28.05.2002

Finite Precision Krylov Methods: Where are the Frontiers?

Jens–Peter M. Zemke

Table Of Contents

The Menagerie of Krylov Methods

- Lanczos based methods (short–term methods)
- Arnoldi based methods (long–term methods)

- eigensolvers $Av = v\lambda$
- linear system solvers: $Ax = b$
    - (quasi-) orthogonal residual approaches: (Q)OR
    - (quasi-) minimal residual approaches: (Q)MR

Extensions:

- Lanczos based methods:
    - look-ahead
    - product-type (LTPMs)
    - applied to normal equations (CGN)
- Arnoldi based methods:
    - restart (thin/thick, explicit/implicit)
    - truncation (standard/optimal)

In the following: $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$

$A \in \mathbb{K}^{n \times n}$	system matrix (usually large, sparse)
$Q \in \mathbb{K}^{n \times n}$	basis matrix used for Krylov subspace
$\hat{Q} \in \mathbb{K}^{n \times n}$	adjoint basis to Q
$I \in \mathbb{K}^{n \times n}$	identity matrix, columns e_j
$T \in \mathbb{K}^{n \times n}$	tridiagonal matrix
$H \in \mathbb{K}^{n \times n}$	Hessenberg matrix
$C \in \mathbb{K}^{n \times n}$	computed (condensed) matrix

First step: iterative transformation to

- tridiagonal form (Lanczos)

$$\hat{Q}^H A Q = T, \quad \hat{Q}^H Q = I$$

- Hessenberg form (Arnoldi)

$$Q^H A Q = H, \quad Q^H Q = I$$

(as attempt to be close to Jordan/Schur normal form)

A Unified Matrix Description of Krylov Methods

Introduce computed (condensed) matrix $C = T, H$

$$Q^{-1}AQ = C \quad \Rightarrow \quad AQ = QC$$

Iteration implied by unreduced Hessenberg structure:

$$AQ_k = Q_{k+1}\underline{C}_k, \quad Q_k = [q_1, \ldots, q_k], \quad \underline{C}_k \in \mathbb{K}^{(k+1)\times k}$$

Stewart: 'Krylov Decomposition'

Iteration spans Krylov subspace ($q = q_1$):

$$\text{span}\{Q_k\} = \mathcal{K}_k = \text{span}\{q, Aq, \ldots, A^{k-1}q\}$$

Purely algebraic (rational) approach: only polynomials involved

Definition: A *Krylov method* is an *iterative method* that returns *approximations* to the desired quantity from a *nested sequence* of *Krylov subspaces*.

Lemma: *Any* possible Krylov method can be expressed in terms of a Krylov decomposition.

Proof:
The natural basis of a Krylov space is given by the *Krylov matrix*

$$K_k = [q, Aq, \ldots, A^{k-1}q]$$

Any other basis Q_k can be expressed as

$$K_k = Q_k B_k, \quad B_k \in \mathbb{K}^{k\times k}, \quad B_k \text{ regular}$$

When the sequence of bases is such that first columns remain unaltered:

$$B_k = R_k$$

with R_k upper triangular, sequence of submatrices (GR decomposition).

Proof cont'd:

Let $q = q_1$ be the starting vector of the Krylov subspace.

$$\begin{aligned} & [q, AK_k] &=& K_{k+1} \\ \Rightarrow \qquad & [q, AQ_k] &=& Q_{k+1} B_{k+1} \begin{pmatrix} 1 & 0 \\ 0 & B_k^{-1} \end{pmatrix} \end{aligned} \tag{1}$$

Observation:

- GR decomposition whenever basis unaltered
- QR decomposition when method based on Arnoldi

Equation (1) gives the Krylov decomposition

$$AQ_k = Q_{k+1} \underline{C}_k = Q_k C_k + q_{k+1} c_{k+1,k} e_k^T$$

Here $\underline{C}_k$ is defined by

$$\begin{pmatrix} \star & \\ 0 & \underline{C}_k \end{pmatrix} = B_{k+1} \begin{pmatrix} 1 & 0 \\ 0 & B_k^{-1} \end{pmatrix}$$

Remark: Previously computed basis vectors unaltered:

- o Krylov decomposition is *GR decomposition*
- o computed matrix $\underline{C}_k$ is *(unreduced) Hessenberg*
- o set $\{\underline{C}_k\}_k$ is *sequence of nested submatrices*

Characterisation of Krylov methods as

$$\begin{aligned} AQ_k = Q_{k+1}\underline{C}_k &= Q_k C_k + q_{k+1} c_{k+1,k} e_k^T \\ &= Q_k C_k + M_k \end{aligned}$$

has impacts.

Lemma: A representation of the basis vectors q_j is given by

$$\left(\prod_{j=1}^{k} c_{j+1,j} \right) q_{k+1} = \chi_{C_k}(A)\, q.$$

Proof: Some commutative algebra on the 'Sylvester equation' form

$$(I_k \otimes A - C_k^T \otimes I_n)\mathrm{vec}(Q_k) = \mathrm{vec}(M_k).$$

Perturbed Krylov Decompositions

A Krylov decomposition analogue holds true in finite precision:

$$\begin{aligned} AQ_k = Q_{k+1}\underline{C}_k - F_k &= Q_k C_k + q_{k+1} c_{k+1,k} e_k^T - F_k \\ &= Q_k C_k + M_k - F_k \end{aligned}$$

We have to investigate the impacts of the method on

- the structure of the basis Q_k (local orthogonality/duality)
- the structure of the computed C_k, $\underline{C}_k$
- the size/structure of the error term $-F_k$

Convergence theory:

- is usually based on inductively proven properties:
  orthogonality, bi-orthogonality, A-conjugacy, ...

What can be said about these properties?

'Standard' error analysis:

- splits into *forward* and *backward* error analysis.

Does this analysis apply to Krylov methods?

A (sligthly) different introduction to (well-known) Krylov methods:

Eigenproblem solvers:

- compute the Krylov decomposition:

$$AQ_k = Q_k C_k + M_k$$

- solve a *small structured* eigenvalue problem:

$$C_k S_k = S_k J(\Theta_k)$$

- prolong the eigenvectors:

$$Y_k = Q_k S_k$$

- use *Ritz pair* as approximate eigenpair:

$$AY_k - Y_k J(\Theta_k) = M_k S_k = q_{k+1} c_{k+1,k} e_k^T S_k$$

In these methods the Krylov decomposition is used *explicitly*.

Examples of Krylov method eigenproblem solvers are the methods of Lanczos and Arnoldi. We are only interested in the decompositional part.

Arnoldi's method uses an *orthonormal* basis:
CGS-Arnoldi, MGS-Arnoldi, Householder-Arnoldi, Givens-Arnoldi, ...

MGS-Arnoldi proceeds as follows ($\underline{k} \equiv \{1, \ldots, k\}$):

A and r_0 given
for $k \in \mathbb{N}$ **do** – outer loop
  $h_{k,k-1} \leftarrow \|r_{k-1}\|$ – compute last moment
  $q_k \leftarrow r_{k-1}/h_{k,k-1}$ – normalisation
  $r_k \leftarrow Aq_k$ – expand Krylov subspace
  for $j \in \underline{k}$ **do** – inner loop
    $h_{jk} \leftarrow \langle q_j, r_k \rangle$ – compute moments
    $r_k \leftarrow r_k - q_j h_{jk}$ – purge residual vector
  end for
end for

In all variants $C_k = H_k$ is (unreduced) Hessenberg, F_k is small ($\approx \|A\|\epsilon$)

Lanczos' method uses two *bi-orthogonal* bases:
Symmetric Lanczos, Symplectic Lanczos, Day's Variant, ...

Let $\hat{A} \equiv A^H$ and $\hat{T}_k \equiv T_k^H$. Then the bases fulfil

$$\begin{aligned} AQ_k &= Q_{k+1}\underline{T}_k = Q_k T_k + r_{k+1} e_k^T = Q_k T_k + q_{k+1}\beta_k e_k^T \\ \hat{A}\hat{Q}_k &= \hat{Q}_{k+1}\underline{\hat{T}}_k = \hat{Q}_k \hat{T}_k + \hat{r}_{k+1} e_k^T = \hat{Q}_k \hat{T}_k + \hat{q}_{k+1}\overline{\gamma}_k e_k^T \end{aligned}$$

Here

$$T_k = \begin{pmatrix} \alpha_1 & \gamma_1 & & \\ \beta_1 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \gamma_{k-1} \\ & & \beta_{k-1} & \alpha_k \end{pmatrix}.$$

Unique algorithm: any choice of β_k, γ_k such that

$$\beta_k \gamma_k = \langle \hat{r}_k, r_k \rangle$$

Unified pseudo-code for several Lanczos-algorithms:

A, r_0 and $\hat{r}_0$ given
for $k \in \mathbb{N}$ **do** – outer loop
- $\beta_{k-1}\gamma_{k-1} \leftarrow \langle \hat{r}_{k-1}, r_{k-1} \rangle$ – compute last moments
- $q_k \leftarrow r_{k-1}/\beta_{k-1}$ – *right* normalisation
- $\hat{q}_k \leftarrow \hat{r}_{k-1}/\overline{\gamma}_{k-1}$ – *left* normalisation
- $r_k \leftarrow Aq_k$ – expand *right* Krylov subspace
- $\hat{r}_k \leftarrow \hat{A}\hat{q}_k$ – expand *left* Krylov subspace
- $\alpha_k \leftarrow \langle \hat{q}_k, r_k \rangle = \langle \hat{r}_k, q_k \rangle$ – compute middle moment
- $r_k \leftarrow r_k - \alpha_k q_k - \gamma_{k-1} q_{k-1}$ – purge *right* residual vector
- $\hat{r}_k \leftarrow \hat{r}_k - \overline{\alpha}_k \hat{q}_k - \overline{\beta}_{k-1} \hat{q}_{k-1}$ – purge *left* residual vector

end for

Remark: $A = A^H$, $r_0 = \hat{r}_0$ and $\gamma_k = \overline{\beta}_k \quad \Rightarrow \quad$ Lanczos = Arnoldi (CG)

In most variants $C_k = T_k$ is tridiagonal, F_k, $\hat{F}_k$ are small compared to A and the length of the columns of Q_k, $\hat{Q}_k$

The Krylov method linear system solvers can be distinguished into OR and MR methods. OR methods are more close to the eigenproblem methods.

(Q)OR linear system solvers, direct approach:

- compute the Krylov decomposition:

$$AQ_k = Q_k C_k + M_k, \quad q_1 = b/\|b\|$$

- solve a *small structured* linear system:

$$C_k z_k = \|b\| e_1$$

- prolong the solution:

$$x_k = Q_k z_k$$

- use this as approximate solution:

$$Ax_k - Q_k\|b\|e_1 = Ax_k - b = M_k z_k = q_{k+1} c_{k+1,k} e_k^T z_k$$

In these methods the Krylov decomposition is used *explicitly*.

The MR methods use the larger (non-square) matrix $\underline{C}_k$.

(Q)MR linear system solvers, direct approach:

- compute the Krylov decomposition:

$$AQ_k = Q_{k+1}\underline{C}_k, \quad q_1 = b/\|b\|$$

- solve a *small structured* minimal residual linear system:

$$z_k = \arg\min_z \|\underline{C}_k z - \|b\|e_1\|$$

- prolong the solution:

$$x_k = Q_k z_k$$

- use this as approximate solution:

$$\|Ax_k - b\| = \|Q_{k+1}(\underline{C}_k z_k - \|b\|e_1)\|$$

In these methods the Krylov decomposition is used *explicitly*.

When the basis Q_{k+1} is *orthonormal*, the computed solution is the *minimal residual solution* in the Krylov subspace.

The solution of the small system is often based on decompositions.

Examples include:

(Q)OR: FOM, SymmLQ, SymmBK, QOR

(Q)MR: GMRES, MinRes, QMR

Krylov decompositions are quite similar to Richardson iteration:

$$x_{k+1} = (I - A)x_k + r_0 \quad \Leftrightarrow \quad Ax_k - r_0 = x_k - x_{k+1}$$
$$AX_k - R_0 = X_{k+1}\underline{B}_k \quad \Leftrightarrow \quad -R_k = X_{k+1}\underline{B}_k$$

The column sums of $\underline{B}_k$ are zero. Similarly, for the Chebychev polynomial acceleration on the interval $(-1, 1)$, we obtain the recurrence

$$-R_k D_k = X_{k+1}\underline{T}_k,$$

where $\underline{T}_k$ is a tridiagonal matrix with zero column sums.

Inspired by Richardson, Chebychev, or more general, polynomial acceleration, we seek $\underline{C}_k$ with zero column sums. This will enable us to discard the step of the solution of the small linear system.

Let a generic Krylov decomposition ($c_{m+1,m} = 0$)

$$AQ_m = Q_m C_m$$

be given. Suppose C_m non-singular unreduced Hessenberg. The system $y^T C_m = e_m^T$ has a unique solution y. Of course $y^T \underline{C}_{m-1} = 0$.

When $y(k) = 0$, C_{k-1} is singular, since then $\underline{C}_{m-1}$ with kth row deleted must be singular and is given by

$$\begin{pmatrix} C_{k-1} & \star \\ 0 & R \end{pmatrix}$$

Remark: When C_k is singular and Q_k is orthonormal, zero is in the field of values of A,

$$0 = z^H C_k z = z^H Q_k^H A Q_k z = y^H A y.$$

When all $y(k)$ are non-zero, we can scale C_m by $D = \text{diag}(y)$:

$$C_m^{(0)} = DC_mD^{-1}.$$

Suppose further that $q_1 = b$. Then the (Q)OR solution is given by

$$C_k^{(0)} z_k = e_1.$$

Observation: The scaling implies that

$$e^T \underline{C}_{m-1}^{(0)} = 0, \quad \Leftrightarrow \quad e^T C_k^{(0)} = -c_{k+1,k}^{(0)} e_k^T \tag{2}$$

holds true, with $e = (1, \ldots, 1)^T$ of appropriate length.

This, in turn, implies that the residuals satisfy

$$-r_k = Ax_k - b = q_{k+1} c_{k+1,k}^{(0)} e_k^T z_k = -q_{k+1} e^T C_k^{(0)} (C_k^{(0)})^{-1} e_1 = -q_{k+1}$$

We re-write the decomposition as

$$AR_k = R_{k+1}\underline{C}_k^{(0)} \tag{3}$$

When we apply A^{-1} to this equation, we obtain by equation (2)

$$\begin{aligned} R_k &= A^{-1}R_{k+1}\underline{C}_k^{(0)} \\ &= [x - x_0, \ldots, x - x_k]\underline{C}_k^{(0)} \\ &= (xe^T - X_{k+1})\underline{C}_k^{(0)} \\ R_k &= -X_{k+1}\underline{C}_k^{(0)} \end{aligned} \tag{4}$$

Equations (3) and (4) define the class of methods known as **Orthores**.

Examples include:

(Q)OR: Orthores (Arnoldi), CG-Ores (Lanczos), Biores (Lanczos)

(Q)MR: CR-Ores (Lanczos), QMR

Many methods derived thus far can be handled using the following lemma:

Lemma: The error in methods that are based on the *direct computation* of the Krylov decomposition fulfils

$$|F_k| \leq \gamma_n |A||Q_k| + \gamma_{k+1}|Q_{k+1}||\underline{C}_k|$$

Here γ_n is given by $n\epsilon/(1-n\epsilon)$ where ϵ denotes the machine precision, in IEEE arithmetic double precision given by $\epsilon = 2^{-53} \approx 1.11 \cdot 10^{-16}$.

This class of methods includes Orthores methods, CGS-Arnoldi and MGS-Arnoldi based methods (FOM, GMRES) and many Lanczos variants.

This *a posteriori* result is similar to the well-known result on LR decomposition (Higham 1996, Accuracy and Stability of Numerical Algorithms).

The proof, like the proof for the LR decomposition, is based on Lemma 8.4 in Higham's textbook.

Orthores yet another way: equation (2) re-written

$$\begin{pmatrix} 1 & & & 0 \\ 1 & 1 & & \\ \vdots & \vdots & \ddots & \\ 1 & 1 & \cdots & 1 \end{pmatrix} C_k^{(0)} = D_k M_k^H$$

where M_k^H upper triangular, unit diagonal and

$$D_k = -\text{diag}(c_{2,1}^{(0)}, \ldots, c_{k+1,k}^{(0)})$$

Observe

$$\begin{pmatrix} 1 & & & 0 \\ 1 & 1 & & \\ \vdots & \vdots & \ddots & \\ 1 & 1 & \cdots & 1 \end{pmatrix} = L_k^{-1} = \begin{pmatrix} 1 & & & 0 \\ -1 & \ddots & & \\ & \ddots & 1 & \\ 0 & & -1 & 1 \end{pmatrix}^{-1}.$$

Orthores computes (implicitly) an LDMT decomposition

$$C_k^{(0)} = L_k D_k M_k^H$$

We insert this decomposition into the Krylov decomposition and re-write it:

$$\begin{aligned} AR_m &= R_m L_m D_m M_m^H \\ AR_m M_m^{-H} D_m^{-1} &= R_m L_m \end{aligned}$$

The columns of the basis

$$P_m = R_m M_m^{-H} \qquad \Rightarrow \qquad R_k = P_k M_k^H \tag{5}$$

are termed *direction vectors*.

Equation (5) together with the equations

$$AP_k D_k^{-1} = R_{k+1}\underline{L}_k = R_k L_k - r_k e_k^T \tag{6}$$

$$P_k D_k^{-1} = -X_{k+1}\underline{L}_k = -X_k L_k + x_k e_k^T \tag{7}$$

forms the class of methods known as **Orthomin**.

Orthomin includes:

(Q)OR: CG-Omin, Biomin

(Q)MR: Orthomin, CR-Omin, QMR

Orthomin methods (this includes the usual variant of CG) come in form of two coupled recurrences. We refer to this computation of

$$AR_k = R_{k+1}\underline{C}_k^{(0)} - F_k$$

as a **split Krylov decomposition**.

Lemma: In **split Krylov decompositions** the error term fulfils:

$$-F_k = AF_k^{(1)} + F_k^{(2)} L_k^{-1} C_k^{(0)}.$$

The error terms come from the coupled recurrences,

$$AP_k D_k^{-1} = R_{k+1}\underline{L}_k + F_k^{(1)}, \qquad R_k = P_k M_k^H + F_k^{(2)}.$$

We mention **Orthodir**, a class of methods based on a different scaling.

A large class of methods is based on a transformation of a Lanczos variant to not use the transpose A^T or Hermitian A^H. This class is known as Lanczos-type product methods, LTPMs and results (implicitly) in Krylov decompositions.

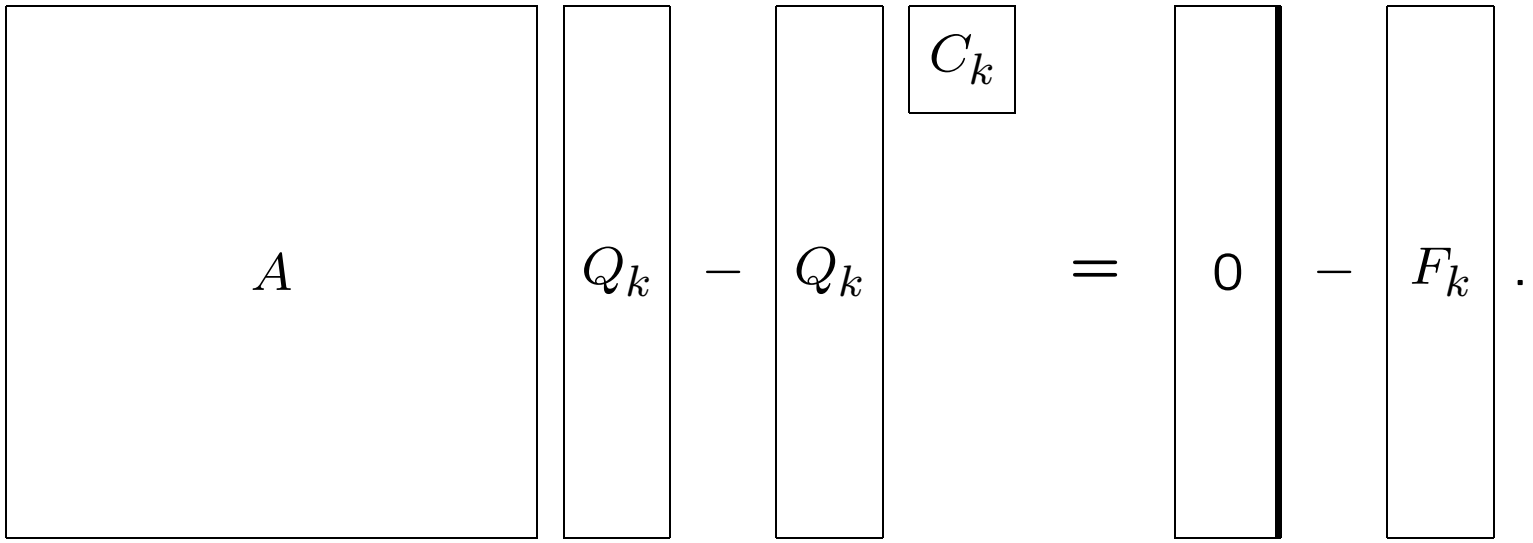
We just remark:

- o C_k is Hessenberg and depends on $O(k)$ values
- o not every basis vector must be a (quasi-) residual
- o Orthores, Orthomin and Orthodir variants exist
- o F_k depends on complicated expressions

Examples include:

CGS, CGS2, shifted CGS,
BiCGSTAB, BICG$\times$MR2, BiCGSTAB2, BiCGstab(ℓ),
TFQMR, QMRCGSTAB, ...

All methods fit pictorially into:



This is a perturbed Krylov decomposition, as subspace equation.

Examination of the methods has to be done according to

- methods directly based on the Krylov decomposition
- methods based on a split Krylov decomposition
- LTPMs

The matrix C_k plays a crucial role:

- C_k is Hessenberg or even tridiagonal (basics),
- C_k may be blocked or banded (block Krylov methods),
- C_k may have humps, spikes, ... (more sophisticated)

The error analysis and convergence theory splits further up:

- knowledge on Hessenberg (tridiagonal) matrices
- knowledge on orthogonality, duality, conjugacy, ...

We start with results on Hessenberg matrices.

A Short Excursion on Matrix Structure

$J_\wedge$	Jordan matrix of A
V	right eigenvector-matrix, $AV = VJ_\wedge$
$\widehat{V} \equiv V^{-H}$	left eigenvector-matrix, $\widehat{V}^H A = J_\wedge \widehat{V}^H$
$\check{V} \equiv V^{-T}$	alternate left eigenvector-matrix, $\check{A}^T A = J_\wedge \check{V}^T$
$\chi_A(\lambda) \equiv \det(\lambda I - A)$	characteristic polynomial of A
$R(\lambda) \equiv (\lambda I - A)^{-1}$	resolvent
$\mathcal{C}_k(A)$	kth compound matrix of A
$\mathrm{adj}\,(A)$	classical adjoint, adjugate of A
A_{ij}	A with row i and column j deleted
S, S_i	sign matrices

The adjoint of $\lambda I - A$ fulfils

$$\mathrm{adj}\,(\lambda I - A)(\lambda I - A) = \det(\lambda I - A) I.$$

Suppose that λ is not contained in the spectrum of A.

We form the resolvent of λ and obtain

$$\begin{aligned} \operatorname{adj}(\lambda I - A) &= \det(\lambda I - A)\, R(\lambda) \\ &= V\left(\chi_A(\lambda)\, J_{\lambda-\Lambda}^{-1}\right) \widehat{V}^H. \end{aligned}$$

The shifted and inverted Jordan matrix looks like

$$J_{\lambda-\lambda_i}^{-1} = S_i E_i S_i \equiv S_i \begin{pmatrix} (\lambda-\lambda_i)^{-1} & (\lambda-\lambda_i)^{-2} & \dots & (\lambda-\lambda_i)^{-k} \\ & (\lambda-\lambda_i)^{-1} & & \\ & & \ddots & \vdots \\ & & & (\lambda-\lambda_i)^{-1} \end{pmatrix} S_i,$$

The multiplication with the characteristic polynomial allows to cancel the terms with negative exponent.

The resulting expression is a source of eigenvalue – eigenvector relations.

We express the adjugate with the aid of compound matrices,

$$\operatorname{adj} A \equiv S\mathcal{C}_{n-1}(A^T)S.$$

Then we have equality

$$\begin{aligned} P \equiv \mathcal{C}_{n-1}(\lambda I - A^T) &= (SVS)\, G\, (S\widehat{V}^H S) \\ &\equiv (SVS)\, \chi_A(\lambda)\, E\, (S\widehat{V}^H S). \end{aligned}$$

The elements of the compound matrix P are *polynomials* in λ of the form

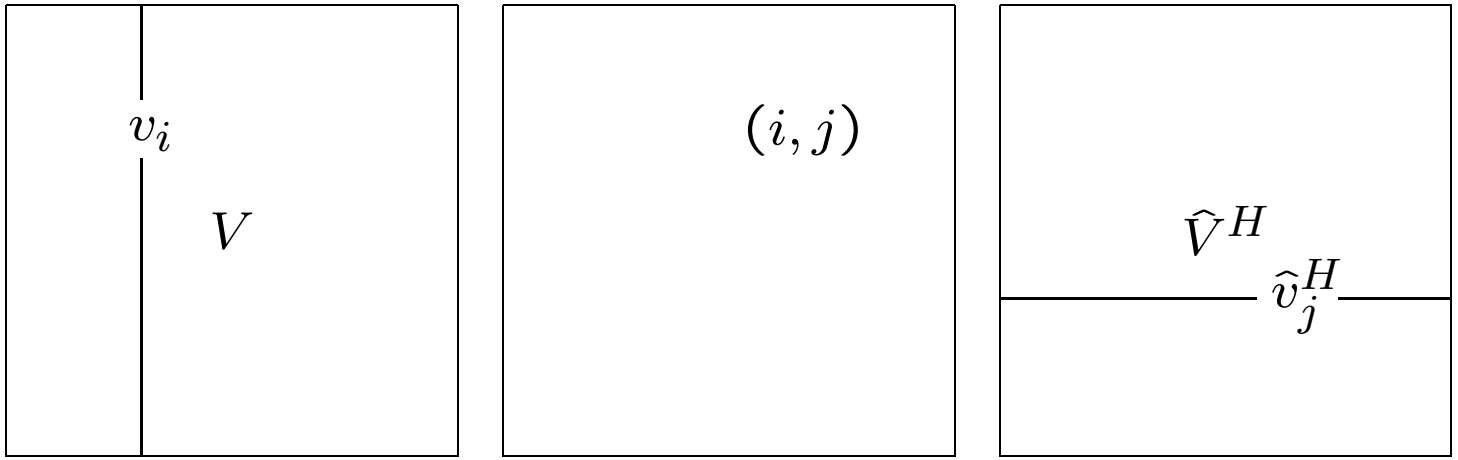
$$p_{ij} = p_{ij}(\lambda; A) \equiv \det L_{ji}, \qquad \text{where} \qquad L \equiv \lambda I - A.$$

The elements of G are obviously given by *rational functions* in λ, since

$$G = \chi_A(\lambda) \cdot (\oplus_i E_i).$$

Many terms cancel, the elements of G are *polynomials*. We divide by the maximal factor and compute the limes $\lambda \to \lambda_i$.

The choice of eigenvectors is based on the non-zero positions i, j in the matrix (the sign matrices are left out):



Amongst others, the well-known result on eigenvalue – eigenvector relations by Thompson and McEnteggert is included. This is one of the basic results used in Paige's analysis of the finite precision symmetric Lanczos method.

We consider here only the special case of non-derogatory eigenvalues.

Theorem: Let $A \in \mathbb{K}^{n \times n}$. Let $\lambda_l = \lambda_{l+1} = \ldots = \lambda_{l+k}$ be a geometrically simple eigenvalue of A. Let $k+1$ be the algebraic multiplicity of λ. Let $\widehat{v}_l^H$ and v_{l+k} be the corresponding left and right eigenvectors with appropriate normalization.

Then

$$v_{jl}\breve{v}_{i,l+k} = (-1)^{(j+i+k)} \frac{p_{ji}(\lambda_l; A)}{\prod_{\lambda_s \neq \lambda_l} (\lambda_l - \lambda_s)}$$

holds true.

The minus one stems from the sign matrices, the polynomial from the definition of the adjoint as matrix of cofactors and the denominator by division with the maximal factor.

This setting matches every eigenvalue of non-derogatory A.

Unreduced Hessenberg matrices are non-derogatory matrices. This is easily seen by a simple rank argument. In the following let $H = H_m$ be unreduced Hessenberg of size $m \times m$,

$$\text{rank}(H - \theta I) \geq m - 1.$$

Many polynomials can be evaluated in case of Hessenberg matrices:

Theorem: The polynomial p_{ji}, $i \leq j$ has degree $(i-1)+(m-j)$ and can be evaluated as follows:

$$\begin{aligned} p_{ji}(\theta; H) &= \begin{vmatrix} \theta I - H_{1:i-1} & & \star \\ & R_{i+1:j-1} & \\ 0 & & \theta I - H_{j+1:m} \end{vmatrix} \\ &= (-1)^{i+j} \chi_{H_{1:i-1}}(\theta) \prod \text{diag}(H_{i:j}, -1) \chi_{H_{j+1:m}}(\theta). \end{aligned}$$

Denote by $\mathcal{H}(m)$ the set of unreduced Hessenberg matrices of size $m \times m$. The general result on eigenvalue – eigenvector relations can be simplified to read:

Theorem: Let $H \in \mathcal{H}(m)$. Let $i \leq j$. Let θ be an eigenvalue of H with multiplicity $k+1$. Let s be the unique left eigenvector and $\hat{s}^H$ be the unique right eigenvector to eigenvalue θ.

Then

$$(-1)^k \, \breve{s}(i) s(j) = \left[\frac{\chi_{H_{1:i-1}} \chi_{H_{j+1:m}}}{\chi_{H_{1:m}}^{(k+1)}}(\theta) \right] \prod_{l=i}^{j-1} h_{l+1,l} \tag{8}$$

holds true.

Remark: We ignored the implicit scaling in the eigenvectors imposed by the choice of eigenvector-matrices, i.e. by $\breve{S}^T S = I$.

Among these relations of special interest is the case of index pairs (i,m), $(1,m)$ and $(1,m)$, $(1,j)$:

$$
\begin{aligned}
(-1)^k\ \breve{s}(i)s(m) &= \left[\frac{\chi_{H_{1:i-1}}}{\chi^{(k+1)}_{H_{1:m}}}(\theta)\right]\prod_{l=i}^{m-1} h_{l+1,l},\\
(-1)^k\,\breve{s}(1)s(m) &= \left[\frac{1}{\chi^{(k+1)}_{H_{1:m}}}(\theta)\right]\prod_{l=1}^{m-1} h_{l+1,l},\\
(-1)^k\ \ \breve{s}(1)\,s(j) &= \left[\frac{\chi_{H_{j+1:m}}}{\chi^{(k+1)}_{H_{1:m}}}(\theta)\right]\prod_{l=1}^{j-1} h_{l+1,l}.
\end{aligned}
$$

These relations are used to derive relations between eigenvalues and *one* eigenvector.

They are also of interest for the understanding of the convergence of Krylov methods, at least in context of Krylov eigensolvers.

Theorem: Let $H \in \mathcal{H}(m)$. Let θ be an eigenvalue of H. Then $\widehat{s} = \overline{\breve{s}}$ defined by non-zero $\breve{s}(1)$ and the relations

$$\frac{\breve{s}(i)}{\breve{s}(1)} = \frac{\chi_{H_{i-1}}(\theta)}{\prod_{l=1}^{i-1} h_{l+1,l}} \quad \forall\ i \in \underline{m},$$

is (up to scaling) the unique left eigenvector of H to eigenvalue θ.

Theorem: Let $H \in \mathcal{H}(m)$. Let θ be an eigenvalue of H. Then s defined by non-zero $s(m)$ and the relations

$$\frac{s(j)}{s(m)} = \frac{\chi_{H_{j+1:m}}(\theta)}{\prod_{l=j+1}^{m} h_{l,l-1}} \quad \forall\ j \in \underline{m},$$

is (up to scaling) the unique right eigenvector of H to eigenvalue θ.

Since the polynomials remain unchanged, merely the eigenvalue moves, this helps to explain convergence behaviour (even in finite precision).

The derivation of the theorems proves that the last component of s and the first component of $\widehat{s}^H$ are both non-zero.

Alternate (direct) proof by contradiction: s right eigenvector with last component zero,

$$Hs = s\theta, \quad e_m \perp s.$$

Last row of H orthogonal to s. H is unreduced upper Hessenberg. This implies $e_{m-1} \perp s$. By induction all components are zero. The proof for the left eigenvector is analogous.

These results are of interest in the *understanding of the convergence*.

Next we focus (shortly) on results important for backward error analysis.

We can prove the following *residual bounds*:

Theorem: Let $H \in \mathcal{H}(m)$. Split $H = H_{1:m}$ into

$$H_{1:m} = \begin{pmatrix} H_{1:k} & \star \\ h_{k+1,k} e_1 e_k^T & H_{k+1:m} \end{pmatrix} \equiv \begin{pmatrix} H_{1:k} & \star \\ M & H_{k+1:m} \end{pmatrix}.$$

Consider the *prolonged right eigenvectors of the leading part* $H_{1:k}$ as approximate right eigenvectors of H. The residual is given by

$$\begin{pmatrix} H_{1:k} & \star \\ M & H_{k+1:m} \end{pmatrix} \begin{pmatrix} S_{1:k} \\ 0 \end{pmatrix} - \begin{pmatrix} S_{1:k} \\ 0 \end{pmatrix} J_{1:k} = h_{k+1,k} e_{k+1} e_k^T S_{1:k}.$$

The *prolonged left eigenvectors of the trailing part* $H_{k+1:m}$ have the residual

$$\begin{pmatrix} 0 \\ \widehat{S}_{k+1:m} \end{pmatrix}^H \begin{pmatrix} H_{1:k} & \star \\ M & H_{k+1:m} \end{pmatrix} - J_{k+1:m} \begin{pmatrix} 0 \\ \widehat{S}_{k+1:m} \end{pmatrix}^H = h_{k+1,k} \widehat{S}_{k+1:m}^H e_1 e_k^T.$$

Error Analysis Revisited

Error analysis often is based on loss of orthogonality (bi-orthogonality). We introduce the matrix $W_k = \hat{Q}_k^H Q_k$.

Theorem: In a long-term recurrence the loss of orthogonality fulfils

$$W_{k+1}\underline{C}_k = \hat{Q}_{k+1}^H A Q_k + \hat{Q}_{k+1}^H F_k.$$

In two coupled short-term recurrences the loss of orthogonality additionally fulfils

$$\underline{\hat{C}}_k^H W_{k+1} = \hat{Q}_k^H A Q_{k+1} + \hat{F}_k^H Q_{k+1}.$$

This implies the fundamental relation

$$\underline{\hat{C}}_k^H W_{k+1,k} - W_{k,k+1}\underline{C}_k = \hat{F}_k^H Q_k - \hat{Q}_k^H F_k.$$

We transform these equations to a form, such that we can see the error sources more clearly.

Theorem: In case of a long-term recurrence the relation

$$(W_{k+1} - I_{k+1})\underline{C}_k = \left(\hat{Q}_{k+1}^H A Q_k - \underline{C}_k\right) + \hat{Q}_{k+1}^H F_k$$

holds true.

In case of two coupled short-term recurrences the relation

$$\underline{\hat{C}}_k^H (W_{k+1,k} - I_{k+1,k}) - (W_{k,k+1} - I_{k,k+1})\underline{C}_k = C_k - \hat{C}_k^H + \hat{F}_k^H Q_k - \hat{Q}_k^H F_k$$

holds true.

In infinite precision $W_k = I_k$, $\hat{Q}_k^H A Q_k = C_k$, $\hat{C}_k^H = C_k$ and $F_k = 0$.

The Hessenberg structure enables an iteration of the loss of orthogonality similar to the basis vector iteration.

We re-order the matrix equations of the last slide. The newest quantity is brought to the left.

Theorem: The matrix expression of the loss of bi-orthogonality is given by

$$\begin{aligned}\hat{Q}_k^H M_k &= \hat{Q}_k^H A Q_k - \hat{Q}_k^H Q_k C_k + \hat{Q}_k^H F_k \\ &= (\hat{Q}_k^H A Q_k - C_k) - (W_k - I_k) C_k + \hat{Q}_k^H F_k.\end{aligned}$$

In two-sided methods the loss of bi-orthogonality fulfils

$$\begin{aligned}\hat{Q}_k^H M_k - \hat{M}_k^H Q_k &= \left(\hat{C}_k^H - C_k\right) \\ &+ \hat{C}_k^H (W_k - I_k) - (W_k - I_k) C_k \\ &- \left(\hat{Q}_k^H F_k - \hat{F}_k^H Q_k\right).\end{aligned}$$

We state this as a recurrence on *vectors* as a corollary.

Corollary: The loss of bi-orthogonality is governed by the vector recurrence

$$\begin{aligned}\hat{Q}_k^H q_{k+1} &= \left(\hat{Q}_k^H A q_k - \hat{Q}_k^H Q_k c_k + \hat{Q}_k^H f_k\right) c_{k+1,k}^{-1} \\ &= \left((\hat{Q}_k^H A q_k - c_k) - (W_k - I_k)c_k + \hat{Q}_k^H f_k\right) c_{k+1,k}^{-1}.\end{aligned}$$

This is a recurrence on the *columns* of the matrix $W_m - I_m$.

In two-sided methods the loss of bi-orthogonality fulfils the vector recurrence

$$\hat{Q}_k^H q_{k+1} = \left(\hat{C}_k^H \hat{Q}_k^H q_k - \hat{Q}_k^H Q_k c_k + \hat{Q}_k^H f_k - \hat{F}_k^H q_k + \hat{M}_k^H q_k\right) c_{k+1,k}^{-1}.$$

This is a recurrence on the *columns* of the matrix $W_m - I_m$. Analogously we obtain a recurrence on the *rows* of $W_m - I_m$.

There are two well-known ways to proceed:

- o **additive splitting** of W_k
- o **multiplicative splitting** of W_k

We use the equation

$$\begin{aligned} (W_{k+1} - I_{k+1})\underline{C}_k &= \hat{Q}_{k+1}^H A Q_k - \underline{C}_k + \hat{Q}_k^H F_k \qquad (9) \\ (\underline{C}_k^T \otimes I_{k+1})\mathrm{vec}(W_{k+1} - I_{k+1}) &= \mathrm{vec}(\hat{Q}_{k+1}^H A Q_k - \underline{C}_k) + \mathrm{vec}(\hat{Q}_k^H F_k). \end{aligned}$$

When $\hat{Q}_k = Q_k$ (Arnoldi, symmetric Lanczos):

- o control on the accuracy of computed moments $c_{ij} \neq 0$, $i \leq j$
- o control on the accuracy of the normalisation

Additive split on $\mathrm{vec}(W_{k+1} - I_{k+1})$ $\Rightarrow$ measure *local orthogonality*.

As example consider Arnoldi:

number of unknowns:	$1 + 2 + \cdots + k + (k+1)$
'small' equations from (9):	$1 + 2 + \cdots + k$
equations from normalisation:	$k + 1$

For short-term methods (Lanczos) we use the additive splitting of

$$W_k = L_k + D_k + R_k$$

into diagonal and strictly lower and upper part.

This splitting is inserted into

$$\begin{aligned} \widehat{Q}_k^H M_k - \widehat{M}_k^H Q_k &= \left(\widehat{C}_k^H - C_k\right) \\ &+ \widehat{C}_k^H (W_k - I_k) - (W_k - I_k) C_k \\ &- \left(\widehat{Q}_k^H F_k - \widehat{F}_k^H Q_k\right), \end{aligned}$$

to obtain expressions on the loss of (bi-) orthogonality, based on the computed Ritz pairs. Here $\widehat{C}_k^H = C_k$, i.e. $C_k = T_k$ is tridiagonal.

Local orthogonality is important for this type of analysis.

The analysis was first carried out by Paige for the symmetric Lanczos method and by Bai for the non-symmetric Lanczos method.

The multiplicative splitting approach does not distinguish between long-term and short-term. The drawback is the strong assumption on W_k:

Suppose that W_k can be triangular decomposed, $W_k = \hat{R}_k^H R_k$. Then

$$\begin{aligned} \hat{Q}_k^H A Q_k - W_k C_k &= \hat{Q}_k^H M_k - \hat{Q}_k^H F_k \\ \hat{R}_k^{-H} \hat{Q}_k^H A Q_k R_k^{-1} - R_k C_k R_k^{-1} &= \hat{R}_k^{-H} \hat{Q}_k^H M_k R_k^{-1} - \hat{R}_k^{-H} \hat{Q}_k^H F_k R_k^{-1} \end{aligned}$$

Define $C_k^{\text{sim}} \equiv R_k C_k R_k^{-1}$.

Remark: C_k^{sim} in all cases is Hessenberg, even for tridiagonal C_k.

Define the *exact oblique projection*

$$(C_k^{\text{exact}}, I_k) \equiv (\hat{P}_k^H A P_k, \hat{P}_k^H P_k) \equiv (\hat{R}_k^{-H} \hat{Q}_k^H A Q_k R_k^{-1}, \hat{R}_k^{-H} W_k R_k^{-1}).$$

This proves that C_k^{sim} is a perturbation of C_k^{exact}.

Theorem: The matrix C_k is similar to C_k^{sim}, which is an additive perturbation of C_k^{exact}, an exact oblique projection of A:

$$\begin{aligned} C_k^{\text{exact}} - C_k^{\text{sim}} &= c_{k+1,k}\hat{R}_k^{-H}\hat{Q}_k^H q_{k+1} e_k^T R_k^{-1} - \hat{R}_k^{-H}\hat{Q}_k^H F_k R_k^{-1} \\ &= \frac{c_{k+1,k}}{r_{kk}} \begin{pmatrix} r_{1,k+1} \\ \vdots \\ r_{k,k+1} \end{pmatrix} e_k^T - \hat{P}_k^H F_k R_k^{-1}. \end{aligned}$$

The deviation can be bounded normwise by

$$\|C_k^{\text{exact}} - C_k^{\text{sim}}\|_2 \leq \|\hat{R}_k^{-1}\|_2 \|R_k^{-1}\|_2 (\|c_{k+1,k}\hat{Q}_k^H q_{k+1}\|_2 - \|\hat{Q}_k^H F_k\|_2)$$

To obtain useful bounds we have to measure the *growth factor* of the LR decomposition and the vector recurrence of the loss of orthogonality.

This type of analysis was carried out by Simon (symmetric Lanczos) and Day (non-symmetric Lanczos). The analysis results in

semi-orthogonalisation and *semi-duality* techniques.

The error analysis applies to all methods. No real backward results are contained. The only backward result is the backward error analysis by Greenbaum. Nevertheless, in her analysis is a substantiable gap between *proven* and *observed* behaviour.

The re-orthogonalisation techniques are not easily adoptable to linear system solvers.

The error analysis thus far was only considered with W_k, the matrix of the loss of orthogonality. Part of the analysis is based on eigenvalues and eigenvectors.

The (Q)MR methods would better be analysed in terms of the SVD. We mention the analysis of GMRES by Rozložník.

It is possible to analyse the recurrence of the basis vectors q_j instead of the loss of orthogonality.

For simplicity we assume that the perturbed Krylov decomposition

$$M_k = AQ_k - Q_k C_k + F_k$$

is diagonalisable, i.e. that A and C_k are diagonalisable.

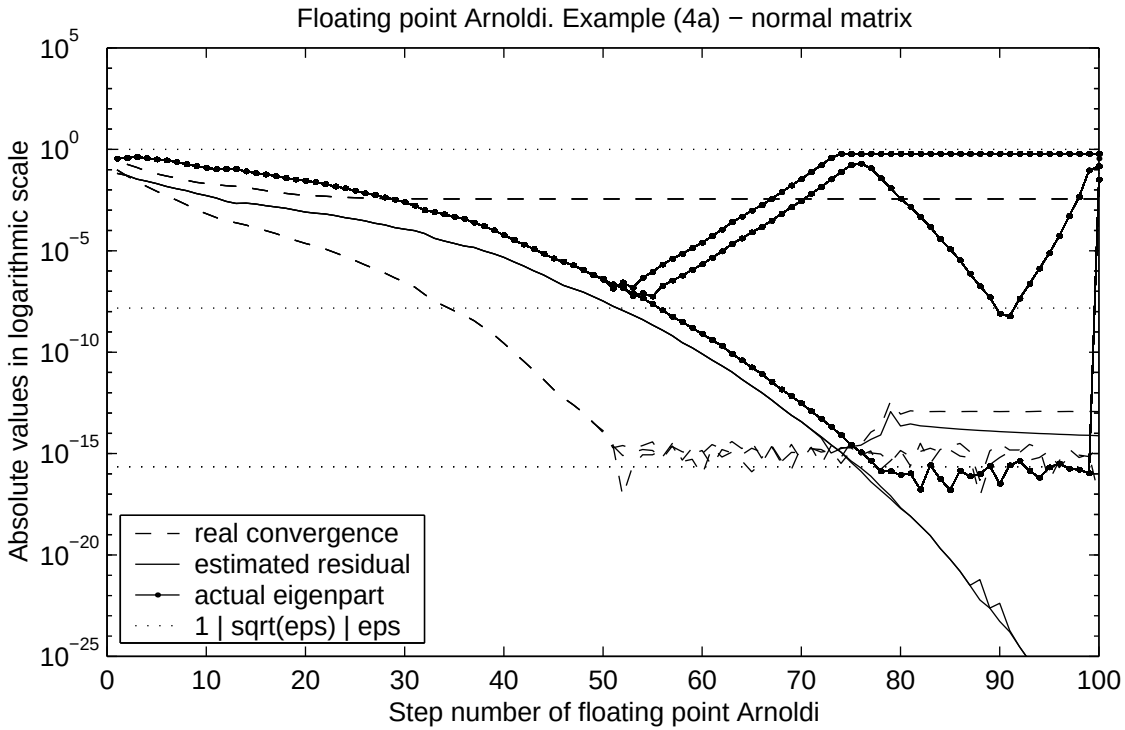
Theorem: The recurrence of the basis vectors in eigenparts is given by

$$\widehat{v}_i^H q_{k+1} = \frac{\left(\lambda_i - \theta_j\right) \widehat{v}_i^H y_j + \widehat{v}_i^H F_k s_j}{c_{k+1,k} s_{kj}} \quad \forall\ i, j(, k).$$

This *local error amplification formula* consists of:

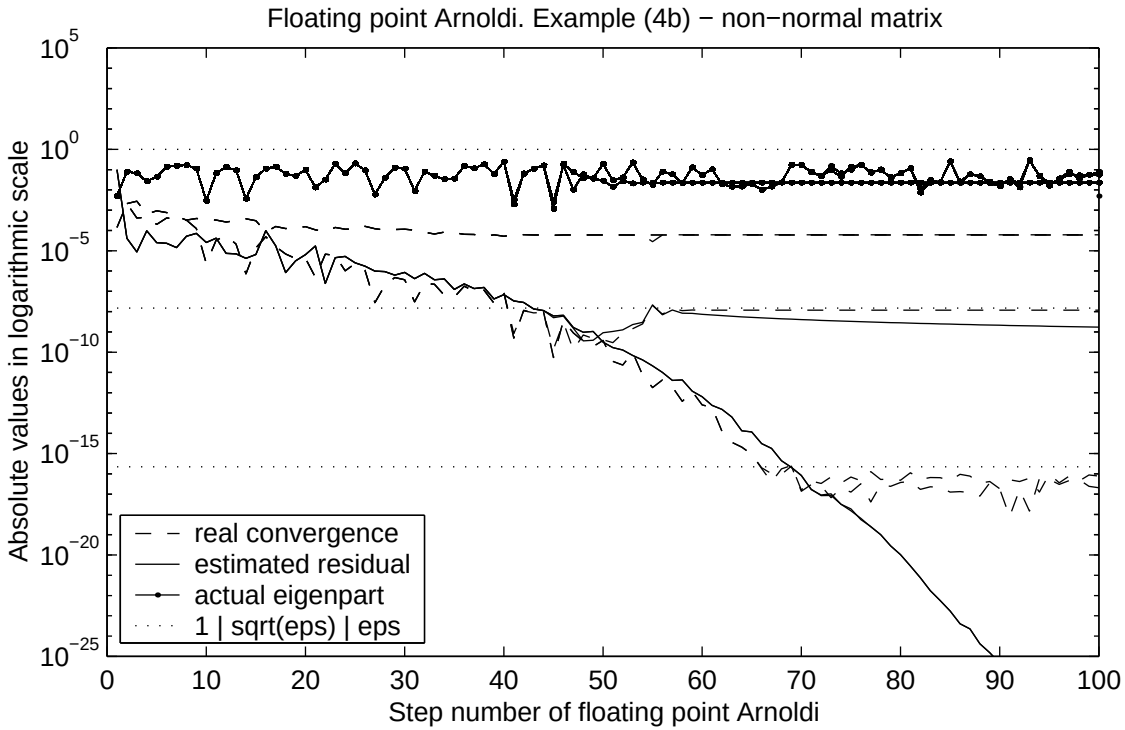
- the left eigenpart of q_{k+1}: $\widehat{v}_i^H q_{k+1}$,
- a measure of convergence: $\left(\lambda_i - \theta_j\right) \widehat{v}_i^H y_j$,
- an error term: $\widehat{v}_i^H F_k s_j$,
- an amplification factor: $c_{k+1,k} s_{kj}$.

$A \in \mathbb{R}^{100\times 100}$ normal, eigenvalues equidistant in $[0, 1]$.



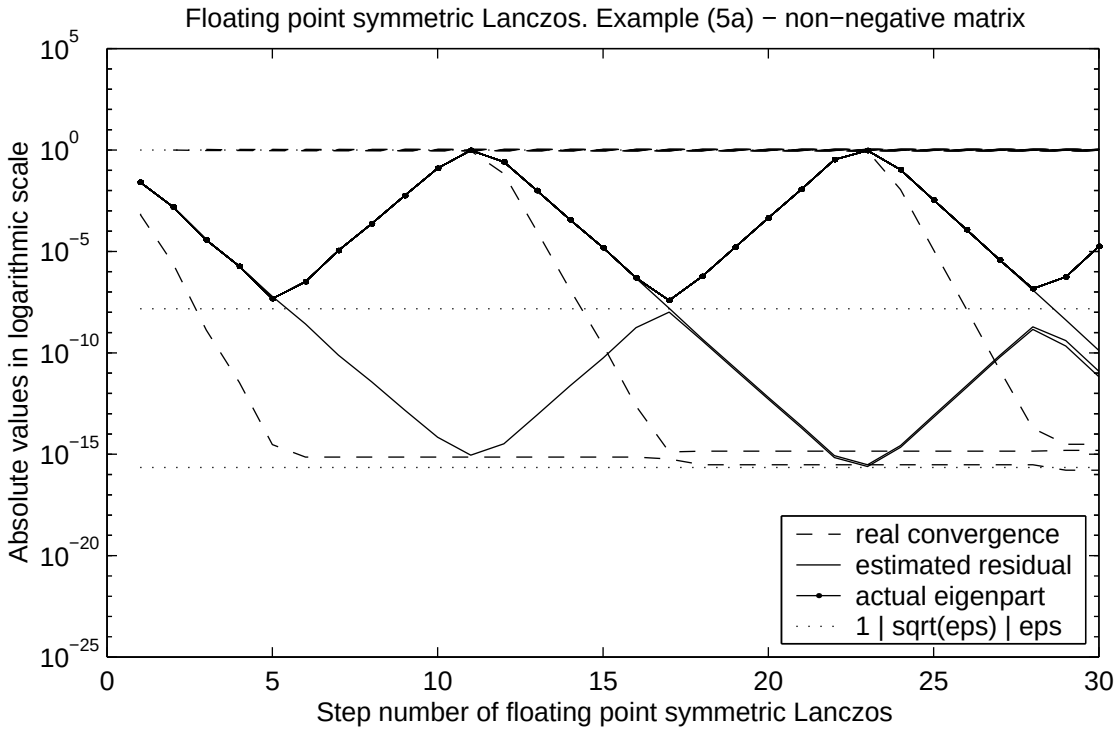
Behaviour of CGS-Arnoldi, MGS-Arnoldi, DO-Arnoldi, convergence to largest eigenvalue.

$A \in \mathbb{R}^{100\times 100}$ non-normal, eigenvalues equidistant in $[0, 1]$.



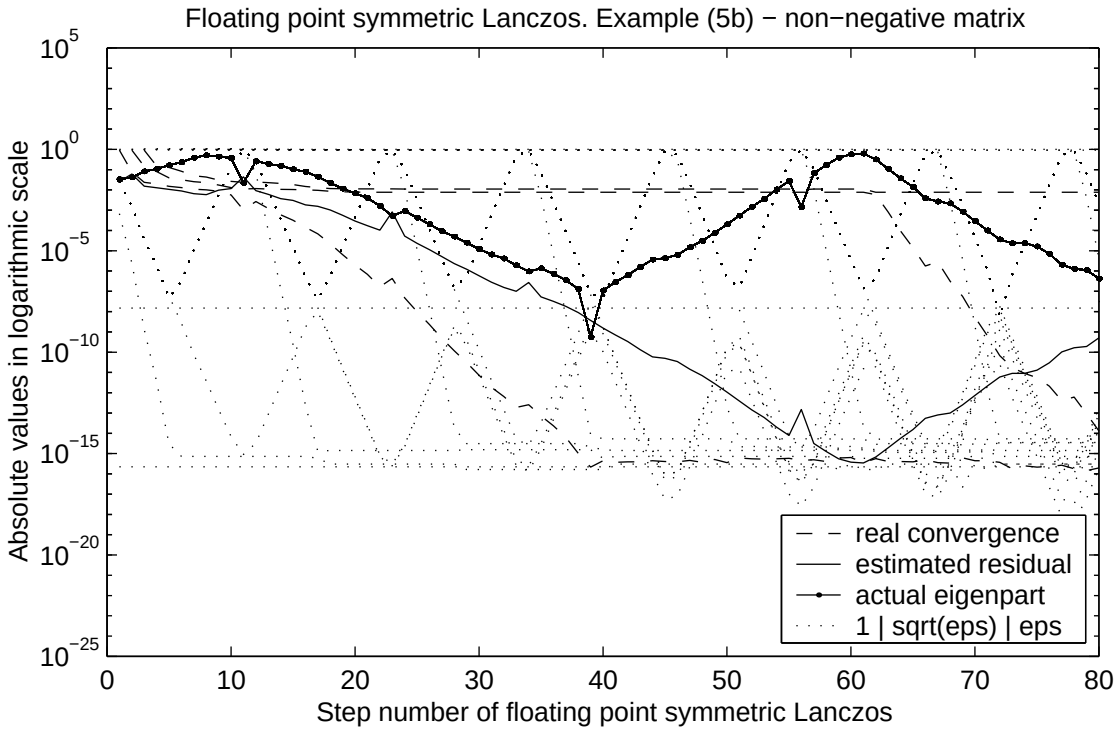
Behaviour of CGS-Arnoldi, MGS-Arnoldi, DO-Arnoldi, convergence to largest eigenvalue.

$A = A^T \in \mathbb{R}^{100\times 100}$, random entries in $[0, 1]$. Perron root well separated.



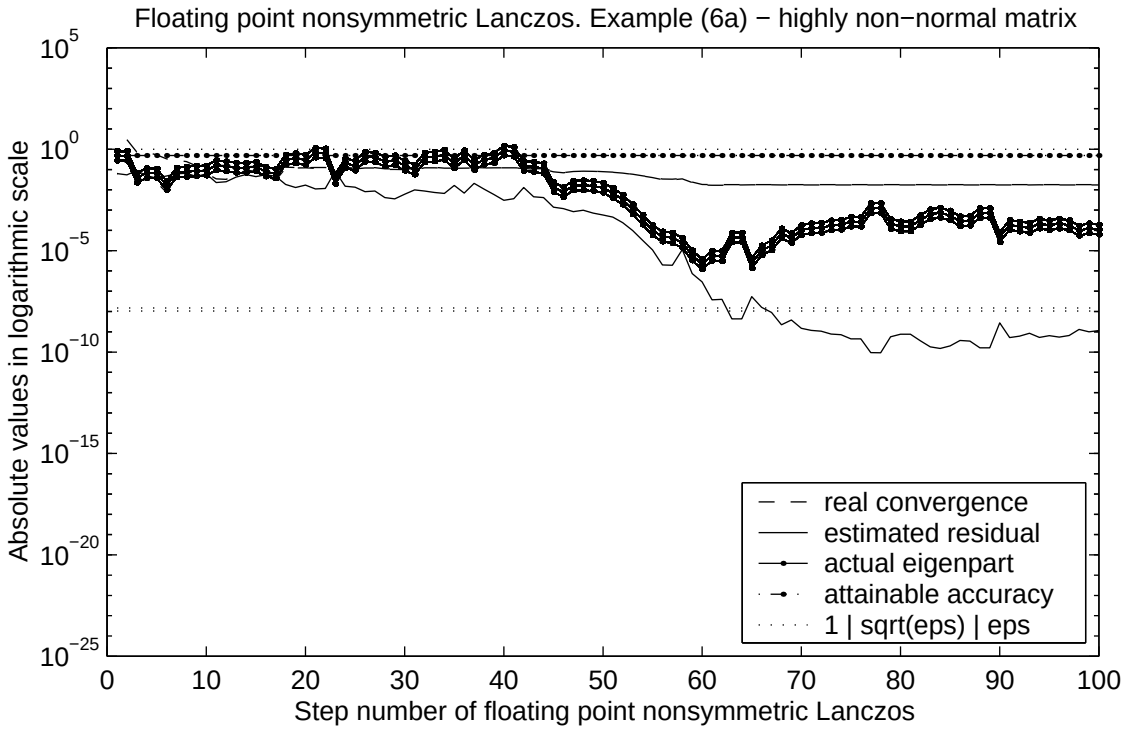
Behaviour of symmetric Lanczos, convergence to eigenvalue of largest modulus.

$A = A^T \in \mathbb{R}^{100\times 100}$, random entries in $[0, 1]$. Perron root well separated.



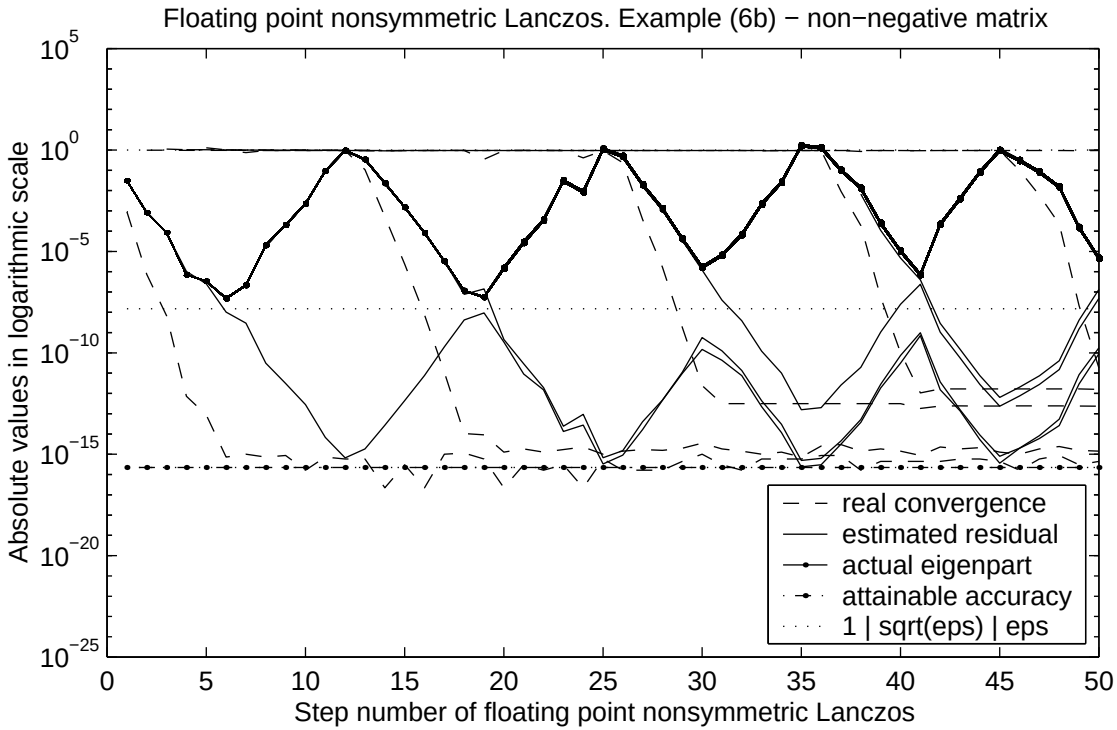
Behaviour of symmetric Lanczos, convergence to eigenvalue of largest and second largest modulus.

$A \in \mathbb{R}^{100\times 100}$, zero below fourth subdiagonal, randomly chosen between $[0, 1]$ elsewhere. A highly non-normal.



Non-symmetric Lanczos with $|\beta_j| = |\gamma_j|$. Convergence to eigenvalue of largest modulus. Left deviation, right deviation and geometric mean plotted.

$A \in \mathbb{R}^{100\times 100}$, random entries in $[0,1]$. Perron root well separated.



Behaviour of non-symmetric Lanczos, convergence to eigenvalue of largest modulus.

The formula depends on the *Ritz pair* of the actual step. Using the eigenvector basis we can get rid of the *Ritz vector*:

$$I = SS^{-1} = S\check{S}^T \quad \Rightarrow \quad e_l = S\check{S}^T e_l \equiv \sum_{j=1}^{k} \check{s}_{lj} s_j.$$

Theorem: The recurrence between vectors q_l and q_{k+1} is given by

$$\left[\sum_{j=1}^{k} \frac{c_{k+1,k} s_{kj} \check{s}_{lj}}{\lambda_i - \theta_j}\right] \hat{v}_i^H q_{k+1} = \hat{v}_i^H q_l + \hat{v}_i^H F_k \left[\sum_{j=1}^{k} \left(\frac{\check{s}_{lj}}{\lambda_i - \theta_j}\right) s_j\right].$$

For $l = 1$ we obtain a formula that reveals how the errors affect the recurrence from the beginning:

$$\left[\sum_{j=1}^{k} \frac{c_{k+1,k} s_{kj} \check{s}_{1j}}{\lambda_i - \theta_j}\right] \hat{v}_i^H q_{k+1} = \hat{v}_i^H q_1 + \hat{v}_i^H F_k \left[\sum_{j=1}^{k} \left(\frac{\check{s}_{1j}}{\lambda_i - \theta_j}\right) s_j\right].$$

Interpretation: The size of the deviation depends on the *size* of the *first component* of the *left* eigenvector $\widehat{s}_j$ of C_k and the *shape and size* of the *right* eigenvector s_j.

Next step: Application of the eigenvector – eigenvalue relation

$$(-1)^k \breve{s}(i)s(j) = \left[\frac{\chi_{H_{1:i-1}}\chi_{H_{j+1:m}}}{\chi^{(k+1)}_{H_{1:m}}}(\theta)\right] \prod_{l=i}^{j-1} h_{l+1,l}.$$

Theorem: The recurrence between basis vectors q_1 and q_{k+1} can be described by

$$\left[\sum_{j=1}^{k} \frac{\prod_{p=1}^{k} c_{p+1,p}}{\prod_{s\neq j}\left(\theta_s-\theta_j\right)\left(\lambda_i-\theta_j\right)}\right] \widehat{v}_i^H q_{k+1} = \widehat{v}_i^H q_1 + \widehat{v}_i^H F_k \left[\sum_{j=1}^{k}\left(\frac{\breve{s}_{1j}}{\lambda_i-\theta_j}\right)s_j\right]$$

A result from polynomial interpolation (Lagrange):

$$\begin{aligned}\sum_{j=1}^{k} \frac{1}{\prod_{l\neq j}\left(\theta_j-\theta_l\right)\left(\lambda_i-\theta_j\right)} &= \frac{1}{\chi_{C_k}(\lambda_i)} \sum_{j=1}^{k} \frac{\prod_{l\neq j}(\lambda_i-\theta_l)}{\prod_{l\neq j}\left(\theta_j-\theta_l\right)} \\ &= \frac{1}{\chi_{C_k}(\lambda_i)}\end{aligned}$$

Thus the following theorem holds true:

Theorem: The recurrence between basis vectors q_1 and q_{k+1} can be described by

$$\hat{v}_i^H q_{k+1} = \frac{\chi_{C_k}(\lambda_i)}{\prod_{p=1}^{k} c_{p+1,p}} \left(\hat{v}_i^H q_1 + \hat{v}_i^H F_k \left[\sum_{j=1}^{k} \left(\frac{\check{s}_{1j}}{\lambda_i-\theta_j} \right) s_j \right] \right).$$

Similarly we can get rid of the eigenvectors s_j in the error term:

$$e_l^T \left[\sum_{j=1}^{k} \left(\frac{\check{s}_{1j}}{\lambda_i - \theta_j} \right) s_j \right] = \sum_{j=1}^{k} \left(\frac{\check{s}_{1j} s_{lj}}{\lambda_i - \theta_j} \right) = \frac{\prod_{p=1}^{l} c_{p+1,p} \chi_{C_{l+1:k}}(\lambda_i)}{\chi_{C_k}(\lambda_i)}$$

This results in the following theorem:

Theorem: The recurrence between basis vectors q_1 and q_{k+1} can be described by

$$\begin{aligned} \hat{v}_i^H q_{k+1} &= \frac{\chi_{C_k}(\lambda_i)}{\prod_{p=1}^{k} c_{p+1,p}} \left(\hat{v}_i^H q_1 + \hat{v}_i^H \sum_{l=1}^{k} \frac{\prod_{p=1}^{l} c_{p+1,p} \chi_{C_{l+1:k}}(\lambda_i)}{\chi_{C_k}(\lambda_i)} f_l \right) \\ &= \frac{\chi_{C_k}(\lambda_i)}{\prod_{p=1}^{k} c_{p+1,p}} \hat{v}_i^H q_1 + \sum_{l=1}^{k} \left(\frac{\chi_{C_{l+1:k}}(\lambda_i)}{\prod_{p=l+1}^{k} c_{p+1,p}} \hat{v}_i^H f_l \right). \end{aligned}$$

Multiplication by the right eigenvectors v_i and summation gives the familiar result

Theorem: The recurrence of the basis vectors of a finite precision Krylov method can be described by

$$q_{k+1} = \frac{\chi_{C_k}(A)}{\prod_{p=1}^{k} c_{p+1,p}} \, q_1 + \sum_{l=1}^{k} \left(\frac{\chi_{C_{l+1:k}}(A)}{\prod_{p=l+1}^{k} c_{p+1,p}} \, f_l \right).$$

This result holds true even for non-diagonalisable matrices A, C_k.

The method can be interpreted as an *additive mixture* of several instances of the same method with several starting vectors.

A *severe deviation* occurs when one of the characteristic polynomials $\chi_{C_{l+1:k}}(A)$ becomes large compared to $\chi_{C_k}(A)$.

Open Questions

- Can Krylov methods be forward or backward stable?
- If so, which can?
- Are there any matrices A for which Krylov methods are stable?
- Does the stability depend on the starting vector?