Ron Boisvert

remains alive, and will carry us forward well into the 21st century.

evidence that the intellectual excitement kindled by that collaboration
the Conjugate Gradient method. The agenda for this meeting is ample

collaboration that characterized the seminal work of Hestenes and Stiefel on
continue to be inspired by the technical excellence and the spirit of

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

Institute of Standards and Technology in Gaithersburg, MD, USA.

Greetings
Work supported in part by the National Science Foundation

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

oleary@cs.umd.edu

University of Maryland

Institute for Advanced Computer Studies

Computer Science Dept. and

Dianne P. O'Leary

Methods of Krylov Subspace Methods

Toward Understanding the Convergence
Stiefel worked on Cg

150th Anniversary (– 2) of the founding of ETH Zurich, where Edvard
Magnus Hestenes worked on Cg.

called the National Institute of Standards and Technology, where
100th Anniversary (+ 1) of the U.S. National Bureau of Standards, now

101st Anniv. Chicago (Chicagö)

70th Anniversary of Magnus Hestenes' Ph.D. degree (University of
70th Anniversary of Gene Golub’s birth

60th Anniversary (– 1) of Edvard Stiefel’s habilitation degree (ETH)

50th Anniversary of the classic paper on the conjugate gradient (CG)

2002: A Banne Year
Second and last palindrome year we expect to see in our lifetimes.

Stiefel worked on CG

150th Anniversary (-2) of the founding of ETH Zurich, where Eduard

Magnus Hestenes worked on CG
called the National Institute of Standards and Technology, where

100th Anniversary (+1) of the U.S. National Bureau of Standards, now

100th Anniversary (Chicago)

70th Anniversary of Magnus Hestenes, Ph.D. degree (University of

70th Anniversary of Gene Golub, birth

60th Anniversary (-1) of Eduard Stiefel, habilitation degree (ETH)

Algorithm

50th Anniversary of the classic paper on the conjugate gradient (CG)

2002: A Banner Year
Convergence of GMRs

Convergence of conjugate gradients

The Plan
\[ 0 = (0)^x \]

Solution is

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

\[ \{ q_{1+m}V, \ldots, qV, q_0 \} \text{span} = (q, V)^u X \]

We denote the Kronecker subspace of dimension \( u \) by

\[ (u)^x V - q = (u)^d \]

We define the residual for the linear system by

\[ 1 = \|q\| \text{ so that } I. \]

We normalize the problem so that

where \( A \in \mathbb{R}^{u \times u} \) and \( q \in \mathbb{R}^u \), and

\[ q = \star \]

We solve the linear system

Notation
Convergence of Conjugate Gradients
Eduard Stiefel (1909-1978), of ETH, a visitor to NBS.

associated with the Institute for Numerical Analysis, part of NBS.

Magnus Hestenes (1906-1991), a faculty member at UCLA who became

the Journal of Research of the NBS.

Hestenes and Stiefel (1952) presented the Conjugate Gradient Algorithm in

The Conjugate Gradient Algorithm
have been by R. Hayes, U. Houshtrassser, and M. Stein."

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

Their account of how the paper came to be written
— 4th algorithm
— continued fractions
— relaxation algorithms

Stiefel

steepest descent for solving linear systems
— discouraging numerical experience by George Forsythe in using
— advised by a Harvard professor that it was too obvious for publication
— 1936: developed an algorithm for constructing conjugate bases, but
— variational theory and optimal control

Hestenes

Two distinct voices in the paper
Relation to Lanczos algorithm and continued fractions. 

Algebraic formulation of preconditioning. 

Solution if $A$ is rank deficient. 

Remedy for loss of orthogonality. 

Smoothing initial residual. 

Round-off error analysis. 

Monotonicity properties. 

$\text{iterations} + 2 \geq \text{(iterations)}$. 

By 1958: 10x10 grid Laplace equation in II Chebyshev 

use as iterative method: solves 106 difference equations in 90 

direct method: finite termination. 

Assume that $A$ is Hermitian positive definite. 

The Scope of the 1Q52 Paper
citig in particular the pioneering work of Hestenes, Stiefel, and Lanczos. Subspace Iteration as one of the Top 10 Algorithms of the 20th Century, Computer Society and the American Institute of Physics, named Krylov Computing in Science and Engineering, a publication of the IEEE. • Lanczos’ eigenvalue algorithm – the Conjugate Gradient Algorithm – creation of Bose-Einstein condensation – a highly successful consumer Information Series – ASCII • Among them: 100 most significant achievements. Among them: recently celebrated its centennial by picking its NIST Science Citation Index lists over 800 citations between 1983 and 1999. • Recent Recognition of the Algorithm
\[ (\forall x) H \left( \frac{I - \gamma \wedge + 1}{1 - \gamma \wedge - 1} \right) \forall \geq (w)x) H \]

where \( H \)

Kannel (1966)-Daniell (1965) theory

Superlinear convergence for \( I \) completely continuous operator.

Linear convergence for general operators.

Hayes (1954): Hilbert spaces

\[ (q, A) \wedge \forall \wedge \forall \wedge \forall \wedge \forall \]

over the Kylov subspace \( A \)

\[ (x - (w)x) A H (x - (w)x) = (w)x H \]

CG minimizes the error function

Convergence Analyses of Conjugate Gradients
and a right hand side of ones.

\[ w, \ldots, w' = f \left( \frac{w z}{\nu(1 - f^2 z)} \right) \cos \approx f \nu \]

Roots of the scaled and shifted Chebyshev polynomials, for example, the \( m \times m \) diagonal matrix that has eigenvalues equal to the

Therefore, the worst case for conjugate gradient convergence is, for

In place of the minimization,

\[ (z \cos \text{arccos } w = (z)^w \mathcal{I} \]

versions of the Chebyshev polynomials

To provide an upper bound on \( \langle w, x \rangle \mathcal{H} \), they use scaled and shifted

\[ q(1 - \nu (\nu d)^H q(1 - \nu (\nu d)^H (\mu > (d) \text{element})) = (\mu, x) \mathcal{H} \]

\( \forall \) times \( q \), and therefore

\( \forall \) times the Kaniel-Daniil bounds rest on the fact that \( x \) is a polynomial in

What is the worst case convergence for CG?
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.
The progress can be discouragingly small until the 17th iteration.  

**Bad news:** The progress makes some progress at each iteration.  

**Good news:** The progress makes some progress at each iteration.  

**Summary:** Conjugate Gradient Convergence
Convergence of the GMRES Algorithm
over the Krylov subspace $K_n(q, \mathcal{V})$.

$$(x - (w)x) V_H V_H^T (x - (w)x)$$

CMRES (Saad, Schultz, 1986) minimizes the error function for simplicity, we'll assume that $A$ is nonsingular. For CMRES, we drop the assumption that $A$ is Hermitian positive definite.

The CMRES Algorithm
and is guaranteed to make progress at each step.

When \( A \) is Hermitian or real symmetric, GMRES is equivalent to MINRES.

In general, \( \| \cdot \| \)-conditioning of \( A \) can have a negative impact on convergence.

Matrix with \( A \) in the corresponding analysis is related to the G case.

If \( A \) is Hermitian or, more generally, normal, then \( A \) is an orthogonal.

\( w \) is a polynomial of degree \( m \).

where \( \Lambda \) is the condition number of the matrix of eigenvectors of \( A \) and

\[
\| (f^T)^{1/2} \text{max}_{1=1}^{\infty} (\Lambda)^{1/m} \| \Lambda \| \geq \| qI - \Lambda(V)^{wd} \Lambda \|^{1/2} \geq \frac{\| 0, e \|}{\| w, l \|}
\]

Convergence bound:

**Convergence Analysis of GMRES**
radius $s$ (Eiermann 1993).

When the field of values of $A$ is contained in a disk centered at $c$ with
\[
\left(\frac{c}{s}\right)^2 \geq \frac{\|A\|}{\|A_{h}\|}
\]

Convergence bounds can be derived from the field of values of a matrix.


Convergence curve for GMRES applied to some problem, with arbitrary
convergence.

Any monotonically nonincreasing curve that goes to zero is the

Some Clues to understanding GMRES convergence
We want to understand stagnation better.

\[ 0 = (I-u)x = \cdots = (I)x = (0)x \]

Progress for \( n \) iterations:

Well-known in which GMRES completely stagnates, failing to make any

In GMRES, we do not have this nice property. In fact, examples are

orthogonal to the gradient of the function minimized.

Each iteration, because the new component of the Krylov subspace is never

CC and MINRES are guaranteed to make progress, however minimal, at

What is the worst case convergence for GMRES?
exact solution to the problem. If \( n - 1 \), then this is complete stagnation, and then \( x \) will be the

\[(u) x = \cdots = (1) x = (0) x\]

\[0 = (u) x = \cdots = (1) x = (0) x\]

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

Joint work with Ilya Zavorin and Howard Elman.

Stagnation of CRES
\[ q_{I-1} \Lambda = \hat{h} \text{ and let } \Lambda V \Lambda = \mathcal{V} \text{ be the eigenvalue decomposition of } \mathcal{V} \]
where $\mathbf{e}$ is the vector of ones.

\[
\begin{pmatrix}
\mathbf{e}^\top \mathbf{V} & \cdots & \mathbf{e} \mathbf{V} \\
\vdots & \ddots & \vdots \\
\mathbf{e} \mathbf{V} & \cdots & \mathbf{e} \mathbf{V}
\end{pmatrix} = \mathbf{I}^{u+w}
\]

and

\[
\forall \mathbf{1}_{1+w} \in L[0, \ldots, 1, 0, \ldots, 1] = \mathbf{e} \mathbf{1}
\]

where $\mathbf{X}$ is the diagonal matrix formed from the entries of $\mathbf{v}$.

(1)

\[
\mathbf{\mathbf{f}^\top I} = \mathbf{f}^\top \mathbf{\Lambda}_H \mathbf{\Lambda}_H^{1+w} Z
\]

Theorem: Characterizing Partial Stagnation

Characterizing Partial Stagnation
our stagnation system. \[ I_{1} = q_{I}^{1+u} H Y \]

I only noted that expression gives

I\[ I \]

\[ \text{Therefore, CMRES stagnates at step } u \text{ if and only if } q \text{ is orthogonal to the subspace orthogonal to the span of the columns of } AV^{u}. \]

This means that the resulting residual \[ w R \]

\[ [q_{I-u}, \ldots, q_{A}, q] = w Y \]

in the span of the columns of \[ x \]

\[ \text{Proof: At the } u \text{th step, CMRES minimizes the residual over all vectors } x. \]
(2) \( \left( \frac{\chi - \chi}{\chi} \prod_{i=1}^{n} \text{const}_{1+u} (1- \right) = \lambda u \)

where \( n \) is a vector derived from the eigenvalues \( \lambda \):

\( n = \lambda \Lambda \Lambda \Lambda \)

Complete stagnation occurs iff

\( \cdot \in \left( \lambda \Lambda \Lambda \Lambda^{1+u} \right) \)

system

If \( n = I + um \), then \( u \) is invertible, and we can rewrite the stagnation.

Characterizing Complete Stagnation
condition number is greater than or equal to some critical value \( \lambda_c \).

Then \( A \) is completely 

Let the eigenvalues of \( A \) be \( 1 \) and \( \lambda > 1 \).

expressions characterizing stagnation.

In certain simple cases, for example for \( n = 2 \), we can get closed-form

Illustration for \( n = 2 \)
Illustration for $n = \mathcal{Z}$: Contours of $y = \phi(\theta, \nu^x, x_0)$.
\( f_{\Lambda} = q \) and \( n_{\Lambda} = q \). Use \( q = x_H V \) and only if it

Complete characterization for a matrix \( A \) implies complete characterization for \( V_A \).

A consequence of the characterization.
If \( \Lambda \) is such that the corresponding \( n \) contains complex or real negative entries, then there is no right-hand side for which CMRES stagnates.

and the phase angles \( \theta \) are arbitrary.

\[
\begin{align*}
u, & \ldots, I = I, \\
& e^{\theta_i n} \Lambda = \Lambda
\end{align*}
\]

\( \Lambda \) is CMRES stagnates for \( q \), where

\[
\begin{align*}
u, & \ldots, I = I, \\
& e^{n} = e^{\Lambda}
\end{align*}
\]

which is a system of decoupled equations of the form,

\[
\begin{align*}
& n = \Lambda
\end{align*}
\]

Therefore, for normal matrices,

\[
\begin{align*}
& u, \ldots, I = I, \\
& e^{n} = e^{\Lambda}
\end{align*}
\]

In this case, the stagnation system simplifies to

\[
\begin{align*}
& n = \Lambda
\end{align*}
\]

A normal matrix \( \Lambda \) is one whose eigenvector matrix \( \Lambda \) is unitary.

Complete Stagnation of Normal Matrices
Theorem of Marshall and Olkin tells us that such a scaling matrix exists. If $n > 0$, then a matrix $A$ has row sums $n$. Since $n$ is equivalent to finding a diagonal scaling. Solving the equation $n = \mu_i A$, where $n$ is real, then it is symmetric positive definite. Solving the

$\frac{1}{\mu_i} = \lambda_i A = \mu_i A$, where $\mu_i \in \mathbb{R}$.

**Theorem:** Suppose we have a vector $\lambda \in \mathbb{C}$ with distinct elements such matrices with the same eigenvalues.

A normal matrix does imply stagnation of an entire family of

Does Normal Stagnation Imply Non-Normal Stagnation?
can be an eigenvalue for a completely stagnating matrix. For any nonzero scalar \( c \), every complex number \( \lambda \) of \( cI - A \) is an eigenvalue of \( A \). Therefore, every solution of \( \lambda_1 \) is so that the following equation holds:

\[
\begin{pmatrix}
\frac{t}{\lambda_1} \\
\frac{t}{\lambda_2}
\end{pmatrix}
\begin{bmatrix}
I + u(1 - I) &=& \ell n
\begin{pmatrix}
\frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_2} \\
\frac{t}{\lambda_1}
\end{pmatrix}
\end{bmatrix}
\]

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

\[
\lambda_1 \geq \begin{pmatrix}
\frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_2} \\
\frac{t}{\lambda_1}
\end{pmatrix}
\begin{bmatrix}
I + u(1 - I) &=& \ell n > 0
\end{bmatrix}
\]

Consequently, stagnating eigenvalue distributions will stagnate for any eigenvector matrix satisfying real, if GMR-1, 3. Constraining Stagnating Eigenvalue Distributions
Some Stagñating Eigenvalue Distributions for $n = 3$, $\Lambda_{H} \Lambda_{H}^{\dagger}$ Real
Theorem: These are the only unitary matrices for which complete stagnation can occur.

The converse can be established using the stagnation system.

The complex plane, uniformly over the unit circle in the complex plane, completely stagnate when applied to a unitary matrix $A$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

A normal matrix $A$ is unitary if its eigenvalues satisfy

$\forall \lambda_i \in \mathbb{C}, \exists \phi \in \mathbb{R}, u \in \mathbb{C}, u \neq 0 : \lambda_i = e^{\phi i u}$
right-hand sides. It is possible to construct real matrices $A$ that never completely stagnate eigenvalues. Where $P$ is a permutation matrix that depends on the ordering of $n = \hat{p} \Lambda \hat{d} \Lambda$

Experimentation: Considerably simplifying analyses and numerical system in $\hat{p}$, the stagnation system can be written as a polynomial when $A$ is real, the stagnation of real matrices

Complete Stagnation of Real Matrices
and if other right-hand sides, none of them real.

\[
\begin{bmatrix}
-1.8679775 & -1.2644748 \\
0.7086397 & 1.5089330 \\
-1.2084970 & -0.3414864 \\
1.5564116 & 1.5564116
\end{bmatrix} = \vec{f}
\]

stagnates for

\[
\mathbf{Y} = (1.0000000, -0.7658066, -0.2656295, 0.8705227)
\]

and eigenvalues

\[
\begin{bmatrix}
-0.4333364 & -0.4893898 & -0.1323115 \\
0.5155494 & 0.7499413 & 0.6984230 \\
0.3213318 & 0.4920391 & 0.7586559 \\
0.390634 & -0.087875 & 0.2414875 \\
0.3998204 & 0.4306034 & -0.4306034
\end{bmatrix} = \Lambda
\]

Example: The matrix with eigenvectors
Conclusions