The effect of aggressive early deflation on the convergence of the QR algorithm

Daniel Kressner

TU Berlin

kressner@math.ethz.ch

Harrachov 2007

Computational Methods with Applications

08/25/07

QR and QZ algorithms

QR algorithm	standard method for solving dense nonsymmetric eigenvalue problems (`eig(A)`)
QZ algorithm	standard method for solving dense generalized eigenvalue problems (`eig(A,B)`)

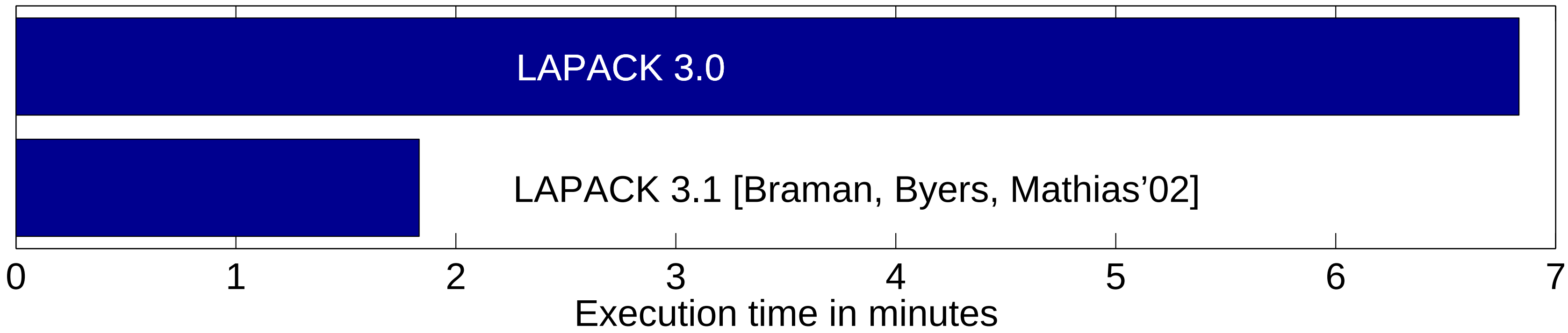
Both have recently undergone significant improvements.

QR and QZ algorithms

QR algorithm	standard method for solving dense nonsymmetric eigenvalue problems (`eig(A)`)
QZ algorithm	standard method for solving dense generalized eigenvalue problems (`eig(A,B)`)

Both have recently undergone significant improvements.

QR for 2961×2961 matrix (Matrix Market's PDE2961):

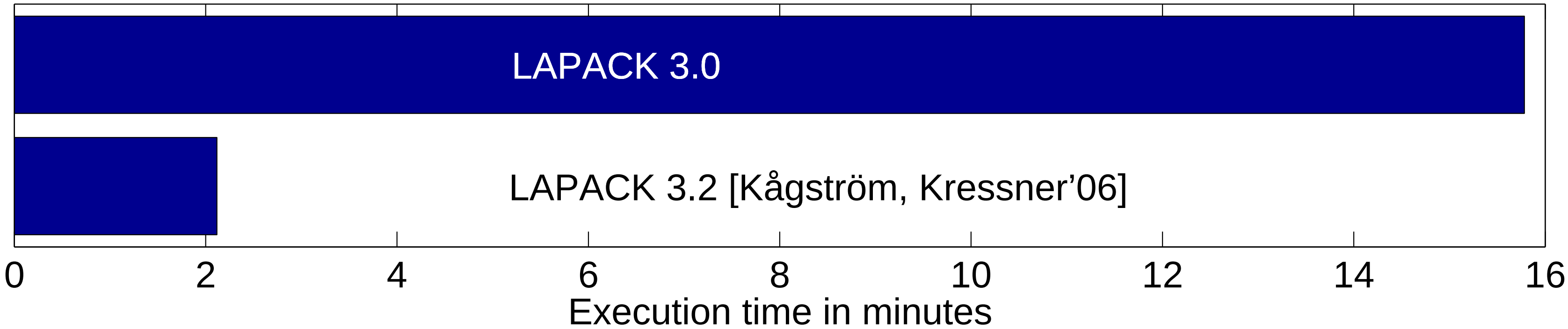


QR and QZ algorithms

QR algorithm	standard method for solving dense nonsymmetric eigenvalue problems (`eig(A)`)
QZ algorithm	standard method for solving dense generalized eigenvalue problems (`eig(A,B)`)

Both have recently undergone significant improvements.

QZ for 3600×3600 matrix pencil (Matrix Market's BCSST21):

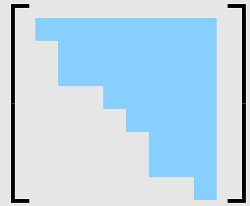


New ingredients: block multishift + advanced deflation techniques.

Goal of QR algorithm

For real $n \times n$ matrix A, compute orthogonal Q s.t.

$$Q^T A Q = T =$$



,

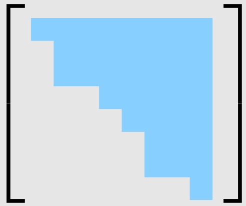
where T is in real Schur form.

- Diagonal of T yields eigenvalues of A.
- First k columns of Q span invariant subspace $\mathcal{X}$: $A\mathcal{X} \subseteq \mathcal{X}$.

Goal of QR algorithm

For real $n \times n$ matrix A, compute orthogonal Q s.t.

$$Q^T A Q = T =$$



,

where T is in real Schur form.

- Diagonal of T yields eigenvalues of A.
- First k columns of Q span invariant subspace $\mathcal{X}$: $A\mathcal{X} \subseteq \mathcal{X}$.

Goal of this talk

- Show intimate relation between QR algorithm and Krylov subspace techniques.
- Provide intuitive explanation and convergence bounds capturing observed improvements.

The basic multishift QR algorithm

QR generates sequence of orthogonally similar matrices

$$A_0, A_1, A_2, \ldots$$

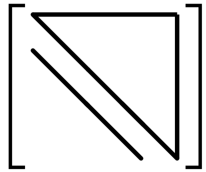
The basic multishift QR algorithm

QR generates sequence of orthogonally similar matrices

$$A_0, A_1, A_2, \ldots$$

(i) Initial reduction to Hessenberg form

$$A_0 \leftarrow Q_0^{\mathsf{T}} A Q_0 =$$



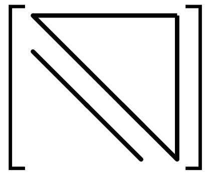
The basic multishift QR algorithm

QR generates sequence of orthogonally similar matrices

$$A_0, A_1, A_2, \ldots$$

(i) Initial reduction to Hessenberg form

$$A_0 \leftarrow Q_0^{\mathsf{T}} A Q_0 =$$



(ii) QR iterations (preserve Hessenberg form)

for $i \leftarrow 1, 2, \ldots$

Select $m \ll n$ shifts $\sigma_1, \ldots, \sigma_m$.

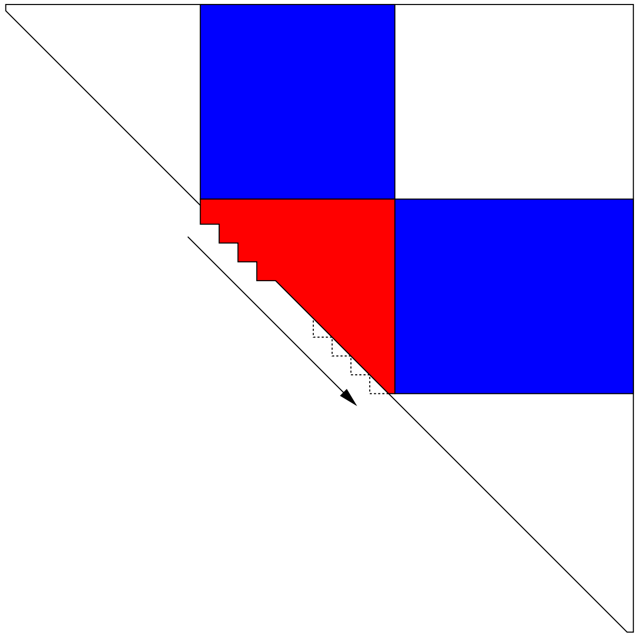
QR factorization $p_i(A_{i-1}) = Q_i R_i$ with $p_i(z) = \prod_{j=1}^{m} (z - \sigma_j)$.

Update $A_i \leftarrow Q_i^{\mathsf{T}} A_{i-1} Q_i$.

end for

Careful implementation: implicit, bulge chasing.

Idea behind blocked multishift QR



Do computations only locally (red area), delay and accumulate updates of rest (blue area).

- Bulk of computation becomes level BLAS 3 (matrix-matrix multiplications).
- Typically reduces execution time by factor 2–3.

[Lang'99], [Braman, Byers, Mathias'02].

Progress made by multishift QR iterations

`A = hess(randn(11));` After 1 iteration $\ldots$

0	1	1	1	0	1	1	1	1	1	0
1	0	1	0	0	1	0	0	0	1	1
	1	0	0	−1	1	1	1	1	1	1
		1	0	1	1	0	0	1	0	0
			1	1	0	0	1	1	1	1
				1	1	1	0	1	0	0
					1	−1	1	0	1	0
						0	0	0	1	0
							1	0	0	1
								0	0	0
									−1	1

shifts $=$ eigenvalues of bottom right 4×4 submatrix.

Progress made by multishift QR iterations

`A = hess(randn(11));` After 2 iterations $\ldots$

0	1	0	1	1	0	1	0	1	1	1
1	0	−2	0	1	0	0	0	1	0	0
	0	1	1	1	0	1	−1	0	1	0
		1	1	0	0	1	0	1	0	1
			1	1	1	0	−1	1	1	0
				0	1	0	0	1	0	0
					1	0	1	0	0	0
						−1	0	1	0	0
							0	0	0	0
								0	1	1
									−2	1

shifts $=$ eigenvalues of bottom right 4×4 submatrix.

Progress made by multishift QR iterations

`A = hess(randn(11));` After 3 iterations $\ldots$

1	1	0	−1	1	0	1	0	0	1	1
1	0	−1	−1	0	0	−1	1	0	1	−1
	0	1	1	1	0	1	0	0	1	1
		1	1	0	0	1	1	1	−2	1
			0	0	1	0	1	0	1	0
				1	0	0	0	1	0	0
					1	1	0	0	0	1
						−5	0	0	0	1
							0	0	0	0
								−2	1	1
									−3	1

shifts $=$ eigenvalues of bottom right 4×4 submatrix.

Progress made by multishift QR iterations

`A = hess(randn(11));` After 4 iterations $\ldots$

0	1	0	0	1	0	1	0	0	1	1
1	0	0	−2	1	0	0	0	1	1	0
	0	1	1	1	1	1	0	0	1	0
		1	1	0	1	0	1	1	0	1
			−1	1	1	1	1	0	1	1
				1	0	0	1	0	−1	0
					0	1	−1	1	0	0
						−10	0	0	0	1
							0	0	−1	0
								−7	1	1
									−3	1

shifts $=$ eigenvalues of bottom right 4×4 submatrix.

Progress made by multishift QR iterations

`A = hess(randn(11));` After 5 iterations $\ldots$

0	1	0	0	0	0	0	1	0	0	0
1	0	0	−1	1	1	0	0	0	1	−1
	0	1	1	1	0	1	0	0	1	−1
		1	1	1	1	1	1	0	0	0
			−1	1	1	1	1	−1	1	0
				0	1	1	0	0	1	0
					0	1	−1	1	0	−1
						−22	0	0	0	0
							0	0	0	0
								−4	1	1
									0	0

shifts $=$ eigenvalues of bottom right 4×4 submatrix.

Idea of aggressive early deflation

`A = hess(randn(11));` After 4 iterations $\ldots$

0	1	0	0	1	0	1	0	0	1	1
1	0	0	−2	1	0	0	0	1	1	0
	0	1	1	1	1	1	0	0	1	0
		1	1	0	1	0	1	1	0	1
			−1	1	1	1	1	0	1	1
				1	0	0	1	0	−1	0
					0	1	−1	1	0	0
						−10	0	0	0	1
							0	0	−1	0
								−7	1	1
									−3	1

It turns out that some eigenvalues have already converged.

Idea of aggressive early deflation

Compute Schur form of bottom 4×4 submatrix ...

0	1	0	0	1	0	1	0	0	1	1
1	0	0	−2	1	0	0	1	−1	1	0
	0	1	1	1	1	1	0	−1	1	0
		1	1	0	1	0	1	1	0	1
			−1	1	1	1	1	1	1	1
				1	0	0	0	1	−1	0
					0	1	1	0	0	0
						−11	0	0	0	0
						−11	1	0	0	1
						−19			1	1
						−21			−4	1

Two eigenvalues can be deflated!

Idea of aggressive early deflation

Or: Compute Schur form of bottom 9×9 submatrix ...

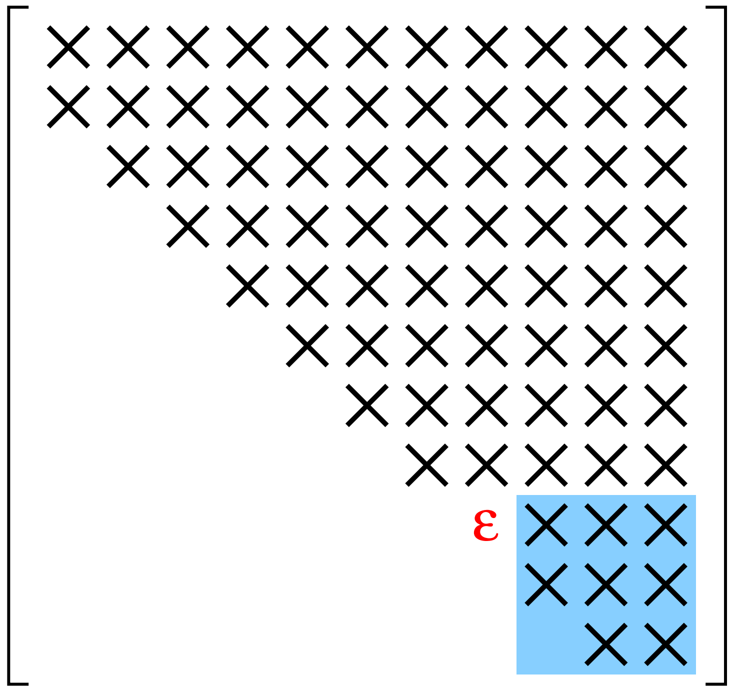
0	1	0	−1	1	1	0	0	0	1	1
1	0	0	0	1	0	0	1	−1	1	0
	0	1	1	1	0	1	1	0	1	−2
	0	1	1	1	1	1	1	0	1	1
	−2			1	1	1	1	1	1	1
	−3				1	1	0	0	0	0
	−4				0	1	0	0	0	0
	−14						0	0	0	0
	−14						1	0	0	1
	−22								1	1
	−24								−4	1

Four eigenvalues can be deflated!

A closer look at the Hessenberg matrix

After i iterations $\ldots$

$$A\hat{Q}_i = \hat{Q}_i$$



Observed: If shifts are eigenvalues of bottom right $m \times m$ submatrix:

$\varepsilon \longrightarrow 0$ locally quadratically.

A closer look at the Hessenberg matrix

After i iterations ...

$$A^{\mathsf{T}}\hat{Q}_i = \hat{Q}_i \begin{bmatrix} \times & \times & & & & & & & & & \\ \times & \times & \times & & & & & & & & \\ \times & \times & \times & \times & & & & & & & \\ \times & \times & \times & \times & \times & & & & & & \\ \times & \times & \times & \times & \times & \times & & & & & \\ \times & \times & \times & \times & \times & \times & \times & & & & \\ \times & \times & \times & \times & \times & \times & \times & \times & & & \\ \times & \times & \times & \times & \times & \times & \times & \times & \varepsilon & & \\ \times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \\ \times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \end{bmatrix}$$

For notational convenience: transpose

A closer look at the Hessenberg matrix

After i iterations ...

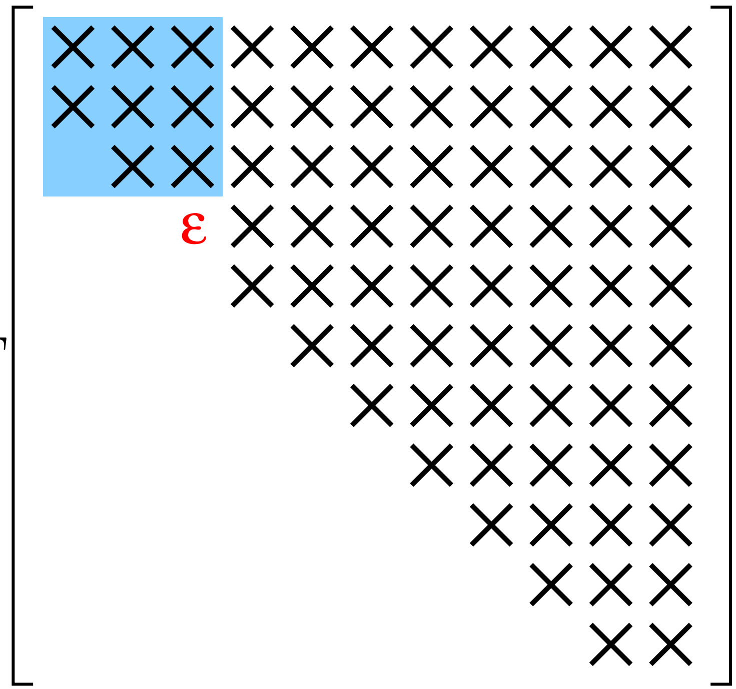
$$
A^{\mathsf{T}}\hat{Q}_i F = \hat{Q}_i F \left[\begin{array}{ccccccccccc}
\times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
 & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
 & & \varepsilon & \times & \times & \times & \times & \times & \times & \times & \times \\
 & & & \times & \times & \times & \times & \times & \times & \times & \times \\
 & & & & \times & \times & \times & \times & \times & \times & \times \\
 & & & & & \times & \times & \times & \times & \times & \times \\
 & & & & & & \times & \times & \times & \times & \times \\
 & & & & & & & \times & \times & \times & \times \\
 & & & & & & & & \times & \times & \times \\
 & & & & & & & & & \times & \times
\end{array}\right]
$$

For notational convenience: transpose and flip rows+columns.

A closer look at the Hessenberg matrix

After i iterations ...

$$A^{\mathsf{T}} \hat{Q}_i F = \hat{Q}_i F$$



For notational convenience: transpose and flip rows+columns.

Set

$$\hat{Q}_i F = \left[u_1^{(i)}, u_2^{(i)}, \ldots, u_n^{(i)} \right].$$

Connection to Krylov subspaces

Then

$$A^{\mathsf{T}} \left[u_1^{(i)}, \ldots, u_m^{(i)} \right] = \left[u_1^{(i)}, \ldots, u_m^{(i)} \right] \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ & \times & \times \end{bmatrix} + u_{m+1}^{(i)} \begin{bmatrix} 0 & 0 & \varepsilon \end{bmatrix}$$

Connection to Krylov subspaces

Then

$$A^{\mathsf{T}} \left[u_1^{(i)}, \ldots, u_m^{(i)} \right] = \left[u_1^{(i)}, \ldots, u_m^{(i)} \right] \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ & \times & \times \end{bmatrix} + u_{m+1}^{(i)} \left[\begin{array}{ccc} 0 & 0 & \varepsilon \end{array} \right]$$

This is an Arnoldi decomposition and reveals

$$\text{span}\left\{ u_1^{(i)}, \ldots, u_m^{(i)} \right\} = \text{span}\left\{ u_1^{(i)}, A^{\mathsf{T}} u_1^{(i)}, \ldots, (A^{\mathsf{T}})^{m-1} u_1^{(i)} \right\} := \mathcal{X}_i.$$

Moreover, $\varepsilon \to 0$ iff $\mathcal{X}_i$ converges to an invariant subspace $\mathcal{X}$ of A^{T}.

Formal convergence result [Watkins/Elsner'91]

Let cols of X form orthonormal basis for $\mathcal{X}$,
$X_\perp$ form orthonormal basis for $\mathcal{X}^\perp$.

Corresponding block Schur decomposition:

$$A\,[X_\perp, X] = [X_\perp, X] \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}.$$

If QR algorithm converges then

$$d\big(\mathcal{X}, \mathcal{X}_i\big) \leq C\, \|p_i(A_{11})^{-1}\|\, \|p_i(A_{22})\| d\big(\mathcal{X}, \mathcal{X}_{i-1}\big),$$

where p_i is the shift polynomial used in iteration i.

Formal convergence result [Watkins/Elsner'91]

Let cols of X form orthonormal basis for $\mathcal{X}$,
$X_\perp$ form orthonormal basis for $\mathcal{X}^\perp$.

Corresponding block Schur decomposition:

$$A\,[X_\perp, X] = [X_\perp, X] \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}.$$

If QR algorithm converges then

$$d\big(\mathcal{X}, \mathcal{X}_i\big) \le C\,\|p_i(A_{11})^{-1}\|\,\|p_i(A_{22})\| d\big(\mathcal{X}, \mathcal{X}_{i-1}\big),$$

where p_i is the shift polynomial used in iteration i.

Chosen shifts converge to eigenvalues of A_{22}

$$\rightsquigarrow \; \|p_i(A_{22})\| \le \tilde{C}\, d\big(\mathcal{X}, \mathcal{X}_{i-1}\big) \; \rightsquigarrow \; d\big(\mathcal{X}, \mathcal{X}_i\big) \le \hat{C}\, d\big(\mathcal{X}, \mathcal{X}_{i-1}\big)^2.$$

Intermediate summary

- QR generates sequence of m-dimensional Krylov subspaces $\mathcal{X}_i$ converging locally quadratically to invariant subspace $\mathcal{X}$ of A^T.
- Convergence bound determined by maximal distance between *all* shifts and the eigenvalues beloning to $\mathcal{X}$.

Intermediate summary

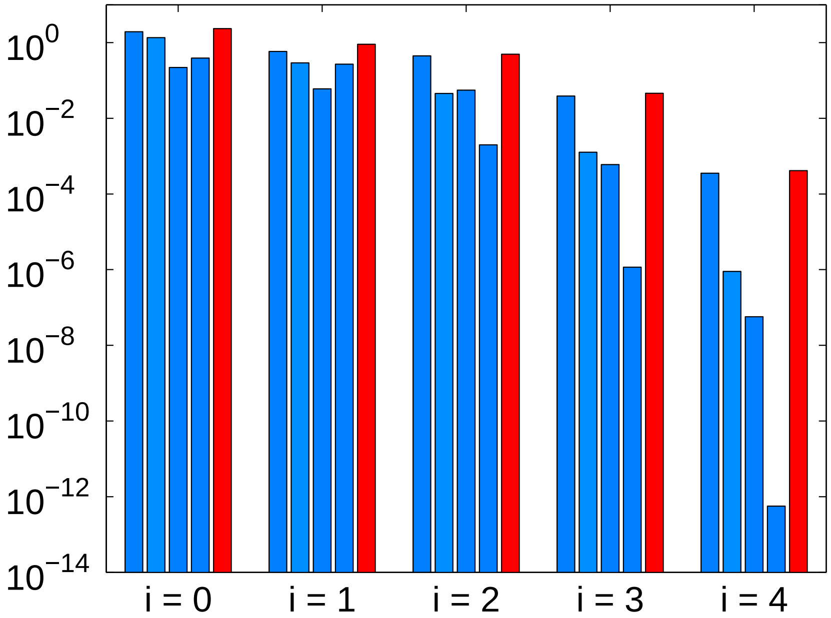
- QR generates sequence of m-dimensional Krylov subspaces $\mathcal{X}_i$ converging locally quadratically to invariant subspace $\mathcal{X}$ of A^T.
- Convergence bound determined by maximal distance between *all* shifts and the eigenvalues beloning to $\mathcal{X}$.

Questions:

- Does $\mathcal{X}_i$ contain much better approximations to individual eigenvectors?
- Does it make sense to consider Krylov subspaces larger than m?

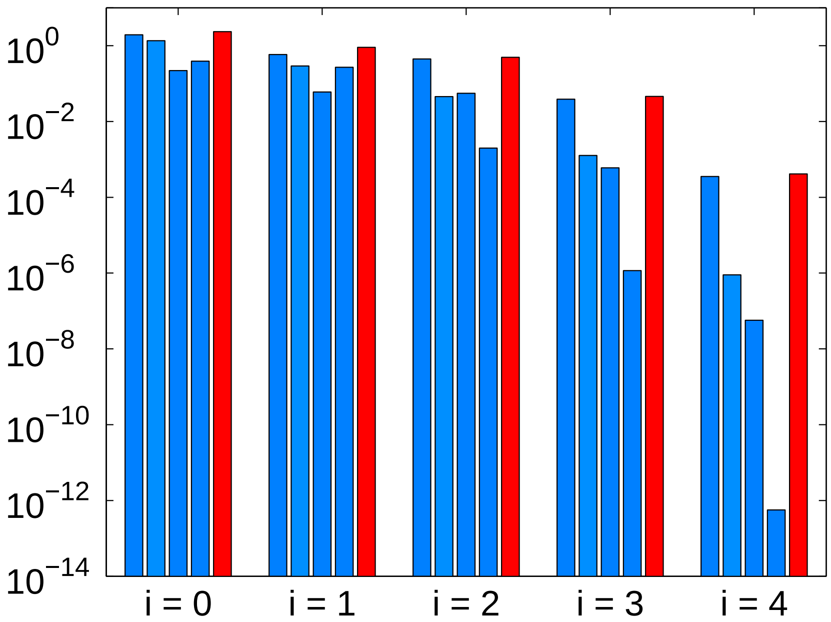
Example ($n = 250, m = 4$)

Convergence of individual Ritz vectors $x_1, \ldots, x_4$ from $\mathcal{X}_i$ to eigenvectors vs. ε:



Example ($n = 250, m = 4$)

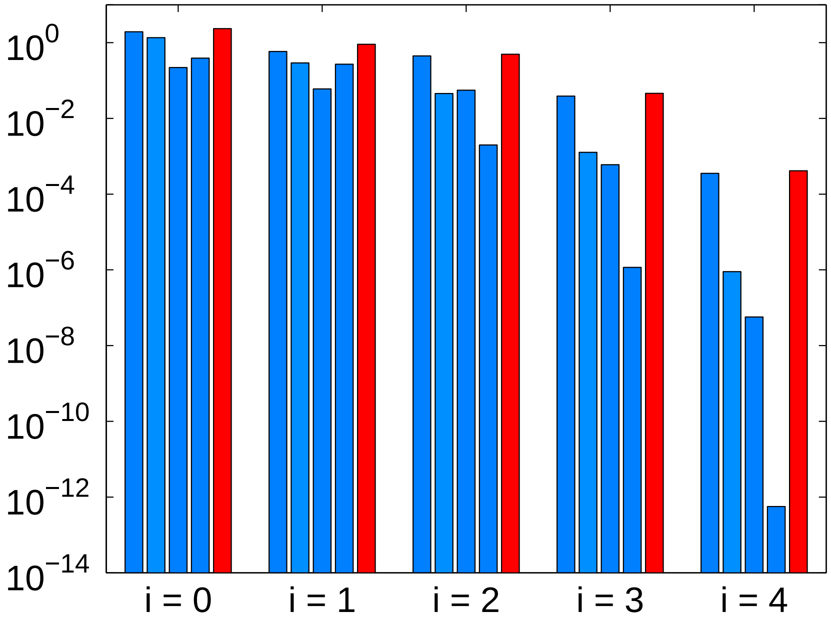
Convergence of individual Ritz vectors $x_1, \dots, x_4$ from $\mathcal{X}_i$ to eigenvectors vs. ε:



- Some Ritz vectors converge much faster than others.
- ε determined by the poorest Ritz vector approx.

Example ($n = 250, m = 4$)

Convergence of individual Ritz vectors $x_1, \ldots, x_4$ from $\mathcal{X}_i$ to eigenvectors vs. ε:



- Classical deflation (LAPACK 3.0) based on ε; cannot benefit from faster convergence of individual Ritz vectors.

Convergence result for individual eigenvectors

Let $v_1, \dots, v_m$ be left eigenvectors of A belonging to eigenvalues $\lambda_1, \dots, \lambda_m$ contained in A_{22}.

[K.'06]: For each v_j,

$$d\big(v_j, \mathcal{X}_i\big) \le C \, \|p_i(A_{11})^{-1}\| \, |p_i(\lambda_j)| \, d\big(v_j, \mathcal{X}_{i-1}\big).$$

- $|p_i(\lambda_j)|$ can be much smaller than $\|p_i(A_{22})\|$;
- only one shift needs to converge in order to obtain rapid convergence.

The Krylov-Schur algorithm [Stewart'01]

$$A^{\mathsf{T}} \left[u_1^{(i)}, \ldots, u_m^{(i)} \right] = \left[u_1^{(i)}, \ldots, u_m^{(i)} \right] \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ & \times & \times \end{bmatrix} + u_{m+1}^{(i)} \left[\begin{array}{ccc} 0 & 0 & \varepsilon \end{array} \right]$$

Compute Schur decomposition of Hessenberg factor:

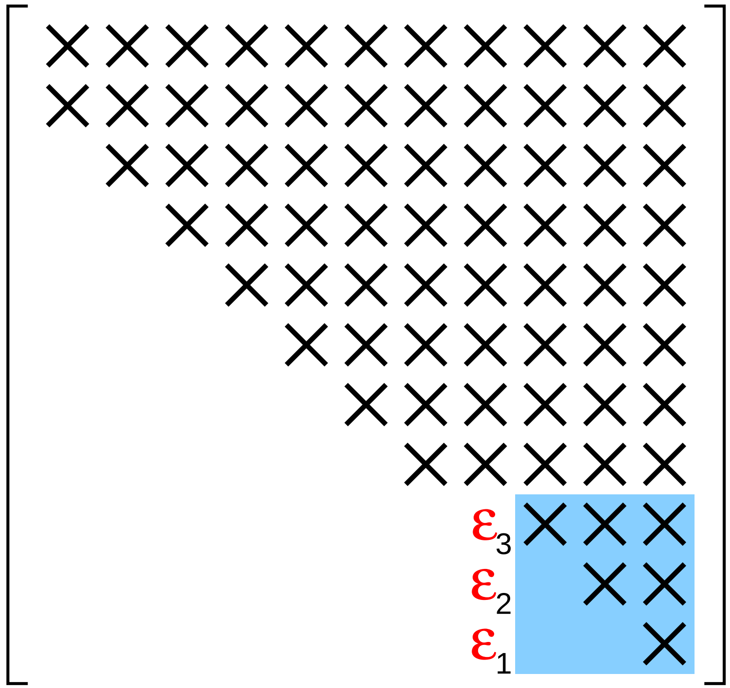
$$A^{\mathsf{T}} \left[\tilde{u}_1^{(i)}, \ldots, \tilde{u}_m^{(i)} \right] = \left[\tilde{u}_1^{(i)}, \ldots, \tilde{u}_m^{(i)} \right] \begin{bmatrix} \times & \times & \times \\ & \times & \times \\ & & \times \end{bmatrix} + u_{m+1}^{(i)} \left[\begin{array}{ccc} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 \end{array} \right]$$

- If $|\varepsilon_1| \le \mathbf{u} \|A\|_F$, converged Ritz vector $\tilde{u}_1^{(i)}$ is deflated by setting $\varepsilon_1 = 0$.
- If $|\varepsilon_1| > \mathbf{u} \|A\|_F$, test other Ritz vectors by reordering Schur form.

The Krylov-Schur algorithm on A

$$
\begin{bmatrix}
\times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
 & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
 & & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
 & & & \times & \times & \times & \times & \times & \times & \times & \times \\
 & & & & \times & \times & \times & \times & \times & \times & \times \\
 & & & & & \times & \times & \times & \times & \times & \times \\
 & & & & & & \times & \times & \times & \times & \times \\
 & & & & & & & \varepsilon & \times & \times & \times \\
 & & & & & & & & \times & \times & \times \\
 & & & & & & & & & \times & \times
\end{bmatrix}
$$

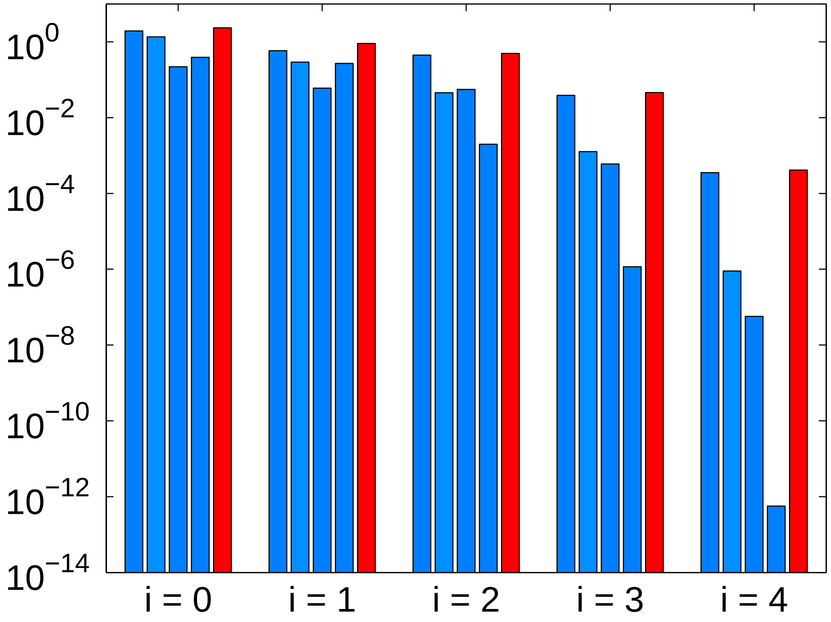
The Krylov-Schur algorithm on A



Re-interpreting Krylov-Schur algorithm for $\mathcal{X}_i$ in terms of orthogonal operations on A reveals theoretical and numerical equivalence to aggressive early deflation [Braman/Byers/Mathias'02] as implemented in LAPACK 3.1.

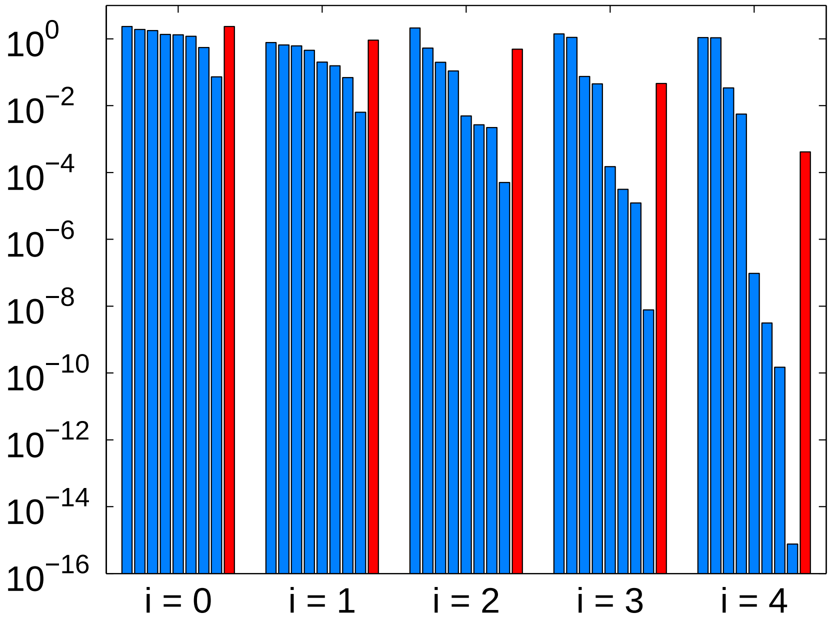
Example ($n = 250, m = 4$)

Convergence of individual Ritz vectors $x_1, \ldots, x_4$ from $\text{span}\{u_1^{(i)}, \ldots, u_4^{(i)}\}$ to eigenvectors vs. ε:



Example ($n = 250, m = 4$)

Convergence of individual Ritz vectors $x_1, \ldots, x_8$ from Krylov subspace $\mathrm{span}\{u_1^{(i)}, \ldots, u_8^{(i)}\}$ to eigenvectors vs. ε:



Larger $\mathcal{X}_i$ add Krylov subspace acceleration.

Convergence with Krylov subspace acceleration

Let $v_1, \dots, v_m$ be left eigenvectors of A belonging to eigenvalues $\lambda_1, \dots, \lambda_m$ contained in A_{22}. Let $w \geq m$ be dim of Krylov subspace

$$\mathcal{X}_i = \operatorname{span}\{u_1^{(i)}, \dots, u_w^{(i)}\}.$$

Then [K'06]:

$$d\big(v_j, \mathcal{X}_i\big) \leq C\, \|p_i(A_{11})^{-1}\|\, |p_i(\lambda_j)| \inf_{\phi \in \mathcal{P}_{w-m}} \frac{\|\phi(A_{11})\|}{|\phi(\lambda_j)|} d\big(v_j, \mathcal{X}_{i-1}\big).$$

Convergence with Krylov subspace acceleration

Let $v_1, \dots, v_m$ be left eigenvectors of A belonging to eigenvalues $\lambda_1, \dots, \lambda_m$ contained in A_{22}. Let $w \geq m$ be dim of Krylov subspace

$$\mathcal{X}_i = \mathrm{span}\{u_1^{(i)}, \dots, u_w^{(i)}\}.$$

Then [K'06]:

$$d\big(v_j, \mathcal{X}_i\big) \leq C \, \|p_i(A_{11})^{-1}\| \, |p_i(\lambda_j)| \inf_{\phi \in \mathcal{P}_{w-m}} \frac{\|\phi(A_{11})\|}{|\phi(\lambda_j)|} d\big(v_j, \mathcal{X}_{i-1}\big).$$

- Extra term $\inf_{\phi \in \mathcal{P}_{w-m}} \|\phi(A_{11})\| / |\phi(\lambda_j)|$ can be expected to become very small as $w \gg m$.
- Need extra assumptions on eigenvalue distribution of A_{11} to obtain quantitative results using, e.g., elementary potential theory [Beattie/Embree/Rossi'04].

Aggressive early deflation in practice

Choice of m and w crictical for obtaining good performance.

Current setting in LAPACK's QR and QZ based on experiments with random matrices, e.g.,

$$500 \leq n \leq 3000: \quad m = 64, \quad w = 96.$$

Usually not optimal for practical examples.

Aggressive early deflation in practice

Choice of m and w crictical for obtaining good performance.

Current setting in LAPACK's QR and QZ based on experiments with random matrices, e.g.,

$$500 \leq n \leq 3000: \quad m = 64, \quad w = 96.$$

Usually not optimal for practical examples.

Ongoing work

- develop heuristics for dynamical choice of w based on convergence bounds;
- replace ScaLAPACK's QR by two-level recursive aggressive early deflation approach.

Conclusions

- Aggressive early deflation $=$ extraction of Ritz vectors from Krylov subspaces associated with the QR algorithm.
- Can be turned into improved convergence bounds and intuitive explanations why aggressive early deflation works so well.

References:

C. A. Beattie, M. Embree, and J. Rossi. Convergence of restarted Krylov subspaces to invariant subspaces. *SIAM J. Matrix Anal. Appl.*, 25(4):1074–1109, 2004.

K. Braman, R. Byers, and R. Mathias. The multishift QR algorithm. II. Aggressive early deflation. *SIAM J. Matrix Anal. Appl.*, 23(4):948–973, 2002.

B. Kågström and D. Kressner. Multishift variants of the QZ algorithm with aggressive early deflation. SIAM J. Matrix Anal. Appl., 29(1):199-227, 2006.

D. Kressner. The effect of aggressive early deflation on the convergence of the QR algorithm. Uminf report, Department of Computing Science, Umeå University, 2006.