Applied Krylov subspace methods

Jens-Peter M. Zemke
zemke@tu-harburg.de

joint work with:
Martin Gutknecht (IDREig);
Martin van Gijzen & Gerard Sleijpen (QMRIDR);
Olaf Rendel & Anisa Rizvanolli (classification of IDR);
Chris Paige & Ivo Panayotov (augmented backward error analysis).

Institut für Mathematik
Technische Universität Hamburg-Harburg

July 18th, 14:30-15:30

TUHH
Technische Universität Hamburg-Harburg

Outline

Classification of Krylov subspace methods

- Krylov/Hessenberg
- Arnoldi-based
- Lanczos-based
- Sonneveld-based

Connections

- Interpolation
- Approximation

Applications

- RQI and the Opitz-Larkin Method
- QMRIDR & IDREig
- Augmented Backward Error Analysis

Part I

We give an algorithmically oriented approach to Krylov subspace methods, the first method using Krylov subspaces dates to 1931, by Krylov (sic).

In our approach Krylov subspace methods are divided into three classes:

- Arnoldi-based methods (first by Hessenberg, 1940),
- Lanczos-based methods (first by Stieltjes, 1884), and
- Sonneveld-based methods (first by Bouwer, 1950).

Basics

Krylov subspaces:

$$\mathcal{K}_k := \mathcal{K}_k(\mathbf{A},\mathbf{q}) := \text{span}\left\{\mathbf{q},\mathbf{A}\mathbf{q},\mathbf{A}^2\mathbf{q},\ldots,\mathbf{A}^{k-1}\mathbf{q}\right\} = \{p_{k-1}(\mathbf{A})\mathbf{q} \mid p_{k-1} \in \Pi_{k-1}\}$$

spanned by columns of Krylov matrix

$$\mathbf{K}_k := \mathbf{K}_k(\mathbf{A},\mathbf{q}) := \left(\mathbf{q},\mathbf{A}\mathbf{q},\mathbf{A}^2\mathbf{q},\ldots,\mathbf{A}^{k-1}\mathbf{q}\right).$$

Krylov subspace methods based on ideas by:

Hessenberg: CMRH; costly;

Lanczos: CG, BiCG, QMR; short recurrence, look-ahead, transpose;

Arnoldi: GMRes; long recurrence, optimal, costly, truncation & restart;

Sonneveld: IDR, CGS, BiCGStab, BiCGStab(ℓ), IDR(s), IDR(s)Stab(ℓ); short recurrence, ~~transpose~~, {unstable,cheap}–{stable,costly}

We subsume Hessenberg and Arnoldi as "Arnoldi-based".

Hessenberg decompositions

Arnoldi- and Lanczos-based methods $\rightsquigarrow$ Hessenberg decomposition:

$$\mathbf{A}\mathbf{Q}_k + \mathbf{F}_k = \mathbf{Q}_{k+1}\underline{\mathbf{H}}_k. \quad \text{(Lanczos: } \underline{\mathbf{H}}_k = \underline{\mathbf{T}}_k,\ 2\times)$$

Sonneveld-based methods $\rightsquigarrow$ generalized Hessenberg decomposition:

$$\mathbf{A}\mathbf{V}_k + \widehat{\mathbf{F}}_k = \mathbf{A}\mathbf{G}_k\mathbf{U}_k + \mathbf{F}_k = \mathbf{G}_{k+1}\underline{\mathbf{H}}_k, \qquad \mathbf{V}_k := \mathbf{G}_k\mathbf{U}_k + \widetilde{\mathbf{F}}_k.$$

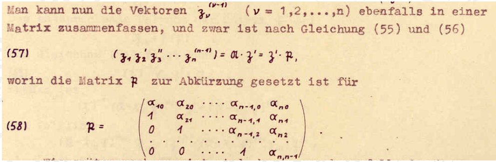
Three remarks:

- Structure: $\underline{\mathbf{H}}_k \in \mathbb{C}^{(k+1)\times k}$ always unreduced extended Hessenberg;
- Generalization: $\mathbf{I}_k \rightsquigarrow \mathbf{U}_k \in \mathbb{C}^{k\times k}$ upper triangular;
- Mnemonic for names of matrices in Sonneveld-based methods: IDR(s)-coauthor "van Gijzen" $\rightsquigarrow$ first $\mathbf{V}_k$, then $\mathbf{G}_k$.

Finite precision or inexact method $\rightsquigarrow$ perturbations $\mathbf{F}_k$, $\mathbf{F}_k = \widehat{\mathbf{F}}_k + \mathbf{A}\widetilde{\mathbf{F}}_k$.

Karl Hessenberg & “his” matrix + decomposition

„Behandlung linearer Eigenwertaufgaben mit Hilfe der Hamilton-Cayleyschen Gleichung“, Karl Hessenberg, 1. Bericht der Reihe „Numerische Verfahren“, July, 23rd 1940, page 23:



Man kann nun die Vektoren $\mathfrak{z}_\nu^{(\nu-1)}$ $(\nu = 1,2,\dots,n)$ ebenfalls in einer Matrix zusammenfassen, und zwar ist nach Gleichung (55) und (56)

(57) $$(\mathfrak{z}_1\, \mathfrak{z}_2'\, \mathfrak{z}_3'' \cdots \mathfrak{z}_n^{(n-1)}) = \mathfrak{A}\cdot\mathfrak{z}' = \mathfrak{z}'\cdot\mathfrak{P},$$

worin die Matrix $\mathfrak{P}$ zur Abkürzung gesetzt ist für

(58) $$\mathfrak{P} = \begin{pmatrix} \alpha_{10} & \alpha_{20} & \cdots & \alpha_{n-1,0} & \alpha_{n0} \\ 1 & \alpha_{21} & \cdots & \alpha_{n-1,1} & \alpha_{n1} \\ 0 & 1 & \cdots & \alpha_{n-1,2} & \alpha_{n2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & \alpha_{n,n-1} \end{pmatrix}.$$

- Hessenberg decomposition, Eqn. (57),
- Hessenberg matrix, Eqn. (58).

Karl Hessenberg (* September 8th, 1904, † February 22nd, 1959)

OR and MR for linear systems ($\mathbf{A}\mathbf{x} = \mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$)

Residuals of OR and MR approximation ($\mathbf{Q}_k\mathbf{e}_1\|\mathbf{r}_0\| = \mathbf{Q}_{k+1}\underline{\mathbf{e}}_1\|\mathbf{r}_0\| = \mathbf{r}_0$)

$$\mathbf{x}_k := \mathbf{Q}_k\mathbf{z}_k \quad \text{and} \quad \underline{\mathbf{x}}_k := \mathbf{Q}_k\underline{\mathbf{z}}_k$$

with coefficient vectors

$$\mathbf{z}_k := \mathbf{H}_k^{-1}\mathbf{e}_1\|\mathbf{r}_0\| \quad \text{and} \quad \underline{\mathbf{z}}_k := \underline{\mathbf{H}}_k^{\dagger}\underline{\mathbf{e}}_1\|\mathbf{r}_0\|$$

satisfy

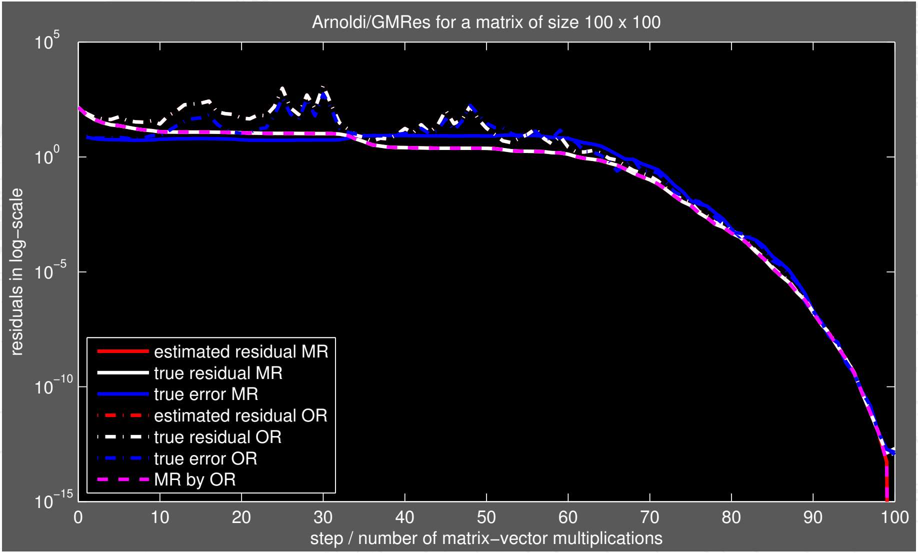
$$\mathbf{r}_k := \mathbf{r}_0 - \mathbf{A}\mathbf{x}_k = \mathcal{R}_k(\mathbf{A})\mathbf{r}_0 \quad \text{and} \quad \underline{\mathbf{r}}_k := \mathbf{r}_0 - \mathbf{A}\underline{\mathbf{x}}_k = \underline{\mathcal{R}}_k(\mathbf{A})\mathbf{r}_0.$$

Residual polynomials $\mathcal{R}_k$, $\underline{\mathcal{R}}_k$ given by

$$\mathcal{R}_k(z) := \det\left(\mathbf{I}_k - z\mathbf{H}_k^{-1}\mathbf{I}_k\right) \quad \text{and} \quad \underline{\mathcal{R}}_k(z) := \det\left(\mathbf{I}_k - z\underline{\mathbf{H}}_k^{\dagger}\underline{\mathbf{I}}_k\right).$$

Convergence of OR and MR depends on (harmonic) Ritz values.

Arnoldi/GMRes



OR and MR for eigenpairs

Well known: Ritz pairs $\rightsquigarrow$ OR eigenpairs $(\theta_j, \mathbf{y}_j)$,

$$\mathbf{y}_j := \mathbf{Q}_k \mathbf{s}_j, \quad \text{where} \quad \mathbf{H}_k \mathbf{s}_j = \mathbf{s}_j \theta_j, \quad 1 \leqslant j \leqslant k.$$

Known: (shifted) harmonic Ritz pairs $(\underline{\theta}_j, \underline{\mathbf{y}}_j)$,

$$\underline{\mathbf{y}}_j := \mathbf{Q}_k \underline{\mathbf{s}}_j, \quad \text{where} \quad \mathbf{I}_k \underline{\mathbf{s}}_j = (\underline{\mathbf{H}}_k - \tau \underline{\mathbf{I}}_k)^\dagger \underline{\mathbf{I}}_k \underline{\mathbf{s}}_j (\underline{\theta}_j - \tau), \quad 1 \leqslant j \leqslant k.$$

Less known: ρ-values, refined extraction, combinations thereof.

Mostly unknown: MR eigenpairs $(\grave{\theta}, \grave{\mathbf{y}} = \mathbf{Q}_k \grave{\mathbf{s}})$,

$$\frac{\|(\grave{\theta}\underline{\mathbf{I}}_k - \underline{\mathbf{H}}_k)\grave{\mathbf{s}}\|}{\|\grave{\mathbf{s}}\|} := \min_{z \in \mathbb{C}, \mathbf{s} \in \mathbb{C}^k, \|\mathbf{s}\|=1} \operatorname{loc} \frac{\|(z\underline{\mathbf{I}}_k - \underline{\mathbf{H}}_k)\mathbf{s}\|}{\|\mathbf{s}\|},$$

Lehmann: MR by minimization over shifts in harmonic Ritz & ρ-values.

A graphical representation

We associate with every real or complex approximate eigenpair $(\tilde{\theta}, \tilde{\mathbf{y}} = \mathbf{Q}_k \tilde{\mathbf{s}})$ a point (z, w) in the plane $\mathbb{R} \times \mathbb{R}$ or $\mathbb{C} \times \mathbb{R}$:

$$z = \tilde{\theta}, \qquad w = \frac{\|(\tilde{\theta}\underline{\mathbf{I}}_k - \underline{\mathbf{H}}_k)\tilde{\mathbf{s}}\|}{\|\tilde{\mathbf{s}}\|}. \tag{1}$$

The former gives the approximate eigenvalue, the latter gives the norm of the (quasi-)residual of the approximate eigenpair.

The norm of the residual of $(\tilde{\theta}, \tilde{\mathbf{y}})$ gives the backward error, i.e.,

$$w = \min\left\{\|\Delta\mathbf{A}\| \,:\, (\mathbf{A} + \Delta\mathbf{A})\tilde{\mathbf{y}} = \tilde{\mathbf{y}}\,\tilde{\theta}\right\}. \tag{2}$$

Remark 1: Without additional knowledge a small backward error is the best we can achieve.

Remark 2: There exist “graphical” bounds for general and “Rayleigh” approximations.

A beautiful example

As an example we use

$$\underline{\mathbf{H}}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3}$$

Its Ritz values are given by

$$\theta_{1,3} = \mp\sqrt{2} \approx \mp 1.41421356, \quad \theta_2 = 0, \tag{4}$$

its harmonic Ritz values are given by

$$\underline{\theta}_{1,3} = \mp\sqrt{2} \approx \mp 1.41421356, \quad \underline{\theta}_2 = \infty, \tag{5}$$

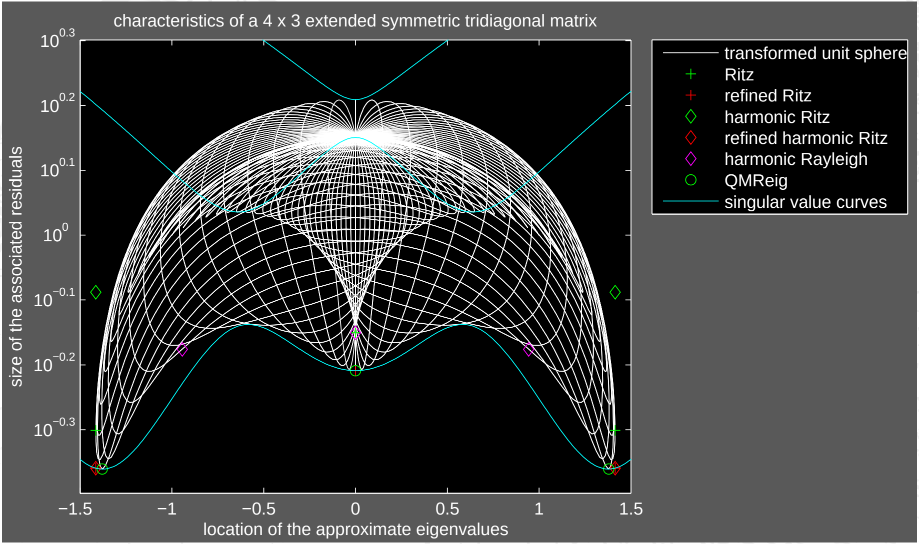
its ρ-values (Rayleigh quotients with harmonic Ritz vectors) are given by

$$\rho_{1,3} = \mp\sqrt{2} \cdot \frac{2}{3} \approx \mp 0.9428090, \quad \rho_2 = 0, \tag{6}$$

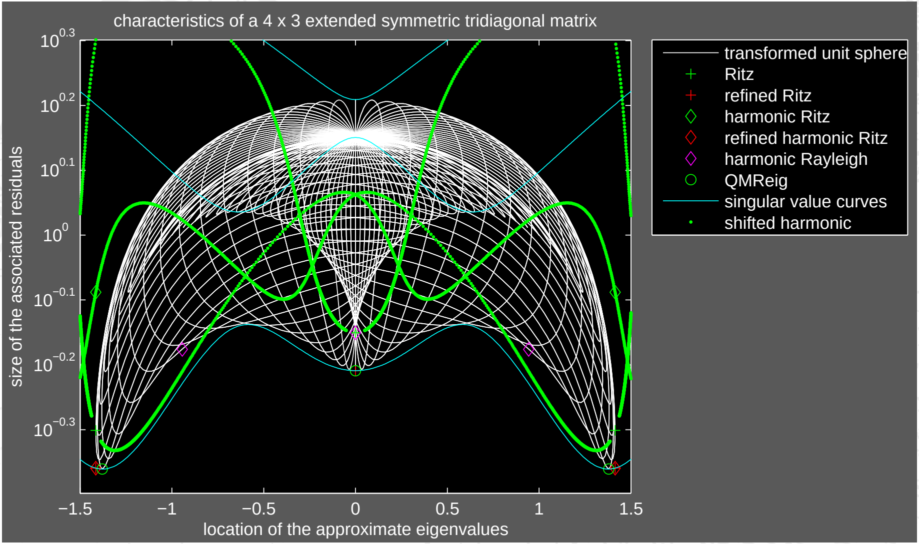
and its MR eigenvalues are given by (where $y = 276081 + 21504\sqrt{2}i$)

$$\grave{\theta}_{1,3} = \mp\frac{\sqrt{2}}{16}\sqrt{113 + 2\mathrm{Re}\,\sqrt[3]{y}} \approx \mp 1.37898323557, \quad \grave{\theta}_2 = 0. \tag{7}$$

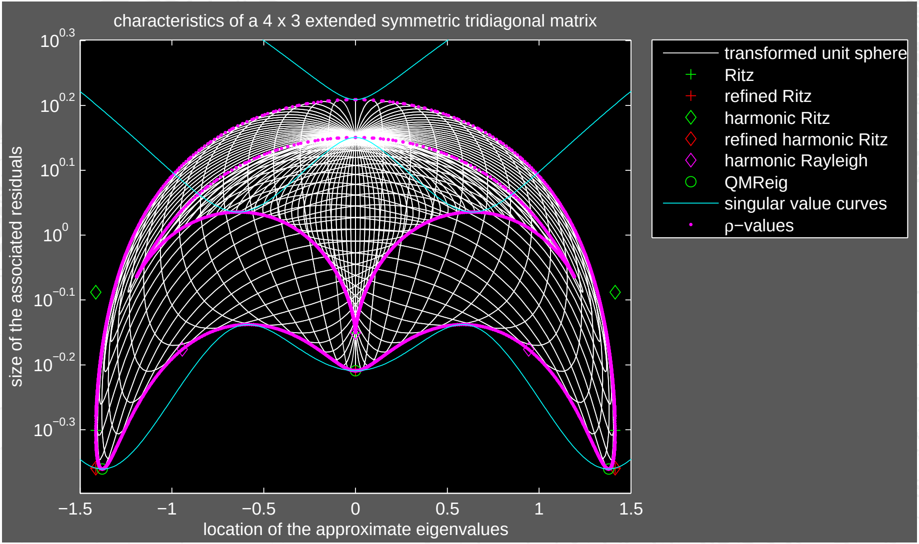
A beautiful example



A beautiful example



A beautiful example



OR and MR for Sonneveld-based methods

Generalized Hessenberg decomposition:

$$\mathbf{A}\mathbf{V}_k = \mathbf{A}\mathbf{G}_k\mathbf{U}_k = \mathbf{G}_{k+1}\underline{\mathbf{H}}_k, \quad \mathbf{V}_k := \mathbf{G}_k\mathbf{U}_k.$$

Rules of thumb: Use $\mathbf{V}_k$, not $\mathbf{G}_k$ as “basis”; insert $\mathbf{U}_k$ appropriately.

Sonneveld OR ($\mathbf{H}_k$ regular):

$$\mathbf{z}_k := \mathbf{H}_k^{-1}\mathbf{e}_1\|\mathbf{r}_0\|, \quad \mathbf{x}_k := \mathbf{V}_k\mathbf{z}_k = \mathbf{G}_k\mathbf{U}_k\mathbf{z}_k.$$

Sonneveld MR:

$$\underline{\mathbf{z}}_k := \underline{\mathbf{H}}_k^{\dagger}\underline{\mathbf{e}}_1\|\mathbf{r}_0\|, \quad \underline{\mathbf{x}}_k := \mathbf{V}_k\underline{\mathbf{z}}_k = \mathbf{G}_k\mathbf{U}_k\underline{\mathbf{z}}_k.$$

Sonneveld Ritz:

$$\mathbf{H}_k\mathbf{s}_j = \theta_j\mathbf{U}_k\mathbf{s}_j, \quad \mathbf{y}_j := \mathbf{V}_k\mathbf{s}_j = \mathbf{G}_k\mathbf{U}_k\mathbf{s}_j.$$

Sonneveld (shifted) harmonic Ritz:

$$\mathbf{I}_k\underline{\mathbf{s}}_j = (\underline{\theta}_j - \tau)\left(\underline{\mathbf{H}}_k - \tau\underline{\mathbf{U}}_k\right)^{\dagger}\underline{\mathbf{U}}_k\underline{\mathbf{s}}_j, \quad \underline{\mathbf{y}}_j := \mathbf{V}_k\underline{\mathbf{s}}_j = \mathbf{G}_k\mathbf{U}_k\underline{\mathbf{s}}_j.$$

Beyond “classical” Krylov subspace methods

Generalizations:

$$\mathcal{F}_k := \mathcal{F}_k(\mathbf{A}, \mathbf{q}) := \{f_{k-1}(\mathbf{A})\mathbf{q} \mid f_{k-1} \text{ structured, e.g., rational}\}.$$

Rational methods:

- Rational Krylov (Ruhe);
- Rayleigh Quotient Iteration (RQI); Lord Rayleigh’s original iteration.

Word of warning: I consider these to be Krylov subspace methods.

Partial motivation:

- can be captured by a generalized Hessenberg decomposition.

Single vector Krylov subspace methods (von Mises 1929, Wielandt 1944; Bernoulli 1728 $\rightsquigarrow$ Frobenius companion matrices):

- Power method (von Mises 1929),
- (Shifted) Inverse Iteration (Wielandt 1944).

Hessenberg structure

Krylov subspace method $\rightsquigarrow$ Hessenberg (tridiagonal) matrices:

- first occurrence: Wronski (one step of Laplace expansion),
- various links to (bi)orthogonal polynomials,
- interesting polynomial recursions (Schweins),
- low-rank structure: Asplund, ...

Schwein's recurrence for determinants: (Schweins, 1825, Erste Abtheilung, IV. Abschnitt, §154, Seite 361, Gleichung (560)):

$$(z\mathbf{I}_k - \mathbf{H}_k)\boldsymbol{\nu}_k(z) = \mathbf{e}_1 \frac{\chi_k(z)}{\prod_{\ell=1}^{k} h_{\ell+1,\ell}}, \quad (\check{\boldsymbol{\nu}}_k(z))^{\mathsf{T}}(z\mathbf{I}_k - \mathbf{H}_k) = \frac{\chi_k(z)}{\prod_{\ell=1}^{k} h_{\ell+1,\ell}} \mathbf{e}_k^{\mathsf{T}},$$

with polynomial vectors ($\chi_{i:j}(z) := \det\left(z\mathbf{I}_{j-i+1} - \mathbf{H}_{i:j}\right)$)

$$\mathbf{e}_i^{\mathsf{T}}\boldsymbol{\nu}_k(z) := \frac{\chi_{i+1:k}(z)}{\prod_{\ell=i+1}^{k} h_{\ell,\ell-1}}, \quad \mathbf{e}_i^{\mathsf{T}}\check{\boldsymbol{\nu}}_k(z) := \frac{\chi_{1:i-1}(z)}{\prod_{\ell=1}^{i-1} h_{\ell+1,\ell}},$$

$\rightsquigarrow$ Adjugate; inverse; eigenvectors and principal vectors; nullspace.

Linear independence ⇝ orthonormality

Krylov matrix $\mathbf{K}_{k+1}(\mathbf{A},\mathbf{q})$ rank deficient (k minimal) ⇝ minimal polynomial μ_k:

$$\mathbf{K}_k(\mathbf{A},\mathbf{q})\mathbf{c} = \mathbf{A}^k\mathbf{q} \quad \Rightarrow \quad \mu_k(\mathbf{A})\mathbf{q} = \mathbf{o}_n, \quad \mu_k(z) = z^k - \sum_{i=1}^{k} c_i z^{i-1}.$$

Eigenvalues, Inverse:

$$\mathbf{A}\mathbf{K}_k = \mathbf{K}_k\mathbf{F}_k, \quad \mathbf{F}_k := \begin{pmatrix} \mathbf{o}_{k-1}^T & \mathbf{c} \\ \mathbf{I}_{k-1} & \end{pmatrix}, \quad \mathbf{A}\mathbf{x} = \mathbf{A}(\mathbf{A}^{k-1}\mathbf{q} - \sum_{i=2}^{k} c_i\mathbf{A}^{i-2}\mathbf{q}) = \mathbf{q}c_1.$$

Natural idea: use linearly independent vectors for some other basis $\mathbf{Q}_k$.

⇝ nested basis transformation: $\mathbf{K}_{k+1} = \mathbf{Q}_{k+1}\mathbf{R}_{k+1}$ with $\mathbf{R}_{k+1}$ upper triangular.

Hessenberg: LU decomposition: $\mathbf{K}_{k+1} = \mathbf{L}_{k+1}\mathbf{R}_{k+1}, r_{ii} = 1, 1 \leqslant i \leqslant k+1$.

Arnoldi: orthonormal basis, i.e., QR decomposition: $\mathbf{K}_{k+1} = \mathbf{Q}_{k+1}\mathbf{R}_{k+1}$.

Extended Hessenberg matrix as quotient: $(\mathbf{e}_1, \underline{\mathbf{H}}_k) = \mathbf{R}_{k+1}\begin{pmatrix} 1 & \mathbf{o}_k^T \\ \mathbf{o}_k & \mathbf{R}_k^{-1} \end{pmatrix}$.

Arnoldi based on orthogonal projection: minimal coeffs $\mathbf{c}$ ⇝ "optimal".

Arnoldi

Arnoldi decomposition:

$$\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1}\underline{\mathbf{H}}_k.$$

Construction:

$$\underline{\mathbf{H}}_0 = [\ ];\ \mathbf{Q}_1 = \mathbf{q}_1 = \mathbf{q}/\|\mathbf{q}\|;$$

for i=1:k do

$$\mathbf{r} = \mathbf{A}\mathbf{q}_i;$$
$$\mathbf{h}_i = \mathbf{Q}_i^{\mathsf{H}}\mathbf{r};$$
$$\mathbf{r} = \mathbf{r} - \mathbf{Q}_i\mathbf{h}_i;$$
$$h_{i+1,i} = \|\mathbf{r}\|;$$
$$\mathbf{q}_{i+1} = \mathbf{r}/h_{i+1,i};$$
$$\underline{\mathbf{H}}_i = \begin{pmatrix} \underline{\mathbf{H}}_{i-1} & \mathbf{h}_i \\ \mathbf{o}_{i-1}^{\mathsf{T}} & h_{i+1,i} \end{pmatrix};$$
$$\mathbf{Q}_{i+1} = (\mathbf{Q}_i, \mathbf{q}_{i+1});$$

done

Gram-Schmidt variant.
Others possible.

Other inner products or semi-inner products possible.

Linear independence using less vectors

Lanczos: biorthonormal bases $\rightsquigarrow \widehat{\mathbf{Q}}_{k+1}^{\mathsf{H}}\mathbf{Q}_{k+1} = \mathbf{I}_{k+1}$ of

$$\mathcal{K}_k := \mathcal{K}_k(\mathbf{A}, \mathbf{q}) := \text{span}\left\{\mathbf{q}, \mathbf{A}\mathbf{q}, \mathbf{A}^2\mathbf{q}, \ldots, \mathbf{A}^{k-1}\mathbf{q}\right\} = \{p_{k-1}(\mathbf{A})\mathbf{q} \mid p_{k-1} \in \Pi_{k-1}\},$$
$$\widehat{\mathcal{K}}_k := \mathcal{K}_k(\mathbf{A}^{\mathsf{H}}, \widehat{\mathbf{q}}) := \text{span}\{\widehat{\mathbf{q}}, \mathbf{A}^{\mathsf{H}}\widehat{\mathbf{q}}, \mathbf{A}^{2\mathsf{H}}\widehat{\mathbf{q}}, \ldots, \mathbf{A}^{(k-1)\mathsf{H}}\widehat{\mathbf{q}}\}.$$

Based on three-term recurrence for the solutions η_k, $\widetilde{\eta}_k$ of the Hankel systems

$$\mathbf{C}_{k+1}\begin{pmatrix}\eta_k\\1\end{pmatrix} = \mathbf{e}_{k+1}h_k, \quad \widetilde{\mathbf{C}}_{k+2}\begin{pmatrix}\widetilde{\eta}_k\\1\end{pmatrix} = \mathbf{e}_{k+1}\widetilde{h}_{k+1},$$

$$\mathbf{C}_{k+1} = \widehat{\mathbf{K}}_{k+1}^{\mathsf{H}}\mathbf{K}_{k+1} = \begin{pmatrix} c_0 & c_1 & c_2 & \cdots & c_k \\ c_1 & c_2 & c_3 & \cdots & c_{k+1} \\ c_2 & c_3 & c_4 & \cdots & c_{k+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_k & c_{k+1} & c_{k+2} & \cdots & c_{2k} \end{pmatrix}, \quad c_i = \widehat{\mathbf{q}}^{\mathsf{H}}\mathbf{A}^i\mathbf{q},$$

where $\widetilde{\mathbf{C}}_{k+2}$ is $\mathbf{C}_{k+2}$ w/o first row & last column.

Modern implementations

(Example of) Lanczos decompositions:

$$\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1}\underline{\mathbf{T}}_k, \quad \mathbf{A}^{\mathsf{H}}\widehat{\mathbf{Q}}_k = \widehat{\mathbf{Q}}_{k+1}\underline{\widehat{\mathbf{T}}}_k, \quad \widehat{\mathbf{Q}}_{k+1}^{\mathsf{H}}\mathbf{Q}_{k+1} = \mathbf{I}_{k+1}, \quad \mathbf{T}_k^{\mathsf{H}} = \widehat{\mathbf{T}}_k.$$

Implementation nowadays usually based on two-sided Gram-Schmidt:

$$\begin{aligned} \mathbf{r} &= \mathbf{A}\ \ \mathbf{q}_k - \mathbf{q}_k\alpha_k - \mathbf{q}_{k-1}\overline{\widehat{\beta}_k}, \\ \widehat{\mathbf{r}} &= \mathbf{A}^{\mathsf{H}}\widehat{\mathbf{q}}_k - \widehat{\mathbf{q}}_k\overline{\alpha_k} - \widehat{\mathbf{q}}_{k-1}\overline{\beta_k}, \end{aligned} \quad \overline{\widehat{\beta}_{k+1}}\beta_{k+1} = \langle\widehat{\mathbf{r}}, \mathbf{r}\rangle, \quad \begin{aligned} \mathbf{q}_{k+1} &= \mathbf{r}/\beta_{k+1}, \\ \widehat{\mathbf{q}}_{k+1} &= \widehat{\mathbf{r}}/\widehat{\beta}_{k+1}. \end{aligned}$$

- Hankel matrices may become singular vs. inner products may be zero: need for look-ahead.
- Problems with incurable breakdown (in finite fields): $\rightsquigarrow$ Taylor's mismatch theorem.

Avoiding the use of the transpose

Lanczos method can be generalized:

- block variants $\rightsquigarrow$ ℓ left- and right-hand starting vectors;
- block variants with different number of left- and right-hand starting vectors $\rightsquigarrow$ applications in model reduction.

Variants denoted by Lanczos(ℓ, s), ℓ denotes number of the left-hand starting vectors and s denotes number of right-hand starting vectors. Linear systems: left (block) Krylov subspace is not used to compute approximations.

- Brower, 1950: scalars c_i can be formed using only powers of $\mathbf{A}$, no need for transpose, but $n \rightsquigarrow 2n$;
- Sonneveld, 1979: Birth of "Induced Dimension Reduction";
- Sonneveld, 1989: $\langle \overline{p}(\mathbf{A}^{\mathsf{H}})\hat{\mathbf{r}}_0, q(\mathbf{A})\mathbf{r}_0 \rangle = \langle \hat{\mathbf{r}}_0, p(\mathbf{A})q(\mathbf{A})\mathbf{r}_0 \rangle$;
- Famous classical examples of Sonneveld-based methods: CGS, BiCGStab, Wiedemann's method (for finite fields);
- Lanczos$(s, 1)$ without transpose: IDR(s) & Sonneveld spaces.

IDR(s)

IDR spaces:

$$\mathcal{G}_0 := \mathcal{K}(\mathbf{A}, \mathbf{q}), \qquad \text{(full Krylov subspace)}$$
$$\mathcal{G}_j := (\mathbf{A} - \mu_j \mathbf{I})(\mathcal{G}_{j-1} \cap \mathcal{S}), \quad j \geqslant 1, \quad \mu_j \in \mathbb{C},$$

where

$$\operatorname{codim}(\mathcal{S}) = s, \quad \text{e.g.,} \quad \mathcal{S} = \operatorname{span}\{\widetilde{\mathbf{R}}_0\}^\perp, \quad \widetilde{\mathbf{R}}_0 \in \mathbb{C}^{n \times s}.$$

Interpreted as Sonneveld spaces (Sleijpen, Sonneveld, van Gijzen 2010):

$$\mathcal{G}_j = \mathcal{S}_j(P_j, \mathbf{A}, \widetilde{\mathbf{R}}_0) := \left\{ M_j(\mathbf{A})\mathbf{v} \mid \mathbf{v} \perp \mathcal{K}_j(\mathbf{A}^{\mathsf{H}}, \widetilde{\mathbf{R}}_0), \mathbf{v} \in \mathcal{G}_0 \right\},$$
$$M_j(z) := \prod_{i=1}^{j} (z - \mu_i).$$

Image of shrinking space: Induced Dimension Reduction.

IDR(s)

IDR spaces nested:

$$\{\mathbf{o}\} = \mathcal{G}_{\text{jmax}} \subsetneq \cdots \subsetneq \mathcal{G}_{j+1} \subsetneq \mathcal{G}_j \subsetneq \mathcal{G}_{j-1} \subsetneq \cdots \subsetneq \mathcal{G}_2 \subsetneq \mathcal{G}_1 \subsetneq \mathcal{G}_0.$$

How many vectors in $\mathcal{G}_j \setminus \mathcal{G}_{j+1}$? In generic case, $s+1$.

Stable basis: Partially orthonormalize basis vectors $\mathbf{g}_k$, $1 \leqslant k \leqslant n$:

Arnoldi: compute orthonormal basis of $\mathcal{K}_{s+1} \subset \mathcal{G}_0$,

$$\mathbf{A}\mathbf{G}_s = \mathbf{G}_{s+1}\underline{\mathbf{H}}_s.$$

"Lanczos": perform intersection $\mathcal{G}_j \cap \mathcal{S}$, map, and orthonormalize,

$$\mathbf{v}_k = \sum_{i=k-s}^{k} \mathbf{g}_i \gamma_i, \quad \widetilde{\mathbf{R}}_0^{\mathsf{H}} \mathbf{v}_k = \mathbf{o}_s, \quad k \geqslant s+1,$$

$$\mathbf{g}_{k+1}\nu_{k+1} = (\mathbf{A} - \mu_j \mathbf{I})\mathbf{v}_k - \sum_{i=k-j(s+1)-1}^{k} \mathbf{g}_i \nu_i, \quad j = \left\lfloor \frac{k-1}{s+1} \right\rfloor.$$

IDR(s)

Generalized Hessenberg decomposition:

$$\mathbf{A}\mathbf{V}_k = \mathbf{A}\mathbf{G}_k\mathbf{U}_k = \mathbf{G}_{k+1}\underline{\mathbf{H}}_k,$$

where $\mathbf{U}_k \in \mathbb{C}^{k\times k}$ upper triangular.

Structure of Sonneveld pencils:

$$\mathbf{H}_k = \begin{pmatrix}
\times & \times & \times & \times & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ \\
+ & \times & \times & \times & \times & \circ & \circ & \circ & \circ & \circ & \circ & \circ \\
\circ & + & \times & \times & \times & \times & \circ & \circ & \circ & \circ & \circ & \circ \\
\circ & \circ & + & \times & \times & \times & \times & \circ & \circ & \circ & \circ & \circ \\
\circ & \circ & \circ & + & \times & \times & \times & \times & \circ & \circ & \circ & \circ \\
\circ & \circ & \circ & \circ & + & \times & \times & \times & \times & \circ & \circ & \circ \\
\circ & \circ & \circ & \circ & \circ & + & \times & \times & \times & \times & \circ & \circ \\
\circ & \circ & \circ & \circ & \circ & \circ & + & \times & \times & \times & \times & \circ \\
\circ & \circ & \circ & \circ & \circ & \circ & \circ & + & \times & \times & \times & \times \\
\circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & + & \times & \times & \times \\
\circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & + & \times & \times \\
\circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & + & \times
\end{pmatrix}, \quad
\begin{pmatrix}
\times & \times & \times & \times & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ \\
\circ & \times & \times & \times & \times & \circ & \circ & \circ & \circ & \circ & \circ & \circ \\
\circ & \circ & \times & \times & \times & \times & \circ & \circ & \circ & \circ & \circ & \circ \\
\circ & \circ & \circ & \times & \times & \times & \times & \circ & \circ & \circ & \circ & \circ \\
\circ & \circ & \circ & \circ & \times & \times & \times & \times & \circ & \circ & \circ & \circ \\
\circ & \circ & \circ & \circ & \circ & \times & \times & \times & \times & \circ & \circ & \circ \\
\circ & \circ & \circ & \circ & \circ & \circ & \times & \times & \times & \times & \circ & \circ \\
\circ & \circ & \circ & \circ & \circ & \circ & \circ & \times & \times & \times & \times & \circ \\
\circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \times & \times & \times & \times \\
\circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \times & \times & \times \\
\circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \times & \times \\
\circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \times
\end{pmatrix} = \mathbf{U}_k$$

Part II

The connections between

- Krylov subspace methods and
- (generalized) Hessenberg decompositions

on the one hand, and

- polynomials,
- interpolation &
- approximation

on the other are established.

First: Relations between the three approaches to Krylov subspace methods.

Connections between the three approaches

(Generalized) Hessenberg decompositions:

$$\begin{aligned} \text{Arnoldi:} \quad & \mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1}\underline{\mathbf{H}}_k, \\ \text{Lanczos:} \quad & \mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1}\underline{\mathbf{T}}_k, \quad \mathbf{A}^{\mathsf{H}}\widehat{\mathbf{Q}}_k = \widehat{\mathbf{Q}}_{k+1}\underline{\widehat{\mathbf{T}}}_k, \\ \text{Sonneveld:} \quad & \mathbf{A}\mathbf{V}_k = \mathbf{A}\mathbf{G}_k\mathbf{U}_k = \mathbf{G}_{k+1}\underline{\mathbf{H}}_k, \quad \mathbf{V}_k = \mathbf{G}_k\mathbf{U}_k. \end{aligned}$$

- Arnoldi and Lanczos ($\widehat{\mathbf{q}} = \mathbf{q}$) are the same (so-called symmetric Lanczos) for Hermitean matrices (pencil $(\mathbf{K}, \mathbf{M})$: $(\mathbf{K} - \sigma\mathbf{M})^{-1}\mathbf{M}$ is $\mathbf{M}$-symmetric):

$$\mathbf{H}_k = \mathbf{Q}_k^{\mathsf{H}}\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_k^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{Q}_k = (\mathbf{Q}_k^{\mathsf{H}}\mathbf{A}\mathbf{Q}_k)^{\mathsf{H}} = \mathbf{H}_k^{\mathsf{H}} = \mathbf{T}_k;$$

- Lanczos is typically slower in terms of matrix-vector multiplies, faster in terms of computing time, but less stable than Arnoldi;
- Sonneveld is Lanczos multiplied with extra polynomials;
- Sonneveld with varying s fills the gap between Lanczos and Arnoldi, reduces risk of breakdown.

Introducing: polynomials

For simplicity we only consider perturbed methods that satisfy

$$\mathbf{A}\mathbf{Q}_k + \mathbf{F}_k = \mathbf{Q}_{k+1}\underline{\mathbf{H}}_k.$$

Polynomials based on computed $\mathbf{H}_k$ or $\underline{\mathbf{H}}_k$ $\rightsquigarrow$ useful properties.

Polynomials named by property. In (Zemke, 2007) we considered the following five types of polynomials:

- basis polynomials $\mathcal{B}_k$,
- adjugate polynomials $\mathcal{A}_k$,
- Lagrange interpolation polynomials $\mathcal{L}_k[z^{-1}]$ and $\underline{\mathcal{L}}_k[z^{-1}]$,
- Lagrange interpolation polynomials $\mathcal{L}_k[1-\delta_{z0}]$ and $\underline{\mathcal{L}}_k[1-\delta_{z0}]$,
- residual polynomials $\mathcal{R}_k$ and $\underline{\mathcal{R}}_k$.

We restrict ourselves to $\mathcal{A}_k$, $\mathcal{L}_k[z^{-1}]$, $\mathcal{L}_k[1-\delta_{z0}]$ and $\mathcal{R}_k$.

Adjugate polynomials

First we consider certain bivariate polynomials – the adjugate polynomials.

- Property:

$$\mathcal{A}_k(z, \mathbf{H}_k) = \mathrm{adj}(z\mathbf{I}_k - \mathbf{H}_k).$$

- Implies (Schweins, 1825; Zemke, 2006)

$$\mathcal{A}_k(\theta_j, \mathbf{H}_k)\mathbf{e}_1 = \mathbf{s}_j, \qquad \mathbf{H}_k\mathbf{s}_j = \mathbf{s}_j\theta_j$$

  for all eigenvalues (Ritz values) θ_j of $\mathbf{H}_k$.

- Definition:

$$\mathcal{A}_k(\theta, z) := \frac{\chi_k(\theta) - \chi_k(z)}{\theta - z}, \qquad \chi_k(z) := \det(z\mathbf{I}_k - \mathbf{H}_k).$$

- Generalization:

$$\mathcal{A}_{\ell+1:k}(\theta, z) := \frac{\chi_{\ell+1:k}(\theta) - \chi_{\ell+1:k}(z)}{\theta - z}, \qquad \ell = 0, 1, \ldots, k.$$

Adjugate polynomials and Ritz vectors

Theorem (Ritz vectors)

Let $\mathbf{H}_k\mathbf{S}_\theta = \mathbf{S}_\theta\mathbf{J}_\theta$ (for a certain $\mathbf{S}_\theta$). Let the Ritz matrix be given by $\mathbf{Y}_\theta := \mathbf{Q}_k\mathbf{S}_\theta$. Then

$$\operatorname{vec}(\mathbf{Y}_\theta) = \begin{pmatrix} \mathcal{A}_k(\theta,\mathbf{A}) \\ \mathcal{A}_k'(\theta,\mathbf{A}) \\ \vdots \\ \dfrac{\mathcal{A}_k^{(\alpha-1)}(\theta,\mathbf{A})}{(\alpha-1)!} \end{pmatrix} \mathbf{q}_1 + \sum_{\ell=1}^{k} \prod_{j=1}^{\ell-1} h_{j+1,j} \begin{pmatrix} \mathcal{A}_{\ell+1:k}(\theta,\mathbf{A}) \\ \mathcal{A}_{\ell+1:k}'(\theta,\mathbf{A}) \\ \vdots \\ \dfrac{\mathcal{A}_{\ell+1:k}^{(\alpha-1)}(\theta,\mathbf{A})}{(\alpha-1)!} \end{pmatrix} \mathbf{f}_\ell, \qquad (8)$$

with derivation with respect to the shift θ.

We might scale differently such that (here only for approximate eigenvectors)

$$\mathbf{y} = \frac{\mathcal{A}_k(\theta,\mathbf{A})}{\prod_{j=1}^{k-1} h_{j+1,j}}\mathbf{q}_1 + \sum_{\ell=1}^{k} \frac{\mathcal{A}_{\ell+1:k}(\theta,\mathbf{A})}{\prod_{j=\ell+1}^{k-1} h_{j+1,j}} \cdot \frac{\mathbf{f}_\ell}{h_{\ell+1,\ell}}.$$

Lagrange polynomials

We consider Lagrange interpolation polynomials interpolating the inverse and a singularly perturbed identity.

The Lagrange interpolation of the inverse is denoted by $\mathcal{L}_k[z^{-1}](z)$.

- Property:

$$\mathcal{L}_k[z^{-1}](\mathbf{H}_k) = \mathbf{H}_k^{-1}.$$

- Definition:

$$\mathcal{L}_k[z^{-1}](z) := \frac{\chi_k(0) - \chi_k(z)}{z\chi_k(0)} = -\frac{\mathcal{A}_k(0, z)}{\chi_k(0)}.$$

- Generalization:

$$\mathcal{L}_{\ell+1:k}[z^{-1}](z) := \frac{\chi_{\ell+1:k}(0) - \chi_{\ell+1:k}(z)}{z\chi_{\ell+1:k}(0)} = -\frac{\mathcal{A}_{\ell+1:k}(0, z)}{\chi_{\ell+1:k}(0)}, \quad \ell = 0, 1, \ldots, k.$$

Lagrange polynomials and OR iterates

Theorem (OR iterates)

Suppose that all $\mathbf{H}_{\ell+1:k}$ *are regular. Define* $\mathbf{z}_k := \mathbf{H}_k^{-1}\mathbf{e}_1\|\mathbf{r}_0\|$ *and* $\mathbf{x}_k := \mathbf{Q}_k\mathbf{z}_k$*. Then*

$$\mathbf{x}_k = \mathcal{L}_k[z^{-1}](\mathbf{A})\mathbf{r}_0 - \sum_{\ell=1}^{k} \mathcal{L}_{\ell+1:k}[z^{-1}](\mathbf{A})\,\mathbf{f}_\ell z_{\ell k}. \tag{9}$$

Really sloppily speaking, in case of convergence,

$$\mathbf{x}_\infty = \mathbf{A}^{-1}\mathbf{r}_0 + \mathbf{A}^{-1}\mathbf{F}_\infty\mathbf{z}_\infty = \mathbf{A}^{-1}(\mathbf{r}_0 + \mathbf{F}_\infty\mathbf{z}_\infty).$$

Proving convergence is the hard task.

Lagrange polynomials (continued)

We consider Lagrange interpolation polynomials interpolating the inverse and a singularly perturbed identity.

The Lagrange interpolation of the singularly perturbed identity is denoted by $\mathcal{L}_k^0[1-\delta_{z0}](z)$.

- Properties:

$$\mathcal{L}_k^0[1-\delta_{z0}](\mathbf{H}_k) = \mathbf{I}_k, \qquad \mathcal{L}_k^0[1-\delta_{z0}](0) = 0.$$

- Definition:

$$\mathcal{L}_k^0[1-\delta_{z0}](z) := \frac{\chi_k(0)-\chi_k(z)}{\chi_k(0)} = \mathcal{L}_k[z^{-1}](z)z.$$

- Generalization ($\ell = 0, 1, \ldots, k$):

$$\mathcal{L}_{\ell+1:k}^0[1-\delta_{z0}](z) := \frac{\chi_{\ell+1:k}(0)-\chi_{\ell+1:k}(z)}{\chi_{\ell+1:k}(0)} = \mathcal{L}_{\ell+1:k}[z^{-1}](z)z.$$

Residual polynomials

Well-known residual polynomials (Stiefel, 1955), denoted by $\mathcal{R}_k(z)$.

- Properties:

$$\mathcal{R}_k(\mathbf{H}_k) = \mathbf{O}_k, \qquad \mathcal{R}_k(0) = 1.$$

- Definition:

$$\mathcal{R}_k(z) := \frac{\chi_k(z)}{\chi_k(0)} = 1 - \mathcal{L}_k^0[1 - \delta_{z0}](z) = \det\left(\mathbf{I}_k - z\mathbf{H}_k^{-1}\right).$$

- Generalization ($\ell = 0, 1, \ldots, k$):

$$\mathcal{R}_{\ell+1:k}(z) := \frac{\chi_{\ell+1:k}(z)}{\chi_{\ell+1:k}(0)} = 1 - \mathcal{L}_{\ell+1:k}^0[1 - \delta_{z0}](z).$$

Two types of polynomials $\rightsquigarrow$ two expressions for the OR residuals.

Residual polynomials and OR residuals

Theorem (OR residuals)

Suppose $\mathbf{q}_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$ *and let all* $\mathbf{H}_{\ell+1:k}$ *be invertible. Let* $\mathbf{x}_k$ *denote the OR iterate and* $\mathbf{r}_k = \mathbf{r}_0 - \mathbf{A}\mathbf{x}_k$ *the corresponding OR residual.*
Then

$$\begin{aligned} \mathbf{r}_k &= \mathcal{R}_k(\mathbf{A})\mathbf{r}_0 + \sum_{\ell=1}^{k} \mathcal{L}^0_{\ell+1:k}[1-\delta_{z0}](\mathbf{A})\,\mathbf{f}_\ell z_{\ell k} \\ &= \mathcal{R}_k(\mathbf{A})\mathbf{r}_0 - \sum_{\ell=1}^{k} \mathcal{R}_{\ell+1:k}(\mathbf{A})\,\mathbf{f}_\ell z_{\ell k} + \mathbf{F}_k\mathbf{z}_k. \end{aligned} \tag{10}$$

First expression: related to perturbation amplification.

Second expression: related to the attainable accuracy.

The connection to approximation theory

OR and MR perform polynomial approximation. Best understood: case $\mathbf{Q}_{k+1}$ orthonormal, i.e., Arnoldi/GMRES.

OR = Arnoldi/symmetric Lanczos:

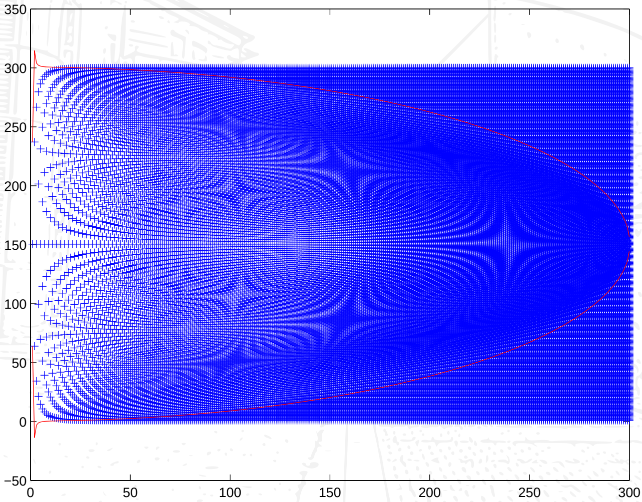
$$\min_{p\in\Pi_k} \|p(\mathbf{A})\mathbf{q}\|, \quad p(z) = z^k + \cdots \quad \Rightarrow \quad p(z) = \chi_k(z) = \det(z\mathbf{I}_k - \mathbf{H}_k).$$

MR = GMRES/MINRES:

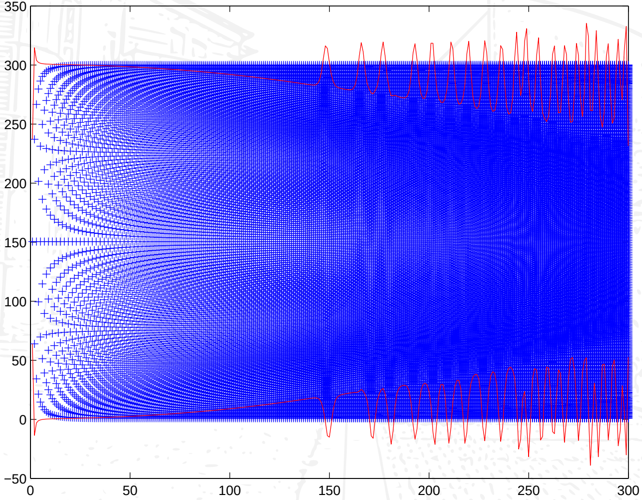
$$\min_{p\in\Pi_k} \|p(\mathbf{A})\mathbf{q}\|, \quad p(z) = 1 \ + \cdots \quad \Rightarrow \quad p(z) = \underline{\chi}_k(z) = \det(z\mathbf{I}_k - \underline{\mathbf{H}}_k^{\dagger}\underline{\mathbf{I}}_k).$$

- Others: Sonneveld $\approx$ Lanczos $\approx$ Arnoldi;
- Link to Potential Theory via Green's functions;
- Potential Theory: also for eigenvalue approximations.

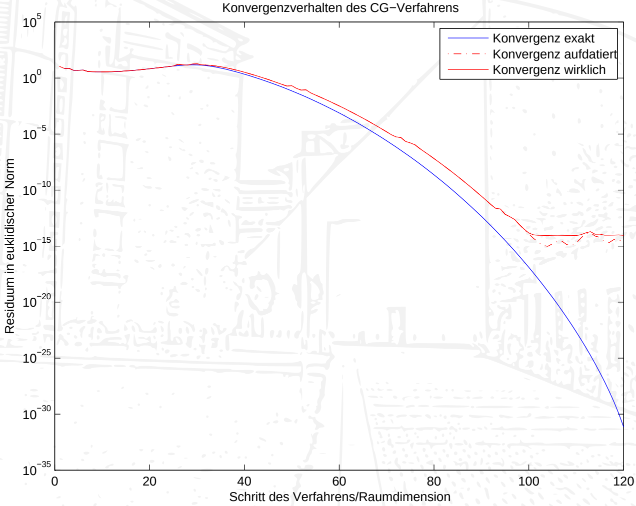
Eigenvalue convergence



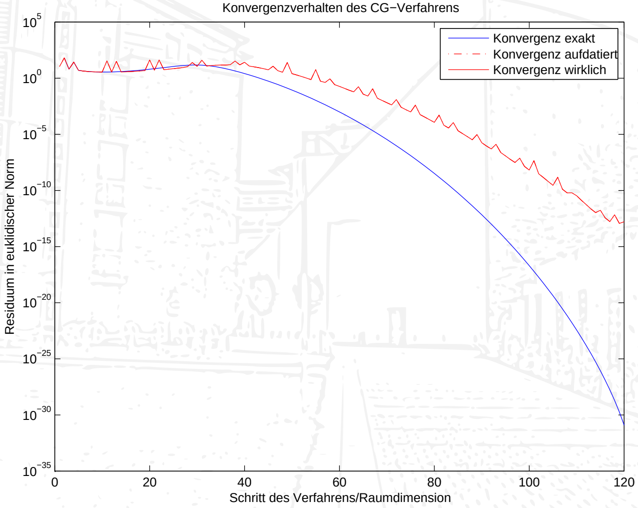
Eigenvalue convergence in finite precision



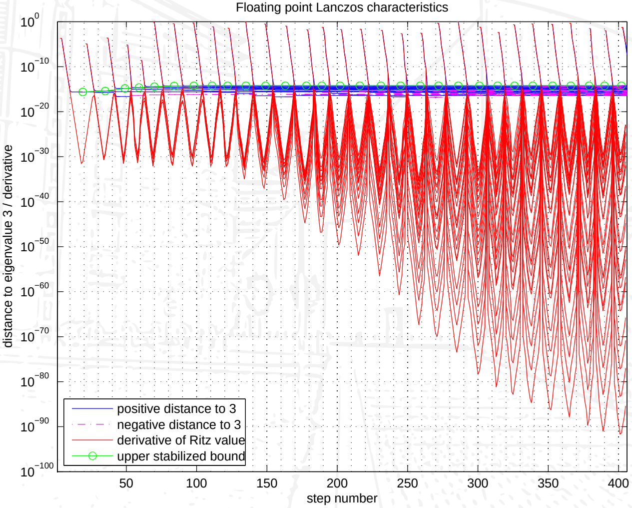
Convergence of CG, first example



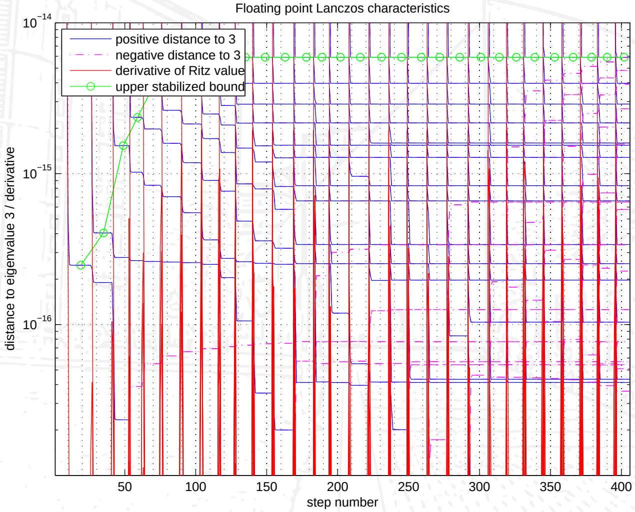
Convergence of CG, second example . . .



Characteristics of floating point Lanczos



Characteristics of floating point Lanczos; details



Part III

As an example we consider a deep link between Rayleigh Quotient Iteration (RQI) and the Opitz-Larkin Method (OLM).

We briefly sketch some recent developments in two fascinating areas:

- Progress in methods based on the principle of Induced Dimension Reduction (IDR), and the
- Augmented backward error analysis of Lanczos methods.

Original RQI

In the second edition of the first volume of his book "The Theory of Sound" (Strutt, 1894), John William Strutt, 3rd Baron Rayleigh, included on page 110 the following passage:

> The stationary property of the roots of Lagrange's determinant (3) § 84, suggests a general method of approximating to their values. Beginning with assumed rough approximations to the ratios $A_1 : A_2 : A_3 \ldots\ldots$ we may calculate a first approximation to p^2 from
>
> $$p^2 = \frac{\frac{1}{2} c_{11} A_1^2 + \frac{1}{2} c_{22} A_2^2 + \ldots + c_{12} A_1 A_2 + \ldots}{\frac{1}{2} a_{11} A_1^2 + \frac{1}{2} a_{22} A_2^2 + \ldots + a_{12} A_1 A_2 + \ldots} \quad \ldots\ldots (3).$$
>
> With this value of p^2 we may recalculate the ratios $A_1 : A_2 \ldots$ from any $(m-1)$ of equations (5) § 84, then again by application of (3) determine an improved value of p^2, and so on.]

Original RQI

In modern notation, Lord Rayleigh starts with an approximate eigenvector $\mathbf{v}_k$, $k = 0$, of a Hermitean matrix (Hermitean pencil), computes its Rayleigh quotient

$$\rho(\mathbf{v}_k) := \frac{\mathbf{v}_k^{\mathsf{H}} \mathbf{A} \mathbf{v}_k}{\mathbf{v}_k^{\mathsf{H}} \mathbf{v}_k},$$

and iterates for some suitably chosen $j \in \{1, 2, \ldots, n\}$,

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{e}_j}{\|(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{e}_j\|}, \quad k = 0, 1, \ldots$$

where j may vary, depending on the computed approximate eigenvector.

The Rayleigh quotient uniquely solves the least squares problem

$$\rho(\mathbf{v}_k) = \operatorname{argmin}_{\rho \in \mathbb{C}} \|\mathbf{A}\mathbf{v}_k - \mathbf{v}_k \rho\|.$$

Inverse Iteration

Closely connected to RQI is inverse iteration (Wielandt, 1944). In its most basic variant the shift τ is never updated, but the right-hand side is replaced by the latest approximate eigenvector:

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \tau \mathbf{I}_n)^{-1}\mathbf{v}_k}{\|(\mathbf{A} - \tau \mathbf{I}_n)^{-1}\mathbf{v}_k\|}, \quad k = 0, 1, \ldots$$

The shift can be updated by using the approximate eigenvalues obtained by the shift update strategy

$$\tau_{k+1} := \tau_k + \frac{1}{\mathbf{e}_j^\mathsf{T}(\mathbf{A} - \tau_k \mathbf{I}_n)^{-1}\mathbf{v}_k}.$$

The latter variant is described in (Wielandt, 1944, Seite 9, Formel (20)) and converges locally quadratically.

Modern variants of RQI

Combination gives (symmetric/Hermitean) RQI:

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k}{\|(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k\|}, \quad k = 0, 1, \ldots$$

This iteration is also used for nonsymmetric $\mathbf{A}$.

Crandall was the first who investigated the three variants (the original Rayleigh quotient iteration; inverse iteration with fixed shift; symmetric RQI), see (Crandall, 1951).

Ostrowski proved that unsymmetric RQI still has a quadratic convergence rate, (Ostrowski, 1959b). In (Ostrowski, 1959a), he devised two-sided RQI:

$$\rho(\mathbf{w}_k, \mathbf{v}_k) := \frac{\mathbf{w}_k^{\mathsf{H}}\mathbf{A}\mathbf{v}_k}{\mathbf{w}_k^{\mathsf{H}}\mathbf{v}_k}, \qquad \begin{aligned} \mathbf{v}_{k+1} &= (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k, \\ \mathbf{w}_{k+1} &= (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k)\mathbf{I}_n)^{-\mathsf{H}}\mathbf{w}_k, \end{aligned} \qquad k = 0, 1, \ldots$$

This trick recovers the cubic convergence rate of RQI at the expense of an additional system. Parlett's alternating RQI preserves monotonicity.

Classical methods

Methods for the computation of a root of a rational function

$$f : \mathbb{C} \to \mathbb{C}, \quad f(z) := \frac{p(z)}{q(z)}, \quad p, q \in \mathbb{P}_m$$

include Newton's method

$$z_{k+1} = z_k - \frac{f(z_k)}{f'(z_k)}$$

and the secant method:

$$z_{k+1} = z_k - \frac{f(z_k)}{[z_k, z_{k-1}]f}.$$

The secant method has R-order of convergence given by the golden ratio

$$\phi := \frac{1+\sqrt{5}}{2} \approx 1.618.$$

Two steps of the secant method are as costly as one step of Newton's method. This makes the secant method the winner:

$$\phi^2 = \phi + 1 \approx 2.618 > 2.$$

Schröder's and König's methods

Newton's method has been generalized to incorporate higher order derivatives and to exhibit a higher order of convergence. Well-known generalized Newton's methods are Halley's and Laguerre's methods.

In 1870 E. Schröder from Pforzheim came up with two infinite families of generalizations (Schröder, 1870). In 1884 Julius König proved a theorem on the limiting behavior of certain ratios of Taylor coefficients (König, 1884), enabling a simpler derivation of Schröder's family A_ω^λ with $\lambda = 0$.

This family is nowadays known as "König's method":

$$z_{k+1} = z_k + s\frac{(1/f)^{(s-1)}(z_k)}{(1/f)^{(s)}(z_k)}, \quad s = 1, 2, \ldots$$

König's method for $s = 1$ is Newton's method,

$$z_{k+1} = z_k + \frac{(1/f)(z_k)}{(1/f)'(z_k)} = z_k - \frac{1/f(z_k)}{f'(z_k)/(f(z_k))^2} = z_k - \frac{f(z_k)}{f'(z_k)}.$$

The Opitz-Larkin method

There is a natural extension of König's method using divided differences in place of the derivatives. This natural extension (without the connection to König's method) was published in 1958 by Günter Opitz in a two-page article in ZAMM.

He published few additional papers on the subject (including his most famous "Steigungsmatrizen" paper). A more complete presentation can be found in his "Habilitationsschrift". There, he even pointed out the connection to König's method.

Independently, 23 years later F. M. Larkin re-developed Opitz' method, see (Larkin, 1981) and the predecessor (Larkin, 1980).

We will refer to this method as the Opitz-Larkin method. The Opitz-Larkin method is based on iterations of the form

$$x_{k+1} = z_k + \frac{[z_1, z_2, \ldots, z_{k-1}](1/f)}{[z_1, z_2, \ldots, z_{k-1}, z_k](1/f)}.$$

The Opitz-Larkin method

Mostly, the z_i are all distinct and the next iterate is used as new evaluation point $z_{k+1} = x_{k+1}$,

$$z_{k+1} = z_k + \frac{[z_1, z_2, \dots, z_{k-1}](1/f)}{[z_1, z_2, \dots, z_{k-1}, z_k](1/f)}.$$

This variant of the Opitz-Larkin method converges with R-order 2.

Frequently, the Opitz-Larkin method is used with truncation:

$$z_{k+1} = z_k + \frac{[z_{k-p}, \dots, z_{k-1}](1/f)}{[z_{k-p}, \dots, z_{k-1}, z_k](1/f)},$$

see (Opitz, 1958, Seite 277, Gleichung (9)) and (Larkin, 1981, Section 4, pages 98–99).

The Opitz-Larkin method

It is possible to use confluent divided differences, i.e., multiple points of evaluation, i.e., higher order derivatives of $1/f$.

When we use only confluent divided differences in the truncated Opitz-Larkin method with truncation parameter $p = s$, we recover König's method:

$$z_{k+1} = z_k + \frac{[\overbrace{z_k, \dots, z_k}^{s}](1/f)}{[\underbrace{z_k, \dots, z_k, z_k}_{s+1}](1/f)}$$

$$= z_k + \frac{(1/f)^{(s-1)}(z_k)/(s-1)!}{(1/f)^{(s)}(z_k)/s!} = z_k + s\frac{(1/f)^{(s-1)}(z_k)}{(1/f)^{(s)}(z_k)}.$$

The Opitz-Larkin method

Truncated Opitz-Larkin with $p = 1$ is the secant method,

$$
\begin{aligned}
z_{k+1} &= z_k + \frac{[z_{k-1}](1/f)}{[z_{k-1}, z_k](1/f)} \\
&= z_k + \frac{1}{f(z_{k-1})} \cdot \frac{z_{k-1} - z_k}{1/f(z_{k-1}) - 1/f(z_k)} \\
&= z_k + \frac{f(z_k)f(z_{k-1})}{f(z_{k-1})} \cdot \frac{z_{k-1} - z_k}{f(z_k) - f(z_{k-1})} \\
&= z_k - \frac{f(z_k)}{[z_{k-1}, z_k]f}.
\end{aligned}
$$

Confluent truncated Opitz-Larkin with $p = 1$ is Newton's method.

The Opitz-Larkin method

In general, the Opitz-Larkin method is closely connected to rational interpolation of the inverse function (Larkin, 1981, Theorem 1, page 96):

Theorem (Larkin 1981)

If, for any integer $k > 1$, there exists a rational function of the form

$$r_k(z) = \frac{q_d(z)}{z - \alpha}, \quad \forall\, z,$$

where q_d is a polynomial of degree $d \leqslant k - 2$, such that $q_d(\alpha) \neq 0$ and

$$r_k(z_j) = f(z_j)^{-1}, \quad j = 1, 2, \ldots, k,$$

then

$$z_k + \frac{[z_1, z_2, \ldots, z_{k-1}](1/f)}{[z_1, z_2, \ldots, z_{k-1}, z_k](1/f)} = \alpha.$$

Simplification

We set ${}^{z}\mathbf{H}_n := (z\mathbf{I}_n - \mathbf{H}_n)$. By the first resolvent identity (Chatelin, 1993)

$$
\begin{aligned}
({}^{z_1}\mathbf{H}_n)^{-1}({}^{z_2}\mathbf{H}_n)^{-1} &= (z_1\mathbf{I}_n - \mathbf{H}_n)^{-1}(z_2\mathbf{I}_n - \mathbf{H}_n)^{-1} && \text{(11a)}\\
&= \frac{({}^{z_1}\mathbf{H}_n)^{-1} - ({}^{z_2}\mathbf{H}_n)^{-1}}{z_2 - z_1} = -[z_1, z_2]({}^{z}\mathbf{H}_n)^{-1}. && \text{(11b)}
\end{aligned}
$$

The first resolvent identity is based on the trivial observation that

$$(z_2\mathbf{I}_n - \mathbf{H}_n) - (z_1\mathbf{I}_n - \mathbf{H}_n) = (z_2 - z_1)\mathbf{I}_n.$$

Generalization (see also (Dekker and Traub, 1971)):

$$\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1} = (-1)^{k-1}[z_1, \ldots, z_k]({}^{z}\mathbf{H}_n)^{-1}. \tag{12}$$

Confluent divided differences are well-defined.

Simplification

For simplicity we assume that $\mathbf{H}_n$ is unreduced. We denote products of sub-diagonal elements of the unreduced Hessenberg matrices $\mathbf{H}_n \in \mathbb{C}^{n\times n}$ by

$$h_{i:j} := \prod_{\ell=i}^{j} h_{\ell+1,\ell}.$$

Polynomial vectors $\boldsymbol{\nu}$ and $\check{\boldsymbol{\nu}}$ are defined by

$$\boldsymbol{\nu}(z) := \left(\frac{\chi_{j+1:n}(z)}{h_{j:n-1}}\right)_{j=1}^{n} \quad \text{and} \quad \check{\boldsymbol{\nu}}(z) := \left(\frac{\chi_{1:j-1}(z)}{h_{1:j-1}}\right)_{j=1}^{n}. \tag{13}$$

The elements are $\nu_j(z)$ and $\check{\nu}_j(z)$, $j = 1, \ldots, n$. Observe that $\nu_n \equiv 1 \equiv \check{\nu}_1$.

The polynomials $\chi_{i:j}$ are the characteristic polynomials of submatrices of $\mathbf{H}_n$,

$$\chi_{i:j}(z) := \det\left({}^{z}\mathbf{H}_{i:j}\right) = \det\left(z\mathbf{I}_{j-i+1} - \mathbf{H}_{i:j}\right).$$

Simplification

For z in the resolvent set

$$({}^{z}\mathbf{H}_n)\boldsymbol{\nu}(z) = \frac{\chi(z)}{h_{1:n-1}}\mathbf{e}_1 \quad \Leftrightarrow \quad \frac{\boldsymbol{\nu}(z)h_{1:n-1}}{\chi(z)} = ({}^{z}\mathbf{H}_n)^{-1}\mathbf{e}_1, \tag{14a}$$

$$\check{\boldsymbol{\nu}}(z)^{\mathsf{T}}({}^{z}\mathbf{H}_n) = \mathbf{e}_n^{\mathsf{T}}\frac{\chi(z)}{h_{1:n-1}} \quad \Leftrightarrow \quad \frac{h_{1:n-1}\check{\boldsymbol{\nu}}(z)^{\mathsf{T}}}{\chi(z)} = \mathbf{e}_n^{\mathsf{T}}({}^{z}\mathbf{H}_n)^{-1}. \tag{14b}$$

The repeated application of resolvents to $\mathbf{e}_1$ results in

$$\left(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\right)\mathbf{e}_1 = (-1)^{k-1}[z_1,\ldots,z_k]({}^{z}\mathbf{H}_n)^{-1}\mathbf{e}_1 \tag{15}$$

$$= (-1)^{k-1}[z_1,\ldots,z_k]\frac{\boldsymbol{\nu}(z)h_{1:n-1}}{\chi(z)}. \tag{16}$$

Note that $z\mathbf{I}_n - {}^{z}\mathbf{H}_n = z\mathbf{I}_n - (z\mathbf{I}_n - \mathbf{H}_n) = \mathbf{H}_n$, i.e., $\mathbf{H}_n({}^{z}\mathbf{H}_n)^{-1} = z({}^{z}\mathbf{H}_n)^{-1} - \mathbf{I}_n$.

Simplification

For the sake of eased understanding, we look at inverse iteration with a two-sided Rayleigh quotient where the left vector is the last standard unit vector $\mathbf{e}_n^\mathsf{T}$. For this method we have the iterates

$$\mathbf{v}_{k+1} = \Bigl(\prod_{i=1}^{k} ({}^{z_i}\mathbf{H}_n)^{-1}\Bigr)\mathbf{e}_1, \quad x_{k+1} = \frac{\mathbf{e}_n^\mathsf{T}\mathbf{H}_n\mathbf{v}_{k+1}}{\mathbf{e}_n^\mathsf{T}\mathbf{v}_{k+1}},$$

and thus the approximate eigenvalues are given by the Opitz-Larkin method:

$$x_{k+1} = \frac{\mathbf{e}_n^\mathsf{T}\mathbf{H}_n\bigl(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\bigr)\mathbf{e}_1}{\mathbf{e}_n^\mathsf{T}\bigl(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\bigr)\mathbf{e}_1} = \frac{\mathbf{e}_n^\mathsf{T}(z_k\mathbf{I}_n - ({}^{z_k}\mathbf{H}_n))\bigl(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\bigr)\mathbf{e}_1}{\mathbf{e}_n^\mathsf{T}\bigl(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\bigr)\mathbf{e}_1} \tag{17a}$$

$$= z_k - \frac{\mathbf{e}_n^\mathsf{T}\,{}^{z_k}\mathbf{H}_n\bigl(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\bigr)\mathbf{e}_1}{\mathbf{e}_n^\mathsf{T}\bigl(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\bigr)\mathbf{e}_1} = z_k - \frac{\mathbf{e}_n^\mathsf{T}\bigl(\prod_{i=1}^{k-1}({}^{z_i}\mathbf{H}_n)^{-1}\bigr)\mathbf{e}_1}{\mathbf{e}_n^\mathsf{T}\bigl(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\bigr)\mathbf{e}_1} \tag{17b}$$

$$= z_k + \frac{[z_1, \ldots, z_{k-1}](1/\chi)}{[z_1, \ldots, z_{k-1}, z_k](1/\chi)}. \tag{17c}$$

Simplification

When we update the shifts by choosing $z_{k+1} = x_{k+1}$ we obtain the standard variant of the Opitz-Larkin method. This method has asymptotically second order convergence against the roots of the characteristic polynomial χ.

Inverse iteration with fixed shift $\tau = z_1 = z_2 = \ldots = z_k$ results in the recurrence

$$x_{k+1} = \tau + \frac{[\tau, \ldots, \tau](1/\chi)}{[\tau, \ldots, \tau, \tau](1/\chi)} = \tau + k \frac{(1/\chi)^{(k-1)}(\tau)}{(1/\chi)^{(k)}(\tau)}. \quad (18)$$

Inverse iteration with fixed shift performs one step of König's method. Restarting inverse iteration every s steps with updated shift given by the current eigenvalue approximation converges with order s (divided by steps: linearly).

Symmetric RQI is very pleasant to analyze, likely-wise is two-sided RQI, but unsymmetric RQI (and thus, the QR algorithm) and alternating RQI do not fit into the picture.

Simplification

The original Rayleigh quotient iteration (Strutt, 1894) with the symmetric Rayleigh quotient and, because of the symmetry, a tridiagonal Hermitean Hessenberg matrix $\mathbf{H}_n$, gives the update

$$z_{k+1} = \frac{\mathbf{e}_1^\mathsf{T} ({}^{z_k}\mathbf{H}_n)^{-\mathsf{H}} \mathbf{H}_n ({}^{z_k}\mathbf{H}_n)^{-1} \mathbf{e}_1}{\mathbf{e}_1^\mathsf{T} ({}^{z_k}\mathbf{H}_n)^{-\mathsf{H}} ({}^{z_k}\mathbf{H}_n)^{-1} \mathbf{e}_1} = \frac{\mathbf{e}_1^\mathsf{T} \mathbf{H}_n ({}^{z_k}\mathbf{H}_n)^{-2} \mathbf{e}_1}{\mathbf{e}_1^\mathsf{T} ({}^{z_k}\mathbf{H}_n)^{-2} \mathbf{e}_1} \tag{19a}$$

$$= \frac{\mathbf{e}_1^\mathsf{T} (z_k \mathbf{I}_n - {}^{z_k}\mathbf{H}_n) ({}^{z_k}\mathbf{H}_n)^{-2} \mathbf{e}_1}{\mathbf{e}_1^\mathsf{T} ({}^{z_k}\mathbf{H}_n)^{-2} \mathbf{e}_1} \tag{19b}$$

$$= z_k - \frac{\mathbf{e}_1^\mathsf{T} ({}^{z_k}\mathbf{H}_n)^{-1} \mathbf{e}_1}{\mathbf{e}_1^\mathsf{T} ({}^{z_k}\mathbf{H}_n)^{-2} \mathbf{e}_1} = z_k + \frac{[z_k](\chi_{2:n}/\chi)}{[z_k, z_k](\chi_{2:n}/\chi)} \tag{19c}$$

$$= z_k - \frac{r(z_k)}{r'(z_k)}, \qquad r(z) := \frac{\chi(z)}{\chi_{2:n}(z)}. \tag{19d}$$

This is Newton's method on the meromorphic function r. As the poles of this meromorphic function are the eigenvalues of a submatrix, they interlace by Cauchy's interlace theorem the roots, which are the eigenvalues.

Simplification

Symmetric RQI for Hermitean matrices gives the update

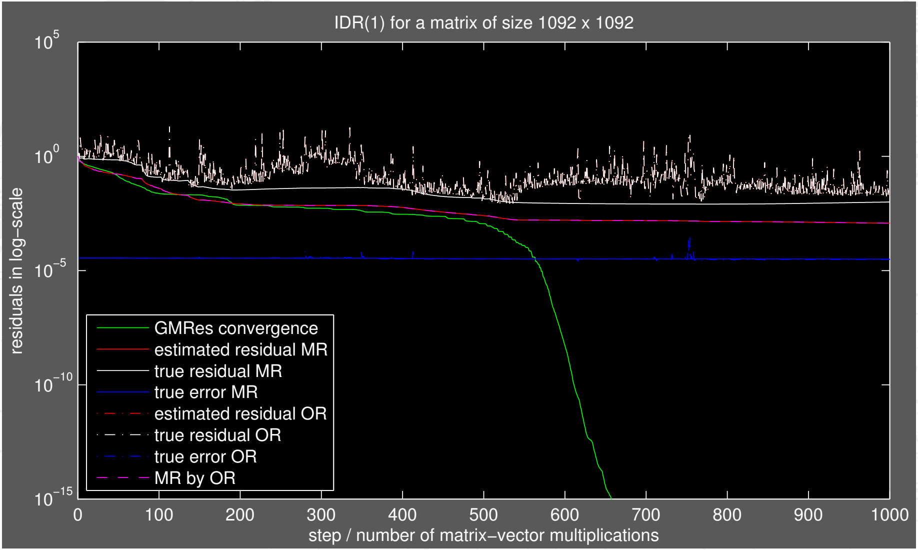
$$z_{k+1} = z_k + \frac{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k](\chi_{2:n}/\chi)}{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k, z_k](\chi_{2:n}/\chi)}. \tag{20}$$

This update has by a result of Tornheim asymptotically a cubic convergence rate. We have to compute the limit of the real root of the equations

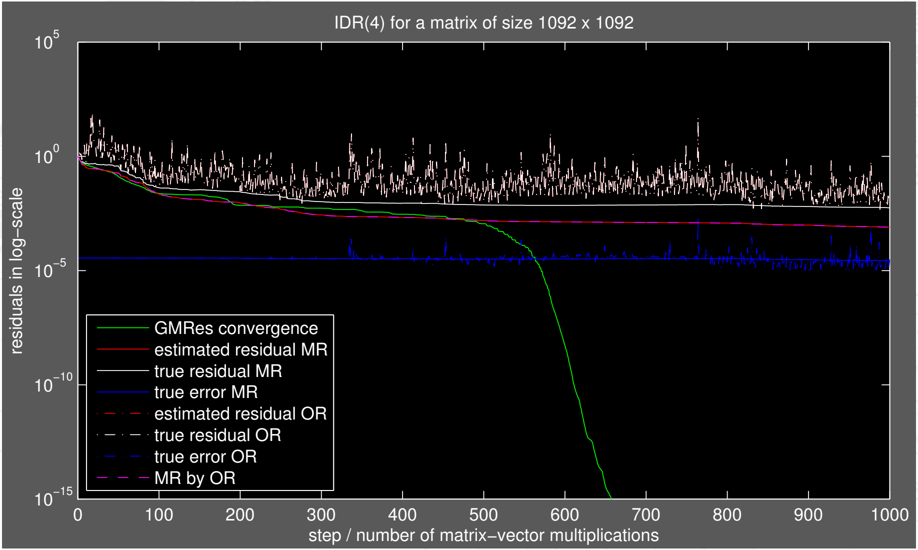
$$x^k - 2x^{k-1} - 2x^{k-2} - \cdots - 2 = 0, \quad k = 1, \ldots$$

This is the maximal eigenvalue of a Hessenberg matrix with one in the lower diagonal and two in the last column. The approximate eigenvector of all ones to the approximate eigenvalue 3 gives the backward error $1/\sqrt{k}$ and the only positive real eigenvalue of the matrix is well separated, the other eigenvalues lie close to a circle of radius one around zero.

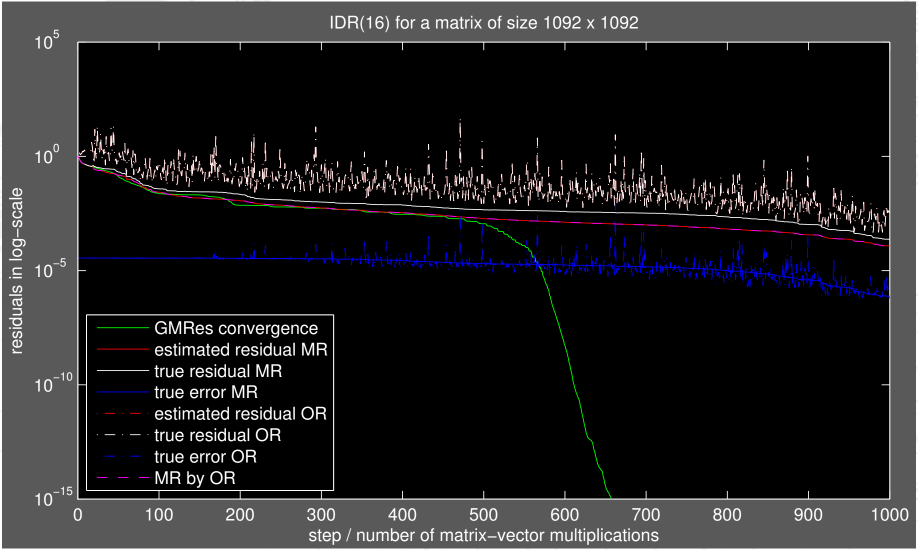
Load applied to structure, $\mathbf{K} \in \mathbb{R}^{1092\times 1092}$, IDR(1)



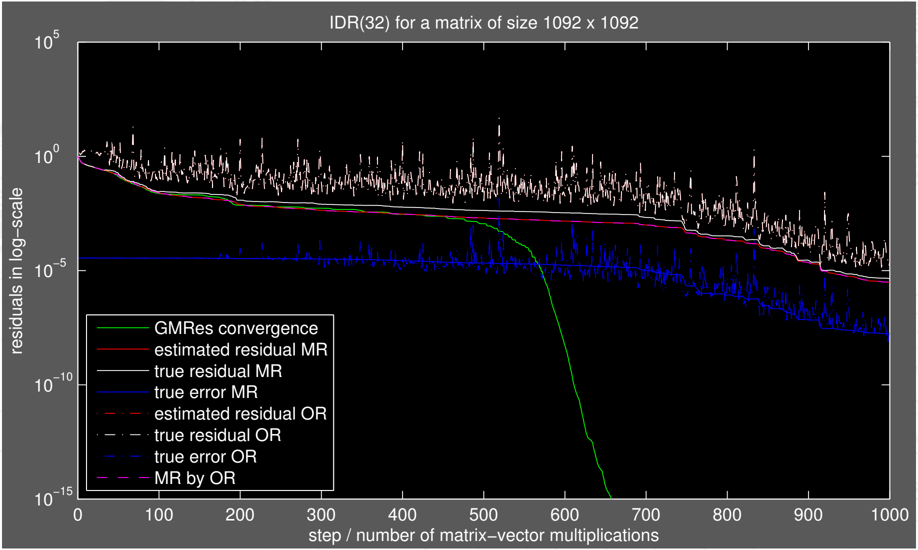
Load applied to structure, $\mathbf{K} \in \mathbb{R}^{1092 \times 1092}$, IDR(4)



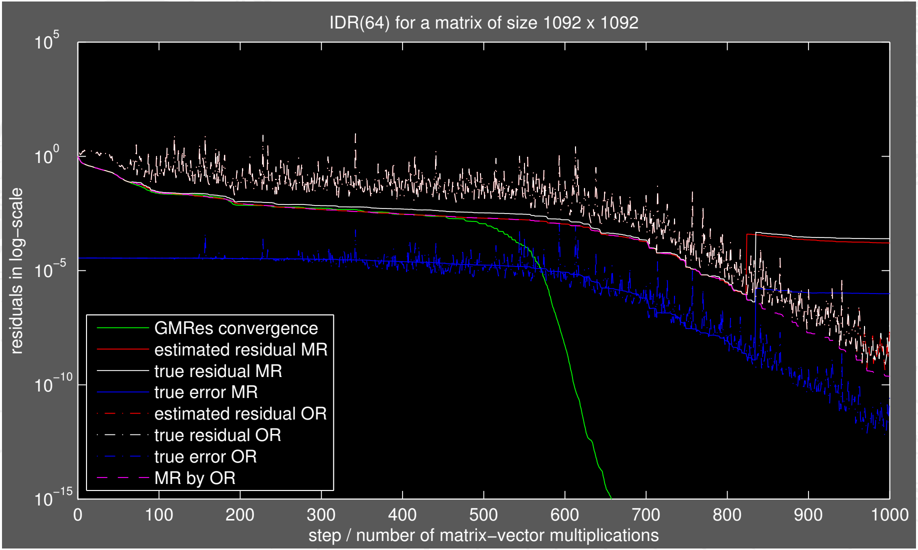
Load applied to structure, $\mathbf{K} \in \mathbb{R}^{1092 \times 1092}$, IDR(16)



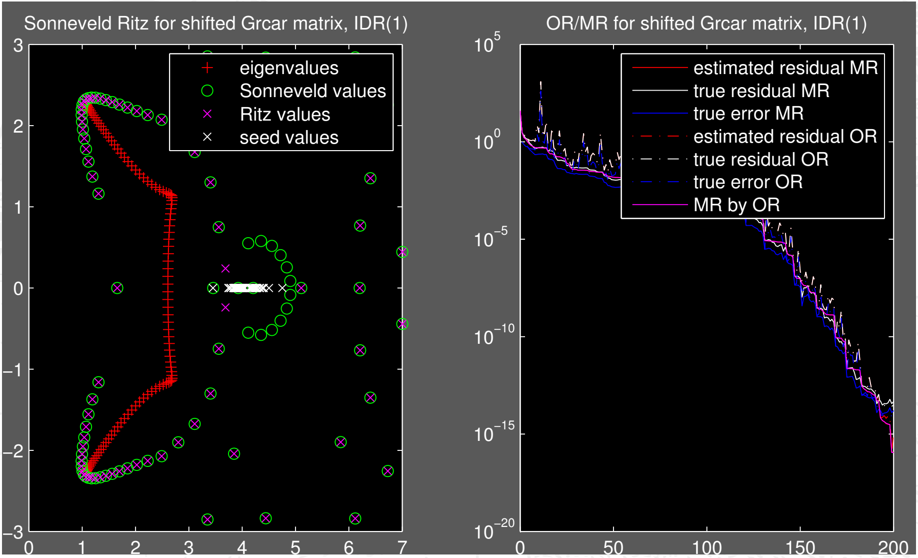
Load applied to structure, $\mathbf{K} \in \mathbb{R}^{1092\times 1092}$, IDR(32)



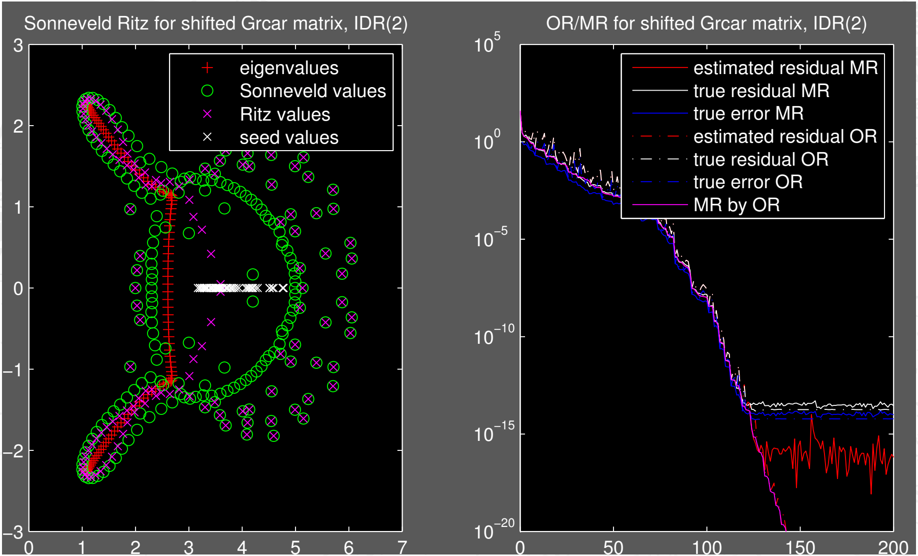
Load applied to structure, $\mathbf{K} \in \mathbb{R}^{1092 \times 1092}$, IDR(64)



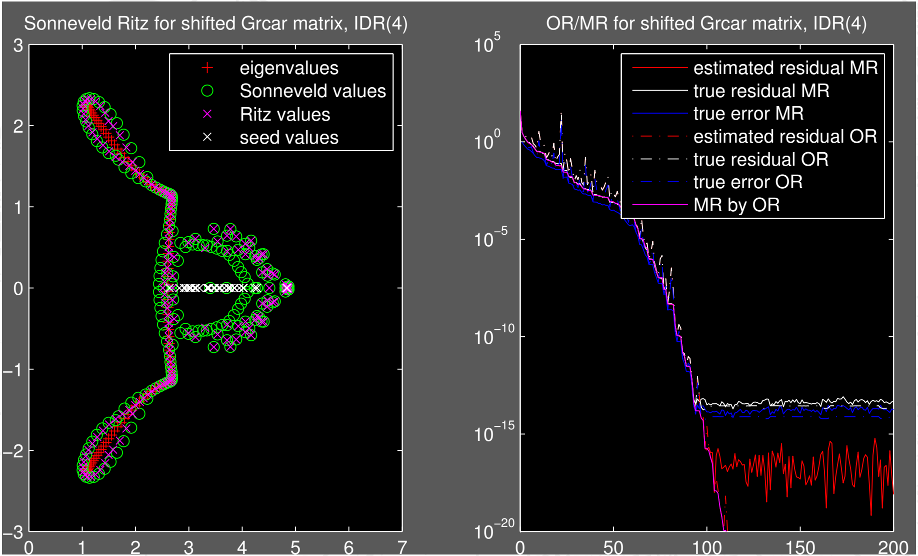
Shifted Grcar matrix; IDR(1)



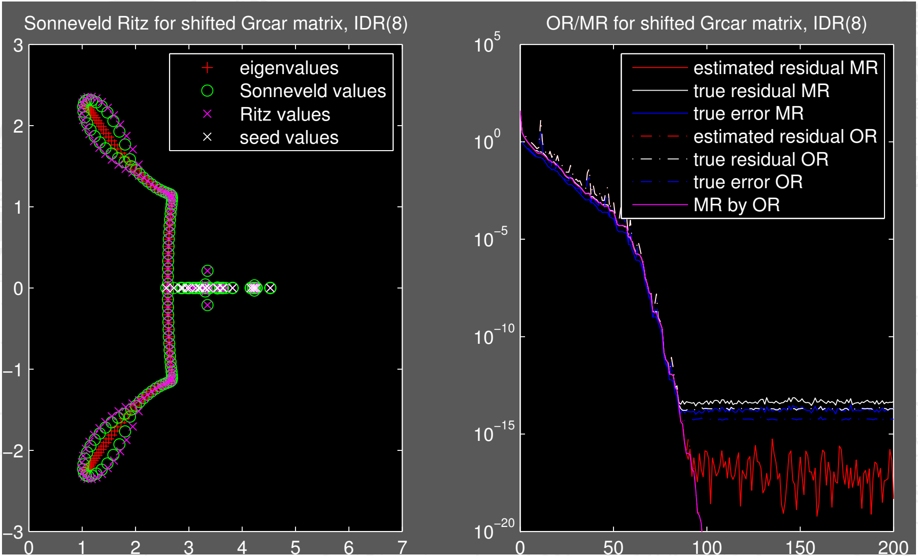
Shifted Grcar matrix; IDR(2)



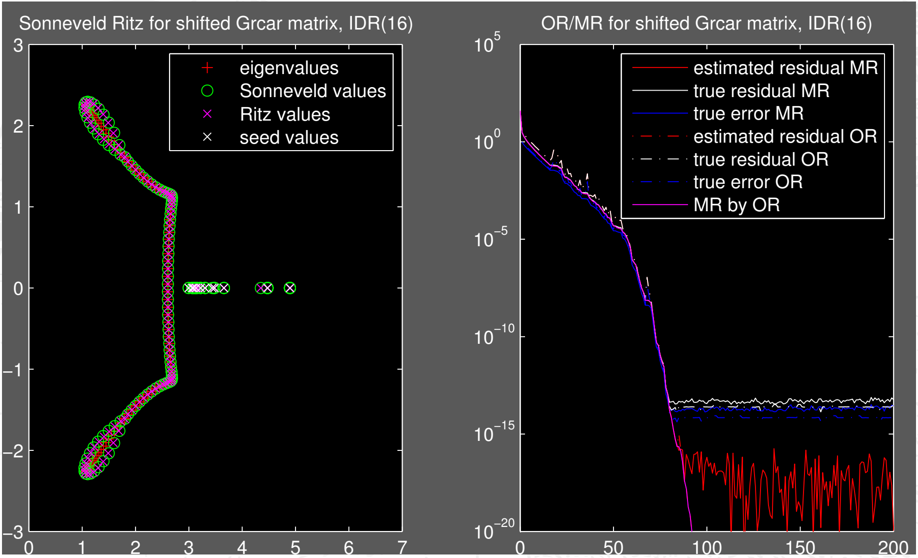
Shifted Grcar matrix; IDR(4)



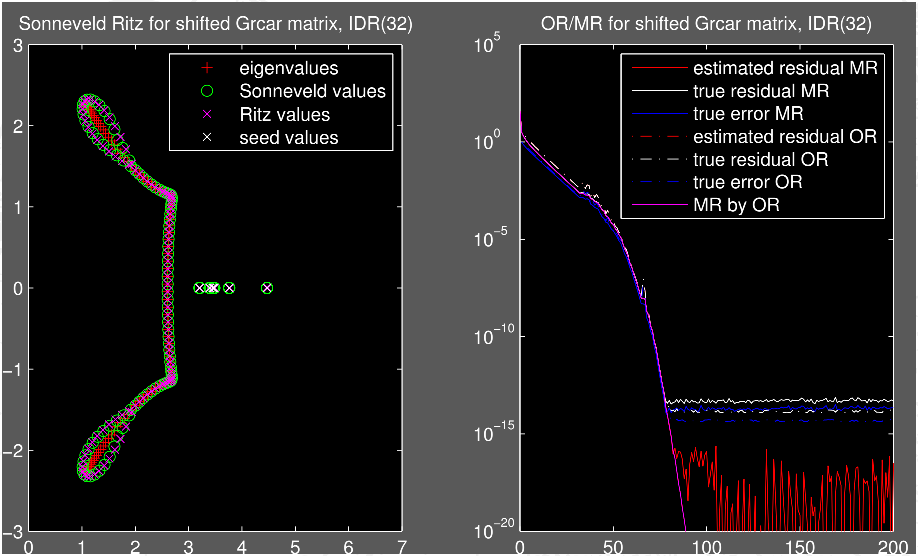
Shifted Grcar matrix; IDR(8)



Shifted Grcar matrix; IDR(16)



Shifted Grcar matrix; IDR(32)



Behaviour of perturbed Krylov subspace methods

Every observed behaviour that occurs in a perturbed method can also be observed in unperturbed methods w/ orthonormal basis vectors.

Hessenberg decomposition:

$$\mathbf{H}_n \mathbf{I}_{n,k} = \mathbf{I}_{n,k+1} \underline{\mathbf{H}}_k.$$

Generalized Hessenberg decomposition:

$$(\mathbf{H}_n \mathbf{U}_n^{-1}) \mathbf{I}_{n,k} \mathbf{U}_k = \mathbf{I}_{n,k+1} \underline{\mathbf{H}}_k.$$

Bad news: Impossible to distinguish effects of perturbation from startling behaviour due to strange data.

Analysis of perturbed Krylov subspace methods

Suppose that

$$\mathbf{A}\mathbf{Q}_k + \mathbf{F}_k = \mathbf{Q}_{k+1}\underline{\mathbf{T}}_k, \quad \mathbf{A}^{\mathsf{H}} = \mathbf{A}, \quad \mathbf{T}_k^{\mathsf{H}} = \mathbf{T}_k.$$

Set

$$\operatorname{diag}(\mathbf{T}_k, \mathbf{A}) := \begin{pmatrix} \mathbf{T}_k & \mathbf{O}_{k,n} \\ \mathbf{O}_{n,k} & \mathbf{A} \end{pmatrix} \in \mathbb{C}^{(k+n)\times(k+n)}, \quad \mathbf{T}_k \in \mathbb{C}^{k\times k}, \quad \mathbf{A} \in \mathbb{C}^{n\times n}.$$

Paige used augmented backward error analysis for symmetric Lanczos in finite precision:

$$(\operatorname{diag}(\mathbf{T}_k, \mathbf{A}) + \mathbf{H})\,\widetilde{\mathbf{Q}}_k = \widetilde{\mathbf{Q}}_{k+1}\underline{\mathbf{T}}_k, \quad \widetilde{\mathbf{Q}}_k^{\mathsf{H}}\widetilde{\mathbf{Q}}_k = \mathbf{I}_k.$$

Here, $\mathbf{H}$ is a “small” perturbation if $\mathbf{F}_k$ is small and local orthonormality is given. Error-free process for perturbed strange matrix.

Extended to two-sided Lanczos by Paige, Panayotov and Z., 2012.

Conclusion and Outlook

- I sketched the three main families of Krylov subspace methods.
- I highlighted the rôle of Hessenberg matrices and the resulting structure.
- The relations to interpolation and approximation have been stated.
- Convergence analysis is split into convergence of vectorial quantities and convergence of (harmonic) Ritz values.
- I gave some insight into some deep link to classical root-finding and presented some current developments.
- I (hopefully) convinced you that finite-dimensional aspects are still quite complicated in nature, but very interesting, and gave some hints, which Krylov subspace methods you could use in your application.

Thank you very much for attending our Kickoff meeting!

This talk is partially based on the following technical reports:

Eigenvalue computations based on IDR, Martin H. Gutknecht and Z., Bericht 145, Institut für Numerische Simulation, TUHH, 2010,

Flexible and multi-shift induced dimension reduction algorithms for solving large sparse linear systems, Martin B. van Gijzen, Gerard L.G. Sleijpen, and Z., Bericht 156, Institut für Numerische Simulation, TUHH, 2011,

IDR: A new generation of Krylov subspace methods?, Olaf Rendel, Anisa Rizvanolli, and Z., Bericht 161, Institut für Mathematik, TUHH, 2012.

Chatelin, F. (1993).
Eigenvalues of matrices.
John Wiley & Sons Ltd., Chichester.
With exercises by Mario Ahués and the author, Translated from the French and with additional material by Walter Ledermann.

Crandall, S. H. (1951).
Iterative procedures related to relaxation methods for eigenvalue problems.
Proceedings of the Royal Society. London. Series A. Mathematical, Physical and Engineering Sciences, 207:416–423.

Dekker, T. J. and Traub, J. F. (1971).
The shifted QR algorithm for Hermitian matrices.
Linear Algebra and Appl., 4:137–154.

König, J. (1884).
Ueber eine Eigenschaft der Potenzreihen.
Mathematische Annalen, 23:447–449.
Scanned article online available at the Göttinger Digitalisierungszentrum (GDZ): http://resolver.sub.uni-goettingen.de/purl?GDZPPN002248026.

Larkin, F. M. (1980).
Root-finding by fitting rational functions.
Mathematics of Computation, 35(151):803–816.

Larkin, F. M. (1981).
Root finding by divided differences.
Numer. Math., 37(1):93–104.

Opitz, G. (1958).
Gleichungsauflösung mittels einer speziellen Interpolation.
Z. Angew. Math. Mech., 38:276–277.

Ostrowski, A. M. (1959a).
On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. III. (Generalized Rayleigh quotient and characteristic roots with linear elementary divisors).
Arch. Rational Mech. Anal., 3:325–340.

Ostrowski, A. M. (1959b).
On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. V. Usual Rayleigh quotient for non-Hermitian matrices and linear elementary divisors.
Arch. Rational Mech. Anal., 3:472–481.

Schröder, E. (1870).
Ueber unendlich viele Algorithmen zur Auflösung der Gleichungen.
Math. Ann., 2(2):317–365.

Schweins, F. (1825).
Theorie der Differenzen und Differentiale.
Verlag der Universitäts-Buchhandlung von C. F. Winter, Heidelberg.
Digitized by Google, online available at
`http://books.google.de/books?id=dntNAAAAMAAJ`.

Stiefel, E. (1955).
Relaxationsmethoden bester Strategie zur Lösung linearer Gleichungssysteme.
Comment. Math. Helv., 29:157–179.
Digitized by GDZ, online available at http://www-gdz.sub.uni-goettingen.de/cgi-bin/digbib.cgi?PPN358147735_0029.

Strutt, J. W. (1894).
The Theory of Sound, volume one.
Macmillan, London, second edition.

Wielandt, H. (1944).
Beiträge zur mathematischen Behandlung komplexer Eigenwertprobleme, Teil V: Bestimmung höherer Eigenwerte durch gebrochene Iteration.
Bericht B44/J/37, Aerodynamische Versuchsanstalt Göttingen.

Zemke, J.-P. M. (2006).
Hessenberg eigenvalue–eigenmatrix relations.
Linear Algebra Appl., 414(2-3):589–606.

Zemke, J.-P. M. (2007).
Abstract perturbed Krylov methods.
Linear Algebra Appl., 424(2-3):405–434.