Abstract

Each iterate generated by the Generalized Conjugate Gradient Method of Concus and Golub [1] and Widlund [3] is shown to be the best approximation to the solution from a certain affine subspace (although *not* from the "natural" affine Krylov subspace). This property is used to improve the error bounds given by Widlund [3] and Hageman, Luk, and Young [2].

A Note on the Generalized Conjugate Gradient Method

Stanley C. Eisenstat

Research Report #228
March 1982

This research was supported in part by ONR Grant N00014-76-C-0277 and NSF Grant MCS-8104874.

1. Introduction

The Generalized Conjugate Gradient Method of Concus and Golub [1] and Widlund [3] is an iterative method for solving a system of linear equations $Ax = b$ when the coefficient matrix A is real and has positive definite symmetric part $M = (A+A^t)/2$:

LET $x^{(0)}$ BE GIVEN AND SET $x^{(-1)} = 0$.

FOR $m = 0$ STEP 1 UNTIL "CONVERGENCE" DO

SOLVE $Mv^{(m)} = b - Ax^{(m)}$

COMPUTE[1] $\rho_m = (Mv^{(m)}, v^{(m)})$

IF $m = 0$ THEN

SET $\omega_{m+1} = 1$

ELSE

COMPUTE $\omega_{m+1} = [1 + \rho_m/(\rho_{m-1}\omega_m)]^{-1}$

COMPUTE $x^{(m+1)} = x^{(m-1)} + \omega_{m+1}(v^{(m)} + x^{(m)} - x^{(m-1)})$

Let $A = M-N$, whence $-N = (A-A^t)/2$ is the skew-symmetric part of A, and let $K = M^{-1}N$. Then it can be shown that the iterate $x^{(m)}$ lies in the affine Krylov subspace

$$x^{(0)} + \text{Span}\{v^{(0)}, Kv^{(0)}, K^2v^{(0)}, \ldots, K^{m-1}v^{(0)}\} \equiv x^{(0)} + S_m$$

and is characterized by the Galerkin condition

$$(z, Ae^{(m)}) = 0 \qquad \text{for all } z \in S_m, \tag{1.1}$$

where $e^{(m)} \equiv x^{(m)} - x$ (see [3]). Moreover,

$$x^{(m)} = x + p_m(K)e^{(0)} \tag{1.2}$$

where $p_m(\mu)$ is an even (odd) polynomial of degree at most m for m even (odd) and $p_m(1) = 1$ (see [3]).

In this paper, we show that $x^{(m)}$ is the *best* approximation to x from a certain m-dimensional affine subspace (but *not* from the affine Krylov subspace $x^{(0)} + S_m$) and use this property to improve the error bounds given by Widlund [3] and Hageman, Luk, and Young [2].

---

[1] (y,z) denotes the Euclidean inner-product.

Notation: $(y,z)_M$ denotes the M-inner product (My,z) and $\|z\|_M$ denotes the corresponding norm. Note that

$$(Ky,z)_M \;=\; (Ny,z) \;=\; -(y,Nz) \;=\; -(My,M^{-1}Nz) \;=\; -(y,Kz)_M$$

so that K is skew-symmetric with respect to $(\cdot,\cdot)_M$ and $(Kz,z)_M = 0$ for all z.

2. An Alternative Characterization

In this section, we show that the iterate $x^{(m)}$ generated by the Generalized Conjugate Gradient Method is the best approximation to x with respect to a certain m-dimensional affine subspace, but not with respect to the affine Krylov subspace $x^{(0)}+S_m$ (unless $x^{(m)} = x$). The cases m even ($= 2k$) and m odd ($= 2k+1$) are treated separately.

Theorem 2.1: $x^{(2k)} \in x^{(0)}+(I+K)S_{2k}$ and

$$(z,\, x^{(2k)}-x)_M \;=\; 0 \qquad \text{for all } z \in (I+K)S_{2k},$$

whence

$$\|x^{(2k)}-x\|_M \;=\; \min\; \{\|y-x\|_M \mid y \in x^{(0)}+(I+K)S_{2k}\}\;.$$

Proof:

Since $p_{2k}(-1) = p_{2k}(1) = 1$ (recall that p_{2k} is even), $p_{2k}(\mu)$ can be written in the form

$$p_{2k}(\mu) \;=\; 1 \;+\; (1+\mu)\,\pi_{2k-2}(\mu)\,(1-\mu)$$

where $\pi_{2k-2}(\mu)$ is a polynomial of degree at most $2k-2$. Therefore, by (1.2),

$$\begin{aligned} x^{(2k)} &= x \;+\; e^{(0)} \;+\; (I+K)\,\pi_{2k-2}(K)\,(I-K)e^{(0)} \\ &= x^{(0)} \;-\; (I+K)\,\pi_{2k-2}(K)v^{(0)} \\ &\in x^{(0)}+(I+K)S_{2k}\;. \end{aligned}$$

If $z \in (I+K)S_{2k}$, then $z = (I+K)u$ for some $u \in S_{2k}$ and

$$(z,\, x^{(2k)}-x)_M \;=\; (M(I+K)u,\, e^{(2k)}) \;=\; (u,\, Ae^{(2k)}) \;=\; 0$$

by the Galerkin condition (1.1). □

However, $x^{(2k)}$ is *not* the best approximation to x from $x^{(0)}+S_{2k}$. To see this, note that

$$(v^{(0)}, x^{(2k)}-x)_M = -((I-K)e^{(0)}, e^{(2k)})_M$$
$$= -(e^{(2k)}, e^{(2k)})_M + (e^{(2k)}-e^{(0)}, e^{(2k)})_M + (Ke^{(0)}, p_{2k}(K)e^{(0)})_M .$$

By Theorem 2.1, $e^{(2k)}-e^{(0)} = x^{(2k)}-x^{(0)} \in (I+K)S_{2k}$ and the second term vanishes. Since K is skew-symmetric with respect to $(\cdot,\cdot)_M$ and p_{2k} is even, the third term also vanishes. Therefore, $v^{(0)} \in S_{2k}$ but

$$(v^{(0)}, x^{(2k)}-x)_M = -(e^{(2k)}, e^{(2k)})_M \neq 0 ,$$

unless $x^{(2k)} = x$.

Theorem 2.2: $x^{(2k+1)} \in x^{(0)}+v^{(0)}+(I+K)S_{2k+1}$ and

$$(z, x^{(2k+1)}-x)_M = 0 \qquad \text{for all } z \in (I+K)S_{2k+1},$$

whence

$$\|x^{(2k+1)}-x\|_M = \min \{\|y-x\|_M \mid y \in x^{(0)}+v^{(0)}+(I+K)S_{2k+1}\} .$$

Proof:

Since $p_{2k+1}(1) = 1$ and $p_{2k+1}(-1) = -p_{2k+1}(1) = -1$ (recall that p_{2k+1} is odd), $p_{2k+1}(\mu)$ can be written in the form

$$p_{2k+1}(\mu) = \mu + (1+\mu)\, \pi_{2k-1}(\mu)\, (1-\mu)$$

where $\pi_{2k-1}(\mu)$ is an odd polynomial of degree at most $2k-1$. Therefore, by (1.2),

$$x^{(2k+1)} = x + Ke^{(0)} + (I+K)\, \pi_{2k-1}(K)\, (I-K)e^{(0)}$$
$$= x^{(0)} - (I-K)e^{(0)} - (I+K)\, \pi_{2k-1}(K)v^{(0)}$$
$$= x^{(0)} + v^{(0)} - (I+K)\, \pi_{2k-1}(K)v^{(0)}$$
$$\in x^{(0)}+v^{(0)}+(I+K)S_{2k+1} .$$

If $z \in (I+K)S_{2k+1}$, then $z = (I+K)u$ for some $u \in S_{2k+1}$ and

$$(z, x^{(2k+1)}-x)_M = (M(I+K)u, e^{(2k+1)}) = (u, Ae^{(2k+1)}) = 0$$

by the Galerkin condition (1.1). □

Again, $x^{(2k+1)}$ is *not* the best approximation to x from $x^{(0)}+S_{2k+1}$. To see this, note that

$$\begin{aligned}(v^{(0)}, x^{(2k+1)}-x)_M &= -((I-K)e^{(0)}, e^{(2k+1)})_M \\ &= (e^{(2k+1)}, e^{(2k+1)})_M - (e^{(2k+1)}-Ke^{(0)}, e^{(2k+1)})_M \\ &\qquad - (e^{(0)}, p_{2k+1}(K)e^{(0)})_M \ .\end{aligned}$$

By Theorem 2.2, $e^{(2k+1)}-Ke^{(0)} = x^{(2k+1)}-x^{(0)}-v^{(0)} \in (I+K)S_{2k}$ and the second term vanishes. Since K is skew-symmetric with respect to $(\cdot,\cdot)_M$ and p_{2k+1} is odd, the third term also vanishes. Therefore, $v^{(0)} \in S_{2k+1}$ but

$$(v^{(0)}, x^{(2k+1)}-x)_M = (e^{(2k+1)}, e^{(2k+1)})_M \neq 0 \ ,$$

unless $x^{(2k+1)} = x$.

3. Error Bounds

In this section, we use the best approximation property of the iterates $\{x^{(m)}\}$ to prove error bounds for the Generalized Conjugate Gradient Method.

Theorem 3.1:

$$\|x^{(m)}-x\|_M \leq \|q_m(K)(x^{(0)}-x)\|_M$$

for any real polynomial $q_m(\mu)$ of degree at most m satisfying $q_m(1) = 1$ and $q_m(-1) = (-1)^m$.

Proof:

Let $y \equiv x + q_m(K)e^{(0)}$. Then it can be shown that $y \in x^{(0)}+(I+K)S_m$ if m is even (see the first part of the proof of Theorem 2.1) and that $y \in x^{(0)}+v^{(0)}+(I+K)S_m$ if m is odd (see the first part of the proof of Theorem 2.2). Therefore, using either Theorem 2.1 or Theorem 2.2,

$$\|x^{(m)}-x\|_M \leq \|y-x\|_M = \|q_m(K)(x^{(0)}-x)\|_M \ .$$

□

Let $\sigma(K)$ denote the spectrum of K. Since K is skew-symmetric with respect to $(\cdot,\cdot)_M$, it can be shown that

$$\text{Re } \mu = 0, \quad |Im\ \mu| \leq \|K\|_M \equiv \Lambda$$

for any $\mu \in \sigma(K)$, and that

$$\|q_m(K)\|_M = \max_{\mu \in \sigma(K)} |q_m(\mu)|$$

for any real polynomial $q_m(\mu)$.

Corollary 3.2:

$$\|x^{(m)} - x\|_M \leq \frac{2}{R(\Lambda)^m + [-R(\Lambda)]^{-m}} \|x^{(0)} - x\|_M$$

where $R(\Lambda) = \Lambda^{-1} + \sqrt{\Lambda^{-2} + 1}$.

Proof:

Let $q_m(\mu) = T_m(i\Lambda^{-1}\mu)/T_m(i\Lambda^{-1})$ where $T_m(z)$ is the m^{th} Chebyshev polynomial. Since $T_m(z)$ is even (odd) when m is even (odd), $q_m(\mu)$ is a real polynomial which satisfies the conditions of Theorem 3.1 so that

$$\|x^{(m)} - x\|_M \leq \|q_m(K)(x^{(0)} - x)\|_M \leq \|q_m(K)\|_M \|x^{(0)} - x\|_M .$$

But

$$\|q_m(K)\|_M = \max_{\mu \in \sigma(K)} \frac{|T_m(i\Lambda^{-1}\mu)|}{|T_m(i\Lambda^{-1})|} \leq \frac{1}{|T_m(i\Lambda^{-1})|}$$

since $-1 \leq i\Lambda^{-1}\mu \leq +1$ for all $\mu \in \sigma(K)$ and $|T_m(z)| \leq 1$ for $-1 \leq z \leq +1$. Moreover, it can be shown that

$$T_m(i\Lambda^{-1}) = \frac{i^m}{2} [R(\Lambda)^m + [-R(\Lambda)]^{-m}] .$$

Therefore, since $R(\Lambda) > 1$,

$$\|x^{(m)} - x\|_M \leq \frac{2}{R(\Lambda)^m + [-R(\Lambda)]^{-m}} \|x^{(0)} - x\|_M$$

□

Hageman, Luk, and Young [2] proved Corollary 3.2 for m even by observing that the even iterates can also be generated by applying conjugate gradient acceleration to a certain symmetrizable "double" method. Widlund [3] proved somewhat weaker bounds for general m using a standard argument for Galerkin methods.

The best approximation property and the nesting of the subspaces $\{S_m\}$ guarantees that $\{\|e^{(2k)}\|_M\}$ and $\{\|e^{(2k+1)}\|_M\}$ are both monotone decreasing. Widlund [3] gives a direct proof. The following result shows that both sequences must converge at the same rate, contradicting the experimental results reported in [3].

Corollary 3.3:

$$\Lambda^{-1}\,\|x^{(m+1)}-x\|_M \;\le\; \|x^{(m)}-x\|_M \;\le\; \Lambda\,\|x^{(m-1)}-x\|_M \qquad \text{for all } m \ge 1.$$

Proof:

It suffices to prove the right-hand inequality. Since $q_m(\mu) = \mu p_{m-1}(\mu)$ satisfies the conditions of Theorem 3.1,

$$\begin{aligned} \|x^{(m)}-x\|_M &\le \|q_m(K)(x^{(0)}-x)\|_M \\ &\le \|K\|_M\,\|p_{m-1}(K)e^{(0)}\|_M \\ &= \Lambda\,\|x^{(m-1)}-x\|_M \;. \end{aligned}$$

□

References

[1] Paul Concus and Gene H. Golub. A generalized conjugate gradient method for nonsymmetric systems of linear equations. In R. Glowinski and J. L. Lions, Editors, *Lecture Notes in Economics and Mathematical Systems, Volume 134*, Springer Verlag, 1976, pp. 56-65.

[2] L. A. Hageman, Franklin T. Luk, and David M. Young. On the equivalence of certain iterative acceleration methods. *SIAM Journal on Numerical Analysis* 17:852-873 (1980).

[3] Olof Widlund. A Lanczos method for a class of nonsymmetric systems of linear equations. *SIAM Journal on Numerical Analysis* 15:801-812 (1978).