On rational Krylov sequences

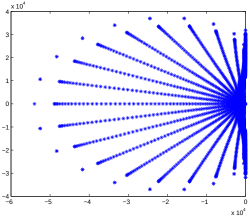
Karl Meerbergen

K.U. Leuven

Rolling waves – December 15–16, 2008

Outline

Eigenvalue problems



- $Ax = \lambda x$
- $Ax = \lambda Bx$
- $Ax + \lambda Bx + \lambda^2 Cx = 0$

Eigenvalue problems

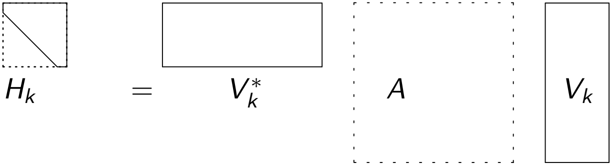
Large scale problems arise in

- Chemistry
- Analysis of vibrations (acoustics, structures, fluids)
- Stability analysis of nonlinear dynamical systems
- Markov chains
- Graph problems (e.g. Google matrix)

Small scale methods not feasible (QR), and only a small number of eigenvalues wanted

Krylov sequences

- Krylov sequence: $\{v_1, Av_1, A^2v_1, \dots, A^{k-1}v_1\}$
- Krylov space: $\mathrm{span}\{v_1, Av_1, A^2v_1, \dots, A^{k-1}v_1\}$
- Is a space of polynomials in A times v_1
- Orthonormal basis: $V_k = [v_1, \dots, v_k]$
- Hessenberg matrix $H_k = V_k^* A V_k$



- Computed by the Arnoldi method
- For Hermitian matrices: Lanczos method

Ritz values and Ritz vectors

- Hessenberg matrix $H_k = V_k^* A V_k$
- Relation:

$$AV_k - V_k H_k = F_k$$

  where F_k has rank one.
- Eigenvalues $(\hat{\lambda})$ and eigenvectors $(\hat{x} = V_k z)$ are computed so that

$$r = A\hat{x} - \hat{\lambda}\hat{x} \perp \mathrm{Range}(V_k)$$

  so,

$$H_k z = \hat{\lambda} z$$

- $\hat{\lambda}$ is called Ritz value
- $\hat{x}$ is called Ritz vector

Loss of orthogonality

- The Lanczos method uses a three term recurrence relation
- The Arnoldi method uses (modified) Gram-Schmidt
- In exact arithmetic, Krylov bases are orthogonal
- In finite precision arithmetic, bases may be far from orthogonal
- For linear system solvers (CG, GMRES, SYMMLQ, FOM), this does not pose major difficulties
- For eigenvalue problems, this is a more serious problem
  $\Rightarrow$ reorthogonalization

Rational Krylov sequences

- Rational Krylov sequence:
  - $\{v_1, (A-\mu_1 I)^{-1}v_1, (A-\mu_1 I)^{-1}(A-\mu_2 I)^{-1}v_1, \ldots\}$
  - $\{v_1, (A-\sigma_1 I)^{-1}v_1, (A-\sigma_2 I)^{-1}v_1, \ldots\}$
- Is a space of rational polynomials in A times b
- Orthonormal basis: $V_k = [v_1, \ldots, v_k]$
  (space is assumed of dimension k)
- Relation:

$$AV_kH_k - V_kL_k = F_k$$

  with H_k and L_k Hessenberg matrices
- Ritz pairs computed from

$$H_kz = \lambda L_kz$$

Rational Krylov sequences for generalized problem

- Rational Krylov sequence:
  - $\{v_1, (A - \mu_1 B)^{-1} B v_1, (A - \mu_1 B)^{-1} B (A - \mu_2 B)^{-1} B v_1, \ldots\}$
  - $\{v_1, (A - \sigma_1 B)^{-1} B v_1, (A - \sigma_2 B)^{-1} B v_1, \ldots\}$
- Orthonormal basis: $V_k = [v_1, \ldots, v_k]$
- Relation:

$$AV_k H_k - BV_k L_k = F_k$$

The Bultheel sequence

- Gorik De Samblanx (and M.)
  - Rational approximations to the exponential
  - (Rational) Krylov with implicit restarts
  - (Rational) Krylov with inexact matrix inversion
- M.
  - Implicit restarts for symmetric matrices
- Karl Deckers
  - Rational Lanczos sequences and orthogonal rational functions

Difficulties with RKS

Numerical analysts are always making trouble:

- Growing subspace: requires more and more memory
- Rational functions: requires inversion of matrices, i.e. solution of linear systems
- Choice of pole (potential theory, see Beckermann this morning, also Beattie & Embree)
- Rounding errors and choice of poles.

Explicit restarting

- Storage cost of vectors methods is a limitation of Krylov methods $\rightarrow$ restarting needed
- Explicit restart:
  - Choose a new v_1 starting vector, e.g. a Ritz vector from the current Krylov space
  - Choose a new pole μ_1, e.g. near a Ritz value (but not too close)
- Explicit restarting throws away the whole Krylov space:
  - Directions thrown away are recomputed = waste of effort

Selective reorthogonalization

- Eigenvalues do not converge with the same speed.
- Converged eigenvectors taken outside the Krylov space by explicit restarting return due to rounding errors.
- Therefore, the Krylov vectors must be orthogonalized against converged eigenvectors.
- This is called selective reorthogonalization.

Implicit restarting (De Samblanx & M. & B.)

- New idea (Sorensen 1992)
- Old space: k vectors V_k with

$$AV_k H_k - BV_k L_k = F_k$$

- New space: p vectors $W_p = V_k Q$ with $Q^* Q = I$
- New Hessenberg matrices $T_p = Q^* H_k Q$ and $K_p = Q^* L_k Q$

so that

$$AW_p T_p - BW_p K_p = G_p$$

Implicit restarting

Theorem

- *Compute* $Q \in \mathbf{C}^{k\times p}$ *from* $k-p$ *QR (QZ) steps with shifts* $\nu_1, \ldots, \nu_{k-p}$:

$$Q_j R_j = H - \nu_j L$$

- $Q = Q_1 \cdots Q_{k-p}$
- $\mathrm{Range}(W_p) = \psi(B^{-1}A)\mathrm{Range}(V_p)$ *with*

$$\psi(\xi) = \Pi_{j=1}^{k-p} \frac{\xi - \nu_j}{\xi - \mu_{p+j}}$$

- $w_1 = \psi(B^{-1}A)v_1 / \|\psi(B^{-1}A)v_1\|$
- *The poles of the restarted RKS are*

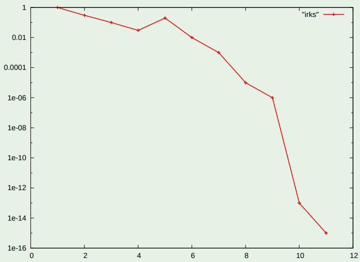
$$\mu_{p+1}, \ldots, \mu_k$$

Implicit restarting

Exact shifts:

- Shifts are 'unwanted' eigenvalues of (H_k, L_k)
- The p eigenpairs of (T_p, K_p) are eigenpairs of (H_k, L_k)

Example (1D Brusselator equations ($n = 968$))



- Residual norm right-most Ritz value
- No difference with full RKS (without restart)
- corresponds to [Morgan 1996]

Implicit restarting

For the extreme eigenvalues of $Ax = \lambda Bx$ with singular B, we may have to take special care to avoid the computation of the infinite eigenvalue

Use filter polynomial with $\nu_1 = \infty$

Poles (Rayleigh quotient)

Iteration	without Implicit restart	with implicit restart
3	8.432	-8.4677
6	19.751	-9.4883
9	74.83	-9.4883

- Infinite eigenvalue may be introduced due to rounding errors
- Implicit restart remove infinite eigenvalue
- Numerical stability can be proven (under certain conditions) [M. & Spence 1997]

Inexact rational Krylov [Lehoucq & M. 1998]

- On each iteration of RKS:
- Solve linear system $(A - \mu_j B)w_j = Bv_j$
- When an iterative solver is used, we usually have a large residual s_j: $(A - \mu_j B)w_j - Bv_j = s_j$
- The RKS relation then becomes:

$$AV_k H_k - BV_k L_k = F_k + S_k$$

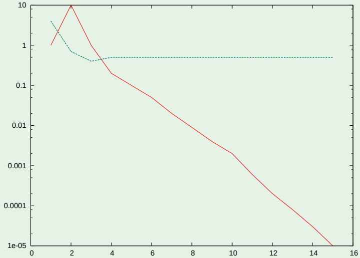
  where $S_k e_j$ is the residual of the linear solver at iteration j.
- The solution may be corrupted ...

Inexact rational Krylov

- ... unless we use the Cayley transform:
- Solve linear system $(A - \mu_j B)w_j = (A - \hat{\lambda}_j B)\hat{x}_j$ instead of $(A - \mu_j B)w_j = Bv_j$

Example (Olmstead model)

- Use 20 iterations of Gauss-Seidel iteration



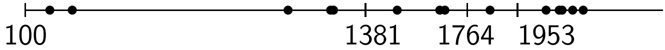
Choice of pole and rounding errors [M. 2000]

Example from poro-elastic material: plate of $0.4m \times 0.4m \times 0.06m$.

Algorithm:

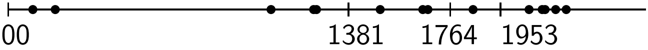
- Perform $k = 50$ iterations with the same pole
- Compute eigenvalues and eigenvectors
- Remove unwanted eigenvalues and reduce the space to $p = 25$
- Change the pole and expand the Krylov space from dimension p to k

restart	converged	pole
1	5	1381.
2	12	1764.
3	17	1953.
4	25	



Choice of pole and rounding errors (cont.)

- The poles are chosen in between eigenvalues
- A pole close to an eigenvalue amplifies rounding errors so that only one eigenvalue can be accurately computed



Experiment:

- Select the pole equal to a Ritz value (as in Rayleigh quotient iteration): $\mu_2 = 1264.112$.
- Before the change of pole, we have an error on the recurrence relation of the order of 10^{-11}.
- Our theory expects to keep only one accurate digit in the RKS relation.
- Experiments show we keep no accuracy at all.

Where is RKS?

- RKS is not used in computations for eigenvalue problems
- RKS is used for modelreduction
- Reasons?
  - It came too late for the symmetric eigenvalue problem
  - Codes in structural analysis were developed in the early nineties based on Lanczos method
  - I have tried to develop a code but . . .

Conclusions

- Rational Krylov is an extension of Krylov with great potential
- Implicit restarts can be used
- Pole selection is partially understood
- Iterative solver can be used (but I do not recommend it)

Industrial example parameterized linear system with NASTRAN

- Traditional computation
  - For each frequency, perform factorization of $K - \omega^2 M$ and solve
- Lanczos computation
  - One matrix factorization of $K - \sigma M$ and solve
  - k solves.