# Krylov Subspace Methods:
# Characteristic Properties Inherited in Finite Precision.

Jens-Peter M. Zemke

Dagstuhl Seminar 03421:

**Theoretical and Computational Aspects of Matrix Algorithms**

12.10.–17.10.2003

# Outline.

Krylov Subspace Methods 3

Krylov Eigenproblem Solvers 5

Krylov (Q)OR Methods 7

Method of Proof (sketched) 9

Krylov Method Properties 10–16

- o Eigenproblem Solvers 10
- o (Q)OR Methods 13

Conclusion & Outview 17

## Krylov Subspace Methods.

o Lanczos based methods (short–term methods)
o Arnoldi based methods (long–term methods)

o eigensolvers: $Av = v\lambda$
o linear system solvers: $Ax = b$
    o (quasi-) orthogonal residual approaches: (Q)OR
    o (quasi-) minimal residual approaches: (Q)MR

Extensions:
o Lanczos based methods:
    o look-ahead
    o product-type (LTPMs)
    o applied to normal equations (CGN)
o Arnoldi based methods:
    o restart (thin/thick, explicit/implicit)
    o truncation (standard/optimal)

Every (basic) Krylov method can be written as

$$AQ_k = Q_{k+1}\underline{C}_k - F_k,$$

where $\underline{C}_k \in \mathbb{C}^{k+1\times k}$ is (unreduced upper) Hessenberg.

This formulation is of interest in (Q)MR methods like GMRES.
$\Rightarrow$ brings in singular values, pseudo inverses and (total) least squares.

We prefer the slightly re-written version

$$AQ_k - Q_k C_k = M_k - F_k,$$

with the rank-one update $M_k = q_{k+1} c_{k+1,k} e_k^T$.

This formulation has the advantage that Hessenberg $C_k$ is *square*.
$\Rightarrow$ we can continue with eigendecompositions and inverses.

Disadvantage: applies only to eigensolvers and (Q)OR methods.

## Krylov Eigenproblem Solvers.

Assumption: $A$, $C_k$ diagonalizable (makes life easier).

Eigendecompositions of $A$ and $C_k$:

$$AV = V\Lambda, \qquad C_k S_k = S_k \Theta_k.$$

Left Eigenmatrices for $A$:

$$\begin{aligned} \widehat{V} &\equiv V^{-H} &\Rightarrow\quad \widehat{V}^H A &= \Lambda \widehat{V}^H, \\ \check{V} &\equiv V^{-T} &\Rightarrow\quad \check{V}^T A &= \Lambda \check{V}^T. \end{aligned}$$

Left Eigenmatrices for $C_k$:

$$\begin{aligned} \widehat{S}_k &\equiv S_k^{-H} &\Rightarrow\quad \widehat{S}_k^H C_k &= \Theta_k \widehat{S}_k^H, \\ \check{S}_k &\equiv S_k^{-T} &\Rightarrow\quad \check{S}_k^T C_k &= \Theta_k \check{S}_k^T. \end{aligned}$$

## Relations between small system and large system (AEP).

Computable (right) Ritz pair $(\theta_j, y_j) \equiv (\theta_j, Q_k s_j)$:

$$(A - \theta_j I) y_j = q_{k+1} c_{k+1,k} s_{kj} - F_k s_j.$$

Incomputable (left) pair $(\lambda_i, \tilde{s}_i^H) \equiv (\lambda_i, \hat{v}_i^H Q_k)$:

$$\tilde{s}_i^H (\lambda_i I - C_k) = \hat{v}_i^H q_{k+1} c_{k+1,k} e_k^T - \hat{v}_i^H F_k.$$

When $M_k$ and $F_k$ small, small (relative) backward errors

$$\eta(\theta_j, y_j) = \frac{\| q_{k+1} c_{k+1,k} s_{kj} - F_k s_j \|}{\|A\| \|Q_k s_j\|},$$

$$\eta(\lambda_i, \tilde{s}_i^H) = \frac{\| \hat{v}_i^H q_{k+1} c_{k+1,k} e_k^T - \hat{v}_i^H F_k \|}{\|C_k\| \|\hat{v}_i^H Q_k\|}.$$

No hope for both to be small: usually one small, other large:

$$\hat{v}_i^H q_{k+1} c_{k+1,k} s_{kj} = \left(\lambda_i - \theta_j\right) \hat{v}_i^H y_j + \hat{v}_i^H F_k s_j.$$

## Krylov (Q)OR Methods.

Assumption: $A$, $C_k$ invertible (makes life easier).

Solution of linear system $(A, b)$ given starting approximation $x_0$:

$$Ax = r_0, \qquad r_0 = b - Ax_0.$$

Special Krylov subspace: starting vector

$$q_1 = \frac{r_0}{\|r_0\|}.$$

Define (Q)OR approximation by

$$x_k \equiv Q_k z_k, \qquad \frac{z_k}{\|r_0\|} \equiv C_k^{-1} e_1.$$

Residual of (Q)OR approximation given by

$$r_k = r_0 - Ax_k = -q_{k+1} c_{k+1,k} z_{kk} + F_k z_k.$$

(Relative) backward error of (Q)OR approximation given by

$$\eta(x_k) = \frac{\|q_{k+1} c_{k+1,k} z_{kk} - F_k z_k\|}{\|A\| \|Q_k z_k\| + \|r_0\|}.$$

Therefore: aim at $M_k$ small (and $F_k$ too).

Understanding **inexact Krylov methods**: $Q_k$ no problem. Observe

$$\frac{z_k}{\|r_0\|} = C_k^{-1} e_1 \quad \Rightarrow \quad \frac{z_{lk}}{\|r_0\|} = e_l^T C_k^{-1} e_1.$$

Everything fine as long as

$$\|f_l\| \approx O\left(\frac{\epsilon}{z_{lk}}\right) \approx O\left(\frac{\|r_0\|\epsilon}{e_l^T C_k^{-1} e_1}\right).$$

## Method of Proof (sketched).

Starting point is diagonalized form of Hessenberg decomposition:

$$\hat{v}_i^H q_{k+1} = \frac{\left(\lambda_i - \theta_j\right) \hat{v}_i^H y_j + \hat{v}_i^H F_k s_j}{c_{k+1,k} s_{kj}} \quad \forall\ i, j(, k).$$

Toolkit (in order of appearance):

- o tricky summation along $j$
- o Hessenberg eigenvalue-eigenvector relations (HEER)
- o Lagrange polynomial interpolation
- o glueing it all together

Results: Explicit expressions reflecting influences of

- o starting vector
- o error vectors

## Krylov Method Properties (Eigenproblem Solvers).

Starting point re-written:

$$\left[\frac{c_{k+1,k}s_{kj}}{\lambda_i - \theta_j}\right]\hat{v}_i^H q_{k+1} = \hat{v}_i^H Q_k s_j + \left[\frac{1}{\lambda_i - \theta_j}\right]\hat{v}_i^H F_k s_j.$$

Observe:

$$e_1 = I_k e_1 = S_k S_k^{-1} e_1 = S_k \check{S}_k^T e_1 = \sum_{j=1}^{k} \check{s}_{1j} s_j.$$

Dependence of $k+1$st "basis" vector on starting vector and error vectors:

$$\left[\sum_{j=1}^{k} \frac{c_{k+1,k}\check{s}_{1j}s_{kj}}{\lambda_i - \theta_j}\right]\hat{v}_i^H q_{k+1} = \hat{v}_i^H q_1 + \hat{v}_i^H F_k \left[\sum_{j=1}^{k}\left(\frac{\check{s}_{1j}}{\lambda_i - \theta_j}\right) s_j\right].$$

We obtain a relation in terms of Ritz values **and** Ritz vector components.

Plug-in HEER:

$$\left[\sum_{j=1}^{k} \frac{\prod_{\ell=1}^{k} c_{\ell+1,\ell}}{\chi'_{C_k}(\theta_j)\left(\lambda_i - \theta_j\right)}\right] \hat{v}_i^H q_{k+1} = \hat{v}_i^H q_1 + \sum_{l=1}^{k} \left[\sum_{j=1}^{k} \frac{\left(\prod_{\ell=1}^{l-1} c_{\ell+1,\ell}\right) \chi_{C_{l+1:k}}(\theta_j)}{\chi'_{C_k}(\theta_j)(\lambda_i - \theta_j)}\right] \hat{v}_i^H f_l.$$

We obtain a relation solely in terms of Ritz values.

Use Lagrange interpolation:

$$\left[\frac{\prod_{\ell=1}^{k} c_{\ell+1,\ell}}{\chi_{C_k}(\lambda_i)}\right] \hat{v}_i^H q_{k+1} = \hat{v}_i^H q_1 + \sum_{l=1}^{k} \left[\frac{\left(\prod_{\ell=1}^{l-1} c_{\ell+1,\ell}\right) \chi_{C_{l+1:k}}(\lambda_i)}{\chi_{C_k}(\lambda_i)}\right] \hat{v}_i^H f_l.$$

We obtain a relation in terms of characteristic (Ritz) polynomials.

Division by first factor results in explicit expression for "basis" vectors:

$$\hat{v}_i^H q_{k+1} = \left( \frac{\chi_{C_k}(\lambda_i)}{\prod_{\ell=1}^{k} c_{\ell+1,\ell}} \right) \hat{v}_i^H q_1 + \sum_{l=1}^{k} \left( \frac{\chi_{C_{l+1:k}}(\lambda_i)}{\prod_{\ell=l+1}^{k} c_{\ell+1,\ell}} \left( \frac{\hat{v}_i^H f_l}{c_{l+1,l}} \right) \right). \qquad (1)$$

Multiplication by $v_i$ and summation yields:

$$q_{k+1} = \left( \frac{\chi_{C_k}(A)}{\prod_{\ell=1}^{k} c_{\ell+1,\ell}} \right) q_1 + \sum_{l=1}^{k} \left( \frac{\chi_{C_{l+1:k}}(A)}{\prod_{\ell=l+1}^{k} c_{\ell+1,\ell}} \left( \frac{f_l}{c_{l+1,l}} \right) \right). \qquad (2)$$

**Theorem:** (blue: infinite precision, blue & red: finite precision)

The "basis" vectors constructed by a (finite precision) Krylov method fulfill eqns. (1) and (2).

## Krylov Method Properties ((Q)OR Methods).

Starting point re-written:

$$\left[\frac{c_{k+1,k}s_{kj}}{\lambda_i - \theta_j}\right] \hat{v}_i^H q_{k+1} = \hat{v}_i^H Q_k s_j + \left[\frac{1}{\lambda_i - \theta_j}\right] \hat{v}_i^H F_k s_j.$$

Observe:

$$\frac{z_k}{\|r_0\|} = C_k^{-1} e_1 = S_k \Theta_k^{-1} \breve{S}_k^T = \sum_{j=1}^{k} \frac{\breve{s}_{1j}}{\theta_j} s_j.$$

Dependence of $k$th "best" approximation on "basis" and error vectors:

$$\frac{\hat{v}_i^H x_k}{\|r_0\|} = \left[\sum_{j=1}^{k} \frac{c_{k+1,k}\breve{s}_{1j}s_{kj}}{(\lambda_i - \theta_j)\,\theta_j}\right] \hat{v}_i^H q_{k+1} - \hat{v}_i^H F_k \left[\sum_{j=1}^{k} \left(\frac{\breve{s}_{1j}}{(\lambda_i - \theta_j)\,\theta_j}\right) s_j\right].$$

We obtain a relation in terms of Ritz values **and** Ritz vector components.

Plug-in HEER:

$$\frac{\hat{v}_i^H x_k}{\|r_0\|} = \left[ \sum_{j=1}^{k} \frac{\left(\prod_{\ell=1}^{k} c_{\ell+1,\ell}\right)}{\chi'_{C_k}(\theta_j)(\lambda_i - \theta_j)\theta_j} \right] \hat{v}_i^H q_{k+1}$$

$$- \left[ \sum_{l=1}^{k} \sum_{j=1}^{k} \left( \frac{\left(\prod_{\ell=1}^{l-1} c_{\ell+1,\ell}\right) \chi_{C_{l+1:k}}(\theta_j)}{\chi'_{C_k}(\theta_j)(\lambda_i - \theta_j)\theta_j} \right) \hat{v}_i^H f_l \right].$$

We obtain a relation solely in terms of Ritz values.

Use Lagrange interpolation:

$$\frac{\hat{v}_i^H x_k}{\|r_0\|} = \mathcal{L}_k[x^{-1}](\lambda_i) \left( \frac{\prod_{\ell=1}^{k} c_{\ell+1,\ell}}{\chi_{C_k}(\lambda_i)} \right) \hat{v}_i^H q_{k+1}$$

$$- \left[ \sum_{l=1}^{k} \left[ \mathcal{L}_k[x^{-1} \chi_{C_{l+1:k}}(x)](\lambda_i) \right] \left( \frac{\prod_{\ell=1}^{l-1} c_{\ell+1,\ell}}{\chi_{C_k}(\lambda_i)} \right) \hat{v}_i^H f_l \right].$$

We obtain a relation in terms of interpolation polynomials.

We insert the explicit expression for the "basis" vectors and re-formulate:

$$\frac{\hat{v}_i^H x_k}{\|r_0\|} = \mathcal{L}_k[x^{-1}](\lambda_i)\hat{v}_i^H q_1 + \left[\sum_{l=1}^{k} \left( \frac{\left(\prod_{\ell=1}^{l-1} c_{\ell+1,\ell}\right) \omega_l(\lambda_i)}{\chi_{C_k}(0)} \right) \hat{v}_i^H f_l \right],$$

where polynomials $\omega_l$ are defined by

$$\omega_l(x) \equiv \sum_{s=1}^{k-l} \frac{\chi_{C_{l+1:k}}^{(s)}(x)}{s!} x^{s-1}.$$

(Explicit proof omitted.)

**Theorem.** In matrix form (with careful & appropriate interpretation):

$$x_k = \mathcal{L}_k[x^{-1}](A)r_0 + \|r_0\| \left[\sum_{l=1}^{k} \left( \frac{\left(\prod_{\ell=1}^{l-1} c_{\ell+1,\ell}\right) \omega_l(A)}{\chi_{C_k}(0)} \right) f_l \right].$$

Proceeding this manner, we obtain an explicit expression for residuals:

$$r_k = \frac{\chi_{C_k}(A)}{\chi_{C_k}(0)} r_0 + \|r_0\| \sum_{l=1}^{k} \left[ \frac{\left(\prod_{\ell=1}^{l-1} c_{\ell+1,\ell}\right) \left[\chi_{C_{l+1:k}}(A) - \chi_{C_{l+1:k}}(0)\right]}{\chi_{C_k}(0)} f_l \right].$$

This implies the following explicit expression for error vectors:

$$(x - x_k) = \frac{\chi_{C_k}(A)}{\chi_{C_k}(0)} (x - x_0) + \|r_0\| \sum_{l=1}^{k} \left[ \frac{\left(\prod_{\ell=1}^{l-1} c_{\ell+1,\ell}\right) \varepsilon_l(A)}{\chi_{C_k}(0)} f_l \right],$$

where *polynomials* $\varepsilon_l$ are defined by

$$\varepsilon_l(x) = \frac{\chi_{C_{l+1:k}}(x) - \chi_{C_{l+1:k}}(0)}{x}.$$

(Explicit proof omitted.)

## Conclusion & Outview.

Pros:

o We have constructed a variety of formulae for finite precision Krylov methods that have the same “look & feel” as the “corresponding” (more or less well known) formulae for infinite precision Krylov methods.

Cons:

o The formulae need genuine interpretation to be useful, and (seem to) neglect second order error effects (or replace them by another point of view, i.e., Krylov methods as Lagrange interpolation).

o The (by far) more interesting case of (Q)**M**R Krylov subspace methods is not included by now. Room for improvements. Any suggestions?