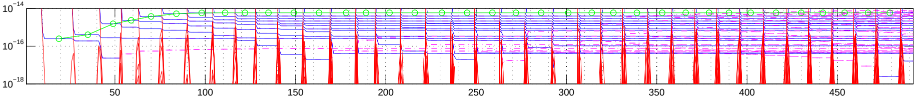
Towards a deeper understanding of Chris Paige's error analysis of the finite precision Lanczos process

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

Institut für Numerische Simulation
Technische Universität Hamburg-Harburg

ICS of CAS, September 24th, 2009



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, Anne Greenbaum, the Lanczos process

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).

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

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

and the starting vector

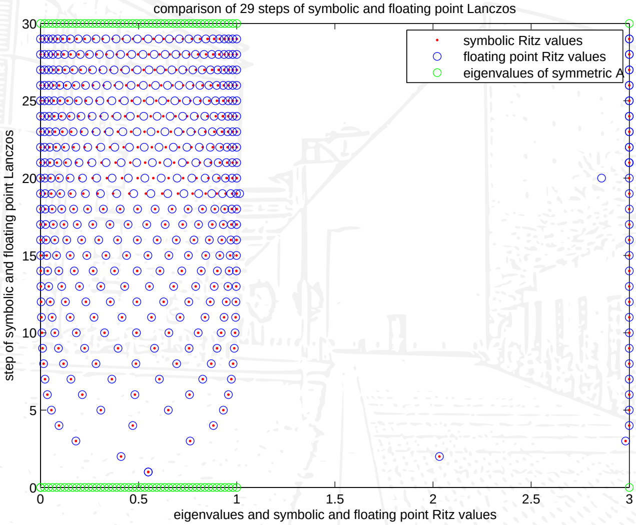
$$\mathbf{e} = \texttt{ones(51,1)}$$

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

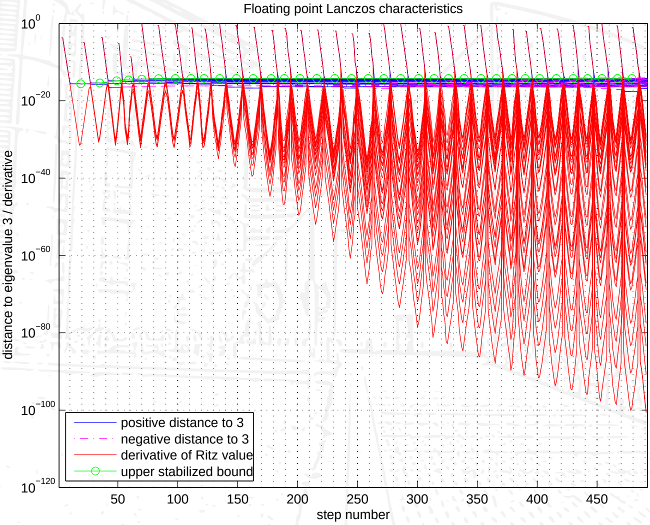
At step 10 the first Ritz value has converged (up to machine precision) to the eigenvalue 3, at step 27 the second one has converged. Detoriation reaches a maximum at step $19 = \lceil (10 + 27)/2 \rceil$.

Eigenvalues and eigenvectors are computed using MRRR, i.e., LAPACK's routine DSTEGR, since MATLAB's eig (using LAPACK's DSYEV, i.e., the QR algorithm implemented as DSTEQR) fails in delivering accurate eigenvectors. Additionally, we heavily used the symbolic toolbox, i.e., MAPLE.

The finite precision behavior



The finite precision behavior



Hessenberg decompositions

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

$$\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1}\underline{\mathbf{H}}_k = \mathbf{Q}_k\mathbf{H}_k + \mathbf{q}_{k+1}h_{k+1,k}\mathbf{e}_k^{\mathsf{T}}. \tag{1}$$

Here, $\mathbf{H}_k$ 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

$$\mathbf{A}\mathbf{Q}_k + \mathbf{F}_k = \mathbf{Q}_{k+1}\underline{\mathbf{H}}_k = \mathbf{Q}_k\mathbf{H}_k + \mathbf{q}_{k+1}h_{k+1,k}\mathbf{e}_k^{\mathsf{T}}. \tag{2}$$

The matrix $\mathbf{H}_k$ of the perturbed variant will, in general, still be unreduced.

Hessenberg decompositions

In (Z, 2007) we did consider in an abstract manner the matrix equation

$$\mathbf{A}\mathbf{Q}_k + \mathbf{F}_k = \mathbf{Q}_{k+1}\underline{\mathbf{H}}_k = \mathbf{Q}_k\mathbf{H}_k + \mathbf{q}_{k+1}h_{k+1,k}\mathbf{e}_k^{\mathsf{T}} \tag{3}$$

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 θ_j,
- and the angles between Ritz vectors and eigenvectors of $\mathbf{A}$.

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).

Hessenberg decompositions

In case of the symmetric Lanczos process we have

$$\mathbf{A}\mathbf{Q}_k + \mathbf{F}_k = \mathbf{Q}_{k+1}\underline{\mathbf{T}}_k = \mathbf{Q}_k\mathbf{T}_k + \mathbf{q}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}}, \tag{4}$$

where

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

The elements of the tridiagonal matrix $\mathbf{T}_k$ are denoted by

$$\mathbf{T}_k = \begin{pmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_{k-1} \\ & & \beta_{k-1} & \alpha_k \end{pmatrix}, \quad \beta_j > 0 \quad \forall\, 1 \leqslant j \leqslant k. \tag{5}$$

(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 $\mathbf{A}$.

The first one is the original Lanczos Hessenberg decomposition

$$\mathbf{A}\mathbf{Q}_k + \mathbf{F}_k = \mathbf{Q}_{k+1}\underline{\mathbf{T}}_k = \mathbf{Q}_k\mathbf{T}_k + \mathbf{q}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}}. \qquad \text{(HessA1)}$$

With a given eigenpair $\mathbf{v}_i^{\mathsf{H}}\mathbf{A} = \lambda_i\mathbf{v}_i^{\mathsf{H}}$ and a given Ritz value θ_j we define

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

Then we obtain the shifted Lanczos Hessenberg decomposition

$$\tilde{\mathbf{A}}\mathbf{Q}_k + \tilde{\mathbf{F}}_k = \mathbf{Q}_{k+1}\underline{\mathbf{T}}_k = \mathbf{Q}_k\mathbf{T}_k + \mathbf{q}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}}. \qquad \text{(HessA2)}$$

This Hessenberg decomposition is interesting especially in case that $\lambda_i - \theta_j$ is “small”, i.e., “comparable” to $\mathbf{F}_k$.

Hessenberg decompositions

The next two Hessenberg decompositions are based on $\mathbf{T}_k$ in place of $\mathbf{A}$. These form the essential part of Chris Paige's analysis.

Let $\mathbf{W}_{k+1} := \mathbf{Q}_k^{\mathsf{H}}\mathbf{Q}_{k+1}$, define $\mathbf{G}_k := \mathbf{e}_k\beta_k\mathbf{q}_{k+1}^{\mathsf{H}}\mathbf{Q}_k + \mathbf{Q}_k^{\mathsf{H}}\mathbf{F}_k - \mathbf{F}_k^{\mathsf{H}}\mathbf{Q}_k$. Then

$$\mathbf{T}_k\mathbf{W}_k + \mathbf{G}_k = \mathbf{W}_{k+1}\underline{\mathbf{T}}_k = \mathbf{W}_k\mathbf{T}_k + \mathbf{w}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}}. \qquad \text{(HessT1)}$$

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

Let $\mathbf{W}_{k+1} = \mathbf{R}_k^{\mathsf{H}} + \mathbf{D}_k + \mathbf{R}_{k+1}$ with $\mathbf{R}_{k+1} = \operatorname{sut}(\mathbf{W}_{k+1})$ and $\mathbf{D}_k$ diagonal. Then

$$\mathbf{T}_k\mathbf{R}_k + \mathbf{E}_k = \mathbf{R}_{k+1}\underline{\mathbf{T}}_k = \mathbf{R}_k\mathbf{T}_k + \mathbf{r}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}} \qquad \text{(HessT2)}$$

with $\mathbf{E}_k$ upper triangular and small if $\mathbf{F}_k$ is small and local orthogonality is preserved.

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 $(\check{\boldsymbol{\nu}}(z)^{\mathsf{T}} = \hat{\boldsymbol{\nu}}(z)^{\mathsf{H}})$

$$(z\mathbf{I}_k - \mathbf{T}_k)\boldsymbol{\nu}(z) = \mathbf{e}_1 \frac{\chi(z)}{\beta_{1:k-1}}, \quad \check{\boldsymbol{\nu}}(z)^{\mathsf{T}}(z\mathbf{I}_k - \mathbf{T}_k) = \frac{\chi(z)}{\beta_{1:k-1}}\mathbf{e}_n^{\mathsf{T}} \tag{7}$$

with $\chi(z) := \det(z\mathbf{I}_k - \mathbf{T}_k)$ 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

$$\hat{\boldsymbol{\nu}}(z)^{\mathsf{H}}\boldsymbol{\nu}(w) = \frac{\chi[z,w]}{\beta_{1:k-1}} = \frac{1}{\beta_{1:k-1}} \begin{cases} \dfrac{\chi(z)-\chi(w)}{z-w}, & z \neq w \\ \chi'(z), & z = w. \end{cases} \tag{8}$$

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 $\mathbf{A}$ and $\mathbf{T}_k$ we know that the left and right eigenvector are parallel and can be scaled to unit length.

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

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

$$\boldsymbol{\nu}(z) := \left(\frac{\chi_{j+1:k}(z)}{\beta_{j:k-1}}\right)_{j=1}^{k} \quad \text{and} \quad \check{\boldsymbol{\nu}}(z) := \left(\frac{\chi_{1:j-1}(z)}{\beta_{1:j-1}}\right)_{j=1}^{k}, \tag{9}$$

where

$$\chi_{i:j}(z) := \det\left(z\mathbf{I}_{j-i+1} - \mathbf{T}_{i:j}\right) \quad \text{and} \quad \beta_{i:j} := \prod_{\ell=i}^{j} \beta_\ell, \quad 0 \leqslant i \leqslant j < k. \tag{10}$$

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

Hessenberg eigenvectors and eigenvector derivatives

Unit length eigenvectors $\mathbf{s}_j$ of $\mathbf{T}_k$ to the eigenvalue θ_j are defined by

$$\mathbf{s}_j := \frac{\boldsymbol{\nu}_j}{\|\boldsymbol{\nu}_j\|_2}. \tag{11}$$

This ensures that the last component s_{kj} of $\mathbf{s}_j$ is positive and given by

$$s_{kj} = \frac{1}{\|\boldsymbol{\nu}_j\|_2} = \frac{1}{\|\boldsymbol{\nu}(\theta_j)\|_2} > 0. \tag{12}$$

In case of an error-free process we have with the Ritz vector $\mathbf{y}_j := \mathbf{Q}_k\mathbf{s}_j$ the (backward- and forward-error) bound

$$\min_{\lambda} |\lambda - \theta_j| \leqslant \frac{\|\mathbf{A}\mathbf{y}_j - \mathbf{y}_j\theta_j\|_2}{\|\mathbf{y}_j\|_2} = \beta_k s_{kj}. \tag{13}$$

Chris Paige's approach

Chris Paige bounded the deviation of $\|\mathbf{y}_j\|_2$ from one by something of the form

$$\big|\|\mathbf{y}_j\|_2^2 - 1\big| \leqslant \frac{O(\mathbf{F}_k)}{\min_{\ell \neq j} |\theta_j - \theta_\ell|}. \quad (14)$$

The length of the Ritz vector $\mathbf{y}_j$ is close to one as long as the perturbation term is small and no other Ritz value is close to θ_j.

People working in perturbation theory immediately recognize that the right-hand side (14) measures the sensitivity of the eigenvector $\mathbf{s}_j$ of $\mathbf{T}_k$ to perturbations of size $O(\mathbf{F}_k)$ in the matrix $\mathbf{T}_k$.

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

$$y_j^{(k)T} R_k y_j^{(k)} = -\sum_{t=1}^{k-1} \eta_{t+1,j}^{(k)} \sum_{r=1}^{t} \frac{\varepsilon_{rr}^{(t)}}{\beta_{t+1}\eta_{tr}^{(t)}} y_j^{(k)T} \begin{bmatrix} y_r^{(t)} \\ 0 \end{bmatrix} \tag{3.19}$$

$$= -\sum_{t=1}^{k-1} \left(\eta_{t+1,j}^{(k)}\right)^2 \sum_{r=1}^{t} \frac{\varepsilon_{rr}^{(t)}}{\mu_j^{(k)} - \mu_r^{(t)}} \tag{3.20}$$

$$= -\sum_{t=1}^{k-1} \sum_{r=1}^{t} \frac{\varepsilon_{rr}^{(t)}}{\mu_j^{(k)} - \mu_{s(r)}^{(k)}} \prod_{\substack{i=1 \\ i \neq j \\ i \neq s(r)}}^{k} \delta_i(t+1, j, k). \tag{3.21}$$

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 $\delta_{s(r)}(t+1, j, k)$ cancels with $1/(\mu_j^{(k)} - \mu_r^{(t)})$ 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 $\delta_{tr} = \beta_{t+1}|\eta_{tr}^{(t)}|$, and some $\mu_r^{(t)}$ has therefore stabilized. Equation (3.20) shows that some $\mu_r^{(t)}$ must be close to $\mu_j^{(k)}$, 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 $\mu_j^{(k)}$, so that $\mu_j^{(k)}$ cannot be a well-separated eigenvalue of T_k. Conversely if $\mu_j^{(k)}$ is a well-separated eigenvalue of T_k, then (3.16) holds, and if $\mu_j^{(k)}$ 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.

Chris Paige used the splitting

$$\begin{aligned} \mathbf{y}_j^{\mathsf{H}} \mathbf{y}_j &= \mathbf{s}_j^{\mathsf{H}} \mathbf{Q}_k^{\mathsf{H}} \mathbf{Q}_k \mathbf{s}_j \\ &= \mathbf{s}_j^{\mathsf{H}} (\mathbf{R}_k^{\mathsf{H}} + \mathbf{D}_k + \mathbf{R}_k) \mathbf{s}_j \\ &= 1 + \mathbf{s}_j^{\mathsf{H}} (\mathbf{D}_k - \mathbf{I}_k) \mathbf{s}_j^{\mathsf{H}} \\ &\quad + 2\mathrm{Re}\,(\mathbf{s}_j^{\mathsf{H}} \mathbf{R}_k \mathbf{s}_j) \end{aligned} \tag{15}$$

Caution: notational changes!

C. Paige:		this talk:
$z_j^{(k)}$	$\Leftrightarrow$	$\mathbf{y}_j$
$y_j^{(k)}$	$\Leftrightarrow$	$\mathbf{s}_j$
$\beta_{k+1}\eta_{kj}^{(k)}$	$\Leftrightarrow$	$\beta_k s_{kj}$
$\mu_j^{(k)}$	$\Leftrightarrow$	$\theta_j^{(k)} = \theta_j$

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 $\mathbf{T}_k$ is involved:

$$\mathbf{T}_k \mathbf{W}_k + \mathbf{G}_k = \mathbf{W}_{k+1} \underline{\mathbf{T}}_k = \mathbf{W}_k \mathbf{T}_k + \mathbf{w}_{k+1} \beta_k \mathbf{e}_k^{\mathsf{T}}. \tag{HessT1}$$

Here, the basis vectors $\mathbf{w}_j$ describe the loss of orthogonality and the perturbation term has a large rank-one part (i.e., large last row),

$$\begin{aligned} \mathbf{W}_{k+1} &:= \mathbf{Q}_k^{\mathsf{H}} \mathbf{Q}_{k+1}, \\ \mathbf{G}_k &:= \mathbf{e}_k \beta_k \mathbf{q}_{k+1}^{\mathsf{H}} \mathbf{Q}_k + \mathbf{Q}_k^{\mathsf{H}} \mathbf{F}_k - \mathbf{F}_k^{\mathsf{H}} \mathbf{Q}_k. \end{aligned} \tag{16}$$

Chris Paige's approach

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

$$\mathbf{A}\mathbf{Q}_k + \mathbf{F}_k = \mathbf{Q}_{k+1}\underline{\mathbf{T}}_k = \mathbf{Q}_k\mathbf{T}_k + \mathbf{q}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}}, \qquad \text{(HessA1)}$$

with $\mathbf{Q}_k^{\mathsf{H}}$ from the left gives

$$\mathbf{Q}_k^{\mathsf{H}}\mathbf{A}\mathbf{Q}_k + \mathbf{Q}_k^{\mathsf{H}}\mathbf{F}_k = \mathbf{Q}_k^{\mathsf{H}}\mathbf{Q}_k\mathbf{T}_k + \mathbf{Q}_k^{\mathsf{H}}\mathbf{q}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}}, \qquad (17)$$

and (17)−(17)$^{\mathsf{H}}$ gives

$$\mathbf{T}_k\mathbf{W}_k + \mathbf{G}_k = \mathbf{W}_{k+1}\underline{\mathbf{T}}_k = \mathbf{W}_k\mathbf{T}_k + \mathbf{w}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}} \qquad \text{(HessT1)}$$

with

$$\mathbf{G}_k = \mathbf{e}_k\beta_k\mathbf{q}_{k+1}^{\mathsf{H}}\mathbf{Q}_k + \mathbf{Q}_k^{\mathsf{H}}\mathbf{F}_k - \mathbf{F}_k^{\mathsf{H}}\mathbf{Q}_k, \qquad (18)$$

since $\mathbf{A} = \mathbf{A}^{\mathsf{H}}$ and $\mathbf{T}_k = \mathbf{T}_k^{\mathsf{T}}$ 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:

$$
\begin{aligned}
\mathbf{y}_j^{\mathsf{H}}\mathbf{y}_j = \mathbf{s}_j\mathbf{Q}_k^{\mathsf{H}}\mathbf{Q}_k\mathbf{s}_j = \mathbf{s}_j^{\mathsf{H}}\mathbf{W}_k\mathbf{s}_j &= \frac{\beta_{1:k-1}}{\omega(\theta_j)}\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{W}_k\boldsymbol{\nu}(\theta_j) \\
&= \frac{1}{\omega(\theta_j)}\left(\mathcal{A}_k(\theta_j,\theta_j)\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{Q}_k^{\mathsf{H}}\mathbf{q}_1 + \sum_{\ell=1}^{k}\beta_{1:\ell-1}\mathcal{A}_{\ell+1:k}(\theta_j,\theta_j)\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{g}_\ell\right) \\
&= \hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{Q}_k^{\mathsf{H}}\mathbf{q}_1 + \sum_{\ell=1}^{k}\frac{\beta_{1:k-1}}{\omega(\theta_j)}\nu_\ell'(\theta_j)\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{g}_\ell \\
&= \hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{Q}_k^{\mathsf{H}}\mathbf{q}_1 + \frac{\boldsymbol{\nu}(\theta_j)^{\mathsf{H}}\mathbf{G}_k\boldsymbol{\nu}'(\theta_j)}{\boldsymbol{\nu}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)} \\
&= \hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{Q}_k^{\mathsf{H}}\mathbf{q}_1 + \frac{\beta_k\mathbf{q}_{k+1}^{\mathsf{H}}\mathbf{Q}_k\boldsymbol{\nu}'(\theta_j)}{\boldsymbol{\nu}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)} + \frac{\boldsymbol{\nu}(\theta_j)^{\mathsf{H}}(\mathbf{Q}_k^{\mathsf{H}}\mathbf{F}_k - \mathbf{F}_k^{\mathsf{H}}\mathbf{Q}_k)\boldsymbol{\nu}'(\theta_j)}{\boldsymbol{\nu}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)}.
\end{aligned}
\tag{19}
$$

Here, $\omega(\theta_j) := \chi'(\theta_j)$ and $\mathcal{A}_{\ell+1:k}(z,w) := \chi_{\ell+1:k}[z,w] = \beta_{\ell:k-1}\nu_\ell[z,w]$.

Chris Paige's approach

We consider the terms in this representation of $\|\mathbf{y}_j\|_2^2$. We start with the first term.

In the exact case, i.e., if $\mathbf{Q}_k$ is orthonormal,

$$\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{Q}_k^{\mathsf{H}}\mathbf{q}_1 = 1, \quad \text{since} \quad \hat{\nu}_1(z) \equiv 1. \tag{20}$$

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

$$\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{Q}_k^{\mathsf{H}}\mathbf{q}_1 = \sum_{l=1}^{k} \frac{\chi_{1:l-1}(\theta_j)}{\beta_{1:l-1}} \mathbf{q}_l^{\mathsf{H}}\mathbf{q}_1. \tag{21}$$

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

Chris Paige's approach

Both other terms in our expression for $\|\mathbf{y}_j\|_2^2$ are of the form

$$\frac{\boldsymbol{\nu}(\theta_j)^{\mathsf{H}}\mathbf{X}_k\boldsymbol{\nu}'(\theta_j)}{\boldsymbol{\nu}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)} = \frac{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{X}_k\boldsymbol{\nu}'(\theta_j)}{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)}. \tag{22}$$

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

For those not familiar with eigenvector perturbations:

$$|\sin\angle(\boldsymbol{\nu}(\theta_j+\Delta\theta_j),\boldsymbol{\nu}(\theta_j))| = \frac{\|\mathbf{P}_{\boldsymbol{\nu}(\theta_j)^{\perp}}\boldsymbol{\nu}(\theta_j+\Delta\theta_j)\|_2}{\|\boldsymbol{\nu}(\theta_j+\Delta\theta_j)\|_2} \tag{23}$$

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

$$|\sin\angle(\boldsymbol{\nu}(\theta_j+\Delta\theta_j),\boldsymbol{\nu}(\theta_j))| = \frac{\|\mathbf{P}_{\boldsymbol{\nu}(\theta_j)^\perp}\boldsymbol{\nu}'(\theta_j)\|_2}{\|\boldsymbol{\nu}(\theta_j)\|_2}|\Delta\theta_j| + O(|\Delta\theta_j|^2). \tag{24}$$

Thus, we need "nice" expressions for

$$\frac{\boldsymbol{\nu}(\theta_i)^{\mathsf{H}}\boldsymbol{\nu}'(\theta_j)}{\|\boldsymbol{\nu}(\theta_i)\|_2\|\boldsymbol{\nu}(\theta_j)\|_2} = \frac{\hat{\boldsymbol{\nu}}(\theta_i)^{\mathsf{H}}\boldsymbol{\nu}'(\theta_j)}{\|\hat{\boldsymbol{\nu}}(\theta_i)\|_2\|\boldsymbol{\nu}(\theta_j)\|_2}. \tag{25}$$

It turns out to be easy to obtain analytic expressions for

$$\frac{\hat{\boldsymbol{\nu}}(\theta_i)^{\mathsf{H}}\boldsymbol{\nu}'(\theta_j)}{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)} = \begin{cases} \dfrac{1}{\theta_j-\theta_i}, & j\neq i, \\ \displaystyle\sum_{\ell\neq j}\dfrac{1}{\theta_j-\theta_\ell}, & j=i. \end{cases} \tag{26}$$

Chris Paige's approach

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

$$|\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)| = \|\hat{\boldsymbol{\nu}}(\theta_j)\|_2\|\boldsymbol{\nu}(\theta_j)\|_2. \tag{27}$$

Thus, we need an expression for

$$\begin{aligned}\frac{|\boldsymbol{\nu}(\theta_i)^{\mathsf{H}}\boldsymbol{\nu}'(\theta_j)|}{\|\boldsymbol{\nu}(\theta_i)\|_2\|\boldsymbol{\nu}(\theta_j)\|_2} &= \frac{\|\hat{\boldsymbol{\nu}}(\theta_j)\|_2}{\|\hat{\boldsymbol{\nu}}(\theta_i)\|_2}\frac{|\hat{\boldsymbol{\nu}}(\theta_i)^{\mathsf{H}}\boldsymbol{\nu}'(\theta_j)|}{|\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)|} \\ &= \begin{cases} \dfrac{\|\hat{\boldsymbol{\nu}}(\theta_j)\|_2}{\|\hat{\boldsymbol{\nu}}(\theta_i)\|_2}\dfrac{1}{|\theta_j-\theta_i|}, & j \neq i, \\ \left|\displaystyle\sum_{\ell\neq j}\frac{1}{\theta_j-\theta_\ell}\right|, & j = i. \end{cases}\end{aligned} \tag{28}$$

Chris Paige's approach

Observe that the norms of the eigenvectors

$$\|\hat{\boldsymbol{\nu}}(\theta_j)\|_2^2 = \frac{1}{s_{1j}^2} \tag{29}$$

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

$$\begin{aligned} s_{kj}^2 &= \frac{\chi_{1:k-1}(\theta_j)}{\omega(\theta_j)} = \frac{1}{\|\boldsymbol{\nu}(\theta_j)\|_2^2}, \\ s_{1j}^2 &= \frac{\chi_{2:k}(\theta_j)}{\omega(\theta_j)} = \frac{1}{\|\hat{\boldsymbol{\nu}}(\theta_j)\|_2^2}, \end{aligned} \tag{30}$$

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

$$\omega(z) = \prod_{\ell \neq j} (z - \theta_\ell). \tag{31}$$

Chris Paige's approach

By classical perturbation theory

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

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

$$\begin{aligned}\sin^2\angle(\hat{\boldsymbol{\nu}}(\theta_j), \boldsymbol{\nu}(\theta_j) + \boldsymbol{\nu}'(\theta_j)\Delta\theta_j) &= \frac{\|\mathbf{P}_{\hat{\boldsymbol{\nu}}(\theta_j)^\perp}\boldsymbol{\nu}'(\theta_j)\|_2^2}{\|\boldsymbol{\nu}(\theta_j)\|_2^2}|\Delta\theta_j|^2 + O(|\Delta\theta_j|^3) \\ &= \frac{|\Delta\theta_j|^2}{s_{1j}^2}\sum_{\ell\neq j}\frac{s_{1\ell}^2}{(\theta_j - \theta_\ell)^2} + O(|\Delta\theta_j|^3).\end{aligned} \quad (33)$$

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

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

valid for all Ritz values θ_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 $\|\hat{\boldsymbol{\nu}}(\theta_j)\|_2$ and $\|\boldsymbol{\nu}(\theta_j)\|_2$ has to be "large",

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

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

$$s_{1j}s_{kj} = \frac{\beta_{1:k-1}}{\omega(\theta_j)} = \frac{1}{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)}. \tag{36}$$

Chris Paige's approach

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

$$|s_{1j}s_{kj}| < \frac{1}{2}, \tag{37}$$

since for a vector $\mathbf{x}$ with non-zero structure as follows,

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

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

$$\mathbf{s}_j^{\mathsf{T}} \left(\beta_\ell \mathbf{e}_{\ell+1} = (\mathbf{T}_k - \alpha_\ell)\mathbf{e}_\ell - \beta_{\ell-1}\mathbf{e}_{\ell-1}\right). \tag{39}$$

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.

We reconsider

$$\frac{\boldsymbol{\nu}(\theta_j)^{\mathsf{H}}\mathbf{X}_k\boldsymbol{\nu}'(\theta_j)}{\boldsymbol{\nu}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)} = \frac{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{X}_k\boldsymbol{\nu}'(\theta_j)}{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)}. \tag{40}$$

Inserting the identity matrix gives

$$\begin{aligned}\frac{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{X}_k\boldsymbol{\nu}'(\theta_j)}{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)} &= \sum_{i=1}^{k}\frac{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{X}_k\boldsymbol{\nu}(\theta_i)}{\hat{\boldsymbol{\nu}}(\theta_i)^{\mathsf{H}}\boldsymbol{\nu}(\theta_i)}\,\frac{\hat{\boldsymbol{\nu}}(\theta_i)^{\mathsf{H}}\boldsymbol{\nu}'(\theta_j)}{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)} \\ &= \sum_{i\neq j}\frac{1}{\theta_j-\theta_i}\left(\frac{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{X}_k\boldsymbol{\nu}(\theta_i)}{\hat{\boldsymbol{\nu}}(\theta_i)^{\mathsf{H}}\boldsymbol{\nu}(\theta_i)} + \frac{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{X}_k\boldsymbol{\nu}(\theta_j)}{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)}\right).\end{aligned} \tag{41}$$

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

We only used the first Hessenberg decomposition with $\mathbf{T}_k$. We can stick closer to what Chris Paige did, and use the second one:

$$\mathbf{T}_k\mathbf{R}_k + \mathbf{E}_k = \mathbf{R}_{k+1}\underline{\mathbf{T}}_k = \mathbf{R}_k\mathbf{T}_k + \mathbf{r}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}}. \qquad \text{(HessT2)}$$

Here, $\mathbf{E}_k$ is upper triangular, and $\mathbf{W}_{k+1} = \mathbf{R}_k^{\mathsf{H}} + \mathbf{D}_k + \mathbf{R}_{k+1}$ with $\mathbf{R}_{k+1} = \text{sut}(\mathbf{W}_{k+1})$ and $\mathbf{D}_k$ diagonal.

Chris Paige proved that $\mathbf{E}_k$ is "small".

Based on the identity

$$\|\mathbf{y}_j\|_2^2 - 1 = \mathbf{s}_j^{\mathsf{H}}(\mathbf{D}_k - \mathbf{I}_k)\mathbf{s}_j + 2\text{Re}\left(\mathbf{s}_j^{\mathsf{H}}\mathbf{R}_k\mathbf{s}_j\right) \qquad (42)$$

Chris Paige bounded the deviation of $\|\mathbf{y}_j\|$ from one.

Chris Paige's approach

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

$$\begin{aligned} \mathbf{s}_j^{\mathsf{H}} \mathbf{R}_k \mathbf{s}_j &= \frac{\beta_{1:k-1}}{\omega(\theta_j)} \hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}} \mathbf{R}_k \boldsymbol{\nu}(\theta_j) \\ &= \frac{1}{\omega(\theta_j)} \left(\mathcal{A}_k(\theta_j, \theta_j) \hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}} \mathbf{r}_1 + \sum_{\ell=1}^{k} \beta_{1:\ell-1} \mathcal{A}_{\ell+1:k}(\theta_j, \theta_j) \hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}} \mathbf{E}_k \mathbf{e}_\ell \right) \\ &= \sum_{\ell=1}^{k} \frac{\beta_{1:k-1}}{\omega(\theta_j)} \nu'_\ell(\theta_j) \hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}} \mathbf{E}_k \mathbf{e}_\ell = \frac{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}} \mathbf{E}_k \boldsymbol{\nu}'(\theta_j)}{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}} \boldsymbol{\nu}(\theta_j)}. \end{aligned} \tag{43}$$

Thus,

$$\|\mathbf{y}_j\|_2^2 - 1 = \mathbf{s}_j^{\mathsf{H}} (\mathbf{D}_k - \mathbf{I}_k) \mathbf{s}_j + 2\mathrm{Re} \left(\frac{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}} \mathbf{E}_k \boldsymbol{\nu}'(\theta_j)}{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}} \boldsymbol{\nu}(\theta_j)} \right). \tag{44}$$

Chris Paige's approach

We can reformulate this by our "perturbation analysis":

$$\frac{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{E}_k\boldsymbol{\nu}'(\theta_j)}{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)} = \sum_{\ell\neq j}\frac{1}{\theta_j-\theta_i}\left(\frac{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{E}_k\boldsymbol{\nu}(\theta_i)}{\hat{\boldsymbol{\nu}}(\theta_i)^{\mathsf{H}}\boldsymbol{\nu}(\theta_i)} + \frac{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\mathbf{E}_k\boldsymbol{\nu}(\theta_j)}{\hat{\boldsymbol{\nu}}(\theta_j)^{\mathsf{H}}\boldsymbol{\nu}(\theta_j)}\right). \tag{45}$$

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

$$\mathbf{s}_j^{\mathsf{H}}\left(\mathbf{T}_k\mathbf{R}_k + \mathbf{E}_k = \mathbf{R}_k\mathbf{T}_k + \mathbf{r}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}}\right)\mathbf{s}_j, \tag{46}$$

proves that loss of orthogonality and "convergence" go hand in hand,

$$\epsilon_{jj}^{(k)} = \mathbf{s}_j^{\mathsf{H}}\mathbf{Q}_k^{\mathsf{H}}\mathbf{q}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}}\mathbf{s}_j = \mathbf{y}_j^{\mathsf{H}}\mathbf{q}_{k+1}\beta_k s_{kj}. \tag{47}$$

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 $\mathbf{A}$. This can be seen with the aid of the shifted Hessenberg decomposition.

A new approach

Consider the shifted Lanczos Hessenberg decomposition

$$\tilde{\mathbf{A}}\mathbf{Q}_k + \tilde{\mathbf{F}}_k = \mathbf{Q}_{k+1}\underline{\mathbf{T}}_k = \mathbf{Q}_k\mathbf{T}_k + \mathbf{q}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}} \tag{HessA2}$$

where for a given eigenpair $\mathbf{A}\mathbf{v}_i = \mathbf{v}_i\lambda_i$ and a given Ritz value θ_j we defined

$$\tilde{\mathbf{A}} := \mathbf{A} - (\lambda_i - \theta_j)\mathbf{v}_i\mathbf{v}_i^{\mathsf{H}} \quad \text{and} \quad \tilde{\mathbf{F}}_k := (\lambda_i - \theta_j)\mathbf{v}_i\mathbf{v}_i^{\mathsf{H}}\mathbf{Q}_k + \mathbf{F}_k. \tag{48}$$

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

$$\mathbf{v}_i^{\mathsf{H}}\tilde{\mathbf{A}} = \mathbf{v}_i^{\mathsf{H}}(\mathbf{A} - (\lambda_i - \theta_j)\mathbf{v}_i\mathbf{v}_i^{\mathsf{H}}) = \lambda_i\mathbf{v}_i^{\mathsf{H}} - (\lambda_i - \theta_j)\mathbf{v}_i^{\mathsf{H}}\mathbf{v}_i\mathbf{v}_i^{\mathsf{H}} = \theta_j\mathbf{v}_i^{\mathsf{H}}, \tag{49}$$

i.e., $\mathbf{v}_i$ is a left eigenvector to the eigenvalue θ_j.

A new approach

The angle between the eigenvector $\mathbf{v}_i$ and a scaled Ritz vector is given by

$$\frac{\beta_{1:k-1}}{\omega(\theta_j)}\mathbf{v}_i^{\mathsf{H}}\mathbf{Q}_k\boldsymbol{\nu}(\theta_j) = \mathbf{v}_i^{\mathsf{H}}\mathbf{q}_1 + \frac{\beta_{1:k-1}}{\omega(\theta_j)}\mathbf{v}_i^{\mathsf{H}}\tilde{\mathbf{F}}_k\boldsymbol{\nu}'(\theta_j), \tag{50}$$

in other words,

$$\begin{aligned}\mathbf{v}_i^{\mathsf{H}}\mathbf{Q}_k\boldsymbol{\nu}(\theta_j) &= \frac{\omega(\theta_j)}{\beta_{1:k-1}}\mathbf{v}_i^{\mathsf{H}}\mathbf{q}_1 + \mathbf{v}_i^{\mathsf{H}}\tilde{\mathbf{F}}_k\boldsymbol{\nu}'(\theta_j)\\ &= \frac{\omega(\theta_j)}{\beta_{1:k-1}}\mathbf{v}_i^{\mathsf{H}}\mathbf{q}_1 + (\lambda_i - \theta_j)\mathbf{v}_i^{\mathsf{H}}\mathbf{Q}_k\boldsymbol{\nu}'(\theta_j) + \mathbf{v}_i^{\mathsf{H}}\mathbf{F}_k\boldsymbol{\nu}'(\theta_j).\end{aligned} \tag{51}$$

Remark: This relation is correct, no matter how close or far away λ_i and θ_j are. The relation can be obtained using any eigenvalue and any Ritz value.

A new approach

Sorting gives the following anti-Taylor-like approximation,

$$\mathbf{v}_i^{\mathsf{H}}\mathbf{Q}_k\left(\boldsymbol{\nu}(\theta_j) - \boldsymbol{\nu}'(\theta_j)(\lambda_i - \theta_j)\right) = \frac{\omega(\theta_j)}{\beta_{1:k-1}}\mathbf{v}_i^{\mathsf{H}}\mathbf{q}_1 + \mathbf{v}_i^{\mathsf{H}}\mathbf{F}_k\boldsymbol{\nu}'(\theta_j), \tag{52}$$

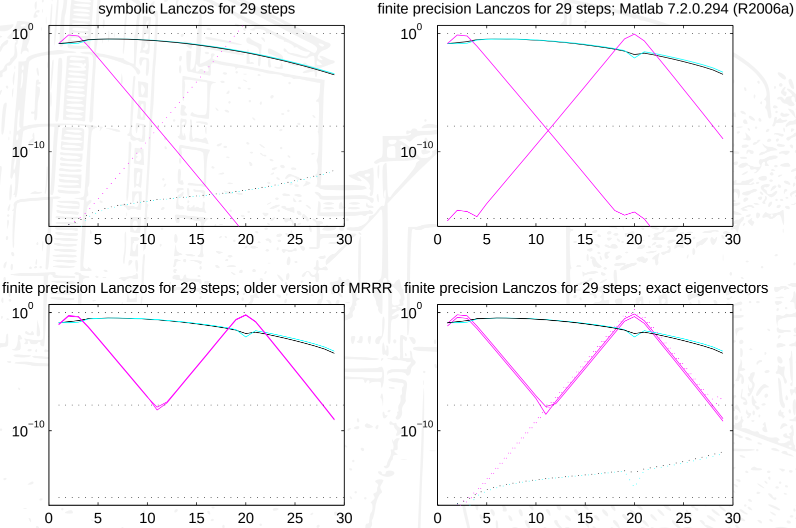
weighted summation over all eigenpairs of $\mathbf{A}$ gives the inexact inverse subspace iteration

$$((\theta_j\mathbf{I}_n - \mathbf{A})\mathbf{Q}_k - \mathbf{F}_k)\boldsymbol{\nu}'(\theta_j) = \frac{\omega(\theta_j)}{\beta_{1:k-1}}\mathbf{q}_1 - \mathbf{Q}_k\boldsymbol{\nu}(\theta_j). \tag{53}$$

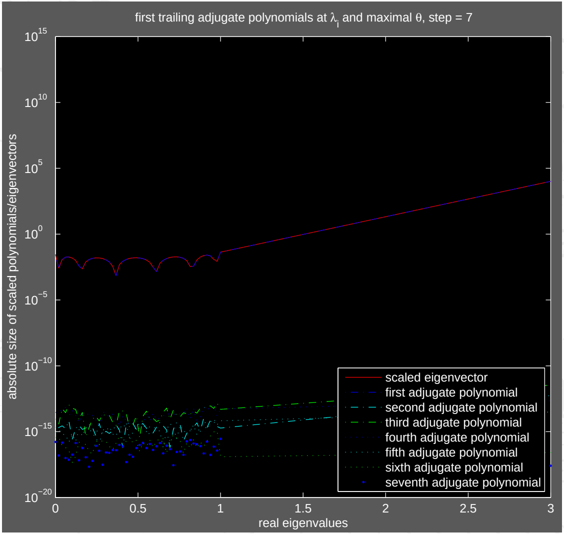
There is a good chance that $\mathbf{Q}_k\boldsymbol{\nu}'(\theta_j)$ is a better candidate for a “Ritz vector” if $\mathbf{Q}_k\boldsymbol{\nu}(\theta_j)$ is “small” and θ_j is close to an eigenvalue of $\mathbf{A}$.

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

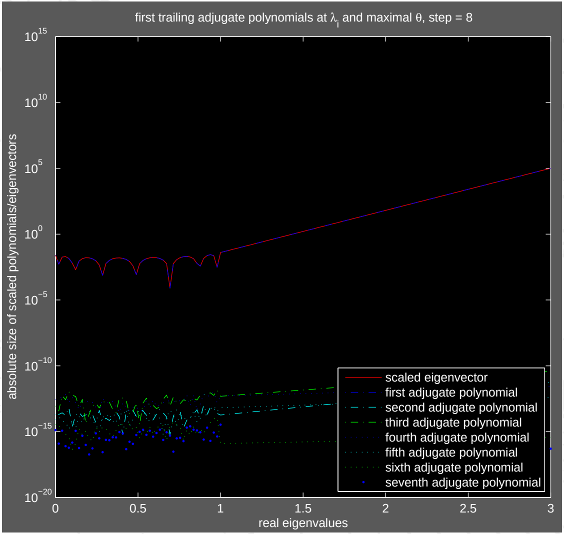
An example from my 2006 GAMM talk



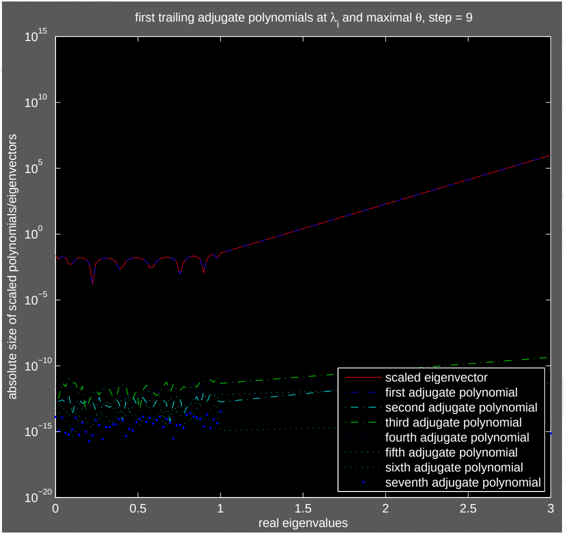
An example from my 2006 GAMM talk



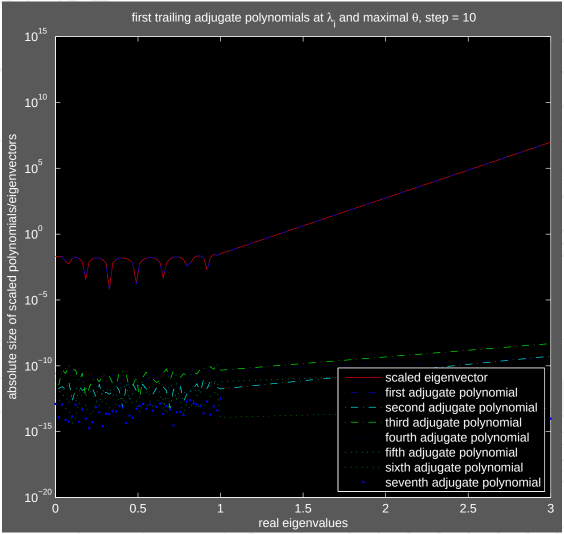
An example from my 2006 GAMM talk



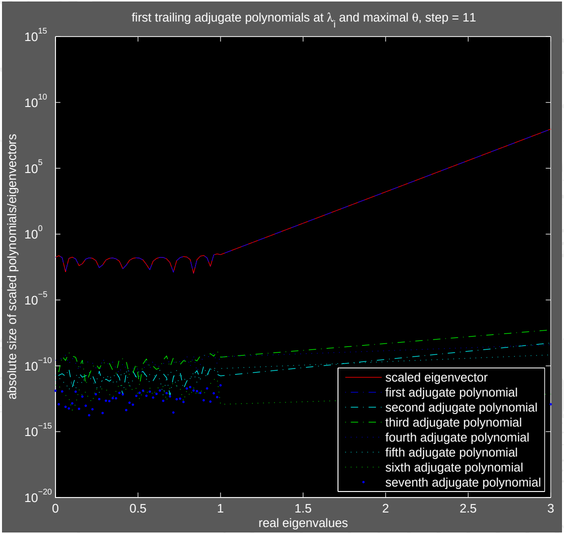
An example from my 2006 GAMM talk



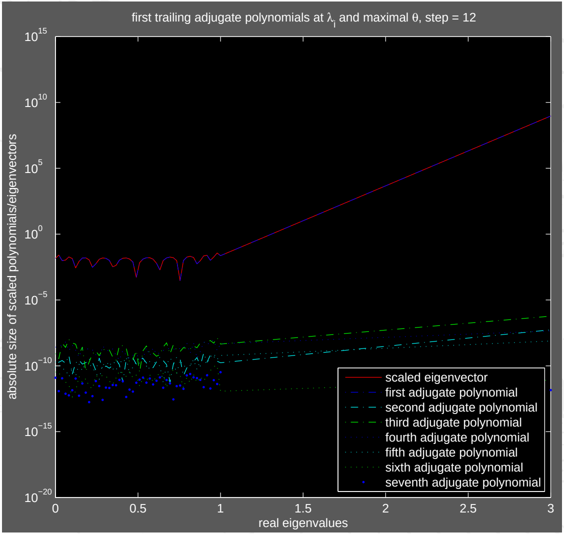
An example from my 2006 GAMM talk



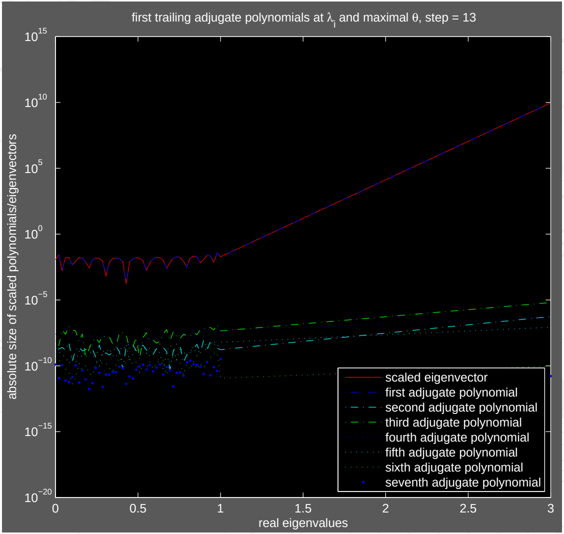
An example from my 2006 GAMM talk



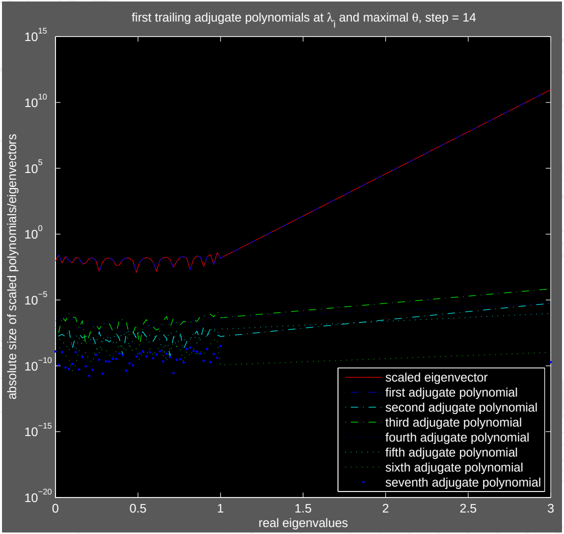
An example from my 2006 GAMM talk



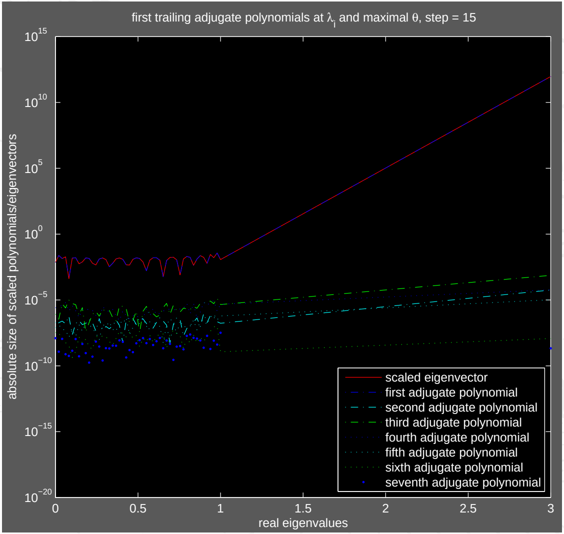
An example from my 2006 GAMM talk



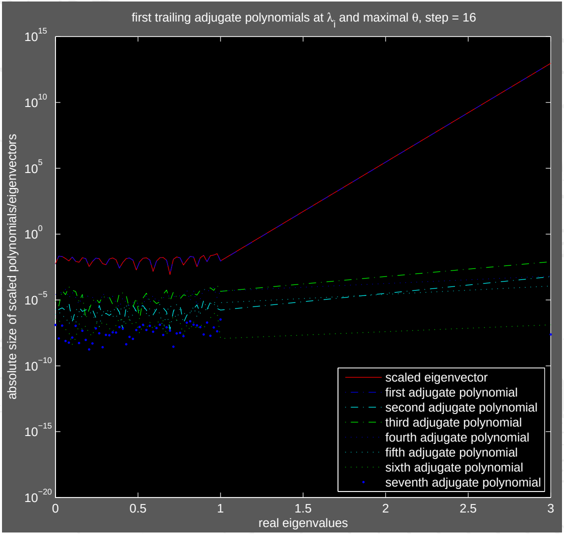
An example from my 2006 GAMM talk



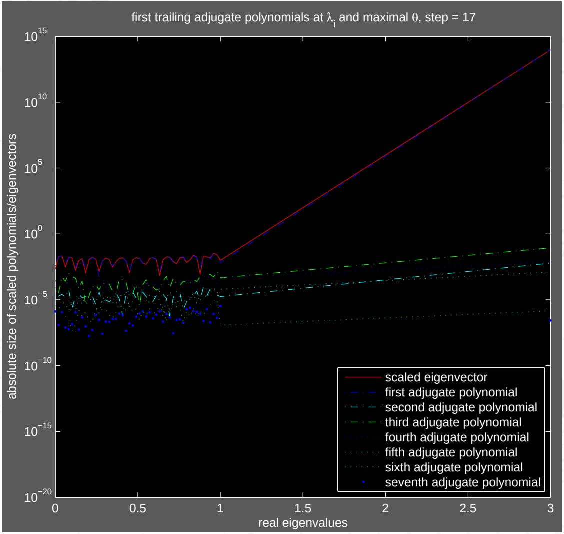
An example from my 2006 GAMM talk



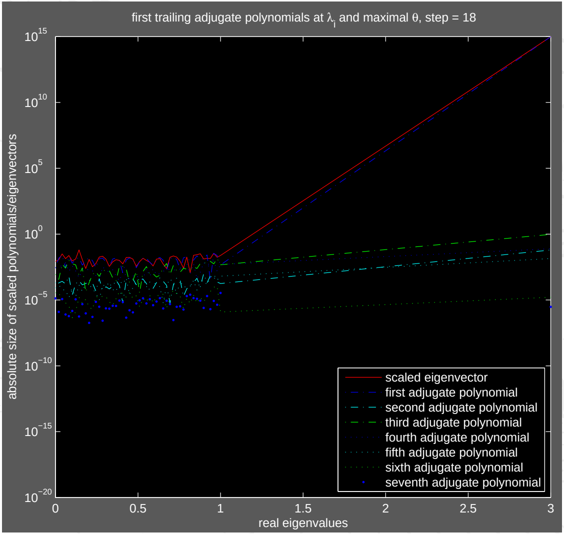
An example from my 2006 GAMM talk



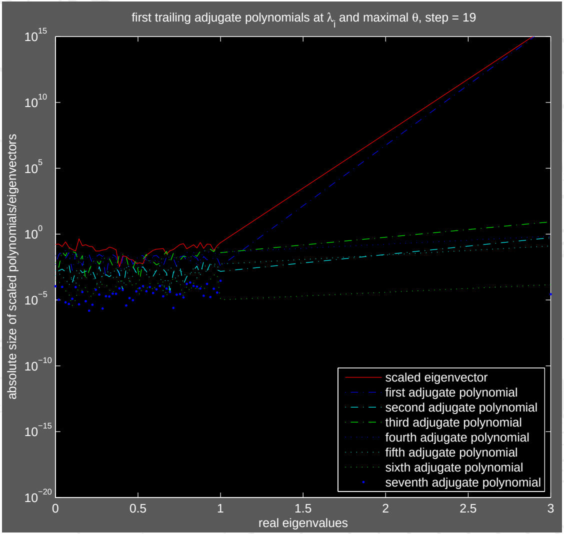
An example from my 2006 GAMM talk



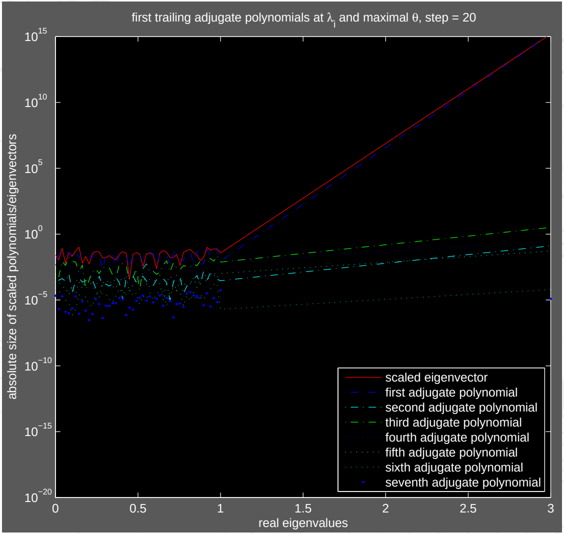
An example from my 2006 GAMM talk



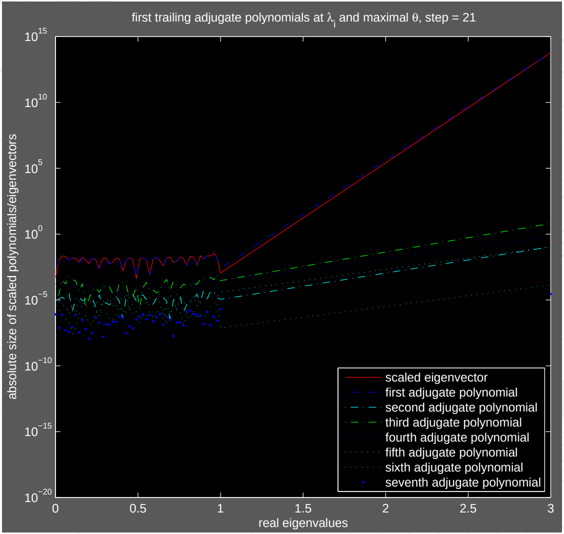
An example from my 2006 GAMM talk



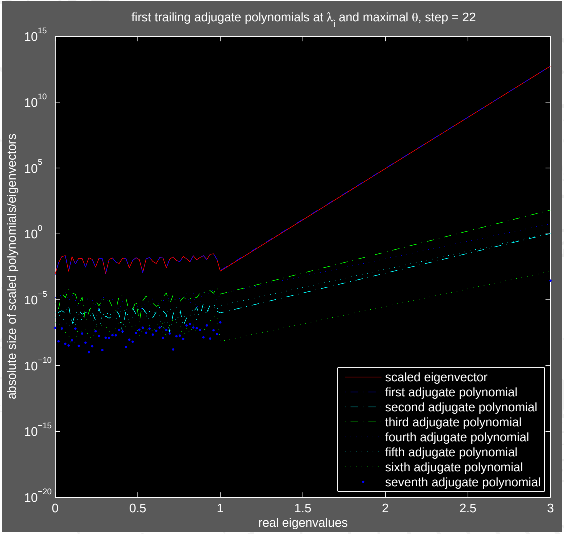
An example from my 2006 GAMM talk



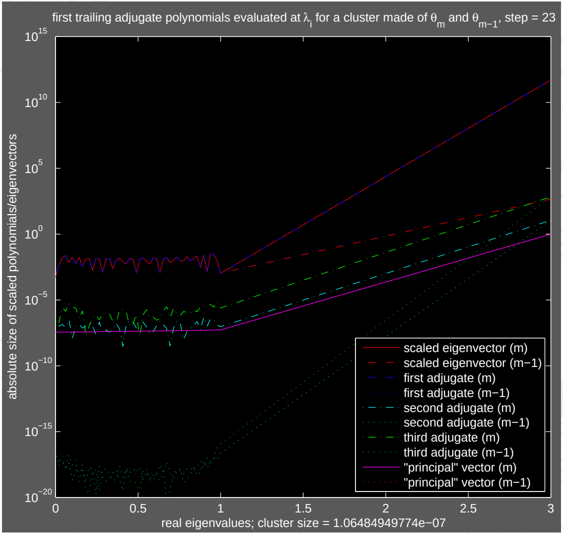
An example from my 2006 GAMM talk



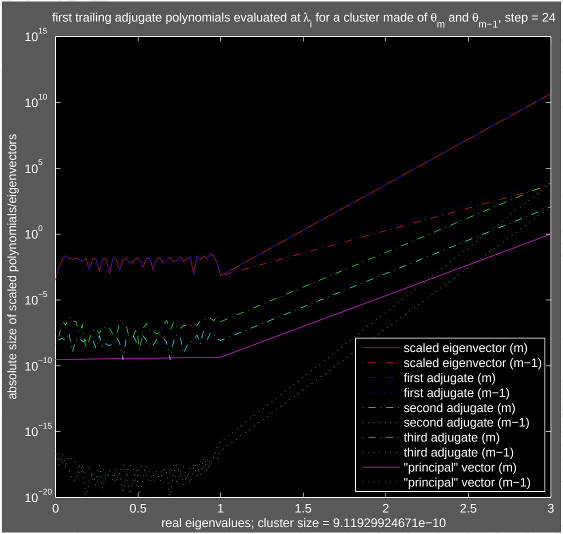
An example from my 2006 GAMM talk



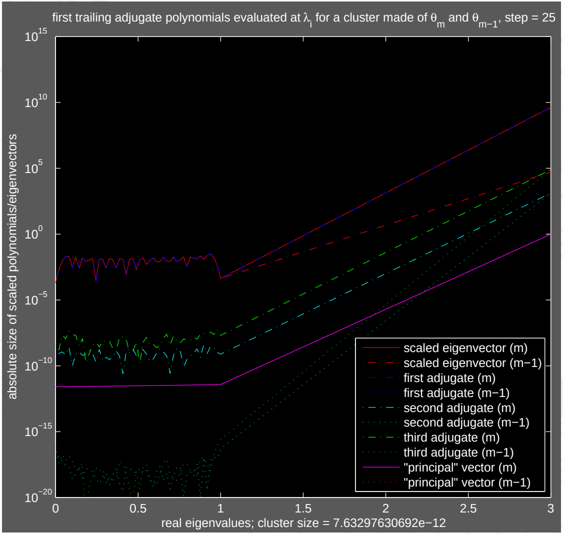
An example from my 2006 GAMM talk



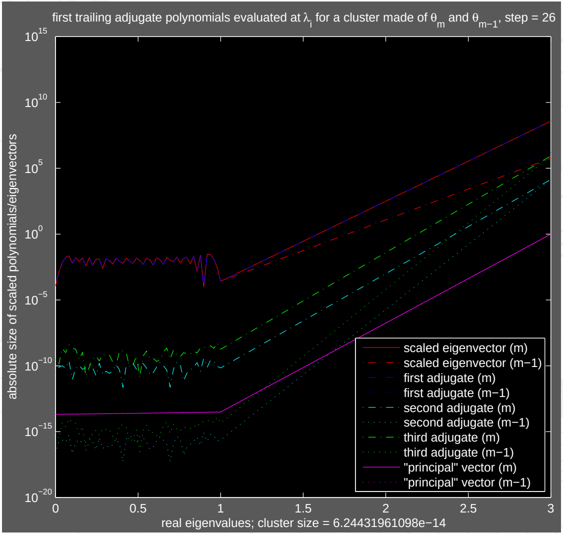
An example from my 2006 GAMM talk



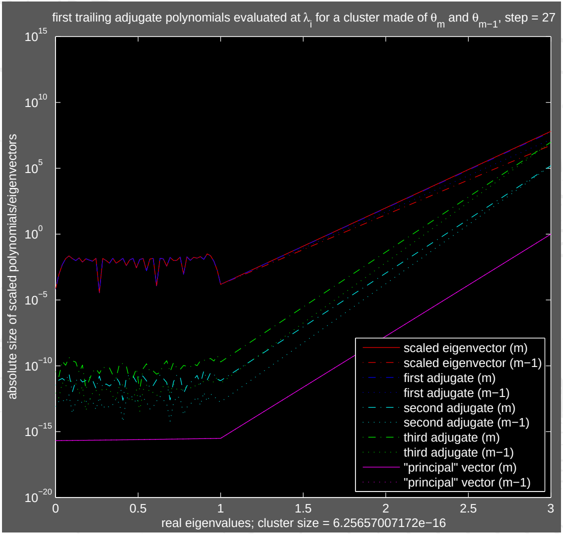
An example from my 2006 GAMM talk



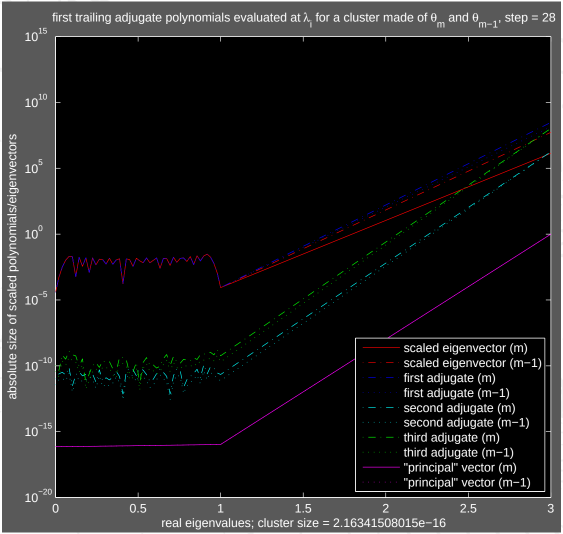
An example from my 2006 GAMM talk



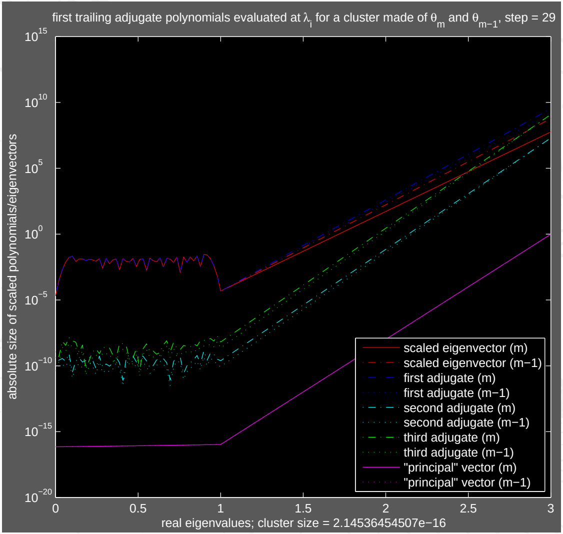
An example from my 2006 GAMM talk



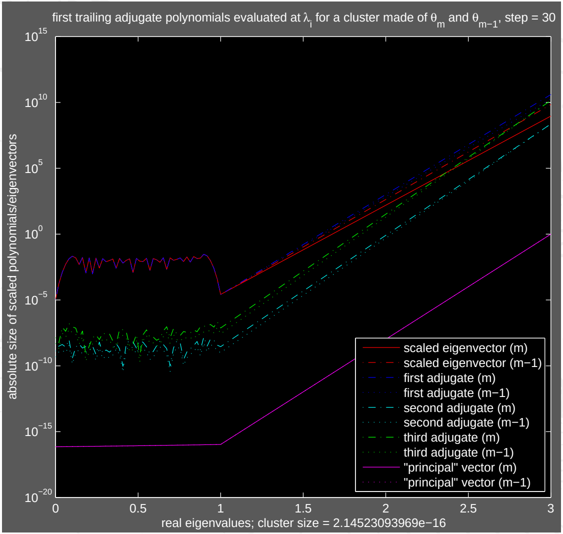
An example from my 2006 GAMM talk



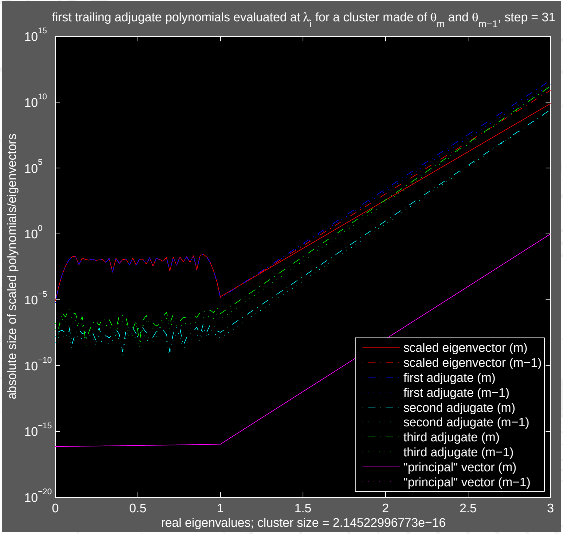
An example from my 2006 GAMM talk



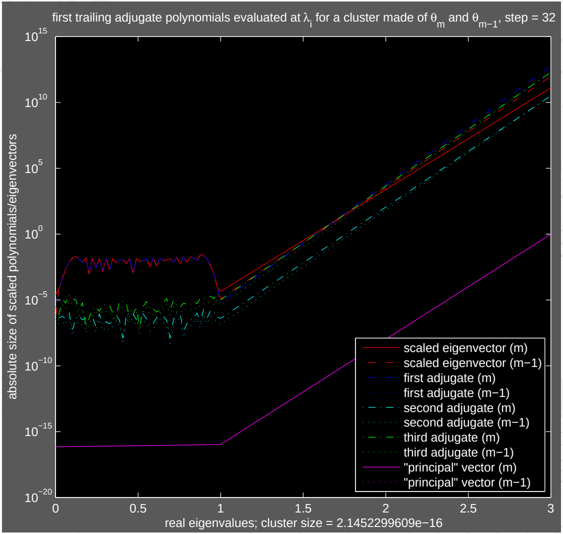
An example from my 2006 GAMM talk



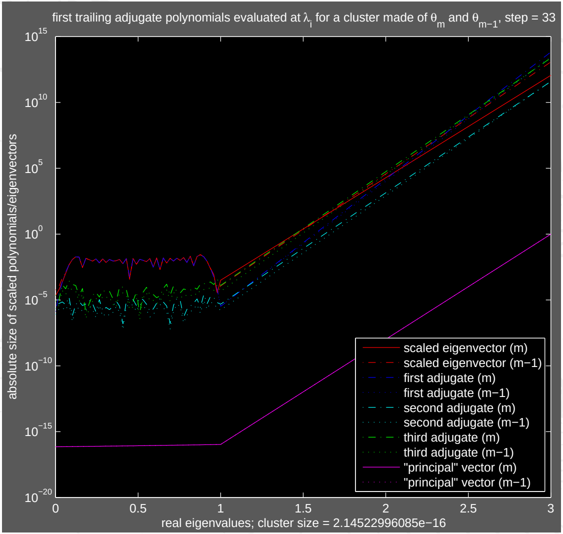
An example from my 2006 GAMM talk



An example from my 2006 GAMM talk



An example from my 2006 GAMM talk



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

$$(z\mathbf{I}_k - \mathbf{T}_k)\frac{\boldsymbol{\nu}^{(\ell)}(z)}{\ell!} + \frac{\boldsymbol{\nu}^{(\ell-1)}(z)}{(\ell-1)!} = \mathbf{e}_1 \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}}. \tag{54}$$

This implies that

$$\begin{aligned}(z\mathbf{Q}_k - \mathbf{Q}_k\mathbf{T}_k)\frac{\boldsymbol{\nu}^{(\ell)}(z)}{\ell!} + \mathbf{Q}_k\frac{\boldsymbol{\nu}^{(\ell-1)}(z)}{(\ell-1)!} &= \mathbf{Q}_k\mathbf{e}_1 \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}} \\ = ((z\mathbf{I}_n - \mathbf{A})\mathbf{Q}_k - \mathbf{F}_k)\frac{\boldsymbol{\nu}^{(\ell)}(z)}{\ell!} + \mathbf{Q}_k\frac{\boldsymbol{\nu}^{(\ell-1)}(z)}{(\ell-1)!} &= \mathbf{q}_1 \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}}.\end{aligned} \tag{55}$$

We have used the fact that the last ℓ components of $\boldsymbol{\nu}^{(\ell)}(z)$ 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 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,

$$((z\mathbf{I}_n - \mathbf{A})\mathbf{Q}_k - \mathbf{F}_k)\Big(\sum_{\ell=0}^{p} a_\ell \frac{\boldsymbol{\nu}^{(\ell)}(z)}{\ell!}\Big) + \mathbf{Q}_k\Big(\sum_{\ell=1}^{p} a_\ell \frac{\boldsymbol{\nu}^{(\ell-1)}(z)}{(\ell-1)!}\Big) = \mathbf{q}_1\Big(\sum_{\ell=1}^{p} a_\ell \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}}\Big). \quad (56)$$

Higher derivatives

We could try to find a linear combination

$$\sum_{\ell=1}^{p} a_\ell \frac{\boldsymbol{\nu}^{(\ell-1)}(z)}{(\ell-1)!} \tag{57}$$

that (almost) lies in the null-space of $\mathbf{Q}_k$. This linear combination of the derivatives would be close to an eigenvector of $\mathbf{A}$, if the corresponding linear combination

$$\sum_{\ell=1}^{p} a_\ell \frac{\chi^{(\ell)}(z)}{\beta_{1:k-1}} \tag{58}$$

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 $\mathbf{A}$ at λ.

Polynomial view on Chris Paige's result

We can consider the parameter-dependent relation

$$(\mathbf{T}_k - z\mathbf{I}_k)\mathbf{R}_k + \mathbf{E}_k = \mathbf{R}_k(\mathbf{T}_k - z\mathbf{I}_k) + \mathbf{r}_{k+1}\beta_k\mathbf{e}_k^{\mathsf{T}}. \tag{59}$$

Remember that $\mathbf{R}_k$ is a strictly upper triangular matrix.

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

$$\hat{\boldsymbol{\nu}}(z)^{\mathsf{H}}\mathbf{E}_k\boldsymbol{\nu}(z) = \hat{\boldsymbol{\nu}}(z)^{\mathsf{H}}\mathbf{r}_{k+1}\beta_k. \tag{60}$$

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

$$\hat{\boldsymbol{\nu}}(z)^{\mathsf{H}}\mathbf{E}_k\boldsymbol{\nu}(z) = \begin{pmatrix} 1 & \cdots & z^{k-1} \end{pmatrix} \begin{pmatrix} \star & \cdots & \star \\ & \ddots & \vdots \\ & & \star \end{pmatrix} \begin{pmatrix} z^{k-1} \\ \vdots \\ 1 \end{pmatrix} \tag{61}$$

Polynomial view on Chris Paige's result

This gives the complete characterization of the loss of orthogonality

$$\mathbf{r}_{k+1}\beta_k = \mathbf{Q}_k^{\mathsf{H}}\mathbf{q}_{k+1}\beta_k \tag{62}$$

at step $k+1$ in terms of the errors $\mathbf{E}_k$.

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}:

$$\mathrm{trace}(\mathbf{E}_k)z^{k-1} + \cdots = \hat{\boldsymbol{\nu}}(z)^{\mathsf{H}}\mathbf{E}_k\boldsymbol{\nu}(z) = \hat{\boldsymbol{\nu}}(z)^{\mathsf{H}}\mathbf{r}_{k+1}\beta_k = \mathbf{q}_k^{\mathsf{H}}\mathbf{q}_{k+1}\beta_k z^{k-1} + \cdots . \tag{63}$$

This is correct. It does not give further insights, but proves that the relation is sound. The diagonal of $\mathbf{E}_k$ is closely related to the local loss of orthogonality.

Polynomial view on Chris Paige's result

Maybe of interest in CG or other OR methods is the relation involving the constant terms, namely

$$\hat{\boldsymbol{\nu}}(0)^{\mathsf{H}} \mathbf{E}_k \boldsymbol{\nu}(0) = \hat{\boldsymbol{\nu}}(0)^{\mathsf{H}} \mathbf{Q}_k^{\mathsf{H}} \mathbf{q}_{k+1} \beta_k. \tag{64}$$

By definition of $\boldsymbol{\nu}(z)$, $\mathbf{z}_k$ defined by

$$\mathbf{z}_k \frac{\chi(0)}{\|\mathbf{r}_0\| \beta_{1:k-1}} := -\boldsymbol{\nu}(0) = -(-\mathbf{T}_k)^{-1} \frac{\chi(0)}{\beta_{1:k-1}} \mathbf{e}_1, \tag{65}$$

where $\mathbf{r}_0 := \mathbf{b} - \mathbf{A}\mathbf{x}_0$ denotes the starting residual, is the kth QOR solution, see (Z, 2007).

At this point the talk comes to its end. The true research can start here.

Conclusion and Outlook

- 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.

The final slide . . .

Děkuji. Once Again.

Greenbaum, A. (1989).
Behavior of slightly perturbed Lanczos and conjugate-gradient recurrences.
Linear Algebra Appl., 113:7–63.

Greenbaum, A. and Strakoš, Z. (1992).
Predicting the behavior of finite precision Lanczos and conjugate gradient computations.
SIAM J. Matrix Anal. Appl., 13(1):121–137.

Meurant, G. (2006).
The Lanczos and Conjugate Gradient Algorithms: From Theory to Finite Precision Computation.
SIAM, Philadelphia.

Meurant, G. and Strakoš, Z. (2006).
The Lanczos and conjugate gradient algorithms in finite precision arithmetic.
Acta Numerica, 15:471–542.

Paige, C. C. (1971).
The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices.
PhD thesis, Institute of Computer Science, London University.

Paige, C. C. (1972).
Computational variants of the Lanczos method for the eigenproblem.
J. Inst. Math. Appl., 10:373–381.

Paige, C. C. (1976).
Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix.
J. Inst. Math. Appl., 18(341–349).

Paige, C. C. (1980).
Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem.
Linear Algebra Appl., 34:235–258.

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

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