Journal of the Nigerian Association of Mathematical Physics Volume 9 (November 2005)

On the convergence of the extended conjugate gradient method for discrete optimal control problems (DOCP)

***J. S. Apanapudor,.**
Department of Mathematics and Computer Science
Delta State University, Abraka, Nigeria.
and
F. O. Otunta.
Department of Mathematics,
University of Benin, Benin City, Nigeria

Abstract

One striking factor that attracts problem solvers to use particular algorithm is its convergence behavior. In this paper, the convergence of the iterates generated by the ECGM for DOCP as proposed by [14] ***is examined.***

Keywords: DOCP, Operator, energy norm, iterates.

pp 493 - 500

1.0 Introduction

Since the early fifties when the Conjugate Gradient method was discovered independently by [5] and [21]; and shortly thereafter, when they jointly published what is considered the seminal reference on CGM [6], there have been lots of improvement on this method. The various improvements in literature today are accompanied with convergence analysis or properties of the method of conjugate Gradients. Among these are [1], [2], [9], [10], [11], [14], [15], [16], [19], [21] and [22]. However most of these discussions were convergence estimates, which depend on values, which are mere, guesses [17].

The quest for further improvement on the CGM led to the extension of the conjugate Gradient method to handle optimal control problems. Several authors have directed their efforts, towards the design and analysis of new algorithms. Among these are [7] and [8]. [16] developed a suitable ECGM algorithm for discrete optimal control problems, which is actually a modification of the basic philosophy of the ECGM for continuous problems, due to [8].

In presenting the proof of the convergence of the ECGM algorithm for DOCP [16] considered the function

$$E(z) = < (z - z^*), H(z - z^*)> \qquad (*)$$

From (*) the author asserted that

$$E(z_n) = < (z_n - z^*), H(z_n)> \qquad (**)$$

and from (**) drew the similarity relation that

$$E(z_{n+1}) = <z_{n+1} - z^*, Hz_{n+1}> \qquad (***)$$

A simple comparison of (*) and (**) and (***) shows that the second term in the inner product of (**) and (***) does not conform with that in (*). This has obscured the beauty of this proof.

*Corresponding author

It is therefore the desire of the authors of this paper to present the proof of the convergence of the ECGM algorithm for DOCP using the energy norm in place of the above function and when this is applied it is obvious that presentation looks sharper and better. To achieve our goal therefore, we shall consider the constrained discrete optimal control problems of the type

$$Mimmize\ \ j(x,u) = \sum_{i=0}^{k}\left[X_i^T PX_i\ + U_i^T QU_i\right] \tag{1.1}$$

Subject to $$X_i\ =\ CX_{i-1} + DU_{i-1}\ ,\ i = 1,\ 2,\ldots,k \tag{1.2}$$

where $X = (x_1,\ x_2,\ldots,x_n)^T$, $U = (u_1,\ u_2\ldots u_n)^T$ are the state and control vectors respectively $x_i \in R^n$, $u_i \in R^m$; $P, Q, C\ and\ D$ are constant matrices; with $P\ and\ Q$ symmetric and positive definite. In adopting a penalty optimization technique for the above problem, [16] proposed the Extended conjugate Gradient Method by linking equations (1.1) and (1.2) with the matrix operator H such that

$$\underset{(x,u)}{Min}\ J(x,u) = \underset{(x,u)}{Min}\langle z,\ Hz\rangle_w$$

$$= \underset{(x,u)}{Min}\sum_{i=1}^{k}\left[x_i^T\ Px_i + u_i^T Qu_j + \varphi\langle x_i - Cx_{i-1}Du_{i-1}, x_i - Cx_{i-1} - Du_{i-1}\rangle\right] \tag{1.3}$$

with $z = \left(x_0, x_{1,}x_2,\ldots,x_k,u_0,u_1,u_2,\ldots,u_k\right), where\,\varphi(\varphi > 0)$ is a penalty constant, w is a real Hilbert space and the symbol $\langle *,*\rangle$ is the usual inner product defined on the Hilbert space in the form $\langle x,y\rangle = x^T y$ where $x\,and\ \ y$ are column vectors of same dimension.

We shall examine the sequence of iterates $\{z_i\}$ from the ECGM algorithm for DOCP. The ECGM algorithm for DOCP is as follows:

Step 1: Choose $z_0 = (x_0,\ u_0)^T$ from w, with x_0 given.

Step 2: For $i = 0$,

Compute $$h_i\ = -\ Hz_i \tag{1.4}$$

$$\alpha_i = \frac{\langle Hz_i,\ Hz_i\rangle}{\langle h_i, Hh_i\rangle} \tag{1.5}$$

$$z_{i+1} = z_i\ + \alpha_i h_i \tag{1.6}$$

$$Hz_{i+1} = Hz_i + \alpha_i Hh_i \tag{1.7}$$

$$\beta_i = \frac{\langle Hz_{i+1}, Hz_{i+1}\rangle}{\langle Hz_i, Hz_i\rangle} \tag{2.8}$$

$$h_{i+1} = -\,Hz_{i+1}\ + \beta_i h_i \tag{1.9}$$

Step 3: If $h_{i+1}\ =\ 0\,, or\ \ i = k,$ Go To step 4

Otherwise set $i = i + 1$ and go to step 2

Step 4: Stop and set $z^* = \left(x^*, u^*\right)$.

In section two of this paper, we shall consider some essential tools that will enhance our understanding of this analysis. Some of these tools like the symmetry, and positive definiteness of the matrix operator H have been discussed in [15], [17] and [18].

Sections three and four dwell on the main thrust of this paper and the conclusion respectively.

2.0 Basic concepts and Lemma.

The quadratic form $f(z)$ is simply a scalar, quadratic function of a vector with the form

$$f(z) = \frac{1}{2} z^T Az - b^T z + c \tag{2.1}$$

where A is a matrix, $z \text{ and } b$ are vectors, and c is a scalar constant.

Definition

The gradient of a quadratic form is defined by

$$f'(z) = \begin{bmatrix} \frac{\partial}{\partial z_1} f(z) \\ \frac{\partial}{\partial z_2} f(z) \\ \vdots \\ \frac{\partial}{\partial z_n f(z)} \end{bmatrix} \tag{2.2}$$

The gradient is a vector field for a given point z and indicates the direction of $f(z)$.

Definition

The error $e_i = z_i - z^*$ is a vector that indicates how far z_i is from the solution $z*$. The residual $r_i = b - Az_i$ indicates how far Az_i is we are from the correct value b. We know that $r_i = -Ae_i$ and our thought of the residual should be the error transformed by A or any operator into the same space as b. More importantly, $r_i = -f'(z_i)$ and so we should always think of the residual as the search direction or direction of the gradient descent. We shall henceforth denote it by h_i, the ith search direction of the gradient descent.

Definition

A line search is a procedure such as that in equation (1.6), that chooses α_i to minimizes $f(z)$ along a line. The value of α_i, the step length will ensure decrease of the iterates.

Definition [23]

A sequence $\{z_i\}$ from an iterative scheme is said to converge super linearly if and only if for the Euclidean norm $\| \bullet \|$, we have

$$\lim_{k \to \infty} \frac{\| z_{k+1} - z^* \|}{\| z_k - z^* \|} \to 0 \tag{2.3}$$

as $k \to \infty$.

Definition

The energy norm [3] is defined as $$\| e \|_A = \left(e^T Ae \right)^{1/2} \tag{2.4}$$

where A is the matrix operator. Minimizing $\| e_i \|_A$ is equivalent to minimizing $f(z)$.[3]

Definition [13]

An iterative scheme is R-superlinearly convergent if $\left\|z_n - z^*\right\| \le (C_n)^n \left\|z^* - z_0\right\|$, where $\lim_{n\to\infty} C_n = 0$ and $\{z_n\}$ is the sequence of approximation of the solution z^* from an arbitrary initial approximation z_0.

***Lemma* 1** [14]

Under the conditions of equation (1.3), *the functional inequality*

$$\|h_i\| \; \|Hh_i\| \; \|z_0\| \ge \|hz_i\|^3 \tag{2.5}$$

holds at each minimization step n *of the Extended conjugate Gradient Method for solving Discrete Optimal Control Problems*.

Proof:

Recall from equation (1.9), that

$$h_i = \beta_{i-1}\, h_{i-1} - Hz_i \tag{2.6}$$

By taking the inner product of h_i *and* Hz_i, (2.6) becomes

$$\langle h_i,\, Hz_i\rangle = \langle \beta_{i-1}h_{i-1} - Hz_i\,,\, Hz_i\rangle = \beta_{i-1}\,\langle h_{i-1},\, Hz_i\rangle - \langle Hz_i\,,\, Hz_i\rangle$$

But $\langle h_{i-1}, Hz_i\rangle = h_{i-1}^T\, hz_i = 0$, since h_{i-1} *and* z_i are orthogonal [17] and [20]. Hence

$$\langle h_i,\, Hz_i\rangle = -\langle Hz_i\,,\, Hz_i\rangle = -\|Hz_i\|^2 \tag{2.7}$$

Since H is symmetric [16],

$$\langle h_i\,,\; Hz_i\rangle = \langle Hh_i\,, z_i\rangle \tag{2.8}$$

From equation (1.6), we have

$$z_i = z_0 + \sum_{j=0}^{i-1} \alpha_j h_j \tag{2.9}$$

Put (2.9) into (2.8) and bearing in mind that $\langle Hh_i,\, h_j\rangle = 0,\ \forall\ i \ne j$, we have,

$$\begin{aligned}\langle Hh_i\,, z_i\rangle &= \langle Hh_i, z_0 + \sum_{j=0}^{i-1}\alpha_j h_i\rangle \\ &= \langle Hh_i,\, z_0\rangle + \sum_{j=0}^{i-1}\alpha_j\,\langle Hh_i,\, h_j\rangle \\ &= \langle Hh_i\,,\, z_0\rangle\,. \end{aligned} \tag{2.10}$$

$$\therefore\ \langle Hh_i, z_i\rangle = \langle h_i\,, Hz_i\rangle = \langle Hh_i,\, z_0\rangle \tag{2.11}$$

Next put (2.11) into (2.7), we have $\langle Hh_i\,,\, z_0\rangle = -\|Hz_i\|^2$, which yields,

$$\|Hh_i\|\,\|z_0\| \ge -\|Hz_i\|^2\,. \tag{2.12}$$

Take the norms of equation (2.6)

$$\|h_i\| = \|\beta_{i-1}\, h_{i-1} - Hz_i\| \ge \|\beta_{i-1}\, h_{i-1}\| - \|Hz_i\|$$

$\|h_i\| - \|\beta_{i-1}\, h_{i-1}\| \ge -\|Hz_i\|$. This clearly indicates that

$$\|h_i\| \geq -\|Hz_i\| \tag{2.13}$$

Combining (2.12) and (2.13), we have

$$\|h_i\| \, \|Hh_i\| \, \|z_0\| \geq \|Hz_i\|^3$$

which completes the proof of the lemma.
In the next section, we present a proof similar to that of [18], but firstly we state the theorem.

3.0 The convergence analysis

Theorem

Suppose that the sequence of iterates $\{z_i\}$ is generated by the Extended Conjugate Gradient Method for Discrete Optimal Control Problems with an arbitrarily selected $z_0 \in W$, then the sequence converges and converges R-superlinearly.

Proof

Let $z_i = (x_i, u_i)^T$, where $x_i, and\ u_i$ are the state and control vectors specified in equation (1.2), be the iterates from the ECGM algorithm. Let $z^* = (x^*, u^*)^T$ denote the value of the control vector with its corresponding trajectory at optimum. Using the energy norm, $\|e_{i+1}\|_H^2$ where H is the matrix operator in equation (1.3), we have,

$$\begin{aligned}
\|e_{i+1}\|_H^2 &= (e_{i+1})^T \, He_{i+1} \\
&= (e_i + \alpha_i h_i)^T \, H(e_i + \alpha_i h_i) && (3.1) \\
&= e_i^T He_i + 2\alpha_i h_i^T He_i + \alpha_i^2 h_i^T Hh_i \\
&= \|e_i\|_H^2 + \frac{2\langle Hz_i, Hz_i\rangle}{\langle h_i, Hh_i\rangle}\left(-h_i^T h_i\right) + \left(\frac{\langle Hz_i, Hz_i\rangle}{\langle h_i, Hh_i\rangle}\right)^2 h_i^T Hh_i \\
&= \|e_i\|_H^2 - 2\frac{\langle Hz_i, Hz_i\rangle}{\langle h_i, Hz_i\rangle}\left(h_i^T h_i\right) + \left(\frac{\langle Hz_i, Hz_i\rangle}{\langle h_i, Hh_i\rangle}\right)^2 h_i^T Hh_i && (3.2) \\
&= \|e_i\|_H^2 - 2\frac{\langle Hz_i, Hz_i\rangle}{\langle h_i, Hz_i\rangle}\langle Hz_i, Hz_i\rangle + \frac{\langle Hz_i, Hz_i\rangle^2}{\langle h_i, Hh_i\rangle} \\
&= \|e_i\|_H^2 - \frac{\langle Hz_i, Hz_i\rangle^2}{\langle h_i, Hh_i\rangle} \\
&= \|e_i\|_H^2 \left[1 - \frac{\langle Hz_i, Hz_i\rangle^2}{\langle h_i, Hh_i\rangle \ \langle e_i, He_i\rangle}\right] && (3.3)
\end{aligned}$$

From the comparative idea of equation (1.6) and $e_{i+1} = e_i + \alpha_i h_i$, (3.3) becomes,

$$\|e_{i+1}\|_H^2 = \|e_i\|_H^2 \left[1 - \frac{\langle Hz_i, Hz_i\rangle^2}{\langle h_i, Hh_i\rangle \ \langle z_i, Hz_i\rangle}\right] \tag{3.4}$$

But

$$\langle z_i, Hz_i\rangle = \langle z_0 + \sum_{j=0}^{i-1}\alpha_j h_j , Hz_i\rangle$$

$$= \langle z_0 , Hz_i\rangle + \sum_{j=0}^{i-1}\alpha_j \langle h_j , Hz_i\rangle$$

$$= \langle z_0, Hz_i\rangle \text{ since } \langle h_j , Hz_i\rangle = 0 \quad \forall i \neq j . \tag{3.5}$$

Also $\langle z_0, Hz_i\rangle \leq /\langle z_0, Hz_i\rangle/$ holds trivially for all z_0 and Hz_i and by the Schwarz's inequality,

$$\langle z_0 , Hz_i\rangle \leq \|z_0\| \; \|Hz_i\| \tag{3.6}$$

Using (3.5) and (3.6) in (3.4), we have

$$\|e_{i+1}\|_H^2 \leq \|e_i\|_H^2 \left[1 - \frac{\langle Hz_i , Hz_i\rangle^2}{\langle h_i, Hh_i\rangle \; \|Hz_i\| \; \|z_0\|}\right]$$

$$= \|e_i\|_H^2 \left[1 - \frac{\|Hz_i\|^2}{\|h_i\| \; \|Hh_i\| \; \|Hz_i\| \; \|z_0\|}\right]$$

$$= \|e_i\|_H^2 \left[1 - \frac{\|Hz_i\|^3}{\|h_i\| \; \|Hh_i\| \; \|z_0\|}\right] \tag{3.7}$$

$$\|e_{i+1}\|_H^2 \leq \|e_i\|_H^2 \; \psi^2 \tag{3.8}$$

if
$$\psi^2 = 1 - \frac{\|z_i\|^3}{\|h_i\| \; \|Hh_i\| \; \|z_0\|} \tag{3.9}$$

$$\frac{\|e_{i+1}\|_H}{\|e_i\|_H} \leq \psi$$

$$\frac{\left\|z_{k+1} - z^*\right\|}{\left\|z_k - z^*\right\|} \leq \psi \tag{3.10}$$

Taking limit of (3.10) $\underset{k\to\infty}{Lim} \sup \frac{\left\|z_{k+1} - z^*\right\|}{\left\|z_k - z^*\right\|} \to 0$ as $k \to \infty$, the LHS tends to zero if $\frac{\|Hz_i\|^3}{\|h_i\| \, \|Hh_i\| \, \|z_0\|} \leq 1$. This shows that the sequence $\{z_i\}$ of iterates generated from the ECGM for DOCP converges super linearly.

Furthermore from (3.8) and (3.9) we have $\|e_i\| \leq \psi_i \; \|e_0\|$. By setting $C_i - \min\psi_i$ we can see that $(c_i)^i \leq \underset{j=0}{\overset{i-1}{\pi}} \psi_j$ such that

$$\|e_i\| \le (c_i)^i \|e_0\| \qquad (3.11)$$

The status of equation (3.9) from the foregoing it is obvious the $\frac{\|Hz_i\|^3}{\|h_i\|\,\|Hh_i\|\,\|z_0\|} \le 1$ holds for each step n of the ECGM algorithm. Thus with large $i, c_i \to 0$ *ie* $\lim_{i\to\infty} C_i = 0$ the end of the proof of the theorem.

4.0 Conclusion

We have in this paper through the lemma and equation (2.4) shown that the sequence $\{z_i\}$of iterates generated by the ECGM algorithm for DOCP converges superlinearly to z^* the optimum value. We have also shown that the ECGM algorithm converges R–superlinearly using the energy norm. The proof obtained through this approach is sharper and better than that by [18].

References

[1] Dimitri, B. P. and Tsitsiklis, J. N. (1997) Gradient Convergence in Gradient Methods. Lab. for info. and Decision Systems Report LIDS – P -2404,Massachusetts Institute of Technology Cambridge.

[2] Crowder, H. P. and Wolfe, P. (1972): Linear Convergence of the Conjugate Gradient Methods IBM.J. of Research and Development No. 16 pp. 431-433.

[3] Griffel, D. H. (1993). Applied functional Analysis Kills Horwood New York.

[4] Golub, G. H. and Van Loan (1989): Matrix Computation. The John-Hopkins University.

[5] Hestenes, M. R. (1973): Iterative Methods for Solving Linear Equation, Journal of Optimization Theory and Applications (1973) Originally Published in 1951 as NAML Report Washington D. C.

[6] Hestenes, M. R. and Stiefel, E. (1952): Methods of Conjugate Gradients for Solving Linear systems, J. of Research of the National Bureau of Standards 49, pp. 409 – 436.

[7] Ibiejugba, M. A. (1980) Computing methods in Optimal Control, Ph.D Thesis in University of Leeds, Leeds England.

[8] Ibiejugba, M. A. and Onumanyi, P. (1984) On a Control Operator and some of its applications J. of Mathematical Analysis and Applications Vol. 103 pp. 31-47.

[9] Kaniel, S, (1966) Estimates for some Computational techniques in Linear Algebra, Mathematics of Computation.20, pp. 369 – 378.

[10] Luenberger, D. G. (1973) Introduction to Linear and Nonlinear Programming Addison-Wesley.

[11] Mahony, R., Absil, Pierre-Antoine, and Andrews, B. (2004) Convergence of the iterates of Descent Methods for Analytic cost functions Austrilian National University Press.

[12] Oliveira, I. B. (2002) A "HUM" Conjugate Gradient Algorithm for Constrained Non-Linear Optimal Control: Terminal and Regulator Problems. A Ph.D Thesis at Massachusetts Institute of Technology.

[13] Ortega, J. M. and Rheinboldt, W. C. (1970) Iterative solution of Non-Linear Equations in several Variables, Academic Press. New York.

[14] Otunta, F. O. (1991) Optimization Techniques for a Class of Regulator Problems A Ph.D. Thesis University of Ilorin Nigeria.

[15] Otunta, F. O. (1996) The Construction of a Control Operator for DOCP. Spectrum Journal Vol. 3 No 1 pp. 82-88.

[16] Otunta, F. O. (1999) An ECGM Algorithm for DOCP J. of Nig. Association of Maths-Physics Vol. 3; pp. 75-88.

[17] Otunta, F. O. (2001) A Convergence estimate for the ECGM. Abacus J. of MAN. Vol. 28, No. 2 pp. 50-57.

[18] Otunta, F. O. (2003) On the Convergence of the ECGM for DOCP. Abacus J. of MAN. Vol. 30 No. 2A pp. 74-80.

[19] Polak, E. (1971) Computational Methods in Optimization Academic Press London.

[20] Shewchuk, J. R. (1994) An Introduction to the Conjugate Gradient Method without Agonizing Pain, Carnegie Mellon University Pittsburgh PA 15213.

[21] Powell, M. J. D. (1976) Some Convergence Properties of the Conjugate Gradient method Mathematical Programming Vol. 11. Pp. 42-49.

[22] Stiefel, E. (1952) Uber einigie Methoden Relaxation-srechnung Zeitschrift fur Anaewandte Mathematik and Physik, no. pp. 1 - 33.

[23] Van der Sluis and Van der Vorst (1986). The rate of convergence of Conjugate Gradients Numerische Methematik, no. 5, pp. 543 - 560.

[24] Wolfe, M.A.(1978): Numerical Methods for Unconstrained Optimization, Van Nostrand Reinhold Company New York.