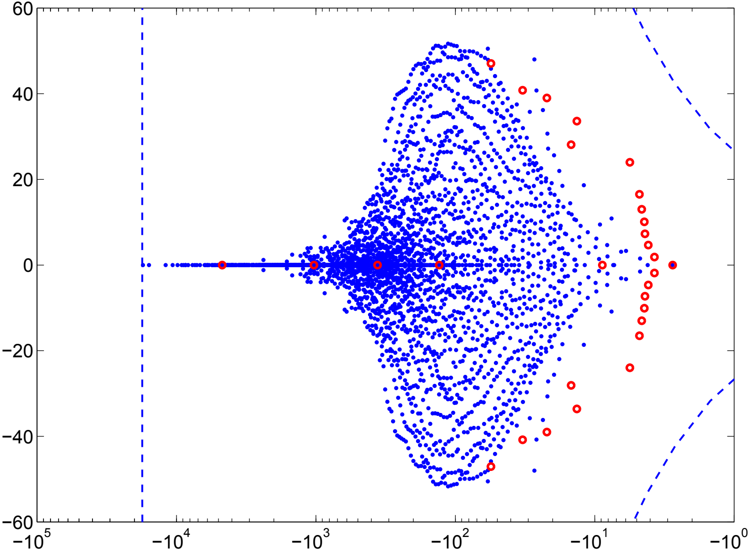
Interpolation nodes of rational Arnoldi ($m = 10$)



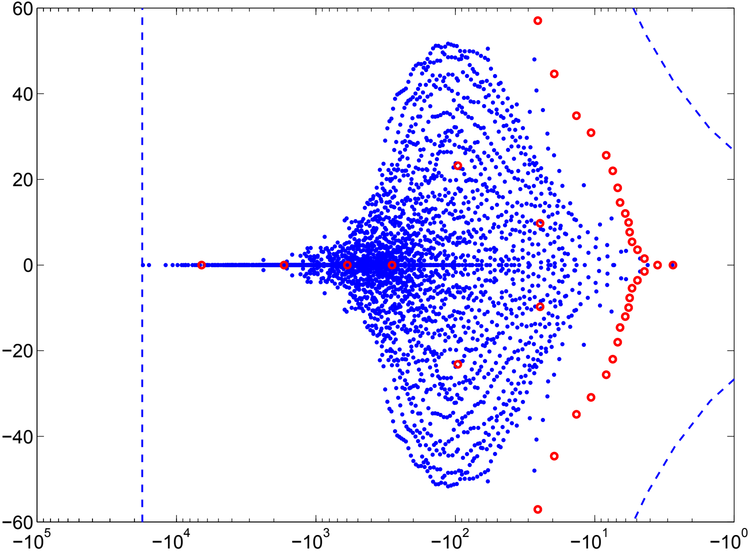
Interpolation nodes of rational Arnoldi ($m = 20$)



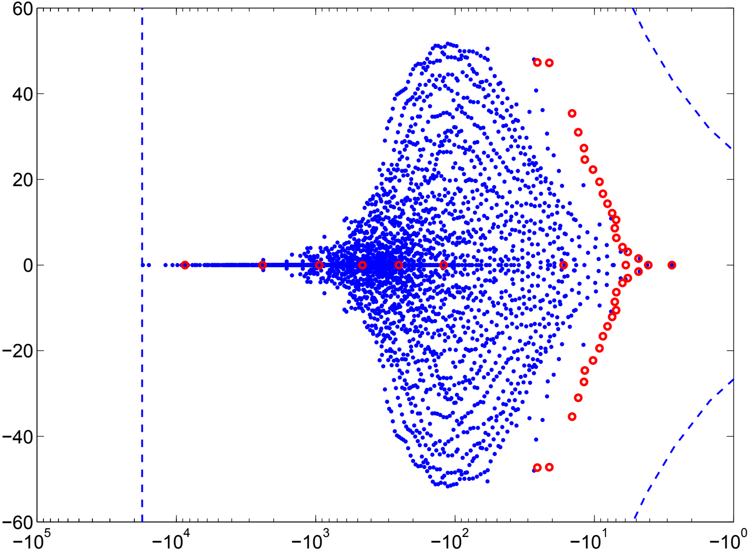
Interpolation nodes of rational Arnoldi ($m = 30$)



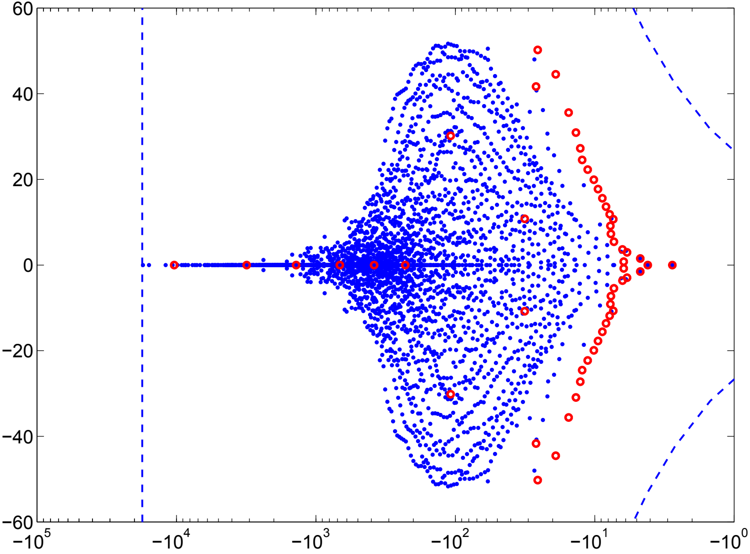
Interpolation nodes of rational Arnoldi ($m = 40$)



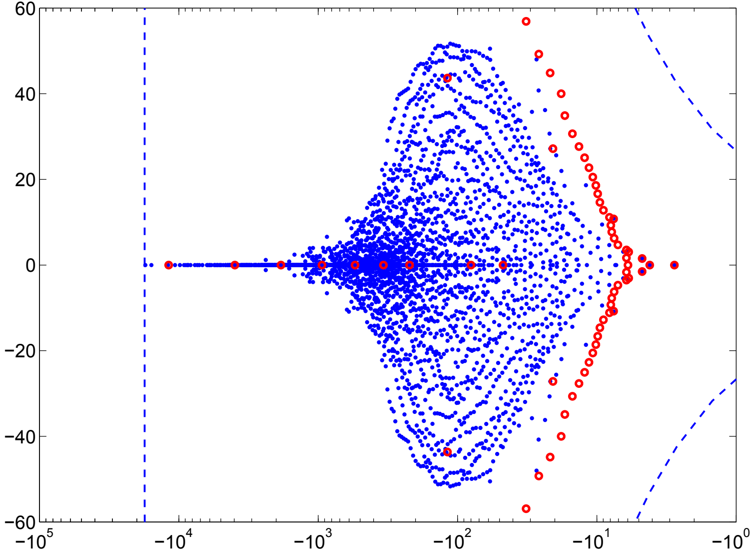
Interpolation nodes of rational Arnoldi ($m = 50$)



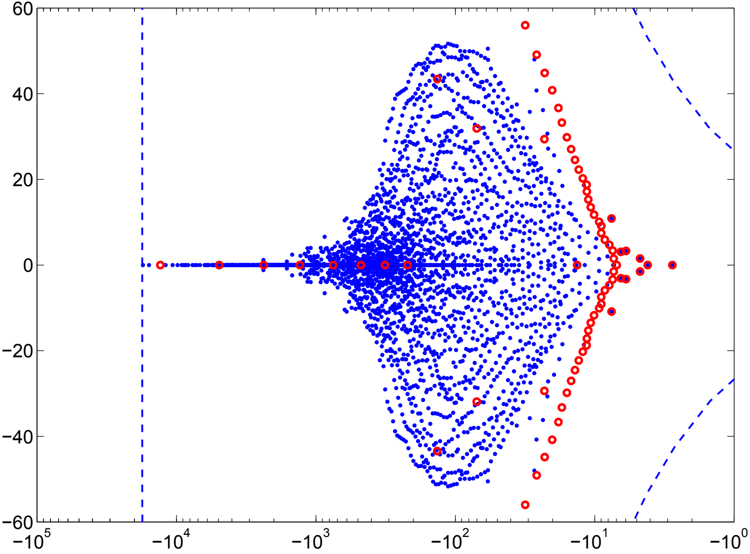
Interpolation nodes of rational Arnoldi ($m = 60$)



Interpolation nodes of rational Arnoldi ($m = 70$)



Interpolation nodes of rational Arnoldi ($m = 80$)



Due to its near-optimality and robustness, the rational Arnoldi method is widely used for approximating $f(A)\boldsymbol{b}$.

Drawback: Computation and storage of orthonormal basis V_m of rational Krylov space for A and $\boldsymbol{b}$.

$\Longrightarrow$ May become expensive especially for problems $f(A)\boldsymbol{b}_j$.

4. A nonorthogonal rational Krylov method

The iteration

$$\begin{aligned} \boldsymbol{v}_1 &= \boldsymbol{b}/\|\boldsymbol{b}\|, \\ \beta_j \boldsymbol{v}_{j+1} &= (I - A/\xi_j)^{-1}(A - \alpha_j I)\boldsymbol{v}_j, \quad j = 1, \ldots, m, \end{aligned}$$

yields a decomposition $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$ with

$V_{m+1} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m+1}]$,

$$\underline{K_m} = \left[\begin{array}{ccccc} 1 & & & & \\ {}^{\beta_1}/_{\xi_1} & 1 & & & \\ & {}^{\beta_2}/_{\xi_2} & \ddots & & \\ & & \ddots & 1 & \\ \hline & & & {}^{\beta_m}/_{\xi_m} & \end{array}\right] \quad \text{and} \quad \underline{H_m} = \left[\begin{array}{cccc} \alpha_1 & & & \\ \beta_1 & \alpha_2 & & \\ & \beta_2 & \ddots & \\ & & \ddots & \alpha_m \\ \hline & & & \beta_m \end{array}\right].$$

4. A nonorthogonal rational Krylov method

The iteration

$$\begin{aligned} \boldsymbol{v}_1 &= \boldsymbol{b}/\|\boldsymbol{b}\|, \\ \beta_j \boldsymbol{v}_{j+1} &= (I - A/\xi_j)^{-1}(A - \alpha_j I)\boldsymbol{v}_j, \quad j = 1, \ldots, m, \end{aligned}$$

can be used for stable rational interpolation:

By the last theorem we know that

$$\boldsymbol{f}_m = V_m f(H_m K_m^{-1})\|\boldsymbol{b}\|\boldsymbol{e}_1 = r_m(A)\boldsymbol{b} = \frac{p_{m-1}}{q_{m-1}}(A)\boldsymbol{b},$$

where r_m interpolates f at $\Lambda(H_m K_m^{-1}) = \{\alpha_1, \ldots, \alpha_m\}$.

4. A nonorthogonal rational Krylov method

The iteration

$$\begin{aligned} \boldsymbol{v}_1 &= \boldsymbol{b}/\|\boldsymbol{b}\|, \\ \beta_j \boldsymbol{v}_{j+1} &= (I - A/\xi_j)^{-1}(A - \alpha_j I)\boldsymbol{v}_j, \quad j = 1, \ldots, m, \end{aligned}$$

can be used for stable rational interpolation:

By the last theorem we know that

$$\boldsymbol{f}_m = V_m f(H_m K_m^{-1})\|\boldsymbol{b}\|\boldsymbol{e}_1 = r_m(A)\boldsymbol{b} = \frac{p_{m-1}}{q_{m-1}}(A)\boldsymbol{b},$$

where r_m interpolates f at $\Lambda(H_m K_m^{-1}) = \{\alpha_1, \ldots, \alpha_m\}$.

We call this **PAIN method** (poles and interpolation nodes).

The PAIN method

- can be interpreted as a restarted rational Krylov method [Afanasjew, Eiermann, Ernst & G. 08],
- requires storage of only 2 long vectors:

$$\boldsymbol{f}_m = \boldsymbol{f}_{m-1} + \gamma_m \boldsymbol{v}_m \quad \text{with} \quad \gamma_m = \boldsymbol{e}_m^T f(H_m K_m^{-1}) \|\boldsymbol{b}\| \boldsymbol{e}_1,$$

- requires no inner products,
- requires choice of interpolation nodes $\{\alpha_j\}$.

The PAIN method

- can be interpreted as a restarted rational Krylov method [Afanasjew, Eiermann, Ernst & G. 08],
- requires storage of only 2 long vectors:

$$\boldsymbol{f}_m = \boldsymbol{f}_{m-1} + \gamma_m \boldsymbol{v}_m \quad \text{with} \quad \gamma_m = \boldsymbol{e}_m^T f(H_m K_m^{-1}) \|\boldsymbol{b}\| \boldsymbol{e}_1,$$

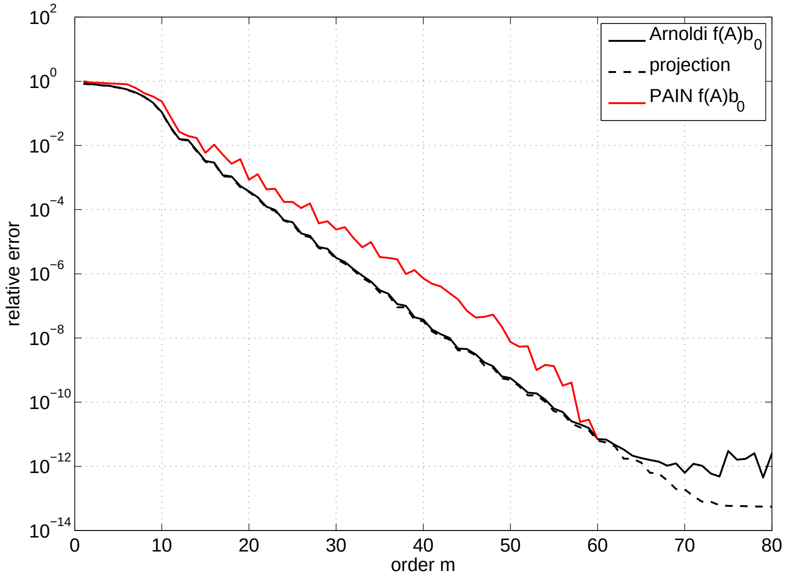
- requires no inner products,
- requires choice of interpolation nodes $\{\alpha_j\}$.

Idea: Use rational Ritz values of Arnoldi as interpolation nodes for PAIN.

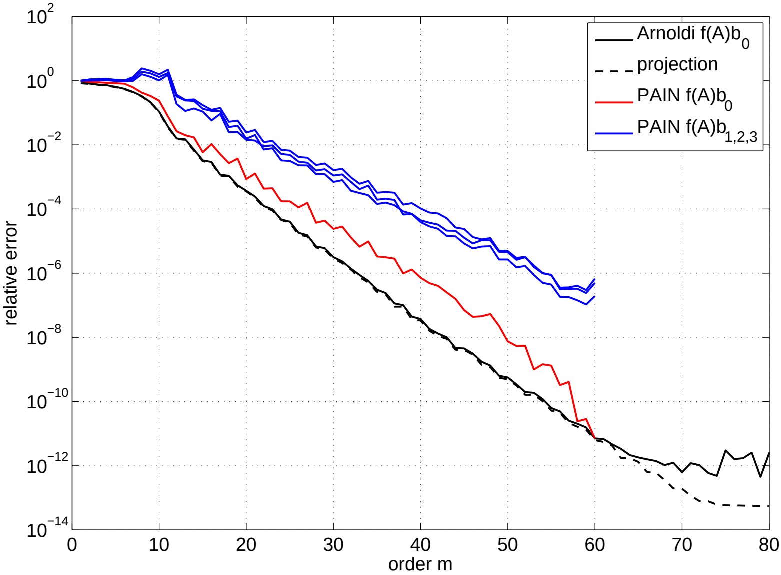
$\Longrightarrow$ **Hybrid rational Krylov methods**

(cf. *hybrid polynomial Krylov methods* for $A\boldsymbol{x} = \boldsymbol{b}$ [Nachtigal et al. 92])

Convergence of PAIN for $f(A)b = f(A)b_0$ using 60 Ritz values



Convergence of PAIN for $f(A)b_j$, $b_j = \texttt{randn}$



Why do we loose 5 digits of accuracy?

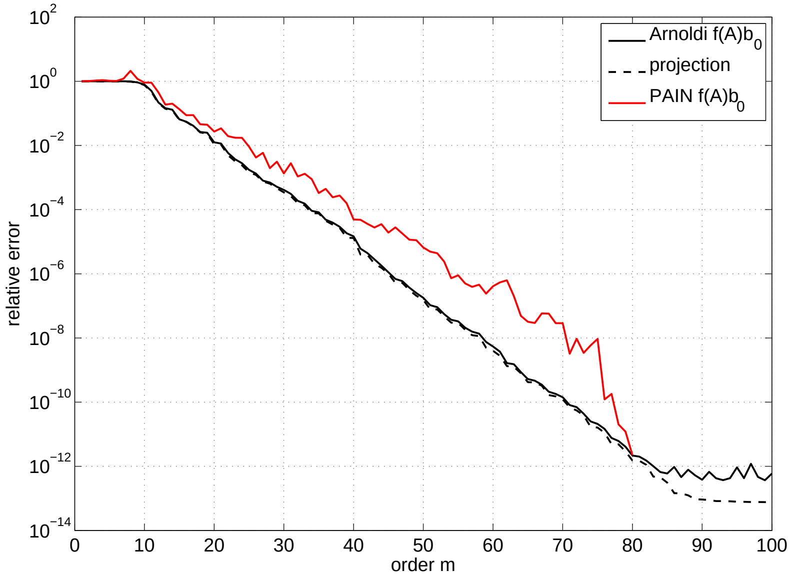
We have implicity reused the rational function r_{60}^* underlying the rational Arnoldi approximation

$$\boldsymbol{f}_{60}^* = r_{60}^*(A)\boldsymbol{b}_0,$$

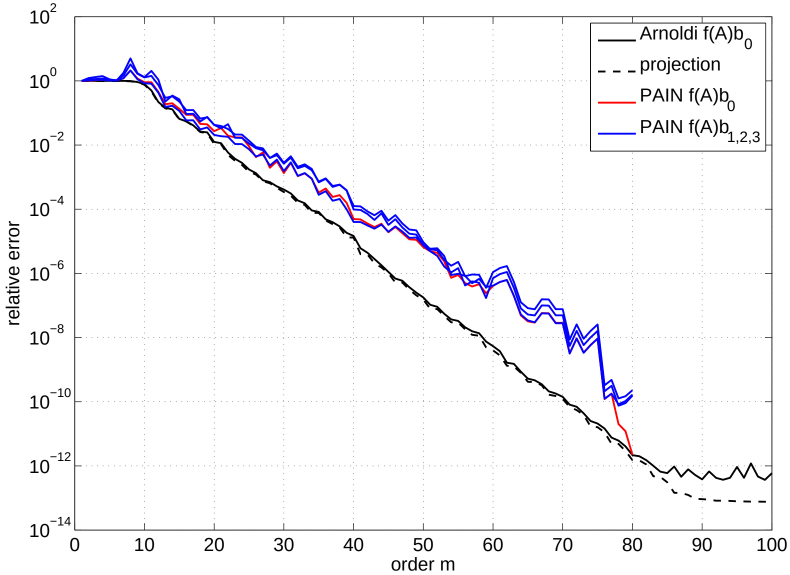
which is near-optimal for $f(A)\boldsymbol{b}_0$ (spectral adaption).

Hence, the "seed" vector $\boldsymbol{b}_0$ should have a similar spectral representation (with respect to A) as $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$.

Convergence of PAIN for $f(A)b_0$, $b_0 = \texttt{randn}$



Convergence of PAIN for $f(A)b_j$, $b_j = \texttt{randn}$



Summary

- Rational Krylov decompositions provide general framework for Krylov methods based on rational interpolation.
- Can design new rational Krylov methods using this framework.
- The rational Arnoldi method is special rational Krylov method.
- PAIN method is particularly simple method with only 2 vectors storage need and no inner products.
- Can combine to hybrid rational Krylov methods for problems of type $f(A)\boldsymbol{b}_j$ (if $\boldsymbol{b}_j$ have similar spectral representation).
- PAIN method can reuse Arnoldi's factorizations of $I - A/\xi_j$.
- Could be interesting, e.g., for parallel exponential integration of initial value problems.

Summary

- Rational Krylov decompositions provide general framework for Krylov methods based on rational interpolation.
- Can design new rational Krylov methods using this framework.
- The rational Arnoldi method is special rational Krylov method.
- PAIN method is particularly simple method with only 2 vectors storage need and no inner products.
- Can combine to hybrid rational Krylov methods for problems of type $f(A)\boldsymbol{b}_j$ (if $\boldsymbol{b}_j$ have similar spectral representation).
- PAIN method can reuse Arnoldi's factorizations of $I - A/\xi_j$.
- Could be interesting, e.g., for parallel exponential integration of initial value problems.

Thesis: "Rational Krylov methods for operator functions".