Journal of the Nigerian Association of Mathematical Physics
Volume 10 (November 2006), 355 -362

Hessian Spectrum to perturbation factor for gradient method algorithm

J. O. Omolehin
Department of Mathematics, University of Ilorin, Ilorin, Nigeria
e-mail:emolehin-joseph@yahoo.com

Abstract

In this paper, the eigen values of the associated Hessian matrix of our control problem are considered for optimal selection of the perturbation factor (λ) or perturbation parameter for gradient method. The perturbation factor is calculated as an n-dimensional vector as against real number. The numerical results generated compare favorably with the existing works.

Keywords: Hessian; Matrix; Gradient; Spectrum; Minimize
C.R. Categories: G.1.7

1.0 Introduction

In any numerical method the focal point is the Convergence rate. The researcher will continue to find a simpler way to speed up the Convergence. Towards this end, this paper considers the Convergence rate of Gradient Method due to Rushell. For detail see [3]. The convergence rate of the method is greatly influenced by the choice of the perturbation term (λ) . Rushell [3] guessed the perturbation arbitrary and the rate of convergence was not consistent. In an attempt to solve this problem of inconsistency, he (Rushell) proposed some complicated hypothesis. However the problem still remained unsolved. Once the problem of this perturbation parameter is solved we are done. The method proposed is very easy to handle.

To solve this problem of convergence rate, this work used the principles of the Conventional Conjugate Gradient Method (CGM) algorithm to calculate the value of the perturbation parameter judiciously.

The conventional CGM algorithm was originally developed by Hestenes and Stiefel [6]. The method is very easy, elegant and precise. It was developed to handle problems in quadratic form. It has a well worked out theory and beautiful convergence profile. That is why some of its properties are being used in this work by the author.

Let us temporarily confine ourselves to the case when the conjugate gradient method is used for quadratic functional minimization.

We consider a quadratic functional of the form $$f(x) = f_0 + \langle a, x \rangle_H + \langle x, Ax \rangle_H \quad (1.1)$$

where A is an nxn symmetric, positive definite operator on the Hilbert Space H , f_0 is a constant and a is a vector in H . The CGM algorithm is described as follows:

If $H \equiv R^n$

Step 1

The first element $x_0 \in H$ of the descent sequence is guessed, while the remaining members of the sequence are computed with the aid of the formulae:

Step 2

$$p_0 = -g_0 = -(a + Ax_0) \quad (1.2)$$

(p_0 is the descent direction, g_0 is the gradient of $f(x)$ at $x = x_0$)

Step 3

$$x_{i+1} = x_i + \alpha_i p_i, \quad \alpha_i = \frac{\langle g_i, g_i \rangle_H}{\langle p_i, p_i \rangle_H} \quad (1.3)$$

(α is the step length)

$$g_{i+1} = g_i + \alpha_i A p_i \quad (1.4)$$

$$p_{i+1} = -g_{i+1} + \beta_i \beta i, \quad \beta_i = \frac{< g_{i+1}, g_{i+1} >}{< g_i, g_i >} \qquad (1.5)$$

Step 4

If $g_i = 0$ for some i terminate the algorithm, else set $i = i+1$ and go to step 3. It has been established in [1] that the CGM will converge in at most n iterations. The rate of convergence was given [1] as

$$E(x_n) \le \left\{ \frac{1 - \frac{m}{M}}{1 + \frac{m}{M}} \right\}^{2n} E(x_0) \qquad (*)$$

2.0 Monotonic convergence rate of CGM Algorithm

The monotonic convergence rate has been established and it will be given here as a theorem. For details see [1].

***Theorem* 2.1**

Under the conditions of problem (1.1) ehe convergence process of CGM algorithm is monotonic in the sense that $\left\| x_{i+1} - x_{i+j} \right\|^2 \le \left\| x_i - x_{i+j} \right\|^2$ (**)

Proof:

We know that $x_{k+1} = x_k + \alpha_k p_k,\ k = 0,1,...$ (2.1

from (6) we obtain the following set of equations:

$x_1 - x_0 = \alpha_0 p_0$, when $k = 0$

$x_2 - x_1 = \alpha_1 p_1$, when $k = 1$

$x_3 - x_2 = \alpha_2 p_2$ when $k = 2$

$\cdots \quad \cdots \quad \cdots \quad \cdots$

$\cdots \quad \cdots \quad \cdots \quad \cdots$ (2.2)

$x_{i-1} - x_{i-2} = \alpha_{i-2} p_{i-2}$, when $k = i-2$, $x_i - x_{i-1} = \alpha_{i-1} p_{i-1}$, when $k = i-1$ the addition of the system of equation (2.2) yields $x_i - x_0 = \sum_{k=0}^{i=1} \alpha_k p_k$ (2.3)

In the same manner as equation (2.3) is derived we can easily establish the following relationship:

$$x_{i+j} - x_i = \sum_{k=l}^{i+j-1} \alpha_k p_k \qquad (2.4)$$

From (2.6) we have $x_{i+1} = x_i + \alpha_i p_i$ (2.5)

Subtracting x_{i+j} from both sides of equation (2.5) we get

$$x_{i+1} - x_{i+j} = x_i + \alpha_i p_i - x_{i+j} \qquad (2.6)$$

Henceforth, to conserve space and for notational convenience we shall drop the subscript H on inner product and norm symbols: from (2.6), we see that

$$\left\| x_{i+1} - x_{i+j} \right\|^2 = \left\| x_i - x_{i+j} + \alpha_i p_i \right\|^2 = \left\| x_i - x_{i+j} \right\|^2 + \alpha_i^2 \left\| p_i \right\|^2 + 2\alpha_i < p_i, x_i - x_{i+j} >$$

$$= \left\| x_i - x_{i+j} \right\|^2 + \alpha_i^2 \left\| p_i \right\|^2 + 2\alpha_i < p_i, -\sum_{k=i}^{i+j-1} \alpha_k p_k >,$$

by virtue of (2.4) $= \left\| x_{i+1} - x_{i+j} \right\|^2 = \left\| x_i - x_{i+j} \right\|^2 + 2\alpha_i \left\| p_i \right\|^2 - 2\alpha_i < p_i, \sum_{k=i}^{i+j-1} \alpha_k p_k >$ (2.7)

To prove our desired result we need to show that

$$\alpha^2 \left\| p_i \right\|^2 - 2\alpha_i < p_i, -\sum_{k=i}^{i+j-1} \alpha_k p_k > = -\gamma \qquad (2.8)$$

where γ is a non-negative quantity. Towards this objective, we proceed as follows:

$$p_0 = -g_0$$

$$p_1 = -g_1 + \beta_0 \; p_0 = -g_1 - \frac{< g_1, g_1 >}{< g_0, g_0 >} g_0$$

$$p_2 = -g_2 + \beta_1 p_1 = -g_2 + \frac{< g_2, g_2 >}{< g_1, g_1 >}\left(-g_1 - \frac{< g_1, g_1 >}{< g_0, g_0 >} g_0 \right) - g_2 - \frac{< g_2, g_2 >}{< g_1, g_1 >} g_1 - \frac{< g_2, g_2 >}{< g_0, g_0 >} g_0$$

and $p_3 = -g_3 - \frac{< g_3, g_3 >}{< g_1, g_1 >} g_1 - \frac{< g_2, g_2 >}{< g_0, g_0 >} g_0$

From which we can see the recursion relationship

$$p_k = - < g_k, g_k > \sum_{i=0}^{k} \frac{g_i}{< g_i, g_i >} \tag{2.9}$$

Now, $< p_i, p_{i+1} > = < - < g_i, g_i > \sum_{j=0}^{i} \frac{g_i}{< g_i, g_i >}, \; - < g_{i+1}, g_{i+1} > \sum_{j=0}^{i+1} \frac{g_j}{< g_j, g_j >}$

$$= < g_i, g_i >< g_{i+1}, g_{i+1} >< \sum_{j=0}^{i} \frac{g_j}{< g_j, g_j >}, \sum_{j=0}^{i+1} \frac{g_j}{< g_j, g_j >} >$$

$$= < g_i, g_i >< g_{i+1}, g_{i+1} >< \frac{g_0}{< g_0, g_0 >} + \frac{g_1}{< g_1, g_1 >} + \ldots + \frac{g_i}{< g_i, g_i >},$$

$$\frac{g_0}{< g_0, g_o >} + \frac{g_1}{< g_1, g_1 >} + \ldots + \frac{g_i}{< g_i, g_i >} + \frac{g_{i+1}}{< g_{i+1}, g_{i+1} >} >$$

$$= < g_i, g_i >< g_{i+1}, g_{i+1} > \left[\frac{< g_0, g_0 >}{< g_0, g_0 >^2} + \frac{< g_1, g_1 >}{< g_1, g_1 >^2} + \ldots + \frac{< g_i, g_i >}{< g_i, g_i >^2} \right]$$

Since $< g_i, g_j > = 0 \quad i = j \rightarrow$ a property of inner product. Similarly,

$$< p_i, p_{i+2} > = < g_i, g_i > \theta_{i+2} \left[\frac{< g_0, g_0 >}{< g_0, g_0 >^2} + \ldots + \frac{< g_1, g_1 >}{< g_1, g_1 >^2} + \ldots + \frac{< g_i, g_i >}{< g_i, g_i >^2} \right]$$

where $\theta_{i+2} = < g_{i+2}, g_{i+2} >$

$$< p_i, p_{i+j-1} > = < g_i, g_i >< g_{i+j-1}, g_{i+j-1} > \left[\frac{< g_0, g_0 >}{< g_0, g_0 >^2} + \frac{< g_1, g_1 >}{< g_1, g_1 >^2} + \ldots + \frac{< g_i, g_i >}{< g_i, g_i >^2} \right] \tag{2.10}$$

Expression (2.9) can then be written in the following ways

$$\alpha_i^2 \| p_i \|^2 - 2\alpha_i < p_i, \sum_{k=i}^{i+j-1} \alpha_k p_k > = \alpha_i^2 \| p_i \|^2 - 2\alpha_i < p_i, \alpha_i p_i + \alpha_{i+1} p_{i+1} + \ldots + \alpha_{i+j-1}, p_{i+j-1} > = \alpha_i^2 \| p_i \|^2 - 2\alpha_i^2 \| p_i \|^2$$

$$-2\alpha_i, \alpha_{i+1} < p_i, p_{i+1} > - \ldots - 2\alpha_i, \alpha_{i+j-1} < p_i, p_{i+j-1} >$$

$$= \Big\{ \alpha_i^2 \| p_i \|^2 - \big[2\alpha_i, \alpha_{i+1} + 2\alpha_i, \alpha_{i+2} + \ldots + 2\alpha_i, \alpha_{i-1} \big] \left[\frac{< g_0, g_0 >}{< g_0, g_0 >^2} + \right.$$

$$\left. + \frac{< g_1, g_1 >}{< g_1, g_1 >^2} + \ldots + \frac{< g_i, g_i >}{< g_i, g_i >^2} \right] \| p_i \|^2 \Big\} = -\gamma \tag{2.11}$$

which is our contention since, by virtue of equations (1.4) and (1.5) each of the components within the chain brackets in (2.11) is non-negative due to the positive definiteness of operator A satisfying (1.3). See Ref.[1]. We can then rewrite (2.7) in the equivalent form

$$\| x_{i+1} - x_{1+j} \|^2 = \| x_i - x_{i+j} \|^2 - \Big\{ \alpha_i^2 \| p_i \|^2 + \big[2\alpha_{i+1} + 2\alpha_i, \alpha_{i+2} + \ldots + 2\alpha_i, \alpha_{i+j-1} \big] \left[\frac{< g_1, g_1 >}{< g_1, g_1 >^2} + \ldots + \frac{< g_i, g_i >}{< g_i, g_i >^2} \right] \| p_i \|^2$$

from where it is clear that: $\|x_{i+1} - x_{i+j}\|^2 \leq \|x_i - x_{i+j}\|^2$, which is the desired result.

We have carefully recalled the Monotonic Convergence property of the conventional gradient method algorithm. Based on properties $(*)$ and $(**)$ we are going to use spectrum analysis by determining the eigen values associated with the Hessian matrix of the problem to consider using Gradient Method. The eigen values will be used as perturbation term for our Gradient Method. In this case our perturbation factor will be a vector of n-dimension as against real positive number used by Rushell [3].

3.0 Gradient Method.

Let us consider the quadratic function $f : IR^n \to R$ which is continuously differentiable in some domain $D \underline{C} IR^n$ and it is assumed that f assumes a local minimum value in D at a point $x \in D_o$ where D_o is the interior of D .Now considering the Taylor's series expansion [3].

$f\left(x - \lambda \frac{\partial f}{\partial x}(x)^T\right) = f(x) + \frac{\partial f}{\partial x}(x)\left(-\lambda \frac{\partial f}{\partial x}(x)^T\right) + \theta(\lambda) = f(x) - \lambda \left\| \frac{\partial f}{\partial x}(x) \right\|_e^2 + \theta(\lambda)$. Where e is the usual Euclidian space, $\|\cdot\|$ is a norm in e. It is assumed that $\frac{\partial f}{\partial x}(x) \neq 0$, 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 IR^n by the recursion equation $x_{k+1} = x_k - \lambda \frac{\partial f}{\partial x}(x)^T, k = 0,1,2,\ldots$

We make an initial guessed value for x_o $[3, pg, 345 - 293]$. The convergence rate of GM has been extensively considered but the associated problem with GM (gradient method) the convergence rate is not stable for the problems considered in [3]. Different values of λ show different types of convergence profiles. There could be a situation where by one of the components of the vector might be converging consistently and rapidly, the other components might not show any pattern of convergence at all. These are the problems we are set to solve with a view to finding the optimal parameter λ that will make the convergence rate of the method more stable. The spectrum analysis of the control operator A in the Conjugate Gradient Method (CGM) algorithm due to Ibiejugba [1] will be employed in this study.

4.0 Hessian approach to spectrum analysis of GM for quadratic functional.

***Theorem* 4.1**

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 eigen values of the control operator A respectively.

Proof:

Recall the problem of the minimization of $f : IR^n \to R$ Given by

$$f(x) = f_o + \langle a, x\rangle_H + \tfrac{1}{2}\langle x, Ax\rangle_H \qquad (4.1)$$

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 (1.1) to obtain

$$\frac{\partial f(x)}{\partial x} = a + Ax \qquad (4.2)$$

Consider

$$x_{k+1} = K_k - \lambda \frac{\partial f(x)^T}{\partial x} \qquad (4.3)$$

Substitute equation (1.2) in (1.3), to obtain

$$x_{k+1} = x_k - \lambda A x_k - \lambda a = (I - \lambda A)x_k - \overline{\lambda} \text{ (} \lambda a = \overline{\lambda} \text{ Constant).} \qquad (4.4)$$

It has been established [3] that there exists an orthogonal matrix p which diagonalizes A, i.e.

$$P^{-1}AP = P^{T}AP \ \text{diag}(\lambda, \lambda, ..., \lambda_n) \quad (4.5)$$

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

$$x_k = p_{yk}, k = 0,1,2,3... \quad (4.6)$$

Therefore $$P_{k+1} = (I - \lambda\Lambda)P_{\succ k} \quad (4.7)$$

(Without loss of generality the constant $\overline{\lambda}$ can be dropped.

$$y_{k+1} = (I - \lambda pAP^{-1})yk = \ \text{diag}(1-\lambda\lambda_1, -\lambda\lambda_2, 1-\lambda\lambda_3, ...1-\lambda\lambda_n y_n) \quad (4.8)$$

$$y^{j}{}_{k+1} = (1-\lambda\lambda_j)^k y_o{}^j, \ \ j = 1,2,...,n \quad (4.9)$$

Thus $$y^{j}{}_{k} = (1-\lambda\lambda_j)^k y_o{}^j \quad j = 1,2,...n \quad (4.10)$$

Therefore $y_k{}^j$ form a geometric progression $y_o{}^j, (1-\lambda\lambda_j)y_o{}^j, (1-\lambda\lambda_j)^2 y_o{}^j, ...$

The rate at which the numbers approach the minimum is dependent on

$$|1-\lambda\lambda_j| \quad (4.11)$$

Ibiejugba et al. established the convergence rate of CGM as

$$\frac{E(x_n)}{E(x_o)} \leq \frac{1-\frac{m}{M}}{1+\frac{m}{M}}, \quad (4.12)$$

Where m and M smallest and largest Eigen values of matrix A respectively. Equate RHS of (4.12) TO (4.11) and simplify. Therefore, we have

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

$$\frac{M-m}{M+m} - 1 = -\lambda\lambda_j \quad (4.13)$$

$$\frac{2m}{\lambda_j} = \lambda \quad (4.14)$$

We take $\lambda_j = 2M$ (for some *j*) and equation (4.14) becomes

$$\lambda = \frac{m}{M} \quad (4.15)$$

Therefore our theorem has been proved.

5.0 Main Result

We now make use of these two results of convergence profile of CGM algorithm:

$$E(x_n) \leq \left\{\frac{1-\frac{m}{M}}{1+\frac{m}{M}}\right\}^{2n} E(x_0) \quad (*)$$

$$\|x_{i+1} - x_{i+j}\|^2 \leq \|x_i - x_{i+j}\|^2 \quad (**)$$

Let us consider a particular problem. Observe that maximization of f is the minimization of $-f$.

Problem:

$$\text{Maximize } f(x,y) = \frac{x}{1+x+x^2} + \frac{\left(y-\frac{y^2}{20}\right)\left(x+\frac{1}{2}\right)}{1+x+x^2}$$

(The exact solution to this problem is $x = 0.5$ and $y = 10$)

Solution:

Let us compute the Hessian matrix as follows:

$$H = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$$

$$\frac{\partial f}{\partial y} = \frac{\left(x + \frac{1}{2}\right)\left(1 - \frac{y}{10}\right)}{1 + x + x^2}$$

$$\frac{\partial f}{\partial x} = \frac{\left(1 - x^2\right) + \left(y - \frac{y^2}{20}\right)\left(\frac{1}{2} - x - x^2\right)}{\left(1 + x + x^2\right)^2}$$

$$H_{12} = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = \frac{1}{1 + x + x^2}\left[1 - \frac{y}{10}\right] - \frac{2x + 1}{\left(1 + x + x^2\right)^2}\left[\left(x + \frac{1}{2}\right)\left(1 - \frac{y}{10}\right)\right] = H_{21}$$

$$\frac{\partial^2 f}{\partial y^2} = -\frac{1}{10}\left[\frac{x + \frac{1}{2}}{1 + x + x^2}\right] = H_{22}$$

$$\frac{\partial^2 f}{\partial x^2} = -\frac{(1+2x)}{1+x+x^2}\left[1 + y - \frac{y^2}{10}\right] - \frac{2x+1}{\left[1+x+x^2\right]^2}\left[1 + \left(y - \frac{y^2}{20}\right)\right] - \left[x + \left(y - \frac{y^2}{20}\right)\left(x + \frac{1}{2}\right)\right]\left[\frac{-6\left(x + x^2\right)}{\left(1+x+x^2\right)^3}\right] = H_{11}$$

To determine the eigen value we proceed as follows:

Let $H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$, $H_\lambda = \begin{vmatrix} H_{11} - \lambda & H_{12} \\ H_{21} & H_{22} - \lambda \end{vmatrix} = 0$

$\Rightarrow (H1_1 - \lambda)(H_{22} - \lambda) - H_{21}H_{12} = 0$

$H_{11}H_{22} - \lambda(H_{11} + H_{22}) + \lambda^2 - H_{21}H_{12} = 0, \quad \lambda^2 - \lambda(H_{11} + H_{22}) + H_{11}H_{22} - H_{21}H_{12} = 0$

Let a =1, b $= -(H_{11} + H_{22})$, c $= H_{11}H_{22} - H_{21}H_{12}$

Therefore $a\lambda^2 + b\lambda + c = 0$

We can now apply quadratic formula to solve for λ. .A program was written in FORTRAN for that purpose.

In the next section we are going to report and analyze our numerical results.

6.0 Numerical Example

We now consider some numerical examples. We are going to focus on the role played by the choice of the initial values and the value of λ . The convergence profile of our minimizing vectors will be given in the table that follows.

Problem:

Maximize $f(x, y) = \dfrac{x}{1 + x + x^2} + \dfrac{\left(y - \frac{y^2}{20}\right)\left(x + \frac{1}{2}\right)}{1 + x + x^2}$ (Analytic solution is given as $x = 0.5, y = 10$)

Table 1: Summary of convergence profile when λ is a real number

Case	Initial values	Minimizing vector	Iteration number
1	$x_o = 1.0$ $y_o = 0.5$ $\lambda = 1.0$	$x = -$ $y = 10.0$	155 155
2	$x_o = 1.0$ $y_o = 0.5$ $\lambda = 0.5$	$x = 0.453$ $y = 9.99$	261 261
3	$x_o = 1.0$ $y_o = 0.5$ variable $\lambda =$ 48.9514	$x = -$ $y = 10.0$	160 160
4	$x_o = 0.6$ $y_o = 1.045499$ variable $\lambda = 0.1228$	$x = 0.46$ $y = 10.0$	91 91
5	$x_o = 0.6$ $y_o = 1.045499$,variable $\lambda = 0.5$	$x = 0.453$ $y = 9.99$	106 221

Table 2: Summary of convergence profile when λ is calculated as a vector

Case	Initial values	Minimizing vector	Iteration Number
1	$x_0 = 1.0$ $y_0 = 0.5$	**Constant** (λ) $x = -0.061$, $\lambda x = 1.25 \times 10^{-1}$ $y = 1.02$ $\lambda y = 1.92 \times 10^{-1}$	105 105
2	$x_0 = 1.0$ $y_0 = 0.5$	**Variable** (λ) $x = 0.5$, $\lambda x = 1.07 \times 10^{-3}$ $y = 6.97$, $\lambda y = 6.77 \times 10^{-2}$ $x = 0.453$, $\lambda x = 5.75 \times 10^{-2}$ $y = 10.0$, $\lambda y = 1.99 \times 10^{-1}$	201 201 1431 1431
3	$x_0 = 0.6$ $y_0 = 1.045499$	**Variable** (λ) $x = 0.5$, $\lambda x = 1.06 \times 10^{-3}$ $y = 7.0$, $\lambda y = 6.7 \times 10^{-2}$ $x = 0.476$, $\lambda x = 1 \times 10^{-5}$ $y = 10.0$, $\lambda y = 5.7 \times 10^{-2}$	183 183 2179 2179

7.0 Analysis of numerical results

Table 1

Since the convergence rate of the Gradient Method depends on the parameter λ and the initial values for updating of our minimizing vector, it is imperative that research should focus emphasis on the optimal selection of these parameters.

In cases 1 and 2 in Table 1 we have the same initial values but different values of λ. When λ = 1.0, the values for x show no sign of convergence but the exact value of y was obtained at the 155th iteration. We now make the value of λ = 0.5 and the exact value of y was obtained at the 261st iteration and the approximate value for x is 0.453 instead of 0.5.

We now calculated the eigen values of the associated Hessian matrix of the problem considered. At each circle of iteration, the value of λ is re calculated through the Hessian matrix operator. In case 3 of Table 1 with initial values $x_o = 1.0$ and $y_o = 0.5$ with judiciously varying λ at each iteration, the exact value of y was obtained at 160th iteration with $\lambda = 48.9514$. However, the convergence of x is discouragingly slow.

When our initial guess is $x_o = 0.6$ and $y_o = 1.045499$, the exact value of $x = 0.5$ and approximate value of $y = 9.99$ was obtained at the 75th iteration. The exact value of $y = 10$ was obtained at the 91st iteration and the approximate value for $x = 0.46$ they are as shown in case 4. If we look at case 5 in Table 1, it is observed that both x and y converge but none of them give the exact value. Their values are $x = 0.453$ and $y = 9.99$ they are acceptable. The value of λ is 0.5 and the values of the initial values are $x_o = 0.6$ and $y_0 = 1.045499$. This is an interesting result and it is better than Russell's [3] results.

Table 2

In case 1 the perturbation is chosen as a constant vector, that is λ does not change at each iteration. The convergence rate is not interesting at all but in case 2 with the same initial values for our minimizing vectors, with variable λ, that is λ is recalculated at each iteration, the exact value of $y = 10$ is obtained at the 1431st iteration. The value of x is 0.453. A close look at case 3 shows more interesting results. The most interesting aspect is that all the variables were descending until convergence occurred. The convergence is monotonic. That is the most interesting aspect of the work. The monotonicity property of CGM algorithm has greatly enhanced the convergence rate of our GM algorithm.

These results are far superior to Rushell result [3] and my earlier results [5]. A FORTRAN program was written for the computation. WATFOR77 was the compiler used.

8. 0 Conclusion

The work considered carefully, the choice of perturbation term (λ) judiciously. That is, the penalty or parameter perturbation term was calculated at each iteration. That is, at each phase of iteration, the method used the current minimizing vector to calculate λ. The convergence rate of the method is far superior to Rushell's result [3] and my earlier result in [5]. The arbitrary selection or calculation of the perturbation term (λ) makes Rushell's work tedious and difficult. See [3].

In this approach, the value of the perturbation λ is very easy to compute. All that is needed to be done is to make an initial guess and the method will automatically recalculate the value of the perturbation term at each face of iteration and it will update itself automatically. The calculation of the perturbation term was calculated as a real number and there is no updating rule. But, in this work, the calculation of the perturbation term is calculated as an n-dimensional.

Finally, the new method is better than the Rushell's result and my earlier work as evident in the tables of values and analysis.

References:

[1]. Ibiejugba, M.A. – Computational methods in optimization, Ph.D. Thesis, University of Leeds, Leeds, U.K., (1985).

[2] . Omolehin, J.O. – On the control of reaction diffusion equation, ph.D. Thesis, University of Ilorin, Ilorin, (1991).

[3]. Russel, David, L. – Optimization theory, New York, W.A. Benjamin, Inc., (1970).

[4] Omolehin, J.O. – Experiment with extended conjugate gradient method algorithm, M.sc. University of Ilorin, Ilorin (1986).

[5]. Omolehin, J.O. – Eigen value perturbation for gradient method, Anaele Stiintifice Ale Universitata "al,L.Cuza, Tomul L I, S.I. Mathematical, f.1, (2005).

[6]. Hestenes, M and Stiefel, E.: Method of Conjugate Gradients for solving linear system, J. Res. Nat. Bus. Standards, Vol.49, pp409-436, 1952.