Institut für Numerische Mathematik und Optimierung

Thick-restart Arnoldi methods for the evaluation of matrix functions

Michael Eiermann

joint work with Oliver G. Ernst and Stefan Güttel

MIMS New Directions Workshop, Functions of Matrices
University of Manchester, May 15-16, 2008

1 Problem

Given: $A \in \mathbb{C}^{n \times n}$, $\boldsymbol{b} \in \mathbb{C}^n$, $\boldsymbol{b} \neq \boldsymbol{0}$, f analytic in neighborhood of $\Lambda(A)$.

Sought: $f(A)\boldsymbol{b}$.

Original Motivation: Numerical simulation of transient electromagnetic (TEM) geophysical exploration (collaboration with Institute of Geophysics):

$$\boldsymbol{u}(t) = \exp(-tA)\boldsymbol{u}_0,$$

where A discretizes $\sigma^{-1}\nabla \times (\mu^{-1}\nabla \times \cdot\,)$ and is large and sparse.

Other Important Applications:
Exponential integrators: $\varphi_0(\lambda) = \exp(-t\lambda)$, $\varphi_{j+1}(\lambda) = \frac{\varphi_j(t\lambda)-\varphi_j(0)}{t\lambda}$.
Lattice quantum chromodynamics: $\operatorname{sign}(\lambda)$.
Time-dependent hyperbolic problems: trigonometric functions.

- The Arnoldi method projects the problem of evaluating $f(A)\boldsymbol{b}$ onto a sequence of m-dimensional Krylov subspaces.
- The cost of storing (and computing) the basis vectors of these spaces increases with m.
- It is possible to restart the Arnoldi method after a fixed dimension m similar to linear systems or eigenproblems.
- Restarting usually results in slower convergence.
- How can we compensate for the loss of information that occurs upon restarting by retaining a judiciously chosen part of the previously generated spaces?

Outline

- Problem
- Three ways to generate Krylov subspace approximations
- Thick restarts
- Convergence
- A numerical example
- Summary

2 Three ways to generate Krylov subspace approximations

- Krylov subspace methods generate approximants $\boldsymbol{y}_m$ of $f(A)\boldsymbol{b}$ with

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

- There are usually based on Arnoldi-like decompositions of A,

$$AW_m = W_m H_m + h_{m+1,m}\boldsymbol{w}_{m+1}\boldsymbol{e}_m^T,$$

  where $\operatorname{colspan}(W_m) = \mathscr{K}_m(A, \boldsymbol{b})$, $\beta W_m \boldsymbol{e}_1 = \boldsymbol{b}$, H_m is unreduced upper Hessenberg.

- Most prominent example: (proper) Arnoldi decomposition, $AV_m = V_m H_m + h_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^T$, where $V_m^H V_m = I_m$.

- restarted Arnoldi: k standard Arnoldi decompositions of A

$$AV_j = V_j H_j + h_{j+1} \boldsymbol{v}_{jm+1} \boldsymbol{e}_m^T, \quad j = 1, 2, \ldots, k,$$

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

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

where $\widehat{V}_k := [V_1 \, V_2 \, \cdots \, V_k] \in \mathbb{C}^{n \times km}$,

$$\widehat{H}_k := \begin{bmatrix} H_1 & & & \\ E_2 & H_2 & & \\ & \ddots & \ddots & \\ & & E_k & H_k \end{bmatrix} \in \mathbb{C}^{km \times km}, \qquad E_j := h_j \boldsymbol{e}_1 \boldsymbol{e}_m^T \in \mathbb{R}^{m \times m},$$

cf. [E. & Ernst, 2006].

Projection: With $\beta \boldsymbol{w}_1 = \boldsymbol{b}$ set

$$\boldsymbol{y}_m^{(1)} := \beta W_m f(H_m) \boldsymbol{e}_1 \in \mathscr{K}_m(A, \boldsymbol{b}).$$

Cauchy integral: $f(A)\boldsymbol{b} = \frac{1}{2\pi i} \int_\Gamma f(\lambda) \boldsymbol{x}(\lambda)\, d\lambda$. Approximate $\boldsymbol{x}(\lambda) := (\lambda I - A)^{-1} \boldsymbol{b}$ by $\boldsymbol{z}_m(\lambda) = \beta W_m (\lambda I - H_m)^{-1} \boldsymbol{e}_1$ and set

$$\boldsymbol{y}_m^{(2)} := \frac{1}{2\pi i} \int_\Gamma f(\lambda) \boldsymbol{z}_m(\lambda)\, d\lambda.$$

Interpolation: Let $p_{m-1} \in \mathscr{P}_{m-1}$ be the interpolating polynomial (in Hermite's sense) for f at the eigenvalues of H_m and set

$$\boldsymbol{y}_m^{(3)} := p_{m-1}(A) \boldsymbol{b}.$$

Theorem 1 $\boldsymbol{y}_m^{(1)} = \boldsymbol{y}_m^{(2)} = \boldsymbol{y}_m^{(3)}$

cf. [Hochbruck & Hochstenbach, 2005]

- Arnoldi approximation [Druskin & Knizhnerman, 1989], [Saad, 1992]

$$\boldsymbol{f}_m = p_{m-1}(A)\boldsymbol{b} = \frac{1}{2\pi i}\int_\Gamma f(\lambda)\boldsymbol{z}_m(\lambda)\,d\lambda = \beta V_m f(H_m)\boldsymbol{e}_1,$$

where $p_{m-1} \in \mathscr{P}_{m-1}$ interpolates f at the Ritz values of A wrt $\mathscr{K}_m(A, \boldsymbol{b})$, where $\boldsymbol{z}_m(\lambda)$ is the FOM approximation to $(\lambda I_n - A)\boldsymbol{x}(\lambda) = \boldsymbol{b}$.

- Restarted Arnoldi approximation [E. & Ernst, 2006], [Niehoff, 2006]

$$\widehat{\boldsymbol{f}}_k = \widehat{p}_k(A)\boldsymbol{b} = \frac{1}{2\pi i}\int_\Gamma f(\lambda)\widehat{\boldsymbol{z}}_k(\lambda)\,d\lambda = \beta \widehat{V}_k f(\widehat{H}_k)\boldsymbol{e}_1,$$

where $\widehat{p}_k \in \mathscr{P}_{km-1}$ interpolates f at the Ritz values of A wrt $\mathscr{K}_m(A, \boldsymbol{v}_{(j-1)m+1})$ $(j = 1, 2, \ldots, k)$, where $\widehat{\boldsymbol{z}}_k(\lambda)$ is the FOM(m) approximation (after k cycles) to $(\lambda I_n - A)\boldsymbol{x}(\lambda) = \boldsymbol{b}$.

3 Thick restarts

- Compensate for the deterioration of convergence of Krylov subspace methods due to restarting by using nearly invariant subspaces to augment the Krylov subspace.
- Identify a subspace which slows convergence, approximate this space and eliminate its influence from the iteration process.
- In practice: Approximate eigenspaces which belong to 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].

Thick-restart procedure

- Starting point: $\mathscr{K}_m(A, \boldsymbol{b})$ with Arnoldi decomposition

$$AV_1 = V_1 H_1 + h_2 \boldsymbol{v}_{m+1} \boldsymbol{e}_m^T.$$

- Compute ℓ-dimensional H_1-invariant subspace,

$$H_1 \, [X_1 \; *] = [X_1 \; *] \begin{bmatrix} U_1 & * \\ O & * \end{bmatrix}$$

(partial Schur decomposition), i.e., $U_1 \in \mathbb{C}^{\ell \times \ell}$ is upper triangular, $X_1 \in \mathbb{C}^{m \times \ell}$ has orthonormal columns.
Set $Y_1 := V_1 X_1$. Then

$$AY_1 = Y_1 U_1 + h_2 \boldsymbol{v}_{m+1} \boldsymbol{u}_1^T, \text{ where } \boldsymbol{u}_1 = X_1^T \boldsymbol{e}_\ell \in \mathbb{C}^\ell \text{ (dense!)}.$$

- Extend by m Arnoldi steps

$$A\,[Y_1\ V_2] = [Y_1\ V_2] \begin{bmatrix} U_1 & G_2 \\ h_2\,\boldsymbol{e}_1\,\boldsymbol{u}_1^T & H_2 \end{bmatrix} + h_3\,\boldsymbol{v}_{2m+1}\,\boldsymbol{e}_{\ell+m}^T,$$

where $[Y_1\ V_2\ \boldsymbol{v}_{2m+1}]$ has orthonormal columns, $V_2\,\boldsymbol{e}_1 = \boldsymbol{v}_{m+1}$, $H_2 \in \mathbb{C}^{m\times m}$ is upper Hessenberg.

- Known [Morgan, 2002]: $\mathrm{colspan}([Y_1\ V_2]) = \mathscr{K}_{\ell+m}(A, s(A)\boldsymbol{b})$, where

$$s(\lambda) = \prod_{\mu\in\Lambda(H_1)\setminus\Lambda(U_1)} (\lambda - \mu) \in \mathscr{P}_{m-\ell}$$

(implicitly restarted Arnoldi [Sorensen, 1992]).

- If $A = A^H$: U_1 is diagonal, H_2 is symmetric tridiagonal, $G_2 = h_2\overline{\boldsymbol{u}}_1\,\boldsymbol{e}_1^T$.

- Second sweep: Compute ℓ-dimensional invariant subspace,

$$\begin{bmatrix} U_1 & G_2 \\ h_2 \boldsymbol{e}_1 \boldsymbol{u}_1^T & H_2 \end{bmatrix} [X_2 \ *] = [X_2 \ *] \begin{bmatrix} U_2 & * \\ O & * \end{bmatrix}.$$

- With $Y_2 := [Y_1 \ V_2] X_2$,

$$AY_2 = Y_2 U_2 + h_3 \boldsymbol{v}_{2m+1} \boldsymbol{u}_2^T, \text{ where } \boldsymbol{u}_2 = X_2^T \boldsymbol{e}_{\ell+m} \in \mathbb{C}^\ell.$$

- Extend by m further Arnoldi steps

$$A [Y_2 \ V_3] = [Y_2 \ V_3] \begin{bmatrix} U_2 & G_3 \\ h_3 \boldsymbol{e}_1 \boldsymbol{u}_2^T & H_3 \end{bmatrix} + h_4 \boldsymbol{v}_{3m+1} \boldsymbol{e}_{\ell+m}^T,$$

where $[Y_2 \ V_3 \ \boldsymbol{v}_{3m+1}]$ has orthonormal columns, $V_3 \boldsymbol{e}_1 = \boldsymbol{v}_{2m+1}$, H_3 is upper Hessenberg.

We glue these decompositions together,

$$A\,[V_1\,Y_1\,V_2\,Y_2\,V_3] = [V_1\,Y_1\,V_2\,Y_2\,V_3]\left[\begin{array}{c||c|c||c|c} H_1 & O & O & O & O \\ \hline\hline O & U_1 & G_2 & O & O \\ \hline E_2 & F_2 & H_2 & O & O \\ \hline\hline O & O & O & U_2 & G_3 \\ \hline O & O & E_3 & F_3 & H_3 \end{array}\right]$$

$$+h_4\,\boldsymbol{v}_{3m+1}\,\boldsymbol{e}_{3m+2\ell}^T.$$

Here $E_j = h_j\,\boldsymbol{e}_1\,\boldsymbol{e}_m^T \in \mathbb{R}^{m\times m}$, $F_j = h_j\,\boldsymbol{e}_1\,\boldsymbol{u}_{j-1}^T \in \mathbb{C}^{m\times \ell}$.

After k sweeps, we arrive at a "thick-restart decomposition"

$$A\widetilde{V}_k = \widetilde{V}_k \widetilde{H}_k + h_{k+1} \boldsymbol{v}_{km+1} \boldsymbol{e}_{\widehat{k}}^T, \text{ where}$$

$$\widetilde{V}_k = [V_1 | Y_1 | V_2 | \cdots | Y_{k-1} | V_k] \in \mathbb{C}^{n \times \widehat{k}} \text{ has linearly dependent columns,}$$

$$\widetilde{H}_k = \left[\begin{array}{c|cc|cc|cc}
H_1 & & & & & & \\ \hline
 & U_1 & G_2 & & & & \\
E_2 & F_2 & H_2 & & & & \\ \hline
 & \ddots & & \ddots & & & \\ \hline
 & & & & & U_{k-1} & G_k \\
 & & & & E_k & F_k & H_k
\end{array}\right] \in \mathbb{C}^{\widehat{k} \times \widehat{k}}$$

is not Hessenberg ($\widehat{k} = km + (k-1)\ell$).

We need km mvm's to construct this decomposition.

From the decomposition $A\widetilde{V}_k = \widetilde{V}_k \widetilde{H}_k + h_{k+1} \boldsymbol{v}_{km+1} \boldsymbol{e}_{\widehat{k}}^T$, we define

$$\widetilde{\boldsymbol{f}}_k := \beta \widetilde{V}_k f(\widetilde{H}_k) \boldsymbol{e}_1.$$

Since $Y_1 = V_1 X_1$ and $Y_j = [Y_{j-1}\ V_j]\, X_j$ $(j = 2, \ldots, k)$, we write

$$\widetilde{V}_k = [V_1\, Y_1\, V_2\, \cdots\, Y_{k-1}\, V_k] = [V_1\, V_2\, \cdots\, V_k]\; C =: \widehat{V}_k C,$$

where $C \in \mathbb{C}^{mk \times \widehat{k}}$ has full row rank.
We have $CC^\dagger = I_{km}$ and $\boldsymbol{e}_{\widehat{k}}^T C^\dagger = \boldsymbol{e}_{km}$. Thus, by inserting

$$A\widehat{V}_k C = \widehat{V}_k C \widetilde{H}_k + h_{k+1} \boldsymbol{v}_{km+1} \boldsymbol{e}_{\widehat{k}}^T$$

or $$A\widehat{V}_k = \widehat{V}_k \left(C \widetilde{H}_k C^\dagger \right) + h_{k+1} \boldsymbol{v}_{km+1} \boldsymbol{e}_{\widehat{k}}^T C^\dagger =: \widehat{V}_k \widehat{H}_k + h_{k+1} \boldsymbol{v}_{km+1} \boldsymbol{e}_{km}^T$$

which is a valid Arnoldi-like decomposition, i.e., $\widehat{H}_k$ is upper Hessenberg and the columns of $\widehat{V}_k$ are linearly independent.

Theorem 2 *Given the thick-restart decomposition*

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

(k-th sweep, i.e., after $k-1$ restarts, ℓ Ritz vectors per restart, m mvm per sweep) and the associated Arnoldi-like decomposition

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

Then

$$\widetilde{\boldsymbol{f}}_k = \beta \, \widetilde{V}_k \, f(\widetilde{H}_k) \boldsymbol{e}_1 = \beta \, \widehat{V}_k \, f(\widehat{H}_k) \boldsymbol{e}_1.$$

Three interpretations:

Theorem 3 *For the thick-restart approximation there holds*

$$\widetilde{\boldsymbol{f}}_k = \beta \widetilde{V}_k f(\widetilde{H}_k)\boldsymbol{e}_1 = \widehat{p}_k(A)\boldsymbol{b} = \frac{1}{2\pi i}\int_\Gamma f(\lambda)\widehat{\boldsymbol{z}}_k^{(m,\ell)}(\lambda)\,d\lambda,$$

where $\widehat{p}_k \in \mathscr{P}_{km-1}$ *interpolates* f *in*

$$\Lambda(\widehat{H}_k) = \Lambda(\widetilde{H}_k) \setminus \left(\cup_{j=1}^{k-1}\Lambda(U_j)\right),$$

and where $\widehat{\boldsymbol{z}}_k^{(m,\ell)}(\lambda)$ *is the approximate solution of* $(\lambda I - A)\boldsymbol{x}(\lambda) = \boldsymbol{b}$ *which is generated by* k *sweeps of* $FOM(m,\ell)$ *(cf. [Morgan, 2002]).*

cf. [Niehoff, 2006]

4 Convergence

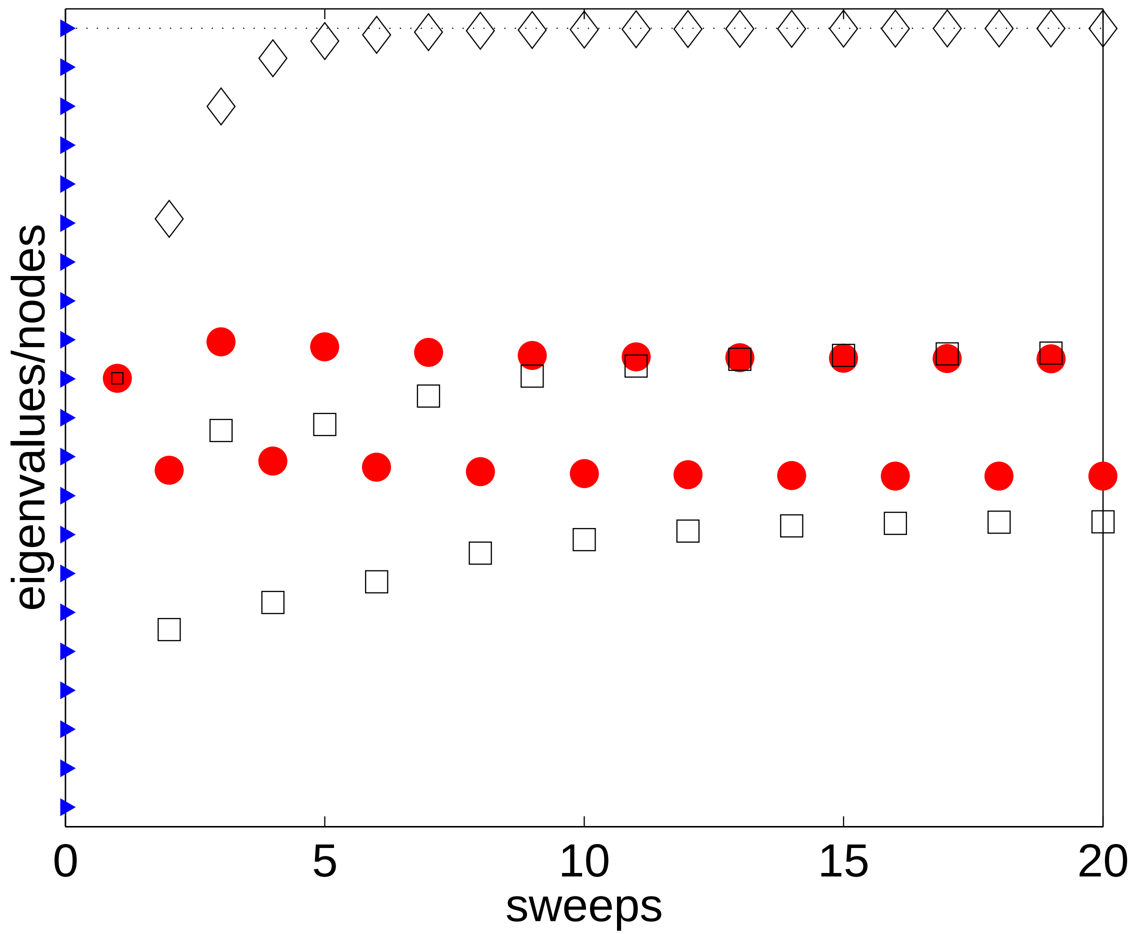
Use the interpretation as an interpolation procedure.

Programm:

1. Where in the complex plane is $\Lambda(\widehat{H}_k)$, the set of interpolation points, located?
2. For which $\lambda \in \mathbb{C}$ do the corresponding interpolation polynomials converge to $f(\lambda)$?

Remarks:

1. This approach works only for (nearly) normal A.
2. The second question is answered, e.g., by [Walsh, 1969].

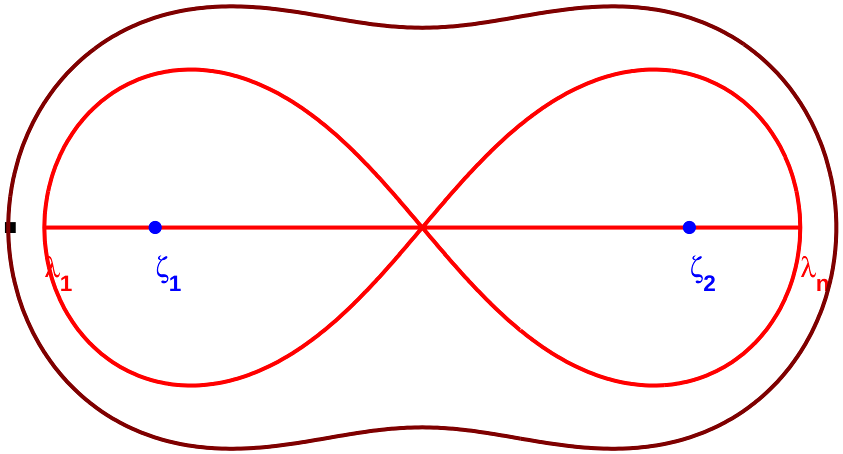


A Hermitian with eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_n$: ▶
Nodes for standard restarted Arnoldi $(m = 1)$: ●
Nodes for thick-restart Arnoldi $(m = 1, \ell = 1)$: $\square$ + last $\diamondsuit$

Theorem 4 (Afanasjew et al., 2008) *Let A be Hermitian. Consider the restarted Arnoldi method with restart length $m = 1$: $\Lambda(\widehat{H}_k) = \{\eta_1, \ldots, \eta_k\}$ and $\Lambda(\widehat{H}_{k+1}) = \{\eta_1, \ldots, \eta_k, \eta_{k+1}\}$. There exists $\alpha \in (0,1)$ (which depends on $\boldsymbol{b}$ and $\Lambda(A)$) such that*

$$\lim_{j\to\infty} \eta_{2j+1} = \zeta_1 = \alpha\lambda_1 + (1-\alpha)\lambda_n,$$

$$\lim_{j\to\infty} \eta_{2j} = \zeta_2 = (1-\alpha)\lambda_1 + \alpha\lambda_n.$$



Theorem 5 (Afanasjew et al., 2008) *Under the conditions of Theorem 4*

$$\limsup_{k\to\infty} \|f(A)\boldsymbol{b} - \widehat{\boldsymbol{f}}_k\|^{1/k} \le \frac{\kappa_A}{\kappa_f}, \text{ where}$$

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

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

If $f(\lambda) = \exp(\tau\lambda)$, $\tau \ne 0$, *then*

$$\limsup_{k\to\infty} \left[k \|f(A)\boldsymbol{b} - \boldsymbol{f}_k\|^{1/k} \right] \le \kappa_A\, |\tau|\, e.$$

In each case, there exist vectors $\boldsymbol{b}$ *such that equality holds.*

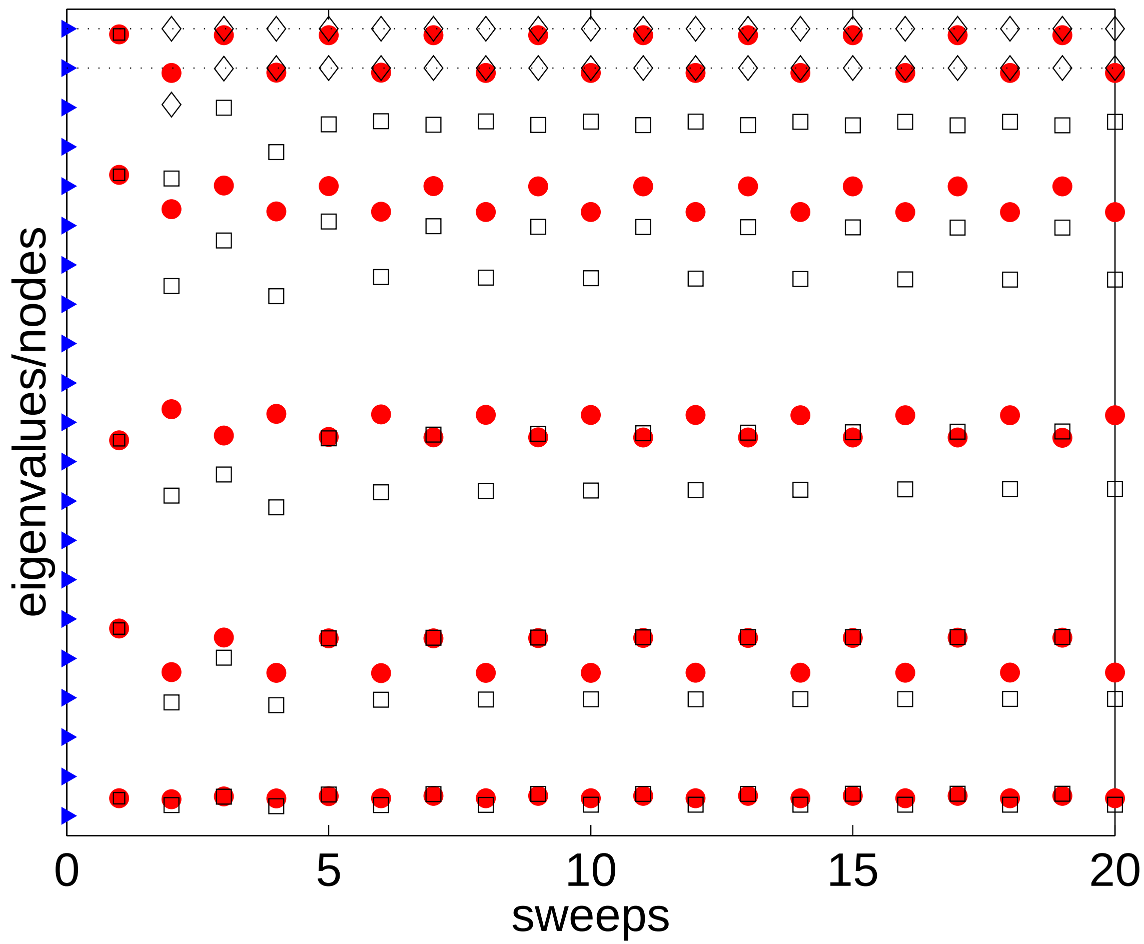
Theorem 6 (E. & Güttel, 2008) *Let A be Hermitian.*
Consider the thick-restarted Arnoldi method with $(m, \ell) = (1, 1)$ with target λ_n.
$\Lambda(\widehat{H}_k) : \eta_1, \eta_2, \ldots, \eta_{k-1}, \eta_k^*$ *and* $\Lambda(\widehat{H}_{k+1}) : \eta_1, \eta_2, \ldots, \eta_{k-1}, \eta_k, \eta_{k+1}^*$.
There exists $\alpha \in (0, 1)$ (which depends on $\boldsymbol{b}$ and $\Lambda(A)$) such that

$$\lim_{j\to\infty} \eta_{2j+1} = \widetilde{\zeta}_1 = \alpha\lambda_1 + (1-\alpha)\lambda_{n-1}$$

$$\lim_{j\to\infty} \eta_{2j} = \widetilde{\zeta}_2 = (1-\alpha)\lambda_1 + \alpha\lambda_{n-1}$$

$$\lim_{j\to\infty} \eta_j^* = \lambda_n.$$

Here, the lemniscates with foci $\widetilde{\zeta}_1$, $\widetilde{\zeta}_2$ determine the convergence behavior.

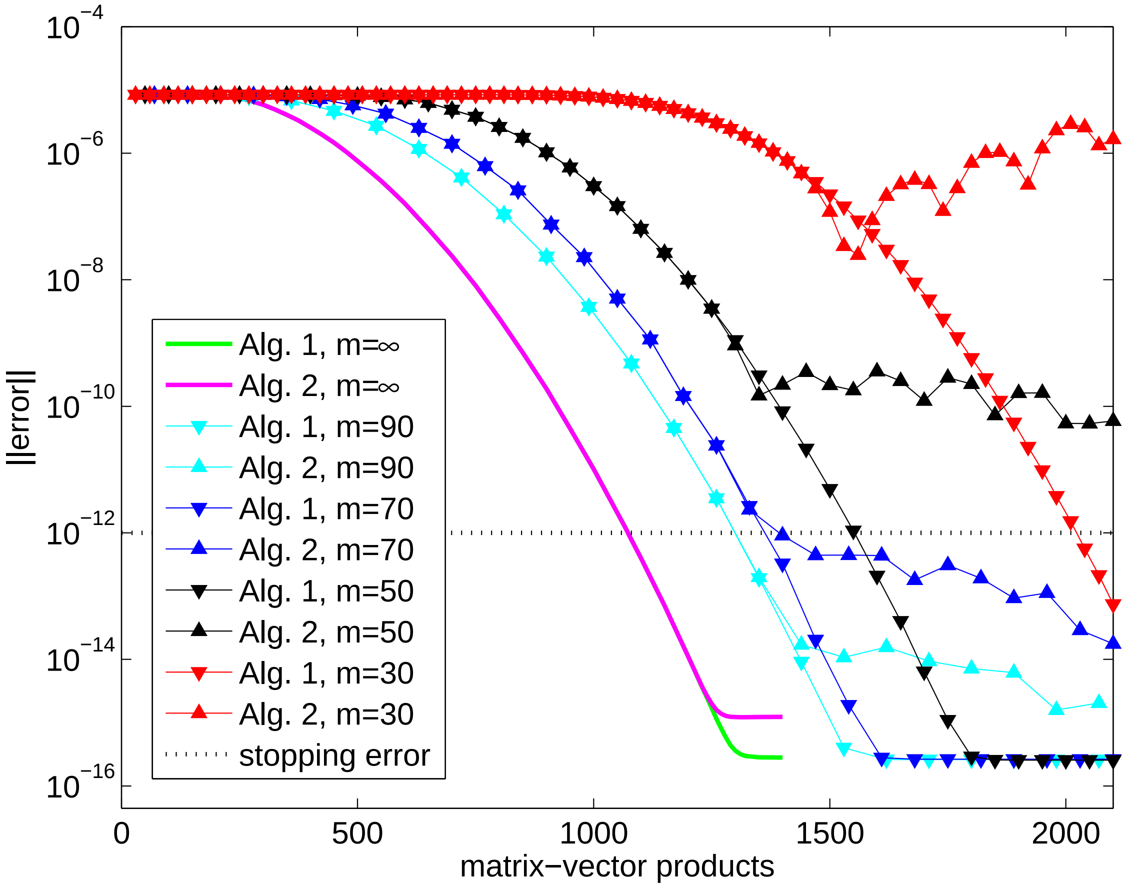


A Hermitian with eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_n$: ▶

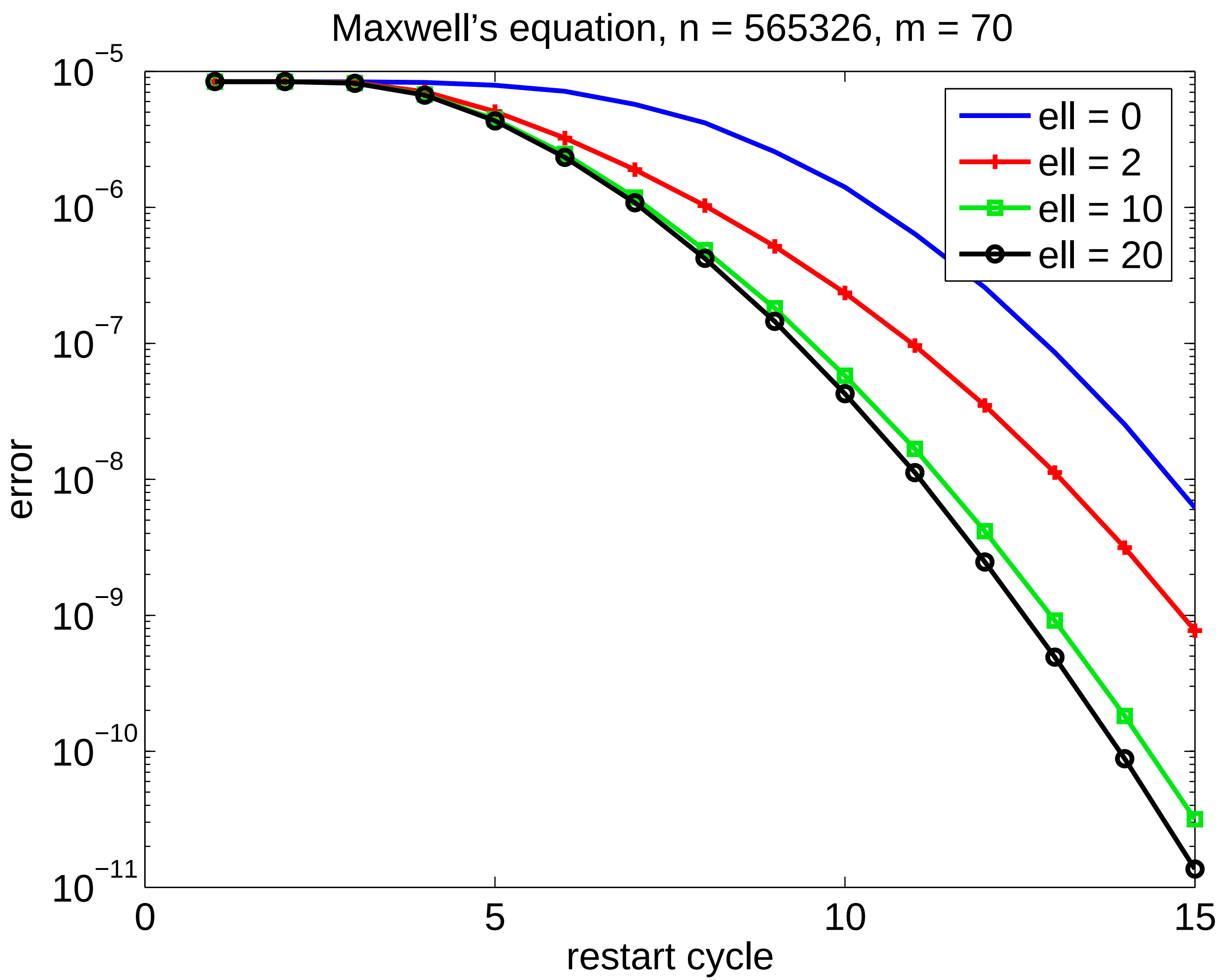
Nodes for standard restarted Arnoldi $(m = 5)$: ●

Nodes for thick-restart Arnoldi $(m = 5, \ell = 2)$: $\square$ + last $\diamondsuit$

5 A numerical example



$\exp(-tA)\boldsymbol{b}$, where $t = 10^{-3}$, A =discrete version of $\sigma^{-1}\nabla\times(\mu^{-1}\nabla\times\dot{\ })$, $\Lambda(A)\in[0,10^8]$, $\dim(A)=565,326$ (see [Afanasjew et al., 2008a])



target = eigenvalues closest to 0

6 Summary

- Restarted Arnoldi methods result in acceptable storage cost even for very large matrices.
- Thick restarts accelerate the convergence.
- There is a stable implementation with constant (low) computational costs per sweep. Necessary: A near best rational approximation to f on $W(A)$ (Faber-Carathéodory-Fejér).
- The asymptotic convergence behavior is (nearly) understood in the Hermitian case.
- Stopping criteria (a posteriori error estimates) are available.