Quasi-Minimal-Residual Eigenpairs

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

Institut für Numerische Simulation
Technische Universität Hamburg-Harburg

01.04.2008
Vortrag zur GAMM-Jahrestagung 2008 in Bremen
S-17: ‘Applied and Numerical Linear Algebra’

Outline

Lehmann's work on eigenvalues

Between 1948 and 1966 N. J. Lehmann published several papers related to "Optimale Eigenwerteinschließungen". Lehmann was interested in selfadjoint and normal linear operators (matrices).

In his works we can find eigenvalue inclusions using the Temple-Quotient, shifted harmonic Ritz values, and the relation of shifted harmonic Ritz to Ritz-Galërkin, all for selfadjoint matrices.

He was interested in generalizing his results [Seite 246, 1963]:

> *"Für Aufgaben mit komplexen Eigenwerten stehen viele der Untersuchungen allerdings noch aus. Mit diesen Problemen befaßt sich eine in Arbeit befindliche Dissertation."*

We extend his approach to general complex square matrices by replacing "Lehmann optimality" by "backward error".

Lehmann's results summarized

Lehmann used the information included in $Q \in \mathbb{C}^{(n,k)}$ and $W = AQ \in \mathbb{C}^{(n,k)}$, where $A \in \mathbb{C}^{(n,n)}$ is selfadjoint. We use a generic $x = Qv \in \mathbb{C}^n$.

Lehmann imposed the least-squares optimality conditions [(5a), 1963]

$$\min = \sigma^2(z) = \frac{\|(A - zI)x\|_2^2}{\|x\|_2^2} = \frac{\|(W - zQ)v\|_2^2}{\|Qv\|_2^2}$$

and thus (by differentation) the eigenvalue (SVD) problem [(8a), 1963]

$$Q^H(A - zI)^H(A - zI)Qv = \sigma^2(z)Q^H Qv.$$

Lehmann was interested in optimal shifts, i.e., shifts z resulting in a minimal radius $\sigma(z)$ of the inclusion. These are [Satz 4, 1963] among the stationary points of $\sigma^2(z)$,

$$\frac{\partial \sigma^2(z)}{\partial z} = 0.$$

Lehmann's little-known results

Differentiating an expression involving the Temple quotient $T_\tau(x)$, he obtained the shifted harmonic Ritz values [(20a)+(28), 1963] of Morgan (1991) and Freund (1992),

$$Q^H(A-\tau I)^H Q v = \frac{1}{\underline{\theta}(\tau)-\tau} Q^H(A-\tau I)^H(A-\tau I)Qv.$$

Lehmann noticed already that poles occur in the shifted harmonic Ritz approach when using the Ritz values as shifts.

He (defined and) noted certain interesting symmetries/properties, namely

$$\tau = z \mp \sigma(z), \qquad \underline{\theta}(\tau) = z \pm \sigma(z), \qquad \text{[(Seite 251),1963]}$$

$$\underline{\theta}(\tau) = T_\tau(\underline{x}), \qquad T_\tau(\underline{x}) = \frac{\underline{x}^H(A-\tau I)^H(A-\tau I)\underline{x}}{\underline{x}^H(A-\tau I)^H\underline{x}} + \tau, \qquad \text{[(15),1963]}$$

$$2z = \tau + \underline{\theta}(\tau), \qquad z^2 - \sigma^2(z) = \tau \cdot \underline{\theta}(\tau). \qquad \text{[(8b)+(21),1963]}$$

Krylov decompositions

We consider a given Krylov decomposition

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

We suppose that

$$A \in \mathbb{C}^{(n,n)}$$ is a general square matrix,

$$Q_{k+1} = \begin{pmatrix} Q_k & q_{k+1} \end{pmatrix} \in \mathbb{C}^{(n,k+1)}$$ is a matrix of basis vectors,

$$\underline{C}_k = \begin{pmatrix} C_k \\ c_{k+1,k} e_k^T \end{pmatrix} \in \mathbb{C}^{(k+1,k)}$$ is an extended Hessenberg matrix.

We assume that $\underline{C}_k$ is unreduced. We do not consider perturbations.

We remark that important parts of the results carry over to general rectangular approximations $\underline{C}_k$ which not necessarily have to be Hessenberg.

QMR eigenpairs

We proceed similar to the QMR approach when applied to linear systems,

$$\min_{z,y=Q_k v} \frac{\|zy - Ay\|}{\|y\|} \leqslant \kappa(Q_{k+1}) \cdot \min_{z,v} \frac{\|(z\underline{I}_k - \underline{C}_k)v\|}{\|v\|}. \tag{2}$$

We always suppose that the columns of Q_{k+1} have been scaled to unit length.

Definition (QMR eigenpair)

The pair $(\grave{\theta}, \grave{y} = Q_k \grave{v})$ is a QMR eigenpair, when

$$\|(\grave{\theta}\underline{I}_k - \underline{C}_k)\grave{v}\| = \operatorname*{min\,loc}_{z\in\mathbb{C}, v\in\mathbb{C}^k, \|v\|=1} \|(z\underline{I}_k - \underline{C}_k)v\|, \tag{3}$$

where “min loc” denotes a (not necessarily strict) local minimum.

QMR eigenpairs: SVD characterization

We denote the SVD of ${}^z\underline{C}_k \equiv z\underline{I}_k - \underline{C}_k$ by $U(z)\Sigma(z)V(z)^H = U\,\Sigma(z)V^H$.

Since for every $z \in \mathbb{C}$

$$\sigma_k(z) = \|\sigma_k(z)u_k\| = \frac{\|{}^z\underline{C}_k v_k\|}{\|v_k\|} = \min_v \frac{\|{}^z\underline{C}_k v\|}{\|v\|},$$

the QMR eigenvalues can be characterized by

$$\grave{\theta} = \underset{z\in\mathbb{C}}{\arg\min\operatorname{loc}}\, \sigma_k(z), \tag{4}$$

the QMR eigenvector can be chosen as a corresponding right singular vector,

$$\grave{v} = v_k(\grave{\theta}), \tag{5}$$

the QMR eigenresidual is given by $\sigma_k(\grave{\theta})$.

QMR eigenpairs: SVD steepest descent

Simple singular values $\sigma(z)$ and corresponding singular vectors v_k, u_k of the complex matrices ${}^z\underline{C}_k = z\underline{I}_k - \underline{C}_k$ are real analytic (Sun, 1988),

$$\sigma(z+w) = \sigma(z) + \sigma_z(z)w + \sigma_{\bar{z}}(z)\overline{w} + O(|w|^2) \tag{6}$$

$$= \sigma(z) + 2\Re((u_k^H \underline{I}_k v_k)w) + O(|w|^2). \tag{7}$$

We obtain steepest descent by subtracting the conjugate of the gradient $\sigma_z(z)$:

$$z_{\text{new}} = z - \alpha\,\overline{u_k^H \underline{I}_k v_k} = z - \alpha\, v_k^H \underline{I}_k^H u_k \tag{8}$$

$$= z - \frac{\alpha}{\sigma_k} v_k^H \underline{I}_k^H (z\underline{I}_k - \underline{C}_k) v_k = z - \frac{\alpha}{\sigma_k} v_k^H (zI_k - C_k) v_k. \tag{9}$$

We note that $\sigma_k(z)$ is the backward error of the approximate eigenvalue z. Setting $\alpha = \sigma_k$ yields alternating projections and is nearly optimal:

$$z_{\text{new}} = v_k^H C_k v_k. \tag{10}$$

QMR eigenpairs: SVD Newton

Steepest descent exhibits linear convergence. The real-analyticity of simple singular values can also be used to adopt Newton's method for stationary points.

Newton's method exhibits the usual locally quadratic convergence behavior, but in most cases for Newton's method good starting values have to be used, better than, say, the Ritz values.

In general Newton's method has problems with clustered and multiple singular values and when far from a solution, as the function to be optimized is almost linear far from stationary points.

The latter problem is resolved when a damped Newton's method is used.

Multiple singular values can not occur in the symmetric case due to the unreduced Hessenberg structure, but still may be pathologically close, compare with results by Lehmann and Wilkinson.

QMR eigenpairs: Grassmannian Optimization

It can be shown that the QMR eigenvectors are minima of real-analytic

$$\begin{aligned} \lambda &: G_1(\mathbb{C}^k) \to \mathbb{R}_{\geqslant 0} \\ \lambda(v) &= \|((v^H C_k v)\underline{I}_k - \underline{C}_k)v\|^2 = v^H(\underline{C}_k)^H \underline{C}_k v - |v^H C_k v|^2. \end{aligned} \tag{11}$$

The stationary points of real-analytic λ are always singular vectors.

When the stationary point is a minimum and the associated singular value is simple, then $\grave{v} = v$ is an QMR eigenvector.

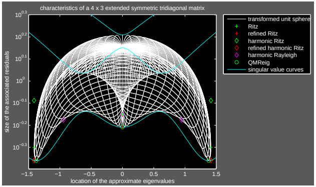
The associated QMR eigenvalue is characterized by $\grave{\theta} = \grave{v}^H C_k \grave{v}$, and the QMR eigenresidual is given by $\sigma(\grave{\theta}) = \sqrt{\lambda(\grave{v})}$.

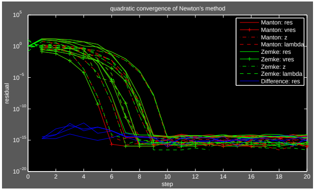
We experimented with steepest descent and Newton's method for the minimization of the real-analytic λ on the first (complex) Grassmannian in the framework of optimization on Riemannian manifolds (as recently developed by Smith; Edelman, Arias & Smith; Manton).

QMR eigenpairs: Grassmannian Newton

```
for j = 1:convergence

  [Q,R] = qr(v); W = Q(:,2:k); v = Q(:,1);

  z = v'*Ck*v; zuCk = z*uIk-uCk;
  zuCkW = zuCk*W; zuCkv = zuCk*v;
  slambda = norm(zuCkv);
  y = zuCkW'*zuCkv;
  grad = 2*[real(y);imag(y)];
  res = norm(grad);

  A = zuCkW'*zuCkW; zCk = uIk'*zuCk;
  g1 = (zCk*W)'*v; r1 = real(g1); c1 = imag(g1);
  g2 = W'*(zCk*v); r2 = real(g2); c2 = imag(g2);
  outer1 = [r1+r2;c1+c2];
  outer2 = [c2-c1;r1-r2];
  Hesse = 2*[real(A) imag(A)';...
             imag(A) real(A)]-...
          2*slambda^2*I-...
          2*outer1*outer1'-2*outer2*outer2';
  ab = Hesse\grad;
  u = -W*(ab(1:k-1)+i*ab(k:2*k-2));
  normu = norm(u);
  v = v*cos(normu)+u*sin(normu)/normu;

end
```

(This is to convince you that the code is short enough to fit on one page.)





A graphical representation

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

$$z = \tilde{\theta}, \quad w = \frac{\|(\tilde{\theta}\underline{I}_k - \underline{C}_k)\tilde{s}\|}{\|\tilde{s}\|}. \tag{12}$$

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

The norm of the residual of an eigenpair gives the backward error, i.e.,

$$w = \min \left\{ \|\Delta A\| \;:\; (A + \Delta A)\tilde{y} = \tilde{y}\,\tilde{\theta} \right\}. \tag{13}$$

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

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

A beautiful example

As an example we use

$$\underline{C}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{14}$$

Its Ritz values are given by

$$\theta_{1,3} = \mp\sqrt{2} \approx \mp 1.41421356, \quad \theta_2 = 0, \tag{15}$$

its harmonic Ritz values are given by

$$\underline{\theta}_{1,3} = \mp\sqrt{2} \approx \mp 1.41421356, \quad \underline{\theta}_2 = \infty, \tag{16}$$

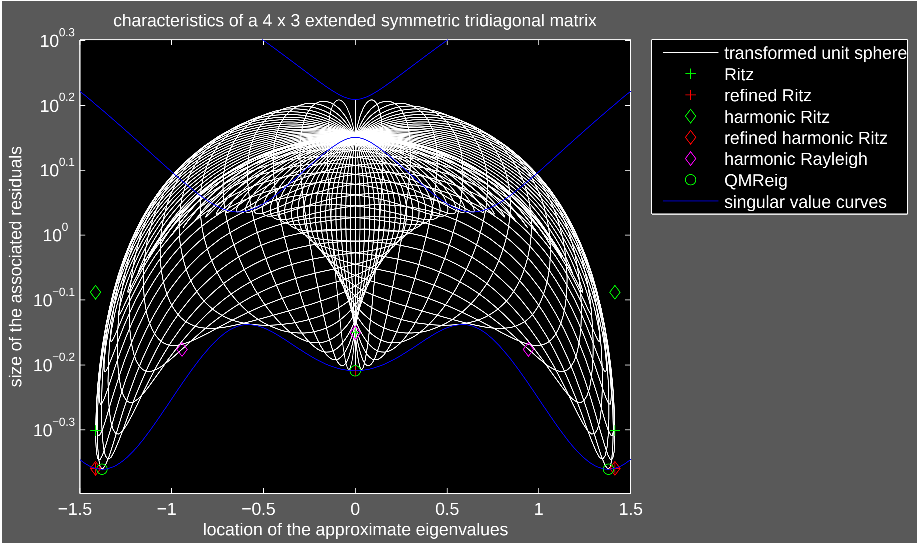
its ρ-values (Rayleigh quotients with harmonic Ritz vectors) are given by

$$\rho_{1,3} = \mp\sqrt{2}\cdot\frac{2}{3} \approx \mp 0.9428090, \quad \rho_2 = 0, \tag{17}$$

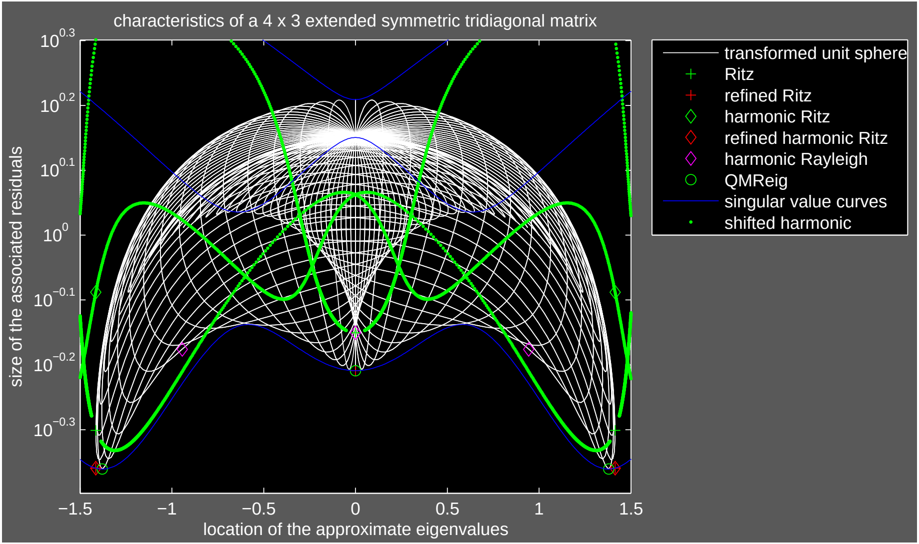
and its QMR eigenvalues are given by (where $y = 276081 + 21504\sqrt{2}i$)

$$\grave{\theta}_{1,3} = \mp\frac{\sqrt{2}}{16}\sqrt{113 + 2\Re\sqrt[3]{y}} \approx \mp 1.37898323557, \quad \grave{\theta}_2 = 0. \tag{18}$$

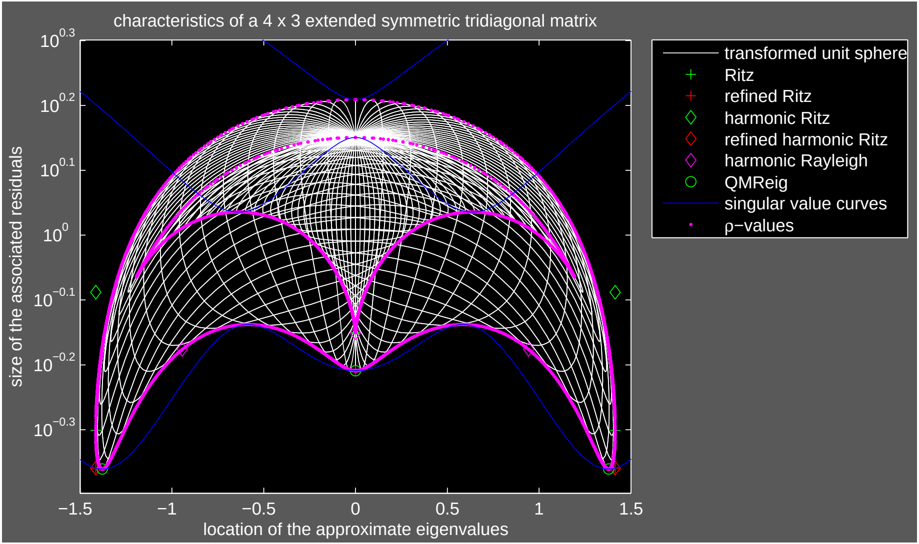
A beautiful example



A beautiful example



A beautiful example



A beautiful example

Why are the ρ-values on the "borders" of the transformed unit sphere?

In the symmetric case it is easy to characterize these "borders" and to prove that the vectors defining them are indeed harmonic Ritz vectors for two certain shifts. These shifts as well as the harmonic Ritz values are expressed using stationary points along straight lines in the graphical interpretation ...

Given the harmonic Ritz pair, it is even easier to find the direction along which the vector is a stationary point.

The QMR eigenvectors are harmonic Ritz vectors, the shifts are given by

$$\tau_{\pm} = \grave{\theta} \pm \sigma_k(\grave{\theta}), \tag{19}$$

in accordance with Lehmann's results for the symmetric case, see also van den Eshof's doctoral thesis (2003).

The Ritz vectors are harmonic Ritz vectors with shifts $\tau_{\pm} = \pm\infty$.

A Jordan block: infinitely many QMR eigenvalues

A more startling example is

$$\underline{C}_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{20}$$

We have Jordan blocks at $\theta = 0$, $\underline{\theta} = \infty$ and $\rho = 0$.

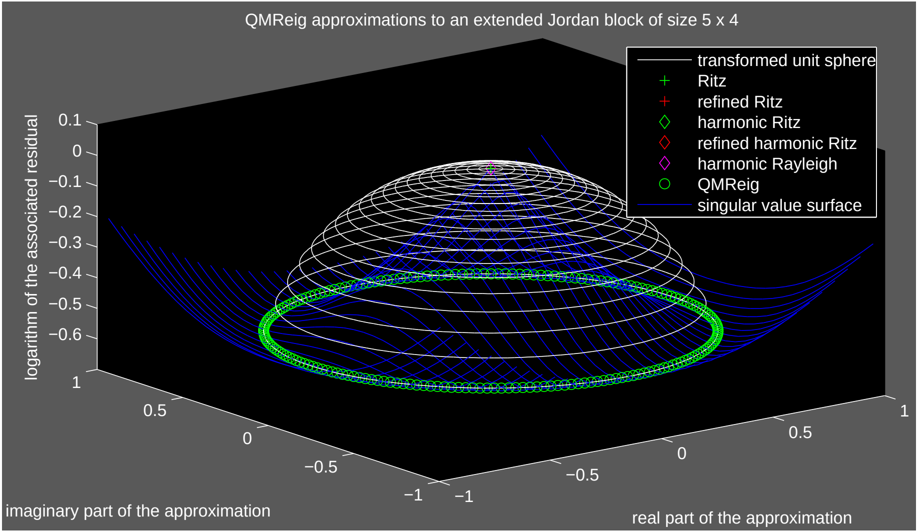
For $k \in \mathbb{N}$ this is an example of an infinite set of QMR eigenvalues,

$$\grave{\theta}_\phi = \cos\left(\frac{\pi}{k+1}\right) e^{i\phi}, \quad \phi \in [0, 2\pi). \tag{21}$$

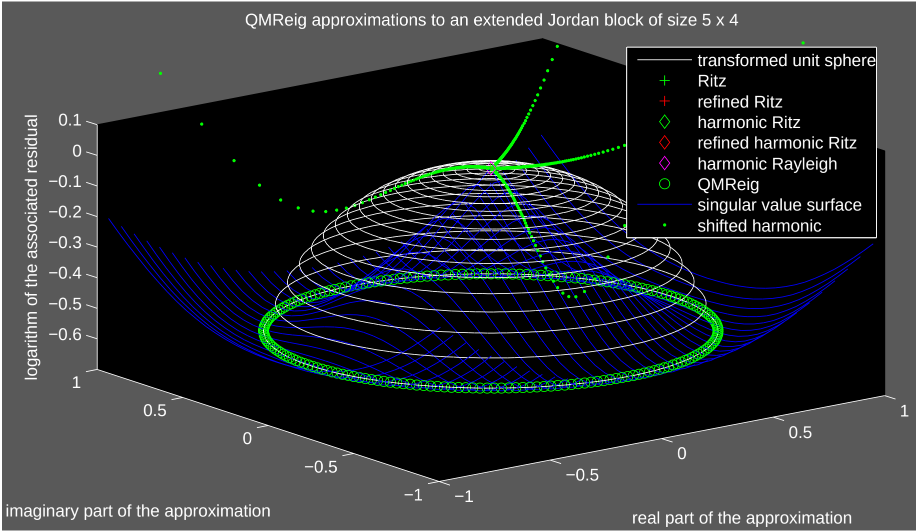
The residual of the corresponding QMR eigenpairs is given by

$$\|(\grave{\theta}_\phi \underline{I}_k - \underline{C}_k)\grave{s}_\phi\| = \sin\left(\frac{\pi}{k+1}\right) \quad \forall\ \phi. \tag{22}$$

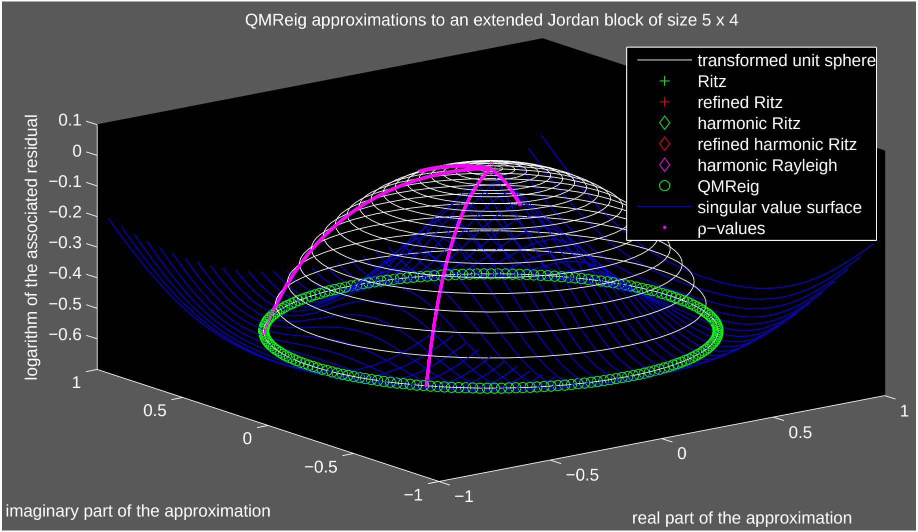
A Jordan block: infinitely many QMR eigenvalues



A Jordan block: infinitely many QMR eigenvalues



A Jordan block: infinitely many QMR eigenvalues



Always remember: It's only locally optimal . . .