A HIGHER ORDER ITERATIVE ALGORITHM FOR MULTIVARIATE OPTIMIZATION PROBLEM

SUVRA KANTI CHAKRABORTY* AND GEETANJALI PANDA

ABSTRACT. In this paper a higher order iterative algorithm is developed for an unconstrained multivariate optimization problem. Taylor expansion of matrix valued function is used to prove the cubic order convergence of the proposed algorithm. The methodology is supported with numerical and graphical illustration.

AMS Mathematics Subject Classification : 90C25, 49M15.
Key words and phrases : Matrix Taylor expansion, Kronecker product, Newton's method, Cubic order convergence.

1. Introduction

Classical Newton method is one of the popular gradient based iterative methods and widely used for its quadratic convergence property. In recent years, a lot of research is going on for developing higher order iterative algorithms which are based on the logic of Newton's method, in different areas of numerical computations. Some important higher order iterative methods for finding the root of a nonlinear equation are seen in the literature. Homeier proposed a modification of Newoton method for finding the zero of univariate functions that converges cubically [6, 7]. Kou et al. have proposed a cubic order convergent algorithm for solving nonlinear equations [8] and also some variant of Ostrowski's method with seventh-order convergence [9]. Chun has contributed on the schemes with fourth order convergence and their family [2, 3]. Liu et. al have proposed eighth order method with high efficiency index [10] and Cordero et. al have proposed sixth and seventh order schemes [4] for finding root of univariate nonlinear equation. Employing any of these iterative methods one can optimize a univariate, nonlinear differentiable function more efficiently. However in this paper, an attempt is made to develop a higher order iterative process for optimizing a multivariate function. For developing this scheme, trapezoidal approximation of definite

integral is used and classical Newton method is considered in an implicit form. Theory of Taylor expansion of matrix valued function has helped to establish the convergence of the algorithm. It is proved that the proposed algorithm has cubic order convergence property.

Calculus of matrix valued functions has been widely used in various fields of mathematics. This theory has been developed in several directions by many researchers like Turnbull [12, 13], Dwyer et. al [5], Vetter [14, 15]. Theory of matrix calculus by Vetter [14, 15], which uses Kronecker algebra, is the most popular one for its consistency and completeness. In the later period it has been adopted by researchers from various fields like system theory [1], sensitivity analysis [18], stochastic perturbation [17], statistical models [20, 16]), econometric forecasting [11], neural network [19] etc. In this paper we use the Taylor expansion of a matrix valued function as developed by Vetter [15] to prove the convergence of our algorithm.

Content of this paper is summarized in the following sections. In Section 2, the new scheme is proposed. In Section 3, detailed convergence analysis of the proposed scheme is given. A comparative study between the classical Newton method and the proposed method is discussed in Section 4. Finally, a table with several test functions and a graphical illustration have been given in Appendix.

2. Proposing a new multivariate optimization algorithm

Consider an optimization problem

$(P) \quad \min_{x \in R^s} \ f(x)$ where, $f : \mathbb{R}^s \to \mathbb{R}$ is a sufficiently differentiable function.

Denote $x^n \in \mathbb{R}^s$ as $x^n = (x_1^n, \ x_2^n, \ x_3^n, \ \ldots, \ x_s^n)$ and $\phi(\theta) = \nabla f(x^n + \theta(x - x^n))$. Then $\phi(0) = \nabla f(x^n)$, $\phi(1) = \nabla f(x)$, and $\phi^{'}(\theta) = [\nabla^2 f(x^n + \theta(x - x^n))](x - x^n)$. From fundamental theorem of calculus,

$$\nabla f(x) - \nabla f(x^n) = \int_0^1 \phi^{'}(\theta) d\theta = \int_0^1 \nabla^2 f(x^n + \theta(x - x^n))(x - x^n) d\theta.$$

So

$$\nabla f(x) = \nabla f(x^n) + \int_0^1 \nabla^2 f(x^n + \theta(x - x^n))(x - x^n) d\theta.$$

Hence

$$\frac{\partial f(x)}{\partial x_i} = \frac{\partial f(x^n)}{\partial x_i} + \sum_{j=1}^s \int_0^1 \frac{\partial^2 f(x^n + \theta(x - x^n))}{\partial x_i \partial x_j} (x_j - x_j^n) d\theta$$

(Here $\int$ is used componentwise)

$$\approx \frac{\partial f(x^n)}{\partial x_i} + \frac{1}{2} \sum_{j=1}^s \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x^n) + \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right] (x_j - x_j^n)$$

(Using Trapezoidal approximation).

Hence $\nabla f(x) \approx \nabla f(x^n) + \frac{1}{2}[\nabla^2 f(x) + \nabla^2 f(x^n)]_{s \times s}(x - x^n)$.
Let x^{n+1} be the root of the equation

$$\nabla f(x^n) + \frac{1}{2}[\nabla^2 f(x) + \nabla^2 f(x^n)]_{s \times s}(x - x^n) = 0.$$

Thus,

$$x^{n+1} \approx x^n - 2[\nabla^2 f(x^{n+1}) + \nabla^2 f(x^n)]^{-1}\nabla f(x^n).$$

This is an implicit functional relation in x^{n+1} at x^n. We replace $\nabla^2 f(x^{n+1})$ by $\nabla^2 f(z^n)$, where z^n is the next iteration point, derived by classical Newton method at x^n. Then the new iteration scheme becomes

$$\begin{aligned} x^{n+1} &= x^n - 2[\nabla^2 f(z^n) + \nabla^2 f(x^n)]^{-1}\nabla f(x^n) \\ &\text{where } z^n = x^n - [\nabla^2 f(x^n)]^{-1}\nabla f(x^n). \end{aligned} \quad (1)$$

3. Convergence analysis of the new scheme

To study the convergence analysis of the new scheme (1), following notations and definitions are explained as prerequisites in Subsection 3.1. In the Subsection 3.2, some new definitions and lemmas are introduced which will be used to prove the convergence theorem in Subsection 3.3.

3.1. Prerequisite.
$I_s = s \times s$ dimensional identity matrix.
$\rho(A) =$ Spectral radius of the matrix A .
$A \otimes B =$ Kronecker product of two matrices A and B.
For matrices $A = (a_{ij})_{m \times n}$ and $B = (b_{ij})_{s \times t}$, $A \otimes B = (a_{ij}B)_{ms \times nt}$.
$A^{\times k} = A \otimes A \otimes \ldots \otimes A$ (The kth Kronecker power of A).
$AB =$ Matrix product of two matrices A and B.
For the matrices A, B, C and D the following properties hold.

(P1) $A \otimes (B + C) = A \otimes B + A \otimes C$.
(P2) $(A + B) \otimes C = A \otimes C + B \otimes C$.
(P3) $(kA) \otimes B = A \otimes (kB) = k(A \otimes B)$, k is a scalar.
(P4) $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.
(P5) $(A \otimes B)(C \otimes D) = AC \otimes BD$, matrix dimension must agree to hold the matrix product AC and BD.
(P6) $(A\ B) \otimes I_s = (A \otimes I_s)(B \otimes I_s)$ (This follows from (P5)).

Definition 3.1 (*Matrix function*). A matrix function $\mathscr{A}_{p \times q} : \mathbb{R}^{s \times t} \to \mathbb{R}^{p \times q}$ maps a matrix of $s \times t$ dimension to a matrix of $p \times q$ dimension.

Definition 3.2 (*Matrix derivative* [15]). The derivative structure of a matrix-valued function $\mathscr{A}_{p \times q}(B)$ with respect to a scalar b_{kl} and with respect to the matrix $B_{s \times t}$ are defined as

$$\mathscr{D}_{b_{kl}}\mathscr{A}(B) = \left(\frac{\partial a_{ij}}{\partial b_{kl}}\right) \begin{matrix} i\downarrow = 1,2,3,\ldots,p \\ j\rightarrow = 1,2,3,\ldots,q \end{matrix} = \begin{pmatrix} \frac{\partial a_{11}}{\partial b_{kl}} & \frac{\partial a_{12}}{\partial b_{kl}} & \cdots & \frac{\partial a_{1q}}{\partial b_{kl}} \\ \frac{\partial a_{21}}{\partial b_{kl}} & \frac{\partial a_{22}}{\partial b_{kl}} & \cdots & \frac{\partial a_{2q}}{\partial b_{kl}} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial a_{p1}}{\partial b_{kl}} & \frac{\partial a_{p2}}{\partial b_{kl}} & \cdots & \frac{\partial a_{pq}}{\partial b_{kl}} \end{pmatrix}$$

$$\mathscr{D}_B\mathscr{A}(B) = (\mathscr{D}_{b_{kl}}\mathscr{A}) \begin{matrix} k\downarrow = 1,2,3,\ldots,s \\ l\rightarrow = 1,2,3,\ldots,t \end{matrix} = \left(\begin{array}{c|c|c|c} \mathscr{D}_{b_{11}}\mathscr{A} & \mathscr{D}_{b_{12}}\mathscr{A} & \cdots & \mathscr{D}_{b_{1t}}\mathscr{A} \\ \hline \mathscr{D}_{b_{2l}}\mathscr{A} & \mathscr{D}_{b_{22}}\mathscr{A} & \cdots & \mathscr{D}_{b_{2t}}\mathscr{A} \\ \hline \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \hline \mathscr{D}_{b_{s1}}\mathscr{A} & \mathscr{D}_{b_{s2}}\mathscr{A} & \cdots & \mathscr{D}_{b_{st}}\mathscr{A} \end{array}\right)$$

Higher order derivatives are given as

$$\mathscr{D}_{B^n}^n \ \mathscr{A}(B) = \underbrace{\mathscr{D}_B(\mathscr{D}_B(\ldots(\mathscr{D}_B \ \mathscr{A}(B))\ldots))}_{} \ n-times. \tag{2}$$

Matrix Taylor Expansion: The Taylor expansion structures for a matrix valued function $\mathscr{A}(u)$ of a column vector $u \in \mathbb{R}^s$ about the column vector $\overline{u} \in \mathbb{R}^s$ described in [15] is:

$$\mathscr{A}_{p\times q}(u) = A(\overline{u}) + \sum_{m=1}^{M} \frac{1}{m!}\mathscr{D}_{u^{T^m}}^m A(\overline{u})((u-\overline{u})^{\times m} \otimes I_q) + R_{M+1}(\overline{u}, u), \tag{3}$$

where $R_{M+1}(\overline{u}, u) = \frac{1}{m!}\int_{\xi=\overline{u}}^{u} \mathscr{D}_{u^{T^{m+1}}}^{m+1} A(\xi)(I_s \otimes (u-\xi)^{\times m} \otimes I_q)(d\xi \otimes I_q)$.

3.2. New Definition and Lemmas.

Definition 3.3. Let $f: \ \mathbb{R}^s \to \mathbb{R}$ be a sufficiently differentiable function, gradient of f be ∇f. A function $\widetilde{\nabla}^n f: \ \mathbb{R}^s \to \mathbb{R}^{s\times s^{n-1}}$ is defined as

$$\widetilde{\nabla}^n f(x) = \mathscr{D}_{x^{T^{n-1}}}^{n-1}(\nabla f(x)), \ n = 2,3,4\ldots,$$

where $\mathscr{D}_{x^{T^{n-1}}}^{n-1}$ is as defined in (2). x^T is the row vector $(x_1, x_2, \ldots, x_s)$.

Lemma 3.4. $\widetilde{\nabla}^2 f = \nabla^2 f$.

Proof. This result follows from Definition 3.3 directly. $\qquad\square$

Lemma 3.5. $\mathscr{D}_{x^{T^n}}^n(\nabla^2 f(x)) = \widetilde{\nabla}^{n+2}f(x)$.

Proof.

$$\begin{aligned} \mathscr{D}_{x^{T^n}}^n(\nabla^2 f) &= \mathscr{D}_{x^{T^n}}^n(\widetilde{\nabla}^2 f(x)) \ \text{(using Lemma 3.4)} \\ &= \mathscr{D}_{x^{T^n}}^n \mathscr{D}(\nabla f(x)) \ \text{(using Definition 3.3)} \\ &= \mathscr{D}_{x^{T^{n+1}}}^{n+1}(\nabla f(x)) \ \text{(using (2))} \\ &= \widetilde{\nabla}^{n+2}f(x) \ \text{(using Definition 3.3)} \end{aligned}$$

$\qquad\square$

Lemma 3.6. *If $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$, then $\| a \otimes b \| = \| a \| \| b \|$ where $\| \, . \, \|$ is Euclidean norm.*

Proof. For $a = (a_1, a_2, \ldots, a_m)$ and $b = (b_1, b_2, \ldots, b_n)$,

$$a \otimes b = (a_1 b_1, a_1 b_2, \ldots, a_1 b_n, a_2 b_1, a_2 b_2, \ldots, a_2 b_n, \ldots, a_m b_1, a_m b_2, \ldots, a_m b_n)^T_{nm \times 1}.$$

So

$$\| a \otimes b \| = \Big\{ a_1^2 \left(b_1^2 + b_2^2 + \ldots + b_n^2 \right) + a_2^2 \left(b_1^2 + b_2^2 + \ldots + b_n^2 \right)$$

$$+ \ldots + a_m^2 (b_1^2 + b_2^2 + \ldots + b_n^2) \Big\}^{1/2}$$

$$= \left(\sum_{i=1}^m a_i^2 \right)^{1/2} \left(\sum_{j=1}^n b_j^2 \right)^{1/2} = \| a \| \, . \, \| b \|$$

$\square$

Lemma 3.7. *If $u \in \mathbb{R}^s$, then $\| u^{\times n} \| = \| u \|^n$ for $n \in \mathbb{N}$.*

Proof. For $n = 1$, $\| u^{\times 1} \| = \| u \| = \| u \|^1$.
Suppose $\| u^{\times k} \| = \| u \|^k$ for some k. Then for $n = k + 1$,

$$\| u^{\times k+1} \| = \| u^{\times k} \otimes u \| = \| u^{\times k} \| \, . \, \| u \| \quad \text{(using Lemma 3.6)}$$

$$= \| u \|^k \, . \, \| u \| = \| u \|^{k+1}$$

$\square$

Lemma 3.8. *If $u \in \mathbb{R}^s$, then $(u^{\times n} \otimes I_s)(u^{\times 1} \otimes I_1) = u^{\times (n+1)}$ for $n \in \mathbb{N}$.*

Proof. For $n = 1$, $(u^{\times 1} \otimes I_s)(u^{\times 1} \otimes I_1) = (u^{\times 1} \otimes I_s)u = u^{\times 2}$.
Suppose $(u^{\times k} \otimes I_s)(u^{\times 1} \otimes I_1) = u^{\times (k+1)}$ for some k. Then for $n = k + 1$,

$$(u^{\times (k+1)} \otimes I_s)(u^{\times 1} \otimes I_1) = ((u \otimes u^{\times k})) \otimes I_s)(u^{\times 1} \otimes I_1)$$

$$= u \otimes (u^{\times k} \otimes I_s)(u^{\times 1} \otimes I_1)$$

$$= u \otimes u^{\times (k+1)} = u^{\times (k+2)}.$$

$\square$

3.3. Third order convergence of the algorithm.

Let $\alpha \in R^s$ be the solution of $\nabla f = 0$. Using matrix Taylor expansion (3) about α, $\nabla f(x)$ and $\nabla^2 f(x)$ can be expressed as

$$\nabla f_{s \times 1}(x) = \nabla f(\alpha) + \sum_{m=1}^M \frac{1}{m!} \mathscr{D}_{x^{T^m}}^m \nabla f(\alpha)((x - \alpha)^{\times m} \otimes I_1) + R_{M+1}^1(\alpha, x),$$

$$\text{where } R_{M+1}^1(\alpha, x) = \frac{1}{m!} \int_{\xi = \alpha}^x \mathscr{D}_{b^{T^{m+1}}}^{m+1} \nabla f(\xi)(I_s \otimes (x - \xi)^{\times m} \otimes I_q)(d\xi \otimes I_q);$$

(4)

$$\nabla^2 f_{s\times s}(x) = \nabla^2 f(\alpha) + \sum_{m=1}^{M} \frac{1}{m!} \mathscr{D}_{x^T m}^m \nabla^2 f(\alpha)((x-\alpha)^{\times m} \otimes I_s) + R_{M+1}^2(\alpha, x),$$

$$\text{where } R_{M+1}^2(\alpha, x) = \frac{1}{m!} \int_{\xi=\alpha}^{x} \mathscr{D}_{b^T m+1}^{m+1} \nabla f(\xi)(I_s \otimes (x-\xi)^{\times m} \otimes I_q)(d\xi \otimes I_q).$$

(5)

Using Definition 3.3 and replacing x by x^n, (4) can be rewritten as

$$\nabla f(x^n) = \nabla f(\alpha) + \sum_{m=1}^{M} \frac{1}{m!} \widetilde{\nabla}^{m+1} f(\alpha)((x^n-\alpha)^{\times m} \otimes I_1) + R_{M+1}^1(x^n, \alpha)$$

$$= \sum_{m=1}^{M} \frac{1}{m!} \widetilde{\nabla}^{m+1} f(\alpha)((x^n-\alpha)^{\times m} \otimes I_1) + R_{M+1}^1(x^n, \alpha)$$

(6)

$$\text{(Since, } \nabla f(\alpha) = 0)$$

Using Lemma 3.5 and replacing x by x^n, (5) can be rewritten as

$$\nabla^2 f(x^n) = \widetilde{\nabla}^2 f(\alpha) + \sum_{m=1}^{M} \frac{1}{m!} \widetilde{\nabla}^{m+2} f(\alpha)((x^n-\alpha)^{\times m} \otimes I_s) + R_{M+1}^2(x^n, \alpha). \quad (7)$$

Denote

$$C_m \triangleq \frac{1}{m!} [\widetilde{\nabla}^2 f(\alpha)]^{-1} [\widetilde{\nabla}^{m+1} f(\alpha)] \tag{8}$$

Neglecting remainder terms for large M and using (8) we rewrite (6) and (7) as

$$\nabla f(x^n) = \widetilde{\nabla}^2 f(\alpha) \sum_{m=1}^{M} C_m((x^n-\alpha)^{\times m} \otimes I_1), \tag{9}$$

$$\nabla^2 f(x^n) = \widetilde{\nabla}^2 f(\alpha) \left[I_s + \sum_{m=1}^{M} (m+1) \, C_{m+1}((x^n-\alpha)^{\times m} \otimes I_s) \right] \tag{10}$$

respectively. Now, (10) can be written as

$$\nabla^2 f(x^n) = \widetilde{\nabla}^2 f(\alpha) \, (I_s + B), \text{ where } B = \sum_{m=1}^{M} (m+1) \, C_{m+1}((x^n-\alpha)^{\times m} \otimes I_s). \tag{11}$$

For large n, x^n is in a sufficiently close neighborhood of α such that $\rho(B) < 1$. So from (11),

$$[\nabla^2 f(x^n)]^{-1} = (I_s+B)^{-1}[\widetilde{\nabla}^2 f(\alpha)]^{-1} = (I_s - B + B^2 - B^3 + \dots)[\widetilde{\nabla}^2 f(\alpha)]^{-1}. \tag{12}$$

From (9) and (12), we have

$$[\nabla^2 f(x^n)]^{-1} \nabla f(x^n)$$

$$= (I_s - B + B^2 - \dots)[\widetilde{\nabla}^2 f(\alpha)]^{-1}[\widetilde{\nabla}^2 f(\alpha)] \sum_{m=1}^{M} C_m((x^n-\alpha)^{\times m} \otimes I_1)$$

$$= (I_s - B + B^2 - \dots) \sum_{m=1}^{M} C_m((x^n-\alpha)^{\times m} \otimes I_1)$$

$$= \sum_{m=1}^{M} C_m((x^n - \alpha)^{\times m} \otimes I_1) + \sum_{k=1}^{\infty} (-1)^k B^k \sum_{m=1}^{M} C_m((x^n - \alpha)^{\times m} \otimes I_1)$$

$$= \sum_{m=1}^{M} C_m((x^n - \alpha)^{\times m} \otimes I_1)$$

$$+ \sum_{k=1}^{\infty} (-1)^k \left[\sum_{m=1}^{M} (m+1) \, C_{m+1}((x^n - \alpha)^{\times m} \otimes I_s) \right]^k \sum_{m=1}^{M} C_m((x^n - \alpha)^{\times m} \otimes I_1)$$

$$= \sum_{m=1}^{M} C_m((x^n - \alpha)^{\times m} \otimes I_1)$$

$$+ \left[- \sum_{m=1}^{M} (m+1) \, C_{m+1}((x^n - \alpha)^{\times m} \otimes I_s) \right] \left[\sum_{m=1}^{M} C_m((x^n - \alpha)^{\times m} \otimes I_1) \right]$$

$$+ \left[\sum_{m=1}^{M} (m+1) \, C_{m+1}((x^n - \alpha)^{\times m} \otimes I_s) \right]^2 \left[\sum_{m=1}^{M} C_m((x^n - \alpha)^{\times m} \otimes I_1) \right] + \ldots$$

Expanding each term in the right hand side of the above expression, we have

$$[\nabla^2 f(x^n)]^{-1} \nabla f(x^n)$$
$$= \left[\, C_1((x^n - \alpha)^{\times 1} \otimes I_1) + C_2((x^n - \alpha)^{\times 2} \otimes I_1) + C_3((x^n - \alpha)^{\times 3} \otimes I_1) + \ldots \right]$$
$$+ \left[\{ -2 \, C_2((x^n - \alpha)^{\times 1} \otimes I_s) - 3 \, C_3((x^n - \alpha)^{\times 2} \otimes I_s) - \ldots \} . \right.$$
$$\left. \{ \, C_1((x^n - \alpha)^{\times 1} \otimes I_1) + C_2((x^n - \alpha)^{\times 2} \otimes I_1) + \ldots \} \right]$$
$$+ \left[\{ \left(2 \, C_2((x^n - \alpha)^{\times 1} \otimes I_s) \right) \left(2 \, C_2((x^n - \alpha)^{\times 1} \otimes I_s) \right) + \ldots \} . \right.$$
$$\left. C_1((x^n - \alpha)^{\times 1} \otimes I_1) + \ldots \right] + \ldots$$

Rearranging the terms in the right side of the above expression according to Kronecker power,

$$[\nabla^2 f(x^n)]^{-1} \nabla f(x^n) = \left[\, C_1((x^n - \alpha)^{\times 1} \otimes I_1) \right] +$$
$$\left[C_2((x^n - \alpha)^{\times 2} \otimes I_1) - 2 \, C_2((x^n - \alpha)^{\times 1} \otimes I_s) \, C_1((x^n - \alpha)^{\times 1} \otimes I_1) \right] +$$
$$\left[C_3((x^n - \alpha)^{\times 3} \otimes I_1) - 2 \, C_2((x^n - \alpha)^{\times 1} \otimes I_s) \, C_2((x^n - \alpha)^{\times 2} \otimes I_1) \right.$$
$$- 3 C_3((x^n - \alpha)^{\times 2} \otimes I_s) \, C_1((x^n - \alpha)^{\times 1} \otimes I_1)$$
$$\left. + 4 C_2((x^n - \alpha)^{\times 1} \otimes I_s) \, C_2((x^n - \alpha)^{\times 1} \otimes I_s) \, C_1((x^n - \alpha)^{\times 1} \otimes I_1) \right] + \ldots$$

$C_1 = \frac{1}{1!} \left[\widetilde{\nabla}^2 f(\alpha) \right]_{s \times s}^{-1} \left[\widetilde{\nabla}^2 f(\alpha) \right]_{s \times s} = I_s$. Using Lemma 3.8, putting $C_1 = I_s$ and rearranging the terms according to Kronecker power in the above expression, we get

$$\left[\nabla^2 f(x^n) \right]^{-1} \nabla f(x^n)$$
$$= (x^n - \alpha)^{\times 1} \otimes I_1 + \left[\, C_2((x^n - \alpha)^{\times 2} \otimes I_1) - 2 \, C_2((x^n - \alpha)^{\times 2} \otimes I_1) \right]$$

$$+ \left[\, C_3((x^n - \alpha)^{\times 3} \otimes I_1) - 2 \, C_2((x^n - \alpha)^{\times 1} \otimes I_s) \, C_2((x^n - \alpha)^{\times 2} \otimes I_1) \right.$$

$$\left. -3 \, C_3((x^n - \alpha)^{\times 3} \otimes I_1) + 4 \, C_2((x^n - \alpha)^{\times 1} \otimes I_s) \, C_2((x^n - \alpha)^{\times 2} \otimes I_1) \right] + \dots \quad (13)$$

$$= (x^n - \alpha)^{\times 1} \otimes I_1 + \left[-C_2((x^n - \alpha)^{\times 2} \otimes I_1) \right]$$

$$+ \left[2 \, C_2((x^n - \alpha)^{\times 1} \otimes I_s) \, C_2((x^n - \alpha)^{\times 2} \otimes I_1) - 2 \, C_3((x^n - \alpha)^{\times 3} \otimes I_1) \right] + \dots$$

z^n is the classical Newton iterate at x^n (See (1)). Replacing x^n by z^n in (10), we get

$$\nabla^2 f(x^n) + \nabla^2 f(z^n)$$

$$= \left[\widetilde{\nabla}^2 f(\alpha) \right] \left[2I_s + \sum_{m=1}^{M} (m+1) \, C_{m+1} \left\{ (x^n - \alpha)^{\times m} + (z^n - \alpha)^{\times m} \right\} \otimes I_s \right].$$

Substituting the expression of z^n from (1) in the above expression, we have

$$\nabla^2 f(x^n) + \nabla^2 f(z^n) = \left[\widetilde{\nabla}^2 f(\alpha) \right] \left[2I_s + \sum_{m=1}^{M} (m+1) \, C_{m+1} \left\{ ((x^n - \alpha)^{\times m} + \right. \right.$$

$$\left. \left. (x^n - [\nabla^2 f(x^n)]^{-1} \nabla f(x^n) - \alpha)^{\times m}) \otimes I_s \right\} \right] \quad (14)$$

Denote

$$D \triangleq x^n - [\nabla^2 f(x^n)]^{-1} \nabla f(x^n) - \alpha$$

$$= C_2((x^n - \alpha)^{\times 2} \otimes I_1) - 2 \, C_2((x^n - \alpha)^{\times 1} \otimes I_s) \, C_2((x^n - \alpha)^{\times 2} \otimes I_1) \quad (15)$$

$$+ 2 \, C_3((x^n - \alpha)^{\times 3} \otimes I_1) + \dots$$

and

$$P \triangleq \frac{1}{2} \sum_{m=1}^{M} (m+1) \, C_{m+1} \left\{ ((x^n - \alpha)^{\times m} + D^{\times m}) \otimes I_s \right\}$$

$$= C_2((x^n - \alpha)^{\times 1} + D^{\times 1}) \otimes I_s + \frac{3}{2} \, C_3((x^n - \alpha)^{\times 2} + D^{\times 2}) \otimes I_s$$

$$+ 2 \, C_4((x^n - \alpha)^{\times 3} + D^{\times 3}) \otimes I_s + \dots$$

One may observe that in the expression of D in (15) the lowest Kronecker power of $(x^n - \alpha)$ is 2. As we are writing the terms which produce at most the third kronecker power of $(x^n - \alpha)$, there is no need of writing $D^{\times 2}$ and $D^{\times 3}$ explicitly. After simplifying, expression for P becomes

$$P = C_2 \left\{ (x^n - \alpha)^{\times 1} \otimes I_s \right\} + C_2 \left\{ \, C_2(x^n - \alpha)^{\times 2} \otimes I_s \right\}$$

$$- C_2 \left\{ \, 2 \, C_2((x^n - \alpha)^{\times 1} \otimes I_s) \, C_2((x^n - \alpha)^{\times 2} \otimes I_1) \, \right\}$$

$$+ C_2 \left\{ \, 2 \, C_3(x^n - \alpha)^{\times 3} \otimes I_s \right\} + \frac{3}{2} \, C_3 \left\{ (x^n - \alpha)^{\times 2} \otimes I_s \right\} \quad (16)$$

$$+ 2 \, C_4 \left\{ (x^n - \alpha)^{\times 3} \otimes I_s \right\} + \dots$$

Hence (14) can be expressed as

$$\frac{1}{2}\left[\nabla^2 f(x^n) + \nabla^2 f(z^n)\right] = \left[\widetilde{\nabla}^2 f(\alpha)\right](I_s + P).$$

So small

$$\left[\frac{1}{2}\left(\nabla^2 f(x^n) + \nabla^2 f(z^n)\right)\right]^{-1}\nabla f(x^n)$$

$$= (I_s + P)^{-1}\left[\widetilde{\nabla}^2 f(\alpha)\right]^{-1}\nabla f(x^n)$$

$$= (I_s + P)^{-1}\sum_{m=1}^{M} C_m((x^n - \alpha)^{\times m} \otimes I_1)$$

(Substituting the value of $\nabla f(x^n)$ from (9))

$$= \left(I_s - P + P^2 - \ldots\right)\sum_{m=1}^{M} C_m((x^n - \alpha)^{\times m} \otimes I_1)$$

(For large n, x^n is in a sufficiently close neighborhood of α, so $\rho(P) < 1$.
Hence $(I_s + P)^{-1} = I_s - P + P^2 - \cdots$. Substituting the value of P,

$$\left[\frac{1}{2}\left(\nabla^2 f(x^n) + \nabla^2 f(z^n)\right)\right]^{-1}\nabla f(x^n)$$

$$= \left[\; C_1((x^n - \alpha)^{\times 1} \otimes I_1) + C_2((x^n - \alpha)^{\times 2} \otimes I_1) + C_3((x^n - \alpha)^{\times 3} \otimes I_1)\right] -$$

$$\left[\left\{\; C_2\left((x^n - \alpha)^{\times 1} \otimes I_s\right) + C_2\left(\; C_2(x^n - \alpha)^{\times 2} \otimes I_s\right) + \frac{3}{2}\left(\; C_3((x^n - \alpha)^{\times 2} \otimes I_s)\right)\right\}\right.$$

$$\left.\left\{\; C_1((x^n - \alpha)^{\times 1} \otimes I_1) + C_2((x^n - \alpha)^{\times 2} \otimes I_1)\right\}\right]$$

$$+ \left[\left\{\; C_2((x^n - \alpha)^{\times 1} \otimes I_s)\; C_2((x^n - \alpha)^{\times 1} \otimes I_s)\right\} . \left\{\; C_1((x^n - \alpha)^{\times 1} \otimes I_1\right\}\right]$$

$$+ \text{ Higher Kronecker Power Terms.}$$

$$= C_1((x^n - \alpha)^{\times 1} \otimes I_1) - C_2\left(\; C_2(x^n - \alpha)^{\times 2} \otimes I_s\right) C_1((x^n - \alpha)^{\times 1} \otimes I_1)$$

$$- \frac{1}{2}\; C_3((x^n - \alpha)^{\times 3} \otimes I_1) + \text{ Higher Kronecker Power Terms.}$$

$$= C_1((x^n - \alpha)^{\times 1} \otimes I_1) - \frac{1}{2}\; C_3((x^n - \alpha)^{\times 3} \otimes I_1)$$

$$- C_2(\; C_2 \otimes I_s)((x^n - \alpha)^{\times 2} \otimes I_s)\; C_1((x^n - \alpha)^{\times 1} \otimes I_1)$$

$$+ \text{ Higher Kronecker Power Terms.}$$

(Since from Property (P6), $C_2(x^n - \alpha)^{\times 2} \otimes I_s = (\; C_2 \otimes I_s)((x^n - \alpha)^{\times 2} \otimes I_s))$

$$= C_1((x^n - \alpha)^{\times 1} \otimes I_1) - \frac{1}{2}\; C_3((x^n - \alpha)^{\times 3} \otimes I_1) - C_2(\; C_2 \otimes I_s)((x^n - \alpha)^{\times 3} \otimes I_1)$$

$$= (x^n - \alpha) - \frac{1}{2}\; C_3((x^n - \alpha)^{\times 3} \otimes I_1)$$

$$- C_2(\; C_2 \otimes I_s)((x^n - \alpha)^{\times 2} \otimes I_s)\; C_1((x^n - \alpha)^{\times 1} \otimes I_1)$$

$$+ \text{ Higher Kronecker Power Terms.}$$

$$= C_1((x^n - \alpha)^{\times 1} \otimes I_1) - C_2(\ C_2 \otimes I_s)((x^n - \alpha)^{\times 3} \otimes I_1)$$
$$- \frac{1}{2}\ C_3((x^n - \alpha)^{\times 3} \otimes I_1) + \text{Higher Kronecker Power Terms.}$$

From the iteration scheme (See (1)) :

$$x^{n+1} - \alpha = x^n - \alpha - \left[\frac{1}{2}\left(\nabla^2 f(x^n) + \nabla^2 f(z^n)\right)\right]^{-1} \nabla f(x^n).$$

$$= \frac{1}{2}\ C_3((x^n - \alpha)^{\times 3} \otimes I_1) +\ C_2(\ C_2 \otimes I_s)((x^n - \alpha)^{\times 3} \otimes I_1)$$
$$+ \text{Higher Kronecker Power Terms.}$$

Denote $e_n = x^n - \alpha$. Then,

$$e_{n+1} = \left[\frac{1}{2}\ C_3 + C_2(\ C_2 \otimes I_s)\right] e_n^{\times 3} + \text{Higher Kronecker Power Terms.}$$

Using Lemma 3.7,

$$\parallel e_{n+1} \parallel\ \leq\ \parallel\ \frac{1}{2}\ C_3 +\ C_2(\ C_2 \otimes I_s) \parallel\ .\ \parallel e_n \parallel^3 + O(\parallel e_n \parallel^4).$$

$$\frac{\parallel e_{n+1} \parallel}{\parallel e_n \parallel^3}\ \leq\ \parallel\ \frac{1}{2}\ C_3 +\ C_2(\ C_2 \otimes I_s) \parallel + O(\parallel e_n \parallel).$$

Since $\parallel e_n \parallel\to 0$, for some large n onwards,

$$\frac{\parallel e_{n+1} \parallel}{\parallel e_n \parallel^3}\ \leq\ \parallel\ \frac{1}{2}\ C_3 +\ C_2(\ C_2 \otimes I_s) \parallel + \epsilon,\ \epsilon \text{ is a small positive real number}$$

or, $\dfrac{\parallel e_{n+1} \parallel}{\parallel e_n \parallel^3}\ \leq r,$ where r is a positive real constant.

This implies that the new scheme has third order convergence. Hence the following result holds.

Theorem 3.9. *Let $f : \mathbb{R}^s \to \mathbb{R}$ be sufficiently differentiable function and locally convex at $\alpha \in \mathbb{R}^s$ such that $\nabla f(\alpha) = 0$. Then the algorithm (1), with initial point x^0, which is sufficiently close to α, converges cubically to the local minimizer α of the problem $\min_{x \in \mathbb{R}^s} f(x)$.*

4. Numerical Result

The new algorithm is executed in MATLAB (version- R2013b) and the numerical computations are summerized in Table 1. One may observe that the total number of iterations in proposed method is less than the total number of iterations in classical Newton method. All the steps of one of these test functions are illustrated graphically in Fig. 1, where it is seen that the proposed process reaches more rapidly than the existing process. CNM denotes classical Newton method and PM denotes proposed method. The Table 1 and Fig.1 are provided in the appendix.

5. Conclusion

Several higher order optimization algorithms exist for single dimension optimization problems in the literature of numerical optimization. Newton, Quasi Newton and Conjugate gradient algorithms, which are used for multidimensional optimization problems have second and super linear rate of convergence. This paper has developed a cubic order iterative algorithm for unconstrained optimization problems in higher dimension. Taylor expansion of matrix valued function is the key concept to prove the convergence of the algorithm. Using this logic the reader may extend the present work to develop similar algorithms for order of convergence more than 3. In the process of developing the recurrence relation, trapezoidal approximation is used. However one may try with other type approximations also.
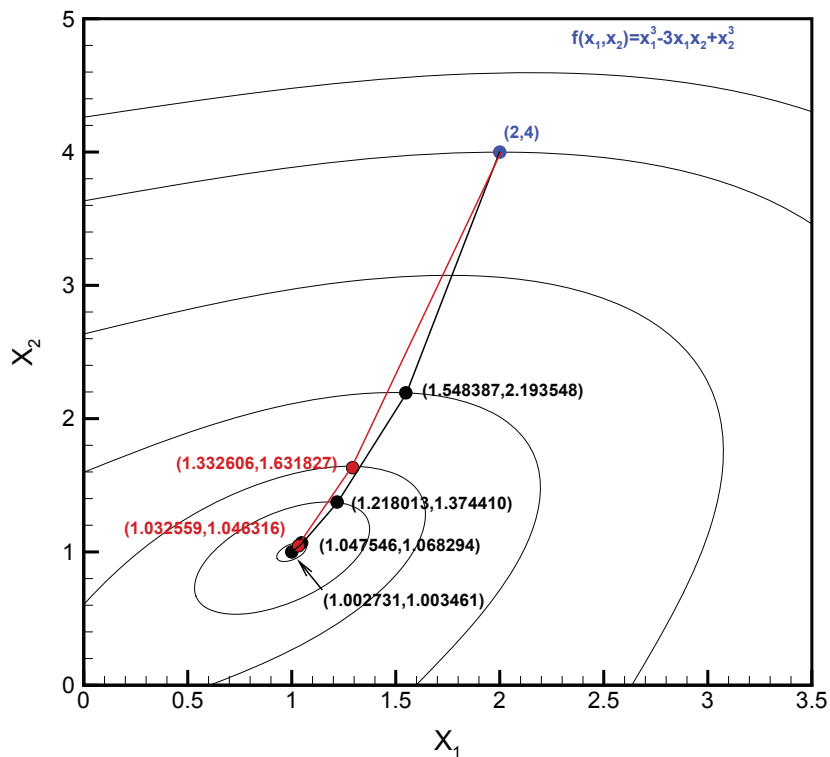
6. Appendix

Figure 1. Comparison between CNM (black) and PM(red)

TABLE 1

$f(x_1, x_2)$	Initial Point	Minimizer	Iterations CNM*	Iterations PM†
$x_1^2 + x_2^3$	$(2, 3)$	$(0, 0)$	22	15
$x_1^3 - 3x_1x_2 + x_2^3$	$(2, 4)$	$(1, 1)$	6	4
$(x_1 - 2)^4 + (x_1 - 2x_2)^2$	$(3, 4)$	$(2, 1)$	33	23
$100\,(x_2 - x_1^2)^2 + (1 - x_1)^2$	$(1.1, 1.2)$	$(1, 1)$	6	5
$\cos(x_1^2 - 3x_2) + \sin(x_1^2 + x_2^2)$	$(1.6, 1.8)$	$(1.376384, 1.678676)$	5	4
$(1.5 - x_1 + x_1x_2)^2 + (2.25 - x_1 + x_1x_2^2)^2$ $+ (2.625 - x_1 + x_1x_2^3)^2$	$(3.5, .4)$	$(3, .5)$	45	37
$2x_1^2 - 1.05x_1^4 + \frac{x_1^6}{6} + x_1x_2 + x_2^2$	$(.4, 1.4)$	$(0,0)$	5	4
$\sin(x_1 + x_2) + (x_1 - x_2)^2$ $-1.5x_1 + 2.5x_2 + 1$	$(-.7, -1.6)$	$(-0.547197, -1.547197)$	4	3
$(x_1 + 2x_2 - 7)^2 + (2x_1 + x_2 - 5)^2$	$(1.6, 2.8)$	$(1,3)$	2	2
$\cos(x_1\cos(x_2))\exp(-((x_1 - \pi)^2 + (x_2 - \pi)^2))$	$(-1.4, -1.5)$	$(-1.4, -1.67425)$	7	5

Tolerance limit : 10^{-6}

* Classical Newton Method, † Proposed Method

References

1. John Brewer, *Kronecker products and matrix calculus in system theory*, Circuits and Systems, IEEE Transactions on **25** (1978), no. 9, 772–781.
2. Changbum Chun, *A family of composite fourth-order iterative methods for solving nonlinear equations*, Applied mathematics and computation **187** (2007), no. 2, 951–956.
3. _____, *Some fourth-order iterative methods for solving nonlinear equations*, Applied Mathematics and Computation **195** (2008), no. 2, 454–459.
4. Alicia Cordero, José L Hueso, Eulalia Martínez, and Juan R Torregrosa, *A family of iterative methods with sixth and seventh order convergence for nonlinear equations*, Mathematical and Computer Modelling **52** (2010), no. 9, 1490–1496.
5. Paul S. Dwyer and M.S. MacPhail, *Symbolic matrix derivatives*, The annals of mathematical statistics **19** (1948), no. 4, 517–534.
6. H.H.H. Homeier, *A modified newton method for rootfinding with cubic convergence*, Journal of Computational and Applied Mathematics **157** (2003), no. 1, 227–230.
7. _____, *On newton-type methods with cubic convergence*, Journal of Computational and Applied Mathematics **176** (2005), no. 2, 425–432.
8. Jisheng Kou, Yitian Li, and Xiuhua Wang, *A modification of newton method with third-order convergence*, Applied Mathematics and Computation **181** (2006), no. 2, 1106–1111.
9. _____, *Some variants of ostrowski's method with seventh-order convergence*, Journal of Computational and Applied Mathematics **209** (2007), no. 2, 153–159.
10. Liping Liu and Xia Wang, *Eighth-order methods with high efficiency index for solving nonlinear equations*, Applied Mathematics and Computation **215** (2010), no. 9, 3449–3454.
11. Alfredo Martinez Estrada, *A treatise on econometric forecasting*, Ph.D. thesis, California Institute of Technology, 2007.
12. H.W. Turnbull, *On differentiating a matrix*, Proceedings of the Edinburgh Mathematical Society (Series 2) **1** (1928), no. 02, 111–128.
13. _____, *A matrix form of taylor's theorem*, Proceedings of the Edinburgh Mathematical Society (Series 2) **2** (1930), no. 01, 33–54.
14. William J. Vetter, *Derivative operations on matrices*, Automatic Control, IEEE Transactions on **15** (1970), no. 2, 241–244.
15. _____, *Matrix calculus operations and taylor expansions*, SIAM review **15** (1973), no. 2, 352–369.
16. Eric Walter and Luc Pronzato, *Identification of parametric models*, Communications and Control Engineering (1997).
17. Zhang Yimin, Suhuan Chen, Qiaoling Liu, and Tieqiang Liu, *Stochastic perturbation finite elements*, Computers & Structures **59** (1996), no. 3, 425–429.
18. Yimin Zhang, Bangchun Wen, and Qiaoling Liu, *Sensitivity analysis of rotor–stator systems with rubbing**, Mechanics of structures and machines **30** (2002), no. 2, 203–211.
19. Y.M. Zhang, L. Zhang, J.X. Zheng, and B.C. Wen, *Neural network for structural stress concentration factors in reliability-based optimization*, Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering **220** (2006), no. 3, 217–224.
20. Qi Zhao and Christian Bohn, *A linearization free recursive prediction error method for combined state and parameter estimation for nonlinear systems*, American Control Conference (ACC), 2013, IEEE, 2013, pp. 899–904.

S. K. Chakraborty received M.Sc. from Indian Institute of Technology, Kharagpur, India and currently pursuing Ph.D. His research interests include numerical optimization.

Department of Mathematics, Indian Institute of Technology, Kharagpur. India.
e-mail: suvrakanti@maths.iitkgp.ernet.in

G. Panda received Ph. D. from Utkal University, India. She is currently Associate Professor at Indian Institute of Technology, Kharagpur, India. Her research interests include convex optimization, numerical optimization and optimization with uncertainty.

Department of Mathematics, Indian Institute of Technology, Kharagpur. India.

e-mail: geetanjali@maths.iitkgp.ernet.in