ANALELE ŞTIINŢIFICE ALE UNIVERSITĂŢII "AL.I. CUZA" IAŞI
Tomul LI, s.I, Matematică, 2005, f.1

EIGEN VALUE PERTURBATION FOR GRADIENT METHOD

BY

J.O. OMOLEHIN

1. Abstract. The ratio of the smallest and largest eigenvalue of the control operator A in quadratic functional is used as a perturbation term for the Gradient Method and the convergence rate is consistent and stable.

Mathematics Subject Classification 2000: 65kxx.

Key words: Gradient, perturbation, convergence, eigenvalues, operator.

2. Introduction. Consider the quadratic functional of the form $f(x) = f_0 + \langle a, x\rangle_H + \frac{1}{2}\langle x, Ax\rangle_H$, where A is an n X n symmetric, positive definite constant matrix operator on the Hilbert space H, a is a vector in H and f_0 is a constant term.

The Conventional Conjugate Gradient Method for the minimization of the above quadratic form has been extensively considered [2]. In this work we consider the extension of our previous work on this problem (see [1]).

In our previous work, we used the result of our spectrum analysis to solve a problem that is not in the class of quadratic form. In that work CGM algorithm is not applicable; only the associated theoretical aspect was used. In the present work both CGM and GM algorithms are available. We shall proceed to give the exact form of gradient method (GM).

Consider the function $f : R^n \to R$ which is continuously differentiable in some domain $D \subseteq R^n$ and it is assumed that f assumes a local minimum value in D at a point $x\varepsilon D^0$, *where* D^0 is the interior of D. Now consider

the Taylor's series expansion.

$$f\left(x-\lambda\frac{\partial f}{\partial x}(x)^T\right) = f(x)+\frac{\partial f}{\partial x}(x)\left(-\lambda\frac{\partial f}{\partial x}(x)^T\right)+\theta(\lambda) = \\ = f(x)-\lambda\left\|\frac{\partial f}{\partial x}(x)\right\|_e^2+\theta(\lambda),$$

e is the usual euclidian space. $\frac{\partial f}{\partial x}(x)\neq 0$, then for sufficiently small $\lambda>0$ we have $f\left(x-\lambda\frac{\partial f}{\partial x}(x)^T\right)< f(x)$. Hence, the gradient method (GM) is defined as the construction of sequence x_k of points in R^n by the recursion equation: $x_{k+1}=x_k-\lambda\frac{\partial f}{\partial x}(x_k)^T, \quad k=0,1,2,\ldots$

Where x_0 is the initial guess value [3, p. 245-293]. The convergence rate of GM based on the choice of λ has been extensively considered [3]. The major problem associated with GM algorithm is the choice of λ [3]. Sometimes one of the components of the vector will be converging rapidly; other components may not show any sign of convergence at all.

Observe that the convergence rate of CGM algorithm [2] is based on the gradient of the functional whereas the convergence rate of GM algorithm is based on the minimizing vector. All these points will be embraced in a theorem in our main result.

3. Main result

Theorem. The convergence rate of GM algorithm for quadratic functional remains stable if $\lambda=\frac{m}{M}$ where m and M are the smallest and largest eigenvalues of the control operator A respectively.

Proof. Recall the problem of the minimization of $f: R^n\to R$ given by

$$f(x)=f_0+\langle a,x\rangle_H+\frac{1}{2}\langle x,Ax\rangle_H, \tag{1}$$

where A is an nxn symmetric positive definite matrix and H is a Hilbert space. In our own case here $H\equiv R^n$. Differentiate (1) to obtain

$$\frac{\partial f(x)}{\partial x}=a+Ax \tag{2}$$

consider

$$x_{k+1}=k_k-\lambda\frac{\partial f(x)^T}{\partial x}. \tag{3}$$

Substitute equation (2) in (3), to obtain

$$x_{k+1} = x_k - \lambda A x_k - \lambda a = (I - \lambda A)x_k - \bar{\lambda}(\lambda a = \bar{\lambda}\, constant). \tag{4}$$

It has been established [1] that there exist an orthogonal matrix p which diagonalizes A, i.e.

$$p^{-1}Ap = p^T Ap = diag(\lambda, \lambda, \ldots, \lambda n), \tag{5}$$

where p^{-1} and p^T are inverse and transpose of p respectively.

$$x_k = py_k, \qquad k = 0, 1, 2, 3. \tag{6}$$

Therefore

$$p_{k+1} = (I - \lambda A)py_k \tag{7}$$

(without loss of generality the constant $\bar{\lambda}$ can be dropped)

$$y_{k+1} = (I - \lambda pAp^{-1})y_k = diag(1 - \lambda\lambda_1, 1 - \lambda\lambda_2, 1 - \lambda\lambda_3, \ldots 1 - \lambda\lambda_n y_n \tag{8}$$

$$y_{k+1}^j = (1 - \lambda\lambda_j)^k y_0^j, \quad j = 1, 2, \ldots, n. \tag{9}$$

Thus

$$y_k^j = (1 - \lambda\lambda_j)^K y_0^j \qquad j = 1, 2, \ldots, n. \tag{10}$$

Therefore y_k^j form a geometric progression $y_0^j, (1-\lambda\lambda_j)y_0^j, (1-\lambda\lambda_j)^2 y_0^j, \ldots$. The rate at which the numbers approach the minimum is dependent on

$$|1 - \lambda\lambda_j|. \tag{11}$$

Ibiejugba et al. established the convergence rate of CGM as

$$\frac{E(x_n)}{E(x_0)} \leq \frac{1 - \frac{m}{M}}{1 + \frac{m}{M}}, \tag{12}$$

where m and m are smallest and largest eigenvalues of matrix A respectively. Equate RHS of (12) to (11) and simplify. Therefore, we have

$$\frac{1 - \frac{m}{M}}{1 + \frac{m}{M}} = 1 - \lambda\lambda_j$$

$$\frac{M-m}{M+m} - 1 = -\lambda\lambda_j \tag{13}$$

$$\frac{2m}{\lambda_j} = \lambda. \tag{14}$$

We take $\lambda_j = 2M$ (for some j) and equation (14) becomes

$$\lambda = \frac{m}{M}. \tag{15}$$

Therefore our theorem has been proved. The next stage of the work is the numerical investigation. □

4. Numerical work. Consider the quadratic functional of the form:

$$f(x) = f0 + \langle a, x\rangle_H + \frac{1}{2}\langle x, Ax\rangle_H, \tag{16}$$

where a, A, f_0 and x are as previously defined. The gradient method for the above problem is

$$x_{k+1} = x_k - \lambda\frac{\partial f(x)^T}{\partial x}. \tag{17}$$

In this work λ is taken to be equal to (m/M) where m and M are as previously defined. The formulation of the problem is as follows:

Minimize $f(x) = f_0 + (a, x)_H + \frac{1}{2}\langle x, Ax\rangle_H$

Problem 1: $a = (1,1)^T, A = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}$

Problem 2: $a = (1,1,1)^T, A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 3 \end{pmatrix}$

Problem 3: $a = (1,1,1,1,1,1)^T, A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 4 & 1 & 2 & 1 & 1 \\ 1 & 1 & 3 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 1 & 0 \\ 1 & 1 & 1 & 1 & 4 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}$

Table 1: $H \equiv R^2$

ITRN	$x \in R^2$
1	$x_1 = -0.2+$ 2716 $x_2 = -0.1928$
2	$x_1 = -0.4322$ $x_2 = -0.1859$
10	$x_1 = -0.7550$ $x_2 = -0.7442$
70	$x_1 = -0.9998$ $x_2 = -0.0001$
90	$x_1 = -1.0000$ $x_2 = -0.000$
95	$x_1 = -1.0000$ $x_2 = -0.000$

Table 2: $H \equiv R^3$

ITRN	$x \in R^3$
1	$x_1 = -0.1637$ $x_2 = -0.1363$ $x_3 = -0.1454$
2	$x_1 = -0.2167$ $x_2 = -0.1502$ $x_3 = -0.1706$
10	$x_1 = -0.4790$ $x_2 = -0.1049$ $x_3 = -0.1606$
70	$x_1 = -0.9700$ $x_2 = -0.0056$ $x_3 = -0.0096$
90	$x_1 = -0.9884$ $x_2 = -0.0022$ $x_3 = -0.0037$
95	$x_1 = -0.9909$ $x_2 = -0.0018$ $x_3 = -0.0029$

Table 3: $H \equiv R^6$

ITRN	$x \in R^3$
1	$x_1 = -0.0450$ $x_2 = -0.0422$ $x_3 = -0.0433$ $x_4 = -0.0439$ $x_5 = -0.0427$ $x_6 = -0.0456$
2	$x_1 = -0.0638$ $x_2 = -0.0557$ $x_3 = -0.0589$ $x_4 = -0.0606$ $x_5 = -0.0573$ $x_6 = -0.0654$
10	$x_1 = -0.1683$ $x_2 = -0.0758$ $x_3 = -0.1063$ $x_4 = -0.1322$ $x_5 = -0.0919$ $x_6 = -0.1871$
70	$x_1 = -0.7125$ $x_2 = -0.2367$ $x_3 = -0.1505$ $x_4 = -0.4052$ $x_5 = -0.1249$ $x_6 = -0.9403$
95	$x_1 = -0.9554$ $x_2 = -0.3878$ $x_3 = -0.3045$ $x_4 = -0.5494$ $x_5 = -0.2313$ $x_6 = -1.2978$

5. Analysis of Result. ITRN = iteration number

The numerical results are tabulated in Tables 1-3 respectively.

It can be seen from Tables 1-3 that the convergence rate is very stable and consistent. It is consistent in the sense that each component in the vector is either increasing steadily or decreasing steadily at every interval of the iteration. This is a break through, because it is very unusual for Gradient Method to behave in that form. The next work will compare this result with the exact solution of the problem considered.

REFERENCES

1. Omolehin, J.O. – *Computational convergence rate of Gradient Method*, NJMA. (Accepted), (1996).
2. Ibiejugba, M.A. – *Computational methods in optimization*, Ph.D Thesis, University of Leeds, Leeds, U.K., (1985).
3. Russel, David, L. – *Optimixation theory*, New York, W.A. Benjamin, Inc., (1970).
4. Omolehin, J.O. – *On the control of reaction diffusion equation*, Ph.D. Thesis, University of Ilorin, Ilorin, (1991).
5. Omolehin, J.O. – *Experiment with extended conjugate gradient method algorithm*, M.sc. University of Ilorin, Ilorin.

Received: 10.V.2004

Mathematics Department,
University of Ilorin,
Ilorin,
NIGERIA,
omolehin_joseph@yahoo.com