Krylov Subspace Methods for $\text{sign}(Q)b$

(where Q is hermitian)

Andreas Frommer

Bergische Universität Wuppertal

Fachbereich Mathematik und Naturwissenschaften

`frommer@math.uni-wuppertal.de`

http://www.math.uni-wuppertal.de/SciComp

joint work with

Henk van der Vorst

Jasper van den Eshof

Katrin Schäfer

Thomas Lippert

Outline

1. matrix methods for $\operatorname{sign}(Q)$
2. Schulz as a Krylov subspace method
3. projection on $K(Q)$ and Lanczos
4. projection on $K(Q^2)$ and Lanczos
5. partial fraction expansions and multishift CG

Definition

$$Q = VJV^{-1} \quad \text{Jordan canonical form}$$

$$J = \mathrm{diag}(J_\ell), \quad J_\ell = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix}$$

Assume $\mathrm{Re}(\lambda) \neq 0$ for all $\lambda \in \mathrm{spec}(Q)$. Then

$$\mathrm{sign}(Q) = V\mathrm{sign}(J)V^{-1}, \quad \mathrm{sign}(J) = \mathrm{diag}(\mathrm{sign}(J_\ell)),$$

where $\mathrm{sign}(J_\ell) = \mathrm{sign}(\lambda) \cdot I$

Note: Q hermitian: $V^{-1} = V^H$, $\mathrm{spec}(Q) \subset \mathbb{R}$.

Matrix Methods

Newton's method

Roberts, 1970: Solve

$$F(X) = 0 \quad \text{where } F(X) = X^2 - I.$$

We have

$$DF(X)H = XH + HX$$

so that Newton's method

$$X_{k+1} = X_k - \Delta_k, \;\; DF(X_k)\Delta_k = F(X_k)$$

gives

$$\Delta_k = \frac{1}{2}\left(X_k - X_k^{-1}\right), \qquad X_{k+1} = \frac{1}{2}\left(X_k + X_k^{-1}\right).$$

Matrix Methods

Newton's method

Roberts, 1970: Solve

$$F(X) = 0 \quad \text{where } F(X) = X^2 - I.$$

We have

$$DF(X)H = XH + HX$$

so that Newton's method

$$X_{k+1} = X_k - \Delta_k, \;\; DF(X_k)\Delta_k = F(X_k)$$

gives

$$\Delta_k = \frac{1}{2}\left(X_k - X_k^{-1}\right), \qquad X_{k+1} = \frac{1}{2}\left(X_k + X_k^{-1}\right).$$

Theorem. Let $X_0 = Q$. Then $\lim_{k\to\infty} X_k = \mathrm{sign}(Q)$ for every Q with $\mathrm{spec}(Q) \cap i\mathbb{R} = \emptyset$. Convergence is quadratic.

Back
Close

Schulz' method

Solve

$$F(X) = 0 \quad \text{where } F(X) = X^{-2} - I.$$

We have

$$DF(X)H = X^{-2}HX^{-1} + X^{-1}HX^{-2}$$

so that Newton's method

$$X_{k+1} = X_k - \Delta_k, \;\; DF(X_k)\Delta_k = F(X_k)$$

gives

$$\Delta_k = \frac{1}{2}\left(X_k^3 - X_k\right), \quad X_{k+1} = \frac{1}{2}X_k\left(3I - X_k^2\right).$$

Schulz' method

Solve

$$F(X) = 0 \quad \text{where } F(X) = X^{-2} - I.$$

We have

$$DF(X)H = X^{-2}HX^{-1} + X^{-1}HX^{-2}$$

so that Newton's method

$$X_{k+1} = X_k - \Delta_k, \;\; DF(X_k)\Delta_k = F(X_k)$$

gives

$$\Delta_k = \frac{1}{2}\left(X_k^3 - X_k\right), \quad X_{k+1} = \frac{1}{2}X_k\left(3I - X_k^2\right).$$

Theorem. Let $X_0 = Q$. Then $\lim_{k\to\infty} X_k = \mathrm{sign}(Q)$ if $\|I - Q^2\| \leq 1$ and $\mathrm{spec}(Q) \cap i\mathbb{R} = \emptyset$. Convergence is quadratic.

Partial Fraction Expansions

Pandey, Kenney and Laub (1990):

$$
\begin{aligned}
S_p &= \left((Q+I)^{2p} - (Q-I)^{2p}\right) \cdot \left((Q+I)^{2p} + (Q-I)^{2p}\right)^{-1} \\
&= \frac{1}{p} Q \sum_{i=1}^{p} \frac{1}{\xi_i} \left(Q^2 + \alpha_i^2 I\right)^{-1},
\end{aligned}
$$

where $\xi_i = \frac{1}{2}\left(1 + \cos\frac{(2i-1)\pi}{2p}\right), \alpha_i^2 = \frac{1}{\xi_i} - 1.$

S_p is an approximation to $\mathrm{sign}(Q)$.
Formula may be iterated.

Zolotarev (long ago): Assume $\mathrm{spec}(Q) \subset [-b, -a] \cup [a, b]$. Then

$$
\begin{aligned}
Z_p &= \delta \cdot Q \prod_{i=1}^{p-1} (Q^2 + c_{2i} I) \cdot \prod_{i=1}^{p} (Q^2 + c_{2i-1} I)^{-1} \\
&= \delta \cdot Q \sum_{i=1}^{p} \omega_i (Q^2 + \alpha_i I)^{-1},
\end{aligned}
$$

where

$$
c_i = \frac{\mathrm{sn}^2\left(iK/(2m); \sqrt{1-(b/a)^2}\right)}{1 - \mathrm{sn}^2\left(iK/(2m); \sqrt{1-(b/a)^2}\right)},
$$

K is the complete elliptic integral.

Motivation for matrix-vector type methods

Neuberger Fermions in QCD. Solve

$$(I - \Gamma_5 \mathrm{sign}(Q))x = b$$

- Q is nearest neighbor coupling on 4-dimensional grid: (very) sparse
- Q is (very) hermitian indefinite
- 12 variables per grid point
- grid is 8^4 to 16^4
- size of system is (very) large: 50 000 to 800 000.

Inner-outer iteration: inner iteration computes $\mathrm{sign}(Q)v$.

2. Schulz as a Krylov Subspace Method

Schulz: $Q_{k+1} = \frac{1}{2}Q_k\left(3I - Q_k^2\right) = p_{3^{k+1}}(Q)$
Consequently

$$Q_{k+1}v = \frac{1}{2}Q_k\left(3v - Q_k^2 v\right) \in K_{3^{k+1}}(Q, v).$$

Related issue: How to draw a fork.

2. Schulz as a Krylov Subspace Method

Schulz: $Q_{k+1} = \frac{1}{2} Q_k \left(3I - Q_k^2\right) = p_{3^{k+1}}(Q)$
Consequently

$$Q_{k+1} v = \frac{1}{2} Q_k \left(3v - Q_k^2 v\right) \in K_{3^{k+1}}(Q, v).$$

Related issue: How to draw a fork.

the linear algebra person's strategy:

2. Schulz as a Krylov Subspace Method

Schulz: $Q_{k+1} = \frac{1}{2} Q_k \left(3I - Q_k^2\right) = p_{3^{k+1}}(Q)$
Consequently

$$Q_{k+1} v = \frac{1}{2} Q_k \left(3v - Q_k^2 v\right) \in K_{3^{k+1}}(Q, v).$$

Related issue: How to draw a fork.

the linear algebra person's strategy:

Back
Close

2. Schulz as a Krylov Subspace Method

Schulz: $Q_{k+1} = \frac{1}{2} Q_k \left(3I - Q_k^2\right) = p_{3^{k+1}}(Q)$
Consequently

$$Q_{k+1} v = \frac{1}{2} Q_k \left(3v - Q_k^2 v\right) \in K_{3^{k+1}}(Q, v).$$

Related issue: How to draw a fork.

the linear algebra person's strategy:

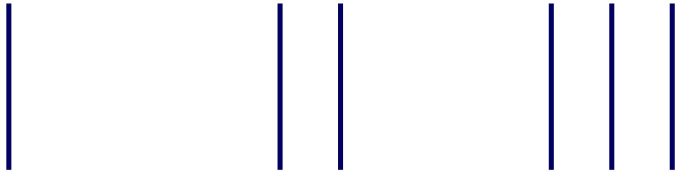
2. Schulz as a Krylov Subspace Method

Schulz: $Q_{k+1} = \frac{1}{2} Q_k \left(3I - Q_k^2\right) = p_{3^{k+1}}(Q)$
Consequently

$$Q_{k+1} v = \frac{1}{2} Q_k \left(3v - Q_k^2 v\right) \in K_{3^{k+1}}(Q, v).$$

Related issue: How to draw a fork.

the linear algebra person's strategy:



Back
Close

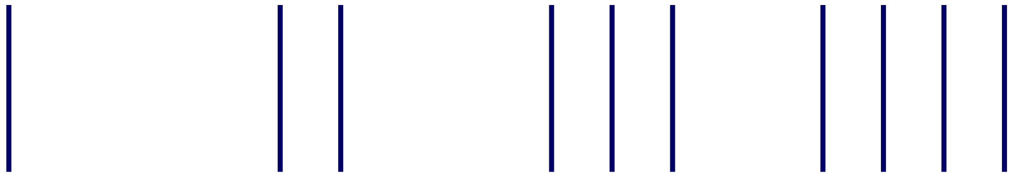
2. Schulz as a Krylov Subspace Method

Schulz: $Q_{k+1} = \frac{1}{2}Q_k\left(3I - Q_k^2\right) = p_{3^{k+1}}(Q)$
Consequently

$$Q_{k+1}v = \frac{1}{2}Q_k\left(3v - Q_k^2 v\right) \in K_{3^{k+1}}(Q, v).$$

Related issue: How to draw a fork.

the linear algebra person's strategy:



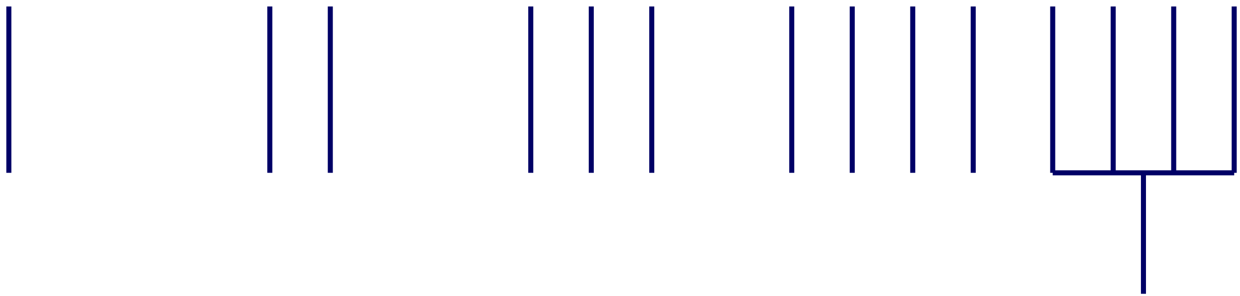
2. Schulz as a Krylov Subspace Method

Schulz: $Q_{k+1} = \frac{1}{2} Q_k \left(3I - Q_k^2\right) = p_{3^{k+1}}(Q)$
Consequently

$$Q_{k+1} v = \frac{1}{2} Q_k \left(3v - Q_k^2 v\right) \in K_{3^{k+1}}(Q, v).$$

Related issue: How to draw a fork.

the linear algebra person's strategy:

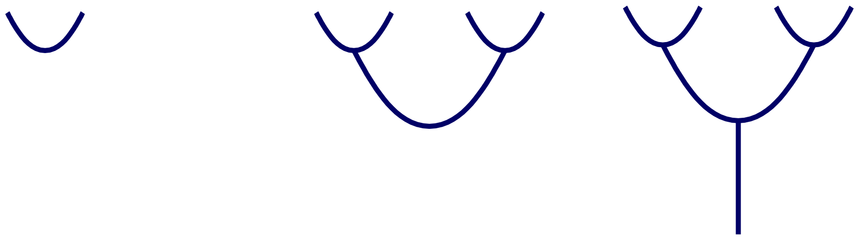


the computer scientist's strategy:

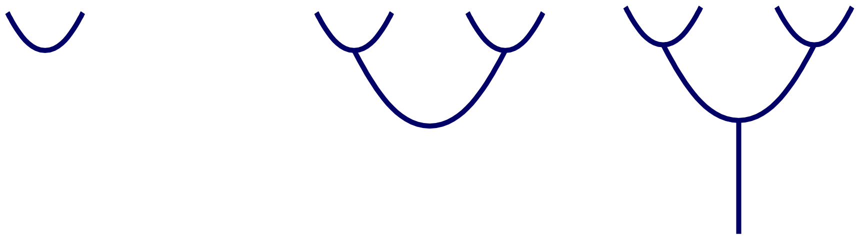
the computer scientist's strategy:

the computer scientist’s strategy:

the computer scientist's strategy:



the computer scientist's strategy:



Recursive computation:

$s = \text{schulz}(v, k) \quad \{$

if $k = 1$

$\qquad s = \frac{1}{2}Q\left(3v - Q^2 v\right)$

else

$\qquad s = \text{schulz}(v, k-1)$

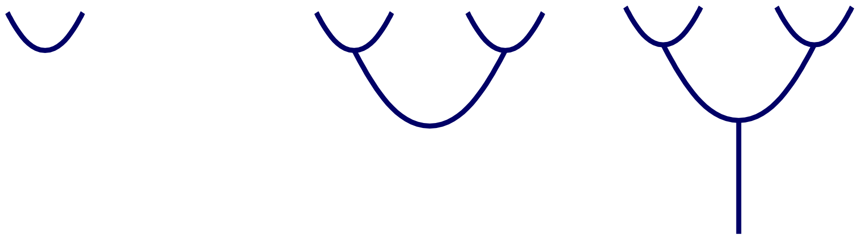
$\qquad s = \text{schulz}(s, k-1)$

$\qquad s = 3v - s$

$\qquad s = (1/2) \cdot \text{schulz}(s, k-1)$

$\}$

the computer scientist's strategy:



Recursive computation:

$s = \text{schulz}(v, k)$ {
if $k = 1$
$\quad s = \frac{1}{2}Q\left(3v - Q^2 v\right)$
else
$\quad s = \text{schulz}(v, k-1)$
$\quad s = \text{schulz}(s, k-1)$
$\quad s = 3v - s$
$\quad s = (1/2) \cdot \text{schulz}(s, k-1)$
}

Properties:

- $\text{schulz}(v, k) = Q_k v$
- $\text{schulz}(v, k) \in K_{3^k}(Q, v)$
  sublinear convergence
- time cost: 3^k MVMs
- storage cost: k vectors
- $k = 6$: $3^k = 729$

Back
Close

3. Projection on $K(Q)$ and Lanczos

From now on: Q is hermitian

The Lanczos process generates an orthonormal basis $v_1, v_2, \ldots, v_m$ for $K_m(Q, v)$

$v_1 = v/\|v\|_2$, $\beta_0 = 0$
for $i = 1, 2, \ldots, k$
  $\tilde{v} = Av_i - \beta_{i-1}v_{i-1}$
  $\alpha_i = v_i^H \tilde{v}$
  $\tilde{v} = \tilde{v} - \alpha_i v_i$
  $\beta_i = \|\tilde{v}\|_2$
  $v_{i+1} = \tilde{v}/\beta_i$

Notation $V_m = [v_1, v_2, \ldots, v_m]$, $T_m = \text{tridiag}(\beta_{m-1}, \alpha_m, \beta_m)$
$\Rightarrow V_m^H Q V_m = T_m$

Approximating **sign**$(Q)v$ from $K_m(Q,v)$

$$\mathrm{sign}(Q) = Q \cdot \left(Q^2\right)^{-1/2}$$

- **variant 1** [Borici 99]: projection on $QK_m(Q,v)$

$$x^m = QV^m \cdot \left(T_m^2\right)^{-1/2} \cdot V_m^H v = \quad (= p_{m+1}(Q)v)$$

Approximating **sign**$(Q)v$ from $K_m(Q,v)$

$$\mathrm{sign}(Q) = Q \cdot \left(Q^2\right)^{-1/2}$$

- **variant 1** [Borici 99]: projection on $QK_m(Q,v)$

$$x^m = QV^m \cdot \left(T_m^2\right)^{-1/2} \cdot V_m^H v = \quad (= p_{m+1}(Q)v)$$

- **variant 2** [van der Vorst 00]: projection on $K_m(Q,v)$

$$x^m = V^m \cdot \mathrm{sign}(T_m) \cdot V_m^H v \quad = V^m \cdot \mathrm{sign}(V_m^H Q V_m) \cdot V_m^H v$$

$(= p_m(Q)v$, p interpolates the Ritz values)

Approximating **sign**$(Q)v$ from $K_m(Q,v)$

$$\text{sign}(Q) = Q \cdot \left(Q^2\right)^{-1/2}$$

- **variant 1** [Borici 99]: projection on $QK_m(Q,v)$

$$x^m = QV^m \cdot \left(T_m^2\right)^{-1/2} \cdot V_m^H v = \quad (= p_{m+1}(Q)v)$$

- **variant 2** [van der Vorst 00]: projection on $K_m(Q,v)$

$$x^m = V^m \cdot \text{sign}(T_m) \cdot V_m^H v \quad = V^m \cdot \text{sign}(V_m^H Q V_m) \cdot V_m^H v$$

($= p_m(Q)v$, p interpolates the Ritz values)

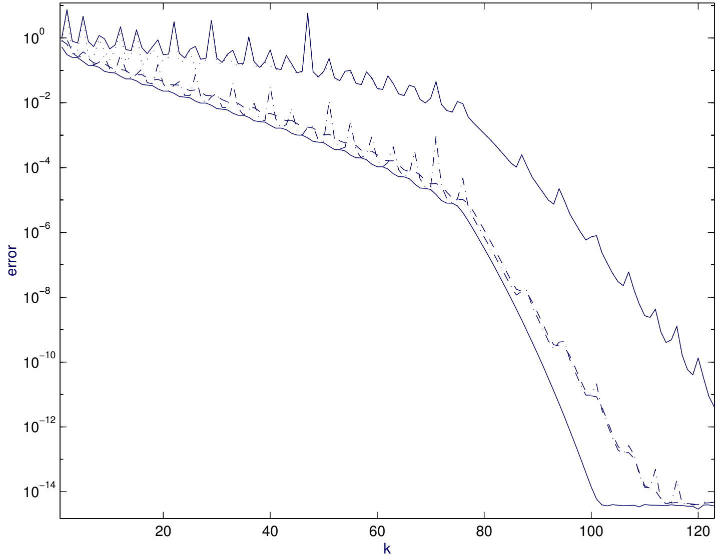
- **variant 3** [van den Eshof et al 02]: projection on $K_m(Q,v)$, harmonic Ritz values

$$x^m = QV^m \cdot \text{sign}(T_m + \beta_m^2 T_m^{-1} e_m) \cdot V_m^H v$$

($= p_m(Q)v$, p interpolates the harmonic Ritz values)

Test example

$$Q = \mathrm{diag}(-30, -29, \ldots, -10, 1, 2, \ldots, 100)$$



optimal
(solid lower)

projection on $QK_m(Q, v)$
(dotted)

projection on $K_m(Q, v)$
(dash-dot)

projection on $K_m(Q, v)$
harmonic Ritz (dash)

norm of the CG residual
(solid top)

Back
Close

4. Projection on $QK(Q^2)$ and Lanczos

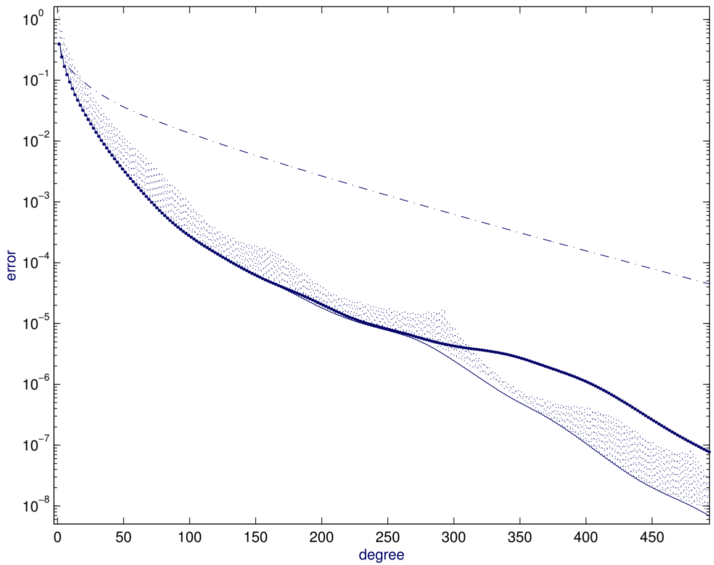
Borici 00: Lanczos for Q^2:

$$Q^2\widehat{V}_k = \widehat{V}_k\widehat{T}_k + \widehat{\beta}_{k+1}\widehat{v}_{k+1}e_k^H,$$

Take

$$x^m = Q\widehat{V}_m\widehat{T}_m^{-1/2}V_m^H v \ \ (= p_{2m+1}(Q)v,\ p_{2m+1} \text{ odd})$$

Q from QCD, 16^4 grid, $\kappa = 0.208$ and $\beta = 6$.



optimal (solid lower)

Chebyshev method (dash-dot)

projection on $K_m(Q, v)$ (dotted)

projection on $QK_m(Q^2, v)$ (solid upper)

Back
Close

Theory

Lemma [van den Eshof et al 02]: Let

$$v = v^+ + v^- \text{ where } \mathrm{sign} Q v^+ = v^+,\ \mathrm{sign} Q v^- = -v^-.$$

- r_m^+: GMRES residual for $Qx = v^+$
- r_m^- GMRES residual for $Qx = b^-$
- odd polynomial $p_{2m+1}(t) = t \cdot q_m(t^2)$
- approximation $x = p_{2m+1}(Q)v$ for $\mathrm{sign}(Q)v$.

Then

$$\|\mathrm{sign} Q v - x\|_2^2 \geq \|r_{2m+1}^+\|_2^2 + \|r_{2m+1}^-\|_2^2.$$

Note: Lower bound goes like $\sqrt{\kappa(Q)}$.

Projection on $QK(Q^2)$ via Lanczos for Q^2:

Theorem [van den Eshof et al. 02]:
r_m residual of CG for $Q^2x = v$, $x^0 = 0$.
Then

$$\|\mathrm{sign}(Q)v - Q\widehat{V}_m\widehat{T}_m^{-1/2}\widehat{V}_m v\|_2 \le \|r_m\|_2.$$

Note: Upper bound goes like $\kappa(Q)$.

Proof: Roberts' integral representation gives

$$\begin{aligned}
&\mathrm{sign}(Q)v - Q\widehat{V}_m\widehat{T}_m^{-1/2}\widehat{V}_m v \\
&\quad = \frac{2}{\pi}\int_0^\infty Q(t^2I+Q^2)^{-1}b - Q\widehat{V}_m(t^2I+\widehat{T}_m)^{-1}\widehat{V}_m^H v \,\mathrm{d}t \\
&\quad = \frac{2}{\pi}\int_0^\infty Q(t^2I+Q^2)^{-1}r_m^{t^2}\,\mathrm{d}t.
\end{aligned}$$

Here $r_m^{t^2} = v-(Q^2+t^2I)\widehat{V}_k(\widehat{T}_k+t^2I)^{-1}e_1$ is the CG residual for $(Q^2+t^2I)x = v$.

Proof: Roberts' integral representation gives

$$\begin{aligned}
&\mathrm{sign}(Q)v - Q\widehat{V}_m\widehat{T}_m^{-1/2}\widehat{V}_m v \\
&\quad = \frac{2}{\pi}\int_0^\infty Q(t^2I + Q^2)^{-1}b - Q\widehat{V}_m(t^2I + \widehat{T}_m)^{-1}\widehat{V}_m^H v \;\mathrm{d}t \\
&\quad = \frac{2}{\pi}\int_0^\infty Q(t^2I + Q^2)^{-1} r_m^{t^2} \;\mathrm{d}t.
\end{aligned}$$

Here $r_m^{t^2} = v - (Q^2 + t^2I)\widehat{V}_k(\widehat{T}_k + t^2I)^{-1}e_1$ is the CG residual for $(Q^2 + t^2I)x = v$.

Use $r_m^{t^2} = \phi_m^{t^2}\cdot r_m^0$, $|\phi_m^{t^2}| < 1$ to get

$$\mathrm{sign}(Q)v - Q\widehat{V}_m\widehat{T}_m^{-1/2}\widehat{V}_m v = X_m r_m^0$$

where $X = \dfrac{2}{\pi}\displaystyle\int_0^\infty Q(t^2I + Q^2)^{-1}\phi_m^{t^2}\;\mathrm{d}t.$

Back
Close

Proof: Roberts' integral representation gives

$$\begin{aligned}
&\mathrm{sign}(Q)v - Q\widehat{V}_m\widehat{T}_m^{-1/2}\widehat{V}_m v \\
&\quad = \frac{2}{\pi}\int_0^\infty Q(t^2I+Q^2)^{-1}b - Q\widehat{V}_m(t^2I+\widehat{T}_m)^{-1}\widehat{V}_m^H v \ \mathrm{d}t \\
&\quad = \frac{2}{\pi}\int_0^\infty Q(t^2I+Q^2)^{-1}r_m^{t^2} \ \mathrm{d}t.
\end{aligned}$$

Here $r_m^{t^2} = v - (Q^2+t^2I)\widehat{V}_k(\widehat{T}_k + t^2I)^{-1}e_1$ is the CG residual for $(Q^2+t^2I)x = v$.

Use $r_m^{t^2} = \phi_m^{t^2}\cdot r_m^0$, $|\phi_m^{t^2}| < 1$ to get

$$\mathrm{sign}(Q)v - Q\widehat{V}_m\widehat{T}_m^{-1/2}\widehat{V}_m v = X_m r_m^0$$

where $X = \dfrac{2}{\pi}\displaystyle\int_0^\infty Q(t^2I+Q^2)^{-1}\phi_m^{t^2} \ \mathrm{d}t.$

But $\mathrm{spec}(X) \subset [-1,1]$.

PFEs and Multishift CG

$$\mathrm{sign}(Q)v \approx \sum_{i=1}^{p} \omega_i Q \left(Q^2 + \tau_i I\right)^{-1} v.$$

$(\tau_i > 0)$.
Solve all p systems $\left(Q^2 + \tau_i I\right) x_i = v$ in one stroke ('multishift CG'), since

$$K_m(Q^2, b) = K_m(Q^2 + \tau_i I, b),\ i = 1, 2, \ldots, m.$$

Computational aspects:

1. **error** is controlled through CG residuals and approximation error of rational approximation
(approx. error needs a, b s.t. spec$(Q) \subset [-b, -a] \cup [a, b]$)

2. **implementation**: perform CG on seed system, update quantities for other systems

3. **stability:** use CGLS-like algorithm for seed (F., Maass 99)

4. update quantities for other systems using the (differential form of the stationary) **qd algorithm** (van den Eshof, Sleijpen 03)

5. **efficiency:** remove converged systems

Comparison of PFEs

- Pandey, Kenney, Laub (PKL)
- Zolotarev
- Edwards, Heller, Narayanan 99 (EHN): $t \cdot w(t^2)$, $w(t)$ best approximation from $R_{m,m}$ to $t^{-1/2}$ on $[a^2, b^2]$ (via the Remez algorithm)

No of poles for accuracy 10^{-6}

b/a	PKL	EHN	Zolotarev
200	19	7	5
1000	42	12	6

Numerical experiments

QCD, 16^4 lattice, 16 processors on ALiCE.

Conf.	1	2	3	4	5

Lanczos/PFE

MVs	2281	1969	1953	1853	1769
time/s	150	131	129	124	118

PFE/CG Zolotarev without removal

MVs	1141	985	977	927	885
time/s	154	125	125	116	102

PFE/CG Zolotarev with removal

MVs	1205	1033	1033	971	927
time/s	122	93	97	87	79

Note: Lanzcos methods need two sweeps (or store all Lanczos vectors)

Conclusions

- Lanczos based projection techniques are often close to optimal

Conclusions

- Lanczos based projection techniques are often close to optimal
- restriction to odd polynomials smoothes convergence curves, but may (in theory, sometimes) be a severe restriction

Conclusions

- Lanczos based projection techniques are often close to optimal
- restriction to odd polynomials smoothes convergence curves, but may (in theory, sometimes) be a severe restriction
- PFEs and projection on $QK_m(Q^2)$ yield error bounds

Conclusions

- Lanczos based projection techniques are often close to optimal
- restriction to odd polynomials smoothes convergence curves, but may (in theory, sometimes) be a severe restriction
- PFEs and projection on $QK_m(Q^2)$ yield error bounds
- schulz(v, k) is a nice idea

Conclusions

- Lanczos based projection techniques are often close to optimal
- restriction to odd polynomials smoothes convergence curves, but may (in theory, sometimes) be a severe restriction
- PFEs and projection on $QK_m(Q^2)$ yield error bounds
- schulz(v, k) is a nice idea
- inner-outer schemes are important in QCD
- Zolotarev is now standard in QCD
- QCD people include deflation techniques