Greetings

Greetings from the Information Technology Laboratory of the National Institute of Standards and Technology in Gaithersburg, MD, USA.

The NIST/ITL Mathematical and Computational Sciences Division is delighted to have the opportunity to co-sponsor this conference with our Swiss colleagues. It was, of course, a US/Swiss, and indeed a NIST/ETH, collaboration of 50 years ago, that we are celebrating here this week. We are indeed proud of the role that our organizational ancestors in the Institute for Numerical Analysis played in bringing to light one of the most significant algorithms of the 20th century.

We continue to be inspired by the technical excellence and the spirit of cooperation that characterized the seminal work of Hestenes and Stiefel on the conjugate gradient method. The agenda for this meeting is ample evidence that the intellectual excitement kindled by that collaboration remains alive, and will carry us forward well into the 21st century.

Ron Boisvert

Toward Understanding the Convergence of Krylov Subspace Methods

Dianne P. O'Leary

Computer Science Dept. and
Institute for Advanced Computer Studies
University of Maryland
oleary@cs.umd.edu

http://www.cs.umd.edu/users/oleary

Work supported in part by the
National Science Foundation

2002: A Banner Year

- 50th Anniversary of the classic paper on the conjugate gradient (cg) algorithm
- 60th Anniversary (- 1) of Eduard Stiefel's habilitation degree (ETH)
- 70th Anniversary of Gene Golub's birth
  70th Anniversary of Magnus Hestenes' Ph.D. degree (University of Chicago)
- 100th Anniversary (+ 1) of the U.S. National Bureau of Standards, now called the National Institute of Standards and Technology, where Magnus Hestenes worked on cg
- 150th Anniversary (- 2) of the founding of ETH Zürich, where Eduard Stiefel worked on cg

2002: A Banner Year

- 50th Anniversary of the classic paper on the conjugate gradient (cg) algorithm
- 60th Anniversary (- 1) of Eduard Stiefel's habilitation degree (ETH)
- 70th Anniversary of Gene Golub's birth
  70th Anniversary of Magnus Hestenes' Ph.D. degree (University of Chicago)
- 100th Anniversary (+ 1) of the U.S. National Bureau of Standards, now called the National Institute of Standards and Technology, where Magnus Hestenes worked on cg
- 150th Anniversary (- 2) of the founding of ETH Zürich, where Eduard Stiefel worked on cg
- Second and last palindrome year we expect to see in our lifetimes.

The Plan

- Convergence of conjugate gradients
- Convergence of GMRES

Notation

- We solve the linear system

$$Ax^* = b$$

  where $A \in \mathcal{C}^{n\times n}$ and $b \in \mathcal{C}^n$.

- We normalize the problem so that $\|b\|_2 = 1$.
- We define the residual for the linear system by

$$r^{(m)} = b - Ax^{(m)}\,.$$

- We denote the Krylov subspace of dimension m by

$$\mathcal{K}_m(A,b) = \mathrm{span}\{b, Ab, \ldots, A^{m-1}b\}\,.$$

- We assume, without loss of generality, that our initial guess for the solution is

$$x^{(0)} = 0\,.$$

Convergence of Conjugate Gradients

The Conjugate Gradient Algorithm

Hestenes and Stiefel (1952) presented the conjugate gradient algorithm in the *Journal of Research of the NBS*.

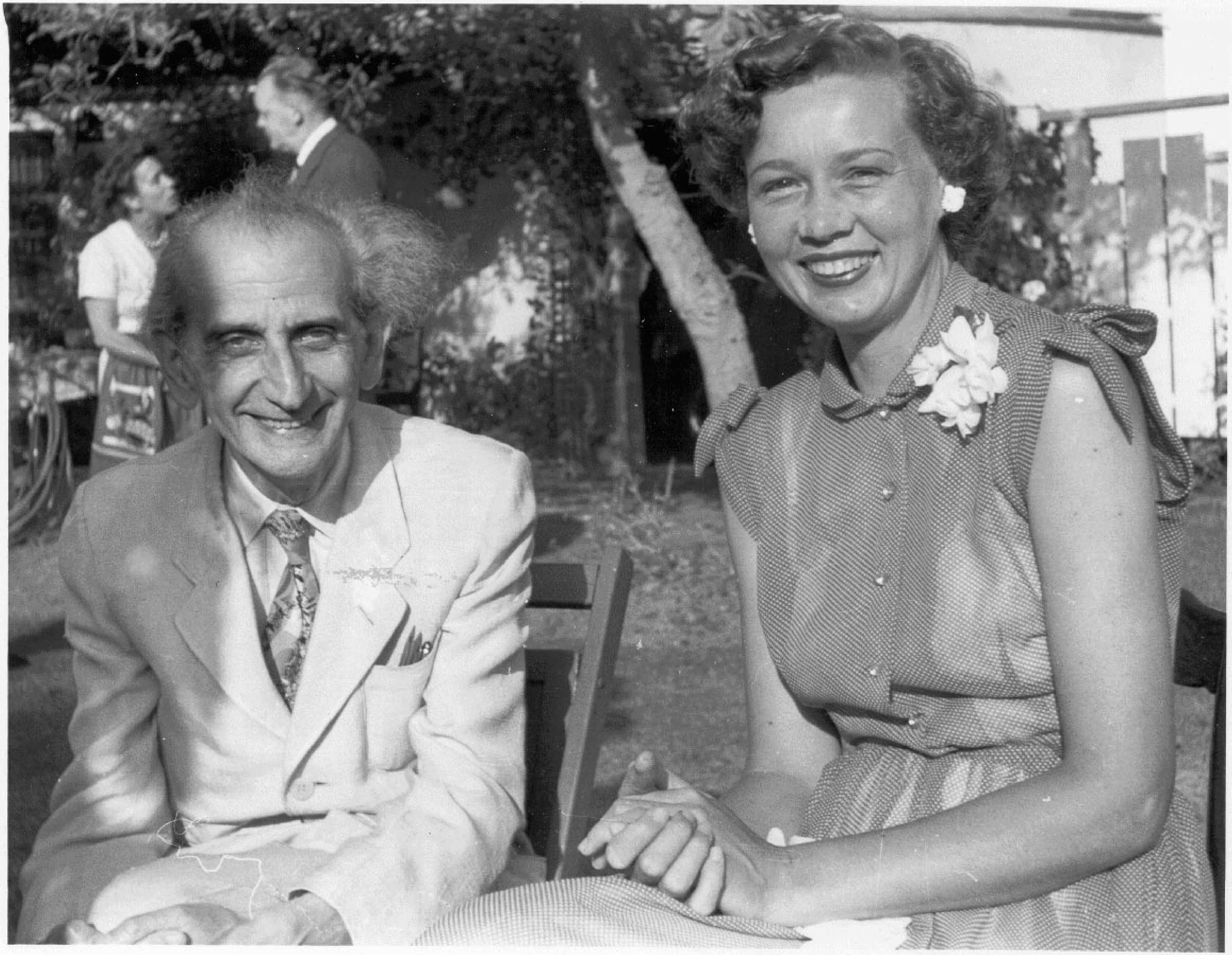
Magnus Hestenes (1906-1991), a faculty member at UCLA who became associated with the Institute for Numerical Analysis, part of NBS.

Eduard Stiefel (1909-1978), of ETH, a visitor to NBS.

Their account of how the paper came to be written

"The method of conjugate gradients was developed independently by E. Stiefel of the Institute of Applied Mathematics at Zurich and by M. R. Hestenes with the cooperation of J. B. Rosser, G. Forsythe, and L. Paige of the Institute for Numerical Analysis, National Bureau of Standards. The present account was prepared jointly by M. R. Hestenes and E. Stiefel during the latter's stay at the National Bureau of Standards. The first papers on this method were given by E. Stiefel [1952] and by M. R. Hestenes [1951]. Reports on this method were given by E. Stiefel and J. B. Rosser at a Symposium on August 23-25, 1951. Recently, C. Lanczos [1952] developed a closely related routine based on his earlier paper on eigenvalue problem [1950]. Examples and numerical tests of the method have been by R. Hayes, U. Hoschstrasser, and M. Stein."

E. Stiefel



Two distinct voices in the paper:

- Hestenes:
  - variational theory and optimal control
  - 1936: developed an algorithm for constructing conjugate bases, but advised by a Harvard professor that it was too obvious for publication
  - discouraging numerical experience by George Forsythe in using steepest descent for solving linear systems.
- Stiefel:
  - relaxation algorithms
  - continued fractions
  - qd algorithm

The Scope of the 1952 Paper

Assume that A is Hermitian positive definite.

- direct method: finite termination.
- use as iterative method: solves 106 “difference equations” in 90 iterations. (By 1958: 10x10 grid Laplace equation in 11 Chebyshev iterations + 2 cg.)
- monotonicity properties.
- round-off error analysis.
- smoothing initial residual.
- remedy for loss of orthogonality.
- solution if A is rank deficient.
- algebraic formulation of preconditioning.
- relation to Lanczos algorithm and continued fractions.

Recent Recognition of the Algorithm

- Science Citation Index lists over 800 citations between 1983 and 1999.
- NIST recently celebrated its centennial by picking its 100 most significant achievements. Among them:
  - ASCII
  - a highly-successful consumer information series
  - creation of Bose-Einstein condensation
  - the Conjugate Gradient Algorithm
  - Lanczos' eigenvalue algorithm
- *Computing in Science and Engineering*, a publication of the IEEE Computer Society and the American Institute of Physics, named Krylov Subspace Iteration as one of the Top 10 Algorithms of the 20th Century, citing in particular the pioneering work of Hestenes, Stiefel, and Lanczos.

Convergence Analysis of Conjugate Gradients

CG minimizes the error function

$$E(x^{(m)}) = (x^{(m)} - x^*)^H A (x^{(m)} - x^*)$$

over the Kylov subspace $\mathcal{K}_m(A, b)$.

Hayes (1954): Hilbert spaces

- linear convergence for general operators
- superlinear convergence for I $+$ completely continuous operator.

Kaniel(1966)-Daniel(1965) theory

$$E(x^{(m)}) \leq 4 \left(\frac{1 - \sqrt{\kappa^{-1}}}{1 + \sqrt{\kappa^{-1}}} \right)^{2m} E(x^{(0)})$$

where $\kappa = \lambda_{max}(A)/\lambda_{min}(A)$.

What is the worst case convergence for CG?

- The Kaniel-Daniel bound rests on the fact that $x^{(m)}$ is a polynomial in A times b, and therefore

$$E(x^{(m)}) = \min_{degree(p)<m} (p(A) - A^{-1})b^H A(p(A) - A^{-1})b\,.$$

- To provide an upper bound on $E(x^{(m)})$, they use scaled and shifted versions of the Chebyshev polynomials

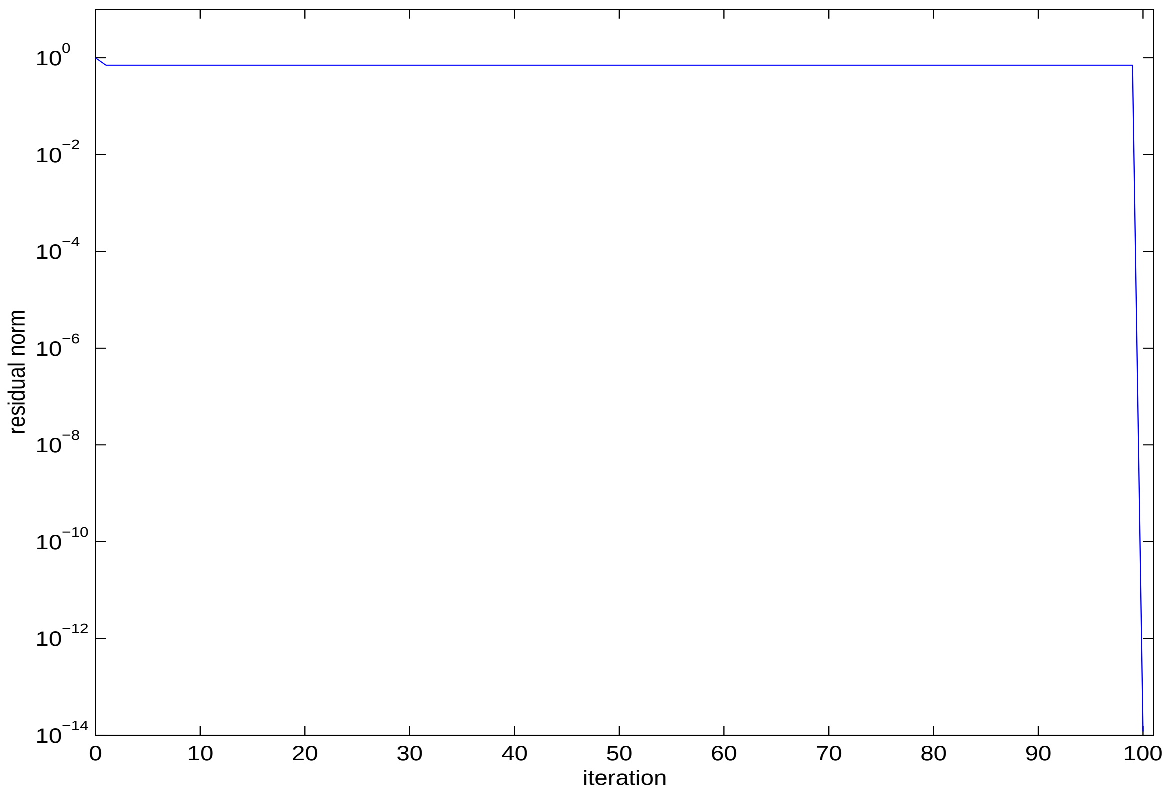
$$T_m(z) = \cos(m \arccos z)$$

in place of the minimization.

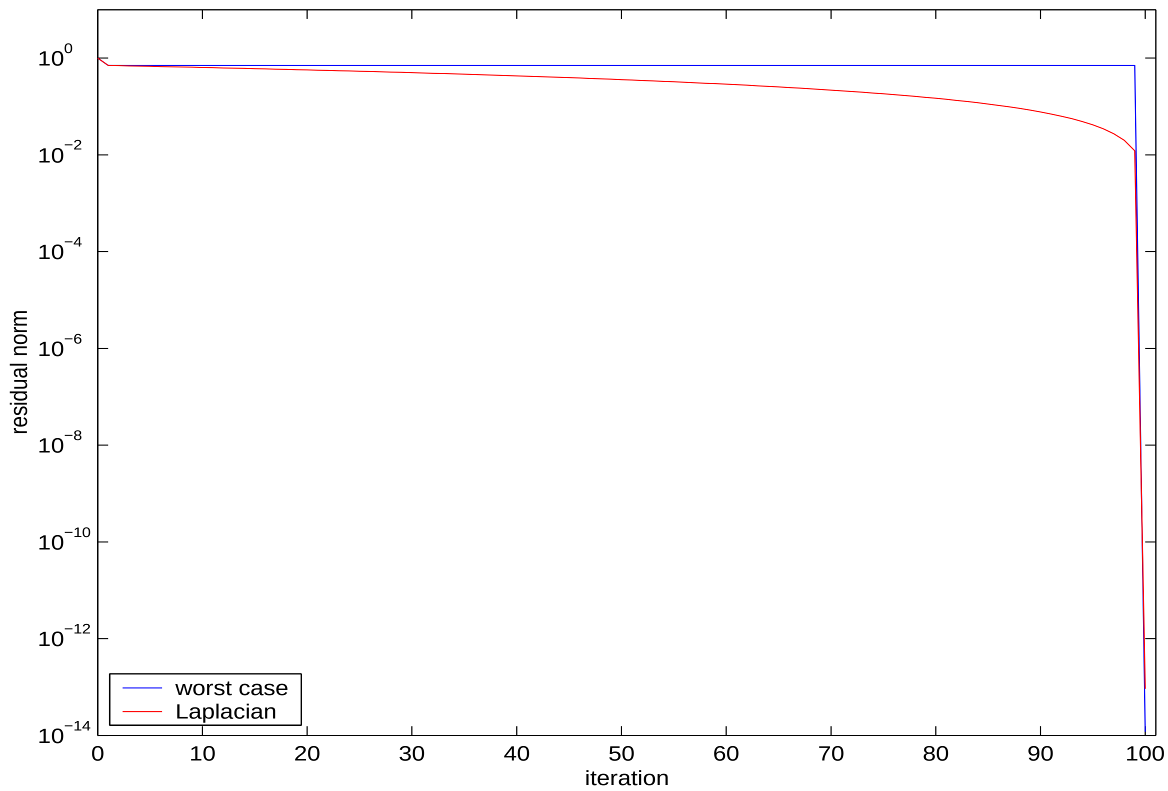
- Therefore, the worst case for conjugate gradient convergence is, for example, the $m \times m$ diagonal matrix that has eigenvalues equal to the roots of the scaled and shifted Chebyshev polynomials,

$$\lambda_j \approx 1 + \cos\left(\frac{(2j-1)\pi}{2m}\right), \quad j = 1, \ldots, m,$$

and a right hand side of ones.



Residuals for this worst-case problem



An unhappy coincidence

The eigenvalues of the 1-d Laplacian are, for large n, almost equal to these worst-case numbers, so convergence is similarly slow.

Conjugate Gradient Convergence

Summary:

- Good news: CG always makes some progress at each iteration.
- Bad news: The progress can be discouragingly small until the nth iteration.

Convergence of the GMRES Algorithm

The GMRES Algorithm

For GMRES, we drop the assumption that A is Hermitian positive definite. For simplicity, we'll assume that A is nonsingular.

GMRES (Saad, Schultz, 1986) minimizes the error function

$$(x^{(m)} - x^*)^H A^H A (x^{(m)} - x^*)$$

over the Kylov subspace $\mathcal{K}_m(A, b)$.

Convergence Analysis of GMRES

Convergence bound:

$$\frac{\|r_m\|}{\|r_0\|} \leq \min_{p_m(0)=1} \|V p_m(\Lambda) V^{-1} b\| \leq \kappa(V) \min_{p_m(0)=1} \max_{\lambda_j} |p_m(\lambda_j)|,$$

where $\kappa(V)$ is the condition number of the matrix of eigenvectors of A and p_m is a polynomial of degree m.

If A is Hermitian, or, more generally, normal, then V is an orthogonal matrix with $\kappa(V) = 1$, so the convergence analysis is related to the cg case.

In general, ill-conditioning of V can have a negative impact on convergence.

When A is Hermitian or real symmetric, GMRES is equivalent to MINRES and is guaranteed to make progress at each step.

Some Clues to understanding GMRES convergence

- Any monotonically nonincreasing curve that goes to zero is the convergence curve for GMRES applied to some problem, with arbitrary eigenvalues. (Greenbaum, Strakos, (and Ptak) 1994, 1996).
- Convergence bounds can be derived from the field of values of a matrix A:

$$\frac{\|r^{(m)}\|}{\|r^{(0)}\|} \leq 2 \left(\frac{s}{|c|} \right)^m$$

when the field of values of A is contained in a disk centered at c with radius s (Eiermann 1993).

What is the worst case convergence for GMRES?

CG and MINRES are guaranteed to make progress, however minimal, at each iteration, because the new component of the Krylov subspace is never orthogonal to the gradient of the function minimized.

In GMRES, we do not have this nice property. In fact, examples are well-known in which GMRES completely stagnates, failing to make any progress for $n-1$ iterations:

$$x^{(0)} = x^{(1)} = \dots x^{(n-1)} = 0\,.$$

We want to understand stagnation better.

Stagnation of GMRES

Joint work with Ilya Zavorin and Howard Elman.

We study an oddity: partial stagnation, in which the GMRES iterates

$$x^{(0)} = x^{(1)} = \ldots = x^{(m)} = 0$$

If $m = n - 1$, then this is complete stagnation, and then $x^{(n)}$ will be the exact solution to the problem.

Characterizing Stagnation

Let the eigendecomposition of A be $A = V\Lambda V^{-1}$, and let $y = V^{-1}b$.

Characterizing Partial Stagnation

Theorem: GMRES $m-$stagnates if and only if y satisfies the stagnation system

$$Z_{m+1}^H \bar{Y} V^H V y = e_1 \,, \tag{1}$$

where Y is the diagonal matrix formed from the entries of y, $e_1 = [1, 0, \ldots, 0]^T \in \mathcal{R}^{m+1}$, and

$$Z_{m+1} = \begin{pmatrix} 1 & \lambda_1 & \ldots & \lambda_1^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \ldots & \lambda_n^m \end{pmatrix} = \left(e \quad \Lambda e \quad \ldots \quad \Lambda^m e \right),$$

where e is the vector of ones.

Proof: At the mth step, GMRES minimizes the residual over all vectors x in the span of the columns of

$$K_m = [b, Ab, \ldots, A^{m-1}b] .$$

This means that the resulting residual r_m is the projection of b onto the subspace orthogonal to the span of the columns of AK_m.

Therefore, GMRES stagnates at step m if and only if b is orthogonal to the columns of AK_m, or, equivalently, orthogonal to the last m columns of K_{m+1}:

$$K_{m+1}^H b = e_1$$

Ipsen noted that $K_{m+1} = VYZ_{m+1}$, and substituting this expression gives our stagnation system. []

Characterizing Complete Stagnation

If $m+1=n$, then Z_{m+1} is invertible, and we can rewrite the stagnation system

$$Z_{m+1}^H \bar{Y} V^H V y = e_1 \,.$$

Complete stagnation occurs iff $\bar{Y} V^H V y = u$

where u is a vector derived from the eigenvalues λ_j:

$$u_j = (-1)^{n+1} \mathrm{conj}\left(\prod_{\substack{k=1\\k\neq j}}^{n} \frac{\lambda_k}{\lambda_j - \lambda_k} \right), \tag{2}$$

Illustration for $n = 2$

In certain simple cases, for example $n = 2$, we can get closed-form expressions characterizing stagnation.

Let the eigenvalues of A be 1 and $\lambda_0 e^{i\theta}$.

Then A is completely stagnating for all eigenvector matrices V whose condition number is greater than or equal to some critical value $\kappa_{cx}(\lambda_0, \theta)$.

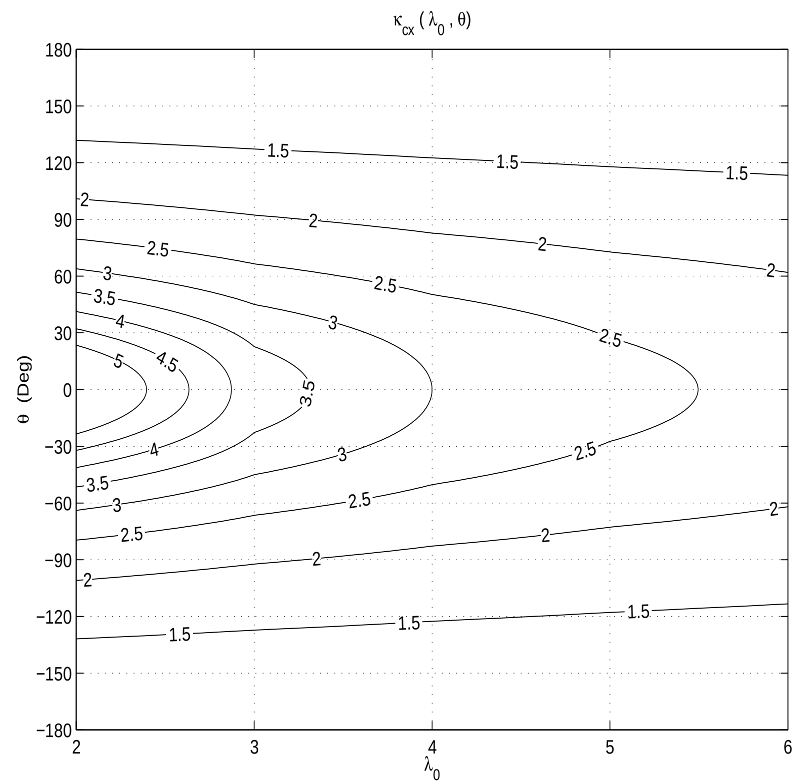


Illustration for $n = 2$: Contours of $\kappa_{cx}(\lambda_0, \theta)$

A Consequence of the Characterization

Complete stagnation for a matrix A implies complete stagnation for A^H:

GMRES completely stagnates for the problem $Ax = b$ if and only if it stagnates for $A^H x = \hat{b}$ where $b = Vy$, $\hat{y} = \bar{Y}^{-1}u$, and $\hat{b} = V^{-H}\hat{y}$.

Complete Stagnation of Normal Matrices

A normal matrix A is one whose eigenvector matrix V is unitary.

In this case, the stagnation system simplifies to

$$\overline{Y} y = u,$$

which is a system of n decoupled equations of the form,

$$|y_j|^2 = u_j, \quad j = 1, \dots, n.$$

Therefore, for normal matrices,

- GMRES stagnates for $b = Vy$, where

  $$y_j = \sqrt{u_j} e^{i\theta_j}, \quad j = 1, \dots, n,$$

  and the phase angles θ_j are arbitrary.
- If λ is such that the corresponding u contains complex or real negative entries, then there is no right-hand side for which GMRES stagnates.

Does Normal Stagnation Imply Non-Normal Stagnation?

Stagnation of a normal matrix does imply stagnation of an entire family of matrices with the same eigenvalues:

Theorem: Suppose we have a vector $\lambda \in \mathcal{C}^n$ with distinct elements such that $u \in \mathcal{R}^n$ satisfies $0 < u_j \leq 1$. Then for any nonsingular eigenvector matrix V with $W = V^H V$ real, GMRES stagnates for $A = V \Lambda V^{-1}$ and $b = Vy$, where $y \in \mathcal{R}^n$ satisfies $YWy = u$.

Proof: If W is real, then it is symmetric positive definite. Solving the stagnation equation $YWy = u$ is equivalent to finding a diagonal scaling matrix Y so that YWY has row sums u. Since $0 < u_j \leq 1$, then a theorem of Marshall and Olkin tells us that such a scaling matrix exists. $\Box$

Constructing Stagnating Eigenvalue Distributions

GMRES will stagnate for any eigenvector matrix satisfying $V^H V$ real, if

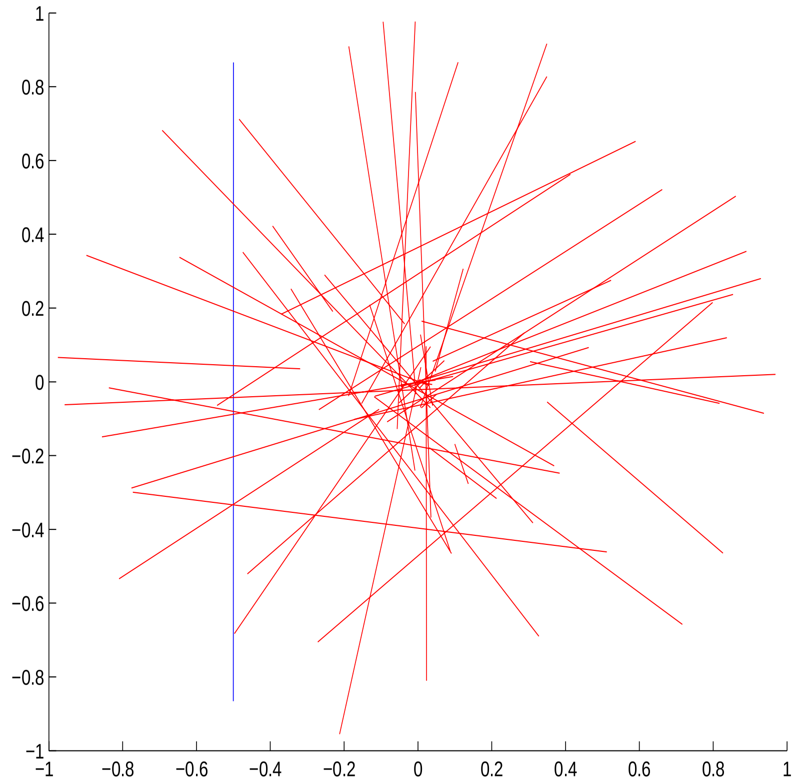
$$0 < u_j = (-1)^{n+1} \mathrm{conj} \left(\prod_{\substack{k=1 \\ k \neq j}}^{n} \frac{\lambda_k}{\lambda_j - \lambda_k} \right) \leq 1$$

Therefore, we can study such eigenvalue distributions by solving the polynomial system

$$\mathrm{conj} \left(\prod_{\substack{k=1 \\ k \neq j}}^{n} (\lambda_j - \lambda_k) \right) u_j = (-1)^{n+1} \mathrm{conj} \left(\prod_{\substack{k=1 \\ k \neq j}}^{n} \lambda_k \right)$$

for choices of $u_j \in [0, 1]$.

An observation: If the elements of λ solve this system, then so do the elements of $c\lambda$ for any nonzero scalar c. Therefore, every complex number can be an eigenvalue for a completely stagnating matrix.



Some Stagnating Eigenvalue Distributions for $n = 3$, $V^H V$ Real

Complete Stagnation for Unitary Matrices

A normal matrix A is unitary iff its eigenvalues satisfy

$$\lambda_j = e^{i\phi_j}, \quad 0 \le \phi_j \le 2\pi, \quad j = 1, \ldots, n.$$

Nachtigal, Reddy, and Trefethen (1992) showed that GMRES can completely stagnate when applied to a unitary matrix A with eigenvalues distributed uniformly over the unit circle in the complex plane.

The converse can be established using the stagnation system.

Theorem: These are the only unitary matrices for which complete stagnation can occur.

Complete Stagnation of Real Matrices

When A is real, the stagnation system can be written as a polynomial system in y, considerably simplifying analysis and numerical experimentation:

$$YPV^TVy = u$$

where P is a permutation matrix that depends on the ordering of eigenvalues.

It is possible to construct real matrices A that never completely stagnate on real right-hand sides but do completely stagnate on some complex right-hand sides.

Example: The matrix with eigenvectors

$$V = \begin{pmatrix} -0.3998204 & 0.2414875 & -0.0877858 & -0.4306034 \\ -0.5786559 & -0.8362391 & 0.4920379 & 0.3213318 \\ 0.6984230 & 0.0537175 & -0.7499413 & 0.5155494 \\ -0.1323115 & 0.4893898 & -0.4333364 & -0.6674844 \end{pmatrix},$$

and eigenvalues

$$\lambda = (1.0000000, -0.7658066, -0.2656295, 0.8705277).$$

stagnates for

$$y = \begin{bmatrix} 1.5564116 + 1.5564116\ i \\ -1.2084570 - 0.3414864\ i \\ 0.7066397 + 1.5089330\ i \\ -1.8679775 - 1.2644748\ i \end{bmatrix}$$

and 15 other right-hand sides, none of them real.

Conclusions

- The convergence of conjugate gradients is quite well understood, so attention focuses on the development of preconditioners to accelerate convergence.
- A comprehensive understanding of the convergence of GMRES and its relatives remains surprisingly illusive and is an additional obstacle to the development of preconditioners.
- By studying the limiting cases – stagnation – we can gain insight into factors that slow convergence.
- In particular, since restarted GMRES often nearly stagnates, we hope to develop better restart strategies.