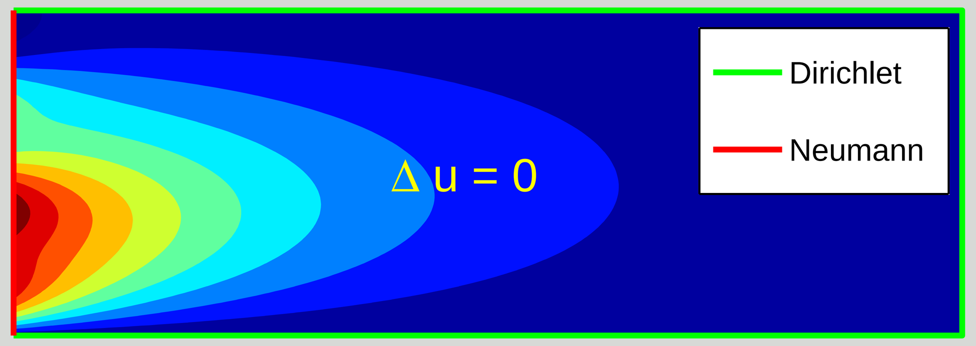
Towards black-box rational Krylov methods for matrix functions $f(A)b$: Automated parameter selection, inexact solves and error estimation

Stefan Güttel (stefan.guettel@unige.ch)

UNIVERSITÉ DE GENÈVE

Introduction

- **Given:** Large sparse matrix $A \in \mathbb{C}^{N\times N}$, vector $b \in \mathbb{C}^N$, function $f(z)$ analytic on the eigenvalues $\Lambda(A)$.
- **Task:** Compute $f(A)b$ without forming $f(A)$ explicitly.
- **Applications:**
  - $A^{-1}b$ is the solution of $Ax = b$,
  - $\exp(tA)b$ is the solution of $u'(t) = Au(t)$, $u(0) = b$,
  - $\cosh(t\sqrt{A})b$ solves $u''(t) = Au(t)$, $u(0) = b$, $u'(0) = 0$,
  - $\exp(t\sqrt{A})b$, $\mathrm{sgn}(A)b$, $\log(A)b$ (see also [8]),
  - $A^{-1/2}b$ for Neumann-to-Dirichlet maps:



The rational Arnoldi method

- **Principle:** Implicitly compute a low-order rational function $r_n(A)b \approx f(A)b$ with prescribed poles $\xi_1, \ldots, \xi_{n-1} \in \mathbb{C}$.
- **Implementation:** Use Ruhe's rational Arnoldi process [10]:

  Set $v_1 := b/\|b\|$.
  For $j = 1, \ldots, n$
    Compute $x_j := (A - \xi_j I)^{-1} v_j$.
    Orthogonalize $w_j := x_j - V_j V_j^H x_j$.
    Set $v_{j+1} := w_j / \|w_j\|$.
- **Output:** Rational Krylov basis $V_{n+1} = [v_1, \ldots, v_{n+1}]$, $V_{n+1}^H V_{n+1} = I_{n+1}$, and rational Arnoldi decomposition

$$AV_{n+1}\underline{K_n} = V_{n+1}\underline{H_n}, \quad \text{with} \quad \{\underline{K_n}, \underline{H_n}\} \subset \mathbb{C}^{(n+1)\times n}.$$

- **Rational Arnoldi approximation of order n is**

$$f_n := V_n f(A_n) V_n^H b, \qquad A_n := V_n^H A V_n.$$

(Note that $f(A_n)$ is a function of a small $n \times n$ matrix.)

Useful properties of f_n

- There exists a rational function $r_n(z)$ with prescribed poles $\xi_1, \ldots, \xi_{n-1}$ such that $f_n = r_n(A)b$.
- This function $r_n(z)$ is a rational interpolant for $f(z)$ with nodes $\Lambda(A_n) = \{\theta_1, \ldots, \theta_n\}$, called *rational Ritz values.*
- Define the nodal function

$$s_n(z) := \frac{(z - \theta_1)\cdots(z - \theta_n)}{(z - \xi_1)\cdots(z - \xi_{n-1})}.$$

Rational Ritz values $\{\theta_j\}$ are optimal in the sense that $\|s_n(A)b\|$ is minimal among all nodal functions.

Three practical problems

1. **How to choose the poles** $\xi_1, \xi_2, \ldots$ **?**
   (Clearly depends on $f(z)$ and on spectral properties of A.)
2. **What happens if** $x_j \approx (A - \xi_j I)^{-1} v_j$ **inexactly?**
   (Residuals are the only practically accessible information.)
3. **How large is the error** $\|f(A)b - f_n\|$**?**
   (For general functions $f(z)$, no residual equation available.)

Adaptive poles for Markov functions

- Can answer Question 1 for the particular function class

$$f(z) = \int_\Gamma \frac{d\gamma(x)}{x - z},$$

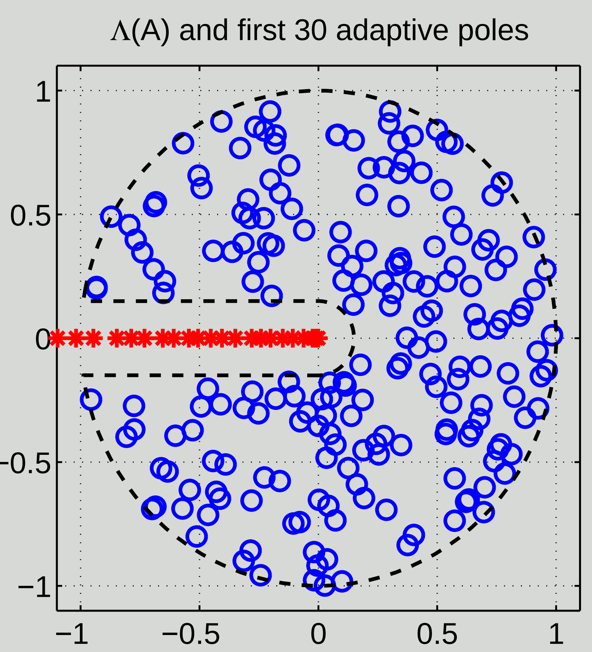
where γ is a (complex) measure with support $\Gamma \subset \mathbb{C}$.
- Can prove (see [2])

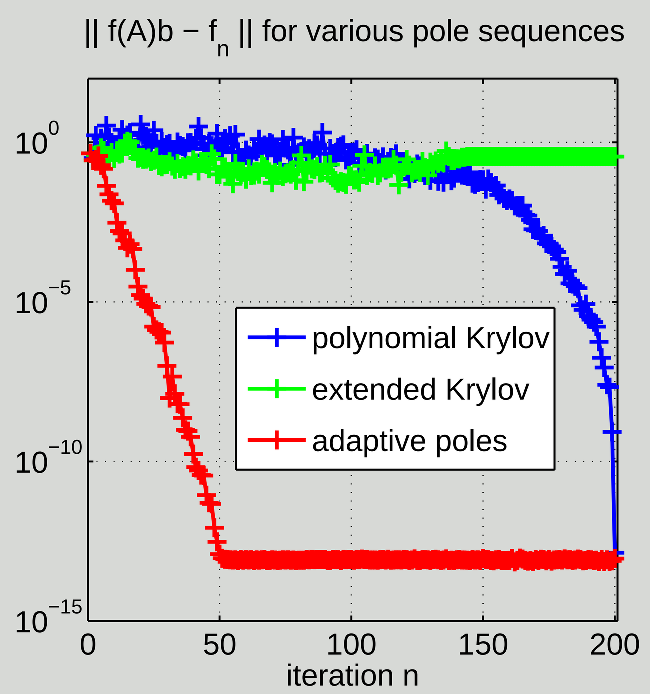
$$\|f(A)b - f_n\| \le \|s_n(A)b\| \cdot \left\| \int_\Gamma \frac{(xI - A)^{-1}}{s_n(x)} \, d\gamma(x) \right\|.$$

- Rational function $s_n(z)$ is explicitly known at iteration n (zeros $\{\theta_j\}$, poles $\{\xi_j\}$), can select the next pole ξ_n as

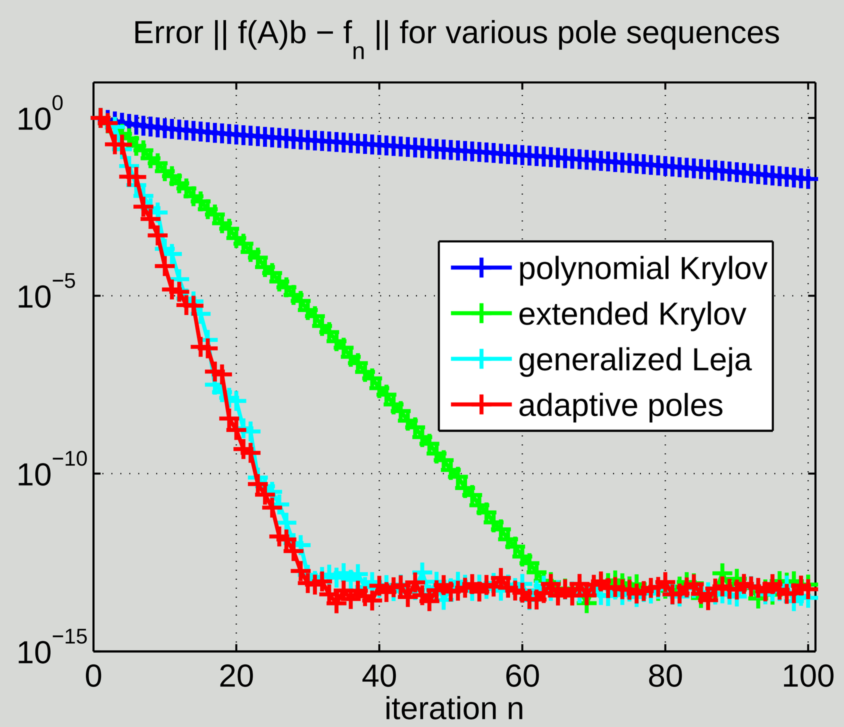
$$\xi_n = \arg\max_{x\in\Gamma} \left| \frac{1}{s_n(x)} \frac{d\gamma}{dx} \right|$$

(cf. [4, 5, 7]).
- **Feature:** So far no assumptions on the matrix A!
- **Example 1:** Compute $\log(A)b$ for highly nonnormal A.





- **Example 2:** Compute $A^{-1/2}b$ for 2D-Laplacian, $N = 10^4$.

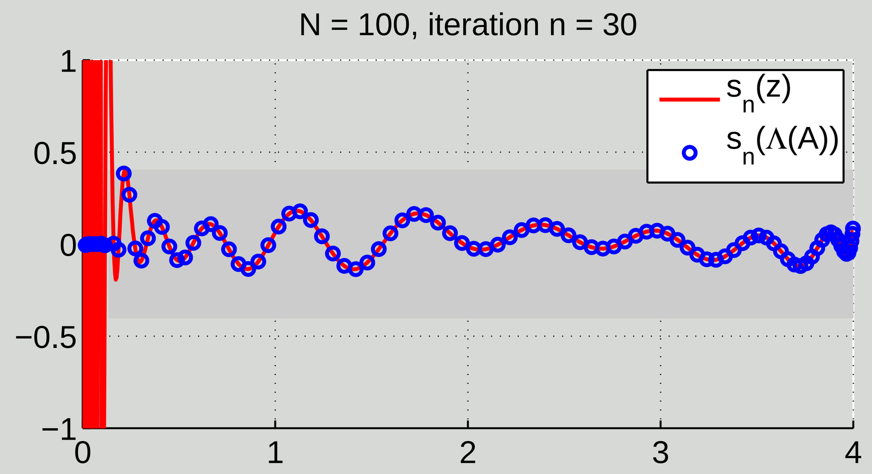


Spectral adaptivity

- Need to understand the (superlinear) decay of $\|s_n(A)b\|$.
- So far only possible for Hermitian A, in which case

$$\|s_n(A)b\| \le \|b\| \max_{z\in\Lambda(A)} |s_n(z)|.$$

- Typical linear error bounds obtained by assuming that $s_n(z)$ is uniformly small on spectral interval $[\lambda_{\min}, \lambda_{\max}]$.
- **This assumption ignores the fine structure of $\Lambda(A)$!**
- **Example:** $A = \mathrm{tridiag}(-1, 2, -1)$, extended Krylov.



Observations: $s_n(z)$ uniformly small on a subinterval S, and Ritz values in $[\lambda_{\min}, \lambda_{\max}] \setminus S$ are close to $\Lambda(A) \setminus S$.

- Optimality and interlacing property of rational Ritz values allow for asymptotic description of their distribution:
  - Let $\Lambda(A)$ be described by a probability measure σ, e.g.,

$$\frac{d\sigma}{dx} = \frac{1}{\pi\sqrt{x(4 - x)}}.$$

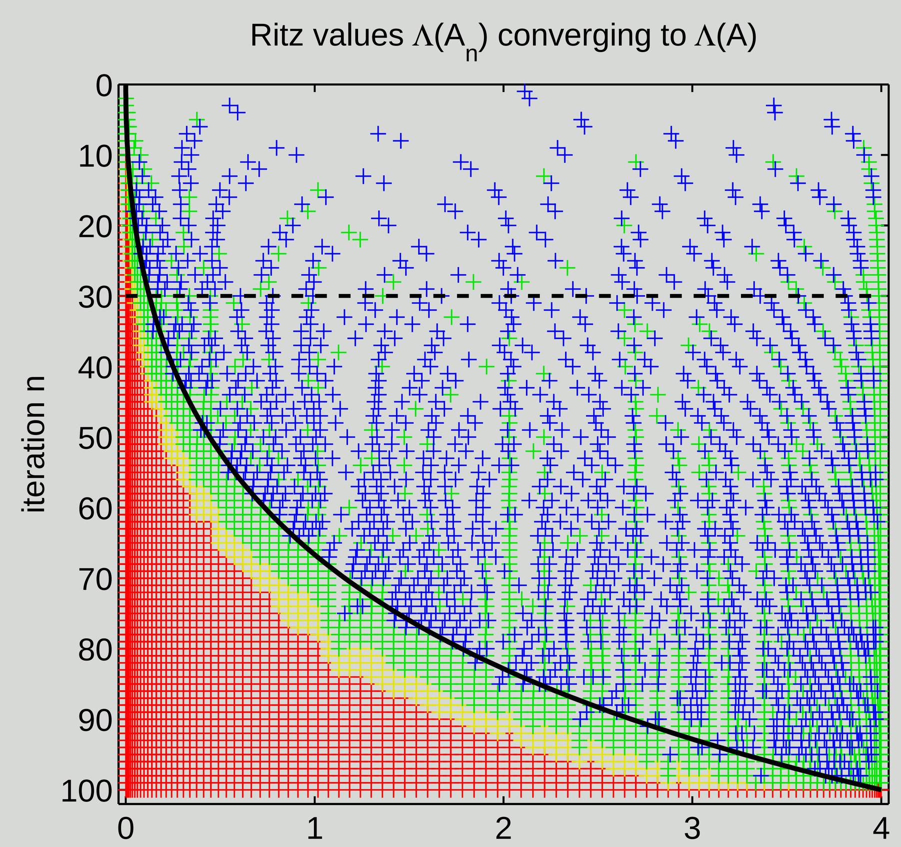
  - Let the poles $\xi_1, \ldots, \xi_n$ be described by a measure ν_t, $\|\nu_t\| = t = n/N$, e.g.,

$$\nu_t = t \cdot (\delta_0 + \delta_\infty)/2.$$

  - Then the distribution of Ritz values $\Lambda(A_n)$ is given as the constrained equilibrium measure $\mu_t \le \sigma$, $\|\mu_t\| = t$, which minimizes the energy $\mu \mapsto I(\mu, \mu) - 2I(\nu_t, \mu)$,

$$I(\mu_1, \mu_2) := \iint \log \frac{1}{|x - y|} \, d\mu_1(x) \, d\mu_2(y).$$

    Moreover, $S(t) := \mathrm{supp}(\sigma - \mu_t) \subseteq [\lambda_{\min}, \lambda_{\max}]$.



Inexact solves & error estimation

- In practical rational Arnoldi implementations the linear systems $(A - \xi_j I)x_j \approx v_j$ are often solved inexactly $\Longrightarrow$ Need to quantify effect on the Arnoldi approximation f_n.
- Collect residuals $r_j := v_j - (A - \xi_j I)x_j$ in $R_n = [r_1, \ldots, r_n]$, and derive an *inexact* rational Arnoldi decomposition

$$AV_{n+1}\underline{K_n} = V_{n+1}\underline{H_n} + R_n,$$

which can be rewritten as an *exact* decomposition [9]

$$(A + E_n)V_{n+1}\underline{K_n} = V_{n+1}\underline{H_n}, \quad E_n := -R_n \underline{K_n}^\dagger V_{n+1}^H.$$

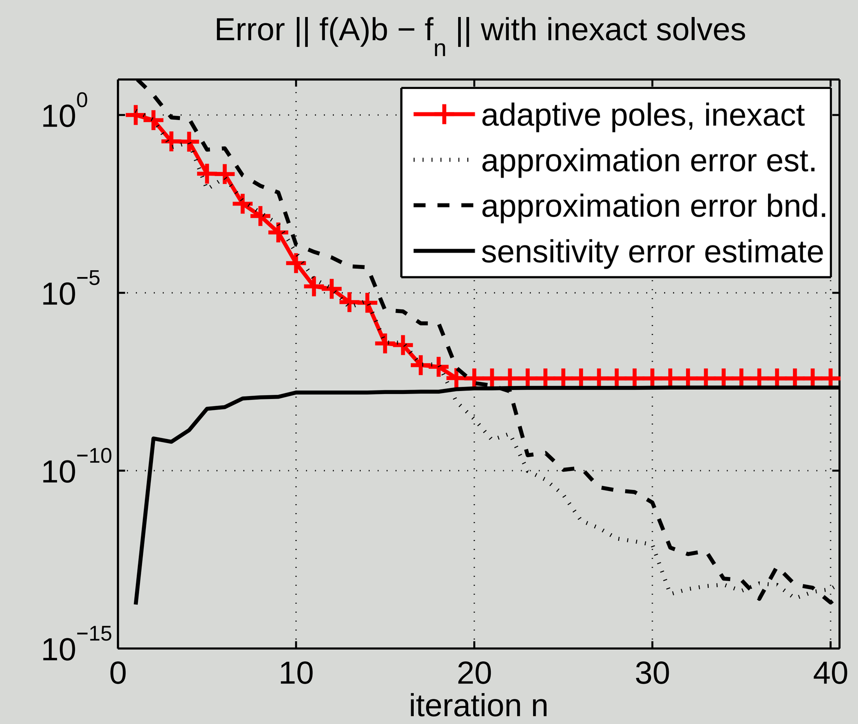
- If the projection $\widetilde{A}_n$ is computed from data $\{\underline{K_n}, \underline{H_n}\}$, the resulting Arnoldi approximation $\widetilde{f}_n$ is close to $f(A + E_n)b$.
- **Idea:** Decompose the error

$$\|f(A)b - \widetilde{f}_n\| \le \underbrace{\|f(A)b - f(A + E_n)b\|}_{\text{sensitivity error}} + \underbrace{\|f(A + E_n)b - \widetilde{f}_n\|}_{\text{approximation error}},$$

and estimate

$$\text{sensitivity error} \approx \|f(V_n^H A V_n)V_n^H b - f(\widetilde{A}_n)V_n^H b\|.$$

- In conjunction with approximation error estimator of [1, 6] we obtain a practical stopping criterion: terminate when the approximation error falls below the sensitivity error.
- **Example:** Compute $A^{-1/2}b$ for 2D-Laplacian, $N = 10^4$, using multigrid solver with `relres` $= 10^{-8}$.



Matlab code

available from `www.guettel.com/markovfunmv`

References

[1] M. Afanasjew, M. Eiermann, O. G. Ernst and S. Güttel, Linear Algebra Appl. **429**, 2293–2314 (2008).
[2] B. Beckermann and S. Güttel, Superlinear convergence of rational Arnoldi, in preparation.
[3] B. Beckermann, S. Güttel and R. Vandebril, SIAM J. Matrix Anal. Appl. **31**, 1740–1774 (2010).
[4] V. Druskin, C. Lieberman and M. Zaslavsky, SIAM J. Sci. Comput. **32**, 2485–2496 (2010).
[5] V. Druskin and V. Simoncini, Systems & Control Letters, to appear.
[6] S. Güttel, Rational Krylov Methods for Operator Functions (PhD thesis, TU Freiberg, 2010).
[7] S. Güttel and L. Knizhnerman, Automated pole selection for Markov functions, submitted.
[8] N. J. Higham, Functions of Matrices. Theory and Computation (SIAM, Philadelphia, 2008).
[9] R. B. Lehoucq and K. Meerbergen, SIAM J. Matrix Anal. Appl. **20**, 131–148 (1998).
[10] A. Ruhe, IMA Vol. Math. Appl. **60**, 149–164 (1994).