Journal of the Nigerian Association of Mathematical Physics
***Volume* 13 (November, 2008), 417 - 422**
© J. of NAMP

Effect of matrix transformation on minimization of quadratic functional

[1]*J. O. Omolehin, [2]M. K. A. Abdulrahman, [1]K. Rauf and [3]A. O. Ameen
[1]Department of Mathematics, University of Ilorin, Ilorin, Nigeria.
[2]Department of Mathematics, Statistics and Computer Science, Kaduna Polytechnic, Kaduna, Nigeria.
[3]Department of Computer Science, University of Ilorin, Ilorin, Nigeria.

Abstract

The convergence profile of the conventional Conjugate Gradient Method (CGM) algorithm is based on the symmetry of the control operator for quadratic functions. This work considers the quadratic functions with non-symetric control operator under suitable matrix transformations: it is proved that the conventional CGM algorithm produces results that are favourably comparable to problems with symmetric control operator equivalent.

C.R. Category: G. 1.7

1.0 Introduction

Consider the quadratic functional of the form:

$$f(x) = f_0 + \langle a, X \rangle_H + \frac{1}{2}\langle X, AX \rangle_H \tag{1.1}$$

where A is an $n \times n$ symmetric positive definite constant matrix operator on Hilbert space H, a is a vector in H and f_0 is a constant term. The conventional CGM algorithm is used as a computation method for the minimization of the above problem and it has been proved that the CGM algorithm enjoys quadratic rate of convergence. That is, the method converges in at most n iterations for an n-dimensional problem. In what follows is the convergence rate of the CGM algorithm.

2.0 Convergence profile of the conventional CGM algorithm

To fully understand our numerical work reported in this work it will be necessary to show the convergence rate of CGM algorithm [1, 4, 6, 8, 9 and 10].
Recall the quadratic functional

$$f(X) = f_o + \langle a, X \rangle_H + \tfrac{1}{2}\langle X, AX \rangle_H$$

where f_0 is constant, H is a Hilbert space, X is a $n \times n$ dimensional vector in H, a positive definite constant matrix operator see [5 and 11].

The convergence rate of CGM algorithm is given as $E(X_n) = \left\{ \dfrac{1 - \frac{m}{M}}{1 + \frac{m}{M}} \right\}^{2n} E(X_o)$, where m and M are the smallest and largest eigen value of A respectively.
Proof: see [4].

[1]*Corresponding author, e-mail-omolehin_joseph@yahoo.com

Define $E(X)=\frac{1}{2}\left\langle (X-X^*), A(X-X^*)\right\rangle_H$

Therefore,

$$E(X)=\frac{1}{2}\left\langle (X-X^*), A(X-X^*)\right\rangle_H=\frac{1}{2}\left\langle X+A^{-1}a, A(X+A^{-1}a)\right\rangle_H=\frac{1}{2}\left\langle X+A^{-1}a, AX+AA^{-1}a\right\rangle_H$$

$$=\frac{1}{2}\left\langle X+A^{-1}a, AX+a, AX+a\right\rangle_H=\frac{1}{2}\left\langle X, AX\right\rangle_H+\frac{1}{2}\left\langle X, a\right\rangle_H+\frac{1}{2}\left\langle A^{-1}a, AX\right\rangle_H+\frac{1}{2}\left\langle A^{-1}a, a\right\rangle_H$$

$$=\frac{1}{2}\left\langle X, AX\right\rangle_H+\frac{1}{2}\left\langle X, a\right\rangle_H+\frac{1}{2}\left\langle A^{-1}a, AX\right\rangle_H+\frac{1}{2}\left\langle -X^*, a\right\rangle_H$$

$$=\frac{1}{2}\left\langle X, AX\right\rangle_H+\frac{1}{2}\left\langle X, a\right\rangle_H+\frac{1}{2}\left\langle X^*, AX^*\right\rangle_H+\frac{1}{2}\left\langle A^{-1}a, AX\right\rangle_H$$

$$E(x)=F(X)-F_o+\frac{1}{2}\left\langle X^*, AX^*\right\rangle_H=F(X)-\tilde{F}(o)$$

Therefore, $E(X)$ is $F(X)$ plus a constant term, hence the convergence of $E(X)$ is considered instead of that of $F(X)$ as from now. Recall that,

$$E(X)=\frac{1}{2}\left\langle X+A^{-1}a, AX+a\right\rangle_H=\frac{1}{2}\left\langle A^{-1}(AX+a), AX+a\right\rangle_H=\frac{1}{2}\left\langle A^{-1}g(X), g(X)\right\rangle_H$$

Hence,

$$E(X_i)-E(X_{i+1})=\frac{1}{2}\left\langle X_i-X^*, A(X_i-X^*)\right\rangle_H-\frac{1}{2}\left\langle X_{i+1}-X^*, A(X_{i+1}-X^*)\right\rangle_H$$

But $X_{i+1}=X_i+\alpha_i p_i$, therefore

$$E(X_i)-E(X_{i+1})=\frac{1}{2}\left\langle X_i-X^*, A(X_i-X^*)\right\rangle-\frac{1}{2}\left\langle X_i+X_i p_i-X^*, A(Xi+\alpha_i p_i-X^*)\right\rangle_H$$

$$=\frac{1}{2}\left\langle X_i-X^*, A(X_i-X^*)\right\rangle_H-\frac{1}{2}\left\langle X_i-X^*, A(X_i-X^*)\right\rangle_H$$

$$-\frac{1}{2}\alpha_i\left\langle p_i, A(X_i+\alpha_i p_i-X^*)\right\rangle_H-\frac{1}{2}\alpha_i\left\langle X_i-X^*, Ap_i\right\rangle_H$$

$$=-\frac{\alpha_i}{2}\left\langle p_i, A(X_i-X^*)\right\rangle_H-\frac{1}{2}\alpha_i\left\langle X_i-X^*, Ap_i\right\rangle_H-\frac{1}{2}\alpha_i\left\langle p_i, A\alpha_o p_o\right\rangle_H$$

$$=-\alpha_i\left\langle p_i, Ax_i+a\right\rangle_H-\frac{1}{2}\alpha_i^{\,2}\left\langle p_i, Ap_i\right\rangle_H-\frac{1}{2}\alpha_i^2\left\langle p_i, Ap_i\right\rangle_H$$

$$=-\alpha_i\left\langle p_i, g_i\right\rangle_H-\frac{1}{2}\alpha_i\left\langle g_i, g_i\right\rangle \text{ since } \alpha_i=\frac{\left\langle g_i, g_i\right\rangle_H}{\left\langle p_i, Ap_i\right\rangle_H}$$

$$=-\alpha_i\left\langle -g_i+\beta_{i-1}p_{i-1}, g_i\right\rangle_H-\frac{1}{2}\alpha_i\left\langle g_i, g_i\right\rangle_H=\alpha_i\left\langle g_i, g_i\right\rangle_H-\alpha_i\beta_{i-1}\left\langle p_{i-1}, g_i\right\rangle_H-\frac{1}{2}\alpha_i\left\langle g_i, g_i\right\rangle_H$$

$$=\alpha_i\langle g_i, g_i\rangle - \frac{1}{2}\alpha_i\langle g_i, g_i\rangle_H$$, (since $\langle p_{i-1}, g_i\rangle_H = 0$ by orthogonality of p_{i-1} and g_i)

$$=\alpha_i\langle g_i, g_i\rangle_H - \frac{1}{2}\alpha_i\langle g_i, g_i\rangle_H = \frac{1}{2}\alpha_i\langle g_i, g_i\rangle_H$$

$$=\frac{\frac{1}{2}\langle g_i, g_i\rangle^2{}_H}{\langle p_i, Ap_i\rangle_H}$$ because $\alpha_i = \dfrac{\langle g_i, g_i\rangle}{\langle p_i, Ap_i\rangle}$. Hence

$$E(X_i) - E(X_{i+1}) = \frac{\langle g_i, g_i\rangle^2{}_H E(X_i)}{\langle p_i, Ap_i\rangle_H \langle g_i, A^{-1}g_i\rangle_H}$$

Using the fact that $g_i = \beta_{i-1}p_{i-1} - p_i$, we get

$\langle g_i, Ag_i\rangle_H = \langle \beta_{i-1}p_{i-1}, A(\beta_{i-1}p_{i-1} - p_i)\rangle_H = \beta_{i-1}{}^2\langle p_{i-1}, Ap_{i-1}\rangle_H + \langle p_i, Ap_i\rangle_H \geq \langle p_i, Ap_i\rangle_H$,

since $\langle p_{i-1}, Ap_{i-1}\rangle_H \geq 0$ (due to the positive definiteness of operator A), $\langle g_i, Ag_i\rangle_H \geq \langle p_i, Ap_i\rangle_H$

Therefore

$$E(X_i) - E(X_{i+1}) \geq \frac{\langle g_i, g_i\rangle^2{}_H E(X_i)}{\langle g_i, Ag_i\rangle_H \langle g_i, A^{-1}g_i\rangle_H}$$

But for a bounded self-adjoint operator in a Hilbert space H, Kantorovich in [7, 2 and 3] established the following inequality

$$\frac{\langle X, X\rangle^2{}_H}{\langle X, AX\rangle_H \langle X, A^{-1}\rangle_H} \geq \frac{4mM}{(m+M)^2},$$ where m and M are respectively the greatest lower and least upper bounds of the spectrum of operator A. using Kantorovich's inequality we obtain

$$E(X_{i+1}) \leq \left\{\frac{1-\frac{m}{M}}{1+\frac{m}{M}}\right\}^{2n} E(X_o)$$

This shows the convergence rate of the CGM. {for all $m \leq M$ }.

3.0 Matrix transformation

Suppose A is a non-symmetric matrix, it is then partitioned into two parts as follows:

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T) \tag{3.1}$$

where the first part is symmetric and the second part is skew symmetric. For subsequent development, the second part can be neglected since the main diagonal elements will be zero. We can now replace our new matrix by B as follows:

$$B = \frac{1}{2}(A + A^T) \tag{3.2}$$

where B is the symmetric transform of A.

4.0 Main results

Suppose the operator A in (1.1) is non-symmetric, we now replace A by B in equation (3.2) to obtain

$$f(x) = f_0 + \langle a, X\rangle_H + \frac{1}{2}\langle X, BX\rangle_H \qquad (4.1)$$

where B is the transform of A.

THEOREM 4.1

Suppose B is the symmetric transform of A in equation (3.2) then,

(i) *the functional* (1.1) *is equivalent to the functional* (4.1)

(ii) $\frac{1}{2}\langle X, AX\rangle_H = \frac{1}{2}\langle X, AX\rangle_H$

Proof

Let $f : \Re^n \to \Re$ then the minimization of f given by

$$f(X) = f_0 + \langle a, X\rangle_H + \frac{1}{2}\langle X, AX\rangle_H$$

and

$$f(X) = f_0 + \langle a, X\rangle_H + \frac{1}{2}\langle X, BX\rangle_H$$

where A and B are non-symmetric nxn matrix and the symmetric transform of A. It is enough to prove (ii). Let,

$$\frac{1}{2}X^T AX = \frac{1}{2}.\begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix}^T.\begin{pmatrix} a_{11} & a_{12} & . & . & a_{1n} \\ a_{21} & a_{22} & . & . & a_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{n1} & a_{n2} & . & . & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix}$$

$$= \frac{1}{2}(a_{11}x_1^2 + a_{22}x_2^2 + \Lambda + a_{nn}x_n^2 + (a_{12} + a_{21})x_1x_2 + \Lambda + (a_{1n} + a_{n1})x_1x_n + (a_{2n} + a_{n2})x_2x_n + \Lambda + (a_{n,n-1} + a_{n-1,n})x_nx_n) \qquad (4.2)$$

Also,

$$\frac{1}{2}X^T BX = \frac{1}{2}.\begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix}^T.\begin{pmatrix} a_{11} & \frac{a_{12}+a_{21}}{2} & . & . & \frac{a_{1n}+a_{n1}}{2} \\ \frac{a_{21}+a_{12}}{2} & a_{22} & . & . & \frac{a_{2n}+a_{n2}}{2} \\ . & . & . & . & . \\ . & . & . & . & . \\ \frac{a_{1n}+a_{n1}}{2} & \frac{a_{2n}+a_{n2}}{2} & . & . & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix}$$

$$\Rightarrow$$

$$= \frac{1}{2}\left(a_{11}x_1^2 + a_{22}x_2^2 + \ldots + a_{nn}x_n^2 + (a_{12} + a_{21})x_1x_2 + \ldots + (a_{1n} + a_{n1})x_1x_n + (a_{2n} + a_{n2})x_2x_n + \ldots + (a_{n,n-1} + a_{n-1,n})x_nx_{n-1}\right) \qquad (4.3)$$

Therefore, the two functional are equivalent. Equations (4.2) and (4.3) are equal and the result follows. The necessary condition for the minimization of the two functionals are

$$\nabla f(X) = a^T + AX = a^T + BX = 0 \qquad (4.4)$$

And the respective optimum will be

$$X^* = -aA^{-1} = -aB^{-1} \qquad (4.5)$$

5.0 Numerical results

Minimize

$$f(X) = f_0 + \langle a, X\rangle_H + \frac{1}{2}\langle X, AX\rangle_H$$

and compare the result with the minimization of

$$f(X) = f_0 + \langle a, X\rangle_H + \frac{1}{2}\langle X, BX\rangle_H$$

For the following problems:

***Problem* 5.1:**

$$f_0 = 1 \quad a = (1,1) \quad A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

***Problem* 5.2:**

$$f_0 = 1 \quad a = (1,1,1) \quad A = \begin{pmatrix} 2 & 2 & 3 \\ 2 & 0 & 1 \\ 2 & 4 & 2 \end{pmatrix}$$

***Problem* 5.3**

$$f_0 = 1 \quad a = (1,1,1,1) \quad A = \begin{pmatrix} 1 & 1 & 2 & 3 \\ 2 & 0 & 4 & 5 \\ 3 & 1 & 2 & 0 \\ 0 & 2 & 1 & 3 \end{pmatrix}$$

***Problem* 5.4:**

$$f_0 = 1 \quad a = (1,1) \quad A = \begin{pmatrix} 3 & 1 \\ 3 & 4 \end{pmatrix}$$

***Problem* 5.5**

$$f_0 = 1 \quad a = (1,1,1) \quad A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 4 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

***Problem* 5.6**

$$f_0 = 1 \quad a = (1,1,1,1) \quad A = \begin{pmatrix} 1 & 1 & 1 & 3 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 3 & 0 \\ 0 & 2 & 1 & 3 \end{pmatrix}$$

Table 5.1: The table below presents the optimal results of quadratic functionals with non-symmetric operators and their symmetric transform equivalent:

	Optimal values with A		Optimal values with B	
	*X**	*f*(*X**)	*X**	*f*(*X**))
Problem1	(-1, 1)	1	(0.667, -0.667)	1
Problem2	(0.5, 0, 0)	0.75	(-0.222, 0, -0.222)	0.778
Problem3	(3.753, -1.251, -5.004, 2.502) $\times 10^{14}$	-2.053×10^{13}	(-0.046, -0.148, -0.234, -0.099)	0.737
Problem4	(-0.3333 ,0)	0.833	(-0.25, -0.125)	0.813
Problem5	(-0.75, -0.25, 0.25)	0.625	(-2.8, 0.4, 0.8)	0.2
Problem6	-(0.235, 0.235, 0.176, 0.118)	0.618	(-5, 0, 1, 2)	0

6.0 Conclusion

*X** is the optimum and *f*(*X**) is the optimal value. It can be seen from the tables that the optimal values for non-symmetric B and the symmetric operator A corresponding to the transform B are favourably comparable. The next study will consider the convergence rate of our transform.

References

[1] Abdulrahman, M. K. A.,(2007). *Conjugate Gradient Method for solving non-symmetric matrix operators* (M. Sc. Dissertation).

[2] Andrew P.S. and Chelsea C.W.(1977). *Optimum System Control (Second Edition),* Prentice-Hall, New Jersey.

[3] Enid R. Pinch (1993). *Optimal Control and the Calculus of Variation*, Oxford University Press, New York.

[4] Ibiejugba M. A. et. al. (2003). *Foundation Postgraduate Course on Computational Methods and Application in Optimization*, National Mathematical Centre, Abuja, Nigeria.

[5] John M. H. and Karin M. (2001). *"PDMATRIX"*-Programs to make matrix positive definite.

[6] Jonathan R. S. (1994). *An Introduction to the Conjugate Gradient Method without Agonizing Pain*

[7] Lawrence Hasdorff (1976). *Gradient Optimization and Nonlinear control*, John Wiley & sons, London.

[8] Lipschutz S. (1987). *Theory and Problems of Linear Algebra (Metric Edition)*, McGraw-Hill, inc. Singapore.

[9] Omolehin J. O. (2006). *Further Result on CGM Algorithm*, The Journal of the Mathematical Association of Nigeria, 33 number 2b, pp374-388.

[10] Omolehin J. O., Rauf, K., (2006). *Computational Capability of CGM Algorithm*, Creative Math. & Inf., 15, pp93-106.

[11] Wayne B, Charles R.J. and Pablo T. (1993). *The Real Positive Definite Completion Problem for a Simple Cycle*, Journal of Linear Algebra and its Application 192 pp 3-31.