TECHNISCHE UNIVERSITÄT BERGAKADEMIE FREIBERG

Institut für Numerische Mathematik und Optimierung

On the Steepest Descent Method for the Evaluation of Matrix Functions

Stefan Güttel

joint work with M. Afanasjew, M. Eiermann, O. G. Ernst

Freiberg, April 2007

Problem

Given

- a matrix $A \in \mathbb{C}^{N \times N}$ (large and sparse),
- a vector $\boldsymbol{b} \in \mathbb{C}^N$, $\|\boldsymbol{b}\|_2 = 1$,
- a scalar function $f : \mathbb{C} \supseteq D \to \mathbb{C}$ which is analytic in $D \supset \Lambda(A)$.

Task: Compute $f(A)\boldsymbol{b}$ in an efficient way considering

- memory requirements,
- convergence speed.

Typical applications

- Solve the linear system of equations $A\boldsymbol{x} = \boldsymbol{b}$.
  The solution is $\boldsymbol{x} = f(A)\boldsymbol{b}$, where $f(z) = 1/z$.
- Solve an ordinary initial value problem $\boldsymbol{y}'(\tau) = A\boldsymbol{y}(\tau), \boldsymbol{y}(0) = \boldsymbol{y}_0$.
  The solution is $\boldsymbol{y}(\tau) = f(A)\boldsymbol{y}_0$, where $f(z) = f_\tau(z) = \exp(\tau z)$.
- The solution of the identification problem in stochastic semigroups requires the evaluation of $f(A)\boldsymbol{b}$ with $f(z) = \log(z)$. [Singer/Spilermann, 1976]
- To simulate Brownian motion of molecules one needs to determine $f(A)\boldsymbol{b}$ with $f(z) = \sqrt{z}$. [Ericsson, 1990]
- The solution of the algebraic Riccati equation (but also simulations in lattice quantum chromodynamics) requires $f(A)\boldsymbol{b}$ where $f(z) = \mathrm{sign}(z)$.

1 Matrix functions

Let

$$\psi_A(z) = \prod_{\lambda \in \Lambda(A)} (z - \lambda)^{d_\lambda}$$

denote the minimal polynomial of A, $\deg(\psi_A) = d$.

Let $p_A \in \mathscr{P}_{d-1}$ satisfy the d Hermite interpolation conditions

$$p_A^{(\nu)}(\lambda) = f^{(\nu)}(\lambda), \qquad \nu = 0, 1, \ldots, d_\lambda - 1$$

for all $\lambda \in \Lambda(A)$.

Then we define

$$f(A) := p_A(A).$$

[Gantmacher, 1959], [Rinehart, 1955]

2 Krylov approximations

The vector $f(A)\boldsymbol{b} = p_A(A)\boldsymbol{b}$ that we seek lies in some Krylov space

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

The polynomial $p_{A,\boldsymbol{b}}$ such that $f(A)\boldsymbol{b} = p_{A,\boldsymbol{b}}(A)\boldsymbol{b}$ is determined by at most d Hermite interpolation conditions at the eigenvalues of A (those relevant for $\boldsymbol{b}$).

Therefore it makes sense to look for approximations (Krylov approximations)

$$f(A)\boldsymbol{b} \approx p_{m-1}(A)\boldsymbol{b} \in \mathscr{K}_m(A, \boldsymbol{b})$$

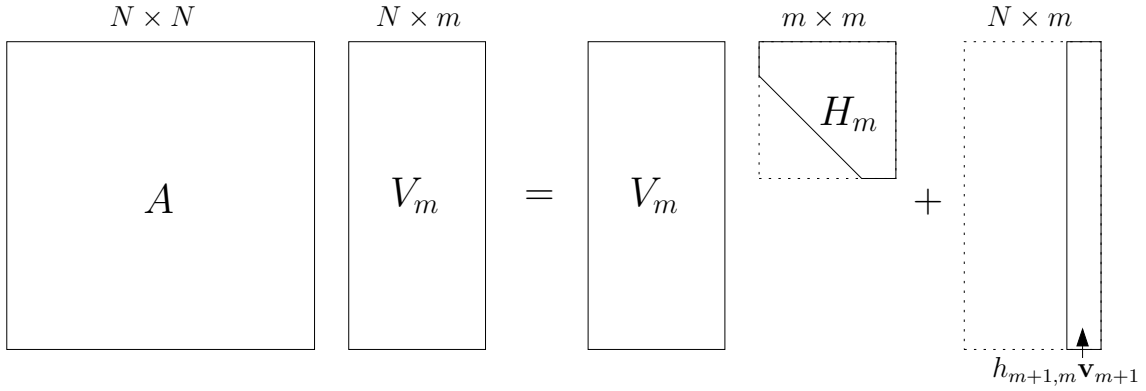
with p_{m-1} determined by interpolation conditions.

The Arnoldi process applied to A with initial vector $\boldsymbol{b}$ yields the Arnoldi decomposition

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

where

$$V_m = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m] \in \mathbb{C}^{N\times m}, \quad \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} \text{ ON basis of } \mathscr{K}_m(A, \boldsymbol{b}),$$
$$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m+1}\} \text{ ON basis of } \mathscr{K}_{m+1}(A, \boldsymbol{b}),$$
$$\boldsymbol{v}_1 = \boldsymbol{b},$$
$$H_m = [h_{i,j}] \in \mathbb{C}^{m\times m} \quad \text{unreduced upper Hessenberg matrix,}$$
$$\boldsymbol{e}_m = [0, \ldots, 0, 1]^T \in \mathbb{R}^m.$$



m-th Arnoldi approximation of $f(A)\boldsymbol{b}$:

$$f(A)\boldsymbol{b} \approx \boldsymbol{f}_m := V_m f(H_m)\boldsymbol{e}_1$$

[Druskin & Knizhnerman, 1989],
[Gallopoulos & Saad, 1992],
[Hochbruck & Lubich, 1995]

Interpretation: By definition there exists a polynomial $p_{m-1} \in \mathscr{P}_{m-1}$ such that

$$f(H_m) = p_{m-1}(H_m).$$

One can show

$$V_m p_{m-1}(H_m)\boldsymbol{e}_1 = p_{m-1}(A)\boldsymbol{b}.$$

Therefore, instead of evaluating $p_A(A)\boldsymbol{b}$ (interpolation at the eigenvalues of A) we compute $\boldsymbol{f}_m = p_{m-1}(A)\boldsymbol{b}$ (interpolation at the eigenvalues of H_m).

Advantage: avoids A (only $\boldsymbol{v} \to A\boldsymbol{v}$), avoids explicit interpolation, requires only evaluation of $f(H_m)$ for (small) matrix H_m.

Drawback: requires basis V_m (extensive storage), even in Hermitian case, high computational costs for Arnoldi decomposition (in the nonsymm. case).

3 Restarting the Arnoldi approximation

- Familiar from Krylov methods for $A\boldsymbol{x} = \boldsymbol{b}$ in non-Hermitian case.
- Idea: construct Arnoldi approximation to $A^{-1}\boldsymbol{b}$ from $\mathscr{K}_k(A, \boldsymbol{b})$:

$$\boldsymbol{x}_k = V_k H_k^{-1} \boldsymbol{e}_1 .$$

- Correction $\boldsymbol{c}$ to $\boldsymbol{x}_k$ obtained by solving the residual equation:

$$A^{-1}\boldsymbol{b} = \boldsymbol{x}_k + \boldsymbol{c}, \qquad A\boldsymbol{c} = \boldsymbol{b} - A\boldsymbol{x}_k =: \boldsymbol{r}_k .$$

- (Approximate) solution of residual equation in new Krylov space $\mathscr{K}_k(A, \boldsymbol{r}_k)$, hence storage requirements are fixed.
- General f (e.g., $f = \exp$): no residual available.

Consider two cycles of restarted Arnoldi for $A, \boldsymbol{b}$:

$$AV_k^{(1)} = V_k^{(1)} H_k^{(1)} + h_{k+1,k}^{(1)} \boldsymbol{v}_{k+1}^{(1)} \boldsymbol{e}_k^T, \qquad \boldsymbol{v}_1^{(1)} = \boldsymbol{b},$$
$$AV_k^{(2)} = V_k^{(2)} H_k^{(2)} + h_{k+1,k}^{(2)} \boldsymbol{v}_{k+1}^{(2)} \boldsymbol{e}_k^T, \qquad \boldsymbol{v}_1^{(2)} = \boldsymbol{v}_{k+1}^{(1)}.$$

Since columns of $W_{2k} := [V_k^{(1)}, V_k^{(2)}]$ form (nonorthogonal) basis of $\mathscr{K}_{2k}(A, \boldsymbol{b})$, we can combine to Arnoldi-like decomposition

$$AW_{2k} = W_{2k} H_{2k} + h_{k+1,k}^{(2)} \boldsymbol{v}_{k+1}^{(2)} \boldsymbol{e}_{2k}^T,$$

$$H_{2k} := \begin{bmatrix} H_k^{(1)} & O \\ h_{k+1,k}^{(1)} \boldsymbol{e}_1 \boldsymbol{e}_k^T & H_k^{(2)} \end{bmatrix}.$$

Arnoldi-like approximation: $\boldsymbol{f}_{2k} = W_{2k} f(H_{2k}) \boldsymbol{e}_1$.

Interpretation: $\boldsymbol{f}_{2k} = p_{2k-1}(A)\boldsymbol{b}$, where p_{2k-1} interpolates f at $\Lambda(H_{2k}) = \Lambda(H_k^{(1)}) \cup \Lambda(H_k^{(2)})$.

[Eiermann & Ernst, 2006]

4 Convergence

The simplest example: $k = 1$

Smallest possible restart length ($k = 1$) leads to

$$AV_m = V_m B_m + \beta_{m+1} \boldsymbol{v}_{m+1} \boldsymbol{e}_m^T$$

with

$$B_m = \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_m & \alpha_m \end{bmatrix}$$

where the columns $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ of V_m are a basis of $\mathscr{K}_m(A, \boldsymbol{b})$ such that $\|\boldsymbol{v}_j\| = 1$, $\boldsymbol{v}_j \perp \boldsymbol{v}_{j+1}$ $(j = 1, 2, \ldots)$.

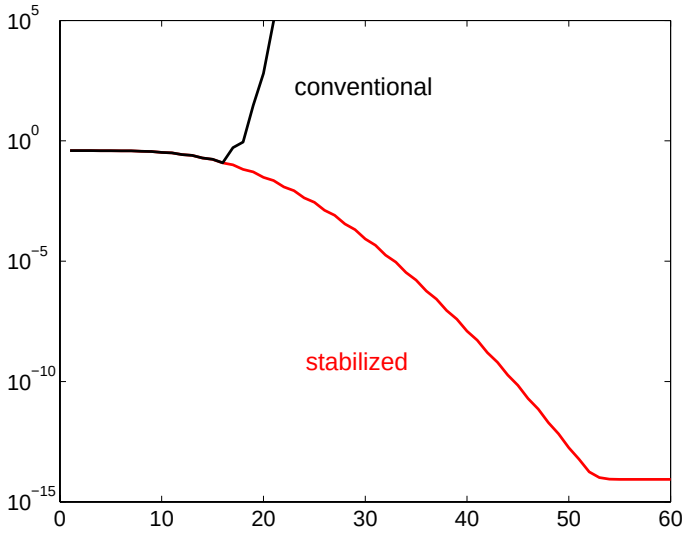
Now, with $\beta_1 = \|\boldsymbol{b}\| = 1$,

$$\boldsymbol{f}_{m+1} = \beta_1 [V_m \ \boldsymbol{v}_{m+1}] f\left(\begin{bmatrix} B_m & \boldsymbol{0} \\ \beta_{m+1}\boldsymbol{e}_m^T & \alpha_{m+1} \end{bmatrix} \right) \boldsymbol{e}_1 = \boldsymbol{f}_m + \beta_1 \left(\boldsymbol{e}_{m+1}^T f(B_{m+1}) \boldsymbol{e}_1 \right) \boldsymbol{v}_{m+1}$$

$$= \boldsymbol{f}_m + \left(\prod_{j=1}^{m+1} \beta_j \right) \ \Delta_f^m(\alpha(1:m+1)) \, \boldsymbol{v}_{m+1}$$

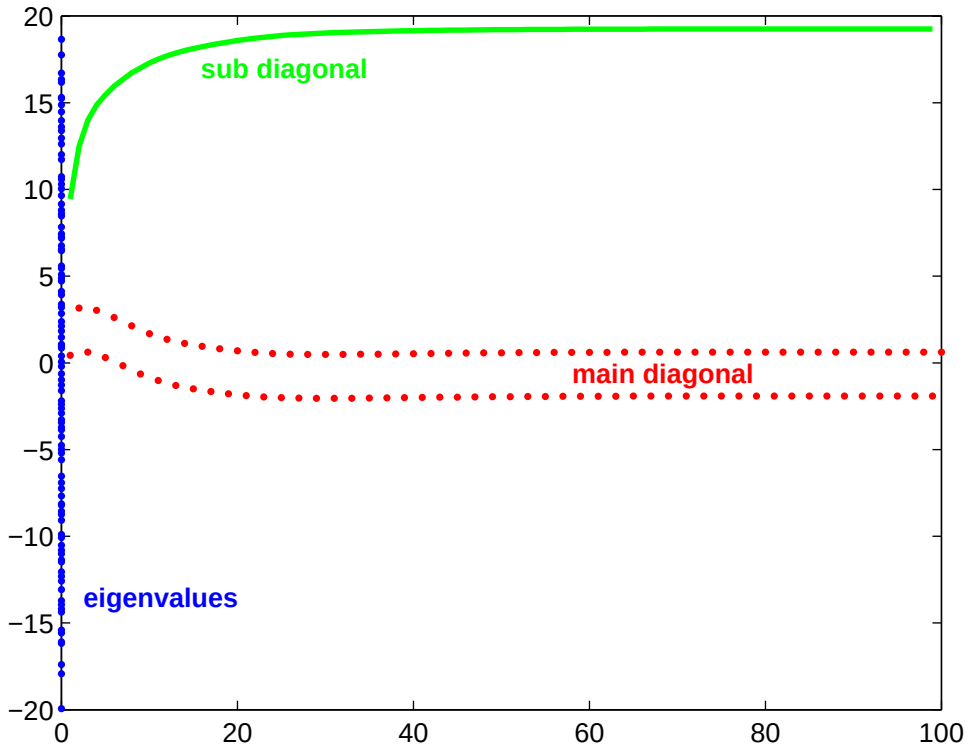
by a Theorem of [Opitz, 1964]. One can show: For $f(z) = 1/z$ and A spd this method coincides with the Steepest Descent Method.

The standard evaluation of the finite differences leads to dramatic cancellation:



Assume that

- A is Hermitian with eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_N$ and eigenvectors $\boldsymbol{x}_j$, $\|\boldsymbol{x}_j\| = 1$.
- $\boldsymbol{b} = \sum_j \gamma_j \boldsymbol{x}_j$, $\|\boldsymbol{b}\|^2 = \sum_j |\gamma_j|^2 = 1$, $\gamma_j \neq 0$.



Theorem 1 *There are $c, s \in \mathbb{R} \setminus \{0\}$, $c^2 + s^2 = 1$, depending on λ_j, $|\gamma_j|^2$ such that*

$$\lim_{m\to\infty} \alpha_{2m} = \zeta_1 := c^2\lambda_1 + s^2\lambda_N,$$
$$\lim_{m\to\infty} \alpha_{2m+1} = \zeta_2 := s^2\lambda_1 + c^2\lambda_N.$$

Thus, if f has finite singularities,

$$\limsup_{m\to\infty} \|f(A)\boldsymbol{b} - \boldsymbol{f}_m\|^{1/m} = \limsup_{m\to\infty} \left[\max_{1\le j\le N} |f(\lambda_j) - q_{m-1}(\lambda_j)|\right]^{1/m} = \frac{\kappa_A}{\kappa_f},$$

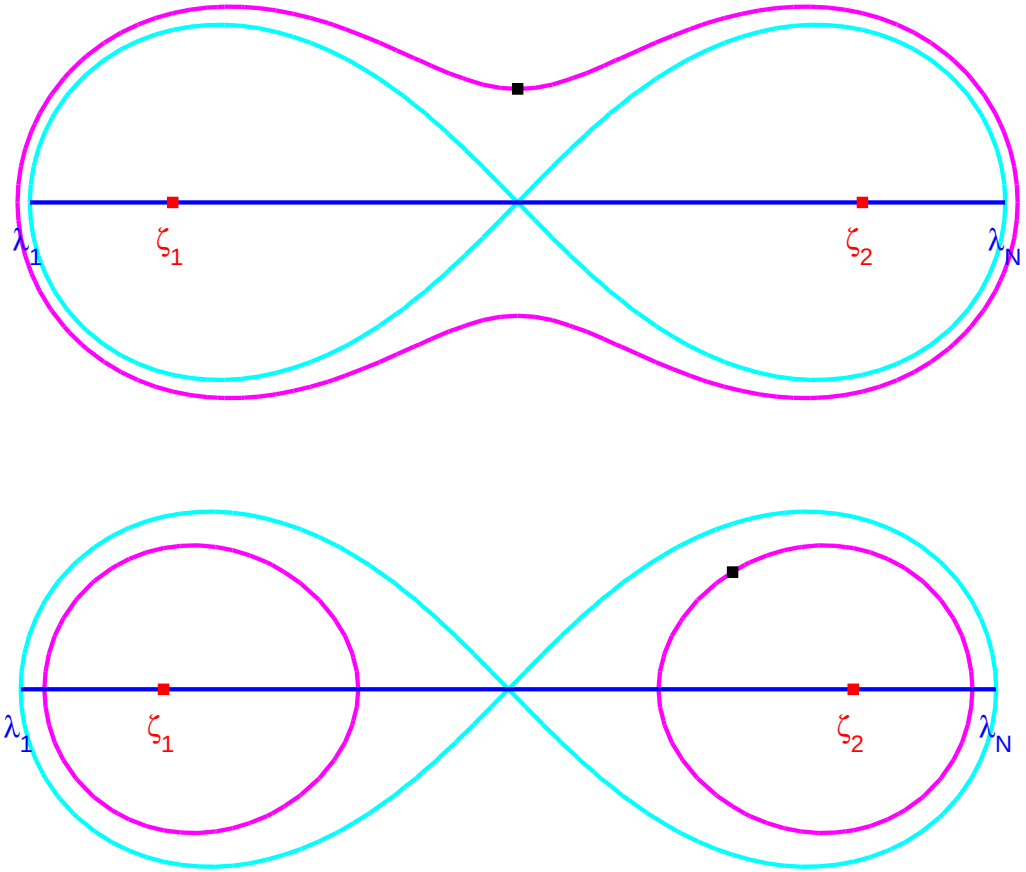
where $q_{m-1} \in \mathscr{P}_{m-1}$ interpolates f in ζ_1 (with multiplicity $\lfloor m/2+1 \rfloor$) and in ζ_2 (with multiplicity $\lfloor m/2 \rfloor$).

$$\Gamma_\rho := \{z \in \mathbb{C} : |z-\zeta_1||z-\zeta_2| = \rho^2\}$$

(lemniscate with foci ζ_1, ζ_2)

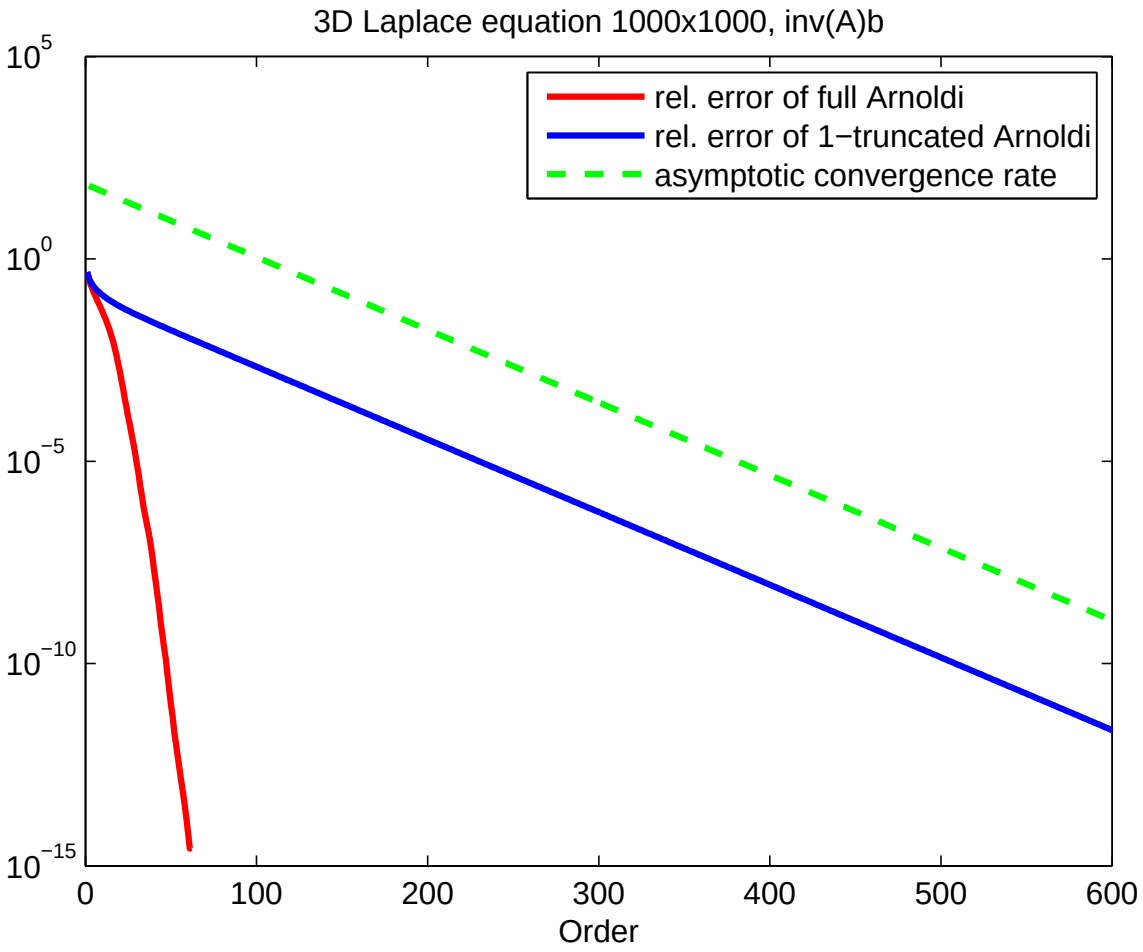
$$\kappa_A := \min\{\rho > 0 : \Lambda(A) \subset \operatorname{int}\Gamma_\rho \cup \Gamma_\rho\},$$

$$\kappa_f := \max\{\rho > 0 : f \text{ analytic in } \operatorname{int}\Gamma_\rho\}.$$



Summary. For $k = 1$, the restarted Arnoldi method is asymptotically equivalent to cyclic interpolation at two nodes which are both convex combinations of λ_1 and λ_N.

If f is an analytic (but not an entire) function then the restarted Arnoldi method converges (diverges) asymptotically like θ^m, where θ is determined by the singularities of f and the spectrum of A.


3D Laplace equation 1000x1000, inv(A)b
rel. error of full Arnoldi
rel. error of 1−truncated Arnoldi
asymptotic convergence rate
10^5
10^0
10^−5
10^−10
10^−15
0
100
200
300
400
500
600
Order

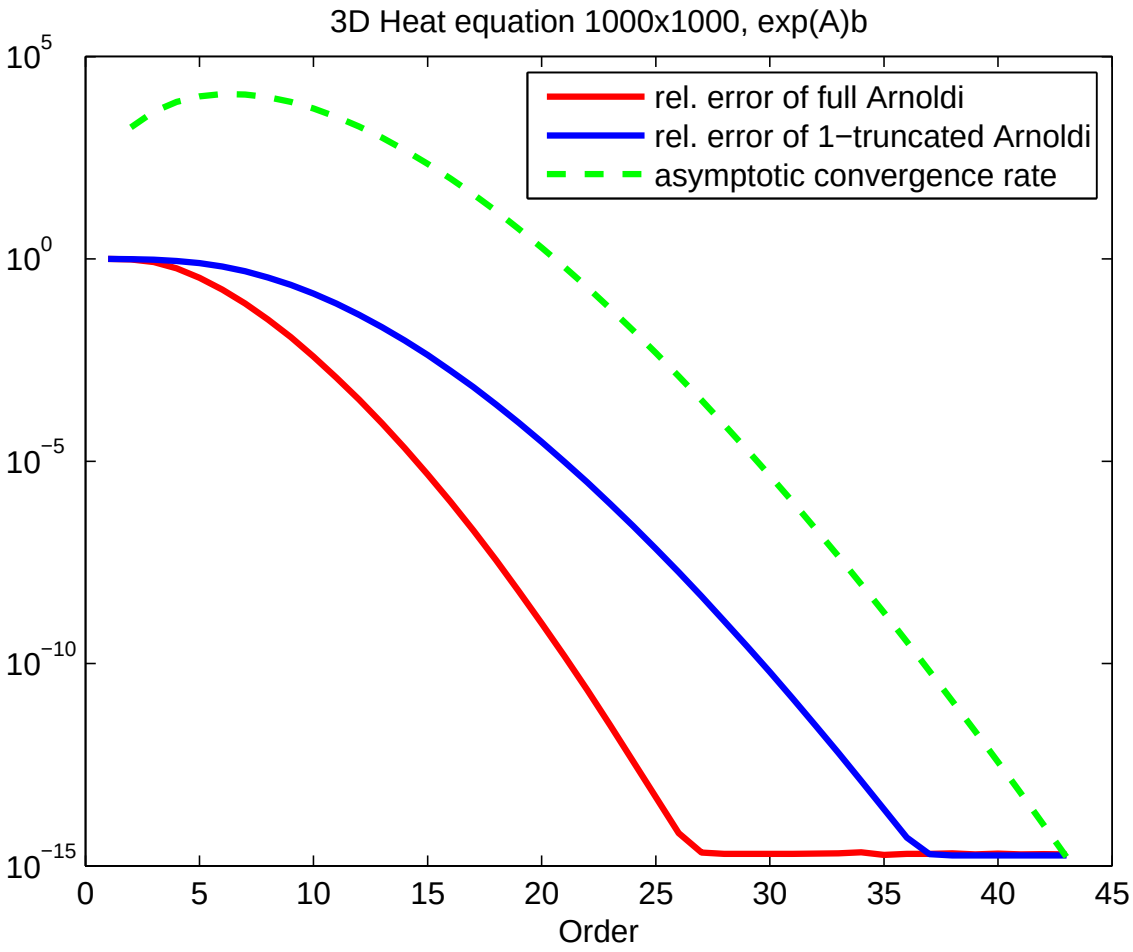
Theorem 2 *If* $f(z) = \exp(\tau z)$, $\tau > 0$, *then*

$$\begin{aligned}\limsup_{m\to\infty} \left[m! \, \|f(A)\boldsymbol{b} - \boldsymbol{f}_m\|\right]^{1/m} &= \limsup_{m\to\infty} \left[m! \left(\max_{1\le j\le N} |f(\lambda_j) - q_{m-1}(\lambda_j)|\right)\right]^{1/m} \\ &= \tau\delta,\end{aligned}$$

where $q_{m-1} \in \mathscr{P}_{m-1}$ *interpolates* f *in* ζ_1 *(with multiplicity* $\lfloor m/2+1 \rfloor$*) and in* ζ_2 *(with multiplicity* $\lfloor m/2 \rfloor$*) and*

$$\delta = \min\left\{\delta \,:\, |\lambda - \zeta_1||\lambda - \zeta_2| \le \delta^2 \quad \forall \lambda \in \Lambda(A)\right\}.$$

Summary. If f is an entire function of order 1 and type τ then the restarted Arnoldi method ($k = 1$) converges (always!) superlinearly like $(\tau\delta/m)^m$.


3D Heat equation 1000x1000, exp(A)b
rel. error of full Arnoldi
rel. error of 1−truncated Arnoldi
asymptotic convergence rate
10^{5}
10^{0}
10^{-5}
10^{-10}
10^{-15}
0
5
10
15
20
25
30
35
40
45
Order

Tools used in proof. Consider the basis vectors

$$\boldsymbol{v}_1 = \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|},\ \boldsymbol{v}_{m+1} = \frac{A\boldsymbol{v}_m - \alpha_m \boldsymbol{v}_m}{\beta_{m+1}} = \frac{A\boldsymbol{v}_m - (\boldsymbol{v}_m^T A\boldsymbol{v}_m)\boldsymbol{v}_m}{\|A\boldsymbol{v}_m - (\boldsymbol{v}_m^T A\boldsymbol{v}_m)\boldsymbol{v}_m\|},\ m = 1, 2, \ldots$$

Write $\boldsymbol{v}_m = \sum_{j=1}^{N} \gamma_{m,j}\boldsymbol{x}_j$, $\sum_j |\gamma_{m,j}|^2 = 1$, and define discrete probability distributions

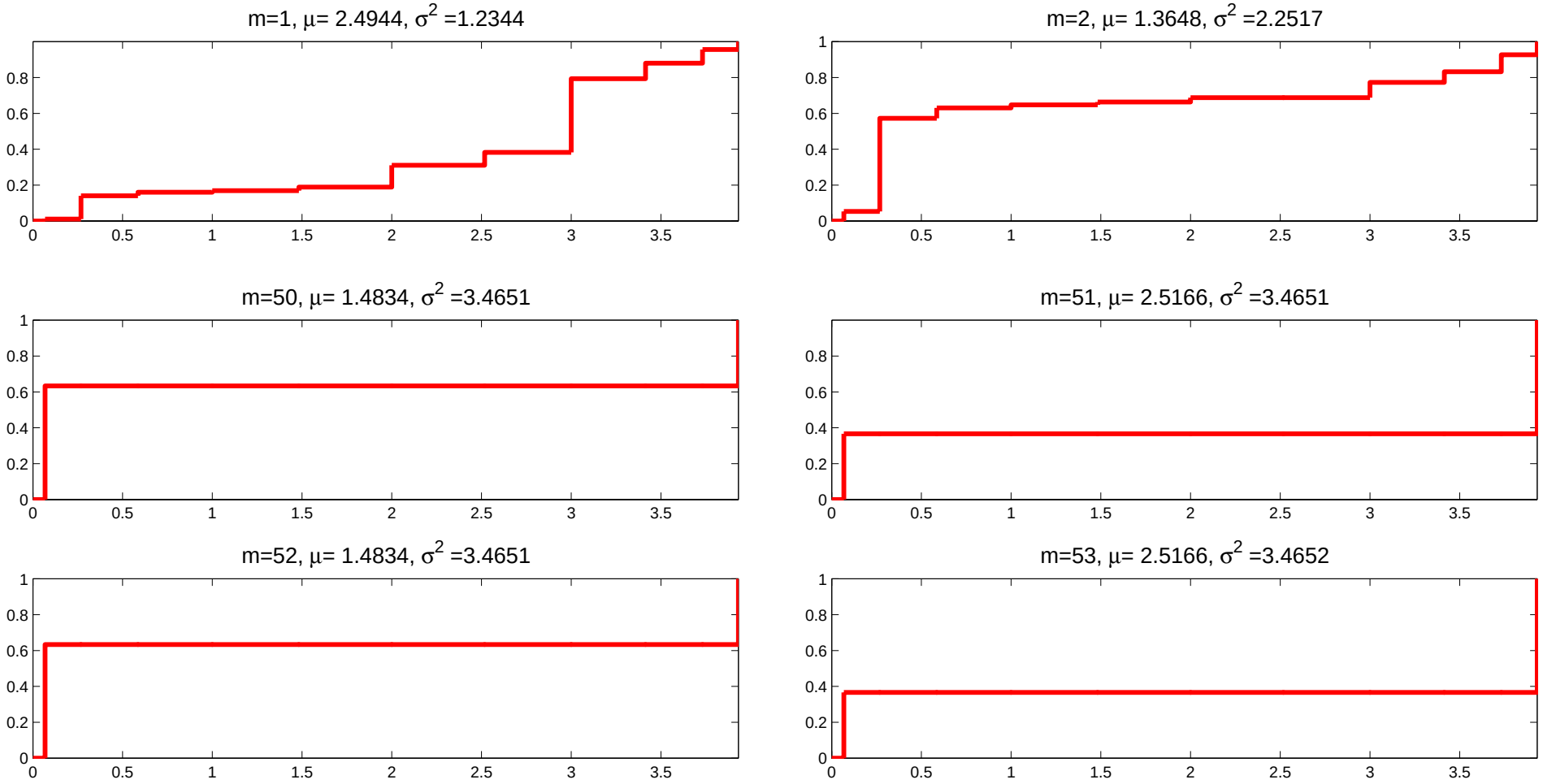
$$w_m(x) := \begin{cases} 0, & x < \lambda_1, \\ \sum_{j=1}^{k} |\gamma_{m,j}|^2, & \lambda_k \le x < \lambda_{k+1}, \\ 1, & \lambda_N \le x \end{cases}$$

with

$$\text{means} \qquad \mu_m = \sum \lambda_j |\gamma_{m,j}|^2 = \boldsymbol{v}_m^T A \boldsymbol{v}_m,$$

$$\text{and variances} \qquad \sigma_m^2 = \sum (\lambda_j - \mu_m)^2 |\gamma_{m,j}|^2 = \|A\boldsymbol{v}_m - (\boldsymbol{v}_m^T A\boldsymbol{v}_m)\boldsymbol{v}_m\|^2.$$

(Note that $\mu_m = \alpha_m$ and $\sigma_m^2 = \beta_{m+1}^2$.)



The (non-linear) transformation $\boldsymbol{v}_{m+1} = T(\boldsymbol{v}_m)$ translates into a transformation $w_{m+1} = T(w_m)$ of the corresponding distributions.

Hirotuko Akaike [1959] showed:

Theorem 3 *For any discrete probability distribution* w*,*

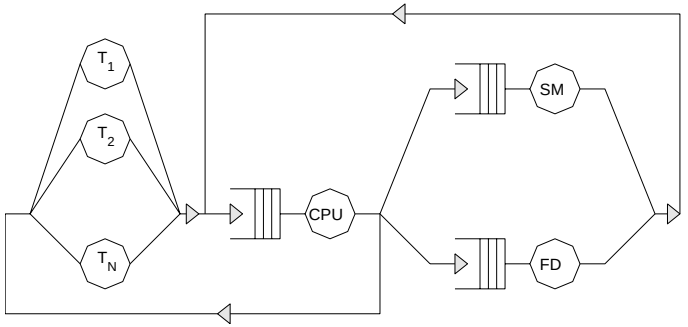
$$\operatorname{Var} T(w) \geq \operatorname{Var} w$$

with equality holding if and only if the support of w *consists of two points.*

The sequence $\{T^m(w)\}$ *ultimately alternates between two probability distributions which are both supported on* λ_1 *and* λ_N*.*

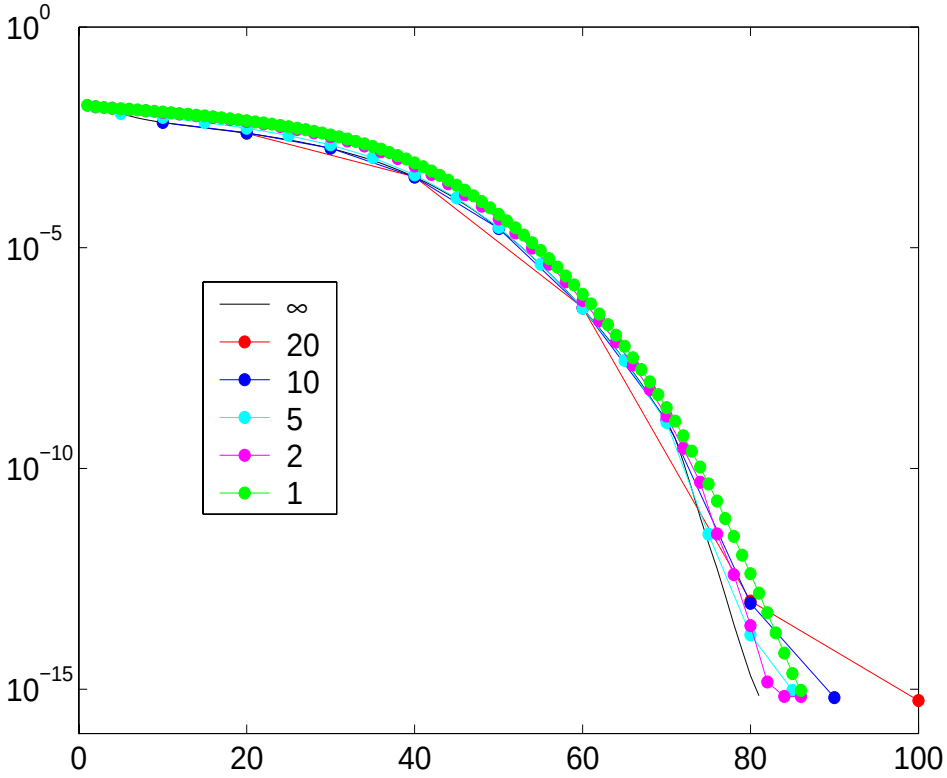
In addition, well known results on the interpolation of analytic functions (see [Walsh, 1969], [Gaier, 1980]) and new results on the interpolation of entire functions are used. □

Summary. For $k = 1$, the restarted Arnoldi method is asymptotically equivalent to interpolation at two nodes which are both convex combinations of λ_1 and λ_N.



Last example: time-continuous Markov chain

[Philippe, Saad & Stewart, 1996]
$N = 62169$, $\rho(A) = 70$.



m	k	time [s]
∞	1	75.4
20	5	11.0
10	9	5.8
5	17	3.7
2	43	3.2
1	86	3.4

Summary

- The presented method requires only minimal storage costs even for very large matrices.
- The asymptotic convergence behavior is (nearly) understood (up to the location of ζ_1, ζ_2).

Interesting questions

- Impact on the convergence analysis of restarted/truncated Arnoldi method for $k > 1$?
- Non-Hermitian, non-normal case?

Vielen Dank für Ihre Aufmerksamkeit...

... und Guten Appetit!