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
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
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
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).
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
Basics

Krylov subspaces:

\[ \mathcal{K}_k := \mathcal{K}_k(A, q) := \text{span} \{ q, Aq, A^2q, \ldots, A^{k-1}q \} = \{ p_{k-1}(A)q \mid p_{k-1} \in \Pi_{k-1} \} \]

spanned by columns of Krylov matrix

\[ K_k := K_k(A, q) := (q, Aq, A^2q, \ldots, A^{k-1}q). \]
Krylov subspaces:

\[ K_k := K_k(A, q) := \text{span} \{ q, Aq, A^2q, \ldots, A^{k-1}q \} = \{ p_{k-1}(A)q \mid p_{k-1} \in \Pi_{k-1} \} \]

spanned by columns of Krylov matrix

\[ K_k := K_k(A, q) := (q, Aq, A^2q, \ldots, A^{k-1}q). \]

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(\( \ell \)), IDR(\( s \)), IDR(\( s \))STAB(\( \ell \));
short recurrence, transpose, \{unstable,cheap\}—\{stable,costly\}
Krylov subspaces:

\[ \mathcal{K}_k := \mathcal{K}_k(A, q) := \text{span}\{q, Aq, A^2q, \ldots, A^{k-1}q\} = \{p_{k-1}(A)q \mid p_{k-1} \in \Pi_{k-1}\} \]

spanned by columns of Krylov matrix

\[ K_k := K_k(A, q) := \{q, Aq, A^2q, \ldots, A^{k-1}q\}. \]

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(\(\ell\)), IDR(\(s\)), IDR(\(s\))STAB(\(\ell\)); short recurrence, transpose, \{unstable,cheap\}—\{stable,costly\}

We subsume Hessenberg and Arnoldi as “Arnoldi-based”.
Hessenberg decompositions

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

$$AQ_k = Q_{k+1}H_k.$$  
(Lanczos: $H_k = T_k$, $2\times$)
Classification of Krylov subspace methods

Krylov/Hessenberg

Hessenberg decompositions

Arnoldi- and Lanczos-based methods \(\rightsquigarrow\) Hessenberg decomposition:

\[
A Q_k = Q_{k+1} H_k. \quad (\text{Lanczos: } H_k = T_k, 2\times)
\]

Sonneveld-based methods \(\rightsquigarrow\) generalized Hessenberg decomposition:

\[
A V_k = A G_k U_k = G_{k+1} H_k, \quad V_k := G_k U_k
\]
Hessenberg decompositions

Arnoldi- and Lanczos-based methods \(\leadsto\) Hessenberg decomposition:

\[
AQ_k = Q_{k+1}H_k. \quad (\text{Lanczos: } H_k = T_k, \ 2\times)
\]

Sonneveld-based methods \(\leadsto\) generalized Hessenberg decomposition:

\[
AV_k = AG_kU_k = G_{k+1}H_k, \quad V_k := G_kU_k.
\]

Three remarks:

▶ Structure: \(H_k \in \mathbb{C}^{(k+1)\times k}\) always unreduced extended Hessenberg;
▶ Generalization: \(I_k \leadsto U_k \in \mathbb{C}^{k\times k}\) upper triangular;
▶ Mnemonic for names of matrices in Sonneveld-based methods: IDR\((s)\)-coauthor “van Gijzen” \(\leadsto\) first \(V_k\), then \(G_k\).
Hessenberg decompositions

Arnoldi- and Lanczos-based methods \(\leadsto\) Hessenberg decomposition:

\[
AQ_k + F_k = Q_{k+1}H_k. \quad \text{(Lanczos: } H_k = T_k, \ 2\times)\]

Sonneveld-based methods \(\leadsto\) generalized Hessenberg decomposition:

\[
AV_k + \hat{F}_k = AG_kU_k + F_k = G_{k+1}H_k, \quad V_k := G_kU_k + \tilde{F}_k.
\]

Three remarks:

- **Structure**: \(H_k \in \mathbb{C}^{(k+1)\times k}\) always unreduced extended Hessenberg;
- **Generalization**: \(I_k \leadsto U_k \in \mathbb{C}^{k\times k}\) upper triangular;
- **Mnemonic for names of matrices in Sonneveld-based methods**: IDR(\(s\))-coauthor “van Gijzen” \(\leadsto\) first \(V_k\), then \(G_k\).

Finite precision or inexact method \(\leadsto\) perturbations \(F_k, F_k = \hat{F}_k + A\tilde{F}_k\).
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:

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

Karl Hessenberg (* September 8th, 1904, † February 22nd, 1959)
Residuals of OR and MR approximation \((Q_k e_1 \|r_0\| = Q_{k+1} e_1 \|r_0\| = r_0)\)

\[ x_k := Q_k z_k \quad \text{and} \quad x_k := Q_k z_k \]

with coefficient vectors

\[ z_k := H_k^{-1} e_1 \|r_0\| \quad \text{and} \quad z_k := H_k^\dagger e_1 \|r_0\| \]
Residuals of OR and MR approximation ($Q_k e_1 \|r_0\| = Q_{k+1} e_1 \|r_0\| = r_0$)

$x_k := Q_k z_k$ and $x_k := Q_k z_k$

with coefficient vectors

$z_k := H^{-1}_k e_1 \|r_0\|$ and $z_k := H^\dagger_k e_1 \|r_0\|

satisfy

$r_k := r_0 - Ax_k = R_k(A)r_0$ and $r_k := r_0 - Ax_k = R_k(A)r_0$. 

Residual polynomials $R_k$, $R_k$ given by

$R_k(z) := \det(I_k - z H^{-1}_k I_k)$ and $R_k(z) := \det(I_k - z H^\dagger_k I_k)$.

Convergence of OR and MR depends on (harmonic) Ritz values.
OR and MR for linear systems \((Ax = r_0 = b - Ax_0)\)

Residuals of OR and MR approximation \((Q_k e_1 \|r_0\| = Q_{k+1} e_1 \|r_0\| = r_0)\)

\[x_k := Q_k z_k \quad \text{and} \quad x_k := Q_k z_k\]

with coefficient vectors

\[z_k := H_k^{-1} e_1 \|r_0\| \quad \text{and} \quad z_k := H_k^\dagger e_1 \|r_0\|\]

satisfy

\[r_k := r_0 - Ax_k = R_k(A)r_0 \quad \text{and} \quad r_k := r_0 - Ax_k = \overline{R}_k(A)r_0.\]

Residual polynomials \(R_k, \overline{R}_k\) given by

\[R_k(z) := \det(I_k - zH_k^{-1}I_k) \quad \text{and} \quad \overline{R}_k(z) := \det(I_k - zH_k^\dagger I_k).\]
OR and MR for linear systems \((Ax = r_0 = b - Ax_0)\)

Residuals of OR and MR approximation \((Q_k e_1 \|r_0\| = Q_{k+1} e_1 \|r_0\| = r_0)\)

\[x_k := Q_k z_k\] and \[x_k := Q_k z_k\]

with coefficient vectors

\[z_k := H_k^{-1} e_1 \|r_0\|\] and \[z_k := H_k^\dagger e_1 \|r_0\|\]

satisfy

\[r_k := r_0 - Ax_k = R_k(A)r_0\] and \[r_k := r_0 - Ax_k = \underline{R}_k(A)r_0.\]

Residual polynomials \(R_k, \underline{R}_k\) given by

\[R_k(z) := \det (I_k - zH_k^{-1}I_k)\] and \[\underline{R}_k(z) := \det (I_k - zH_k^\dagger I_k).\]

Convergence of OR and MR depends on (harmonic) Ritz values.
Arnoldi/GMRes for a matrix of size 100 x 100

- Estimated residual MR
- True residual MR
- True error MR
- Estimated residual OR
- True residual OR
- True error OR
- MR by OR

residuals in log-scale

step / number of matrix-vector multiplications
Well known: Ritz pairs $\rightsquigarrow$ OR eigenpairs $(\theta_j, y_j)$,

\[ y_j := Q_k s_j, \quad \text{where} \quad H_k s_j = s_j \theta_j, \quad 1 \leq j \leq k. \]
OR and MR for eigenpairs

Well known: Ritz pairs $\leadsto$ OR eigenpairs $(\theta_j, y_j)$,

$$y_j := Q_k s_j, \quad \text{where} \quad H_k s_j = s_j \theta_j, \quad 1 \leq j \leq k.$$

Known: (shifted) harmonic Ritz pairs $(\theta_j, y_j)$,

$$y_j := Q_k s_j, \quad \text{where} \quad I_k s_j = (H_k - \tau I_k) I_k s_j (\theta_j - \tau), \quad 1 \leq j \leq k.$$
OR and MR for eigenpairs

Well known: Ritz pairs $\sim$ OR eigenpairs $(\theta_j, y_j)$,

$$ y_j := Q_k s_j, \text{ where } H_k s_j = s_j \theta_j, \quad 1 \leq j \leq k. $$

Known: (shifted) harmonic Ritz pairs $(\theta_j, y_j)$,

$$ y_j := Q_k s_j, \text{ where } I_k s_j = (H_k - \tau I_k)^\dagger I_k s_j (\theta_j - \tau), \quad 1 \leq j \leq k. $$

Less known: $\rho$-values, refined extraction, combinations thereof.
Well known: Ritz pairs $\leadsto$ OR eigenpairs $(\theta_j, y_j)$,

\[ y_j := Q_k s_j, \quad \text{where} \quad H_k s_j = s_j \theta_j, \quad 1 \leq j \leq k. \]

Known: (shifted) harmonic Ritz pairs $(\theta_j, y_j)$,

\[ y_j := Q_k s_j, \quad \text{where} \quad I_k s_j = (H_k - \tau I_k)^\dagger I_k s_j (\theta_j - \tau), \quad 1 \leq j \leq k. \]

Less known: $\rho$-values, refined extraction, combinations thereof.

Mostly unknown: MR eigenpairs $(\hat{\theta}, \hat{y} = Q_k \hat{s})$,

\[ \frac{\| (\hat{\theta} I_k - H_k) \hat{s} \|}{\| \hat{s} \|} := \min_{z \in \mathbb{C}, s \in \mathbb{C}^k, \| s \| = 1} \frac{\| (z I_k - H_k) s \|}{\| s \|}, \]
OR and MR for eigenpairs

Well known: Ritz pairs \( \sim \) OR eigenpairs \((\theta_j, y_j)\),

\[
y_j := Q_k s_j, \quad \text{where} \quad H_k s_j = s_j \theta_j, \quad 1 \leq j \leq k.
\]

Known: (shifted) harmonic Ritz pairs \((\theta_j, y_j)\),

\[
y_j := Q_k s_j, \quad \text{where} \quad I_k s_j = (H_k - \tau I_k)^\dagger I_k s_j (\theta_j - \tau), \quad 1 \leq j \leq k.
\]

Less known: \(\rho\)-values, refined extraction, combinations thereof.

Mostly unknown: MR eigenpairs \((\hat{\theta}, \hat{y} = Q_k \hat{s})\),

\[
\| (\hat{\theta} I_k - H_k) \hat{s} \| \ := \ min \ \text{loc} \left( \frac{\| (z I_k - H_k) s \|}{\| s \|} \right),
\]

Lehmann: MR by minimization over shifts in harmonic Ritz \& \(\rho\)-values.
A graphical representation

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

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 $(\tilde{\theta}, \tilde{y})$ gives the backward error, i.e., $w = \min\{\|\Delta A\|: (A + \Delta A)\tilde{y} = \tilde{y}\tilde{\theta}\}$.

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 graphical representation

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

\[
z = \tilde{\theta}, \quad w = \frac{\| (\tilde{\theta}I_k - H_k)\tilde{s} \|}{\| \tilde{s} \|}.
\] (1)
A graphical representation

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

\[
z = \tilde{\theta}, \quad w = \frac{\left\| (\tilde{\theta}I_k - H_k) s \right\|}{\| s \|}. \tag{1}
\]

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

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

\[
z = \tilde{\theta}, \quad w = \frac{\| (\tilde{\theta}I_k - H_k) \tilde{s} \|}{\| \tilde{s} \|}.
\] (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 \((\tilde{\theta}, \tilde{y})\) gives the backward error, i.e.,

\[
w = \min \left\{ \| \Delta A \| : (A + \Delta A)\tilde{y} = \tilde{y} \tilde{\theta} \right\}.
\] (2)
A graphical representation

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

\[
z = \tilde{\theta}, \quad w = \frac{\| (\tilde{\theta} I_k - H_k) \tilde{s} \|}{\| \tilde{s} \|}.
\]

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 \((\tilde{\theta}, \tilde{y})\) gives the backward error, i.e.,

\[
w = \min \left\{ \| \Delta A \| : (A + \Delta A) \tilde{y} = \tilde{y} \tilde{\theta} \right\}.
\]

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

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

\[
z = \tilde{\theta}, \quad w = \frac{\| (\tilde{\theta} I_k - H_k) \tilde{s} \|}{\| \tilde{s} \|}.
\] (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 \((\tilde{\theta}, \tilde{y})\) gives the backward error, i.e.,

\[
w = \min \left\{ \| \Delta A \| : (A + \Delta A) \tilde{y} = \tilde{y} \tilde{\theta} \right\}.
\] (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

\[
H_3 = \begin{pmatrix}
0 & 1 & 0 \\
1 & 0 & 1 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}.
\] (3)
A beautiful example

As an example we use

$$\mathbf{H}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$  \hspace{1cm} (3)

Its Ritz values are given by

$$\theta_{1,3} = \mp \sqrt{2} \approx \mp 1.41421356, \quad \theta_2 = 0,$$  \hspace{1cm} (4)
A beautiful example

As an example we use

$$H_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$  \hfill (3)

Its Ritz values are given by

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

its harmonic Ritz values are given by

$$\theta'_{1,3} = \mp \sqrt{2} \approx \mp 1.41421356, \quad \theta_2 = \infty,$$  \hfill (5)
A beautiful example

As an example we use

\[ H_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \] (3)

Its Ritz values are given by

\[ \theta_{1,3} = \pm \sqrt{2} \approx \mp 1.41421356, \quad \theta_2 = 0, \] (4)

its harmonic Ritz values are given by

\[ \theta_{1,3} = \pm \sqrt{2} \approx \mp 1.41421356, \quad \theta_2 = \infty, \] (5)

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

\[ \rho_{1,3} = \pm \sqrt{2} \cdot \frac{2}{3} \approx \mp 0.9428090, \quad \rho_2 = 0, \] (6)
A beautiful example

As an example we use

\[ H_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \]  

(3)

Its Ritz values are given by

\[ \theta_{1,3} = \mp \sqrt{2} \approx \mp 1.41421356, \quad \theta_2 = 0, \]  

(4)

its harmonic Ritz values are given by

\[ \tilde{\theta}_{1,3} = \mp \sqrt{2} \approx \mp 1.41421356, \quad \tilde{\theta}_2 = \infty, \]  

(5)

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

\[ \rho_{1,3} = \mp \sqrt{2} \cdot \frac{2}{3} \approx \mp 0.9428090, \quad \rho_2 = 0, \]  

(6)

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

\[ \hat{\theta}_{1,3} = \mp \frac{\sqrt{2}}{16} \sqrt{113 + 2 \text{Re} \sqrt[3]{y}} \approx \mp 1.37898323557, \quad \hat{\theta}_2 = 0. \]  

(7)
A beautiful example

characteristics of a 4 x 3 extended symmetric tridiagonal matrix

- transformed unit sphere
- Ritz
- refined Ritz
- harmonic Ritz
- refined harmonic Ritz
- harmonic Rayleigh
- QMReig
- singular value curves

size of the associated residuals

location of the approximate eigenvalues
A beautiful example

characteristics of a 4 x 3 extended symmetric tridiagonal matrix

- transformed unit sphere
- Ritz
- refined Ritz
- harmonic Ritz
- refined harmonic Ritz
- harmonic Rayleigh
- QMReig
- singular value curves
- shifted harmonic

location of the approximate eigenvalues

size of the associated residuals
A beautiful example

characteristics of a 4 x 3 extended symmetric tridiagonal matrix

- transformed unit sphere
- Ritz
- refined Ritz
- harmonic Ritz
- refined harmonic Ritz
- harmonic Rayleigh
- QMReig

singular value curves
ρ−values

size of the associated residuals

location of the approximate eigenvalues
OR and MR for Sonneveld-based methods

Generalized Hessenberg decomposition:

\[ AV_k = AG_k U_k = G_{k+1} H_k, \quad V_k := G_k U_k. \]
OR and MR for Sonneveld-based methods

Generalized Hessenberg decomposition:

\[ AV_k = AG_k U_k = G_{k+1} H_k, \quad V_k := G_k U_k. \]

Rules of thumb: Use \( V_k \), not \( G_k \) as “basis”; insert \( U_k \) appropriately.
OR and MR for Sonneveld-based methods

Generalized Hessenberg decomposition:

\[ AV_k = AG_k U_k = G_{k+1} H_k, \quad V_k := G_k U_k. \]

Rules of thumb: Use \( V_k \), not \( G_k \) as “basis”; insert \( U_k \) appropriately.

Sonneveld OR (\( H_k \) regular):

\[ z_k := H_k^{-1} e_1 \| r_0 \|, \quad x_k := V_k z_k = G_k U_k z_k. \]
Classification of Krylov subspace methods

OR and MR for Sonneveld-based methods

Generalized Hessenberg decomposition:

\[ AV_k = AG_k U_k = G_{k+1} H_k, \quad V_k := G_k U_k. \]

Rules of thumb: Use \( V_k \), not \( G_k \) as “basis”; insert \( U_k \) appropriately.

Sonneveld OR (\( H_k \) regular):

\[ z_k := H_k^{-1} e_1 \| r_0 \|, \quad x_k := V_k z_k = G_k U_k z_k. \]

Sonneveld MR:

\[ z_k := H_k^{\dagger} e_1 \| r_0 \|, \quad x_k := V_k z_k = G_k U_k z_k. \]
OR and MR for Sonneveld-based methods

Generalized Hessenberg decomposition:

$$AV_k = AG_k U_k = G_{k+1}H_k, \quad V_k := G_k U_k.$$  

Rules of thumb: Use $V_k$, not $G_k$ as “basis”; insert $U_k$ appropriately.

**Sonneveld OR** ($H_k$ regular):

$$z_k := H_k^{-1}e_1 \|r_0\|, \quad x_k := V_k z_k = G_k U_k z_k.$$  

**Sonneveld MR**:

$$z_k := H_k^{\dagger} e_1 \|r_0\|, \quad x_k := V_k z_k = G_k U_k z_k.$$  

**Sonneveld Ritz**:

$$H_k s_j = \theta_j U_k s_j, \quad y_j := V_k s_j = G_k U_k s_j.$$
Classification of Krylov subspace methods

OR and MR for Sonneveld-based methods

Generalized Hessenberg decomposition:

$$AV_k = AG_kU_k = G_{k+1}H_k, \quad V_k := G_kU_k.$$  

Rules of thumb: Use $V_k$, not $G_k$ as “basis”; insert $U_k$ appropriately.

**Sonneveld OR** ($H_k$ regular):

$$z_k := H_k^{-1}e_1 \|r_0\|, \quad x_k := V_kz_k = G_kU_kz_k.$$  

**Sonneveld MR**:

$$z_k := H_k^+e_1 \|r_0\|, \quad x_k := V_kz_k = G_kU_kz_k.$$  

**Sonneveld Ritz**:

$$H_ks_j = \theta_jU_ks_j, \quad y_j := V_ks_j = G_kU_ks_j.$$  

**Sonneveld (shifted) harmonic Ritz**:

$$I_ks_j = (\theta_j - \tau)(H_k - \tau U_k)^+U_ks_j, \quad y_j := V_ks_j = G_kU_ks_j.$$
Beyond “classical” Krylov subspace methods

Generalizations:

\[ \mathcal{F}_k := \mathcal{F}_k(A, q) := \{ f_{k-1}(A)q \mid f_{k-1} \text{ structured, e.g., rational} \} \]
Beyond “classical” Krylov subspace methods

Generalizations:

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

Rational methods:

- Rational Krylov (Ruhe);
- Rayleigh Quotient Iteration (RQI); Lord Rayleigh’s original iteration.
Beyond “classical” Krylov subspace methods

Generalizations:

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

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.
Beyond “classical” Krylov subspace methods

Generalizations:

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

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.
Beyond “classical” Krylov subspace methods

Generalizations:

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

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 $\leadsto$ Hessenberg (tridiagonal) matrices:
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*, \ldots
Hessenberg structure

Krylov subspace method $\leadsto$ 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)):

\[
(zI_k - H_k) \nu_k(z) = e_1 \frac{\chi_k(z)}{\prod_{\ell=1}^k h_{\ell+1,\ell}}, \quad (\check{\nu}_k(z))^T (zI_k - H_k) = \frac{\chi_k(z)}{\prod_{\ell=1}^k h_{\ell+1,\ell}} e_k^T,
\]

with polynomial vectors ($\chi_{i:j}(z) := \det (zI_{j-i+1} - H_{i:j})$)

\[
e_i^T \nu_k(z) := \frac{\chi_{i+1:k}(z)}{\prod_{\ell=i+1}^k h_{\ell,\ell-1}}, \quad e_i^T \check{\nu}_k(z) := \frac{\chi_{1:i-1}(z)}{\prod_{\ell=1}^{i-1} h_{\ell+1,\ell}},
\]
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)):

\[
(zI_k - H_k)\nu_k(z) = e_1 \frac{\chi_k(z)}{\prod_{\ell=1}^{k} h_{\ell+1,\ell}}, \quad (\tilde{\nu}_k(z))^T(zI_k - H_k) = \frac{\chi_k(z)}{\prod_{\ell=1}^{k} h_{\ell+1,\ell}} e_k^T,
\]

with polynomial vectors \((\chi_{i:j}(z) := \det(zI_{j-i+1} - H_{i:j}))\)

\[
e_i^T \nu_k(z) := \frac{\chi_{i+1:k}(z)}{\prod_{\ell=i+1}^{k} h_{\ell,\ell-1}}, \quad e_i^T \tilde{\nu}_k(z) := \frac{\chi_{1:i-1}(z)}{\prod_{\ell=1}^{i-1} h_{\ell+1,\ell}},
\]

\(\rightsquigarrow\) Adjugate; inverse; eigenvectors and principal vectors; nullspace.
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
Linear independence $\leadsto$ orthonormality

Krylov matrix $K_{k+1}(A, q)$ rank deficient ($k$ minimal) $\leadsto$ minimal polynomial $\mu_k$:

$$K_k(A, q)c = A^k q \ \Rightarrow \ \mu_k(A)q = o_n, \ \mu_k(z) = z^k - \sum_{i=1}^{k} c_i z^{i-1}.$$
Krylov matrix $\mathbf{K}_{k+1}(\mathbf{A}, \mathbf{q})$ rank deficient ($k$ minimal) $\leadsto$ minimal polynomial $\mu_k$:

$$\mathbf{K}_k(\mathbf{A}, \mathbf{q})\mathbf{c} = \mathbf{A}^k \mathbf{q} \Rightarrow \mu_k(\mathbf{A})\mathbf{q} = \mathbf{0}_n, \quad \mu_k(z) = z^k - \sum_{i=1}^{k} c_i z^{i-1}.$$  

Eigenvalues, Inverse:

$$\mathbf{A}\mathbf{K}_k = \mathbf{K}_k \mathbf{F}_k, \quad \mathbf{F}_k := \begin{pmatrix} \mathbf{0}_{k-1}^T \\ \mathbf{I}_{k-1} \end{pmatrix} \mathbf{c}, \quad \mathbf{A}\mathbf{x} = \mathbf{A}(\mathbf{A}^{k-1}\mathbf{q} - \sum_{i=2}^{k} c_i \mathbf{A}^{i-2}\mathbf{q}) = \mathbf{q}c_1.$$
Linear independence $\leadsto$ orthonormality

Krylov matrix $K_{k+1}(A, q)$ rank deficient ($k$ minimal) $\leadsto$ minimal polynomial $\mu_k$:

$$K_k(A, q)c = A^k q \quad \Rightarrow \quad \mu_k(A)q = o_n, \quad \mu_k(z) = z^k - \sum_{i=1}^{k} c_i z^{i-1}. $$

Eigenvalues, Inverse:

$$AK_k = K_k F_k, \quad F_k := \begin{pmatrix} o_{k-1}^T \\ I_{k-1} \end{pmatrix} c, \quad Ax = A (A^{k-1} q - \sum_{i=2}^{k} c_i A^{i-2} q) = q c_1. $$

Natural idea: use **linearly independent vectors** for some other basis $Q_k$. 
Linear independence $\leadsto$ orthonormality

Krylov matrix $K_{k+1}(A, q)$ rank deficient ($k$ minimal) $\leadsto$ minimal polynomial $\mu_k$:

$$K_k(A, q)c = A^k q \Rightarrow \mu_k(A)q = o_n, \quad \mu_k(z) = z^k - \sum_{i=1}^{k} c_i z^{i-1}. $$

Eigenvalues, Inverse:

$$AK_k = K_k F_k, \quad F_k := \begin{pmatrix} o^T_{k-1} \\ I_{k-1} \end{pmatrix} \begin{pmatrix} c \end{pmatrix}, \quad Ax = A(A^{k-1} q - \sum_{i=2}^{k} c_i A^{i-2} q) = q c_1. $$

Natural idea: use linearly independent vectors for some other basis $Q_k$.

$\leadsto$ nested basis transformation: $K_{k+1} = Q_{k+1} R_{k+1}$ with $R_{k+1}$ upper triangular.
Linear independence $\leadsto$ orthonormality

Krylov matrix $K_{k+1}(A, q)$ rank deficient ($k$ minimal) $\leadsto$ minimal polynomial $\mu_k$:

$$K_k(A, q)c = A^k q \Rightarrow \mu_k(A)q = o_n, \quad \mu_k(z) = z^k - \sum_{i=1}^k c_i z^{i-1}.$$ 

Eigenvalues, Inverse:

$$AK_k = K_k F_k, \quad F_k := \begin{pmatrix} o_{k-1}^T & c \end{pmatrix}, \quad Ax = A(A^{k-1}q - \sum_{i=2}^k c_i A^{i-2}q) = qc_1.$$ 

Natural idea: use linearly independent vectors for some other basis $Q_k$.

$\leadsto$ nested basis transformation: $K_{k+1} = Q_{k+1} R_{k+1}$ with $R_{k+1}$ upper triangular.

Hessenberg: LU decomposition: $K_{k+1} = L_{k+1} R_{k+1}, r_{ii} = 1, 1 \leq i \leq k + 1$.

Arnoldi: orthonormal basis, i.e., QR decomposition: $K_{k+1} = Q_{k+1} R_{k+1}$. 

Linear independence $\leadsto$ orthonormality

Krylov matrix $K_{k+1}(A, q)$ rank deficient ($k$ minimal) $\leadsto$ minimal polynomial $\mu_k$:

$$K_k(A, q)c = A^k q \quad \Rightarrow \quad \mu_k(A)q = 0, \quad \mu_k(z) = z^k - \sum_{i=1}^{k} c_i z^{i-1}.$$  

Eigenvalues, Inverse:

$$A K_k = K_k F_k, \quad F_k := \begin{pmatrix} o_{k-1}^T & c \\ I_{k-1} & \end{pmatrix}, \quad A x = A (A^{k-1} q - \sum_{i=2}^{k} c_i A^{i-2} q) = q c_1.$$  

Natural idea: use linearly independent vectors for some other basis $Q_k$.

$\leadsto$ nested basis transformation: $K_{k+1} = Q_{k+1} R_{k+1}$ with $R_{k+1}$ upper triangular.

Hessenberg: LU decomposition: $K_{k+1} = L_{k+1} R_{k+1}, \quad r_{ii} = 1, \quad 1 \leq i \leq k + 1$.

Arnoldi: orthonormal basis, i.e., QR decomposition: $K_{k+1} = Q_{k+1} R_{k+1}$.

Extended Hessenberg matrix as quotient: $(e_1, H_k) = R_{k+1} \begin{pmatrix} 1 \\ o_k^T \\ 0_k \\ R_k^{-1} \end{pmatrix}$. 
Linear independence $\rightarrow$ orthonormality

Krylov matrix $K_{k+1}(A, q)$ rank deficient ($k$ minimal) $\rightarrow$ minimal polynomial $\mu_k$:

$$K_k(A, q)c = A^k q \quad \Rightarrow \quad \mu_k(A)q = o_n, \quad \mu_k(z) = z^k - \sum_{i=1}^{k} c_i z^{i-1}.$$  

Eigenvalues, Inverse:

$$AK_k = K_k F_k, \quad F_k := \left( \begin{array}{c} o_{k-1}^T \\ I_{k-1} \end{array} \right) c, \quad Ax = A(A^{k-1} q - \sum_{i=2}^{k} c_i A^{i-2} q) = qc_1.$$  

Natural idea: use linearly independent vectors for some other basis $Q_k$.

$\rightarrow$ nested basis transformation: $K_{k+1} = Q_{k+1} R_{k+1}$ with $R_{k+1}$ upper triangular.

Hessenberg: LU decomposition: $K_{k+1} = L_{k+1} R_{k+1}, r_{ii} = 1, 1 \leq i \leq k + 1$.

Arnoldi: orthonormal basis, i.e., QR decomposition: $K_{k+1} = Q_{k+1} R_{k+1}$.

Extended Hessenberg matrix as quotient: $(e_1, H_k) = R_{k+1} \left( \begin{array}{c} 1 \\ o_k^T \\ o_k \\ R_k^{-1} \end{array} \right)$.  

Arnoldi based on orthogonal projection: minimal coeffs $c \rightarrow$ “optimal”.
Arnoldi decomposition:

\[ AQ_k = Q_{k+1} H_k. \]
Arnoldi decomposition:

\[ AQ_k = Q_{k+1}H_k. \]

Construction:
Arnoldi decomposition:

\[ AQ_k = Q_{k+1}H_k. \]

Construction:

\[ H_0 = [ ]; Q_1 = q_1 = q/\|q\|; \]

\[ \text{for } i=1:k \text{ do} \]
\[ r = Aq_i; \]
\[ h_i = Q_i^Hr; \]
\[ r = r - Q_ih_i; \]
\[ h_{i+1,i} = \|r\|; \]
\[ q_{i+1} = r/h_{i+1,i}; \]
\[ H_i = \begin{pmatrix} H_{i-1} & h_i \\ o_{i-1}^T & h_{i+1,i} \end{pmatrix}; \]
\[ Q_{i+1} = (Q_i, q_{i+1}); \]

\text{done}
Arnoldi decomposition:

\[ AQ_k = Q_{k+1}H_k. \]

Construction:

\[ H_0 = \begin{bmatrix} \end{bmatrix}; Q_1 = q_1 = q/\|q\|; \]

\[
\text{for } i=1:k \text{ do}
\]

\[ r = Aq_i; \]
\[ h_i = Q_i^H r; \]
\[ r = r - Q_i h_i; \]
\[ h_{i+1,i} = \|r\|; \]
\[ q_{i+1} = r/h_{i+1,i}; \]
\[ H_i = \begin{pmatrix} H_{i-1} & h_i \\ o_{i-1}^T & h_{i+1,i} \end{pmatrix}; \]
\[ Q_{i+1} = (Q_i, q_{i+1}); \]

\text{done}

Gram-Schmidt variant.
Others possible.

Other inner products or semi-inner products possible.
Classification of Krylov subspace methods

Lanczos-based

Outline

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

Connections
Interpolation
Approximation

Applications
RQI and the Opitz-Larkin Method
QMRIDR & IDREig
Augmented Backward Error Analysis
Lanczos: biorthonormal bases $\sim \hat{Q}_{k+1}^H Q_{k+1} = I_{k+1}$ of

\[
\mathcal{K}_k := \mathcal{K}_k(A, q) := \text{span}\{q, Aq, A^2q, \ldots, A^{k-1}q\} = \{p_{k-1}(A)q \mid p_{k-1} \in \Pi_{k-1}\},
\]

\[
\hat{\mathcal{K}}_k := \hat{\mathcal{K}}_k(A^H, \hat{q}) := \text{span}\{\hat{q}, A^H\hat{q}, A^{2H}\hat{q}, \ldots, A^{(k-1)H}\hat{q}\}.
\]
Linear independence using less vectors

Lanczos: biorthonormal bases \( \implies \hat{Q}^H_{k+1} Q_{k+1} = I_{k+1} \) of

\[
\mathcal{K}_k := \mathcal{K}_k(A, q) := \text{span}\{q, Aq, A^2q, \ldots, A^{k-1}q\} = \{p_{k-1}(A)q \mid p_{k-1} \in \Pi_{k-1}\},
\]

\[
\hat{\mathcal{K}}_k := \hat{\mathcal{K}}_k(A^H, \hat{q}) := \text{span}\{\hat{q}, A^H\hat{q}, A^{2H}\hat{q}, \ldots, A^{(k-1)H}\hat{q}\}.
\]

Based on three-term recurrence for the solutions \( \eta_k, \tilde{\eta}_k \) of the Hankel systems

\[
C_{k+1} \begin{pmatrix} \eta_k \\ 1 \end{pmatrix} = e_{k+1}h_k, \quad \tilde{C}_{k+2} \begin{pmatrix} \tilde{\eta}_k \\ 1 \end{pmatrix} = e_{k+1}\tilde{h}_{k+1},
\]

\[
C_{k+1} = \hat{K}^H_{k+1} K_{k+1} = \begin{pmatrix}
  c_0 & c_1 & c_2 & \cdots & c_k \\
  c_1 & c_2 & c_3 & \cdots & c_{k+1} \\
  c_2 & c_3 & c_4 & \cdots & c_{k+2} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  c_k & c_{k+1} & c_{k+2} & \cdots & c_{2k}
\end{pmatrix}, \quad c_i = \hat{q}^H A^i q,
\]

where \( \tilde{C}_{k+2} \) is \( \tilde{C}_{k+2} \) w/o first row & last column.
(Example of) **Lanczos decompositions:**

\[
\begin{align*}
AQ_k &= Q_{k+1}T_k, \\
A^H\hat{Q}_k &= \hat{Q}_{k+1}\hat{T}_k, \\
\hat{Q}_{k+1}^HQ_{k+1} &= I_{k+1}, \\
T_k^H &= \hat{T}_k.
\end{align*}
\]
Modern implementations

(Example of) **Lanczos decompositions:**

\[ AQ_k = Q_{k+1}T_k, \quad A^H\hat{Q}_k = \hat{Q}_{k+1}\hat{T}_k, \quad \hat{Q}_{k+1}Q_{k+1} = I_{k+1}, \quad T_k^H = \hat{T}_k. \]

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

\[ r = Aq_k - q_k\alpha_k - q_{k-1}\beta_k, \quad \beta_{k+1}\beta_{k+1} = \langle \hat{r}, r \rangle, \quad q_{k+1} = r/\beta_{k+1}, \]

\[ \hat{r} = A^H\hat{q}_k - \hat{q}_k\alpha_k - \hat{q}_{k-1}\beta_k, \quad \hat{q}_{k+1} = \hat{r}/\beta_{k+1}. \]
Modern implementations

(Example of) **Lanczos decompositions:**

\[
AQ_k = Q_{k+1}T_k, \quad A^H\hat{Q}_k = \hat{Q}_{k+1}\hat{T}_k, \quad \hat{Q}_{k+1}^HQ_{k+1} = I_{k+1}, \quad T_k^H = \hat{T}_k.
\]

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

\[
\begin{align*}
\mathbf{r} &= A\mathbf{q}_k - \mathbf{q}_k\alpha_k - \mathbf{q}_{k-1}\beta_k, \\
\mathbf{r} &= A^H\hat{\mathbf{q}}_k - \hat{\mathbf{q}}_k\alpha_k - \hat{\mathbf{q}}_{k-1}\beta_k, \\
\hat{\mathbf{r}} &= A^H\hat{\mathbf{q}}_k - \hat{\mathbf{q}}_k\alpha_k - \hat{\mathbf{q}}_{k-1}\beta_k,
\end{align*}
\]

\[
\begin{align*}
\beta_{k+1}\beta_{k+1} &= \langle \mathbf{r}, \mathbf{r} \rangle, \\
\mathbf{q}_{k+1} &= \mathbf{r} / \beta_{k+1}, \\
\hat{\mathbf{q}}_{k+1} &= \hat{\mathbf{r}} / \beta_{k+1}.
\end{align*}
\]

- Hankel matrices may become singular vs. inner products may be zero: need for **look-ahead.**
- Problems with incurable breakdown (in finite fields):
  ~> **Taylor’s mismatch theorem.**
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
Avoiding the use of the transpose

Lanczos method can be generalized:

- block variants $\mapsto \ell$ left- and right-hand starting vectors;
- block variants with different number of left- and right-hand starting vectors $\mapsto$ applications in model reduction.
Avoiding the use of the transpose

Lanczos method can be generalized:

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

Variants denoted by $\text{Lanczos}(\ell, s)$, $\ell$ 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.
Avoiding the use of the transpose

Lanczos method can be generalized:

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

Variants denoted by $\text{Lanczos}(\ell, s)$, $\ell$ 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 $A$, no need for transpose, but $n \sim 2n$;
- **Sonneveld, 1979**: Birth of “Induced Dimension Reduction”;
- **Sonneveld, 1989**: $\langle \bar{p}(A^H)\hat{r}_0, q(A)r_0 \rangle = \langle \hat{r}_0, p(A)q(A)r_0 \rangle$;
- Famous classical examples of Sonneveld-based methods: CGS, BICGSTAB, Wiedemann’s method (for finite fields);
- $\text{Lanczos}(s, 1)$ without transpose: IDR$(s)$ & Sonneveld spaces.
IDR\(_{(s)}\)

IDR spaces:

\[ G_0 := \mathcal{K}(A, q), \quad \text{(full Krylov subspace)} \]
\[ G_j := (A - \mu_j I)(G_{j-1} \cap S), \quad j \geq 1, \quad \mu_j \in \mathbb{C}, \]

where

\[ \text{codim}(S) = s, \quad \text{e.g.,} \quad S = \text{span}\{\tilde{R}_0\}^\perp, \quad \tilde{R}_0 \in \mathbb{C}^{n \times s}. \]
IDR spaces:

\[ G_0 := \mathcal{K}(A, q), \quad \text{(full Krylov subspace)} \]
\[ G_j := (A - \mu_j I)(G_{j-1} \cap S), \quad j \geq 1, \quad \mu_j \in \mathbb{C}, \]

where

\[ \text{codim}(S) = s, \quad \text{e.g.,} \quad S = \text{span}\{\tilde{R}_0\}^\perp, \quad \tilde{R}_0 \in \mathbb{C}^{n \times s}. \]

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

\[ G_j = S_j(P_j, A, \tilde{R}_0) := \left\{ M_j(A)v \mid v \perp \mathcal{K}_j(A^H, \tilde{R}_0), \ v \in G_0 \right\}, \]

\[ M_j(z) := \prod_{i=1}^{j} (z - \mu_i). \]
IDR\((s)\)

**IDR spaces:**

\[
G_0 := \mathcal{K}(A, q), \quad \text{(full Krylov subspace)}
\]

\[
G_j := (A - \mu_j I)(G_{j-1} \cap S), \quad j \geq 1, \quad \mu_j \in \mathbb{C},
\]

where

\[
\text{codim}(S) = s, \quad \text{e.g.,} \quad S = \text{span}\{\tilde{R}_0\}^\perp, \quad \tilde{R}_0 \in \mathbb{C}^{n \times s}.
\]

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

\[
G_j = S_j(P_j, A, \tilde{R}_0) := \left\{ M_j(A)v \mid v \perp \mathcal{K}_j(A^H, \tilde{R}_0), v \in G_0 \right\},
\]

\[
M_j(z) := \prod_{i=1}^{j}(z - \mu_i).
\]

Image of shrinking space: **Induced Dimension Reduction.**
IDR spaces nested:

\[
\{0\} = G_{j_{\max}} \subsetneq \cdots \subsetneq G_{j+1} \subsetneq G_j \subsetneq G_{j-1} \subsetneq \cdots \subsetneq G_2 \subsetneq G_1 \subsetneq G_0.
\]
IDR($s$)

IDR spaces nested:

$$\{0\} = G_{j_{\text{max}}} \subsetneq \cdots \subsetneq G_{j+1} \subsetneq G_j \subsetneq G_{j-1} \subsetneq \cdots \subsetneq G_2 \subsetneq G_1 \subsetneq G_0.$$  

How many vectors in $G_j \setminus G_{j+1}$? In generic case, $s + 1$. 
IDR($s$)

IDR spaces nested:

$$\{0\} = G_{j_{\text{max}}} \subset \cdots \subset G_{j+1} \subset G_j \subset G_{j-1} \subset \cdots \subset G_2 \subset G_1 \subset G_0.$$ 

How many vectors in $G_j \setminus G_{j+1}$? In generic case, $s + 1$.

Stable basis: Partially orthonormalize basis vectors $g_k$, $1 \leq k \leq n$: 
IDR $\left( s \right)$

IDR spaces nested:

$$\{ \mathbf{0} \} = G_{j_{\text{max}}} \subseteq \cdots \subseteq G_{j+1} \subseteq G_{j} \subseteq G_{j-1} \subseteq \cdots \subseteq G_{2} \subseteq G_{1} \subseteq G_{0}. $$

How many vectors in $G_{j} \setminus G_{j+1}$? In generic case, $s + 1$.

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

Arnoldi: compute orthonormal basis of $\mathcal{K}_{s+1} \subset G_{0}$,

$$A_{s}G_{s} = G_{s+1}H_{s}. $$
IDR\(_{(s)}\)

IDR spaces nested:

\[
\{0\} = G_{j_{\text{max}}} \subset \cdots \subset G_{j+1} \subset G_{j} \subset G_{j-1} \subset \cdots \subset G_{2} \subset G_{1} \subset G_{0}.
\]

How many vectors in \(G_{j} \setminus G_{j+1}\)? In generic case, \(s + 1\).

Stable basis: Partially orthonormalize basis vectors \(g_k\), \(1 \leq k \leq n\):

Arnoldi: compute orthonormal basis of \(\mathcal{K}_{s+1} \subset G_0\),

\[
AG_s = G_{s+1}H_s.
\]

“Lanczos”: perform intersection \(G_{j} \cap S\), map, and orthonormalize,

\[
v_k = \sum_{i=k-s}^{k} g_i \gamma_i, \quad \tilde{R}_0^H v_k = 0_s, \quad k \geq s + 1,
\]
IDR\((s)\)

**IDR spaces nested:**

\[ \{o\} = G_{j_{\text{max}}} \subset \cdots \subset G_{j+1} \subset G_{j} \subset G_{j-1} \subset \cdots \subset G_2 \subset G_1 \subset G_0. \]

How many vectors in \(G_j \setminus G_{j+1}\)? In generic case, \(s + 1\).

**Stable basis:** Partially orthonormalize basis vectors \(g_k, 1 \leq k \leq n\):

**Arnoldi:** compute orthonormal basis of \(\mathcal{K}_{s+1} \subset G_0\),

\[ AG_s = G_{s+1}H_s. \]

“Lanczos”: perform intersection \(G_j \cap S\), map, and orthonormalize,

\[ v_k = \sum_{i=k-s}^{k} g_i \gamma_i, \quad \tilde{R}_0^H v_k = o_s, \quad k \geq s + 1, \]

\[ (A - \mu_j I)v_k \quad \text{, } j = \left\lceil \frac{k - 1}{s + 1} \right\rceil. \]
IDR(s)

IDR spaces nested:

\[ \{0\} = G_{j_{\max}} \subset G_{j+1} \subset G_j \subset G_{j-1} \subset \cdots \subset G_2 \subset G_1 \subset G_0. \]

How many vectors in \( G_j \setminus G_{j+1} \)? In generic case, \( s + 1 \).

Stable basis: Partially orthonormalize basis vectors \( g_k \), \( 1 \leq k \leq n \):

Arnoldi: compute orthonormal basis of \( \mathcal{K}_{s+1} \subset G_0 \),

\[ AG_s = G_{s+1}H_s. \]

“Lanczos”: perform intersection \( G_j \cap S \), map, and orthonormalize,

\[ v_k = \sum_{i=k-s}^{k} g_i \gamma_i, \quad \tilde{R}_0^H v_k = o_s, \quad k \geq s + 1, \]

\[ g_{k+1} \nu_{k+1} = (A - \mu_j I)v_k - \sum_{i=k-j(s+1)-1}^{k} g_i \nu_i, \quad j = \left\lfloor \frac{k - 1}{s + 1} \right\rfloor. \]
**Generalized Hessenberg decomposition:**

\[ AV_k = AG_k U_k = G_{k+1} H_k, \]

where \( U_k \in \mathbb{C}^{k \times k} \) upper triangular.
Generalized Hessenberg decomposition:

\[ AV_k = AG_k U_k = G_{k+1} H_k, \]

where \( U_k \in \mathbb{C}^{k \times k} \) upper triangular.

Structure of Sonneveld pencils:

\[
H_k = \begin{pmatrix}
\begin{array}{ccccccccc}
\times & \times & \times & \times & \times & \times & \times & \times & \times \\
+ & \times & \times & \times & \times & \times & \times & \times & \times \\
o & + & \times & \times & \times & \times & \times & \times & \times \\
o & o & + & \times & \times & \times & \times & \times & \times \\
o & o & o & + & \times & \times & \times & \times & \times \\
o & o & o & o & + & \times & \times & \times & \times \\
o & o & o & o & o & + & \times & \times & \times \\
o & o & o & o & o & o & + & \times & \times \\
o & o & o & o & o & o & o & + & \times \\
o & o & o & o & o & o & o & o & + \\
\end{array}
\end{pmatrix}
\]

\[
= U_k
\]
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
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:

Arnoldi: \[ AQ_k = Q_{k+1}H_k, \]

Lanczos: \[ AQ_k = Q_{k+1}T_k, \quad A^H\hat{Q}_k = \hat{Q}_{k+1}\hat{T}_k, \]

Sonneveld: \[ AV_k = AG_kU_k = G_{k+1}H_k, \quad V_k = G_kU_k. \]
Connections between the three approaches

(Generalized) Hessenberg decompositions:

Arnoldi: \[ AQ_k = Q_{k+1}H_k, \]

Lanczos: \[ AQ_k = Q_{k+1}T_k, \quad A^H\hat{Q}_k = \hat{Q}_{k+1}\hat{T}_k, \]

Sonneveld: \[ AV_k = AG_kU_k = G_{k+1}H_k, \quad V_k = G_kU_k. \]

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

\[ H_k = Q_k^H A Q_k = Q_k^H A^H Q_k = (Q_k^H A Q_k)^H = H_k^H = T_k; \]
Connections between the three approaches

(Generalized) Hessenberg decompositions:

Arnoldi: \[ AQ_k = Q_{k+1} H_k, \]
Lanczos: \[ AQ_k = Q_{k+1} T_k, \quad A^H \hat{Q}_k = \hat{Q}_{k+1} \hat{T}_k, \]
Sonneveld: \[ AV_k = AG_k U_k = G_{k+1} H_k, \quad V_k = G_k U_k. \]

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

  \[
  H_k = Q_k^H AQ_k = Q_k^H A^H Q_k = (Q_k^H AQ_k)^H = H_k^H = T_k;
  \]

- Lanczos is typically slower in terms of matrix-vector multiplies, faster in terms of computing time, but less stable than Arnoldi;
Connections between the three approaches

(Generalized) Hessenberg decompositions:

Arnoldi: \[ AQ_k = Q_{k+1}H_k, \]

Lanczos: \[ AQ_k = Q_{k+1}T_k, \quad A^H\hat{Q}_k = \hat{Q}_{k+1}\hat{T}_k, \]

Sonneveld: \[ AV_k = AG_kU_k = G_{k+1}H_k, \quad V_k = G_kU_k. \]

- Arnoldi and Lanczos (\( \hat{q} = q \)) are the same (so-called symmetric Lanczos) for Hermitean matrices (pencil \((K, M)\): \( (K - \sigma M)^{-1}M \) is \( M \)-symmetric):
  \[
  H_k = Q_k^HAQ_k = Q_k^HA^HQ_k = (Q_k^HAQ_k)^H = H_k^H = T_k;
  \]

- 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;
Connections between the three approaches

(Generalized) Hessenberg decompositions:

- **Arnoldi:** \( AQ_k = Q_{k+1} H_k, \)
  \[ AQ_k = Q_{k+1} H_k, \]

- **Lanczos:** \( AQ_k = Q_{k+1} T_k, \quad A^H Q_k = Q_{k+1} T_k, \)
  \[ AQ_k = Q_{k+1} T_k, A^H Q_k = Q_{k+1} T_k, \]

- **Sonneveld:** \( AV_k = AG_k U_k = G_{k+1} H_k, \quad V_k = G_k U_k. \)
  \[ AV_k = AG_k U_k = G_{k+1} H_k, V_k = G_k U_k. \]

- **Arnoldi and Lanczos** \( (\hat{q} = q) \) are the same (so-called **symmetric Lanczos**) for Hermitean matrices (pencil \((K, M)\): \((K - \sigma M)^{-1} M \) is \(M\)-symmetric):
  \[ H_k = Q_k^H AQ_k = Q_k^H A^H Q_k = (Q_k^H AQ_k)^H = H_k^H = T_k; \]

- **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.
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
Introducing: polynomials

For simplicity we only consider perturbed methods that satisfy

\[ AQ_k + F_k = Q_{k+1}H_k. \]

Polynomials based on computed \( H_k \) or \( H_k \leadsto \) useful properties.
Introducing: polynomials

For simplicity we only consider perturbed methods that satisfy

\[ AQ_k + F_k = Q_{k+1} + H_k. \]

Polynomials based on computed \( H_k \) or \( H_k \rightarrow \) useful properties.

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

For simplicity we only consider perturbed methods that satisfy

\[ AQ_k + F_k = Q_{k+1} H_k. \]

Polynomials based on computed \( H_k \) or \( H_k \) use 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 \( \mathcal{L}_k[1 - \delta_z] \),
- Lagrange interpolation polynomials \( \mathcal{L}_k[1 - \delta_{z0}] \) and \( \mathcal{L}_k[1 - \delta_{z0}] \),
- residual polynomials \( \mathcal{R}_k \) and \( \mathcal{R}_k \).
Introducing: polynomials

For simplicity we only consider perturbed methods that satisfy

\[ \mathbf{A} \mathbf{Q}_k + \mathbf{F}_k = \mathbf{Q}_{k+1} \mathbf{H}_k. \]

Polynomials based on computed \( \mathbf{H}_k \) or \( \mathbf{H}_k \) use 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 \( \mathcal{L}_k[z^{-1}] \),
- Lagrange interpolation polynomials \( \mathcal{L}_k[1 - \delta z_0] \) and \( \mathcal{L}_k[1 - \delta z_0] \),
- residual polynomials \( \mathcal{R}_k \) and \( \mathcal{R}_k \).

We restrict ourselves to \( \mathcal{A}_k \), \( \mathcal{L}_k[z^{-1}] \), \( \mathcal{L}_k[1 - \delta z_0] \) and \( \mathcal{R}_k \).
Adjugate polynomials

First we consider certain bivariate polynomials – the **adjugate polynomials**.
First we consider certain bivariate polynomials – the adjugate polynomials.

- **Property:**
  \[ A_k(z, H_k) = \text{adj}(zI_k - H_k). \]
Adjugate polynomials

First we consider certain bivariate polynomials – the adjugate polynomials.

Property:

\[ A_k(z, H_k) = \text{adj}(zI_k - H_k). \]

Implies (Schweins, 1825; Zemke, 2006)

\[ A_k(\theta_j, H_k)e_1 = s_j, \quad H_k s_j = s_j \theta_j \]

for all eigenvalues (Ritz values) \( \theta_j \) of \( H_k \).
Adjugate polynomials

First we consider certain bivariate polynomials – the adjugate polynomials.

Property:
\[ A_k(z, H_k) = \text{adj}(zI_k - H_k). \]

Implies (Schweins, 1825; Zemke, 2006)
\[ A_k(\theta_j, H_k)e_1 = s_j, \quad H_k s_j = s_j \theta_j \]
for all eigenvalues (Ritz values) \( \theta_j \) of \( H_k \).

Definition:
\[ A_k(\theta, z) := \frac{\chi_k(\theta) - \chi_k(z)}{\theta - z}, \quad \chi_k(z) := \text{det}(zI_k - H_k). \]
Adjugate polynomials

First we consider certain bivariate polynomials – the **adjugate polynomials**.

- **Property:**
\[
A_k(z, H_k) = \text{adj}(zI_k - H_k).
\]

- **Implies** (Schweins, 1825; Zemke, 2006)
\[
A_k(\theta_j, H_k)e_1 = s_j, \quad H_k s_j = s_j \theta_j
\]
for all eigenvalues (Ritz values) \( \theta_j \) of \( H_k \).

- **Definition:**
\[
A_k(\theta, z) := \frac{\chi_k(\theta) - \chi_k(z)}{\theta - z}, \quad \chi_k(z) := \text{det} (zI_k - H_k).
\]

- **Generalization:**
\[
A_{\ell+1:k}(\theta, z) := \frac{\chi_{\ell+1:k}(\theta) - \chi_{\ell+1:k}(z)}{\theta - z}, \quad \ell = 0, 1, \ldots, k.
\]
Theorem (Ritz vectors)

Let $H_k S_\theta = S_\theta J_\theta$ (for a certain $S_\theta$). Let the Ritz matrix be given by $Y_\theta := Q_k S_\theta$. Then

\[
\text{vec}(Y_\theta) = \begin{pmatrix}
A_k(\theta, A) \\
A'_k(\theta, A) \\
\vdots \\
A^{(\alpha - 1)}_k(\theta, A) \\
(\alpha - 1)!
\end{pmatrix} q_1 + \sum_{\ell=1}^{k} \prod_{j=1}^{\ell-1} h_{\ell+1,j} \begin{pmatrix}
A_{\ell+1:k}(\theta, A) \\
A'_{\ell+1:k}(\theta, A) \\
\vdots \\
A^{(\alpha - 1)}_{\ell+1:k}(\theta, A) \\
(\alpha - 1)!
\end{pmatrix} f_{\ell},
\]

with derivation with respect to the shift $\theta$. 

(8)
Adjugate polynomials and Ritz vectors

**Theorem (Ritz vectors)**

Let \( H_k S_\theta = S_\theta J_\theta \) (for a certain \( S_\theta \)). Let the Ritz matrix be given by \( Y_\theta := Q_k S_\theta \).

Then

\[
\text{vec}(Y_\theta) = \begin{pmatrix}
A_k(\theta, A) \\
A'_k(\theta, A) \\
\vdots \\
A_k^{(\alpha-1)}(\theta, A) \\
(\alpha - 1)!
\end{pmatrix} q_1 + \sum_{\ell=1}^{k-1} \prod_{j=1}^{\ell-1} h_{j+1,j} \begin{pmatrix}
A_{\ell+1:k}(\theta, A) \\
A'_{\ell+1:k}(\theta, A) \\
\vdots \\
A_{\ell+1:k}^{(\alpha-1)}(\theta, A) \\
(\alpha - 1)!
\end{pmatrix} f_\ell, 
\]

(8)

with derivation with respect to the shift \( \theta \).

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

\[
y = \frac{A_k(\theta, A)}{\prod_{j=1}^{k-1} h_{j+1,j}} q_1 + \sum_{\ell=1}^{k} \frac{A_{\ell+1:k}(\theta, A)}{\prod_{j=\ell+1}^{k-1} h_{j+1,j}} \cdot \frac{f_\ell}{h_{\ell+1,\ell}}.
\]
Lagrange polynomials

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

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

  $$\mathcal{L}_k[z^{-1}](H_k) = H_k^{-1}.$$
We consider Lagrange interpolation polynomials interpolating the inverse and a singularly perturbed identity.

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

**Property:**

$$L_k[z^{-1}](H_k) = H_k^{-1}.$$  

**Definition:**

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

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

- **Generalization:**
  \[
  \mathcal{L}_{\ell+1:k}[z^{-1}](z) := \frac{\chi_{\ell+1:k}(0) - \chi_{\ell+1:k}(z)}{z \chi_{\ell+1:k}(0)} = -\frac{A_{\ell+1:k}(0, z)}{\chi_{\ell+1:k}(0)}, \quad \ell = 0, 1, \ldots, k.
  \]
Theorem (OR iterates)

Suppose that all $H_{\ell+1:k}$ are regular. Define $z_k := H_{k}^{-1}e_1 \|r_0\|$ and $x_k := Q_k z_k$. Then

$$x_k = L_k[z^{-1}](A)r_0 - \sum_{\ell=1}^{k} L_{\ell+1:k}[z^{-1}](A)f_{\ell} z_{\ell k}.$$  \tag{9}
Theorem (OR iterates)

Suppose that all $H_{\ell+1:k}$ are regular. Define $z_k := H_k^{-1}e_1\|r_0\|$ and $x_k := Q_k z_k$.

Then

$x_k = L_k[z^{-1}](A)r_0 - \sum_{\ell=1}^{k} L_{\ell+1:k}[z^{-1}](A)f_{\ell}z_{\ell k}$. \hspace{1cm} (9)

Really sloppily speaking, in case of convergence,

$x_\infty = A^{-1}r_0 + A^{-1}F_\infty z_\infty = A^{-1}(r_0 + F_\infty z_\infty)$. 
Suppose that all $H_{\ell+1:k}$ are regular. Define $z_k := H_k^{-1} e_1 \| r_0 \|$ and $x_k := Q_k z_k$. Then
\[ x_k = L_k[z^{-1}](A)r_0 - \sum_{\ell=1}^{k} L_{\ell+1:k}[z^{-1}](A)f_\ell z_{\ell k}. \] (9)

Really sloppily speaking, in case of convergence,
\[ x_\infty = A^{-1}r_0 + A^{-1}F_\infty z_\infty = A^{-1}(r_0 + F_\infty z_\infty). \]

Proving convergence is the hard task.
We consider Lagrange interpolation polynomials interpolating the inverse and a singularly perturbed identity.
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 \( \mathcal{L}_k^0[1 - \delta z_0](z) \).
We consider Lagrange interpolation polynomials interpolating the inverse and a singularly perturbed identity.

The Lagrange interpolation of the singularly perturbed identity is denoted by $L_k^0[1 - \delta z_0](z)$.

- **Properties:**

  \[ L_k^0[1 - \delta z_0](H_k) = I_k, \quad L_k^0[1 - \delta z_0](0) = 0. \]
We consider Lagrange interpolation polynomials interpolating the inverse and a singularly perturbed identity.

The Lagrange interpolation of the singularly perturbed identity is denoted by $L_k^0[1 - \delta z_0](z)$.

**Properties:**

$\mathcal{L}_k^0[1 - \delta z_0](H_k) = I_k, \quad \mathcal{L}_k^0[1 - \delta z_0](0) = 0.$

**Definition:**

$\mathcal{L}_k^0[1 - \delta z_0](z) := \frac{\chi_k(0) - \chi_k(z)}{\chi_k(0)} = \mathcal{L}_k[z^{-1}](z)z.$
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 $L_k^0[1 - \delta z_0](z)$.

- Properties:
  \[
  L_k^0[1 - \delta z_0](H_k) = I_k, \quad L_k^0[1 - \delta z_0](0) = 0.
  \]

- Definition:
  \[
  L_k^0[1 - \delta z_0](z) := \frac{\chi_k(0) - \chi_k(z)}{\chi_k(0)} = L_k[z^{-1}](z)z.
  \]

- Generalization ($\ell = 0, 1, \ldots, k$):
  \[
  L_{\ell+1:k}^0[1 - \delta z_0](z) := \frac{\chi_{\ell+1:k}(0) - \chi_{\ell+1:k}(z)}{\chi_{\ell+1:k}(0)} = L_{\ell+1:k}[z^{-1}](z)z.
  \]
Residual polynomials

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

- **Properties:**

\[
\mathcal{R}_k(H_k) = O_k, \quad \mathcal{R}_k(0) = 1.
\]
Residual polynomials

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

- **Properties:**
  \[ \mathcal{R}_k(H_k) = O_k, \quad \mathcal{R}_k(0) = 1. \]

- **Definition:**
  \[ \mathcal{R}_k(z) := \frac{\chi_k(z)}{\chi_k(0)} = 1 - L^0_k[1 - \delta z_0](z) = \det(I_k - zH_k^{-1}). \]
Residual polynomials

Well-known residual polynomials (Stiefel, 1955), denoted by $R_k(z)$.

- **Properties:**
  \[ R_k(H_k) = O_k, \quad R_k(0) = 1. \]

- **Definition:**
  \[ R_k(z) := \frac{\chi_k(z)}{\chi_k(0)} = 1 - L_0^k [1 - \delta_z 0](z) = \det (I_k - zH_k^{-1}). \]

- **Generalization** ($\ell = 0, 1, \ldots, k$):
  \[ R_{\ell+1:k}(z) := \frac{\chi_{\ell+1:k}(z)}{\chi_{\ell+1:k}(0)} = 1 - L_{\ell+1:k}^0 [1 - \delta_z 0](z). \]
Residual polynomials

Well-known residual polynomials (Stiefel, 1955), denoted by $R_k(z)$.

- **Properties:**
  \[ R_k(H_k) = O_k, \quad R_k(0) = 1. \]

- **Definition:**
  \[ R_k(z) := \frac{\chi_k(z)}{\chi_k(0)} = 1 - \mathcal{L}^0_k[1 - \delta z_0](z) = \det(I_k - zH_k^{-1}). \]

- **Generalization** ($\ell = 0, 1, \ldots, k$):
  \[ R_{\ell+1:k}(z) := \frac{\chi_{\ell+1:k}(z)}{\chi_{\ell+1:k}(0)} = 1 - \mathcal{L}^0_{\ell+1:k}[1 - \delta z_0](z). \]

Two types of polynomials $\leadsto$ two expressions for the OR residuals.
Suppose $q_1 = r_0/\|r_0\|$ and let all $H_{\ell+1:k}$ be invertible. Let $x_k$ denote the OR iterate and $r_k = r_0 - Ax_k$ the corresponding OR residual. Then

$$r_k = R_k(A)r_0 + \sum_{\ell=1}^{k} \mathcal{L}_{\ell+1:k}^0 [1 - \delta_z](A) f_{\ell} z_{\ell k}$$

(10)
Residual polynomials and OR residuals

Theorem (OR residuals)

Suppose \( q_1 = r_0 / \|r_0\| \) and let all \( H_{\ell+1:k} \) be invertible. Let \( x_k \) denote the OR iterate and \( r_k = r_0 - Ax_k \) the corresponding OR residual. Then

\[
\begin{align*}
    r_k &= R_k(A)r_0 + \sum_{\ell=1}^{k} L_{\ell+1:k}^0 \left[ 1 - \delta_{z0}(A) \right] f_\ell z_{\ell k} \\
    &= R_k(A)r_0 - \sum_{\ell=1}^{k} R_{\ell+1:k}(A) f_\ell z_{\ell k} + F_k z_k.
\end{align*}
\]

(10)

First expression: related to perturbation amplification.
Residual polynomials and OR residuals

Theorem (OR residuals)

Suppose $q_1 = r_0 / \| r_0 \|$ and let all $H_{\ell+1:k}$ be invertible. Let $x_k$ denote the OR iterate and $r_k = r_0 - Ax_k$ the corresponding OR residual. Then

$$r_k = R_k(A)r_0 + \sum_{\ell=1}^{k} L_{\ell+1:k}^0 [1 - \delta z_0](A) f_{\ell} z_{\ell k}$$

(10)

$$= R_k(A)r_0 - \sum_{\ell=1}^{k} R_{\ell+1:k}(A) f_{\ell} z_{\ell k} + F_k z_k.$$

First expression: related to perturbation amplification.
Second expression: related to the attainable accuracy.
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
The connection to approximation theory

OR and MR perform polynomial approximation. Best understood: case $Q_{k+1}$ orthonormal, i.e., Arnoldi/GMRES.
The connection to approximation theory

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

**OR = Arnoldi/symmetric Lanczos:**

$$\min_{p \in \Pi_k} \|p(A)q\|, \quad p(z) = z^k + \cdots \quad \Rightarrow \quad p(z) = \chi_k(z) = \det(zI_k - H_k).$$
The connection to approximation theory

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

\textbf{OR} = \textbf{Arnoldi/symmetric Lanczos}:

$$\min_{p \in \Pi_k} \|p(A)\|, \quad p(z) = z^k + \cdots \Rightarrow p(z) = \chi_k(z) = \det(zI_k - H_k).$$

\textbf{MR} = \textbf{GMRES/MinRes}:

$$\min_{p \in \Pi_k} \|p(A)\|, \quad p(z) = 1 + \cdots \Rightarrow p(z) = \chi_k(z) = \det(zI_k - H_k^\dagger I_k).$$
The connection to approximation theory

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

**OR = Arnoldi/symmetric Lanczos:**

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

**MR = GMRES/MinRes:**

$$\min_{p \in \Pi_k} \|p(A)q\|, \quad p(z) = 1 + \cdots \Rightarrow p(z) = \chi_k(z) = \det(zI_k - H_k^\dagger I_k).$$

- **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

Konvergenzverhalten des CG-Verfahrens

- Konvergenz exakt
- Konvergenz aufdatiert
- Konvergenz wirklich

Residuum in euklidischer Norm

Schritt des Verfahrens/Raumdimension
Convergence of CG, second example...
Characteristics of floating point Lanczos
Connections
Approximation

Characteristics of floating point Lanczos; details

Floating point Lanczos characteristics

positive distance to 3
negative distance to 3
derivative of Ritz value
upper stabilized bound

step number
distance to eigenvalue 3 / derivative

10^{-14}
10^{-15}
10^{-16}

50 100 150 200 250 300 350 400

TUHH
Jens-Peter M. Zemke
Krylov @ TUHH Kickoff 2012
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
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.
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
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:
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 \( A_1 : A_2 : A_3 \ldots \) we may calculate a first approximation to \( p^2 \) from

\[
p^2 = \frac{1}{2} c_{11} A_1^2 + \frac{1}{2} c_{22} A_2^2 + \ldots + c_{12} A_1 A_2 + \ldots \frac{1}{2} a_{11} A_1^2 + \frac{1}{2} a_{22} A_2^2 + \ldots + a_{12} A_1 A_2 + \ldots \ldots (3).
\]

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.]
In modern notation, Lord Rayleigh starts with an approximate eigenvector $v_k$, $k = 0$, of a Hermitian matrix (Hermitian pencil), computes its Rayleigh quotient

$$
\rho(v_k) := \frac{v_k^H A v_k}{v_k^H v_k},
$$

where $j$ may vary, depending on the computed approximate eigenvector. The Rayleigh quotient uniquely solves the least squares problem

$$
\rho(v_k) = \arg\min_{\rho \in \mathbb{C}} \|Av_k - v_k\rho\|.
$$
In modern notation, Lord Rayleigh starts with an approximate eigenvector $v_k$, $k = 0$, of a Hermitian matrix (Hermitian pencil), computes its Rayleigh quotient

$$\rho(v_k) := \frac{v_k^H A v_k}{v_k^H v_k},$$

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

$$v_{k+1} = \frac{(A - \rho(v_k) I_n)^{-1} e_j}{\| (A - \rho(v_k) I_n)^{-1} e_j \|}, \quad k = 0, 1, \ldots$$

where $j$ may vary, depending on the computed approximate eigenvector.
In modern notation, Lord Rayleigh starts with an approximate eigenvector \( v_k, k = 0 \), of a Hermitean matrix (Hermitean pencil), computes its Rayleigh quotient

\[
\rho(v_k) := \frac{v_k^H A v_k}{v_k^H v_k},
\]

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

\[
v_{k+1} = \frac{(A - \rho(v_k)I_n)^{-1}e_j}{\| (A - \rho(v_k)I_n)^{-1}e_j \|}, \quad k = 0, 1, \ldots
\]

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

The Rayleigh quotient uniquely solves the least squares problem

\[
\rho(v_k) = \arg\min_{\rho \in \mathbb{C}} \| A v_k - v_k \rho \|.
\]
Closely connected to RQI is inverse iteration (Wielandt, 1944).
Closely connected to RQI is inverse iteration (Wielandt, 1944). In its most basic variant the shift $\tau$ is never updated, but the right-hand side is replaced by the latest approximate eigenvector:

$$v_{k+1} = \frac{(A - \tau I_n)^{-1}v_k}{\| (A - \tau I_n)^{-1}v_k \|}, \quad k = 0, 1, \ldots$$
Closely connected to RQI is inverse iteration (Wielandt, 1944). In its most basic variant the shift $\tau$ is never updated, but the right-hand side is replaced by the latest approximate eigenvector:

$$v_{k+1} = \frac{(A - \tau I_n)^{-1}v_k}{\| (A - \tau I_n)^{-1}v_k \|}, \quad k = 0, 1, \ldots$$

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

$$\tau_{k+1} := \tau_k + \frac{1}{e_j^T (A - \tau_k I_n)^{-1}v_k}.$$
Inverse Iteration

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

$$v_{k+1} = \frac{(A - \tau I_n)^{-1}v_k}{\| (A - \tau I_n)^{-1}v_k \|}, \quad k = 0, 1, \ldots$$

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

$$\tau_{k+1} := \tau_k + \frac{1}{e_j^T (A - \tau_k I_n)^{-1}v_k}.$$  

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 \( \mathbf{A} \).
Modern variants of RQI

Combination gives (symmetric/Hermitean) RQI:

$$v_{k+1} = \frac{(A - \rho(v_k)I_n)^{-1}v_k}{\| (A - \rho(v_k)I_n)^{-1}v_k \|}, \quad k = 0, 1, \ldots$$

This iteration is also used for nonsymmetric $A$.

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

Combination gives (symmetric/Hermitean) RQI:

\[ v_{k+1} = \frac{(A - \rho(v_k)I_n)^{-1}v_k}{\|A - \rho(v_k)I_n)^{-1}v_k\|}, \quad k = 0, 1, \ldots \]

This iteration is also used for nonsymmetric \( A \).

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:

\[ \rho(w_k, v_k) := \frac{w_k^H A v_k}{w_k^H v_k}, \quad v_{k+1} = (A - \rho(w_k, v_k)I_n)^{-1}v_k, \quad w_{k+1} = (A - \rho(w_k, v_k)I_n)^{-H}w_k, \quad k = 0, 1, \ldots \]
Modern variants of RQI

Combination gives (symmetric/Hermitean) RQI:

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

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:

\[ \rho(\mathbf{w}_k, \mathbf{v}_k) := \frac{\mathbf{w}_k^H \mathbf{A} \mathbf{v}_k}{\mathbf{w}_k^H \mathbf{v}_k}, \quad \mathbf{v}_{k+1} = (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k) \mathbf{I}_n)^{-1} \mathbf{v}_k, \quad \mathbf{w}_{k+1} = (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k) \mathbf{I}_n)^{-H} \mathbf{w}_k, \quad k = 0, 1, \ldots \]

This trick recovers the cubic convergence rate of RQI at the expense of an additional system.
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:

\[ \rho(\mathbf{w}_k, \mathbf{v}_k) := \frac{\mathbf{w}_k^H \mathbf{A} \mathbf{v}_k}{\mathbf{w}_k^H \mathbf{v}_k}, \quad \mathbf{v}_{k+1} = (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k, \quad \mathbf{w}_{k+1} = (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k)\mathbf{I}_n)^{-H}\mathbf{w}_k, \quad k = 0, 1, \ldots \]

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

\[ f : \mathbb{C} \rightarrow \mathbb{C}, \quad f(z) := \frac{p(z)}{q(z)}, \quad p, q \in \mathbb{P}_m \]

include **Newton’s method**

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

and the **secant method**:

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

Methods for the **computation of a root** of a rational function

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

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**

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

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

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:

\[ \phi^2 = \phi + 1 \approx 2.618 > 2. \]
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.
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

\[ z_{k+1} = z_k + s \left( \frac{1}{f(z_k)} \right) \left( s - 1 \right) \left( z_k \right) \left( \frac{1}{f(z_k)} \right) \left( s \right) \left( z_k \right) , \]

\[ s = 1, 2, ... \]

König’s method for \( s = 1 \) is Newton’s method,

\[ z_{k+1} = z_k + \left( \frac{1}{f(z_k)} \right) \left( z_k \right) \left( \frac{1}{f(z_k)} \right) f'(z_k) = z_k - \frac{f(z_k)}{f'(z_k)} f''(z_k) / f(z_k)^2 \]
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 $A^\lambda_\omega$ with $\lambda = 0$. 
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 $A^\lambda_\omega$ with $\lambda = 0$.

This family is nowadays known as “König’s method”:

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

$$z_{k+1} = z_k + \frac{(1/f)(z_k)}{(1/f)'(z_k)} = z_k - \frac{1/f(z_k)}{f'(z_k)/(f(z_k))^2} = z_k - \frac{f(z_k)}{f'(z_k)}.$$
There is a natural extension of König’s method using divided differences in place of the derivatives.
The Opitz-Larkin method

There is a natural extension of König’s method using \textit{divided differences} in place of the \textit{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.
The Opitz-Larkin method

There is a natural extension of König’s method using \textit{divided differences} in place of the \textit{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.
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).
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 $z_{k+1} = x_{k+1}$,

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

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

$$\frac{[z_1, z_2, \ldots, z_{k-1}]}{[z_1, z_2, \ldots, z_{k-1}, z_k]} \left( \frac{1}{f} \right).$$

This variant of the Opitz-Larkin method converges with R-order 2.
Mostly, the $z_i$ are all **distinct** and the next iterate is used as **new evaluation point** $z_{k+1} = x_{k+1}$,

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

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

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

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

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 $1/f$. 
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:

$$z_{k+1} = z_k + \frac{s}{[z_k, \ldots, z_k]} \left( \frac{1}{f} \right)$$

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

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

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

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

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

$$= z_k - \frac{f(z_k)}{[z_{k-1}, z_k]},$$
Truncated Opitz-Larkin with $p = 1$ is the secant method,

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

\[ = z_k + \frac{1}{f(z_{k-1})} \cdot \frac{z_{k-1} - z_k}{1/f(z_{k-1}) - 1/f(z_k)} \]

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

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

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):
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 \( d \leq k - 2 \), such that \( q_d(\alpha) \neq 0 \) and

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

then

\[
z_k + \frac{[z_1, z_2, \ldots, z_{k-1}](1/f)}{[z_1, z_2, \ldots, z_{k-1}, z_k](1/f)} = \alpha.
\]
We set $zH_n := (zI_n - H_n)$. By the first resolvent identity (Chatelin, 1993)

$$
(\frac{1}{z_1H_n} - \frac{1}{z_2H_n}) = (\frac{1}{z_1I_n - H_n} - \frac{1}{z_2I_n - H_n}) = \frac{1}{z_2 - z_1} [-[z_1, z_2](zH_n)^{-1}].
$$

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

$$
\prod_{i=1}^{k} (\frac{1}{z_iH_n}) = \frac{1}{-1^{k-1}} [z_1, ..., z_k](zH_n)^{-1}.
$$
Simplification

We set $z^n_H := (zI_n - H_n)$. By the first resolvent identity (Chatelin, 1993)

$$\frac{\left( z_1^n H_n \right)^{-1} \left( z_2^n H_n \right)^{-1}}{z_2 - z_1} = \left( z_1^n I_n - H_n \right)^{-1} \left( z_2^n I_n - H_n \right)^{-1}$$  

$$= - [z_1, z_2] (z^n H_n)^{-1}. \tag{11b}$$

The first resolvent identity is based on the trivial observation that

$$(z_2^n I_n - H_n) - (z_1^n I_n - H_n) = (z_2 - z_1) I_n.$$
We set $z_n H_n := (z I_n - H_n)$. By the first resolvent identity (Chatelin, 1993)

$$
(z_1 H_n)^{-1} (z_2 H_n)^{-1} = (z_1 I_n - H_n)^{-1} (z_2 I_n - H_n)^{-1} = \frac{(z_1 H_n)^{-1} - (z_2 H_n)^{-1}}{z_2 - z_1} = -[z_1, z_2](z H_n)^{-1}.
$$

The first resolvent identity is based on the trivial observation that

$$(z_2 I_n - H_n) - (z_1 I_n - H_n) = (z_2 - z_1)I_n.$$

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

$$
\prod_{i=1}^{k} (z_i H_n)^{-1} = (-1)^{k-1} [z_1, \ldots, z_k](z H_n)^{-1}.
$$
We set $\mathbf{z}_n := (\mathbf{z}_n \mathbf{I}_n - \mathbf{H}_n)$. By the first resolvent identity (Chatelin, 1993)

$$
(\mathbf{z}_1 \mathbf{H}_n)^{-1} (\mathbf{z}_2 \mathbf{H}_n)^{-1} = (\mathbf{z}_1 \mathbf{I}_n - \mathbf{H}_n)^{-1} (\mathbf{z}_2 \mathbf{I}_n - \mathbf{H}_n)^{-1}
$$

$$
= \frac{(\mathbf{z}_1 \mathbf{H}_n)^{-1} - (\mathbf{z}_2 \mathbf{H}_n)^{-1}}{\mathbf{z}_2 - \mathbf{z}_1} = -[\mathbf{z}_1, \mathbf{z}_2](\mathbf{z} \mathbf{H}_n)^{-1}.
$$

The first resolvent identity is based on the trivial observation that

$$
(\mathbf{z}_2 \mathbf{I}_n - \mathbf{H}_n) - (\mathbf{z}_1 \mathbf{I}_n - \mathbf{H}_n) = (\mathbf{z}_2 - \mathbf{z}_1)\mathbf{I}_n.
$$

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

$$
\prod_{i=1}^{k} (\mathbf{z}_i \mathbf{H}_n)^{-1} = (-1)^{k-1} [\mathbf{z}_1, \ldots, \mathbf{z}_k](\mathbf{z} \mathbf{H}_n)^{-1}.
$$

Confluent divided differences are well-defined.
Simplification

For simplicity we assume that $H_n$ is unReduced.
Simplification

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

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

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

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

Polynomial vectors $\nu$ and $\tilde{\nu}$ are defined by

$$\nu(z) := \left( \frac{\chi_{j+1:n}(z)}{h_{j:n-1}} \right)^n_{j=1} \quad \text{and} \quad \tilde{\nu}(z) := \left( \frac{\chi_{1:j-1}(z)}{h_{1:j-1}} \right)^n_{j=1}. \quad (13)$$
Simplification

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

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

Polynomial vectors $\nu$ and $\tilde{\nu}$ are defined by

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

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

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

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

Polynomial vectors $\nu$ and $\tilde{\nu}$ are defined by

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

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

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

$$\chi_{i:j}(z) := \det (zH_{i:j}) = \det (zI_{j-i+1} - H_{i:j}).$$
For $z$ in the resolvent set

\[(zH_n)\nu(z) = \frac{\chi(z)}{h_{1:n-1}}e_1 \quad \Leftrightarrow \quad \frac{\nu(z)h_{1:n-1}}{\chi(z)} = (zH_n)^{-1}e_1, \quad (14a)\]

\[\tilde{\nu}(z)^T(zH_n) = e_n^T \frac{\chi(z)}{h_{1:n-1}} \quad \Leftrightarrow \quad \frac{h_{1:n-1}\tilde{\nu}(z)^T}{\chi(z)} = e_n^T(zH_n)^{-1}. \quad (14b)\]
Simplification

For $z$ in the resolvent set

$$(\mathbf{zH}_n)\mathbf{v}(z) = \frac{\chi(z)}{h_{1:n-1}}\mathbf{e}_1 \quad \Leftrightarrow \quad \frac{\mathbf{v}(z)h_{1:n-1}}{\chi(z)} = (\mathbf{zH}_n)^{-1}\mathbf{e}_1,$$

$$\mathbf{\tilde{v}}(z)^T(\mathbf{zH}_n) = \mathbf{e}_n^T \frac{\chi(z)}{h_{1:n-1}} \quad \Leftrightarrow \quad \frac{h_{1:n-1}\mathbf{\tilde{v}}(z)^T}{\chi(z)} = \mathbf{e}_n^T(\mathbf{zH}_n)^{-1}.$$  \hspace{1cm} (14a, 14b)

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

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

$$= (-1)^{k-1}[z_1, \ldots, z_k]\frac{\mathbf{v}(z)h_{1:n-1}}{\chi(z)}.$$  \hspace{1cm} (15, 16)
For $z$ in the resolvent set

$$
(\tilde{z}H_n)\nu(z) = \frac{\chi(z)}{h_{1:n-1}} e_1 \iff \frac{\nu(z)h_{1:n-1}}{\chi(z)} = (\tilde{z}H_n)^{-1} e_1, \quad (14a)
$$

$$
\tilde{\nu}(z)^T (\tilde{z}H_n) = e_n^T \frac{\chi(z)}{h_{1:n-1}} \iff \frac{h_{1:n-1} \tilde{\nu}(z)^T}{\chi(z)} = e_n^T (\tilde{z}H_n)^{-1}. \quad (14b)
$$

The repeated application of resolvents to $e_1$ results in

$$
\left(\prod_{i=1}^{k} (\tilde{z}_iH_n)^{-1}\right) e_1 = (-1)^{k-1}[z_1, \ldots, z_k](\tilde{z}H_n)^{-1} e_1 \quad (15)
$$

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

Note that $zI_n - \tilde{z}H_n = zI_n - (zI_n - H_n) = H_n$, i.e., $H_n(\tilde{z}H_n)^{-1} = z(\tilde{z}H_n)^{-1} - I_n$. 
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 $e_n^T$. 
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** $e_n^T$. For this method we have the iterates

$$v_{k+1} = \left( \prod_{i=1}^{k} (z_i H_n)^{-1} \right) e_1, \quad x_{k+1} = \frac{e_n^T H_n v_{k+1}}{e_n^T v_{k+1}},$$
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 $e_n^T$. For this method we have the iterates

$$v_{k+1} = \left( \prod_{i=1}^{k} (z_i H_n)^{-1} \right) e_1, \quad x_{k+1} = \frac{e_n^T H_n v_{k+1}}{e_n^T v_{k+1}},$$

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

$$x_{k+1} = \frac{e_n^T H_n \left( \prod_{i=1}^{k} (z_i H_n)^{-1} \right) e_1}{e_n^T \left( \prod_{i=1}^{k} (z_i H_n)^{-1} \right) e_1} = \frac{e_n^T \left( z_k I_n - (z_k H_n) \right) \left( \prod_{i=1}^{k} (z_i H_n)^{-1} \right) e_1}{e_n^T \left( \prod_{i=1}^{k} (z_i H_n)^{-1} \right) e_1} \quad (17a)$$

$$= z_k - \frac{e_n^T z_k H_n \left( \prod_{i=1}^{k} (z_i H_n)^{-1} \right) e_1}{e_n^T \left( \prod_{i=1}^{k} (z_i H_n)^{-1} \right) e_1} = z_k - \frac{e_n^T \left( \prod_{i=1}^{k-1} (z_i H_n)^{-1} \right) e_1}{e_n^T \left( \prod_{i=1}^{k} (z_i H_n)^{-1} \right) e_1} \quad (17b)$$

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

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

$$x_{k+1} = \tau + \left[ \tau, \ldots, \tau \right] \left( \frac{1}{\chi} \right) \left[ \tau, \ldots, \tau, \tau \right] \left( \frac{1}{\chi} \right) = \tau + k \left( \frac{1}{\chi} \right) \left( k-1 \right) \left( \tau \right) \left( \frac{1}{\chi} \right) \left( k \right) \left( \tau \right).$$

(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.
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 $\chi$.

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

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

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).
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 $\chi$.

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

$$x_{k+1} = \tau + \frac{[\tau, \ldots, \tau](1/\chi)}{[\tau, \ldots, \tau, \tau](1/\chi)} = \tau + k \frac{(1/\chi)^{(k-1)}(\tau)}{(1/\chi)^{(k)}(\tau)}.$$  \hspace{0.5cm} (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.
The original Rayleigh quotient iteration (Strutt, 1894) with the symmetric Rayleigh quotient and, because of the symmetry, a tridiagonal Hermitean Hessenberg matrix $H_n$, gives the update

$$z_{k+1} = \frac{e_1^T(z_k H_n)^{-1}H_n(z_k H_n)^{-1}e_1}{e_1^T(z_k H_n)^{-1}e_1} = \frac{e_1^T H_n(z_k H_n)^{-2}e_1}{e_1^T(z_k H_n)^{-2}e_1} \quad (19a)$$

$$= \frac{e_1^T(z_k I_n - z_k H_n)(z_k H_n)^{-2}e_1}{e_1^T(z_k H_n)^{-2}e_1} \quad (19b)$$

$$= z_k - \frac{e_1^T(z_k H_n)^{-1}e_1}{e_1^T(z_k H_n)^{-2}e_1} = z_k + \frac{[z_k] (\chi_{2:n} / \chi)}{[z_k, z_k] (\chi_{2:n} / \chi)} \quad (19c)$$

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

The original Rayleigh quotient iteration (Strutt, 1894) with the symmetric Rayleigh quotient and, because of the symmetry, a tridiagonal Hermitean Hessenberg matrix $H_n$, gives the update

$$z_{k+1} = \frac{e_1^T(z_k H_n)^{-1} H_n (z_k H_n)^{-1} e_1}{e_1^T(z_k H_n)^{-1} e_1} = \frac{e_1^T H_n (z_k H_n)^{-2} e_1}{e_1^T(z_k H_n)^{-2} e_1}$$

(19a)

$$= \frac{e_1^T(z_k I_n - z_k H_n)(z_k H_n)^{-2} e_1}{e_1^T(z_k H_n)^{-2} e_1}$$

(19b)

$$= z_k - \frac{e_1^T(z_k H_n)^{-1} e_1}{e_1^T(z_k H_n)^{-2} e_1} = z_k + \frac{[z_k] (\chi_{2:n}/\chi)}{\chi(z_k, z_k) (\chi_{2:n}/\chi)}$$

(19c)

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

(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.
Symmetric RQI for Hermitean matrices gives the update

$$z_{k+1} = z_k + \frac{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k]}{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k]}(\chi_2:n/\chi).$$

(20)
Symmetric RQI for Hermitean matrices gives the update

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

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

Symmetric RQI for Hermitean matrices gives the update

$$z_{k+1} = z_k + \frac{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k](\chi_{2:n}/\chi)}{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k, z_k](\chi_{2:n}/\chi)}.$$  (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.
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
Load applied to structure, \( K \in \mathbb{R}^{1092 \times 1092} \), 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)

IDR(16) for a matrix of size 1092 x 1092

---

Jens-Peter M. Zemke
Krylov @ TUHH Kickoff 2012
2012-07-18
Load applied to structure, $K \in \mathbb{R}^{1092 \times 1092}$, IDR(32)

(IDR(32) for a matrix of size 1092 x 1092)

- GMRes convergence
- Estimated residual MR
- True residual MR
- True error MR
- Estimated residual OR
- True residual OR
- True error OR
- MR by OR

![Graph showing residuals in log-scale over steps or number of matrix-vector multiplications](image-url)
Load applied to structure, $K \in \mathbb{R}^{1092 \times 1092}$, IDR(64)
Shifted Grcar matrix; IDR(1)

Sonneveld Ritz for shifted Grcar matrix, IDR(1)

OR/MR for shifted Grcar matrix, IDR(1)

- eigenvalues
- Sonneveld values
- Ritz values
- seed values

estimated residual MR
true residual MR
true error MR
estimated residual OR
true residual OR
true error OR
MR by OR
Shifted Grcar matrix; IDR(2)

Sonneveld Ritz for shifted Grcar matrix, IDR(2)

- eigenvalues
- Sonneveld values
- Ritz values
- seed values

OR/MR for shifted Grcar matrix, IDR(2)

- estimated residual MR
- true residual MR
- true error MR
- estimated residual OR
- true residual OR
- true error OR
- MR by OR
Shifted Grcar matrix; IDR(4)

- Sonneveld Ritz for shifted Grcar matrix, IDR(4)
  - eigenvalues
  - Sonneveld values
  - Ritz values
  - seed values

- OR/MR for shifted Grcar matrix, IDR(4)
  - estimated residual MR
  - true residual MR
  - true error MR
  - estimated residual OR
  - true residual OR
  - true error OR
  - MR by OR
Shifted Grcar matrix; IDR(8)

Sonneveld Ritz for shifted Grcar matrix, IDR(8)

OR/MR for shifted Grcar matrix, IDR(8)

- eigenvalues
- Sonneveld values
- Ritz values
- seed values

estimated residual MR
true residual MR
true error MR
estimated residual OR
true residual OR
true error OR
MR by OR
Shifted Grcar matrix; IDR(16)

Sonneveld Ritz for shifted Grcar matrix, IDR(16)

OR/MR for shifted Grcar matrix, IDR(16)
Shifted Grcar matrix; IDR(32)

- Sonneveld Ritz for shifted Grcar matrix, IDR(32)
  - + eigenvalues
  - O Sonneveld values
  - X Ritz values
  - X seed values

- OR/MR for shifted Grcar matrix, IDR(32)
  - red: estimated residual MR
  - black: true residual MR
  - blue: true error MR
  - dotted red: estimated residual OR
  - dotted black: true residual OR
  - dotted blue: true error OR
  - magenta: MR by OR
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
Every observed behaviour that occurs in a perturbed method can also be observed in unperturbed methods w/ orthonormal basis vectors.
Every observed behaviour that occurs in a perturbed method can also be observed in unperturbed methods with orthonormal basis vectors.

Hessenberg decomposition:

\[ H_n I_{n,k} = I_{n,k+1} H_k. \]
Every observed behaviour that occurs in a perturbed method can also be observed in unperturbed methods with orthonormal basis vectors.

Hessenberg decomposition:

\[ H_n I_{n,k} = I_{n,k+1} H_k. \]

Generalized Hessenberg decomposition:

\[ (H_n U_n^{-1}) I_{n,k} U_k = I_{n,k+1} H_k. \]
Every observed behaviour that occurs in a perturbed method can also be observed in unperturbed methods with orthonormal basis vectors.

Hessenberg decomposition:

\[ H_n I_{n,k} = I_{n,k+1} H_k. \]

Generalized Hessenberg decomposition:

\[ (H_n U_n^{-1}) I_{n,k} U_k = I_{n,k+1} H_k. \]

Bad news: Impossible to distinguish effects of perturbation from startling behaviour due to strange data.
Suppose that
\[ AQ_k + F_k = Q_{k+1} T_k, \quad A^H = A, \quad T_k^H = T_k. \]
Suppose that
\[ AQ_k + F_k = Q_{k+1} T_k, \quad A^\text{H} = A, \quad T_k^\text{H} = T_k. \]

Set
\[ \text{diag}(T_k, A) := \begin{pmatrix} T_k & O_{k,n} \\ O_{n,k} & A \end{pmatrix} \in \mathbb{C}^{(k+n) \times (k+n)}, \quad T_k \in \mathbb{C}^{k \times k}, \quad A \in \mathbb{C}^{n \times n}. \]
Suppose that
\[ AQ_k + F_k = Q_{k+1} T_k, \quad A^H = A, \quad T_k^H = T_k. \]

Set
\[ \text{diag}(T_k, A) := \begin{pmatrix} T_k & O_{k,n} \\ O_{n,k} & A \end{pmatrix} \in \mathbb{C}^{(k+n) \times (k+n)}, \quad T_k \in \mathbb{C}^{k \times k}, \quad A \in \mathbb{C}^{n \times n}. \]

Paige used \textit{augmented backward error analysis} for symmetric Lanczos in finite precision:
\[ (\text{diag}(T_k, A) + H) \tilde{Q}_k = \tilde{Q}_{k+1} T_k, \quad \tilde{Q}_k^H \tilde{Q}_k = I_k. \]
Suppose that
\[
AQ_k + F_k = Q_{k+1}T_k, \quad A^H = A, \quad T_k^H = T_k.
\]
Set
\[
\text{diag}(T_k, A) := \begin{pmatrix} T_k & O_{k,n} \\ O_{n,k} & A \end{pmatrix} \in \mathbb{C}^{(k+n) \times (k+n)}, \quad T_k \in \mathbb{C}^{k \times k}, \quad A \in \mathbb{C}^{n \times n}.
\]
Paige used augmented backward error analysis for symmetric Lanczos in finite precision:
\[
(\text{diag}(T_k, A) + H) \tilde{Q}_k = \tilde{Q}_{k+1}T_k, \quad \tilde{Q}_k^H \tilde{Q}_k = I_k.
\]
Here, \(H\) is a “small” perturbation if \(F_k\) is small and local orthonormality is given. Error-free process for perturbed strange matrix.
Analysis of perturbed Krylov subspace methods

Suppose that
\[
AQ_k + F_k = Q_{k+1}T_k, \quad A^H = A, \quad T_k^H = T_k.
\]
Set
\[
\text{diag}(T_k, A) := \begin{pmatrix} T_k & O_{k,n} \\ O_{n,k} & A \end{pmatrix} \in \mathbb{C}^{(k+n) \times (k+n)}, \quad T_k \in \mathbb{C}^{k \times k}, \quad A \in \mathbb{C}^{n \times n}.
\]

Paige used augmented backward error analysis for symmetric Lanczos in finite precision:
\[
(\text{diag}(T_k, A) + H) \tilde{Q}_k = \tilde{Q}_{k+1}T_k, \quad \tilde{Q}_k^H\tilde{Q}_k = I_k.
\]

Here, \( H \) is a “small” perturbation if \( F_k \) 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.
I sketched the **three main families** of Krylov subspace methods.
I sketched the three main families of Krylov subspace methods. I highlighted the rôle of Hessenberg matrices and the resulting structure.
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.
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 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 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,


*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.


Iterative procedures related to relaxation methods for eigenvalue problems.


The shifted $QR$ algorithm for Hermitian matrices.


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).

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.

Schröder, E. (1870).
Ueber unendlich viele Algorithmen zur Auflösung der Gleichungen.

Schweins, F. (1825).
Theorie der Differenzen und Differentiale.


Hessenberg eigenvalue–eigenmatrix relations.  

Abstract perturbed Krylov methods.  