Towards a deeper understanding of Chris Paige’s error analysis of the finite precision Lanczos process

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

Institut für Numerische Simulation
Technische Universität Hamburg-Harburg

ICS of CAS, September 24th, 2009
Some history

Hessenberg matrices
  Hessenberg decompositions
  Hessenberg eigenvectors

Chris Paige’s approach
  On the length of the Ritz vectors
  Eigenvector sensitivity
  Closer to the original

Our approach
  The shifted decomposition
  About higher derivatives
  The polynomial point of view

His results form the basis of Anne Greenbaum’s celebrated “backward error analysis” (Greenbaum, 1989) of the finite-precision symmetric Lanczos and CG methods, compare with (Greenbaum and Strakoš, 1992).

For an introduction and a general exposition especially on the finite-precision symmetric Lanczos and CG methods see also (Meurant, 2006; Meurant and Strakoš, 2006).

Thus far, this is maybe the only successful error analysis ever carried out for a perturbed short-term Krylov subspace method.
Following his seminal PhD thesis (Paige, 1971), Chris Paige published a sequence of papers (Paige, 1972; Paige, 1976; Paige, 1980) on the error analysis of the finite-precision behavior of the symmetric Lanczos process. His results form the basis of Anne Greenbaum’s celebrated “backward error analysis” (Greenbaum, 1989) of the finite-precision symmetric Lanczos and CG methods, compare with (Greenbaum and Strakoš, 1992).
Following his seminal PhD thesis (Paige, 1971), Chris Paige published a sequence of papers (Paige, 1972; Paige, 1976; Paige, 1980) on the error analysis of the finite-precision behavior of the symmetric Lanczos process. His results form the basis of Anne Greenbaum’s celebrated “backward error analysis” (Greenbaum, 1989) of the finite-precision symmetric Lanczos and CG methods, compare with (Greenbaum and Strakoš, 1992).

For an introduction and a general exposition especially on the finite-precision symmetric Lanczos and CG methods see also (Meurant, 2006; Meurant and Strakoš, 2006).

His results form the basis of Anne Greenbaum’s celebrated “backward error analysis” (Greenbaum, 1989) of the finite-precision symmetric Lanczos and CG methods, compare with (Greenbaum and Strakoš, 1992).

For an introduction and a general exposition especially on the finite-precision symmetric Lanczos and CG methods see also (Meurant, 2006; Meurant and Strakoš, 2006).

Thus far, this is maybe the only successful error analysis ever carried out for a perturbed short-term Krylov subspace method.
An example: Lanczos’ method

We used the diagonal matrix

\[ A = \text{diag}([\text{linspace}(0,1,50),3]) \]

and the starting vector

\[ e = \text{ones}(51,1) \]

in an implementation of Lanczos’ method in MATLAB on a PC conforming to ANSI/IEEE 754 with machine precision \( \text{eps}(1) = 2^{-52} \approx 2.2204 \times 10^{-16} \).
An example: Lanczos’ method

We used the diagonal matrix

$$A = \text{diag}([\text{linspace}(0,1,50),3])$$

and the starting vector

$$e = \text{ones}(51,1)$$

in an implementation of Lanczos’ method in MATLAB on a PC conforming to ANSI/IEEE 754 with machine precision $\text{eps}(1) = 2^{-52} \approx 2.2204 \cdot 10^{-16}$.

At step 10 the first Ritz value has converged (up to machine precision) to the eigenvalue 3, at step 27 the second one has converged. Detoriation reaches a maximum at step $19 = \lceil (10 + 27)/2 \rceil$. 
An example: Lanczos’ method

We used the diagonal matrix

\[
A = \text{diag}([\text{linspace}(0,1,50),3])
\]

and the starting vector

\[
e = \text{ones}(51,1)
\]

in an implementation of Lanczos’ method in MATLAB on a PC conforming to ANSI/IEEE 754 with machine precision \( \varepsilon_1 = 2^{-52} \approx 2.2204 \cdot 10^{-16} \).

At step 10 the first Ritz value has converged (up to machine precision) to the eigenvalue 3, at step 27 the second one has converged. Detoriation reaches a maximum at step \( \lceil (10 + 27)/2 \rceil \).

Eigenvalues and eigenvectors are computed using MRRR, i.e., LAPACK’s routine \texttt{DSTEGR}, since MATLAB’s \texttt{eig} (using LAPACK’s \texttt{DSYEV}, i.e., the QR algorithm implemented as \texttt{DSTEQR}) fails in delivering accurate eigenvectors.
An example: Lanczos’ method

We used the diagonal matrix

\[ A = \text{diag}([\text{linspace}(0,1,50),3]) \]

and the starting vector

\[ e = \text{ones}(51,1) \]

in an implementation of Lanczos’ method in MATLAB on a PC conforming to ANSI/IEEE 754 with machine precision \( \epsilon_1 = 2^{-52} \approx 2.2204 \cdot 10^{-16} \).

At step 10 the first Ritz value has converged (up to machine precision) to the eigenvalue 3, at step 27 the second one has converged. Detoriation reaches a maximum at step \( 19 = \lceil (10 + 27)/2 \rceil \).

Eigenvalues and eigenvectors are computed using MRRR, i.e., LAPACK’s routine \text{DSTEGR}, since MATLAB’s \text{eig} (using LAPACK’s \text{DSYEV}, i.e., the QR algorithm implemented as \text{DSTEQR}) fails in delivering accurate eigenvectors. Additionally, we heavily used the symbolic toolbox, i.e., MAPLE.
The finite precision behavior

comparison of 29 steps of symbolic and floating point Lanczos

- symbolic Ritz values
- floating point Ritz values
- eigenvalues of symmetric $A$
The finite precision behavior

Floating point Lanczos characteristics

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

Graph showing the distance to eigenvalue 3 over step number with various lines representing different characteristics.
Outline

Some history

Hessenberg matrices

Hessenberg decompositions

Hessenberg eigenvectors

Chris Paige’s approach

On the length of the Ritz vectors

Eigenvector sensitivity

Closer to the original

Our approach

The shifted decomposition

About higher derivatives

The polynomial point of view
Essential features of Krylov subspace methods can be described by a Hessenberg decomposition

\[ AQ_k = Q_{k+1} H_k = Q_k H_k + q_{k+1} h_{k+1,k} e_k^T. \]  

(1)

Here, \( H_k \) denotes an unreduced Hessenberg matrix.
Hessenberg decompositions

Essential features of Krylov subspace methods can be described by a Hessenberg decomposition

$$AQ_k = Q_{k+1}H_k = Q_kH_k + q_{k+1}h_{k+1,k}e_k^T.$$ (1)

Here, $H_k$ denotes an unreduced Hessenberg matrix.

In the perturbed case, e.g., in finite precision and/or based on inexact matrix-vector multiplies, we obtain a perturbed Hessenberg decomposition

$$AQ_k + F_k = Q_{k+1}H_k = Q_kH_k + q_{k+1}h_{k+1,k}e_k^T.$$ (2)
Essential features of Krylov subspace methods can be described by a Hessenberg decomposition

\[ AQ_k = Q_{k+1} H_k = Q_k H_k + q_{k+1} h_{k+1,k} e_k^T. \]  

Here, \( H_k \) denotes an unreduced Hessenberg matrix.

In the perturbed case, e.g., in finite precision and/or based on inexact matrix-vector multiplies, we obtain a perturbed Hessenberg decomposition

\[ AQ_k + F_k = Q_{k+1} H_k = Q_k H_k + q_{k+1} h_{k+1,k} e_k^T. \]  

The matrix \( H_k \) of the perturbed variant will, in general, still be unreduced.
In (Z, 2007) we did consider in an abstract manner the matrix equation

\[ \mathbf{A} \mathbf{Q}_k + \mathbf{F}_k = \mathbf{Q}_{k+1} \mathbf{H}_k = \mathbf{Q}_k \mathbf{H}_k + \mathbf{q}_{k+1} h_{k+1,k} \mathbf{e}_k^T \]  

(3)

and came up with polynomial expressions in \( \mathbf{A} \) for

- the basis vectors \( \mathbf{q}_j \),
- the Ritz vectors \( \mathbf{y}_j := \mathbf{Q}_k \mathbf{s}_j \), where \( \mathbf{s}_j \) is an eigenvector of \( \mathbf{H}_k \) to the eigenvalue \( \theta_j \),
- and the angles between Ritz vectors and eigenvectors of \( \mathbf{A} \).
In (Z, 2007) we did consider in an abstract manner the matrix equation

\[ A Q_k + F_k = Q_{k+1} H_k = Q_k H_k + q_{k+1} h_{k+1,k} e_k^T \]  

and came up with polynomial expressions in \( A \) for

- the basis vectors \( q_j \),
- the Ritz vectors \( y_j := Q_k s_j \), where \( s_j \) is an eigenvector of \( H_k \) to the eigenvalue \( \theta_j \),
- and the angles between Ritz vectors and eigenvectors of \( A \).

The results were based on eigenvalue–eigenmatrix relations (Z, 2006).
In (Z, 2007) we did consider in an abstract manner the matrix equation

\[ AQ_k + F_k = Q_{k+1}H_k = Q_kH_k + q_{k+1}h_{k+1,k}e_k^T \]  

(3)

and came up with polynomial expressions in \( A \) for

- the basis vectors \( q_j \),
- the Ritz vectors \( y_j := Q_k s_j \), where \( s_j \) is an eigenvector of \( H_k \) to the eigenvalue \( \theta_j \),
- and the angles between Ritz vectors and eigenvectors of \( A \).

The results were based on eigenvalue–eigenmatrix relations (Z, 2006).

This talk: Application to the (symmetric) Lanczos process (in finite precision); Aim: generalize (Paige, 1971; Paige, 1972; Paige, 1976; Paige, 1980) to the general (non-symmetric) Lanczos process (with general perturbations).
In case of the symmetric Lanczos process we have

\[ AQ_k + F_k = Q_{k+1} T_k = Q_k T_k + q_{k+1} \beta_k e_k^T, \]  

(4)
Hessenberg decompositions

In case of the symmetric Lanczos process we have

$$A Q_k + F_k = Q_{k+1} T_k = Q_k T_k + q_{k+1} \beta_k e_k^T,$$

where

$\triangleright \quad A = A^H \in \mathbb{C}^{n \times n}$ is Hermitean,
In case of the symmetric Lanczos process we have

\[ A Q_k + F_k = Q_{k+1} T_k = Q_k T_k + q_{k+1} \beta_k e_k^T, \]  

(4)

where

- \( A = A^H \in \mathbb{C}^{n \times n} \) is Hermitean,
- \( T_k = T_k^T \in \mathbb{R}^{k \times k} \) is unreduced tridiagonal symmetric,
In case of the symmetric Lanczos process we have

\[ AQ_k + F_k = Q_{k+1} T_k = Q_k T_k + q_{k+1} \beta_k e_k^T, \]

where

- \( A = A^H \in \mathbb{C}^{n \times n} \) is Hermitian,
- \( T_k = T_k^T \in \mathbb{R}^{k \times k} \) is unreduced tridiagonal symmetric,
- \( F_k \in \mathbb{C}^{n \times k} \) is “small”.

The elements of the tridiagonal matrix \( T_k \) are denoted by

\[
\begin{pmatrix}
\alpha_1 & \beta_1 & 0 & \cdots & 0 \\
\beta_1 & \alpha_2 & \beta_2 & \cdots & 0 \\
0 & \beta_2 & \alpha_3 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \beta_{k-1} & \beta_k \\
0 & 0 & 0 & \cdots & \beta_k & \beta_{k+1}
\end{pmatrix},
\]

where \( \beta_j > 0 \) for all \( 1 \leq j \leq k \).

(If off-diagonal elements were negative, impose diagonal scaling.)
Hessenberg matrices

Hessenberg decompositions

In case of the symmetric Lanczos process we have

$$AQ_k + F_k = Q_{k+1}T_k = Q_kT_k + q_{k+1}\beta_ke_k^T,$$

where

- $A = A^H \in \mathbb{C}^{n \times n}$ is Hermitean,
- $T_k = T_k^T \in \mathbb{R}^{k \times k}$ is unreduced tridiagonal symmetric,
- $F_k \in \mathbb{C}^{n \times k}$ is “small”.

The elements of the tridiagonal matrix $T_k$ are denoted by

$$T_k = \begin{pmatrix}
\alpha_1 & \beta_1 & & \\
\beta_1 & \alpha_2 & \ddots & \\
& \ddots & \ddots & \beta_{k-1} \\
& & \beta_{k-1} & \alpha_k
\end{pmatrix}, \quad \beta_j > 0 \quad \forall \ 1 \leq j \leq k.$$

(5)
Hessenberg matrices

Hessenberg decompositions

In case of the symmetric Lanczos process we have

\[ AQ_k + F_k = Q_{k+1} T_k = Q_k T_k + q_{k+1} \beta_k e_k^\top, \]

where

- \( A = A^H \in \mathbb{C}^{n \times n} \) is Hermitian,
- \( T_k = T_k^\top \in \mathbb{R}^{k \times k} \) is unreduced tridiagonal symmetric,
- \( F_k \in \mathbb{C}^{n \times k} \) is “small”.

The elements of the tridiagonal matrix \( T_k \) are denoted by

\[
T_k = \begin{pmatrix}
\alpha_1 & \beta_1 \\
\beta_1 & \alpha_2 & & \\
& \ddots & \ddots & \\
& & \ddots & \beta_{k-1} \\
& & & \beta_{k-1} & \alpha_k
\end{pmatrix}, \quad \beta_j > 0 \quad \forall \ 1 \leq j \leq k.
\]

(If off-diagonal elements were negative, impose diagonal scaling.)
Hessenberg decompositions

We encounter four Hessenberg decompositions in this talk. The first two are based on knowledge of $\mathbf{A}$.
We encounter four Hessenberg decompositions in this talk. The first two are based on knowledge of $A$.

The first one is the **original Lanczos Hessenberg decomposition**

$$AQ_k + F_k = Q_{k+1}T_k = Q_kT_k + q_{k+1}\beta_ke_k^T. \quad (\text{HessA1})$$
Hessenberg decompositions

We encounter four Hessenberg decompositions in this talk. The first two are based on knowledge of $A$.

The first one is the original Lanczos Hessenberg decomposition

$$AQ_k + F_k = Q_{k+1}T_k = Q_kT_k + q_{k+1}\beta_ke_k^T.$$  \hfill (HessA1)

With a given eigenpair $v_i^HA = \lambda_i v_i^H$ and a given Ritz value $\theta_j$ we define

$$\tilde{A} := A - (\lambda_i - \theta_j)\frac{v_i v_i^H}{v_i^H v_i} \quad \text{and} \quad \tilde{F}_k := (\lambda_i - \theta_j)\frac{v_i v_i^H}{v_i^H v_i}Q_k + F_k.$$  \hfill (6)
Hessenberg decompositions

We encounter four Hessenberg decompositions in this talk. The first two are based on knowledge of $A$.

The first one is the original Lanczos Hessenberg decomposition

$$A Q_k + F_k = Q_{k+1} T_k = Q_k T_k + q_{k+1} \beta_k e_k^T. \quad (\text{HessA1})$$

With a given eigenpair $v_i^\mathsf{H} A = \lambda_i v_i^\mathsf{H}$ and a given Ritz value $\theta_j$ we define

$$\tilde{A} := A - (\lambda_i - \theta_j) \frac{v_i v_i^\mathsf{H}}{v_i^\mathsf{H} v_i} \quad \text{and} \quad \tilde{F}_k := (\lambda_i - \theta_j) \frac{v_i v_i^\mathsf{H}}{v_i^\mathsf{H} v_i} Q_k + F_k. \quad (6)$$

Then we obtain the shifted Lanczos Hessenberg decomposition

$$\tilde{A} Q_k + \tilde{F}_k = Q_{k+1} T_k = Q_k T_k + q_{k+1} \beta_k e_k^T. \quad (\text{HessA2})$$
Hessenberg decompositions

We encounter four Hessenberg decompositions in this talk. The first two are based on knowledge of $\mathbf{A}$.

The first one is the original Lanczos Hessenberg decomposition

$$\mathbf{AQ}_k + \mathbf{F}_k = \mathbf{Q}_{k+1} \mathbf{T}_k = \mathbf{Q}_k \mathbf{T}_k + q_{k+1} \beta_k e_k^T. \quad \text{(HessA1)}$$

With a given eigenpair $v_i^H \mathbf{A} = \lambda_i v_i^H$ and a given Ritz value $\theta_j$ we define

$$\tilde{\mathbf{A}} := \mathbf{A} - (\lambda_i - \theta_j) \frac{v_i v_i^H}{v_i^H v_i} \quad \text{and} \quad \tilde{\mathbf{F}}_k := (\lambda_i - \theta_j) \frac{v_i v_i^H}{v_i^H v_i} \mathbf{Q}_k + \mathbf{F}_k. \quad \text{(6)}$$

Then we obtain the shifted Lanczos Hessenberg decomposition

$$\tilde{\mathbf{A}} \mathbf{Q}_k + \tilde{\mathbf{F}}_k = \mathbf{Q}_{k+1} \mathbf{T}_k = \mathbf{Q}_k \mathbf{T}_k + q_{k+1} \beta_k e_k^T. \quad \text{(HessA2)}$$

This Hessenberg decomposition is interesting especially in case that $\lambda_i - \theta_j$ is “small”, i.e., “comparable” to $\mathbf{F}_k$. 
The next two Hessenberg decompositions are based on $T_k$ in place of $A$. These form the essential part of Chris Paige’s analysis.
The next two Hessenberg decompositions are based on $T_k$ in place of $A$. These form the essential part of Chris Paige’s analysis.

Let $W_{k+1} := Q_k^H Q_{k+1}$, define $G_k := e_k \beta_k q_{k+1}^H Q_k + Q_k^H F_k - F_k^H Q_k$. Then

$$T_k W_k + G_k = W_{k+1} T_k = W_k T_k + w_{k+1} \beta_k e_k^T.$$  \hspace{1cm} (HessT1)
The next two Hessenberg decompositions are based on $T_k$ in place of $A$. These form the essential part of Chris Paige’s analysis.

Let $W_{k+1} := Q_k^H Q_{k+1}$, define $G_k := e_k \beta_k q_{k+1}^H Q_k + Q_k^H F_k - F_k^H Q_k$. Then

$$T_k W_k + G_k = W_{k+1} T_k = W_k T_k + w_{k+1} \beta_k e_k^T. \quad \text{(HessT1)}$$

The fourth Hessenberg decomposition, mainly used by Chris Paige, is based on an additive splitting of $W_{k+1}$. 
Hessenberg decompositions

The next two Hessenberg decompositions are based on $T_k$ in place of $A$. These form the essential part of Chris Paige’s analysis.

Let $W_{k+1} := Q_k^H Q_{k+1}$, define $G_k := e_k \beta_k q_{k+1}^H Q_k + Q_k^H F_k - F_k^H Q_k$. Then

$$T_k W_k + G_k = W_{k+1} T_k = W_k T_k + w_{k+1} \beta_k e_k^T.$$  \hspace{1cm} (HessT1)

The fourth Hessenberg decomposition, mainly used by Chris Paige, is based on an additive splitting of $W_{k+1}$.

Let $W_{k+1} = R_k^H + D_k + R_{k+1}$ with $R_{k+1} = \text{sut}(W_{k+1})$ and $D_k$ diagonal. Then

$$T_k R_k + E_k = R_{k+1} T_k = R_k T_k + r_{k+1} \beta_k e_k^T.$$ \hspace{1cm} (HessT2)

with $E_k$ upper triangular and small if $F_k$ is small and local orthogonality is preserved.
Outline

Some history

Hessenberg matrices

Hessenberg decompositions

Hessenberg eigenvectors

Chris Paige’s approach

On the length of the Ritz vectors

Eigenvector sensitivity

Closer to the original

Our approach

The shifted decomposition

About higher derivatives

The polynomial point of view
According to (Z, 2006) we can describe the eigenvectors (and principal vectors) of Hessenberg matrices in terms of certain polynomial vectors.
According to \((Z, 2006)\) we can describe the eigenvectors (and principal vectors) of Hessenberg matrices in terms of certain polynomial vectors.

We have that
\[
(\tilde{\nu}(z)^T = \tilde{\nu}(z)^H)
\]
\[
(zI_k - T_k)\nu(z) = e_1 \frac{\chi(z)}{\beta_{1:k-1}}, \quad \tilde{\nu}(z)^T(zI_k - T_k) = \frac{\chi(z)}{\beta_{1:k-1}}e_n^T
\]

with \(\chi(z) := \det (zI_k - T_k)\) and \(\beta_{1:k-1} := \prod_{j=1}^{k-1} \beta_j > 0\).
Hessenberg eigenvectors and eigenvector derivatives

According to \((Z, 2006)\) we can describe the eigenvectors (and principal vectors) of Hessenberg matrices in terms of certain polynomial vectors. We have that

\[
\begin{align*}
(\mathbf{zI}_k - \mathbf{T}_k) \mathbf{\nu}(z) &= \mathbf{e}_1 \frac{\chi(z)}{\beta_{1:k-1}}, \\
\mathbf{\tilde{\nu}}(z)^T (\mathbf{zI}_k - \mathbf{T}_k) &= \frac{\chi(z)}{\beta_{1:k-1}} \mathbf{e}_n^T
\end{align*}
\] (7)

with \(\chi(z) := \det(\mathbf{zI}_k - \mathbf{T}_k)\) and \(\beta_{1:k-1} := \prod_{j=1}^{k-1} \beta_j > 0\).

Inner products between the left and right eigenvector polynomials are given by

\[
\begin{align*}
\mathbf{\tilde{\nu}}(z)^H \mathbf{\nu}(w) &= \frac{\chi[z, w]}{\beta_{1:k-1}^2} = \frac{1}{\beta_{1:k-1}} \begin{cases} \\
\frac{\chi(z) - \chi(w)}{z - w}, & z \neq w \\
\frac{\chi'(z)}{\chi(z)}, & z = w.
\end{cases}
\end{align*}
\] (8)
According to (Z, 2006) we can describe the eigenvectors (and principal vectors) of Hessenberg matrices in terms of certain polynomial vectors.

We have that

\[
\hat{\nu}(z)^T = \hat{\nu}(z)^H
\]

\[
(zI_k - T_k)\nu(z) = e_1 \frac{\chi(z)}{\beta_{1:k-1}}, \quad \tilde{\nu}(z)^T(zI_k - T_k) = \frac{\chi(z)}{\beta_{1:k-1}} e_n^T
\]

(7)

with \( \chi(z) := \det (zI_k - T_k) \) and \( \beta_{1:k-1} := \prod_{j=1}^{k-1} \beta_j > 0 \).

Inner products between the left and right eigenvector polynomials are given by

\[
\hat{\nu}(z)^H \nu(w) = \frac{\chi[z, w]}{\beta_{1:k-1}} = \frac{1}{\beta_{1:k-1}} \begin{cases} 
\frac{\chi(z) - \chi(w)}{z - w}, & z \neq w \\
\chi'(z), & z = w.
\end{cases}
\]

(8)

In (Z, 2006) we used differentiation and the above relations to construct eigenvectors and corresponding principal vectors.
Hessenberg eigenvectors and eigenvector derivatives

In case of Hermitean/symmetric matrices $A$ and $T_k$ we know that the left and right eigenvector are parallel and can be scaled to unit length.
In case of Hermitian/symmetric matrices $A$ and $T_k$ we know that the left and right eigenvector are parallel and can be scaled to unit length.

Thus, the eigenvector of $A$ used in the shifted Lanczos Hessenberg decomposition is assumed to have unit length, $\|v_i\|_2 = 1$. 
In case of Hermitean/symmetric matrices $A$ and $T_k$ we know that the left and right eigenvector are parallel and can be scaled to unit length.

Thus, the eigenvector of $A$ used in the shifted Lanczos Hessenberg decomposition is assumed to have unit length, $\|v_i\|_2 = 1$.

The right and left eigenvectors $v_j := v(\theta_j)$ and $\tilde{v}_j := \tilde{v}(\theta_j)$ are parallel and non-zero in the first and last entry, as

$$v(z) := \left(\frac{\chi_{j+1:k}(z)}{\beta_{j:k-1}}\right)_{j=1}^k \quad \text{and} \quad \tilde{v}(z) := \left(\frac{\chi_{1:j-1}(z)}{\beta_{1:j-1}}\right)_{j=1}^k,$$

(9)

where

$$\chi_{i:j}(z) := \det(zI_{j-i+1} - T_{i:j}) \quad \text{and} \quad \beta_{i:j} := \prod_{\ell=i}^j \beta_\ell, \quad 0 \leq i \leq j < k.$$

(10)
In case of Hermitean/symmetric matrices $A$ and $T_k$ we know that the left and right eigenvector are parallel and can be scaled to unit length.

Thus, the eigenvector of $A$ used in the shifted Lanczos Hessenberg decomposition is assumed to have unit length, $\|v_i\|_2 = 1$.

The right and left eigenvectors $\nu_j := \nu(\theta_j)$ and $\tilde{\nu}_j := \tilde{\nu}(\theta_j)$ are parallel and non-zero in the first and last entry, as

$$
\nu(z) := \left( \frac{\chi_{j+1:k}(z)}{\beta_{j:k-1}} \right)_{j=1}^k \quad \text{and} \quad \tilde{\nu}(z) := \left( \frac{\chi_{1:j-1}(z)}{\beta_{1:j-1}} \right)_{j=1}^k,
$$

where

$$
\chi_{i:j}(z) := \det(zI_{j-i+1} - T_{i:j}) \quad \text{and} \quad \beta_{i:j} := \prod_{\ell=i}^{j} \beta_{\ell}, \quad 0 \leq i \leq j < k.
$$

To be more precise: $\nu_k(z) \equiv 1 \equiv \tilde{\nu}_1(z)$. 

Unit length eigenvectors $s_j$ of $T_k$ to the eigenvalue $\theta_j$ are defined by

$$s_j := \frac{\nu_j}{\|\nu_j\|_2}. \quad (11)$$
Unit length eigenvectors \( s_j \) of \( T_k \) to the eigenvalue \( \theta_j \) are defined by

\[
s_j := \frac{\nu_j}{\|\nu_j\|_2}.
\]  

(11)

This ensures that the last component \( s_{kj} \) of \( s_j \) is positive and given by

\[
s_{kj} = \frac{1}{\|\nu_j\|_2} = \frac{1}{\|\nu(\theta_j)\|_2} > 0.
\]  

(12)
Unit length eigenvectors $s_j$ of $T_k$ to the eigenvalue $\theta_j$ are defined by

$$s_j := \frac{\nu_j}{\|\nu_j\|_2}. \quad (11)$$

This ensures that the last component $s_{kj}$ of $s_j$ is positive and given by

$$s_{kj} = \frac{1}{\|\nu_j\|_2} = \frac{1}{\|\nu(\theta_j)\|_2} > 0. \quad (12)$$

In case of an error-free process we have with the Ritz vector $y_j := Q_k s_j$ the (backward- and forward-error) bound

$$\min_\lambda |\lambda - \theta_j| \leq \frac{\|Ay_j - y_j \theta_j\|_2}{\|y_j\|_2} = \beta_k s_{kj}. \quad (13)$$
Unit length eigenvectors \( s_j \) of \( T_k \) to the eigenvalue \( \theta_j \) are defined by

\[
s_j := \frac{\nu_j}{\|\nu_j\|_2}.
\] (11)

This ensures that the last component \( s_{kj} \) of \( s_j \) is positive and given by

\[
s_{kj} = \frac{1}{\|\nu_j\|_2} = \frac{1}{\|\nu(\theta_j)\|_2} > 0.
\] (12)

In case of an error-free process we have with the Ritz vector \( y_j := Q_k s_j \) the (backward- and forward-error) bound

\[
\min_{\lambda} |\lambda - \theta_j| \leq \frac{\|Ay_j - y_j \theta_j\|_2}{\|y_j\|_2} = \beta_k s_{kj}.
\] (13)
Outline

Some history

Hessenberg matrices
  Hessenberg decompositions
  Hessenberg eigenvectors

Chris Paige’s approach
  On the length of the Ritz vectors
  Eigenvector sensitivity
  Closer to the original

Our approach
  The shifted decomposition
  About higher derivatives
  The polynomial point of view
Chris Paige bounded the deviation of $\|y_j\|_2$ from one by something of the form
\begin{equation}
|\|y_j\|_2^2 - 1| \leq \frac{O(F_k)}{\min_{\ell \neq j} |\theta_j - \theta_\ell|}.
\end{equation}
Chris Paige’s approach

Chris Paige bounded the deviation of $\|y_j\|_2$ from one by something of the form

$$|\|y_j\|_2^2 - 1| \leq \frac{O(F_k)}{\min_{\ell \neq j} |\theta_j - \theta_\ell|}.$$ (14)

The length of the Ritz vector $y_j$ is close to one as long as the perturbation term is small and no other Ritz value is close to $\theta_j$. 
Chris Paige’s approach

Chris Paige bounded the deviation of \( \|y_j\|_2 \) from one by something of the form

\[
\|y_j\|_2^2 - 1 \leq \frac{O(F_k)}{\min_{\ell \neq j} |\theta_j - \theta_\ell|}. \tag{14}
\]

The length of the Ritz vector \( y_j \) is close to one as long as the perturbation term is small and no other Ritz value is close to \( \theta_j \).

People working in perturbation theory immediately recognize that the right-hand side (14) measures the sensitivity of the eigenvector \( s_j \) of \( T_k \) to perturbations of size \( O(F_k) \) in the matrix \( T_k \).
Chris Paige’s approach

Chris Paige bounded the deviation of $\|y_j\|_2$ from one by something of the form

$$\left|\|y_j\|_2^2 - 1\right| \leq \frac{O(F_k)}{\min_{\ell \neq j} |\theta_j - \theta_\ell|}.$$  \hspace{1cm} (14)

The length of the Ritz vector $y_j$ is close to one as long as the perturbation term is small and no other Ritz value is close to $\theta_j$.

People working in perturbation theory immediately recognize that the right-hand side (14) measures the sensitivity of the eigenvector $s_j$ of $T_k$ to perturbations of size $O(F_k)$ in the matrix $T_k$.

We might guess that it is indeed a perturbation of the eigenvector that causes the deviation. But where to look for this perturbation? Where do we find the underlying sensitivity analysis?
Chris Paige’s approach

\[ y_{ij}^{(k)^TR_k y_{ij}^{(k)}} = - \sum_{t=1}^{k-1} \eta_{t+1, i} \sum_{r=1}^{t} \frac{\epsilon_{rr}^{(t)}}{\beta_{t+1} \eta_{rr}^{(t)}} y_{ij}^{(k)^T} \begin{bmatrix} y_r^{(t)} \\ 0 \end{bmatrix} \quad (3.19) \]

\[ = - \sum_{t=1}^{k-1} \left( \eta_{t+1, i} \right)^2 \sum_{r=1}^{t} \frac{\epsilon_{rr}^{(t)}}{\mu_{i}^{(k)} - \mu_{r}^{(t)}} \quad (3.20) \]

\[ = \sum_{t=1}^{k-1} \sum_{r=1}^{t} \frac{\epsilon_{rr}^{(t)}}{\mu_{i}^{(k)} - \mu_{s(r)}^{(k)}} \prod_{i=1}^{k} \delta_{i}(t+1, j, k). \quad (3.21) \]

The last equation has this form because \( t \) of the \( \nu_{ij}^{(k)} \) in (3.4) are the eigenvalues \( \mu_{i}^{(t)} \). The index \( s(r) \) indicates that the numerator of \( \delta_{s(r)}(t+1, j, k) \) cancels with \( 1/(\mu_{i}^{(k)} - \mu_{r}^{(t)}) \) in (3.20), and we know \( s(r) \neq j \). These three equations give some useful insights. From (3.17), \( ||z_{ij}^{(k)}|| \) will be significantly different from unity only if the right hand sides of these last three equations are large. In this case (3.19) shows there must be a small \( \delta_{re} = \beta_{t+1} |\eta_{rr}^{(t)}| \), and some \( \mu_{i}^{(t)} \) has therefore stabilized. Equation (3.20) shows that some \( \mu_{r}^{(t)} \) must be close to \( \mu_{i}^{(k)} \), and combining this with (3.19) we will show that at least one such \( \mu_{i}^{(t)} \) has stabilized. Finally from (3.21) we see that there is at least one \( \mu_{s}^{(k)} \) close to \( \mu_{i}^{(k)} \), so that \( \mu_{i}^{(k)} \) cannot be a well-separated eigenvalue of \( T_k \). Conversely if \( \mu_{i}^{(k)} \) is a well-separated eigenvalue of \( T_k \), then (3.16) holds, and if \( \mu_{i}^{(k)} \) has stabilized, then it and \( z_{ij}^{(k)} \) are a satisfactory approximation to an eigenvalue-eigenvector pair of \( A \). We will now quantify these results.
Chris Paige's approach

\[ y_j^{(k)^T} R_k y_j^{(k)} = - \sum_{t=1}^{k-1} \eta_{i+1, i}^{(k)} \sum_{r=1}^{t} \frac{\epsilon_{rr}^{(t)}}{\beta_{i+1}^{(t)}} y_j^{(k)} \begin{bmatrix} y_r^{(t)} \\ 0 \end{bmatrix} \]

\[ = - \sum_{t=1}^{k-1} \left( \eta_{i+1, i}^{(k)} \right)^2 \sum_{r=1}^{t} \frac{\epsilon_{rr}^{(t)}}{\mu_i^{(k)} - \mu_r^{(t)}} \]

\[ = - \sum_{t=1}^{k-1} \sum_{r=1}^{t} \frac{\epsilon_{rr}^{(t)}}{\mu_i^{(k)} - \mu_r^{(k)}} \prod_{i=1}^{k} \delta_i(t+1, i, k). \]

The last equation has this form because \( t \) of the \( \nu_i^{(k)} \) in (3.4) are the eigenvalues \( \mu_r^{(t)} \). The index \( s(r) \) indicates that the numerator of \( \delta_{s(r)}(t+1, i, k) \) cancels with \( 1/(\mu_i^{(k)} - \mu_r^{(t)}) \) in (3.20), and we know \( s(r) \neq j \). These three equations give some useful insights. From (3.17), \( ||z_i^{(k)}|| \) will be significantly different from unity only if the right hand sides of these last three equations are large. In this case (3.19) shows there must be a small \( \delta_{r+1} = \beta_{r+1} |\eta_{rr}^{(t)}| \), and some \( \mu_r^{(t)} \) has therefore stabilized. Equation (3.20) shows that some \( \mu_r^{(t)} \) must be close to \( \mu_i^{(k)} \), and combining this with (3.19) we will show that at least one such \( \mu_r^{(t)} \) has stabilized. Finally from (3.21) we see that there is at least one \( \mu_s^{(k)} \) close to \( \mu_i^{(k)} \), so that \( \mu_i^{(k)} \) cannot be a well-separated eigenvalue of \( T_k \). Conversely if \( \mu_i^{(k)} \) is a well-separated eigenvalue of \( T_k \), then (3.16) holds, and if \( \mu_i^{(k)} \) has stabilized, then it and \( z_i^{(k)} \) are a satisfactory approximation to an eigenvalue-eigenvector pair of \( A \). We will now quantify these results.
Chris Paige’s approach

Chris Paige used the splitting

\[ y_j^H y_j = s_j^H Q_k^H Q_k s_j = 1 + s_j^H (D_k - I_k) s_j^H + 2 \text{Re} (s_j^H R_k s_j) \] (15)

C. Paige: this talk:

\[ z_j^{(k)} \iff y_j \]
\[ y_j \iff s_j \]
\[ \beta_{k+1} \eta_{kj}^{(k)} \iff \beta_k s_{kj} \]
\[ \mu_j^{(k)} \iff \theta_j^{(k)} = \theta_j \]

Chris Paige’s approach

\[ y_j^{(k)^T R_k y_j^{(k)}} = - \sum_{t=1}^{k-1} \eta_{t+1, i}^{(k)} \sum_{r=1}^t \frac{e_r^{(t)}}{\beta_{t+1} \eta_{rr}^{(t)}} y_j^{(k)^T} \begin{bmatrix} y_r^{(t)} \\ 0 \end{bmatrix} \]

(3.19)

\[ = - \sum_{t=1}^{k-1} \left( \eta_{t+1, i}^{(k)} \right)^2 \sum_{r=1}^t \frac{e_r^{(t)}}{\mu_i^{(k)} - \mu_r^{(t)}} \]

(3.20)

\[ = - \sum_{t=1}^{k-1} \sum_{r=1}^t \frac{e_r^{(t)}}{\mu_i^{(k)} - \mu_s^{(r)}} \prod_{i=1 \atop i \neq j}^k \delta_i(t+1, j, k). \]

(3.21)

The last equation has this form because \( t \) of the \( \nu_i^{(k)} \) in (3.4) are the eigenvalues \( \mu_i^{(t)} \). The index \( s(r) \) indicates that the numerator of \( \delta_{s(r)}(t+1, j, k) \) cancels with \( 1/(\mu_i^{(k)} - \mu_r^{(t)}) \) in (3.20), and we know \( s(r) \neq j \). These three equations give some useful insights. From (3.17), \( \| z_j^{(k)} \| \) will be significantly different from unity only if the right hand sides of these last three equations are large. In this case (3.19) shows there must be a small \( \delta_{tr} = \beta_{t+1} | \eta_{rr}^{(t)} | \), and some \( \mu_r^{(t)} \) has therefore stabilized. Equation (3.20) shows that some \( \mu_r^{(t)} \) must be close to \( \mu_i^{(k)} \), and combining this with (3.19) we will show that at least one such \( \mu_r^{(t)} \) has stabilized. Finally from (3.21) we see that there is at least one \( \mu_s^{(k)} \) close to \( \mu_i^{(k)} \), so that \( \mu_i^{(k)} \) cannot be a well-separated eigenvalue of \( T_k \). Conversely if \( \mu_i^{(k)} \) is a well-separated eigenvalue of \( T_k \), then (3.16) holds, and if \( \mu_i^{(k)} \) has stabilized, then it and \( z_j^{(k)} \) are a satisfactory approximation to an eigenvalue-eigenvector pair of \( A \). We will now quantify these results.
The error analysis by Chris Paige is beautiful and gives quantified bounds.
Chris Paige’s approach

The error analysis by Chris Paige is beautiful and gives quantified bounds. The approach is by no means straightforward nor easily generalizable.
Chris Paige’s approach

The error analysis by Chris Paige is beautiful and gives quantified bounds. The approach is by no means straightforward nor easily generalizable.

We intend to show that there is hope that a more “natural” way exists to gain understanding. We consider the first Hessenberg decomposition where only $T_k$ is involved:

$$T_k W_k + G_k = W_{k+1} T_k = W_k T_k + w_{k+1} \beta_k e_k^T. \quad \text{(HessT1)}$$
Chris Paige’s approach

The error analysis by Chris Paige is beautiful and gives quantified bounds. The approach is by no means straightforward nor easily generalizable.

We intend to show that there is hope that a more “natural” way exists to gain understanding. We consider the first Hessenberg decomposition where only $T_k$ is involved:

$$T_k W_k + G_k = W_{k+1} T_k = W_k T_k + w_{k+1} \beta_k e_k^T.$$  \hspace{1cm} (HessT1)

Here, the basis vectors $w_j$ describe the loss of orthogonality and the perturbation term has a large rank-one part (i.e., large last row),

$$W_{k+1} := Q_k^H Q_{k+1},$$
$$G_k := e_k \beta_k q_{k+1}^H Q_k + Q_k^H F_k - F_k^H Q_k.$$  \hspace{1cm} (16)
The derivation of (HessT1) is really simple: Multiplication of (HessA1),

\[ AQ_k + F_k = Q_{k+1}T_k = Q_kT_k + q_{k+1}\beta_ke_k^T, \]  

(HessA1)

with \(Q_k^H\) from the left gives

\[ Q_k^HAQ_k + Q_k^HF_k = Q_k^HQ_kT_k + Q_k^HQ_{k+1}\beta_ke_k^T, \]

(17)
The derivation of (HessT1) is really simple: Multiplication of (HessA1),

$$AQ_k + F_k = Q_{k+1} T_k = Q_k T_k + q_{k+1} \beta_k e_k^T,$$

(HessA1)

with $Q_k^H$ from the left gives

$$Q_k^H AQ_k + Q_k^H F_k = Q_k^H Q_k T_k + Q_k^H q_{k+1} \beta_k e_k^T,$$

(17)

and (17) $- (17)^H$ gives

$$T_k W_k + G_k = W_{k+1} T_k = W_k T_k + w_{k+1} \beta_k e_k^T$$

(HessT1)

with

$$G_k = e_k \beta_k q_{k+1}^H Q_k + Q_k^H F_k - F_k^H Q_k,$$

(18)

since $A = A^H$ and $T_k = T_k^T$ are self-adjoint.
Chris Paige’s approach

We can use the results of (Z, 2007) on the angles between eigenvectors and Ritz vectors to obtain the following formula:

\[ y_j^H y_j = s_j Q_k^H Q_k s_j = s_j^H W_k s_j \]

(19)
We can use the results of (Z, 2007) on the angles between eigenvectors and Ritz vectors to obtain the following formula:

\[ y_j^H y_j = s_j Q_k^H Q_k s_j = s_j^H W_k s_j = \frac{\beta_{1:k-1}}{\omega(\theta_j)} \hat{\nu}(\theta_j)^H W_k \nu(\theta_j) \]

(19)

Here, \( \omega(\theta_j) := \chi'(\theta_j) \)
Chris Paige’s approach

We can use the results of \((Z, 2007)\) on the angles between eigenvectors and Ritz vectors to obtain the following formula:

\[
y^H_j y_j = s^H_j Q^H_k Q s_j = s^H_j W k s_j = \frac{\beta_{1:k-1}}{\omega(\theta_j)} \hat{\nu}(\theta_j)^H W_k \nu(\theta_j)
\]

\[
= \frac{1}{\omega(\theta_j)} \left( A_k(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H Q^H_k q_1 + \sum_{\ell=1}^{k} \beta_{1:\ell-1} A_{\ell+1:k}(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H g_{\ell} \right)
\]

\(19\)

Here, \(\omega(\theta_j) := \chi'(\theta_j)\) and \(A_{\ell+1:k}(z, w) := \chi_{\ell+1:k}[z, w]\)
We can use the results of (Z, 2007) on the angles between eigenvectors and Ritz vectors to obtain the following formula:

\[
y_j^H y_j = s_j Q_k^H Q_k s_j = s_j^H W_k s_j = \frac{\beta_{1:k-1}}{\omega(\theta_j)} \hat{\nu}(\theta_j)^H W_k \nu(\theta_j)
\]

\[
= \frac{1}{\omega(\theta_j)} \left( A_k(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H Q_k^H q_1 + \sum_{\ell=1}^k \beta_{1:\ell-1} A_{\ell+1:k}(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H g_\ell \right)
\]

\[
= \hat{\nu}(\theta_j)^H Q_k^H q_1 + \sum_{\ell=1}^k \frac{\beta_{1:k-1}}{\omega(\theta_j)} \nu_\ell'(\theta_j) \hat{\nu}(\theta_j)^H g_\ell
\]

(19)

Here, \(\omega(\theta_j) := \chi'(\theta_j)\) and \(A_{\ell+1:k}(z, w) := \chi_{\ell+1:k}[z, w] = \beta_{\ell:k-1} \nu_\ell[z, w]\).
We can use the results of (Z, 2007) on the angles between eigenvectors and Ritz vectors to obtain the following formula:

\[ y_j^H y_j = s_j Q_k^H Q_k s_j = s_j^H W_k s_j = \frac{\beta_{1:k-1}}{\omega(\theta_j)} \hat{\nu}(\theta_j)^H W_k \nu(\theta_j) \]

\[ = \frac{1}{\omega(\theta_j)} \left( A_k(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H Q_k^H q_1 + \sum_{\ell=1}^{k} \beta_{1:\ell-1} A_{\ell+1:k}(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H g_\ell \right) \]

\[ = \hat{\nu}(\theta_j)^H Q_k^H q_1 + \sum_{\ell=1}^{k} \frac{\beta_{1:k-1}}{\omega(\theta_j)} \nu_\ell'(\theta_j) \hat{\nu}(\theta_j)^H g_\ell \]

\[ = \hat{\nu}(\theta_j)^H Q_k^H q_1 + \frac{\nu(\theta_j)^H G_k \nu'(\theta_j)}{\nu(\theta_j)^H \nu(\theta_j)} \]

Here, \( \omega(\theta_j) := \chi'(\theta_j) \) and \( A_{\ell+1:k}(z, w) := \chi_{\ell+1:k}[z, w] = \beta_{\ell:k-1} \nu_\ell[z, w] \).
Chris Paige’s approach

We can use the results of (Z, 2007) on the angles between eigenvectors and Ritz vectors to obtain the following formula:

\[
y_j^H y_j = s_j Q_k^H Q_k s_j = s_j^H W_k s_j = \frac{\beta_{1:k-1}}{\omega(\theta_j)} \hat{\nu}(\theta_j)^H W_k \nu(\theta_j)
\]

\[
= \frac{1}{\omega(\theta_j)} \left( \mathcal{A}_k(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H Q_k^H q_1 + \sum_{\ell=1}^{k} \beta_{1:k-1} A_{\ell+1:k}(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H g_\ell \right)
\]

\[
= \hat{\nu}(\theta_j)^H Q_k^H q_1 + \sum_{\ell=1}^{k} \frac{\beta_{1:k-1}}{\omega(\theta_j)} \nu'_\ell(\theta_j) \hat{\nu}(\theta_j)^H g_\ell
\]

\[
= \hat{\nu}(\theta_j)^H Q_k^H q_1 + \frac{\nu(\theta_j)^H G_k \nu'(\theta_j)}{\nu(\theta_j)^H \nu(\theta_j)}
\]

\[
= \hat{\nu}(\theta_j)^H Q_k^H q_1 + \frac{\beta_k q_{k+1}^H Q_k \nu'(\theta_j)}{\nu(\theta_j)^H \nu(\theta_j)} + \frac{\nu(\theta_j)^H (Q_k^H F_k - F_k^H Q_k) \nu'(\theta_j)}{\nu(\theta_j)^H \nu(\theta_j)}.
\]

Here, \(\omega(\theta_j) := \chi'(\theta_j)\) and \(A_{\ell+1:k}(z, w) := \chi_{\ell+1:k}[z, w] = \beta_{\ell:k-1} \nu_\ell[z, w]\).
Chris Paige’s approach

We consider the terms in this representation of $\|y_j\|_2^2$. We start with the first term.
Chris Paige’s approach

We consider the terms in this representation of $\|y_j\|_2^2$. We start with the first term.

In the exact case, i.e., if $Q_k$ is orthonormal,

$$\hat{\nu}(\theta_j)^H Q_k^H q_1 = 1, \quad \text{since} \quad \hat{\nu}_1(z) \equiv 1.$$  \hspace{1cm} (20)
We consider the terms in this representation of $\|y_j\|_2^2$. We start with the first term.

In the exact case, i.e., if $Q_k$ is orthonormal,

$$\hat{\nu}(\theta_j)^H Q_k^H q_1 = 1,$$

since $\hat{\nu}_1(z) \equiv 1$. (20)

In the perturbed case the elements in the scalar product are given by

$$\hat{\nu}(\theta_j)^H Q_k^H q_1 = \sum_{l=1}^{k} \frac{\chi_{1:l-1}(\theta_j)}{\beta_{1:l-1}} q_l^H q_1.$$ (21)
Chris Paige’s approach

We consider the terms in this representation of $\|y_j\|_2^2$. We start with the first term.

In the exact case, i.e., if $Q_k$ is orthonormal,

$$\hat{\nu}(\theta_j)^H Q_k^H q_1 = 1, \quad \text{since} \quad \hat{\nu}_1(z) \equiv 1. \quad (20)$$

In the perturbed case the elements in the scalar product are given by

$$\hat{\nu}(\theta_j)^H Q_k^H q_1 = \sum_{l=1}^{k} \frac{\chi_{1:l-1}(\theta_j)}{\beta_{1:l-1}} q_l^H q_1. \quad (21)$$

The term should be of order one plus “small” times “sensitivity”, the ratio measures the “closeness” of older Ritz values to $\theta_j$. At “sensitive” steps we can have a large loss of orthogonality. It is not known how we should prove this assertion.
Outline

Some history

Hessenberg matrices
  Hessenberg decompositions
  Hessenberg eigenvectors

Chris Paige’s approach
  On the length of the Ritz vectors
  **Eigenvector sensitivity**
  Closer to the original

Our approach
  The shifted decomposition
  About higher derivatives
  The polynomial point of view
Chris Paige’s approach

Both other terms in our expression for $\|y_j\|_2^2$ are of the form

$$\frac{\nu(\theta_j)^H X_k \nu'(\theta_j)}{\nu(\theta_j)^H \nu(\theta_j)} = \frac{\hat{\nu}(\theta_j)^H X_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)}.$$  (22)
Both other terms in our expression for $\|y_j\|_2^2$ are of the form

$$\frac{\nu(\theta_j)^H X_k \nu'(\theta_j)}{\nu(\theta_j)^H \nu(\theta_j)} = \frac{\hat{\nu}(\theta_j)^H X_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)}.$$ (22)

This looks like perturbation theory!
Both other terms in our expression for $\|y_j\|^2$ are of the form

$$\frac{\nu(\theta_j)^H X_k \nu'(\theta_j)}{\nu(\theta_j)^H \nu(\theta_j)} = \frac{\hat{\nu}(\theta_j)^H X_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)}.$$  (22)

This looks like perturbation theory! (If we look long enough : - )
Chris Paige’s approach

Both other terms in our expression for \( \| y_j \|^2 \) are of the form

\[
\frac{\nu(\theta_j)^H X_k \nu'(\theta_j)}{\nu(\theta_j)^H \nu(\theta_j)} = \frac{\hat{\nu}(\theta_j)^H X_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)}.
\]

This looks like perturbation theory! (If we look long enough : - )

For those not familiar with eigenvector perturbations:

\[
| \sin \angle (\nu(\theta_j + \Delta \theta_j), \nu(\theta_j)) | = \frac{\| P_{\nu(\theta_j)} \nu(\theta_j + \Delta \theta_j) \|_2}{\| \nu(\theta_j + \Delta \theta_j) \|_2}
\]

measures the sensitivity of the eigenvector to structured perturbations affecting “only” the Ritz value.
Both other terms in our expression for $\|y_j\|_2^2$ are of the form

$$\frac{\nu(\theta_j)^H \mathbf{X}_k \nu'(\theta_j)}{\nu(\theta_j)^H \nu(\theta_j)} = \frac{\nu'(\theta_j)^H \mathbf{X}_k \nu(\theta_j)}{\nu(\theta_j)^H \nu(\theta_j)}.$$  \hspace{1cm} (22)

This looks like perturbation theory! (If we look long enough : -)

For those not familiar with eigenvector perturbations:

$$\left| \sin \angle (\nu(\theta_j + \Delta \theta_j), \nu(\theta_j)) \right| = \frac{\| P_{\nu(\theta_j)} \nu(\theta_j + \Delta \theta_j) \|_2}{\| \nu(\theta_j + \Delta \theta_j) \|_2}$$  \hspace{1cm} (23)

measures the sensitivity of the eigenvector to structured perturbations affecting “only” the Ritz value. The right eigenvector polynomial is not affected if we alter the elements in the first row of $\mathbf{T}_k$. 
Chris Paige’s approach

Using **Taylor expansion** we obtain

\[
| \sin \angle(\nu(\theta_j + \Delta\theta_j), \nu(\theta_j)) | = \frac{\| P_{\nu(\theta_j)} \perp \nu'(\theta_j) \|_2}{\| \nu(\theta_j) \|_2} | \Delta\theta_j | + O(|\Delta\theta_j|^2). \tag{24}
\]
Chris Paige’s approach

Using Taylor expansion we obtain
\[
| \sin \angle(\boldsymbol{\nu}(\theta_j + \Delta \theta_j), \boldsymbol{\nu}(\theta_j)) | = \frac{\| \mathbf{P}_{\boldsymbol{\nu}(\theta_j)} \cdot \boldsymbol{\nu}'(\theta_j) \|^2}{\| \boldsymbol{\nu}(\theta_j) \|^2} | \Delta \theta_j | + O(|\Delta \theta_j|^2). \tag{24}
\]

Thus, we need “nice” expressions for
\[
\frac{\boldsymbol{\nu}(\theta_i) \cdot \boldsymbol{\nu}'(\theta_j)}{\| \boldsymbol{\nu}(\theta_i) \|^2 \| \boldsymbol{\nu}(\theta_j) \|^2} = \frac{\hat{\boldsymbol{\nu}}(\theta_i) \cdot \boldsymbol{\nu}'(\theta_j)}{\| \hat{\boldsymbol{\nu}}(\theta_i) \|^2 \| \boldsymbol{\nu}(\theta_j) \|^2}. \tag{25}
\]
Using Taylor expansion we obtain

\[ | \sin \angle(\nu(\theta_j + \Delta \theta_j), \nu(\theta_j)) | = \frac{\| P_{\nu(\theta_j)} \nu'(\theta_j) \|_2}{\| \nu(\theta_j) \|_2} | \Delta \theta_j | + O(\| \Delta \theta_j \|^2). \quad (24) \]

Thus, we need “nice” expressions for

\[ \frac{\nu(\theta_i)^H \nu'(\theta_j)}{\| \nu(\theta_i) \|_2 \| \nu(\theta_j) \|_2} = \frac{\hat{\nu}(\theta_i)^H \nu'(\theta_j)}{\| \hat{\nu}(\theta_i) \|_2 \| \nu(\theta_j) \|_2}. \quad (25) \]

It turns out to be easy to obtain analytic expressions for

\[ \frac{\hat{\nu}(\theta_i)^H \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} = \begin{cases} 1 & j \neq i, \\ j = i. & \end{cases} \quad (26) \]
Chris Paige’s approach

Using Taylor expansion we obtain

\[ | \sin \angle(\nu(\theta_j + \Delta \theta_j), \nu(\theta_j)) | = \frac{\| P_{\nu(\theta_j)} \nu'(\theta_j) \|^2}{\| \nu(\theta_j) \|^2} | \Delta \theta_j | + O(|\Delta \theta_j|^2). \tag{24} \]

Thus, we need “nice” expressions for

\[ \frac{\nu(\theta_i)^H \nu'(\theta_j)}{\| \nu(\theta_i) \|^2 \| \nu(\theta_j) \|^2} = \frac{\hat{\nu}(\theta_i)^H \nu'(\theta_j)}{\| \hat{\nu}(\theta_i) \|^2 \| \nu(\theta_j) \|^2}. \tag{25} \]

It turns out to be easy to obtain analytic expressions for

\[ \frac{\hat{\nu}(\theta_i)^H \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} = \begin{cases} \frac{1}{\theta_j - \theta_i}, & j \neq i, \\ \frac{1}{\theta_j - \theta_i}, & j = i. \end{cases} \tag{26} \]
Chris Paige’s approach

Using Taylor expansion we obtain

\[
| \sin \angle (\nu(\theta_j + \Delta \theta_j), \nu(\theta_j)) | = \frac{\| P_{\nu(\theta_j) \perp \nu'(\theta_j)} \|_2}{\| \nu(\theta_j) \|_2} | \Delta \theta_j | + O(|\Delta \theta_j|^2). \tag{24}
\]

Thus, we need “nice” expressions for

\[
\frac{\nu(\theta_i)^H \nu'(\theta_j)}{\| \nu(\theta_i) \|_2 \| \nu(\theta_j) \|_2} = \frac{\hat{\nu}(\theta_i)^H \nu'(\theta_j)}{\| \hat{\nu}(\theta_i) \|_2 \| \nu(\theta_j) \|_2}. \tag{25}
\]

It turns out to be easy to obtain analytic expressions for

\[
\frac{\hat{\nu}(\theta_i)^H \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} = \begin{cases} 
1, & j \neq i, \\
\frac{1}{\theta_j - \theta_i}, & j = i.
\end{cases} \tag{26}
\]
Chris Paige’s approach

Since \( \hat{\nu}(\theta_j) \) and \( \nu(\theta_j) \) are parallel, by the Cauchy-Schwarz (in)equality

\[
|\hat{\nu}(\theta_j)^H \nu(\theta_j)| = \|\hat{\nu}(\theta_j)\|_2 \|\nu(\theta_j)\|_2.
\]  
(27)
Since $\hat{\nu}(\theta_j)$ and $\nu(\theta_j)$ are parallel, by the Cauchy-Schwarz (in)equality

$$|\hat{\nu}(\theta_j)^H \nu(\theta_j)| = \|\hat{\nu}(\theta_j)\|_2 \|\nu(\theta_j)\|_2.$$ \hspace{1cm} (27)

Thus, we need an expression for

$$\frac{|\nu(\theta_i)^H \nu'(\theta_j)|}{\|\nu(\theta_i)\|_2 \|\nu(\theta_j)\|_2} = \frac{\|\hat{\nu}(\theta_j)\|_2}{\|\hat{\nu}(\theta_i)\|_2} \frac{|\nu(\theta_i)^H \nu'(\theta_j)|}{|\hat{\nu}(\theta_j)^H \nu(\theta_j)|}$$

$$= \begin{cases} \frac{\|\hat{\nu}(\theta_j)\|_2}{\|\hat{\nu}(\theta_i)\|_2} \frac{1}{|\theta_j - \theta_i|}, & j \neq i, \\ \left| \sum_{\ell \neq j} \frac{1}{\theta_j - \theta_\ell} \right|, & j = i. \end{cases} \hspace{1cm} (28)$$
Chris Paige’s approach

Observe that the norms of the eigenvectors

\[ \| \hat{\nu}(\theta_j) \|_2^2 = \frac{1}{s_{1j}^2} \]  

(29)

are related to the squares of the first components of the normalized eigenvectors, which are the weights in Gaussian quadrature.
Chris Paige’s approach

Observe that the norms of the eigenvectors

\[ \| \hat{\nu}(\theta_j) \|_2^2 = \frac{1}{s_{1j}^2} \]  \hspace{1cm} (29)

are related to the squares of the first components of the normalized eigenvectors, which are the weights in Gaussian quadrature.

In general, we can make use of the relations

\[ s_{kj}^2 = \frac{\chi_{1:k-1}(\theta_j)}{\omega(\theta_j)} = \frac{1}{\| \nu(\theta_j) \|_2^2}, \]  \hspace{1cm} (30)

\[ s_{1j}^2 = \frac{\chi_{2:k}(\theta_j)}{\omega(\theta_j)} = \frac{1}{\| \hat{\nu}(\theta_j) \|_2^2}, \]

where the reduced polynomial \( \omega = \omega_j \) is defined as before by

\[ \omega(z) = \prod_{\ell \neq j} (z - \theta_{\ell}). \]  \hspace{1cm} (31)
Chris Paige’s approach

By classical perturbation theory

\[ | \sin \angle (\hat{\nu}(\theta_j), \nu(\theta_j) + \nu'(\theta_j) \Delta \theta_j) | \lesssim \frac{|\Delta \theta_j|}{\min_{\ell \neq j} |\theta_j - \theta_{\ell}|}. \]  

(32)
Chris Paige’s approach

By classical perturbation theory

\[ | \sin \angle(\hat{\nu}(\theta_j), \nu(\theta_j) + \nu'(\theta_j)\Delta \theta_j)| \lesssim \frac{|\Delta \theta_j|}{\min_{\ell \neq j}|\theta_j - \theta_\ell|}. \]  

(32)

This is not easy to deduce here, we only have seen thus far that

\[ \sin^2 \angle(\hat{\nu}(\theta_j), \nu(\theta_j) + \nu'(\theta_j)\Delta \theta_j) = \frac{\|P_{\hat{\nu}(\theta_j)} - \nu'(\theta_j)\|_2^2}{\||\nu(\theta_j)||_2^2 |\Delta \theta_j|^2 + O(|\Delta \theta_j|^3)} \]  

(33)

\[ = \frac{|\Delta \theta_j|^2}{s_{1j}^2} \sum_{\ell \neq j} \frac{s_{1\ell}^2}{(\theta_j - \theta_\ell)^2} + O(|\Delta \theta_j|^3). \]
By classical perturbation theory

\[
| \sin \angle(\hat{\nu}(\theta_j), \nu(\theta_j) + \nu'(\theta_j)\Delta \theta_j) | \lesssim \frac{|\Delta \theta_j|}{\min_{\ell \neq j} |\theta_j - \theta_\ell|}.
\] (32)

This is not easy to deduce here, we only have seen thus far that

\[
\sin^2 \angle(\hat{\nu}(\theta_j), \nu(\theta_j) + \nu'(\theta_j)\Delta \theta_j) = \frac{\|P_{\hat{\nu}(\theta_j)} - \nu'(\theta_j)\|^2}{\|\nu(\theta_j)\|^2} |\Delta \theta_j|^2 + O(|\Delta \theta_j|^3)
\]

\[
= \frac{|\Delta \theta_j|^2}{s_{1j}^2} \sum_{\ell \neq j} \frac{s_{1\ell}^2}{(\theta_j - \theta_\ell)^2} + O(|\Delta \theta_j|^3). \] (33)

Maybe the relations collected on the following slides will provide helpful.
Chris Paige’s approach

A first tool of trade that works in the symmetric case is the identity

$$\beta_{1:k-1}^2 = \chi_{1:k-1}(\theta_j) \cdot \chi_{2:k}(\theta_j),$$

(34)

valid for all Ritz values $\theta_j$. 

This identity proves that if $\beta_{1:k-1}$ is "moderate", then in case of "large" $\omega(\theta_j)$, at least one of $s_{1j}$ and $s_{kj}$ has to be "small" and thus at least one of $\|\hat{\nu}(\theta_j)\|_2$ and $\|\nu(\theta_j)\|_2$ has to be "large",

$$s_{1j} s_{kj} = \beta_{1:k-1}^2 \omega(\theta_j) = 1 \hat{\nu}(\theta_j) H \nu(\theta_j).$$

(35)

A relation without squares follows easily using (Z, 2006), (Z, 2007) and Cauchy-Schwarz, we have

$$s_{1j} s_{kj} = \beta_{1:k-1}^2 \omega(\theta_j) = \frac{1}{\|\hat{\nu}(\theta_j)\|_2} \|\nu(\theta_j)\|_2.$$
A first tool of trade that works in the symmetric case is the identity

$$\beta_{1:k-1}^2 = \chi_{1:k-1}(\theta_j) \cdot \chi_{2:k}(\theta_j),$$  \hspace{1cm} (34)

valid for all Ritz values $\theta_j$.

This identity proves that if $\beta_{1:k-1}^2$ is “moderate”, then in case of “large” $\omega(\theta_j)$, at least one of $s_{1j}$ and $s_{kj}$ has to be “small” and thus at least one of $\|\hat{\nu}(\theta_j)\|_2$ and $\|\nu(\theta_j)\|_2$ has to be “large”,

$$(s_{1j}s_{kj})^2 = \frac{\beta_{1:k-1}^2}{\omega(\theta_j)^2} = \left(\|\hat{\nu}(\theta_j)\|_2\|\nu(\theta_j)\|_2\right)^{-2}. \hspace{1cm} (35)$$
A first tool of trade that works in the symmetric case is the identity

$$\beta_{1:k-1}^2 = \chi_{1:k-1}(\theta_j) \cdot \chi_{2:k}(\theta_j),$$  \hspace{1cm} (34)

valid for all Ritz values $\theta_j$.

This identity proves that if $\beta_{1:k-1}^2$ is “moderate”, then in case of “large” $\omega(\theta_j)$, at least one of $s_{1j}$ and $s_{kj}$ has to be “small” and thus at least one of $\|\hat{\nu}(\theta_j)\|_2$ and $\|\nu(\theta_j)\|_2$ has to be “large”,

$$(s_{1j}s_{kj})^2 = \frac{\beta_{1:k-1}^2}{\omega(\theta_j)^2} = (\|\hat{\nu}(\theta_j)\|_2\|\nu(\theta_j)\|_2)^{-2}. \hspace{1cm} (35)$$

A relation without squares follows easily using (Z, 2006), (Z, 2007) and Cauchy-Schwarz, we have

$$s_{1j}s_{kj} = \frac{\beta_{1:k-1}}{\omega(\theta_j)} = \frac{1}{\hat{\nu}(\theta_j)^H \nu(\theta_j)}. \hspace{1cm} (36)$$
Chris Paige’s approach

For $k > 3$ we observe that we can obtain the upper bound

$$|s_1 s_k| < \frac{1}{2},$$

(37)
Chris Paige’s approach

For $k > 3$ we observe that we can obtain the upper bound

$$|s_{1j}s_{kj}| < \frac{1}{2},$$

(37)

since for a vector $x$ with non-zero structure as follows,

$$x = \begin{pmatrix} x_0 & 0 & \ddots & 0 \\ 0 & x_0 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & y \end{pmatrix}, \quad \max_{x^2+y^2=1} |xy| = \frac{1}{2}.$$  

(38)
Chris Paige’s approach

For $k > 3$ we observe that we can obtain the upper bound

$$|s_j s_{kj}| < \frac{1}{2},$$

(37)

since for a vector $x$ with non-zero structure as follows,

$$x = \begin{pmatrix} x_0 \\ \vdots \\ 0 \\ y \end{pmatrix}, \quad \max_{x^2+y^2=1} |xy| = \frac{1}{2}. \quad (38)$$

There can not be two consecutive zeros in an eigenvector of a tridiagonal matrix, as then the three-term recurrence would construct only zeros,

$$s_j^T \left( \beta_{\ell+1} e_{\ell+1} = (T_k - \alpha_{\ell}) e_\ell - \beta_{\ell-1} e_{\ell-1} \right). \quad (39)$$
Chris Paige’s approach

For $k > 3$ we observe that we can obtain the upper bound

$$|s_is_j| < \frac{1}{2},$$  \hspace{1cm} (37)

since for a vector $x$ with non-zero structure as follows,

$$x = \begin{pmatrix} x_0 & \ldots & 0 \\ 0 & \ddots & \vdots \\ 0 & \ldots & y \end{pmatrix}, \quad \max_{x^2+y^2=1} |xy| = \frac{1}{2}. \hspace{1cm} (38)$$

There can not be two consecutive zeros in an eigenvector of a tridiagonal matrix, as then the three-term recurrence would construct only zeros,

$$s_j^T (\beta_\ell e_{\ell+1} = (T_k - \alpha_\ell)e_\ell - \beta_{\ell-1}e_{\ell-1}). \hspace{1cm} (39)$$

Thus, $|\omega(\theta_j)| = |\chi'(\theta_j)| > 2/\beta_{1:k-1}$.
Chris Paige’s approach

To give a partial resume: There seems to be a relation to perturbation theory, but it really is not fully understood.
Chris Paige’s approach

To give a partial resume: There seems to be a relation to perturbation theory, but it really is not fully understood.

We reconsider

\[
\frac{\nu(\theta_j)\nu'(\theta_j)}{\nu(\theta_j)^H\nu(\theta_j)} = \frac{\hat{\nu}(\theta_j)^H\nu'(\theta_j)}{\hat{\nu}(\theta_j)^H\nu(\theta_j)}.
\] (40)

(41) Again, we have to treat the norms of the eigenvector polynomials in some (not specified) manner to make this a successful approach.
Chris Paige’s approach

To give a partial resume: There seems to be a relation to perturbation theory, but it really is not fully understood.

We reconsider

\[
\frac{\nu(\theta_j)^H X_k \nu'(\theta_j)}{\nu(\theta_j)^H \nu(\theta_j)} = \frac{\hat{\nu}(\theta_j)^H X_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)}. \tag{40}
\]

Inserting the identity matrix gives

\[
\frac{\hat{\nu}(\theta_j)^H X_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} = \sum_{i=1}^{k} \frac{\hat{\nu}(\theta_j)^H X_k \nu(\theta_i)}{\hat{\nu}(\theta_i)^H \nu(\theta_i)} \frac{\hat{\nu}(\theta_i)^H \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)}.
\]

\[
= \sum_{i \neq j} \frac{1}{\theta_j - \theta_i} \left( \frac{\hat{\nu}(\theta_j)^H X_k \nu(\theta_i)}{\hat{\nu}(\theta_i)^H \nu(\theta_i)} + \frac{\hat{\nu}(\theta_j)^H X_k \nu(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} \right). \tag{41}
\]
Chris Paige’s approach

To give a partial resume: There seems to be a relation to perturbation theory, but it really is not fully understood.

We reconsider

\[
\frac{\mathbf{\nu}(\theta_j)^H \mathbf{X}_k \mathbf{\nu}'(\theta_j)}{\mathbf{\nu}(\theta_j)^H \mathbf{\nu}(\theta_j)} = \frac{\hat{\mathbf{\nu}}(\theta_j)^H \mathbf{X}_k \mathbf{\nu}'(\theta_j)}{\hat{\mathbf{\nu}}(\theta_j)^H \mathbf{\nu}(\theta_j)}.
\]

(40)

Inserting the identity matrix gives

\[
\frac{\hat{\mathbf{\nu}}(\theta_j)^H \mathbf{X}_k \mathbf{\nu}'(\theta_j)}{\hat{\mathbf{\nu}}(\theta_j)^H \mathbf{\nu}(\theta_j)} = \sum_{i=1}^{k} \frac{\hat{\mathbf{\nu}}(\theta_j)^H \mathbf{X}_k \mathbf{\nu}(\theta_i)}{\hat{\mathbf{\nu}}(\theta_i)^H \mathbf{\nu}(\theta_i)} \frac{\hat{\mathbf{\nu}}(\theta_i)^H \mathbf{\nu}'(\theta_j)}{\hat{\mathbf{\nu}}(\theta_j)^H \mathbf{\nu}(\theta_j)}
\]

(41)

\[
= \sum_{i \neq j} \frac{1}{\theta_j - \theta_i} \left( \frac{\hat{\mathbf{\nu}}(\theta_j)^H \mathbf{X}_k \mathbf{\nu}(\theta_i)}{\hat{\mathbf{\nu}}(\theta_i)^H \mathbf{\nu}(\theta_i)} + \frac{\hat{\mathbf{\nu}}(\theta_j)^H \mathbf{X}_k \mathbf{\nu}(\theta_j)}{\hat{\mathbf{\nu}}(\theta_j)^H \mathbf{\nu}(\theta_j)} \right).
\]

Again, we have to treat the norms of the eigenvector polynomials in some (not specified) manner to make this a successful approach.
Outline

Some history

Hessenberg matrices
  Hessenberg decompositions
  Hessenberg eigenvectors

Chris Paige’s approach
  On the length of the Ritz vectors
  Eigenvector sensitivity
  Closer to the original

Our approach
  The shifted decomposition
  About higher derivatives
  The polynomial point of view
Chris Paige’s approach

We only used the first Hessenberg decomposition with $T_k$. We can stick closer to what Chris Paige did, and use the second one:

$$T_k R_k + E_k = R_{k+1} T_k = R_k T_k + r_{k+1} \beta_k e_k^T.$$  \hspace{1cm} (HessT2)
Chris Paige’s approach

We only used the first Hessenberg decomposition with $T_k$. We can stick closer to what Chris Paige did, and use the second one:

$$T_k R_k + E_k = R_{k+1} T_k = R_k T_k + r_{k+1} \beta_k e_k^T. \quad \text{(HessT2)}$$

Here, $E_k$ is upper triangular, and $W_{k+1} = R_k^H + D_k + R_{k+1}$ with $R_{k+1} = \text{sut}(W_{k+1})$ and $D_k$ diagonal.
Chris Paige’s approach

We only used the first Hessenberg decomposition with $T_k$. We can stick closer to what Chris Paige did, and use the second one:

$$T_k R_k + E_k = R_{k+1} T_k = R_k T_k + r_{k+1} \beta_k e_k^T.$$  \hspace{1cm} (HessT2)

Here, $E_k$ is upper triangular, and $W_{k+1} = R_k^H + D_k + R_{k+1}$ with $R_{k+1} = \text{sut}(W_{k+1})$ and $D_k$ diagonal.

Chris Paige proved that $E_k$ is “small”.
Chris Paige’s approach

We only used the first Hessenberg decomposition with $T_k$. We can stick closer to what Chris Paige did, and use the second one:

$$T_k R_k + E_k = R_{k+1} T_k = R_k T_k + r_{k+1} \beta_k e_k^T. \quad \text{(HessT2)}$$

Here, $E_k$ is upper triangular, and $W_{k+1} = R_k^H + D_k + R_{k+1}$ with $R_{k+1} = \text{sut}(W_{k+1})$ and $D_k$ diagonal.

Chris Paige proved that $E_k$ is “small”.

Based on the identity

$$\|y_j\|_2^2 - 1 = s_j^H (D_k - I_k) s_j + 2 \text{Re} (s_j^H R_k s_j) \quad \text{(42)}$$

Chris Paige bounded the deviation of $\|y_j\|$ from one.
We can again use the characterization of the angles to compute his results in terms of the derivative,

\[ s_j^H R_k s_j \]

(43)
Chris Paige’s approach

We can again use the characterization of the angles to compute his results in terms of the derivative,

\[
\mathbf{s}_j^\mathsf{H} \mathbf{R}_k \mathbf{s}_j = \frac{\beta_{1:k-1}}{\omega(\theta_j)} \hat{\nu}(\theta_j)^\mathsf{H} \mathbf{R}_k \nu(\theta_j)
\]

(43)
We can again use the characterization of the angles to compute his results in terms of the derivative,

\[
\begin{align*}
    s_j^H R_k s_j &= \frac{\beta_1:k-1}{\omega(\theta_j)} \hat{\nu}(\theta_j)^H R_k \nu(\theta_j) \\
    &= \frac{1}{\omega(\theta_j)} \left( A_k(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H r_1 + \sum_{\ell=1}^{k} \beta_1:1-1 A_{\ell+1:k}(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H E_k e_\ell \right) 
\end{align*}
\]  

(43)
Chris Paige’s approach

We can again use the characterization of the angles to compute his results in terms of the derivative,

\[ s_j^H R_k s_j = \frac{\beta_{1:k-1}}{\omega(\theta_j)} \hat{\nu}(\theta_j)^H R_k \nu(\theta_j) \]

\[ = \frac{1}{\omega(\theta_j)} \left( A_k(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H r_1 + \sum_{\ell=1}^{k} \beta_{1:\ell-1} A_{\ell+1:k}(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H E_k e_\ell \right) \]

\[ = \sum_{\ell=1}^{k} \frac{\beta_{1:k-1}}{\omega(\theta_j)} \nu'_\ell(\theta_j) \hat{\nu}(\theta_j)^H E_k e_\ell \]
Chris Paige’s approach

We can again use the characterization of the angles to compute his results in terms of the derivative,

$$s_j^H R_k s_j = \frac{\beta_{1:k-1}}{\omega(\theta_j)} \hat{\nu}(\theta_j)^H R_k \nu(\theta_j)$$

$$= \frac{1}{\omega(\theta_j)} \left( A_k(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H r_1 + \sum_{\ell=1}^k \beta_{1:\ell-1} A_{\ell+1:k}(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H E_k e_\ell \right)$$

$$= \sum_{\ell=1}^k \frac{\beta_{1:k-1}}{\omega(\theta_j)} \nu'_\ell(\theta_j) \nu(\theta_j)^H E_k e_\ell = \frac{\hat{\nu}(\theta_j)^H E_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)}.$$
We can again use the characterization of the angles to compute his results in terms of the derivative,

\[
s_j^H R_k s_j = \frac{\beta_{1:k-1}}{\omega(\theta_j)} \hat{\nu}(\theta_j)^H R_k \nu(\theta_j)
\]

\[
= \frac{1}{\omega(\theta_j)} \left( A_k(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H r_1 + \sum_{\ell=1}^{k} \beta_{1:\ell-1} A_{\ell+1:k}(\theta_j, \theta_j) \hat{\nu}(\theta_j)^H E_k e_\ell \right)
\]

\[
= \sum_{\ell=1}^{k} \frac{\beta_{1:k-1}}{\omega(\theta_j)} \nu'_\ell(\theta_j) \hat{\nu}(\theta_j)^H E_k e_\ell = \frac{\hat{\nu}(\theta_j)^H E_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)}.
\]

Thus,

\[
\|y_j\|_2^2 - 1 = s_j^H (D_k - I_k) s_j + 2 \text{Re} \left( \frac{\hat{\nu}(\theta_j)^H E_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} \right).
\]
Chris Paige’s approach

We can reformulate this by our “perturbation analysis”:

\[
\frac{\hat{\nu}(\theta_j)^H E_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} = \sum_{\ell \neq j} \frac{1}{\theta_j - \theta_i} \left( \frac{\hat{\nu}(\theta_j)^H E_k \nu(\theta_i)}{\hat{\nu}(\theta_i)^H \nu(\theta_i)} + \frac{\hat{\nu}(\theta_j)^H E_k \nu(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} \right). \tag{45}
\]
Chris Paige’s approach

We can reformulate this by our “perturbation analysis”:

\[
\frac{\hat{\nu}(\theta_j)^H E_k \nu' (\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} = \sum_{\ell \neq j} \frac{1}{\theta_j - \theta_i} \left( \frac{\hat{\nu}(\theta_j)^H E_k \nu(\theta_i)}{\hat{\nu}(\theta_i)^H \nu(\theta_i)} + \frac{\hat{\nu}(\theta_j)^H E_k \nu(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} \right). \tag{45}
\]

The last ratio in parentheses is a Rayleigh quotient of a small matrix and thus small. Chris Paige denoted this Rayleigh quotient by \( \epsilon_{jj}^{(k)} \).
Chris Paige’s approach

We can reformulate this by our “perturbation analysis”:

$$
\frac{\hat{\nu}(\theta_j)^H E_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} = \sum_{\ell \neq j} \frac{1}{\theta_j - \theta_i} \left( \frac{\hat{\nu}(\theta_j)^H E_k \nu(\theta_i)}{\hat{\nu}(\theta_i)^H \nu(\theta_i)} + \frac{\hat{\nu}(\theta_j)^H E_k \nu(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} \right).
$$

The last ratio in parentheses is a Rayleigh quotient of a small matrix and thus small. Chris Paige denoted this Rayleigh quotient by $\epsilon_{jj}^{(k)}$.

Obviously, using

$$
s_j^H (T_k R_k + E_k = R_k T_k + r_{k+1} \beta_k e_k^T) s_j,
$$

(46)
Chris Paige’s approach

We can reformulate this by our “perturbation analysis”:

\[
\frac{\hat{\nu}(\theta_j)^H E_k \nu'(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} = \sum_{\ell \neq j} \frac{1}{\theta_j - \theta_i} \left( \frac{\hat{\nu}(\theta_j)^H E_k \nu(\theta_i)}{\hat{\nu}(\theta_i)^H \nu(\theta_i)} + \frac{\hat{\nu}(\theta_j)^H E_k \nu(\theta_j)}{\hat{\nu}(\theta_j)^H \nu(\theta_j)} \right). \tag{45}
\]

The last ratio in parentheses is a Rayleigh quotient of a small matrix and thus small. Chris Paige denoted this Rayleigh quotient by \( \epsilon_{jj}^{(k)} \).

Obviously, using

\[
s_j^H \left( T_k R_k + E_k = R_k T_k + r_{k+1} \beta_k e_k^T \right) s_j, \tag{46}
\]

proves that loss of orthogonality and “convergence” go hand in hand,

\[
\epsilon_{jj}^{(k)} = s_j^H Q_k^H q_{k+1} \beta_k e_k^T s_j = y_j^H q_{k+1} \beta_k s_{kj}. \tag{47}
\]
Again, we can express part of the relations in terms of perturbations of eigenvectors, but the first term in the parentheses has not been treated **fully satisfactory**.
Again, we can express part of the relations in terms of perturbations of eigenvectors, but the first term in the parentheses has not been treated fully satisfactory.

Perhaps we need to better understand the derivative of the eigenvector polynomial. In (Z, 2006) it was proven that this vector is the first principal vector if the eigenvalue is multiple, which is never true in our setting.
Chris Paige’s approach

Again, we can express part of the relations in terms of perturbations of eigenvectors, but the first term in the parentheses has not been treated fully satisfactory.

Perhaps we need to better understand the derivative of the eigenvector polynomial. In (Z, 2006) it was proven that this vector is the first principal vector if the eigenvalue is multiple, which is never true in our setting.

It turns out that the derivative of the eigenvector polynomial is in some sense obtained by inverse iteration with shifted $A$. This can be seen with the aid of the shifted Hessenberg decomposition.
Outline

Some history

Hessenberg matrices
  Hessenberg decompositions
  Hessenberg eigenvectors

Chris Paige’s approach
  On the length of the Ritz vectors
  Eigenvector sensitivity
  Closer to the original

Our approach
  The shifted decomposition
  About higher derivatives
  The polynomial point of view
Consider the shifted Lanczos Hessenberg decomposition

\[ \tilde{A}Q_k + \tilde{F}_k = Q_{k+1}T_k = Q_kT_k + q_{k+1}\beta_k e_k^T \]  

(HessA2)
A new approach

Consider the **shifted Lanczos Hessenberg decomposition**

\[
\tilde{A}Q_k + \tilde{F}_k = Q_{k+1}T_k = Q_kT_k + q_{k+1}\beta_ke_k^T
\]  

(HessA2)

where for a given eigenpair \(Av_i = v_i\lambda_i\) and a given Ritz value \(\theta_j\) we defined

\[
\tilde{A} := A - (\lambda_i - \theta_j)v_iv_i^H \quad \text{and} \quad \tilde{F}_k := (\lambda_i - \theta_j)v_iv_i^HQ_k + F_k.
\]  

(48)
Consider the shifted Lanczos Hessenberg decomposition

\[
\tilde{A}Q_k + \tilde{F}_k = Q_{k+1}T_k = Q_kT_k + q_{k+1}\beta_ke_k^T \tag{HessA2}
\]

where for a given eigenpair \(Av_i = v_i\lambda_i\) and a given Ritz value \(\theta_j\) we defined

\[
\tilde{A} := A - (\lambda_i - \theta_j)v_iv_i^H \quad \text{and} \quad \tilde{F}_k := (\lambda_i - \theta_j)v_iv_i^HQ_k + F_k. \tag{48}
\]

This definitions ensure that the Hessenberg decomposition is still balanced and that now

\[
v_i^H\tilde{A} = v_i^H(A - (\lambda_i - \theta_j)v_iv_i^H) = \lambda_i v_i^H - (\lambda_i - \theta_j)v_i^Hv_i^H = \theta_j v_i^H, \tag{49}
\]

i.e., \(v_i\) is a left eigenvector to the eigenvalue \(\theta_j\).
A new approach

The angle between the eigenvector $v_i$ and a scaled Ritz vector is given by

$$\frac{\beta_{1:k-1}}{\omega(\theta_j)} v_i^H Q_k \nu(\theta_j) = v_i^H q_1 + \frac{\beta_{1:k-1}}{\omega(\theta_j)} v_i^H \tilde{F}_k \nu'(\theta_j),$$

(50)
The angle between the eigenvector $v_i$ and a scaled Ritz vector is given by

$$\frac{\beta_{1:k-1}}{\omega(\theta_j)} v_i^H Q_k \nu(\theta_j) = v_i^H q_1 + \frac{\beta_{1:k-1}}{\omega(\theta_j)} v_i^H \tilde{F}_k \nu'(\theta_j),$$

in other words,

$$v_i^H Q_k \nu(\theta_j) = \frac{\omega(\theta_j)}{\beta_{1:k-1}} v_i^H q_1 + v_i^H \tilde{F}_k \nu'(\theta_j)$$

$$= \frac{\omega(\theta_j)}{\beta_{1:k-1}} v_i^H q_1 + (\lambda_i - \theta_j)v_i^H Q_k \nu'(\theta_j) + v_i^H F_k \nu'(\theta_j).$$
Our approach

The shifted decomposition

A new approach

The angle between the eigenvector $v_i$ and a scaled Ritz vector is given by

$$\frac{\beta_{1:k-1}}{\omega(\theta_j)} v_i^H Q_k \nu(\theta_j) = v_i^H q_1 + \frac{\beta_{1:k-1}}{\omega(\theta_j)} v_i^H \tilde{F}_k \nu'(\theta_j),$$

in other words,

$$v_i^H Q_k \nu(\theta_j) = \frac{\omega(\theta_j)}{\beta_{1:k-1}} v_i^H q_1 + v_i^H \tilde{F}_k \nu'(\theta_j)$$

$$= \frac{\omega(\theta_j)}{\beta_{1:k-1}} v_i^H q_1 + (\lambda_i - \theta_j) v_i^H Q_k \nu'(\theta_j) + v_i^H F_k \nu'(\theta_j).$$

Remark: This relation is correct, no matter how close or far away $\lambda_i$ and $\theta_j$ are.
A new approach

The angle between the eigenvector $v_i$ and a scaled Ritz vector is given by

$$\frac{\beta_{1:k-1}}{\omega(\theta_j)} v_i^H Q_k \nu(\theta_j) = v_i^H q_1 + \frac{\beta_{1:k-1}}{\omega(\theta_j)} v_i^H \tilde{F}_k \nu'(\theta_j),$$

in other words,

$$v_i^H Q_k \nu(\theta_j) = \frac{\omega(\theta_j)}{\beta_{1:k-1}} v_i^H q_1 + v_i^H \tilde{F}_k \nu'(\theta_j)$$

$$= \frac{\omega(\theta_j)}{\beta_{1:k-1}} v_i^H q_1 + (\lambda_i - \theta_j) v_i^H Q_k \nu'(\theta_j) + v_i^H F_k \nu'(\theta_j).$$

Remark: This relation is correct, no matter how close or far away $\lambda_i$ and $\theta_j$ are. The relation can be obtained using any eigenvalue and any Ritz value.
A new approach

Sorting gives the following \textit{anti-Taylor-like} approximation,

\[
    v_i^H Q_k (\nu(\theta_j) - \nu'(\theta_j)(\lambda_i - \theta_j)) = \frac{\omega(\theta_j)}{\beta_1: k-1} v_i^H q_1 + v_i^H F_k \nu'(\theta_j),
\]

(52)
A new approach

Sorting gives the following \textit{anti-Taylor-like} approximation,

\begin{equation}
\mathbf{v}_i^H \mathbf{Q}_k \left( \mathbf{\nu}(\theta_j) - \mathbf{\nu}'(\theta_j)(\lambda_i - \theta_j) \right) = \frac{\omega(\theta_j)}{\beta_{1:k-1}} \mathbf{v}_i^H \mathbf{q}_1 + \mathbf{v}_i^H \mathbf{F}_k \mathbf{\nu}'(\theta_j),
\end{equation}

weighted summation over all eigenpairs of $\mathbf{A}$ gives the \textit{inexact inverse subspace iteration}

\begin{equation}
\left((\theta_j \mathbf{I}_n - \mathbf{A}) \mathbf{Q}_k - \mathbf{F}_k\right) \mathbf{\nu}'(\theta_j) = \frac{\omega(\theta_j)}{\beta_{1:k-1}} \mathbf{q}_1 - \mathbf{Q}_k \mathbf{\nu}(\theta_j).
\end{equation}
A new approach

Sorting gives the following anti-Taylor-like approximation,

$$v_i^H Q_k (\boldsymbol{\nu}(\theta_j) - \nu'(\theta_j) (\lambda_i - \theta_j)) = \frac{\omega(\theta_j)}{\beta_{1:k-1}} v_i^H q_1 + v_i^H F_k \nu'(\theta_j),$$

(52)

weighted summation over all eigenpairs of $A$ gives the inexact inverse subspace iteration

$$((\theta_j I_n - A) Q_k - F_k) \nu'(\theta_j) = \frac{\omega(\theta_j)}{\beta_{1:k-1}} q_1 - Q_k \nu(\theta_j).$$

(53)

There is a good chance that $Q_k \nu'(\theta_j)$ is a better candidate for a “Ritz vector” if $Q_k \nu(\theta_j)$ is “small” and $\theta_j$ is close to an eigenvalue of $A$. 
Our approach

The shifted decomposition

A new approach

Sorting gives the following anti-Taylor-like approximation,

\[ v_i^H Q_k (\nu(\theta_j) - \nu'(\theta_j)(\lambda_i - \theta_j)) = \frac{\omega(\theta_j)}{\beta_{1:k-1}} v_i^H q_1 + v_i^H F_k \nu'(\theta_j), \]  

(52)

weighted summation over all eigenpairs of \( A \) gives the inexact inverse subspace iteration

\[ ((\theta_j I_n - A)Q_k - F_k)\nu'(\theta_j) = \frac{\omega(\theta_j)}{\beta_{1:k-1}} q_1 - Q_k \nu(\theta_j). \]  

(53)

There is a good chance that \( Q_k \nu'(\theta_j) \) is a better candidate for a “Ritz vector” if \( Q_k \nu(\theta_j) \) is “small” and \( \theta_j \) is close to an eigenvalue of \( A \).

A mixed numerical-symbolic computation I presented at the GAMM annual meeting 2006 does support this idea in case of a second Ritz copy.
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

symbolic Lanczos for 29 steps

finite precision Lanczos for 29 steps; Matlab 7.2.0.294 (R2006a)

finite precision Lanczos for 29 steps; older version of MRRR

finite precision Lanczos for 29 steps; exact eigenvectors
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

A graph showing the first trailing adjugate polynomials at $\lambda_i$ and maximal $\theta$, step = 7. The x-axis represents real eigenvalues, and the y-axis shows the absolute size of scaled polynomials/eigenvectors. The graph includes lines for different polynomials:

- Red line: scaled eigenvector
- Blue line: first adjugate polynomial
- Teal line: second adjugate polynomial
- Green line: third adjugate polynomial
- Purple line: fourth adjugate polynomial
- Cyan line: fifth adjugate polynomial
- Black line: sixth adjugate polynomial
- Dark green line: seventh adjugate polynomial
Our approach
The shifted decomposition

An example from my 2006 GAMM talk

first trailing adjugate polynomials at $\lambda_i$ and maximal $\theta$, step = 8

real eigenvalues
absolute size of scaled polynomials/eigenvectors

0 0.5 1 1.5 2 2.5 3
10
−20
10
−15
10
−10
10
−5
10
0
10
5
10
10
10
15

scaled eigenvector
first adjugate polynomial
second adjugate polynomial
third adjugate polynomial
fourth adjugate polynomial
fifth adjugate polynomial
sixth adjugate polynomial
seventh adjugate polynomial
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

first trailing adjugate polynomials at $\lambda_i$ and maximal $\theta$, step = 9

real eigenvalues

absolute size of scaled polynomials/eigenvectors

first adjugate polynomial

second adjugate polynomial

third adjugate polynomial

fourth adjugate polynomial

fifth adjugate polynomial

sixth adjugate polynomial

seventh adjugate polynomial
An example from my 2006 GAMM talk

- First trailing adjugate polynomials at $\lambda_i$ and maximal $\theta$, step = 10
- Absolute size of scaled polynomials/eigenvectors
- Real eigenvalues
- Scaling of eigenvectors
- First adjugate polynomial
- Second adjugate polynomial
- Third adjugate polynomial
- Fourth adjugate polynomial
- Fifth adjugate polynomial
- Sixth adjugate polynomial
- Seventh adjugate polynomial

TUHH
Jens-Peter M. Zemke
Chris Paige and the Lanczos process
ICS of CAS, September 24th, 2009
43 / 53
Our approach
The shifted decomposition

An example from my 2006 GAMM talk

first trailing adjugate polynomials at $\lambda_i$ and maximal $\theta$, step = 11

real eigenvalues
absolute size of scaled polynomials/eigenvectors

scaled eigenvector
first adjugate polynomial
second adjugate polynomial
third adjugate polynomial
fourth adjugate polynomial
fifth adjugate polynomial
sixth adjugate polynomial
seventh adjugate polynomial
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

First trailing adjugate polynomials at $\lambda_i$, and maximal $\theta$, step = 12

real eigenvalues

absolute size of scaled polynomials/eigenvectors

- scaled eigenvector
- first adjugate polynomial
- second adjugate polynomial
- third adjugate polynomial
- fourth adjugate polynomial
- fifth adjugate polynomial
- sixth adjugate polynomial
- seventh adjugate polynomial
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

![Graph showing first trailing adjugate polynomials at λi and maximal θ, step = 13](image)

- Real eigenvalues
- Absolute size of scaled polynomials/eigenvectors
- First adjugate polynomial
- Second adjugate polynomial
- Third adjugate polynomial
- Fourth adjugate polynomial
- Fifth adjugate polynomial
- Sixth adjugate polynomial
- Seventh adjugate polynomial
Our approach

The shifted decomposition

An example from my 2006 GAMM talk
Our approach

The shifted decomposition

An example from my 2006 GAMM talk
An example from my 2006 GAMM talk

Our approach

The shifted decomposition

first trailing adjugate polynomials at $\lambda$, and maximal $\theta$, step = 16

- scaled eigenvector
- first adjugate polynomial
- second adjugate polynomial
- third adjugate polynomial
- fourth adjugate polynomial
- fifth adjugate polynomial
- sixth adjugate polynomial
- seventh adjugate polynomial

real eigenvalues

absolute size of scaled polynomials/eigenvectors

TUHH
Jens-Peter M. Zemke
Chris Paige and the Lanczos process
ICS of CAS, September 24th, 2009
An example from my 2006 GAMM talk

Our approach

The shifted decomposition

first trailing adjugate polynomials at \( \lambda_i \) and maximal \( \theta \), step = 17

- scaled eigenvector
- first adjugate polynomial
- second adjugate polynomial
- third adjugate polynomial
- fourth adjugate polynomial
- fifth adjugate polynomial
- sixth adjugate polynomial
- seventh adjugate polynomial

real eigenvalues

absolute size of scaled polynomials/eigenvectors
Our approach
The shifted decomposition

An example from my 2006 GAMM talk

- first trailing adjugate polynomials at $\lambda_i$ and maximal $\theta$, step = 18

- absolute size of scaled polynomials/eigenvectors

- first adjugate polynomial
- second adjugate polynomial
- third adjugate polynomial
- fourth adjugate polynomial
- fifth adjugate polynomial
- sixth adjugate polynomial
- seventh adjugate polynomial

- real eigenvalues

TUHH
Jens-Peter M. Zemke
Chris Paige and the Lanczos process
ICS of CAS, September 24th, 2009
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

first trailing adjugate polynomials at $\lambda_i$ and maximal $\theta$, step = 19

real eigenvalues
absolute size of scaled polynomials/eigenvectors

- scaled eigenvector
- first adjugate polynomial
- second adjugate polynomial
- third adjugate polynomial
- fourth adjugate polynomial
- fifth adjugate polynomial
- sixth adjugate polynomial
- seventh adjugate polynomial

TUHH
Jens-Peter M. Zemke
Chris Paige and the Lanczos process
ICS of CAS, September 24th, 2009
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

- real eigenvalues
- absolute size of scaled polynomials/eigenvectors
- first trailing adjugate polynomials at $\lambda_i$ and maximal $\theta$, step = 20

TUHH
Jens-Peter M. Zemke
Chris Paige and the Lanczos process
ICS of CAS, September 24th, 2009
43 / 53
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

![Graph showing first trailing adjugate polynomials at λ_i and maximal θ, step = 21.](image)

- **Scaled eigenvector**
- **First adjugate polynomial**
- **Second adjugate polynomial**
- **Third adjugate polynomial**
- **Fourth adjugate polynomial**
- **Fifth adjugate polynomial**
- **Sixth adjugate polynomial**
- **Seventh adjugate polynomial**

TUHH

Jens-Peter M. Zemke

Chris Paige and the Lanczos process

ICS of CAS, September 24th, 2009

43 / 53
Our approach
The shifted decomposition

An example from my 2006 GAMM talk

First trailing adjugate polynomials at \( \lambda_i \) and maximal \( \theta \), step = 22

- scaled eigenvector
- first adjugate polynomial
- second adjugate polynomial
- third adjugate polynomial
- fourth adjugate polynomial
- fifth adjugate polynomial
- sixth adjugate polynomial
- seventh adjugate polynomial

real eigenvalues
absolute size of scaled polynomials/eigenvectors
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

first trailing adjugate polynomials evaluated at $\lambda_i$ for a cluster made of $\theta_m$ and $\theta_{m-1}$, step = 23

real eigenvalues; cluster size = 1.06484949774e−07

absolute size of scaled polynomials/eigenvectors

$\lambda_i$ for a cluster made of $\theta_m$ and $\theta_{m-1}$, step = 23

scaled eigenvector (m)

scaled eigenvector (m−1)

first adjugate (m)

first adjugate (m−1)

second adjugate (m)

second adjugate (m−1)

third adjugate (m)

third adjugate (m−1)

"principal" vector (m)

"principal" vector (m−1)
An example from my 2006 GAMM talk

first trailing adjugate polynomials evaluated at $\lambda_i$ for a cluster made of $\theta_m$ and $\theta_{m-1}$, step = 24

- scaled eigenvector (m)
- scaled eigenvector (m−1)
- first adjugate (m)
- first adjugate (m−1)
- second adjugate (m)
- second adjugate (m−1)
- third adjugate (m)
- third adjugate (m−1)
- "principal" vector (m)
- "principal" vector (m−1)
Our approach: The shifted decomposition

An example from my 2006 GAMM talk

First trailing adjugate polynomials evaluated at $\lambda_i$, for a cluster made of $\theta_m$ and $\theta_{m-1}$, step = 25

Real eigenvalues; cluster size = 7.63297630692e−12
An example from my 2006 GAMM talk

Our approach

The shifted decomposition

Real eigenvalues; cluster size = 6.24431961098e−14

Absolute size of scaled polynomials/eigenvectors

First trailing adjugate polynomials evaluated at $\lambda_i$ for a cluster made of $\theta_m$ and $\theta_{m-1}$, step = 26

scaled eigenvector (m)
scaled eigenvector (m−1)
first adjugate (m)
first adjugate (m−1)
second adjugate (m)
second adjugate (m−1)
third adjugate (m)
third adjugate (m−1)
"principal" vector (m)
"principal" vector (m−1)
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

first trailing adjugate polynomials evaluated at $\lambda_i$ for a cluster made of $\theta_m$ and $\theta_{m-1}$, step = 27

real eigenvalues; cluster size = 6.25657007172e−16

absolute size of scaled polynomials/eigenvectors

first adjugate (m)
first adjugate (m−1)
second adjugate (m)
second adjugate (m−1)
third adjugate (m)
third adjugate (m−1)
"principal" vector (m)
"principal" vector (m−1)
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

first trailing adjugate polynomials evaluated at $\lambda_i$ for a cluster made of $\theta_m$ and $\theta_{m-1}$, step = 28

real eigenvalues; cluster size = $2.16341508015\times10^{-16}$

absolute size of scaled polynomials/eigenvectors

real eigenvalues; cluster size = $2.16341508015\times10^{-16}$

"principal" vector (m)
"principal" vector (m−1)

Chris Paige and the Lanczos process

ICS of CAS, September 24th, 2009

43 / 53
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

first trailing adjugate polynomials evaluated at $\lambda_i$ for a cluster made of $\theta_m$ and $\theta_{m-1}$, step = 29

real eigenvalues; cluster size = 2.14536454507e−16

absolute size of scaled polynomials/eigenvectors

0 0.5 1 1.5 2 2.5 3
10
−20
10
−15
10
−10
10
−5
10
0
10
5
10
10
10
15

scaled eigenvector (m)
scaled eigenvector (m−1)
first adjugate (m)
first adjugate (m−1)
second adjugate (m)
second adjugate (m−1)
third adjugate (m)
third adjugate (m−1)
"principal" vector (m)
"principal" vector (m−1)
Our approach
The shifted decomposition

An example from my 2006 GAMM talk
Our approach

The shifted decomposition

An example from my 2006 GAMM talk

first trailing adjugate polynomials evaluated at $\lambda_i$ for a cluster made of $\theta_m$ and $\theta_{m-1}$, step = 31

real eigenvalues; cluster size = 2.14522996773e−16

absolute size of scaled polynomials/eigenvectors

first adjugate (m)  first adjugate (m−1)
second adjugate (m)  second adjugate (m−1)
third adjugate (m)  third adjugate (m−1)
"principal" vector (m)  "principal" vector (m−1)
Our approach
The shifted decomposition

An example from my 2006 GAMM talk
Our approach
The shifted decomposition

An example from my 2006 GAMM talk

first trailing adjugate polynomials evaluated at $\lambda_i$ for a cluster made of $\theta_m$ and $\theta_{m-1}$, step = 33

real eigenvalues; cluster size = 2.14522996085e−16

absolute size of scaled polynomials/eigenvectors

first adjugate (m)
second adjugate (m)
third adjugate (m)
"principal" vector (m)

first adjugate (m−1)
second adjugate (m−1)
third adjugate (m−1)
"principal" vector (m−1)

scaled eigenvector (m)
scaled eigenvector (m−1)
Outline

Some history

Hessenberg matrices
  Hessenberg decompositions
  Hessenberg eigenvectors

Chris Paige’s approach
  On the length of the Ritz vectors
  Eigenvector sensitivity
  Closer to the original

Our approach
  The shifted decomposition
  About higher derivatives
  The polynomial point of view
Higher derivatives

There is an alternative way to prove that the first “principal” Ritz vector is obtained by inexact inverse subspace iteration.
Higher derivatives

There is an alternative way to prove that the first “principal” Ritz vector is obtained by inexact inverse subspace iteration.

For any $z \in \mathbb{C}$ and any $\ell \in \mathbb{N}$ we have that

$$
(zI_k - T_k) \frac{\nu^{(\ell)}(z)}{\ell!} + \frac{\nu^{(\ell-1)}(z)}{(\ell - 1)!} = e_1 \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}}.
$$

(54)
Higher derivatives

There is an alternative way to prove that the first “principal” Ritz vector is obtained by inexact inverse subspace iteration.

For any $z \in \mathbb{C}$ and any $\ell \in \mathbb{N}$ we have that

$$(zI_k - T_k) \frac{\nu^{(\ell)}(z)}{\ell!} + \frac{\nu^{(\ell-1)}(z)}{(\ell - 1)!} = e_1 \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}}. \quad (54)$$

This implies that

$$(zQ_k - Q_k T_k) \frac{\nu^{(\ell)}(z)}{\ell!} + Q_k \frac{\nu^{(\ell-1)}(z)}{(\ell - 1)!} = Q_k e_1 \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}}. \quad (55)$$
Higher derivatives

There is an alternative way to prove that the \textit{first “principal” Ritz vector} is obtained by \textit{inexact inverse subspace iteration}.

For \textit{any} \( z \in \mathbb{C} \) and \textit{any} \( \ell \in \mathbb{N} \) we have that

\[
( zI_k - T_k ) \frac{ \nu^{(\ell)}(z) }{ \ell! } + \frac{ \nu^{(\ell-1)}(z) }{ (\ell - 1)! } = e_1 \frac{ \chi^{(\ell)}(z) }{ \beta_{1:k-1} }.
\]

This implies that

\[
( zQ_k - Q_k T_k ) \frac{ \nu^{(\ell)}(z) }{ \ell! } + Q_k \frac{ \nu^{(\ell-1)}(z) }{ (\ell - 1)! } = Q_k e_1 \frac{ \chi^{(\ell)}(z) }{ \beta_{1:k-1} }.
\]

\[
= ( ( zI_n - A ) Q_k - F_k ) \frac{ \nu^{(\ell)}(z) }{ \ell! } + Q_k \frac{ \nu^{(\ell-1)}(z) }{ (\ell - 1)! } = q_k \frac{ \chi^{(\ell)}(z) }{ \beta_{1:k-1} }.
\]
Higher derivatives

There is an alternative way to prove that the first “principal” Ritz vector is obtained by inexact inverse subspace iteration.

For any $z \in \mathbb{C}$ and any $\ell \in \mathbb{N}$ we have that

$$
(zI_k - T_k) \frac{\nu^{(\ell)}(z)}{\ell!} + \frac{\nu^{(\ell-1)}(z)}{(\ell - 1)!} = e_1 \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}}.
$$

This implies that

$$
(zQ_k - Q_k T_k) \frac{\nu^{(\ell)}(z)}{\ell!} + Q_k \frac{\nu^{(\ell-1)}(z)}{(\ell - 1)!} = Q_k e_1 \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}}
$$

$$
= ((zI_n - A)Q_k - F_k) \frac{\nu^{(\ell)}(z)}{\ell!} + Q_k \frac{\nu^{(\ell-1)}(z)}{(\ell - 1)!} = q_1 \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}}.
$$

We have used the fact that the last $\ell$ components of $\nu^{(\ell)}(z)$ are zero.
Higher derivatives

We can now insert any value for $z$, natural candidates are values in a cluster and the eigenvalue closest to the Ritz value(s) of interest.
We can now insert any value for $z$, natural candidates are values in a cluster and the eigenvalue closest to the Ritz value(s) of interest.

We could use Rolle’s theorem and set $z$ to the unique zero of $\chi^{(m-1)}(z)$ in the cluster interval of Ritz values, where $m$ denotes the number of Ritz values in the cluster.
Higher derivatives

We can now insert any value for \( z \), natural candidates are values in a cluster and the eigenvalue closest to the Ritz value(s) of interest.

We could use Rolle’s theorem and set \( z \) to the unique zero of \( \chi^{(m-1)}(z) \) in the cluster interval of Ritz values, where \( m \) denotes the number of Ritz values in the cluster.

We could use any linear combination of the derivatives for a fixed \( z \), as everything is linear,

\[
((zI_n - A)Q_k - F_k) \left( \sum_{\ell=0}^{p} a_\ell \frac{\nu^{(\ell)}(z)}{\ell!} \right) + Q_k \left( \sum_{\ell=1}^{p} a_\ell \frac{\nu^{(\ell-1)}(z)}{(\ell - 1)!} \right) = q_1 \left( \sum_{\ell=1}^{p} a_\ell \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}} \right). \tag{56}
\]
We could try to find a linear combination

$$
\sum_{\ell=1}^{p} a_{\ell} \nu^{(\ell-1)}(z) / (\ell - 1)!
$$

that (almost) lies in the null-space of $Q_k$.  

(57)
We could try to find a linear combination

\[ \sum_{\ell=1}^{p} a_\ell \frac{v^{(\ell-1)}(z)}{(\ell - 1)!} \]  

(57)

that (almost) lies in the **null-space** of \( Q_k \). This linear combination of the derivatives would be close to an eigenvector of \( A \), if the corresponding linear combination

\[ \sum_{\ell=1}^{p} a_\ell \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}} \]  

(58)

involving the characteristic polynomial is “small”. 
Higher derivatives

We could try to find a linear combination

\[
\sum_{\ell=1}^{p} a_\ell \frac{\nu^{(\ell-1)}(z)}{(\ell - 1)!}
\]

(57)

that (almost) lies in the null-space of \( Q_k \). This linear combination of the derivatives would be close to an eigenvector of \( A \), if the corresponding linear combination

\[
\sum_{\ell=1}^{p} a_\ell \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}}
\]

(58)

involving the characteristic polynomial is “small”.

Another example: Choosing \( p = k \) and \( a_\ell = a_\ell(z) \) appropriately gives the Taylor approximation to, say, the characteristic polynomial of \( A \) at \( \lambda \).
Outline

Some history

Hessenberg matrices
   Hessenberg decompositions
   Hessenberg eigenvectors

Chris Paige’s approach
   On the length of the Ritz vectors
   Eigenvector sensitivity
   Closer to the original

Our approach
   The shifted decomposition
   About higher derivatives
   The polynomial point of view
We can consider the parameter-dependent relation

\[(T_k - zI_k)R_k + E_k = R_k(T_k - zI_k) + r_{k+1}\beta_ke_k^T.\]  

(59)
We can consider the parameter-dependent relation

\[(T_k - zI_k)R_k + E_k = R_k(T_k - zI_k) + r_{k+1} \beta_k e_k^T.\]  

(59)

Remember that \(R_k\) is a strictly upper triangular matrix.
Our approach

The polynomial point of view

Polynomial view on Chris Paige’s result

We can consider the \textit{parameter-dependent relation}

\[
(T_k - zI_k) R_k + E_k = R_k (T_k - zI_k) + r_{k+1} \beta_k e_k^T.
\]  

(59)

Remember that \(R_k\) is a strictly upper triangular matrix.

Application of \(\hat{\nu}(z)^H\) and \(\nu(z)\) gives

\[
\hat{\nu}(z)^H E_k \nu(z) = \hat{\nu}(z)^H r_{k+1} \beta_k.
\]  

(60)
Our approach  The polynomial point of view

Polynomial view on Chris Paige’s result

We can consider the parameter-dependent relation

\[(T_k - zI_k)R_k + E_k = R_k(T_k - zI_k) + r_{k+1} \beta_k e_k^T.\]  \hfill (59)

Remember that \(R_k\) is a strictly upper triangular matrix.

Application of \(\hat{\nu}(z)^H\) and \(\nu(z)\) gives

\[\hat{\nu}(z)^H E_k \nu(z) = \hat{\nu}(z)^H r_{k+1} \beta_k.\]  \hfill (60)

This is an exact polynomial relation with polynomials of degree \(k - 1\), i.e., these are \(k\) linear equations:

\[\hat{\nu}(z)^H E_k \nu(z) = (1 \cdots z^{k-1}) \begin{pmatrix} \star & \cdots & \star \\ \vdots & \ddots & \vdots \\ \star & \cdots & 1 \end{pmatrix} \begin{pmatrix} z^{k-1} \\ \vdots \\ 1 \end{pmatrix}.\]  \hfill (61)
This gives the complete characterization of the loss of orthogonality

\[ r_{k+1} \beta_k = Q_k^H q_{k+1} \beta_k \]  

(62)

at step \( k + 1 \) in terms of the errors \( E_k \).
This gives the **complete characterization** of the loss of orthogonality

\[ r_{k+1} \beta_k = Q_k^H q_{k+1} \beta_k \]  

(62)

at step \( k + 1 \) in terms of the errors \( E_k \).

Well known is this result when \( z = \theta_j \) is any **Ritz value**, but we could compare, say, the coefficients of the highest term \( z^{k-1} \):

\[
\text{trace}(E_k) z^{k-1} + \cdots = \hat{\nu}(z)^H E_k \nu(z) = \hat{\nu}(z)^H r_{k+1} \beta_k = q_k^H q_{k+1} \beta_k z^{k-1} + \cdots
\]  

(63)
This gives the **complete characterization** of the loss of orthogonality

\[ r_{k+1} \beta_k = Q_k^H q_{k+1} \beta_k \]  

(62)

at step \( k + 1 \) in terms of the errors \( E_k \).

Well known is this result when \( z = \theta_j \) is any Ritz value, but we could compare, say, the **coefficients of the highest term** \( z^{k-1} \):

\[
\text{trace} (E_k) z^{k-1} + \cdots = \hat{\boldsymbol{\nu}}(z)^H E_k \boldsymbol{\nu}(z) = \hat{\boldsymbol{\nu}}(z)^H r_{k+1} \beta_k = q_k^H q_{k+1} \beta_k z^{k-1} + \cdots. \tag{63}
\]

This is correct. It does not give further insights, but proves that the relation is sound. The diagonal of \( E_k \) is closely related to the local loss of orthogonality.
Maybe of interest in **CG or other OR methods** is the relation involving the constant terms, namely

\[
\hat{\nu}(0)^{H} E_k \nu(0) = \hat{\nu}(0)^{H} Q_k^{H} q_{k+1} \beta_k. \tag{64}
\]
Maybe of interest in **CG or other OR methods** is the relation involving the constant terms, namely

\[
\hat{\nu}(0)^H E_k \nu(0) = \hat{\nu}(0)^H Q_k^H q_{k+1} \beta_k. \tag{64}
\]

By definition of \(\nu(z)\), \(z_k\) defined by

\[
z_k \frac{\chi(0)}{\|r_0\| \beta_{1:k-1}} := -\nu(0) = -(-T_k)^{-1} \frac{\chi(0)}{\beta_{1:k-1}} e_1, \tag{65}
\]

where \(r_0 := b - Ax_0\) denotes the starting residual, is the \(k\)th QOR solution, see (Z, 2007).
Maybe of interest in **CG or other OR methods** is the relation involving the constant terms, namely

\[ \hat{\nu}(0)^H E_k \nu(0) = \hat{\nu}(0)^H Q_k^H q_{k+1} \beta_k. \]  

(64)

By definition of \( \nu(z) \), \( z_k \) defined by

\[ z_k \frac{\chi(0)}{\| r_0 \| \beta_{1:k-1}} := -\nu(0) = -(-T_k)^{-1} \frac{\chi(0)}{\beta_{1:k-1}} e_1, \]  

(65)

where \( r_0 := b - Ax_0 \) denotes the starting residual, is the \( k \)th QOR solution, see (Z, 2007).

At this point the talk comes to its end. The true research can start here.
We sketched how Chris Paige’s approach of error analysis of the finite Lanczos process seems to be related to eigenvector sensitivity.
We sketched how Chris Paige’s approach of error analysis of the finite Lanczos process seems to be related to eigenvector sensitivity.

We have shown that the analytic representation of eigenvectors as polynomial vectors evaluated at the eigenvalues results in simpler expressions. These are based on differentiation.

The presented relations mostly carry over to the unsymmetric Lanczos process, portions of it should help in distinguishing different implementations of the unsymmetric Lanczos process.
We sketched how Chris Paige’s approach of error analysis of the finite Lanczos process seems to be related to eigenvector sensitivity.

We have shown that the analytic representation of eigenvectors as polynomial vectors evaluated at the eigenvalues results in simpler expressions. These are based on differentiation.

We failed to give a complete error analysis based solely on our polynomial description.
We sketched how Chris Paige’s approach of error analysis of the finite Lanczos process seems to be related to eigenvector sensitivity.

We have shown that the analytic representation of eigenvectors as polynomial vectors evaluated at the eigenvalues results in simpler expressions. These are based on differentiation.

We failed to give a complete error analysis based solely on our polynomial description.

The presented relations mostly carry over to the unsymmetric Lanczos process, portions of it should help in distinguishing different implementations of the unsymmetric Lanczos process.
Děkuji.
Děkuji. Once Again.
Behavior of slightly perturbed Lanczos and conjugate-gradient recurrences.
*Linear Algebra Appl.*, 113:7–63.

Predicting the behavior of finite precision Lanczos and conjugate gradient computations.

*The Lanczos and Conjugate Gradient Algorithms: From Theory to Finite Precision Computation.*
SIAM, Philadelphia.

The Lanczos and conjugate gradient algorithms in finite precision arithmetic.


Hessenberg eigenvalue–eigenmatrix relations.  

Abstract perturbed Krylov methods.  