Linear Systems with Multiple Shifts and Multiple Right-hand Sides

Ron Morgan
Baylor University

Joint with Walter Wilcox, Dean Darnell, Abdou Abdel-Rehim
Baylor University

- Yesterday’s quote of the day:

  “He is a loving dog. He loves to chase cats around trees.”

  (from a 4-year old on the plane)

Salute to Yale

- Yale is the Alma Mater for both 2004 presidential candidates

- Yale endowment is 18 billion

(and more importantly:)

- Yale is the birthplace of GMRES (Saad and Schultz, 1986)

Outline

Evil eigenvalues

Exorcising the evil

Greedy physicists

How much is enough of
a good thing?

Living on the edge

Greedier physicists

Some people don’t like
to start over

Outline

Evil eigenvalues	-	Krylov methods and small eigenvalues
Exorcising the evil	-	Deflated GMRES - Remove small eigenvalues
Greedy physicists	-	Multiple right-hand sides
How much is enough of a good thing?	-	How many eigenvalues are needed for effectively deflating a Wilson matrix?
Living on the edge	-	How helpful is deflating at kappa critical?
Greedier physicists	-	Multiple shifts for each right-hand side.
Some people don't like to start over	-	Deflating non-restarted methods such as BiCGStab, for multiple rhs's.

Mathematicians are STUCK UP.

“We are better, because we can prove things.”

- Nature abhors a …

- Nature abhors a … vacuum.

- Polynomials abhor…

- Polynomials abhor …turning sharp corners.

(They don't like the stress.)

Quote from a polynomial:

"Let's just relax and chill."

- George Bush abhors …
- Bill Clinton abhors …
- American Idol producers abhor …

- George Bush abhors …polysyllabic words.
- Bill Clinton abhors …not being in the White House (no more interns).
- American Idol producers abhor …Sanjaya leaving (ratings may go down).

Krylov subspace:

$$Span\{r_0, Ar_0, A^2 r_0, \ldots, A^{m-1} r_0\}$$

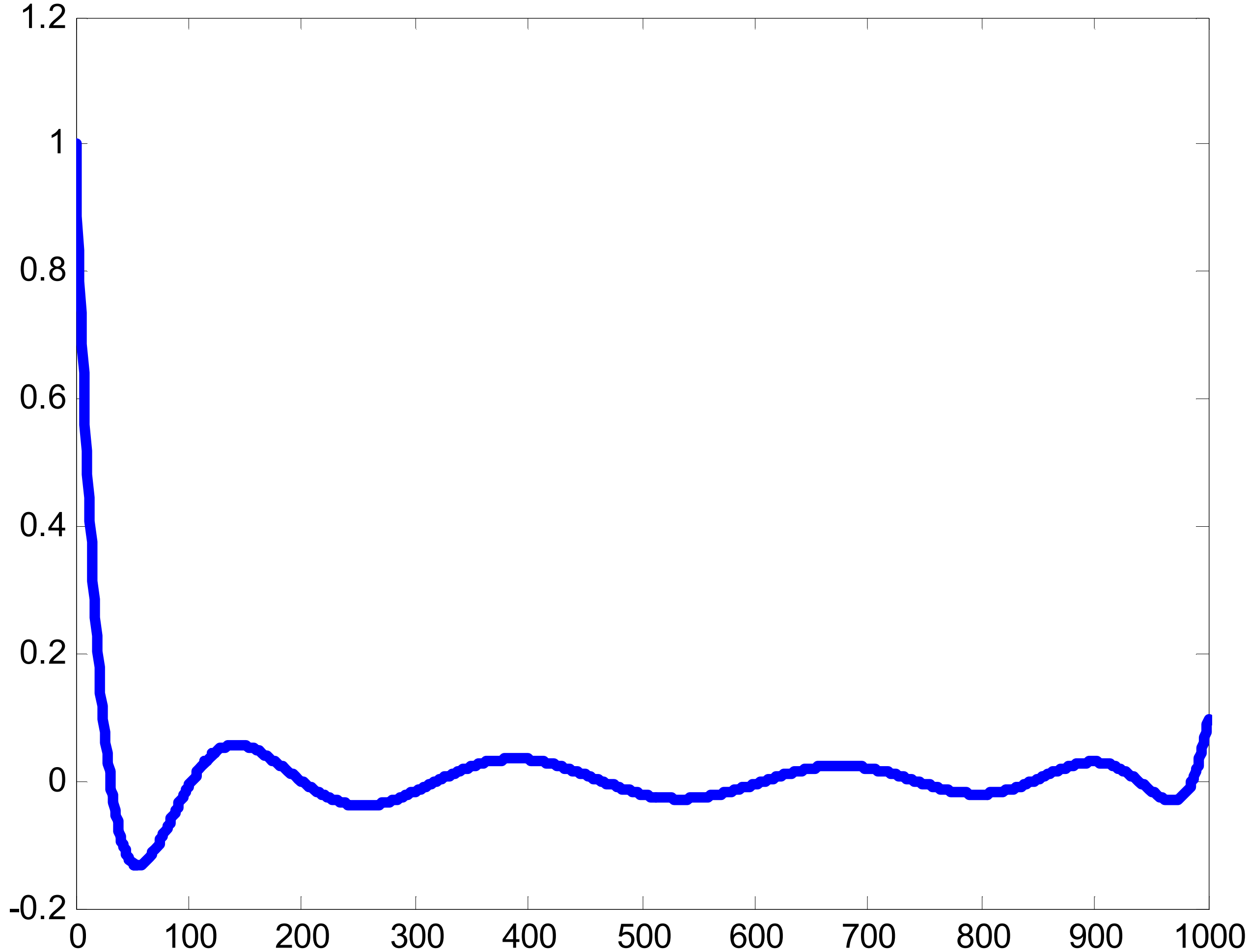
Residual vector:

$$r = q(A) r_0 = \Sigma \beta_i q(\lambda_i) z_i$$

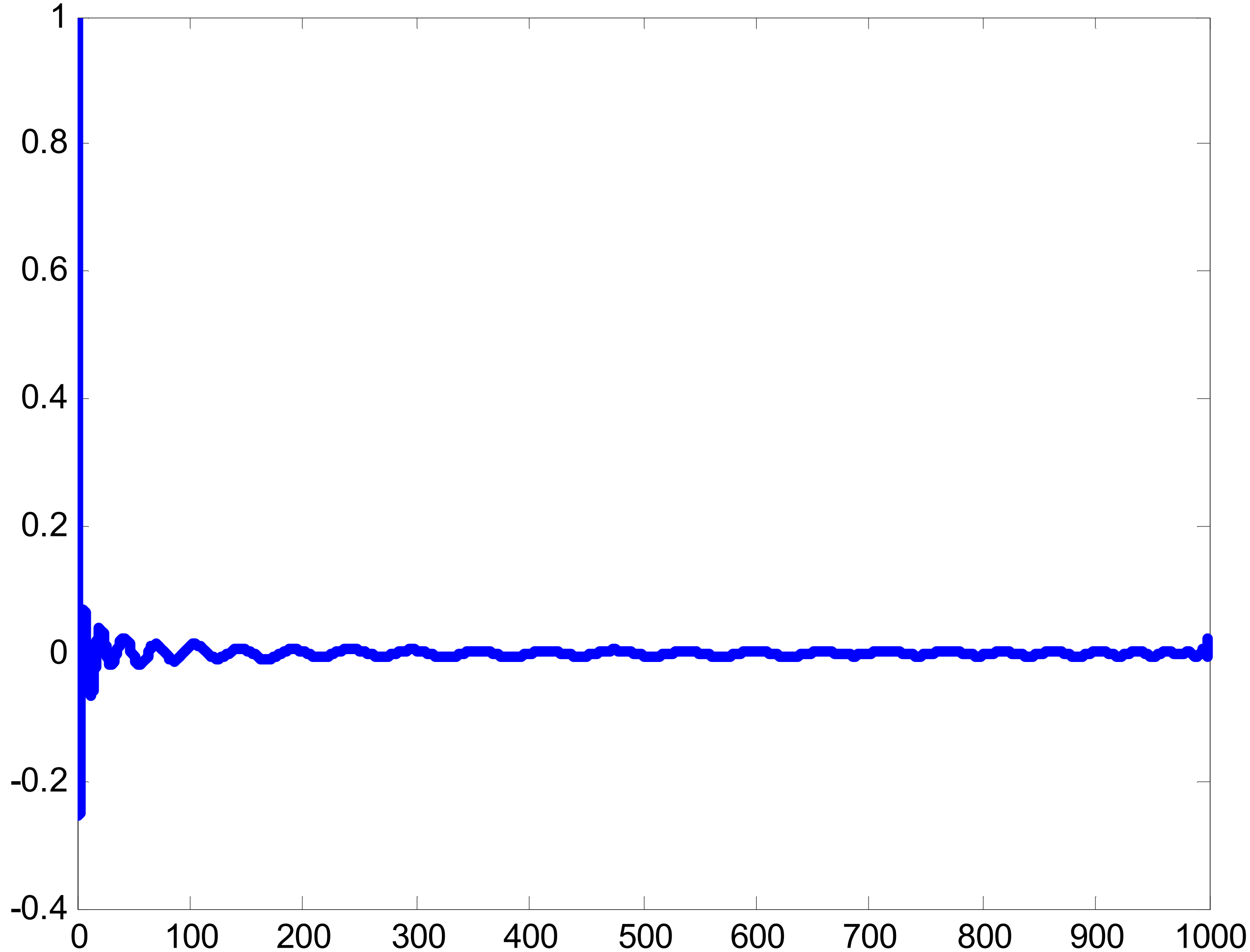
q is poly of degree m or less that has value 1 at 0.

Matrix: bidiagonal, diagonal is 0.1, 1, 2, 3, …999, superdiagonal is all 1's

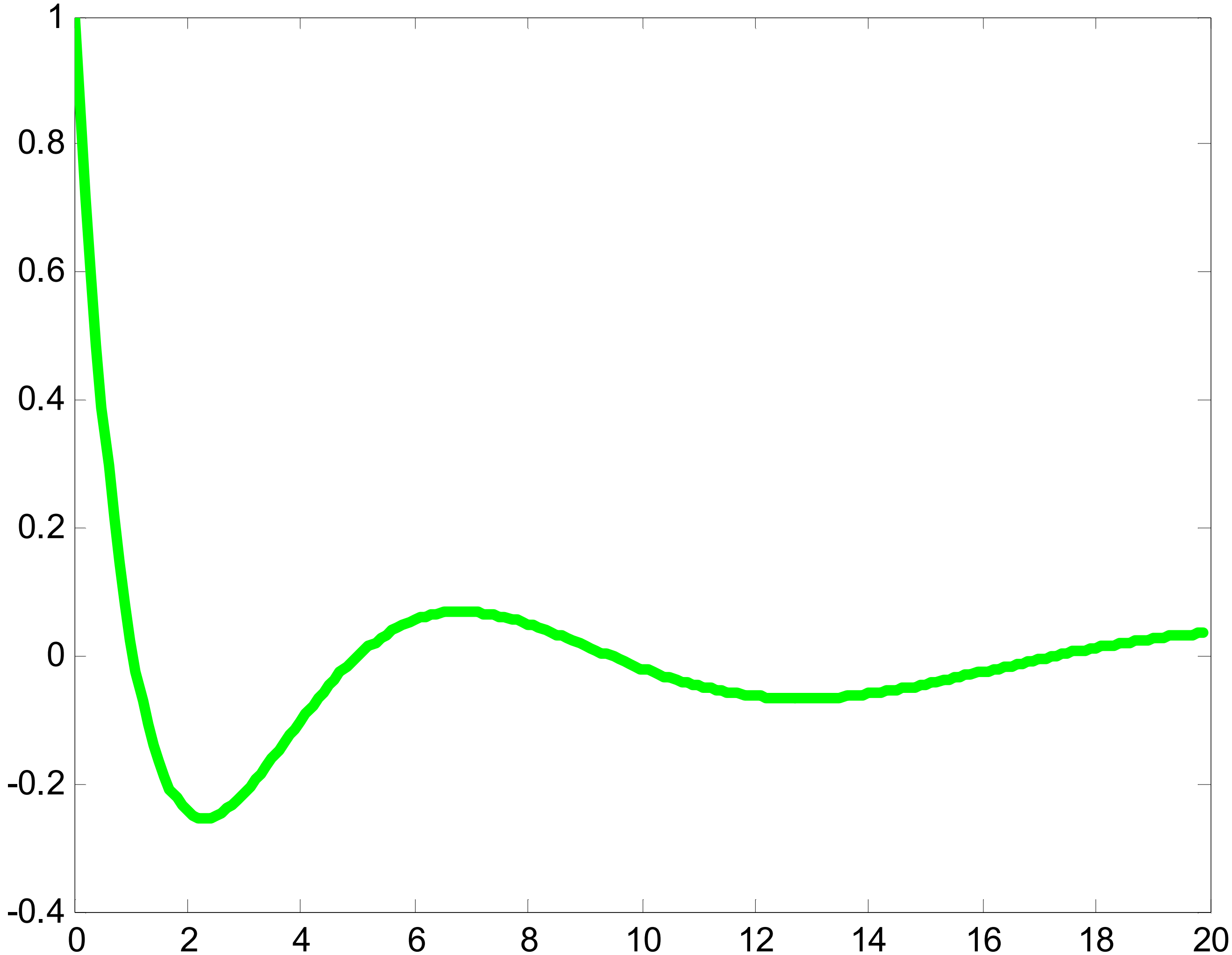
GMRES polynomial of degree 10



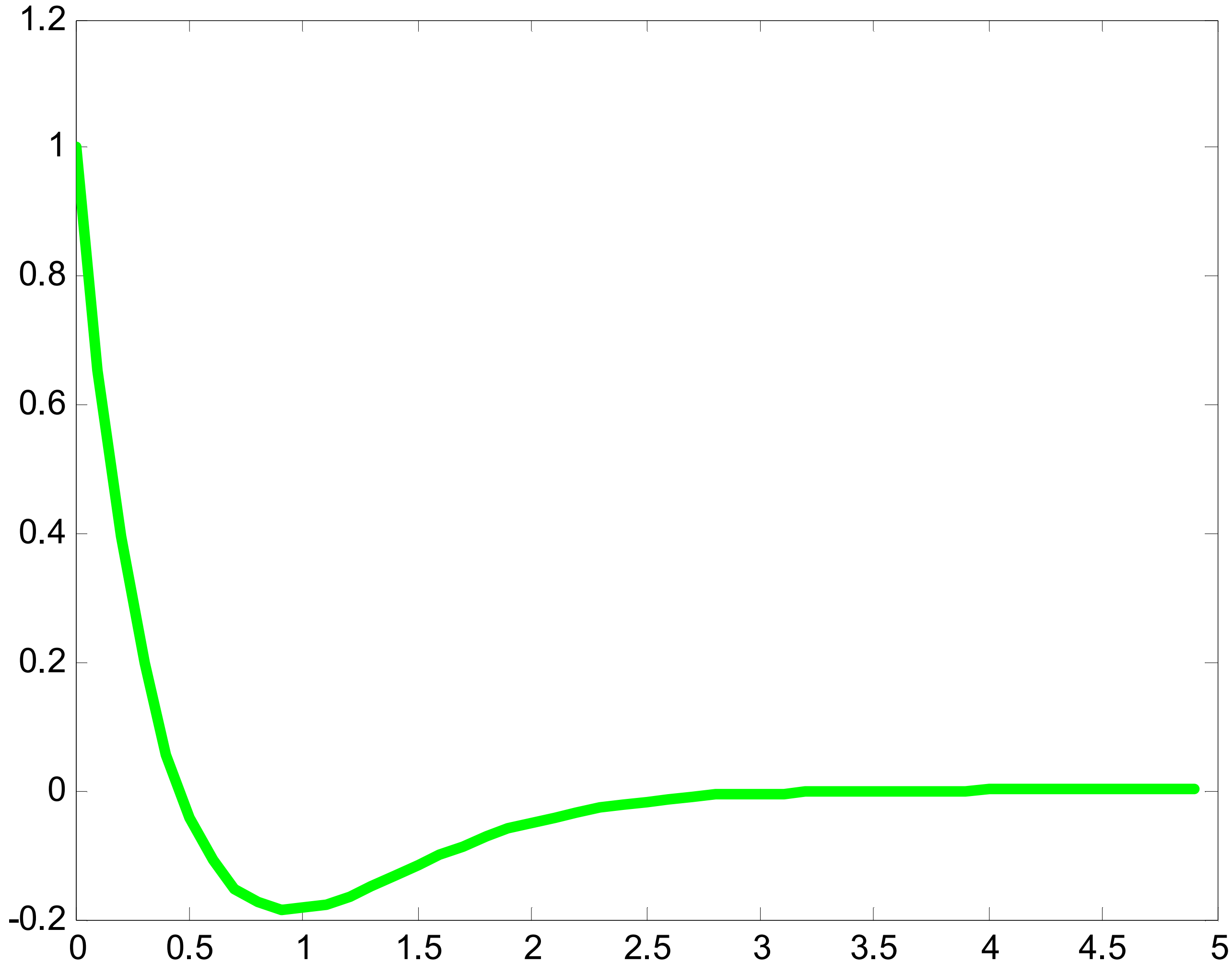
GMRES polynomial of degree 50



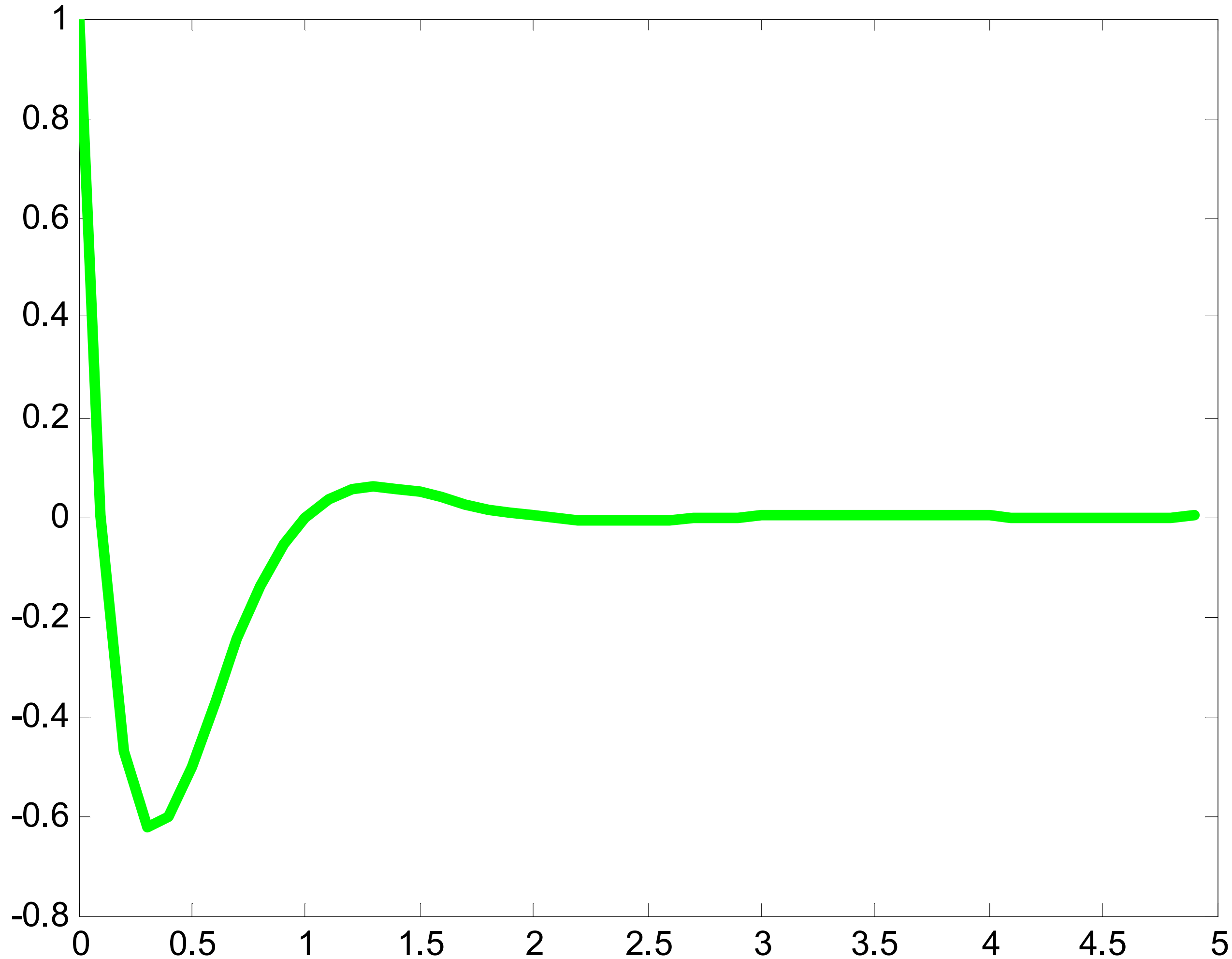
GMRES polynomial of degree 50
(close up view)



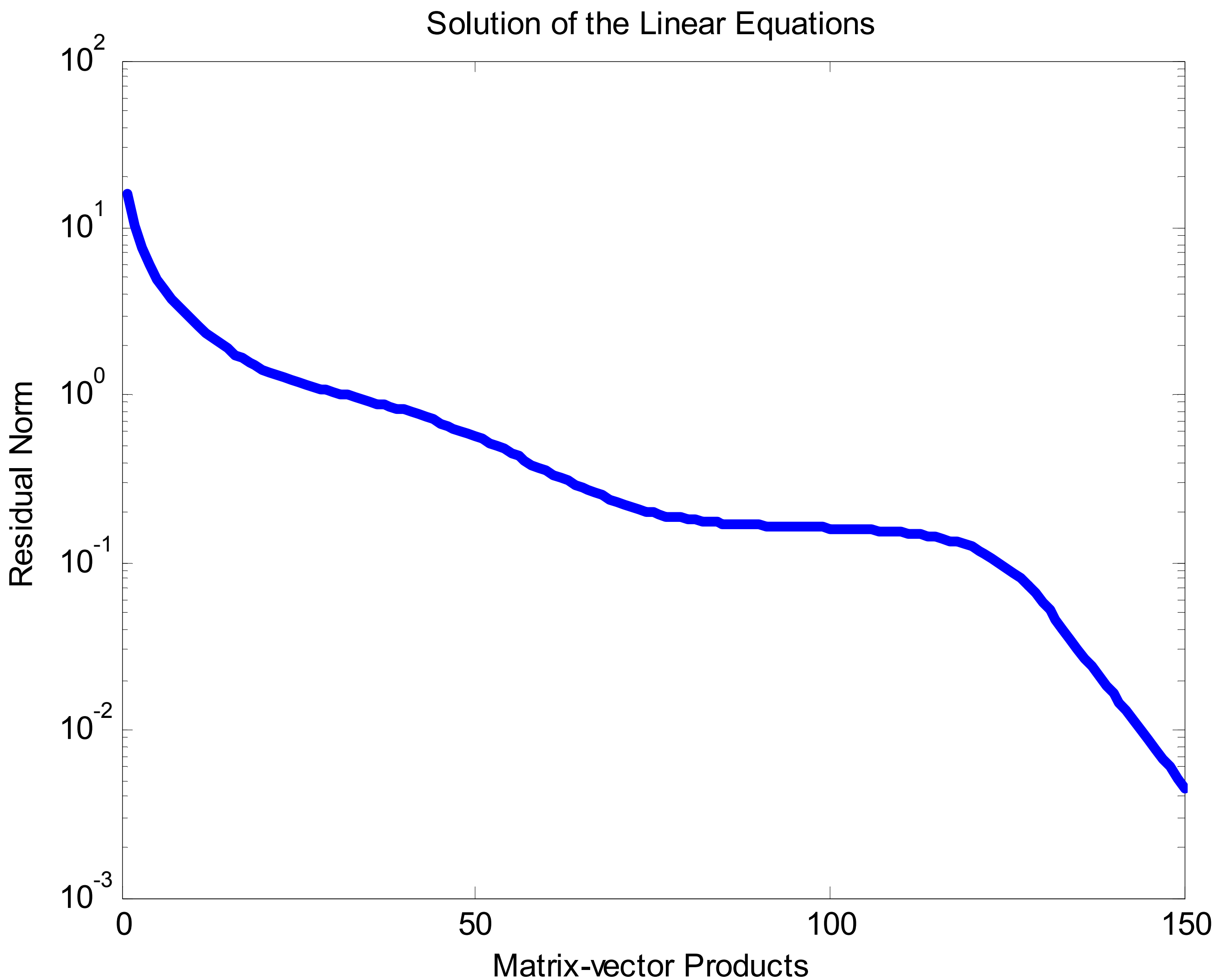
GMRES polynomial of degree 100
(close up view)



GMRES polynomial of degree 150
(close up view)



Residual norm curve



Small eigenvalues and Krylov methods

- SPD matrix: convergence is approximately related to

$$\sqrt{\lambda_n} / \sqrt{\lambda_1}$$

Example: Eigenvalues: 0.1, 1, 2, 3, . . . n

If remove 4 eigenvalues, convergence is improved by factor of about 6 (remove 10 for improvement factor of 10).

Non-restarted methods like CG and BiCGStab naturally remove some eigenvalues as the iteration proceeds – leads to superlinear convergence.

Restarted GMRES often cannot “remove” eigenvalues.

quiz

1) Which spectrum is tougher?

1, 2, 3, … 99, 1000

or

0.01, 0.02, 0.03, … 9.99, 10

2) Indefinite spectrum, how much harder?

1, 2, 3, …999,1000

versus

-1000, -99, -98, …-2, -1, 1, 2, 3, …99, 1000

Stingrays at a sandbar off Grand Cayman

Improving restarted GMRES

- Make restarted GMRES more like a non-restarted method. Save some information at end of each cycle of restarted GMRES and use it to help the next cycle.

What to save?

Save approximations to eigenvectors corresponding to small eigenvalues.

How do we use them?

Two ways:

1) Build preconditioner out of approx. eigenvectors
2) Put approx. eigenvectors in subspace

Build Preconditioner

- Kharchenko and Yeremin, 1995
- Erhel, Burrage and Pohl, 1996

Analyzed by: Eiermann, Ernst and Schneider, 2000

Problem: One cycle may not give good eigenvector approximations.

Methods that improve approximations for preconditioner:

- Baglama, Calvetti, Golub and Reichel, 1998
- Burrage and Erhel, 1998

Add approximate eigenvectors to the subspace.

- Morgan, 1995, etc.

Subspace:

$$Span\{y_1, y_2, \ldots, y_k, r_0, Ar_0, A^2 r_0, \ldots, A^{m-k-1} r_0\}$$

Eigenvector portion + Krylov portion

GMRES with deflated restarting – solve linear equations and compute eigenvalues simultaneously

- Add approximate eigenvectors to the Krylov subspace for the linear equations.
- This essentially removes the corresponding eigenvalues and can thus improve convergence.
- Let y_i's be harmonic Ritz vectors. The subspace is

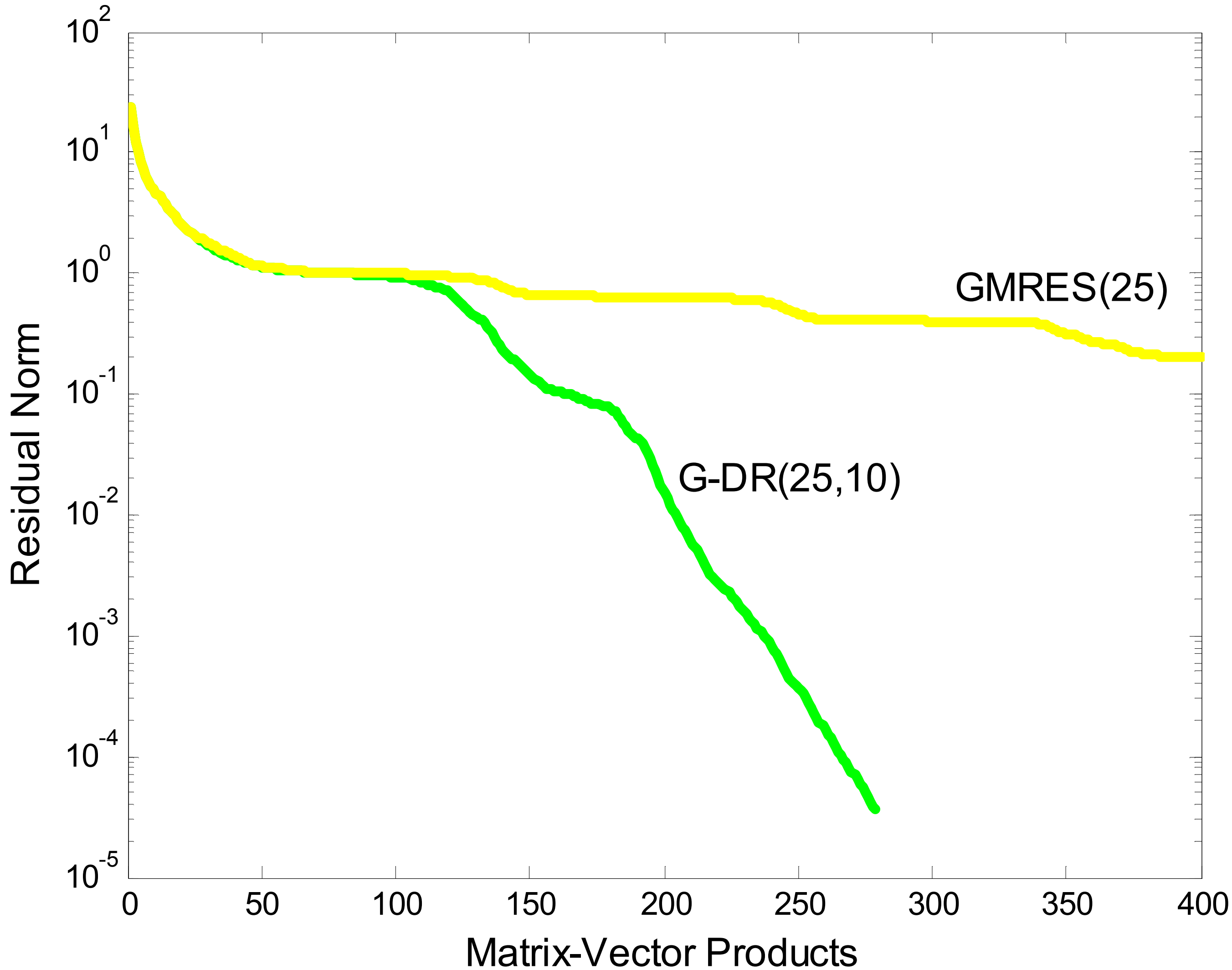
$$Span\{y_1, y_2, \ldots, y_k, r_0, Ar_0, A^2 r_0, \ldots, A^{m-k-1} r_0\}$$

- One deflated GMRES method is called GMRES-DR.
- Mathematically equivalent: GMRES-IR or GMRES with implicit restarting.
- FOM-DR and FOM-IR are equivalent to Sorensen’s Implicitly Restarted Arnoldi (ARPACK) with linear equations solution added on. Also related to Wu and Simon’s Thick Restarted Lanczos.
- GMRES-DR is similar to Implicitly Restarted Arnoldi, but has computation of harmonic Ritz values instead of Ritz values.

GMRES-DR vs. regular restarted GMRES

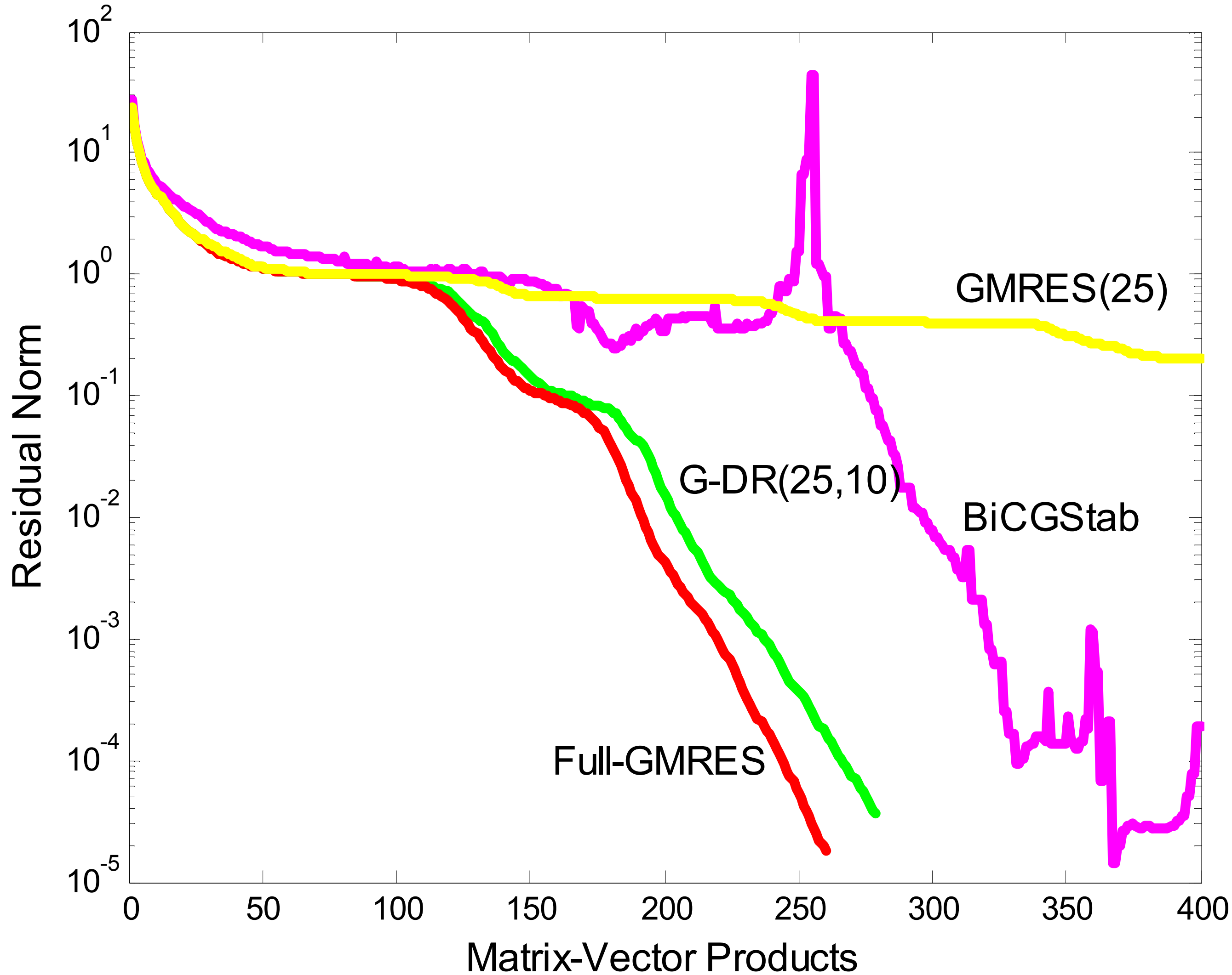
Matrix: bidiagonal, diagonal is 0.1, 1, 2, 3, …1999

superdiagonal is all 1's



GMRES-DR vs. other methods

Matrix: bidiagonal, diagonal is 0.1, 1, 2, 3, …1999
superdiagonal is all 1's

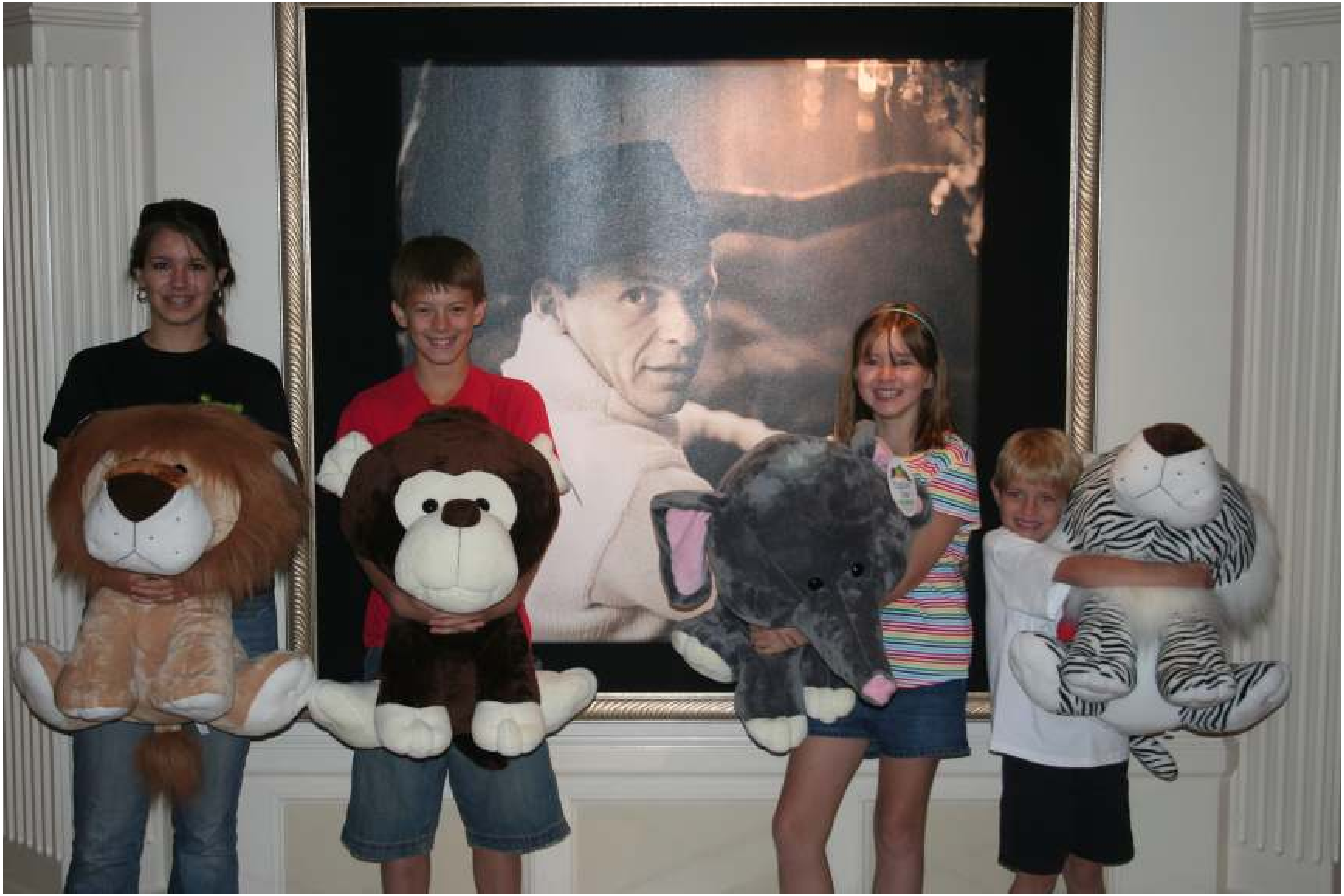


- GMRES-DR creates an Arnoldi-like recurrence,

$$AV_k = V_{k+1}\overline{H}_k$$

where V_k is the *n by k* matrix with columns spanning the approximate eigenvectors, and H_k is small, *k+1 by k*.

- Have both approximate eigenvectors and their products with A in compact storage.



For multiple right-hand sides

- Solve first right-hand side with GMRES-DR
- Use the computed eigenvectors for the other right-hand sides. Method is

  GMRES-Proj:

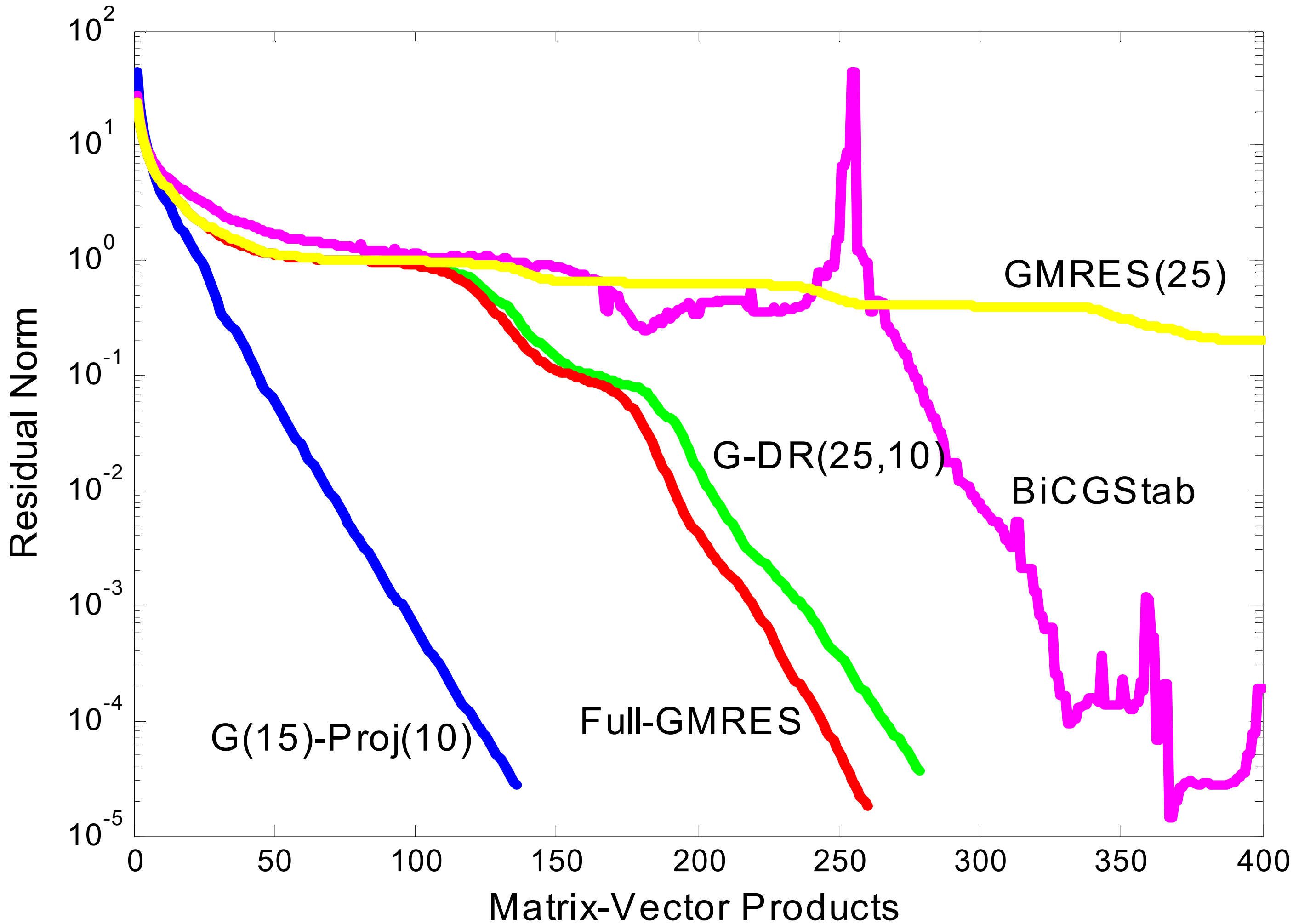
  Alternate:

  1) projection over eigenvectors

  2) cycles of regular GMRES

GMRES-Proj for the 2nd rhs (following GMRES-DR for 1st rhs)

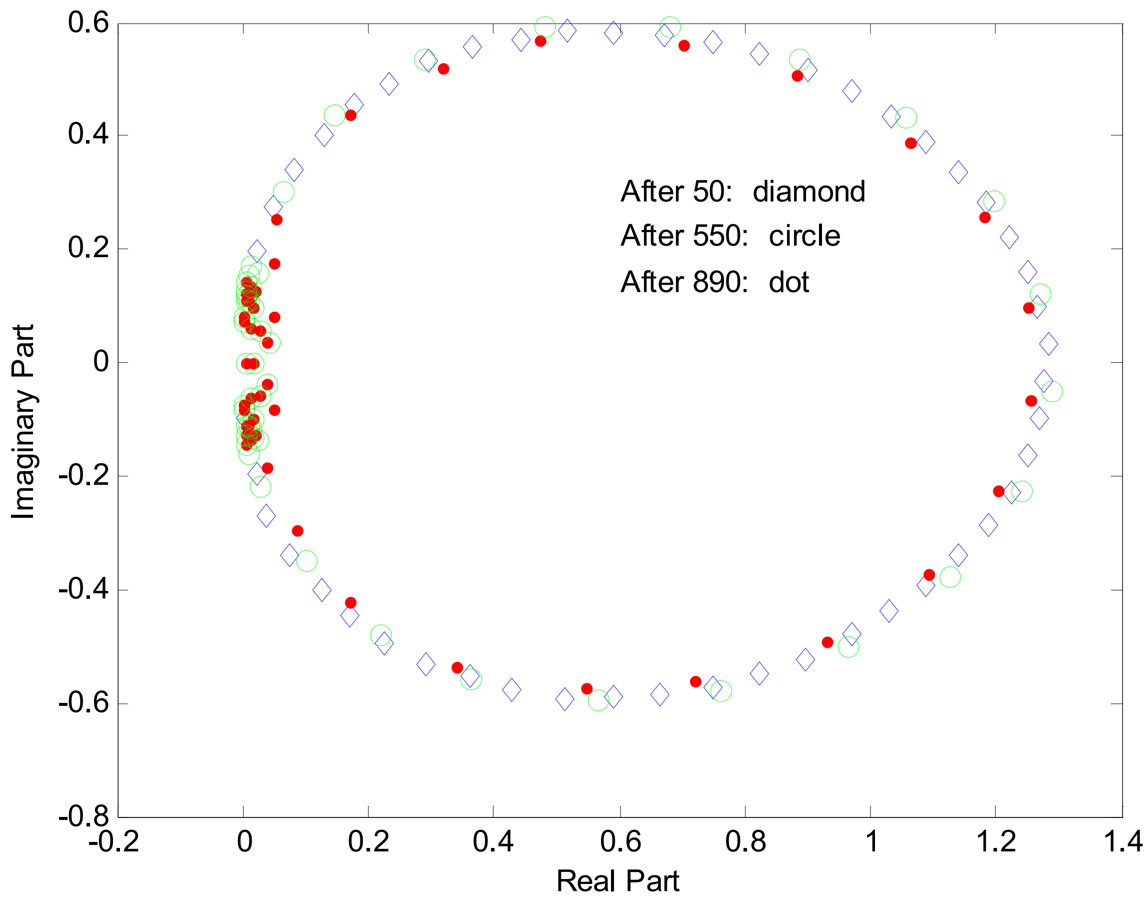
Matrix: bidiagonal, diagonal is 0.1, 1, 2, 3, …1999



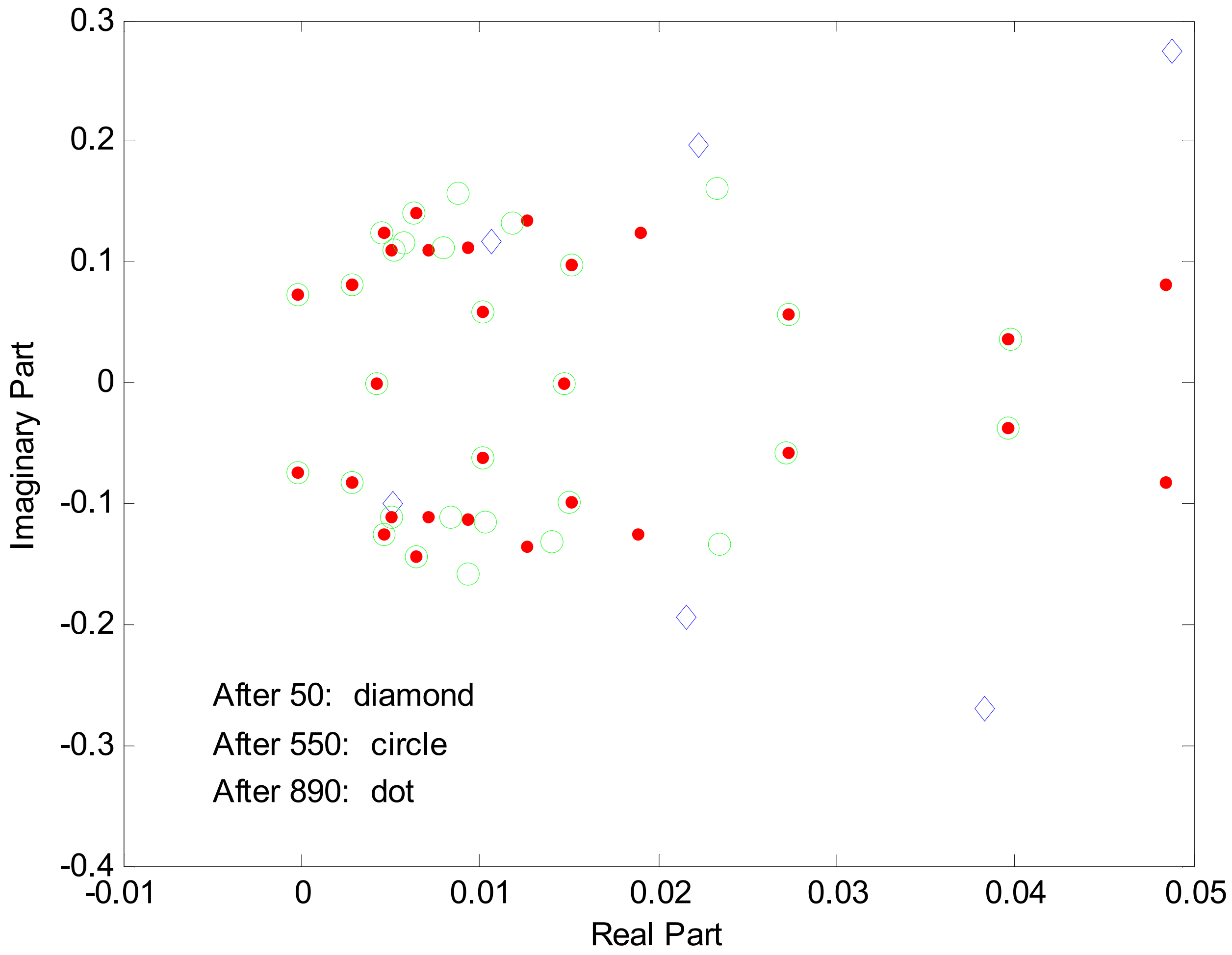
Application from Quantum Chromodynamics (QCD)

- Matrix is size n=¼ million, complex, non-Hermitian
- They want to solve 12 or so right-hand sides for each matrix.
- They also want about 7 shifts ($A-\sigma_i I$) for each right-hand sides – afterall, physicists are greedy.

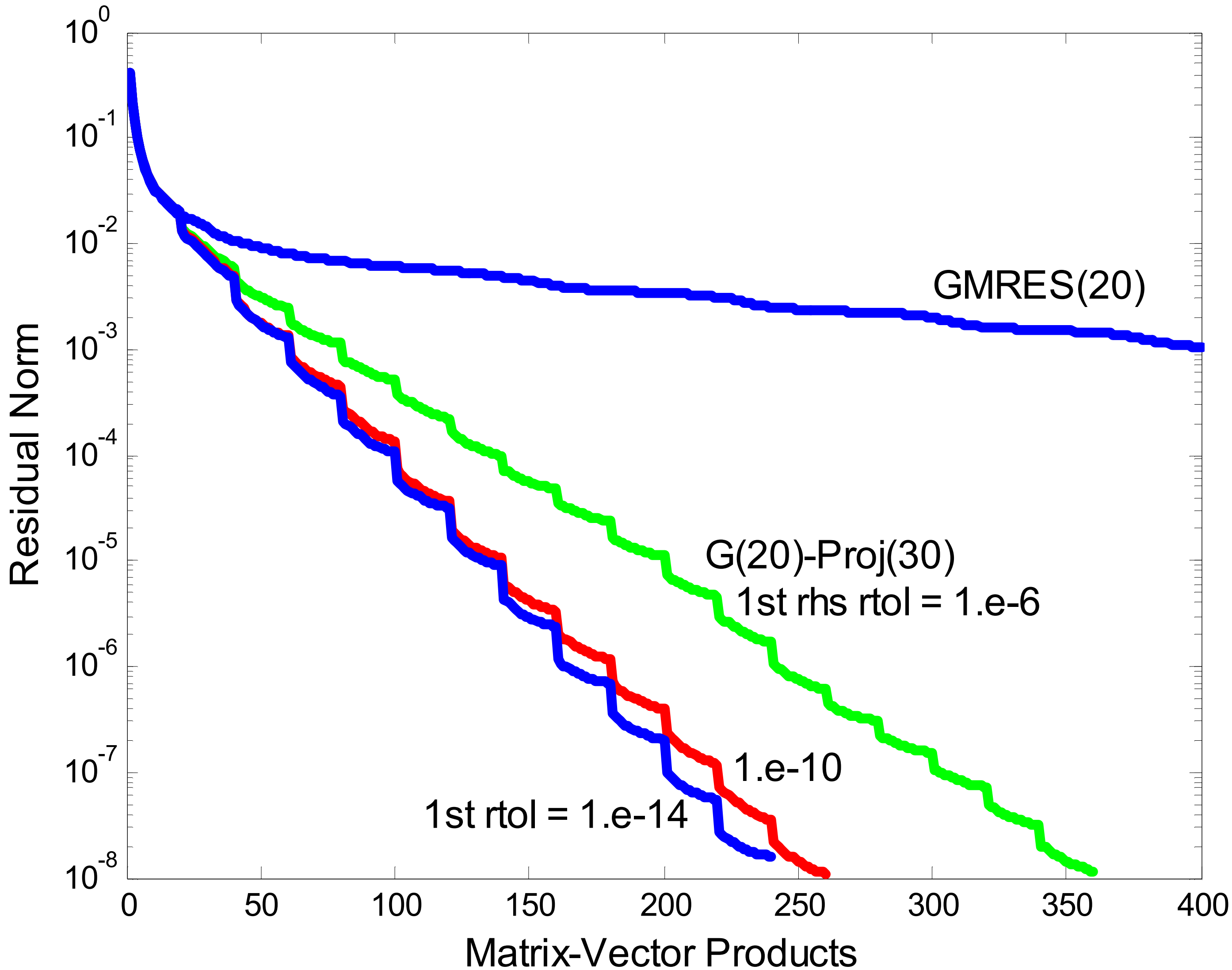
Approximate eigenvalues (harmonic Ritz values) at several points of the GMRES-DR iteration.



Close-up of smallest harmonic Ritz values



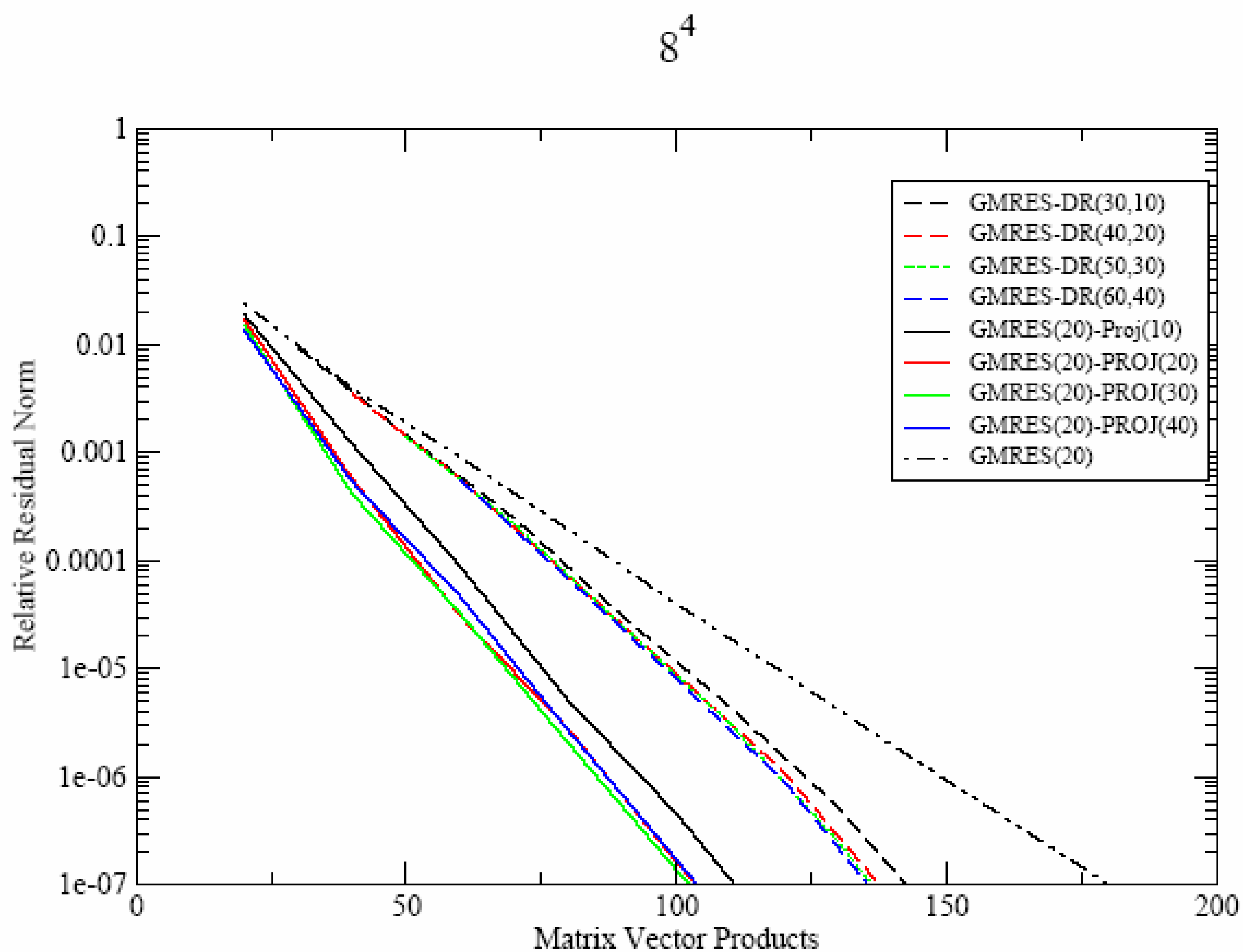
GMRES-Proj for 2nd rhs of QCD matrix

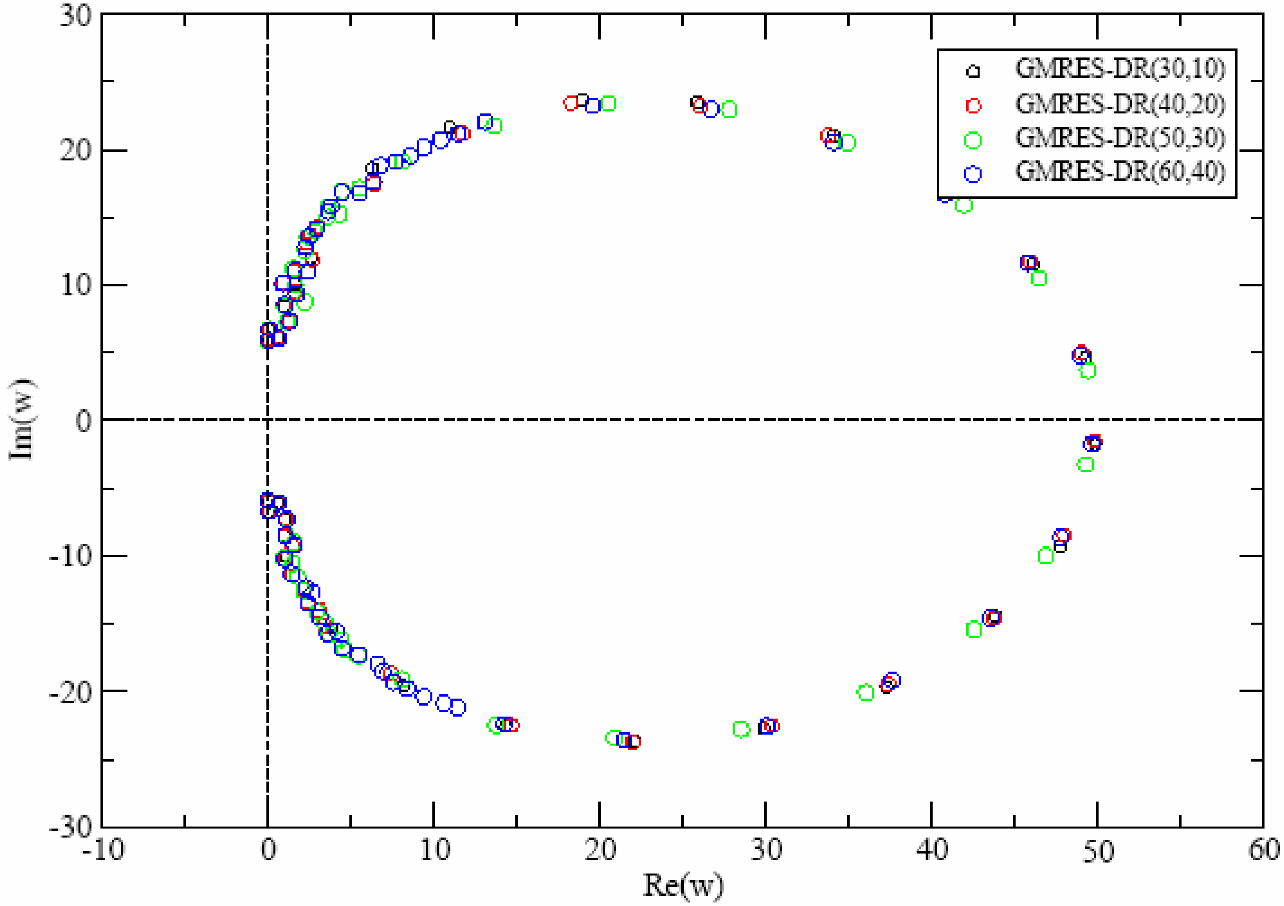


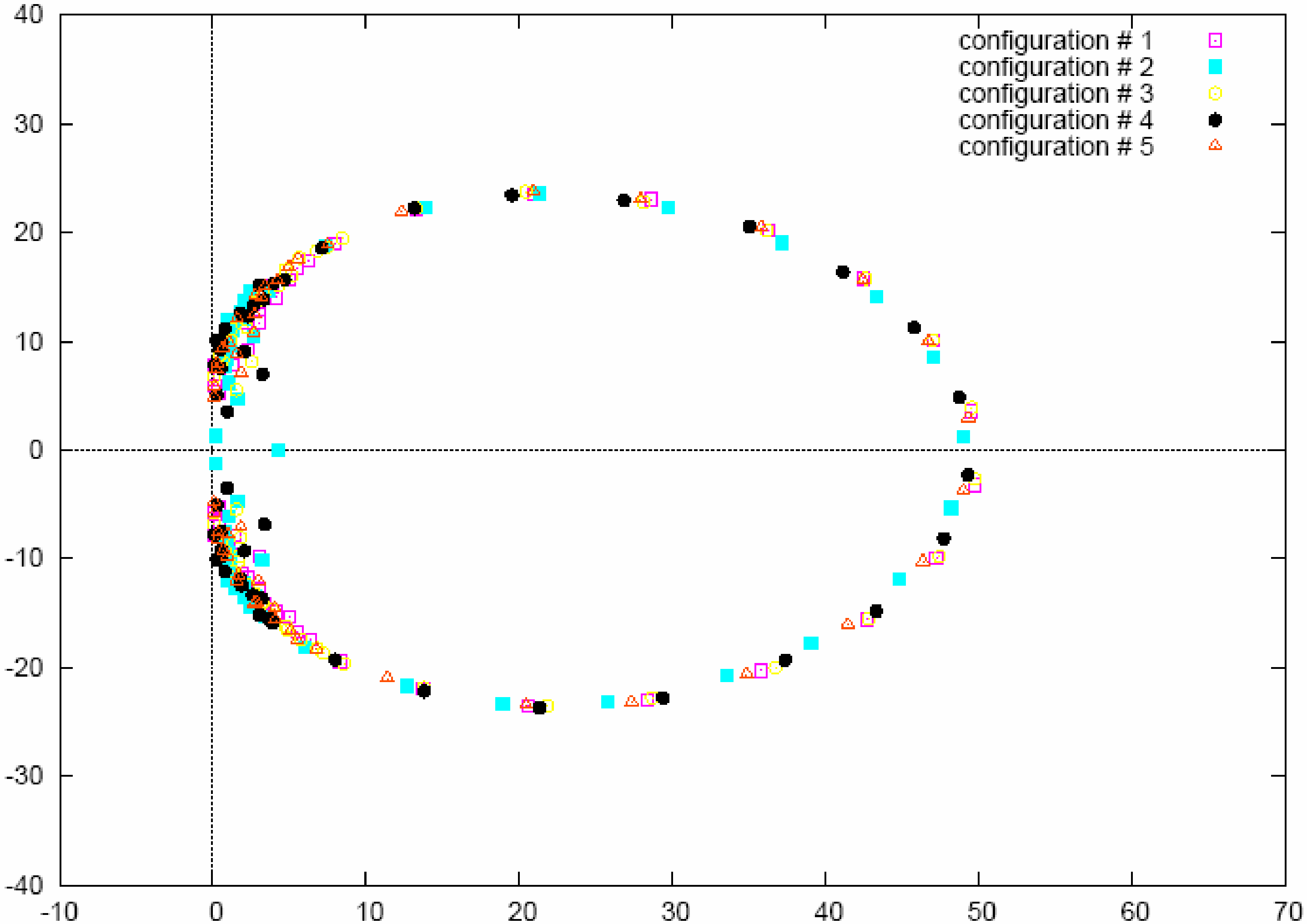
Some questions about Wilson matrix computations.

- How much does deflating eigenvalue generally improve convergence (for the multiple right-hand sides)?
- How many eigenvalues should be deflated for best performance with multiple right-hand sides?
- How does the optimal number of eigenvalues depend on the size of the problem?
- After eigenvalues are deflated, how does the number of iterations vary with the size of the problem?
- Are exceptional configurations more likely for larger problems?

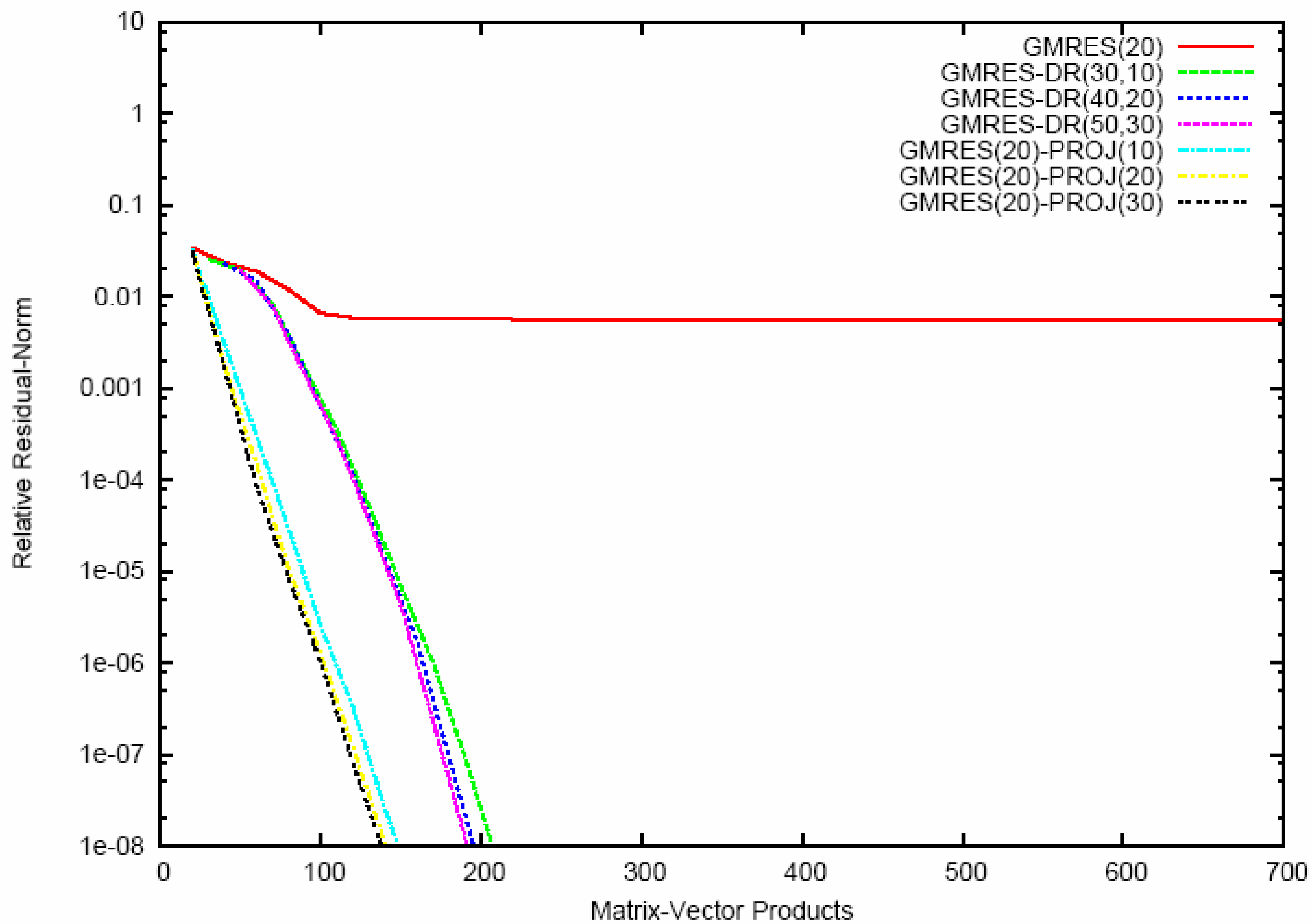
Small problem, 8^4 lattice, matrix is 24578 by 24578
We are approximately at kappa critical.



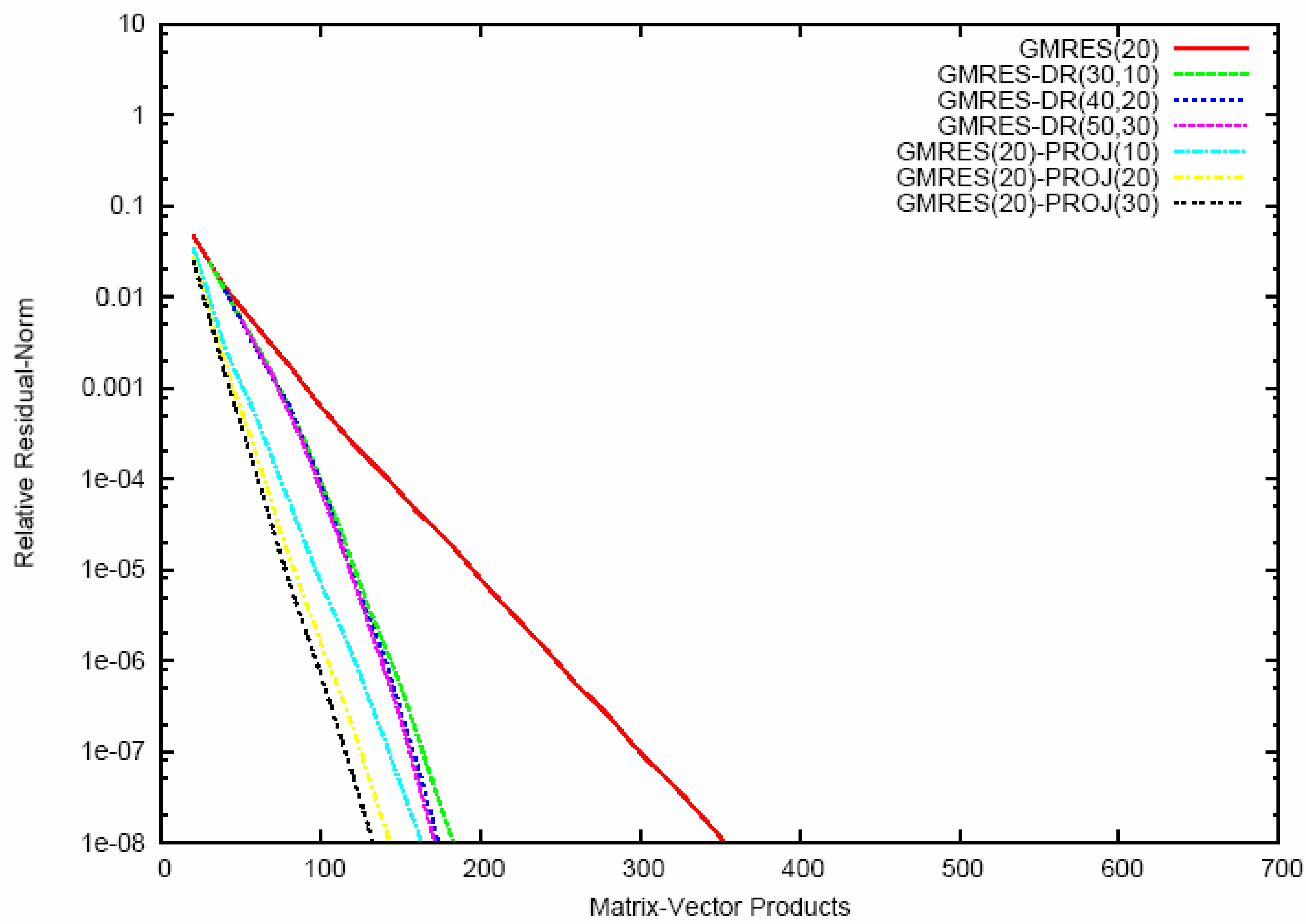

GMRES-DR(30,10)
GMRES-DR(40,20)
GMRES-DR(50,30)
GMRES-DR(60,40)
Im(w)
Re(w)
30
20
10
0
-10
-20
-30
-10
0
10
20
30
40
50
60


configuration # 1
configuration # 2
configuration # 3
configuration # 4
configuration # 5
40
30
20
10
0
-10
-20
-30
-40
-10
0
10
20
30
40
50
60
70

Configuration number 2



Configuration number 4



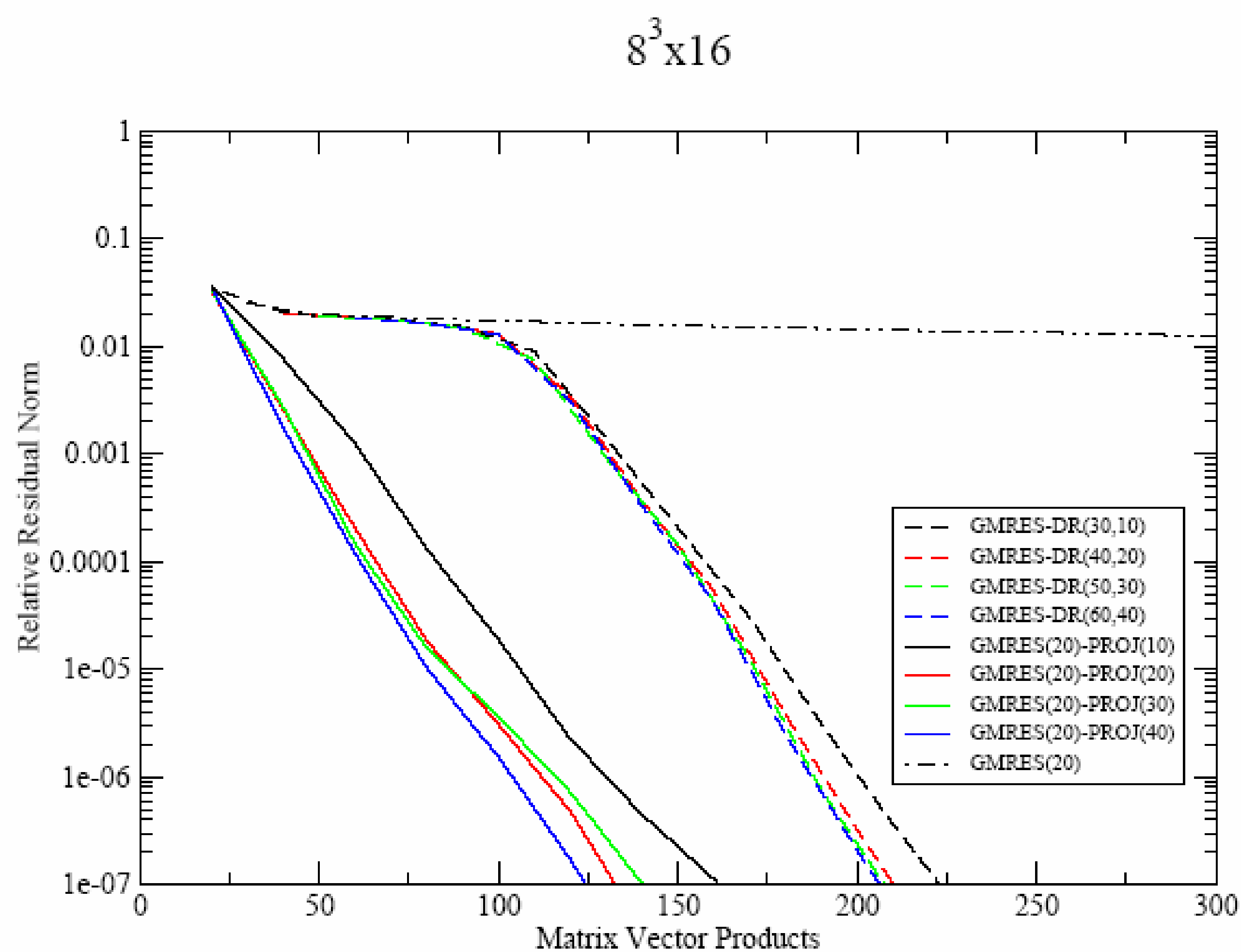
Exceptional Configurations.

Are they

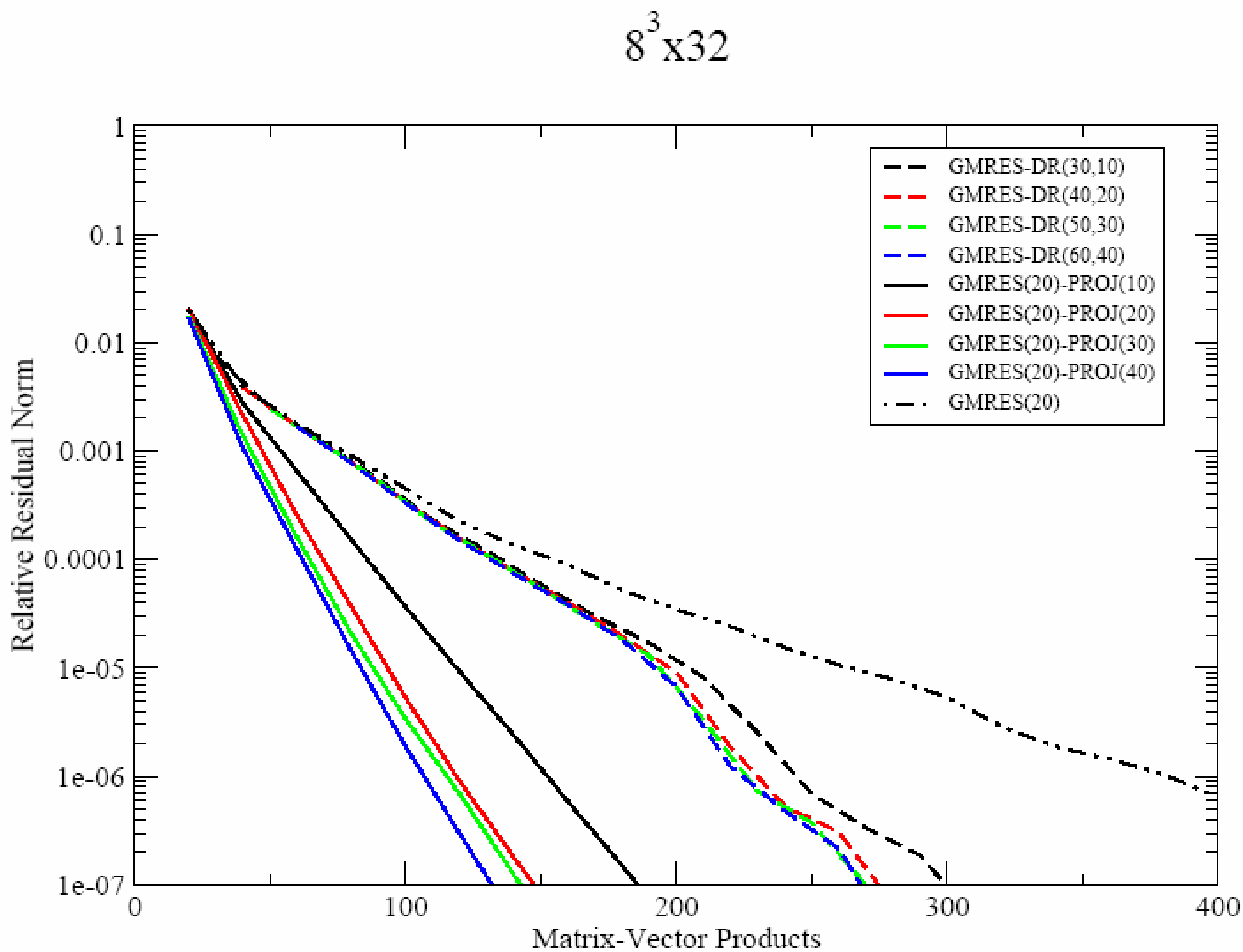
the scourge of QCD calculations,

or

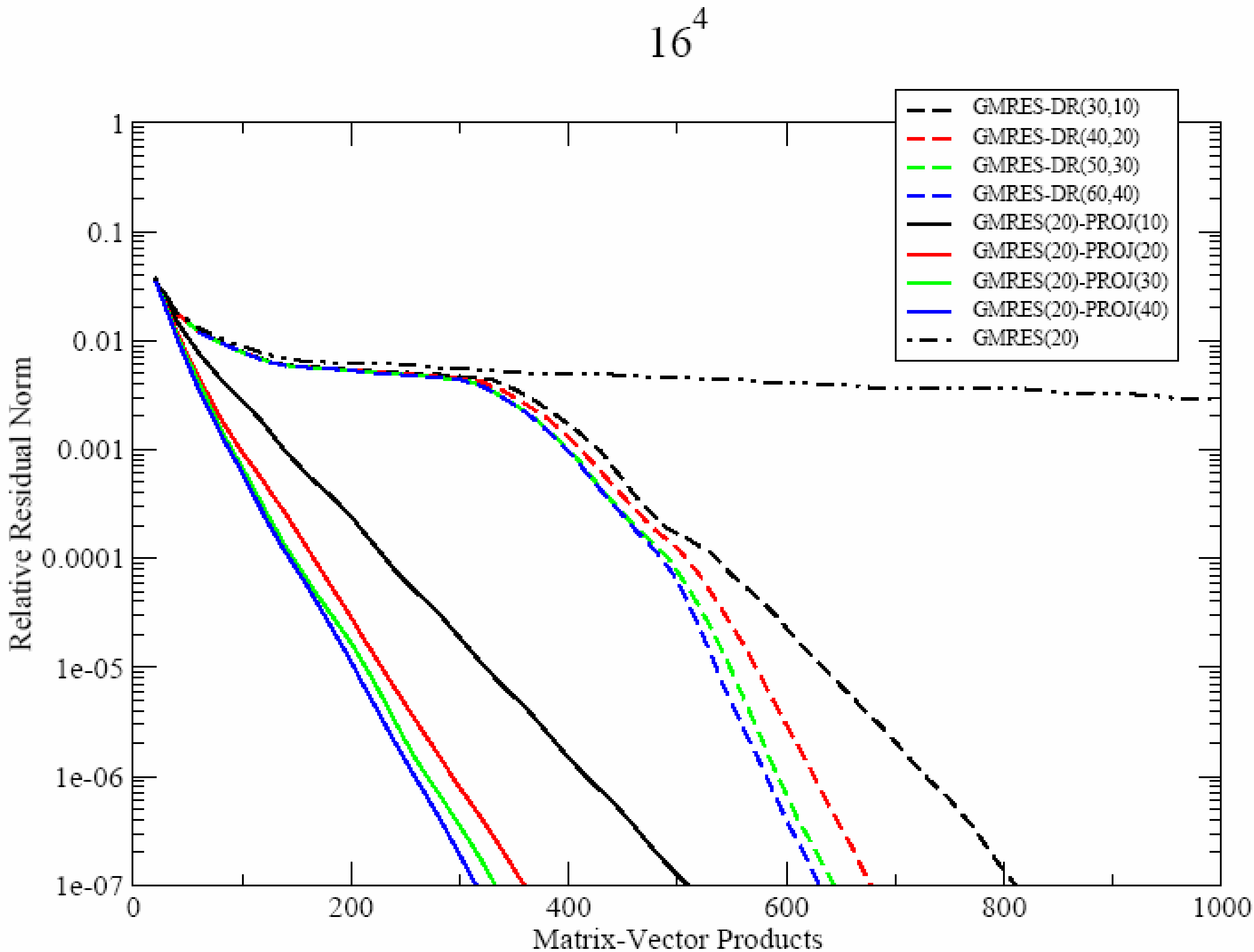
a wonderful opportunity to improve solution methods?

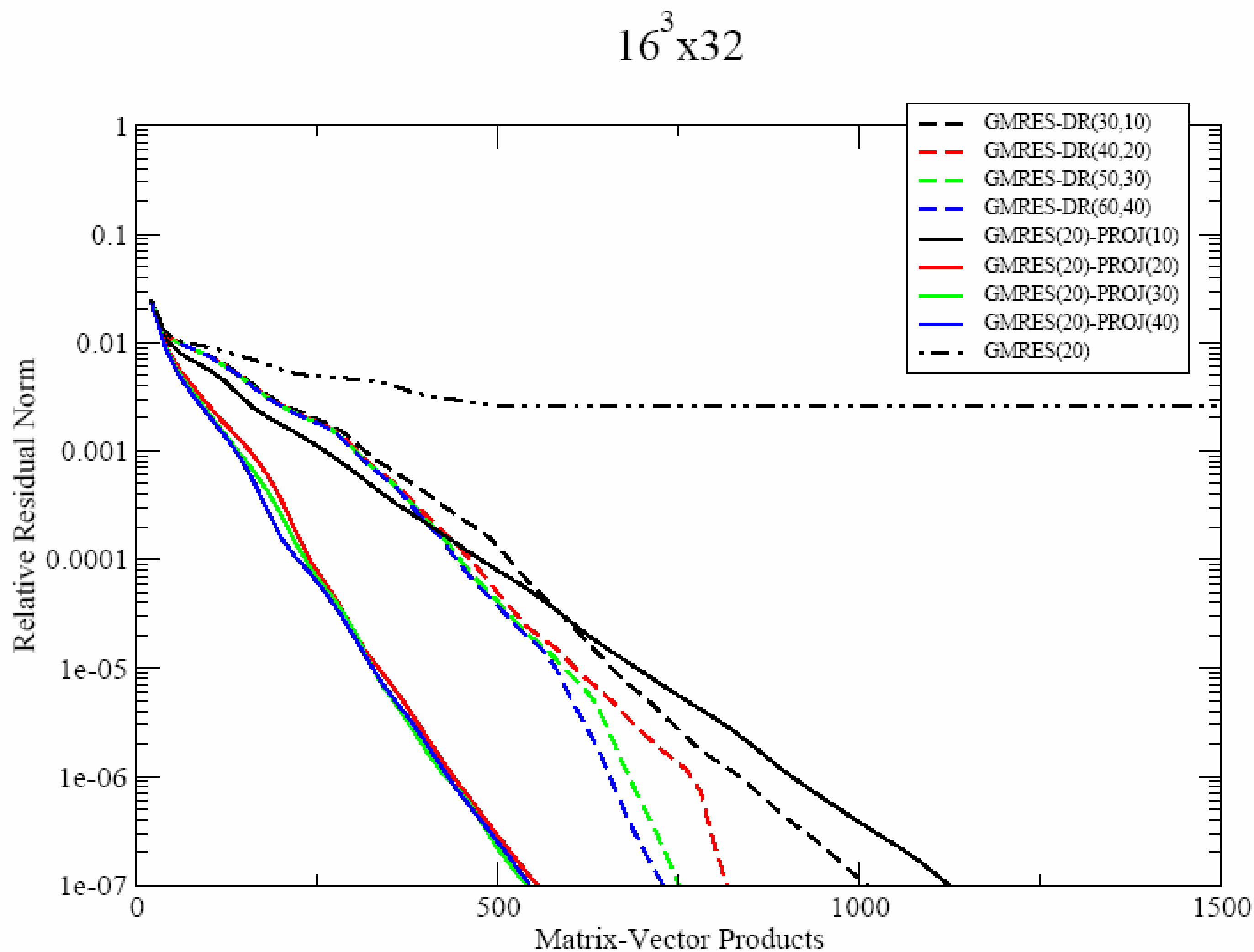


8^3x16
Relative Residual Norm
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
0
50
100
150
200
250
300
Matrix Vector Products
GMRES-DR(30,10)
GMRES-DR(40,20)
GMRES-DR(50,30)
GMRES-DR(60,40)
GMRES(20)-PROJ(10)
GMRES(20)-PROJ(20)
GMRES(20)-PROJ(30)
GMRES(20)-PROJ(40)
GMRES(20)



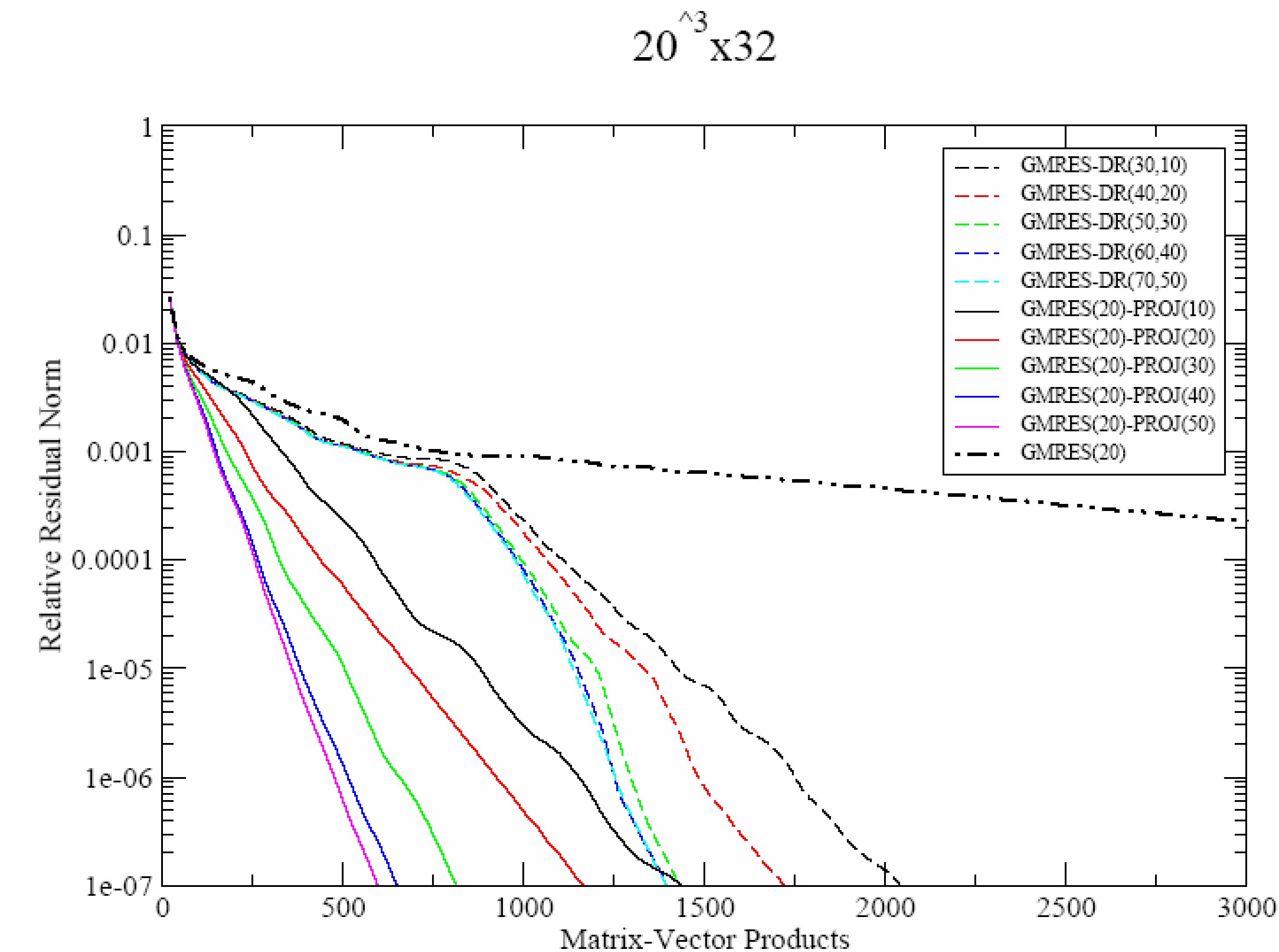
8^3x32
Relative Residual Norm
Matrix-Vector Products
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
0
100
200
300
400
GMRES-DR(30,10)
GMRES-DR(40,20)
GMRES-DR(50,30)
GMRES-DR(60,40)
GMRES(20)-PROJ(10)
GMRES(20)-PROJ(20)
GMRES(20)-PROJ(30)
GMRES(20)-PROJ(40)
GMRES(20)


16^4
GMRES-DR(30,10)
GMRES-DR(40,20)
GMRES-DR(50,30)
GMRES-DR(60,40)
GMRES(20)-PROJ(10)
GMRES(20)-PROJ(20)
GMRES(20)-PROJ(30)
GMRES(20)-PROJ(40)
GMRES(20)
Relative Residual Norm
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
0
200
400
600
800
1000
Matrix-Vector Products



16^3x32
GMRES-DR(30,10)
GMRES-DR(40,20)
GMRES-DR(50,30)
GMRES-DR(60,40)
GMRES(20)-PROJ(10)
GMRES(20)-PROJ(20)
GMRES(20)-PROJ(30)
GMRES(20)-PROJ(40)
GMRES(20)
Relative Residual Norm
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
0
500
1000
1500
Matrix-Vector Products

1,536,000 by 1,536,000 matrix



Some questions about Wilson matrix computations.

- How much does deflating eigenvalue generally improve convergence (for the multiple right-hand sides)?

  Answer: Sometimes a little, but often a lot.

- How many eigenvalues should be deflated for best performance with multiple right-hand sides?

  Answer: Sometimes 10 is plenty, but it may be 40 (or more?).

Some questions about Wilson matrix computations.

- How does the optimal number of eigenvalues depend on the size of the problem?

Answer: it increases, but not nearly proportional to n (fortunately).

n=24,576	*k=10*
n=49,152	*k=10 to 20*
n=98,304	*k=20*
n=393,216	*k=20*
n=786,432	*k=20*
n=1,536,000	*k=40*

Some questions about Wilson matrix computations.

- After eigenvalues are deflated, how does the number of iterations vary with the size of the problem?

Answer: It increases, but not nearly proportional to n (fortunately).

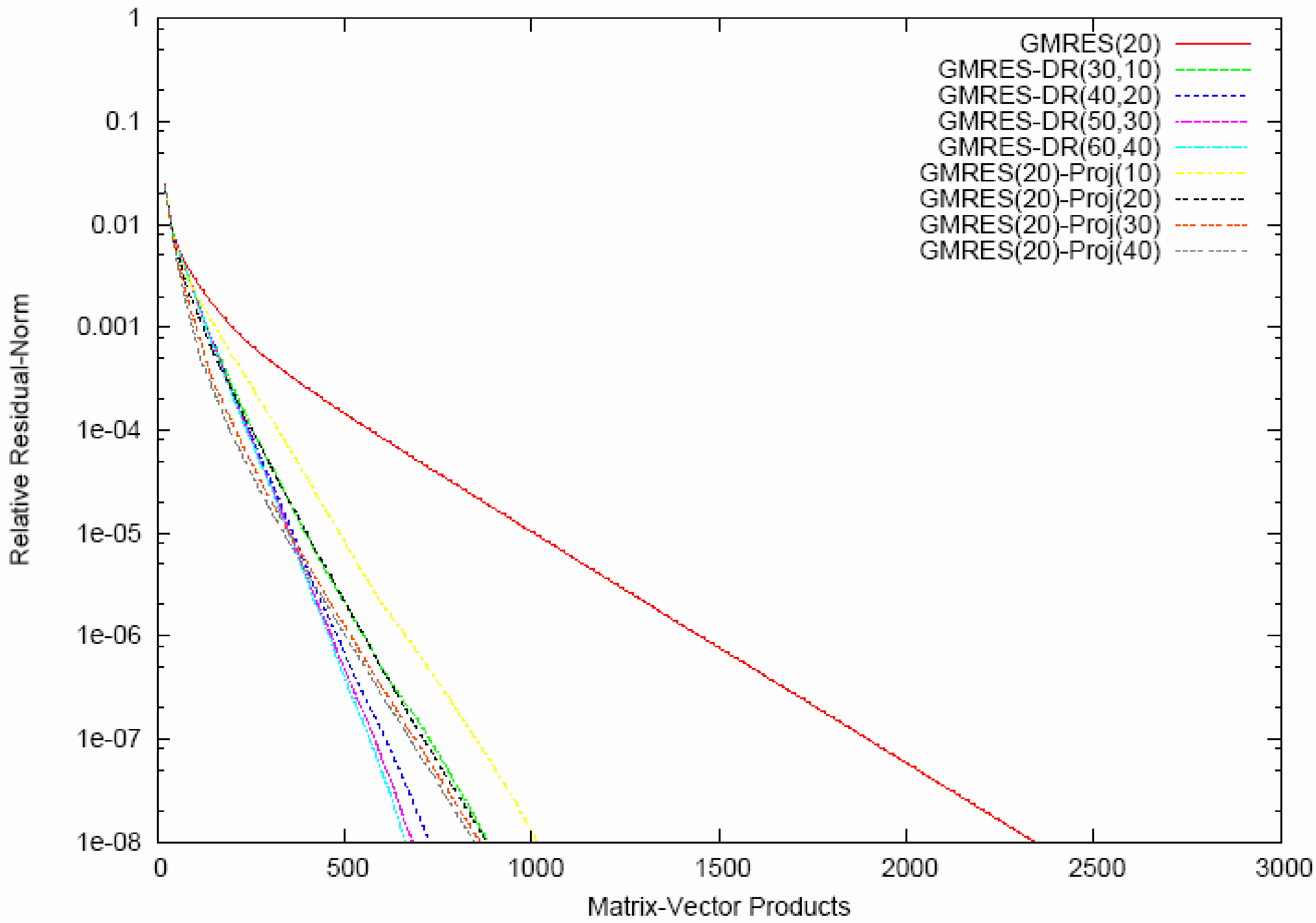
n=24,576	*its = 100*
n=49,152	*its = 125*
n=98,304	*its = 125 - 150*
n=393,216	*its = 325*
n=786,432	*its = 600*
n=1,536,000	*its = 600*

Some questions about Wilson matrix computations.

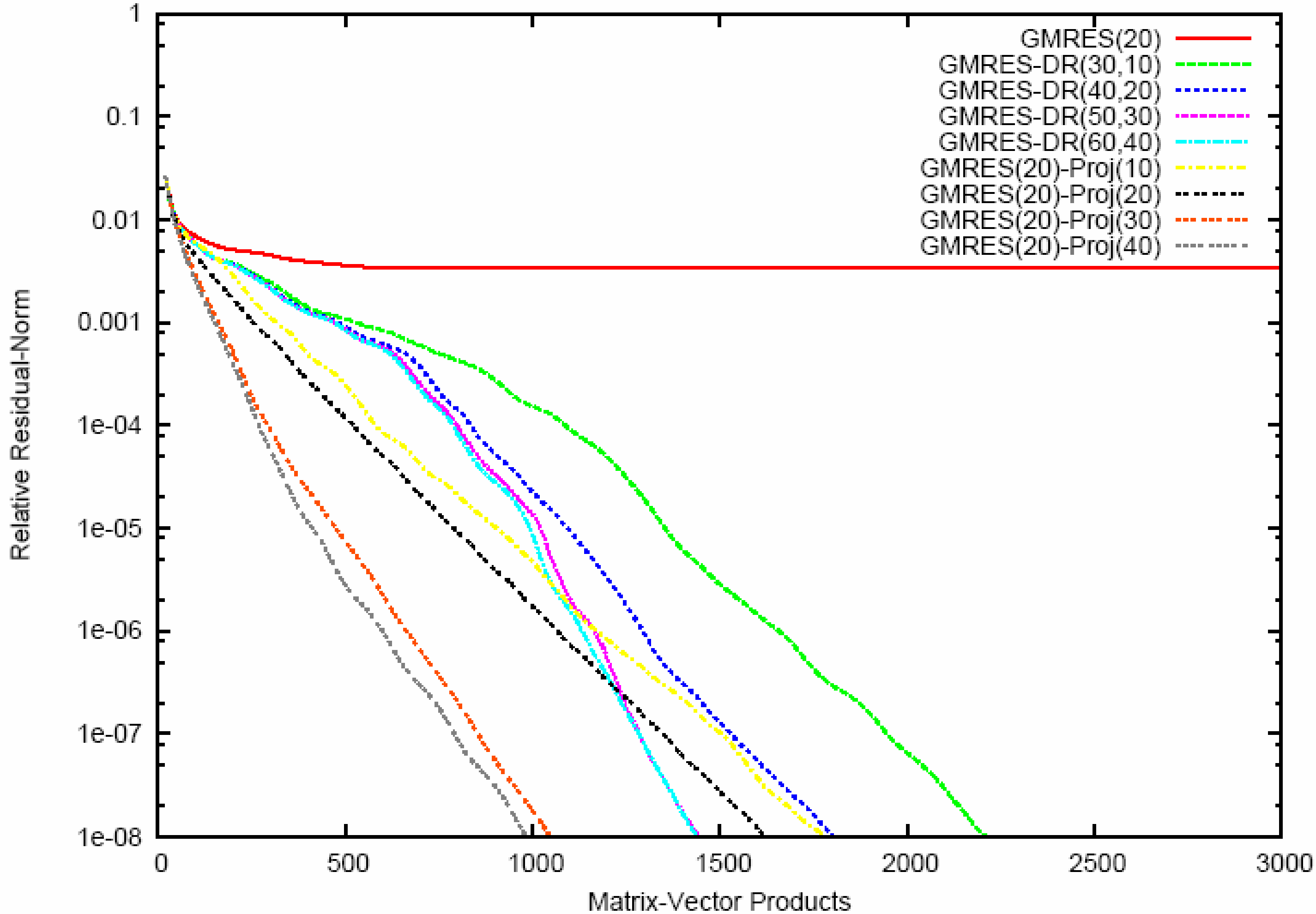
- Are exceptional configurations more likely for larger problems?

Answer: don't know, but it seems likely.

Twisted Mass, 20^3 x 32



Twisted Mass, 20^3 x 32, another configuration



One of Sigfried and Roy’s Tigers

Multi-masses or multiple shifts.

$$(A - \sigma_i I)x_i = b$$

- Krylov subspaces are shift invariant in that $A-\sigma_\iota I$ generates the same Krylov subspace no matter what the shift.

- So the goal is to solve all shifted systems with ONE Krylov subspace.

- For non-restarted methods this has been done, for example; QMR (Freund) and BiCGStab (Frommer).

Restarted methods with multiple shifts.

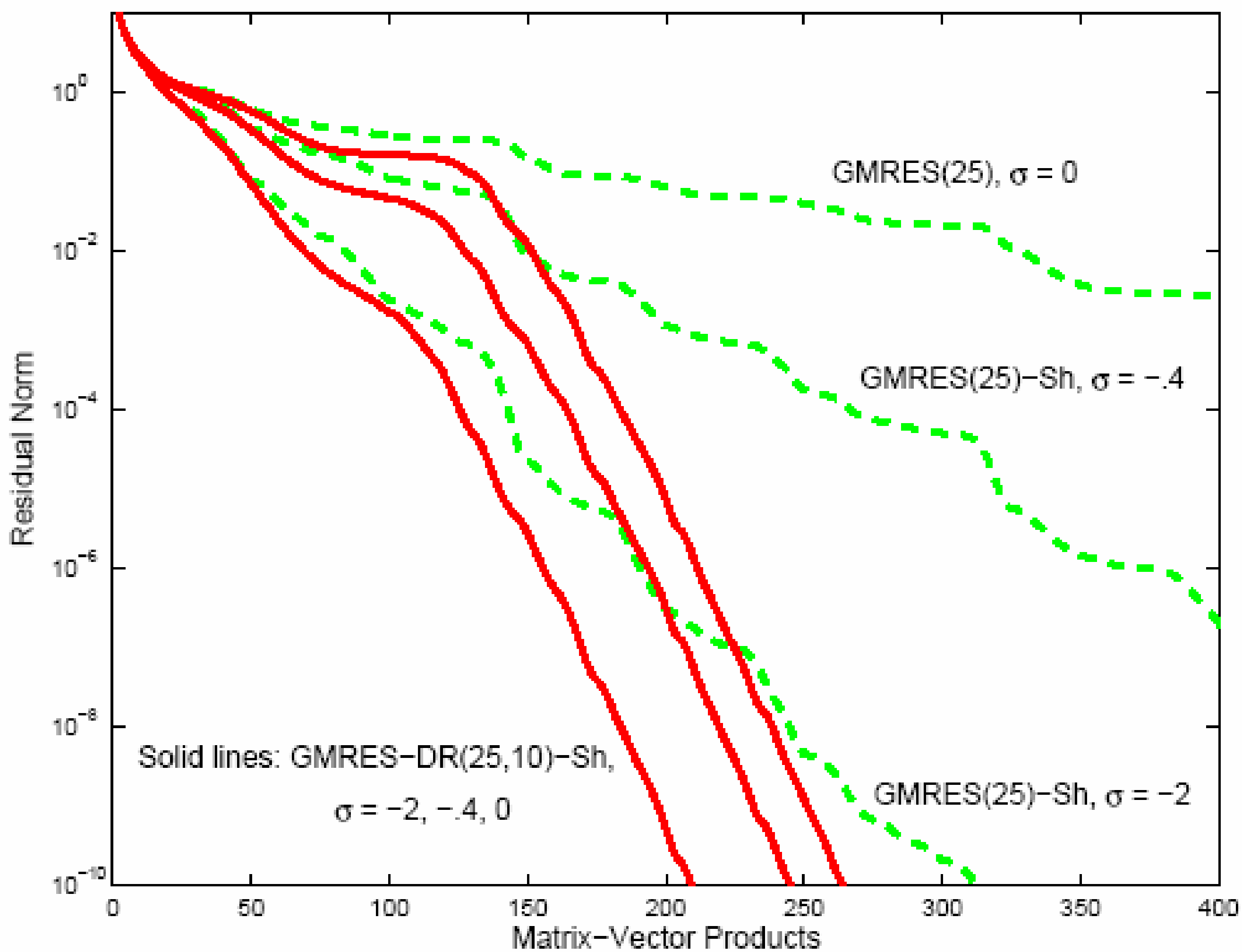
Restarting makes it more difficult.

Frommer and Glassner:

- Force residuals to all be parallel after a restart.
- Minimal residual property is maintained only for the base shift system.
- Can continue using one Krylov subspace for all shifted systems!

GMRES-DR for multiple shifts.

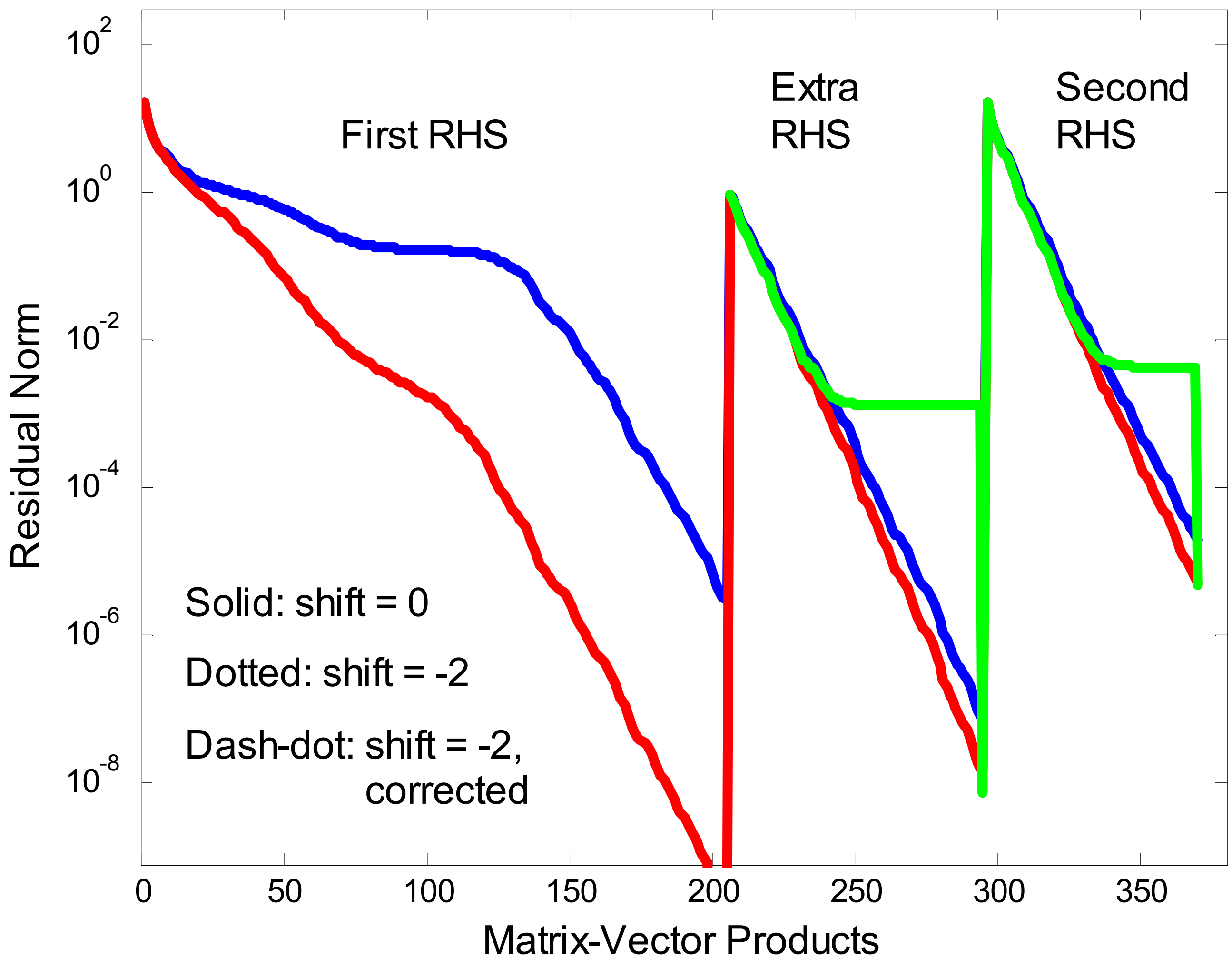
- Subspaces generated by GMRES-DR are combination of approximate eigenvectors portion and Krylov subspace portion, but remarkably when put together, they are Krylov themselves (with a different starting vector).
- So, GMRES-DR can be restarted like GMRES for multiple shifts.



GMRES(25), σ = 0
GMRES(25)−Sh, σ = −.4
GMRES(25)−Sh, σ = −2
Solid lines: GMRES−DR(25,10)−Sh,
σ = −2, −.4, 0
Residual Norm
Matrix−Vector Products
0
50
100
150
200
250
300
350
400
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}

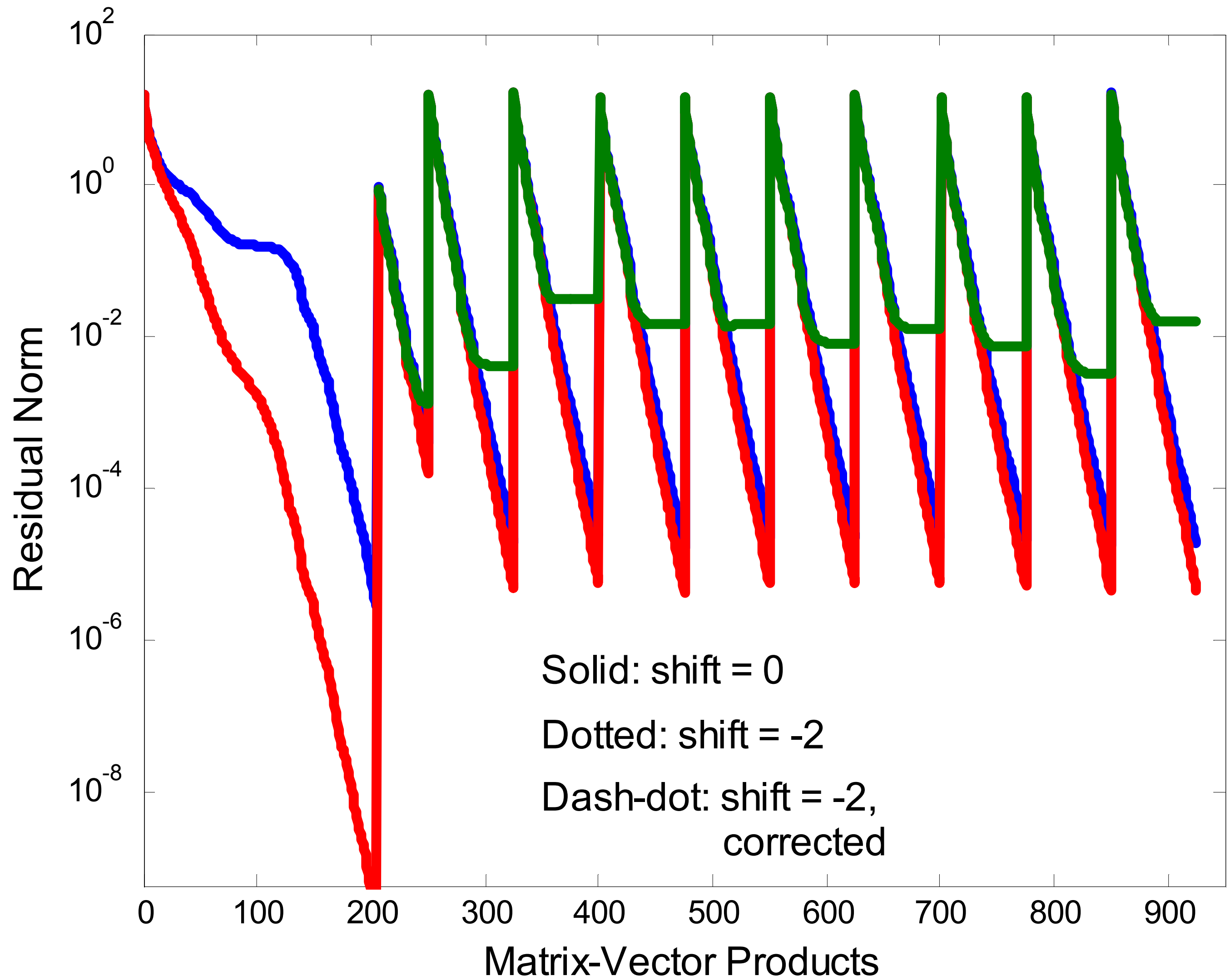
Multiple right-hand sides with multiple shifts and deflation of eigenvalues.

- Deflating eigenvalues messes up solution of multiple shifts for second and subsequent right-hand sides. Cannot keep residual vectors parallel unless have exact eigenvectors.
- Solution: force error to be in the direction of one vector,namely v_{k+1} from $AV_k = V_{k+1}\overline{H}_k$
- Then can correct error at the end.
- Need solution of one extra right-hand side.

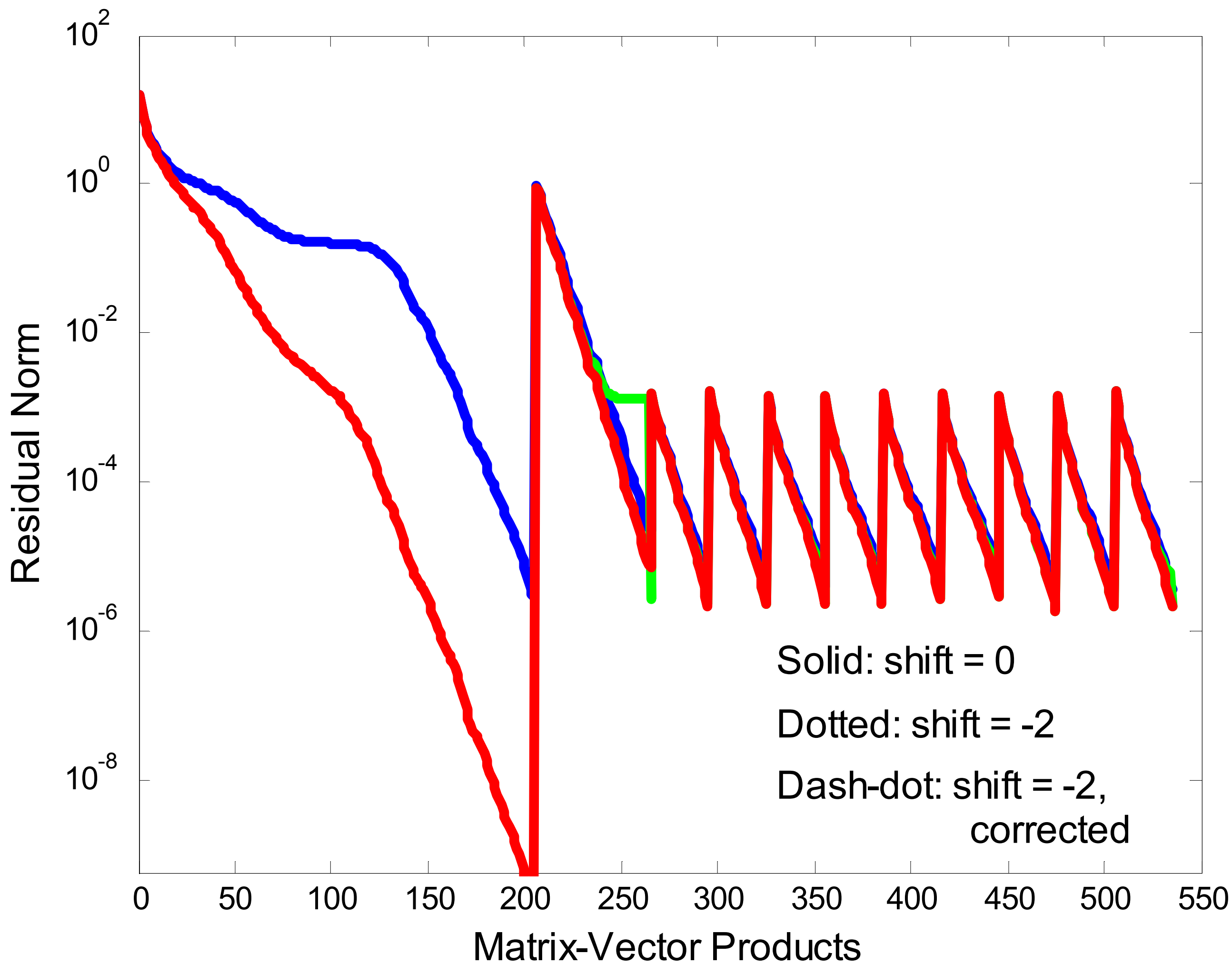


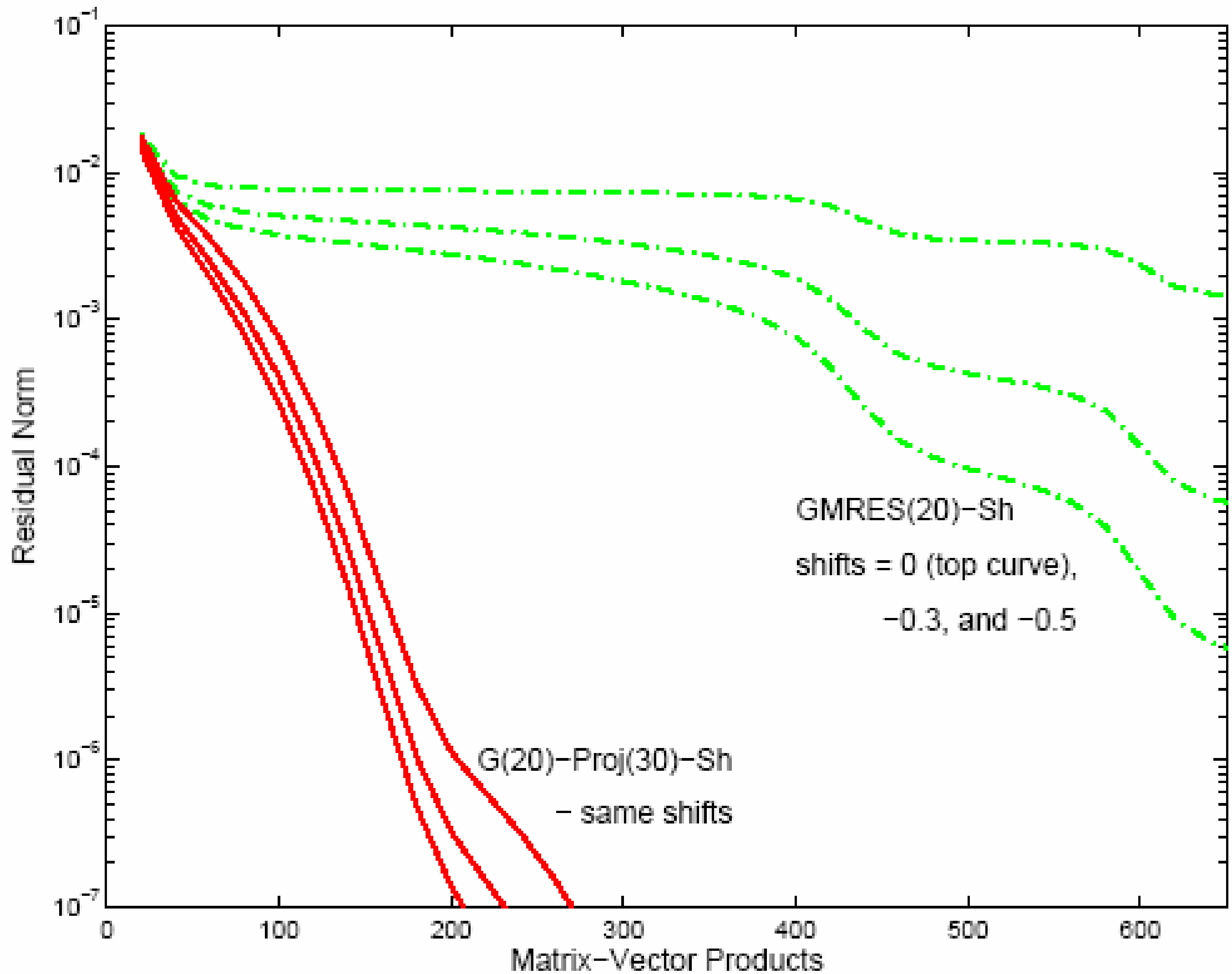
First RHS
Extra
RHS
Second
RHS
Solid: shift = 0
Dotted: shift = -2
Dash-dot: shift = -2,
corrected
Residual Norm
Matrix-Vector Products
10^{2}
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
0
50
100
150
200
250
300
350

Solution of ten right hand sides.



Example with right-hand sides are related to each other.





GMRES(20)−Sh
shifts = 0 (top curve),
−0.3, and −0.5
G(20)−Proj(30)−Sh
− same shifts
Residual Norm
Matrix−Vector Products

What if you don't like restarting, but still want to solve multiple right-hand sides?

- The non-restarted method BiCGStab is popular for QCD problems.
- But what about multiple right-hand sides? Can we still deflate eigenvalues?

Deflated BiCGStab for multiple right-hand sides

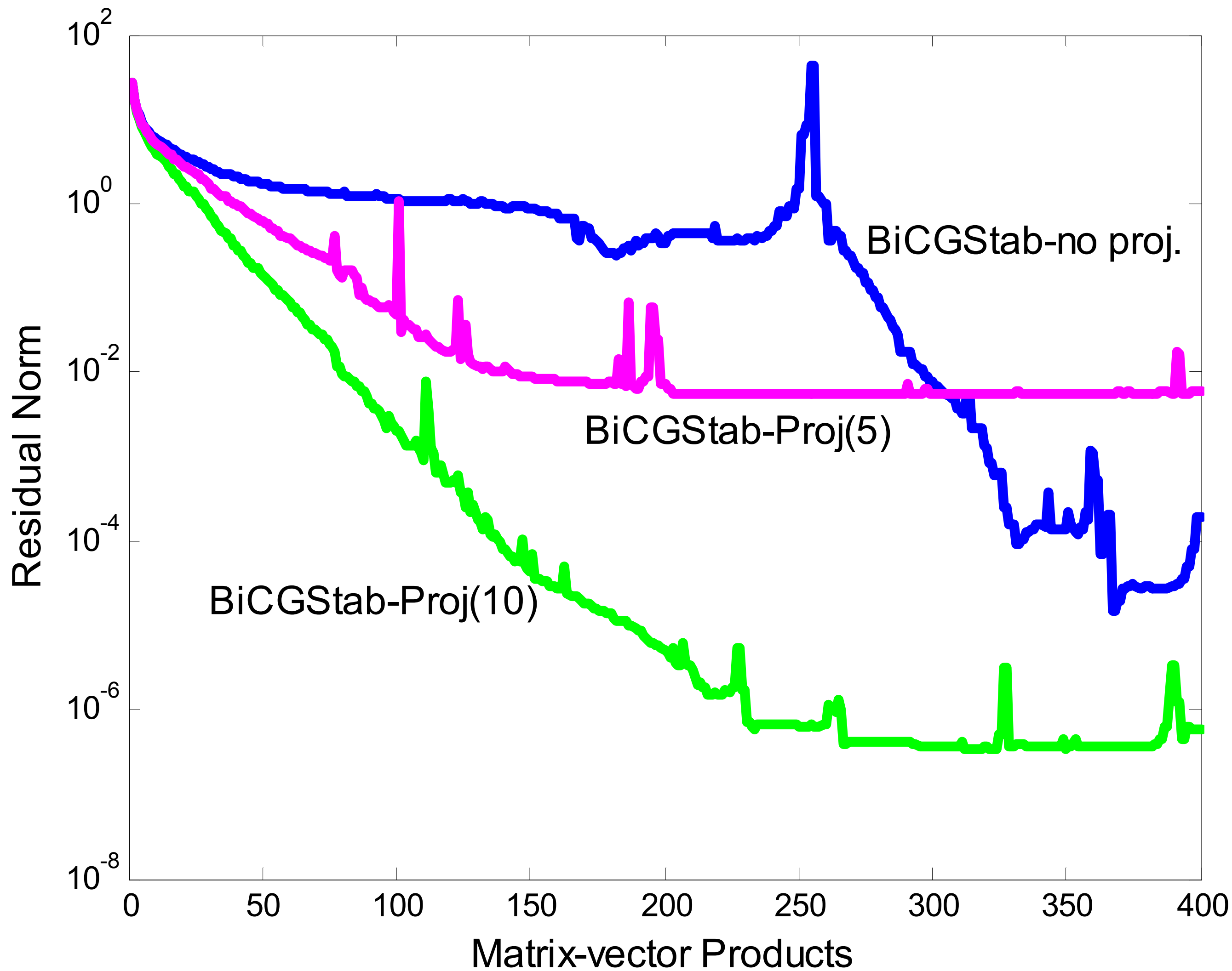
- First RHS: GMRES-DR
- Other RHS's:

  1) Project over eigenvectors (only once)

  2) Run BiCGStab

Preliminary Deflated BiCGStab for the 2nd rhs (following GMRES-DR for 1st rhs)

Matrix: bidiagonal, diagonal is 0.1, 1, 2, 3, …1999



Problem: Projection over eigenvectors is not good enough to last for the entire run of BiCGStab.

Solution: Use a projection over both right and left eigenvectors.

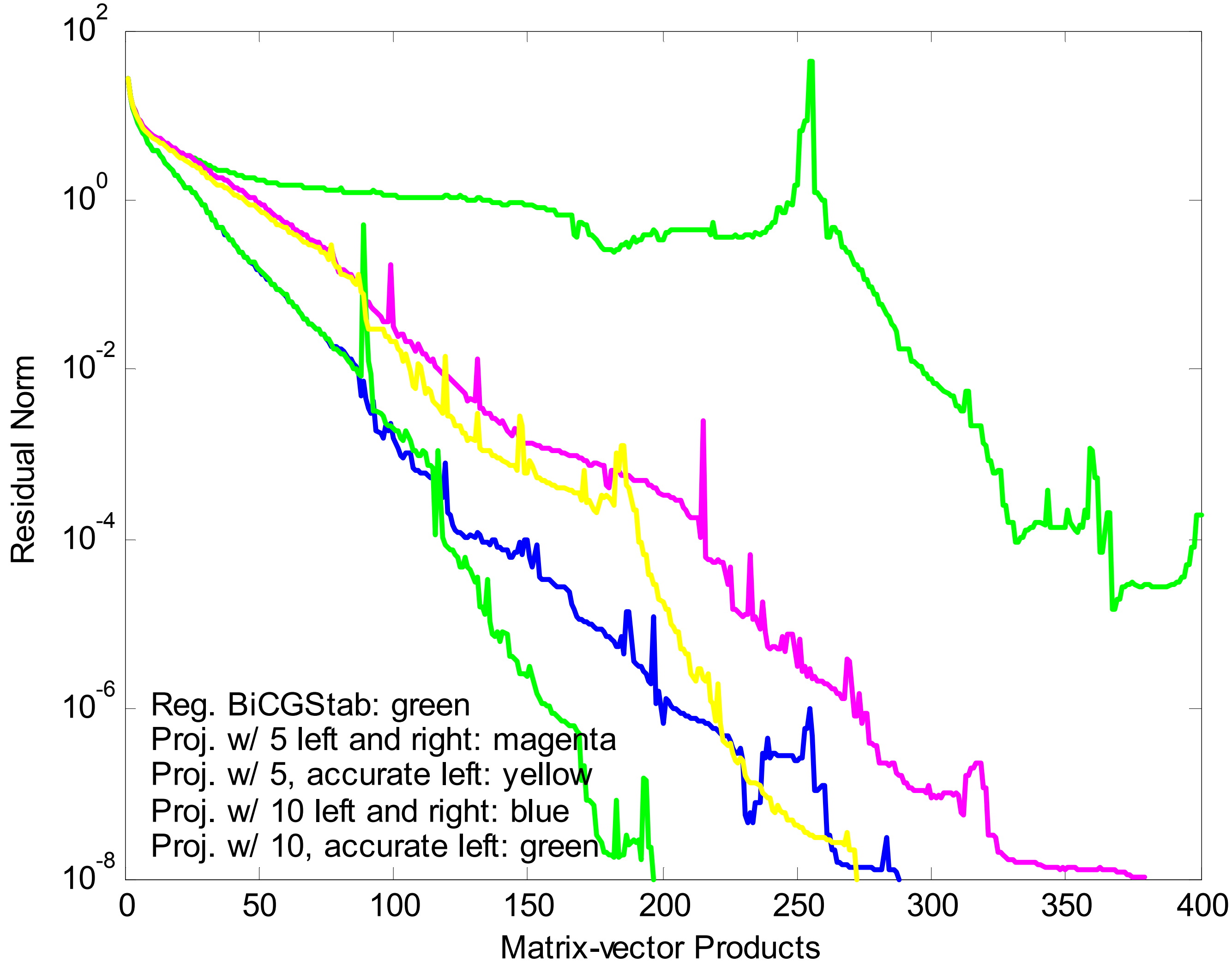
Deflated BiCGStab for the second and subsequent right-hand sides:

1) Project over right and left eigenvectors

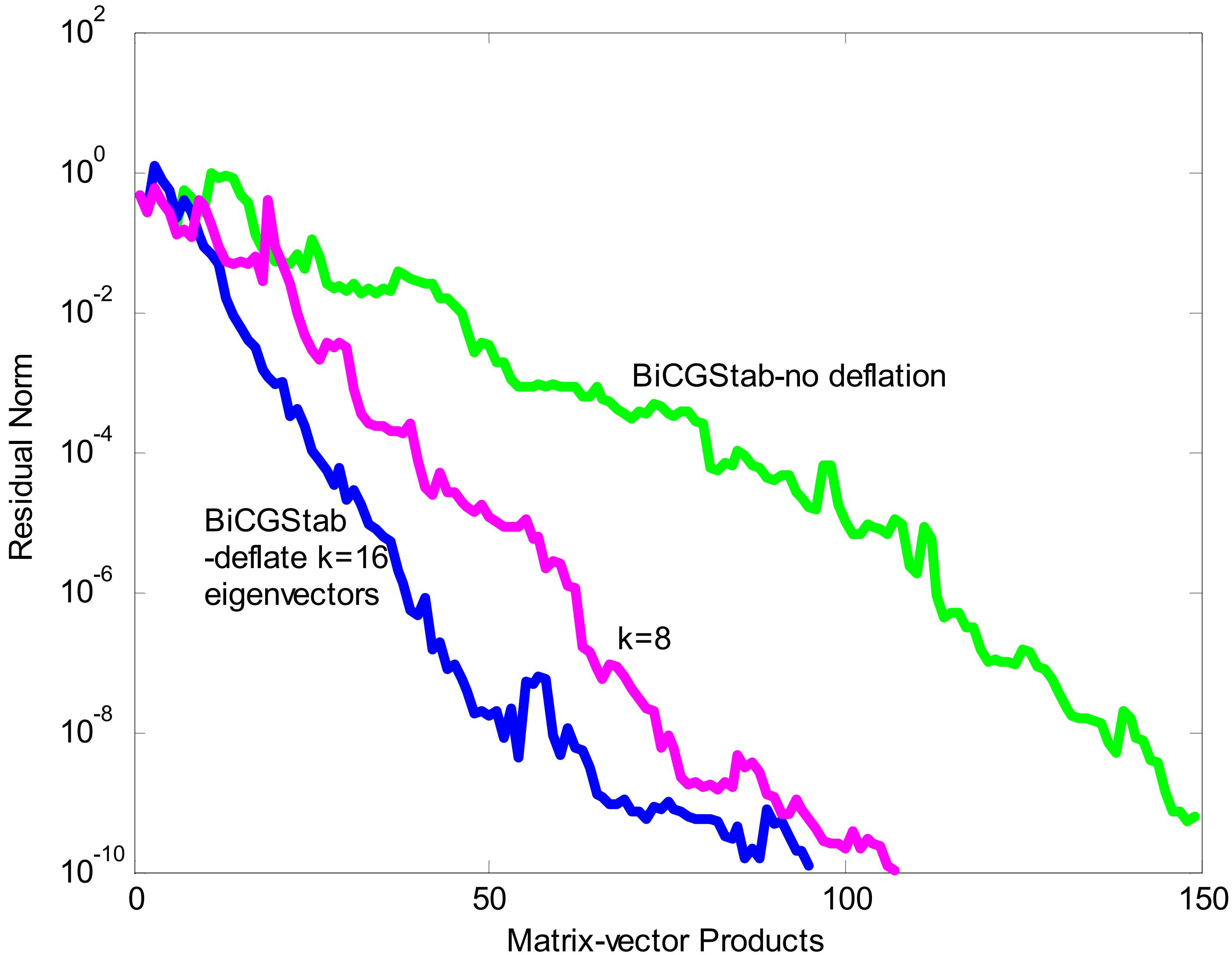
2) Run BiCGStab

Deflated BiCGStab for the 2nd rhs with Left-Right Projection

Matrix: bidiagonal, diagonal is 0.1, 1, 2, 3, …1999



Small QCD matrix. 4^4 lattice (n=1536). Deflated BiCGStab for the second rhs.



How do we get left eigenvectors?

For Wilson QCD matrix, they are for free:

Once you have right eigenvectors, can get left eigenvectors by multiplying by γ_5 matrix.

Computing both right and left eigenvectors

- We have seen you need left eigenvectors for Deflated BiCGStab.

- There are other applications that need left eigenvectors.

Options for computing right and left eigenvectors.

- Use ARPACK - twice.
- Use a Two-sided Arnoldi. Ruhe, Cullum, probably others.
- Nonsymmetric Lanczos
  - if storage is limited, may need to restart
    - see Baglama and Reichel and Kokiopoulou, Bekas and Gallopoulos for implicitly restarted Lanczos bidiagonalization.

Restarted nonsymmetric Lanczos

- For each cycle (except the first), use subspaces:

$$Span\{y_1, y_2, \ldots y_k, r, Ar, A^2 r, \ldots A^{m-k-1} r\} \text{ and}$$

$$Span\{u_1, u_2, \ldots u_k, s, As, A^2 s, \ldots A^{m-k-1} s\}$$

where y_i's are right Ritz vectors,

u_i's are left Ritz vectors,

r is a right eigenvalue residual vector, and

s is a left residual.

Restarted nonsymmetric Lanczos (cont.)

- Build biorthogonal bases for the two subspaces. There are three-term recurrences, after the k+2 vector.

  Let V have the right basis and U have the left basis.
- Find $T = U^T AV$. Form is not tridiagonal for first k+1 by k+1 block (but is sparse).
- From T compute right and left Ritz vectors and use them for the next cycle.

- Similar to the restarted Arnoldi algorithms (such as implicitly restarted Arnoldi), can prove:
  - Both right and left subspaces are Krylov (of course, one with A and the other with A^T).
  - Both contain smaller Krylov subspaces with each approximate eigenvector as starting vector.

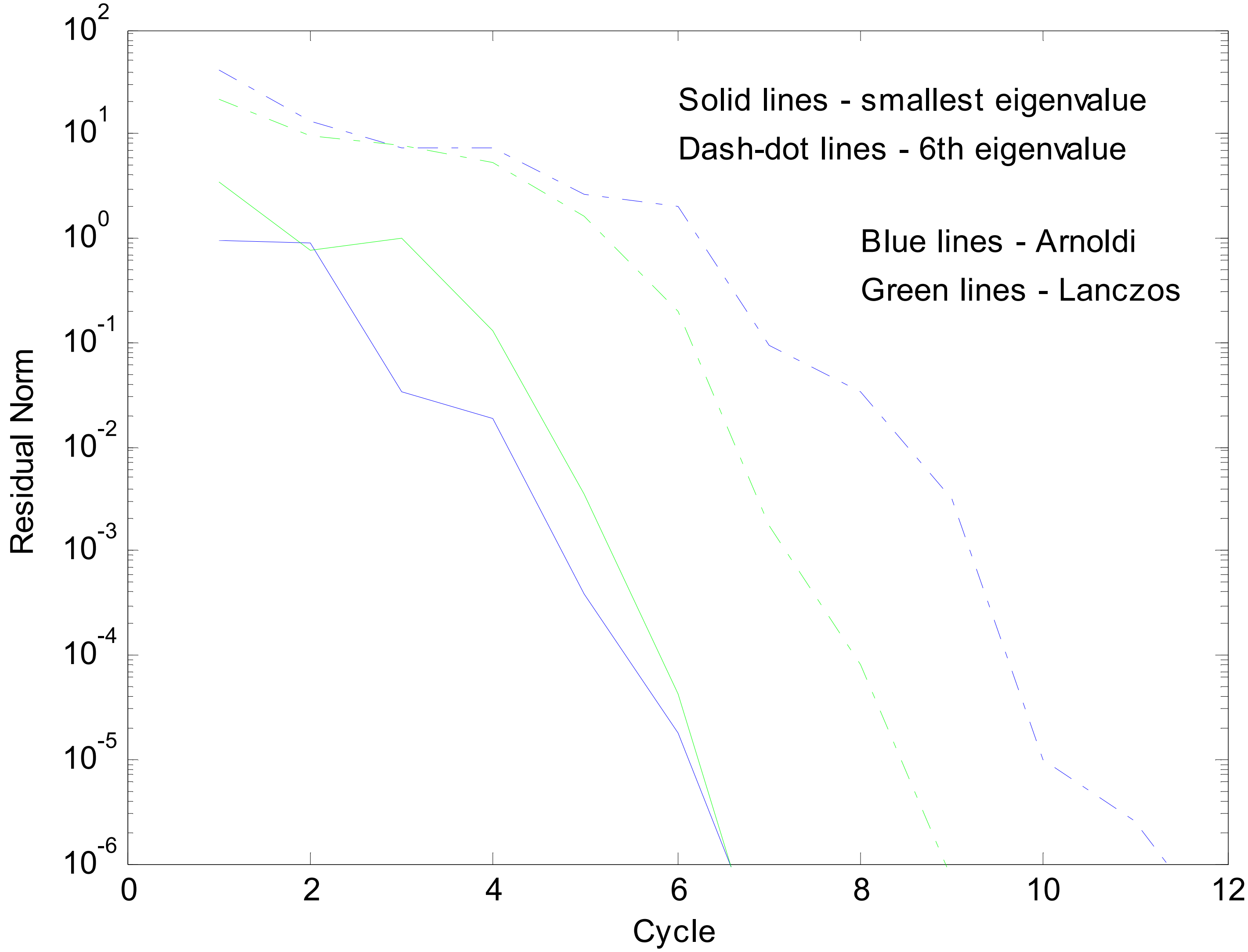
Roundoff error

Some Algorithm Choices:

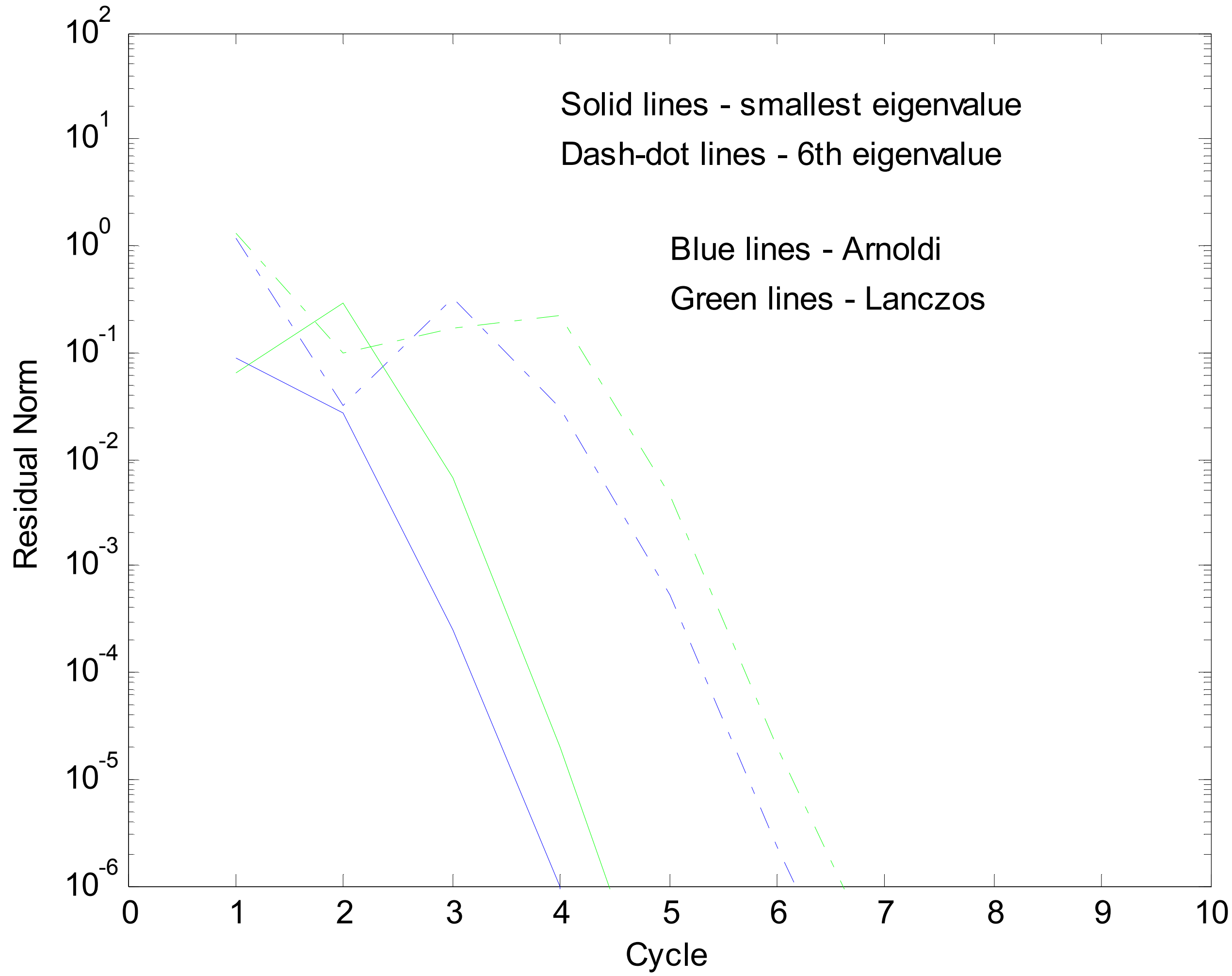
- No reorthogonalization – often unstable
- Partial rebiorthogonalization:
  - First cycle: fully rebiorthogonalize both right and left vectors
  - Other cycles: rebiorthog. only against the first k vectors (so only against the approximate eigenvector portion of the subspace)
- Full rebiorthogonalization

Restarted Lanczos vs. Arnoldi, use (40,10), so m=40 vectors in each subspace, including k=10 Ritz vectors

Matrix: bidiagonal, diagonal is 0.1, 1, 2, 3, …1999



Matrix Sherman4, n=1104, again use m=40, k=10



Cost comparison for Restarted Lanczos(40,10) and Rest. Arnoldi(40,10)

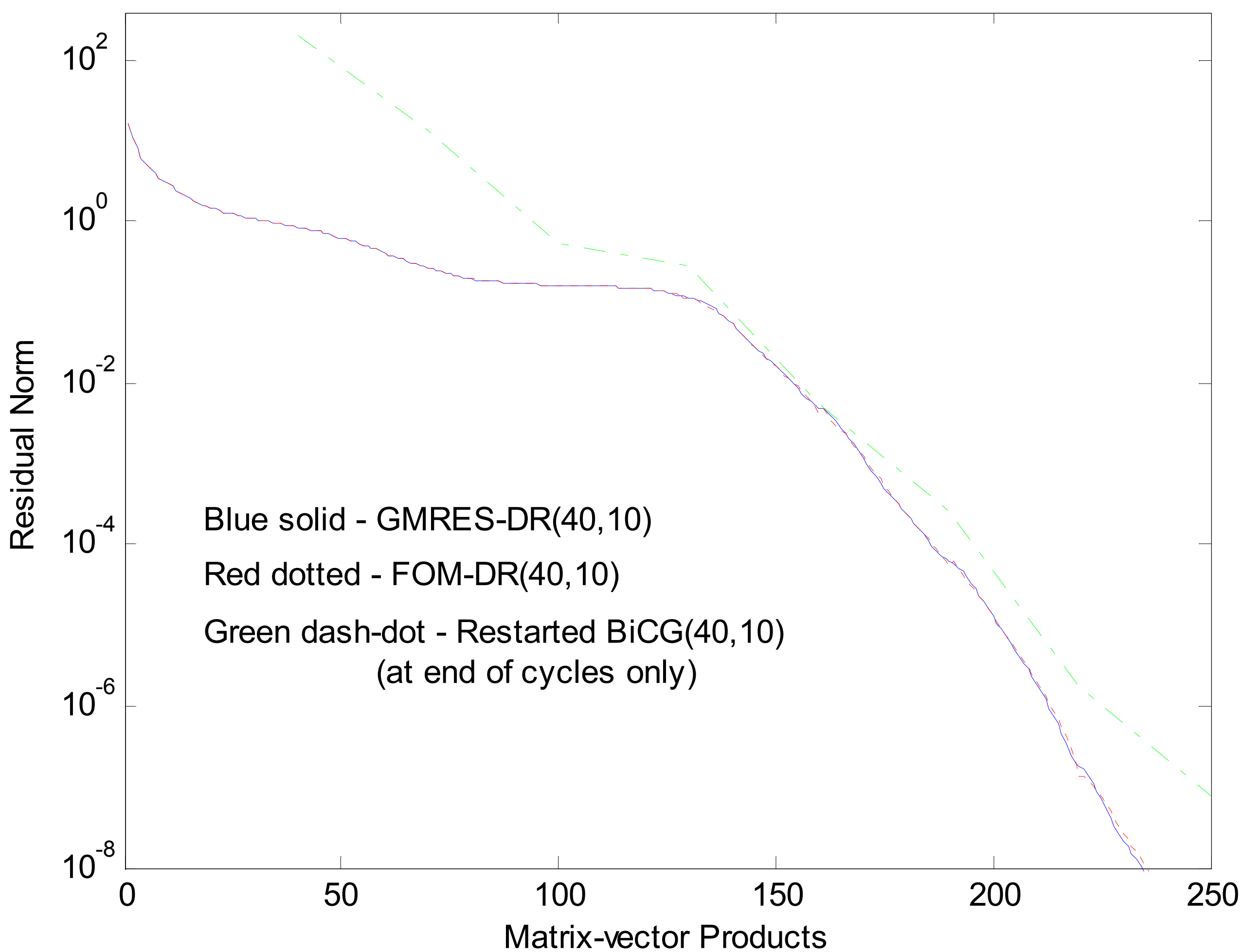
Cost per cycle (measured in vector op's of length n and matrix-vector products):

Lanczos – no reorthog:	1160 v.op's + 60 mvp's
Lancz. – part. reorthog:	1760 v.op's + 60 mvp's
Lancz. – full reorthog:	4160 v.op's+ 60 mvp's
Arnoldi – no reorthog:	1900 v.op's + 30 mvp's
Arnoldi – full reorthog:	3400 v.op's + 30 mvp's
Arnoldi twice – no reorth:	3800 v.op's + 60 mvp's
Arnoldi twice – full reorth:	6800 v.op's + 60 mvp's

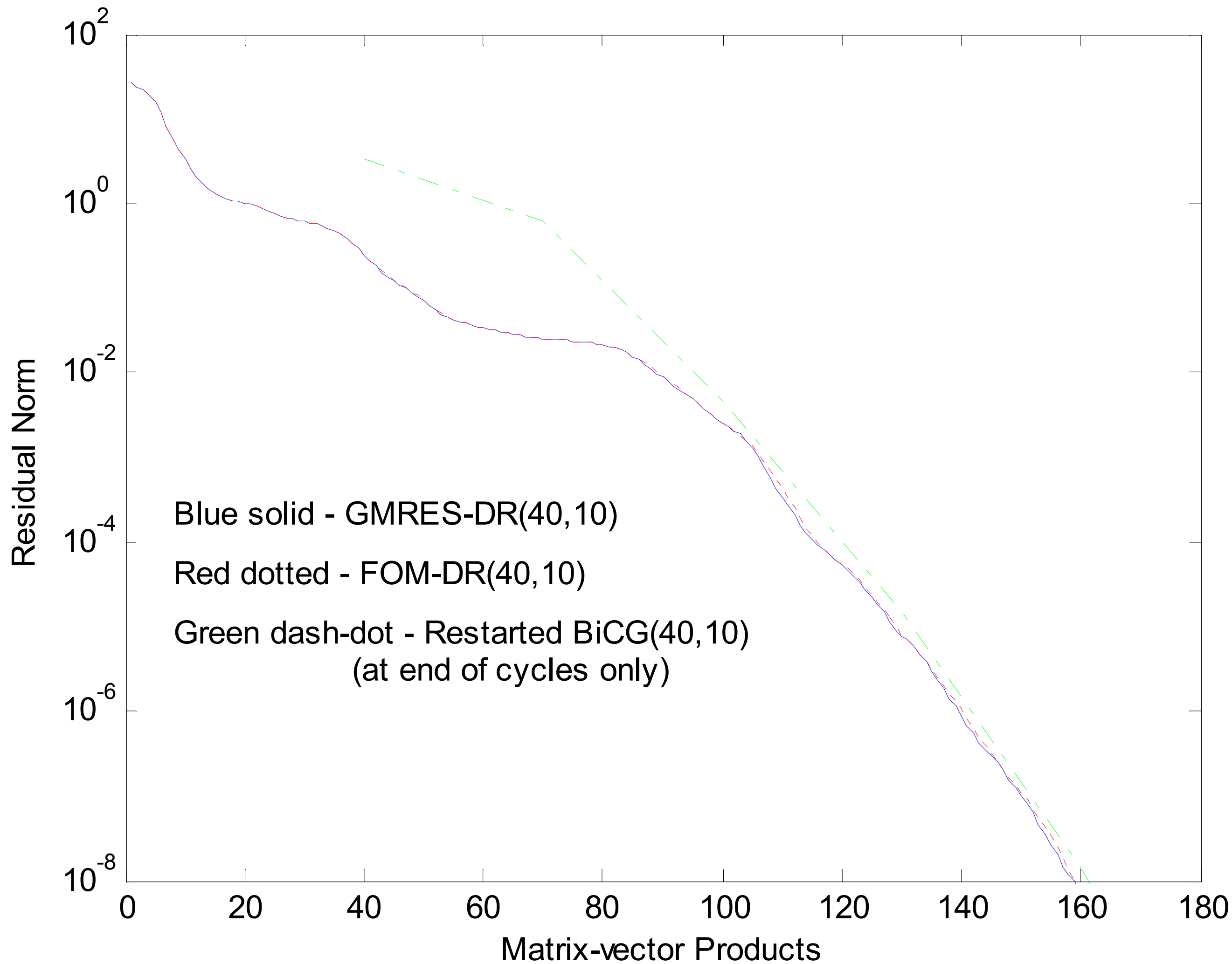
Back to Linear Equations

- New method: Restarted, Deflated BiCG
  - Can solve the linear equations simultaneous with the eigenvalue computations of Restarted Lanczos.
  - Since restart with approximate eigenvectors, convergence is much faster than with regular restarting.

Restarted BiCG vs. GMRES-DR and FOM-DR for the bidiagonal matrix



Restarted BiCG for Sherman4



Possible future projects

- Deflation for other QCD problems.
- Using left eigenvectors for deflated GMRES (seems to help for some small problems and it's easy for QCD).
- Using eigenvectors in a multigrid scheme for QCD problems.
- What if multiple shifts are changing? This happens for some twisted mass problems and also in model reduction, Sylvester's equations, etc.

Conclusion

- Deflated GMRES can be useful for matrices with small eigenvalues. Wilson matrices near kappa critical have small eigenvalues.
- Deflating eigenvalues is especially useful when there are multiple right-hand sides.
- With multiple shifts for multiple right-hand sides, deflating eigenvalues is still possible.
- BiCGStab can be deflated for multiple right-hand sides.