Initial guesses for multi-shift solvers

James Osborn
Argonne National Lab

Lattice 2008
July 14
College of William and Mary

Initial guesses for multi-shift solvers

- multi-shift solvers
- method for initial guess
- multi-source multi-shift solvers
- results
- extensions

multiple shift systems

- solve system of linear equations

  $(A+\sigma_1)\, x_1 = b$

  $(A+\sigma_2)\, x_2 = b$

  ...

  (A is matrix and σ constant (times identity))

- occur frequently lattice QCD
  - multiple masses
  - rational function approximation

multi-shift solvers

- typical solved with a Krylov method (CG)
- share same Krylov space $\{b, Ab, A^2b, ...\}$
- can be solved simultaneously with same number of matrix-vector products as worst conditioned equation

[QMR: R. W. Freund (1993),
MR: U. Glässner, et al. (1996),
BiCG: A. Boriçi (1996),
CG: B. Jegerlehner (1996)]

initial guesses

- want to use prior knowledge to reduce number of iterations
  - restarting solver from approximate solutions
  - projecting approximate low eigenmodes
  - solving similar equations
    - small changes in b
    - small changes in A (chronological inverter [R. Brower, et al. (1995)])

initial guesses

- given initial guesses y_k, construct

  $r_1 = b - (A+\sigma_1)\, y_1$

  $r_2 = b - (A+\sigma_2)\, y_2$

  ...

  want to solve

  $(A+\sigma_1)\, z_1 = r_1$

  $(A+\sigma_2)\, z_2 = r_2$

  ...

- right hand sides (r_k) in general are not the same
- no longer share a Krylov space

initial guesses

- can choose (for 2 shifts)

  $y_1 = (A+\sigma_2)\, w$

  $y_2 = (A+\sigma_1)\, w$

- for some w, then

  $r_1 = r_2 = b - (A+\sigma_1)\,(A+\sigma_2)\, w$

- right hand sides are now the same

approximate solutions

- given approximate solutions

  $v_1 \approx (A+\sigma_1)^{-1}\, b\,, \quad R_1 = b - (A+\sigma_1)\, v_1$

  $v_2 \approx (A+\sigma_2)^{-1}\, b\,, \quad R_2 = b - (A+\sigma_2)\, v_2$

- then

  $w = (v_1 - v_2)/(\sigma_2 - \sigma_1) \approx [(A+\sigma_1)(A+\sigma_2)]^{-1}\, b$

  $r_1 = r_2 = [\,(A+\sigma_2)R_1 - (A+\sigma_1)R_2\,] \,/\, (\sigma_2 - \sigma_1)$

- if v_1, v_2 were exact solutions then restart residual would be zero

two-source two-shift solver

[M. Clark, thesis (2006)]

- to solve

  $(A+\sigma_1)\, x_1 = b_1$

  $(A+\sigma_2)\, x_2 = b_2$

- choose "guesses" y_k such that:

  $b_1 - (A+\sigma_1)\, y_1 = b_2 - (A+\sigma_2)\, y_2$

  $\rightarrow y_1 = y_2 = (b_2 - b_1)/(\sigma_2 - \sigma_1)$

- equivalent to previous slide for $b_k = R_k$

multi-source multi-shift solver

- extension to n equations:
  $(A+\sigma_k)\,x_k = b_k$,   for k=1..n
- choose y_k such that
  $b_k - (A+\sigma_k)\,y_k = r$,   for k=1..n
- set

$$y_k = \sum_{i=0}^{n-2} A^i \; z_{k,i}$$

- equate powers of A and solve for $z_{k,i}$
  in terms of b's

initial guesses: spacial case of multi-source multi-shift solver

- for $b_k = b - (A+\sigma_k)\, v_k$
- equivalent to setting

$$
\begin{aligned}
y_1 &= (A+\sigma_2)(A+\sigma_3) \ldots (A+\sigma_n)\, w \\
y_2 &= (A+\sigma_1)(A+\sigma_3) \ldots (A+\sigma_n)\, w \\
&\ldots \\
y_n &= (A+\sigma_1)(A+\sigma_2) \ldots (A+\sigma_{n-1})\, w
\end{aligned}
$$

- for some $w \in \mathrm{span}(v_k)$, with common residual

  $r = b - (A+\sigma_1)(A+\sigma_2)\ldots(A+\sigma_n)\, w$

results: approximate solutions

- random 32^4 lattice
- preconditioned asqtad operator
  $(m_k^2 - D_{eo} D_{oe})\, x_k = b$
- b: point source
- final $|r|^2 = 1e\text{-}6$
- starting from solutions with $|r|^2 < 1e\text{-}3$
  ($|r_k|^2 < 1e\text{-}6$ for $k>2$)
- $m_k = m_1 \times d^{(k-1)}$
- all work in double precision

results: approximate solutions

m1	d	no guess	n=1	n=2	n=3	n=4	n=5	n=6
0.010	2	683	334 (9.9e-04)	445 (3.6e+04)	488 (1.1e+11)	509 (2.0e+16)	518 (2.5e+20)	523 (2.1e+23)
0.010	√2	683	334 (9.9e-04)	489 (2.3e+05)	575 (1.4e+13)	635 (2.0e+20)	677 (6.7e+26)	985 (5.6e+32)
0.005	2	1365	666 (9.9e-04)	892 (6.0e+05)	978 (2.9e+13)	1018 (8.9e+19)	1039 (1.8e+25)	1214 (2.3e+29)
0.005	√2	1365	666 (9.9e-04)	977 (3.8e+06)	1151 (3.8e+15)	1272 (8.8e+23)	1761 (4.8e+31)	3066 (6.6e+38)
0.002	2	3417	1668 (1.0e-03)	2228 (2.3e+07)	2445 (4.3e+16)	2554 (5.2e+24)	3965 (4.1e+31)	6723 (2.0e+37)
0.002	√2	3417	1668 (1.0e-03)	2445 (1.5e+08)	2885 (5.8e+18)	3246 (5.1e+28)	7210 (1.1e+38)	10000 (5.8e+46)
0.001	2	6830	3353 (1.0e-03)	4464 (3.6e+08)	4903 (1.1e+19)	5328 (2.0e+28)	10000 (2.4e+36)	10000 (1.8e+43)
0.001	√2	6830	3353 (1.0e-03)	4886 (2.3e+09)	5764 (1.4e+21)	8949 (1.9e+32)	10000 (6.4e+42)	10000 (5.2e+52)

- number of iterations (starting $|r|^2$)
- even with large starting residual, needs fewer iterations than no guess
- breaks down for more/smaller shifts due to needed residual reduction being too large

choosing guesses

- general strategies for choosing guesses
  - globally optimize for w
    $r = b - (A+\sigma_1)(A+\sigma_2)\ldots(A+\sigma_n)\, w$
    - minimize norm of residual
    - project out from residual
  - individually optimize
    $r_k = b - (A+\sigma_k)\, w_k$
    then apply multi-source multi-shift algorithm

results: approximate eigenmodes

- random 16^4 lattice
- preconditioned asqtad operator
  $(m_k^2 - D_{eo} D_{oe})\, x_k = b$
- b: point source
- final $|r|^2 = 1e\text{-}6$
- starting from approximate low modes
  (smallest approximate eigenvalue > $4\ \lambda_{min}$)
- $m_k = m_1 \times d^{(k-1)}$
- all work in double precision

results: approximate eigenmodes

m1	d	no guess	n=1	n=2	n=3	n=4	n=5	n=6
0.010	2	665	626 (1.4e+00)	650 (4.9e+06)	677 (1.1e+13)	693 (2.0e+18)	705 (2.5e+22)	708 (2.1e+25)
0.010	√2	665	626 (1.4e+00)	657 (1.2e+07)	700 (3.4e+14)	740 (3.4e+21)	772 (9.8e+27)	1078 (7.5e+33)
0.005	2	1328	1136 (1.9e+00)	1185 (1.2e+08)	1241 (4.2e+15)	1280 (1.2e+22)	1303 (2.3e+27)	1605 (3.0e+31)
0.005	√2	1328	1136 (1.9e+00)	1200 (2.9e+08)	1288 (1.2e+17)	1369 (1.8e+25)	1946 (7.9e+32)	3263 (9.5e+39)
0.002	2	3309	2157 (2.6e+00)	2268 (3.1e+09)	2402 (3.2e+18)	2539 (3.3e+26)	4776 (2.4e+33)	7549 (1.2e+39)
0.002	√2	3309	2157 (2.6e+00)	2291 (5.3e+09)	2479 (3.9e+19)	3339 (1.6e+29)	7274 (2.5e+38)	10000 (1.1e+47)
0.001	2	6563	3627 (2.9e+00)	3800 (1.3e+10)	4028 (1.4e+20)	5969 (2.2e+29)	10000 (2.6e+37)	10000 (2.1e+44)
0.001	√2	6563	3627 (2.9e+00)	3805 (1.9e+10)	4099 (1.9e+21)	8618 (1.6e+32)	10000 (4.4e+42)	10000 (3.5e+52)

- number of iterations (starting $|r|^2$)
- small increase in iterations for more masses or smaller spacings
- still breaks down for more/smaller shifts due to needed residual reduction being too large

extensions

- residuals don't have to be same, only collinear
  $b_k - (A+\sigma_k)\, y_k = \alpha_k\, r$,   for k=1..n
  - may help when restarting from certain Krylov methods
- can also use higher powers of A in y_k
- may need to actually project higher modes from residual to get large reduction

conclusions

- algorithm for solving systems with multiple sources each with a different shift
- can be applied to provide initial guesses to multi-shift solvers
- initial residuals are large, but convergence is still faster
- breaks down at some point when going to more and/or smaller shifts
- need to find ways to reduce residual while preserving improvement in iterations