

12. A Modification of the Gradient Method and Function Extremization

By Tatsuo NODA

Department of Applied Mathematics,
Toyama Prefectural College of Technology

(Communicated by Kôzaku YOSIDA, M. J. A., Feb. 13, 1989)

1. Introduction. Let $x = (x_1, x_2, \dots, x_n)$ be a vector in R^n and D a region contained in R^n . Let $f(x)$ be a real-valued nonlinear function defined on D . Define an n -dimensional vector $\nabla f(x)$ and an $n \times n$ matrix $H(x)$ by

$$\nabla f(x) = (\partial f(x) / \partial x_i) \quad (1 \leq i \leq n)$$

and

$$H(x) = (\partial^2 f(x) / \partial x_j \partial x_k) \quad (1 \leq j, k \leq n).$$

For a vector x , we shall use the norm defined by

$$\|x\| = \left(\sum_{i=1}^n x_i^2 \right)^{1/2}.$$

The Euclidean norm and the spectral norm of an $n \times n$ matrix $A = (a_{ij})$, denoted by $\|A\|$ and $\|A\|_s$, are defined as

$$\|A\| = \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 \right)^{1/2}$$

and

$$\|A\|_s = \lambda^{1/2},$$

respectively, where λ is the maximum eigenvalue of A^*A and A^* is the transposed matrix of A .

Throughout this paper, we shall assume the following three conditions.

(A.1) $f(x)$ is two times continuously differentiable on D .

(A.2) There exists a point $\bar{x} \in D$ satisfying $\nabla f(x) = 0$.

(A.3) The $n \times n$ symmetric matrix $H(\bar{x})$ is positive definite.

Let $U(\bar{x}; \delta) = \{x; \|x - \bar{x}\| < \delta\}$ be a neighbourhood of \bar{x} .

The following well-known theorem gives a sufficient condition for finding a local minimum of $f(x)$.

Theorem 1 ([3, Theorem 8.3]). *In addition to conditions (A.1)–(A.3), suppose that the following condition (A.4) holds.*

(A.4) α is a constant satisfying $0 < \alpha < \frac{2}{\|H(\bar{x})\|_s}$.

Under conditions (A.1)–(A.4), there exists a neighbourhood $U(\bar{x}; \delta_0) \subset D$ such that, for arbitrary $x^{(0)} \in U(\bar{x}; \delta_0)$,

$$x^{(k)} \longrightarrow \bar{x} \quad \text{as } k \longrightarrow \infty,$$

where the $x^{(k)}$ are generated by the gradient method

$$(1.1) \quad x^{(k+1)} = x^{(k)} - \alpha \nabla f(x^{(k)}).$$

The purpose of this paper is to show Theorem 2 by considering an

iteration method different from the above iteration method (1.1). Theorem 2 is a modification of Theorem 1. And we also present a numerical example in order to show the efficiency of our iteration method.

2. Statement of results. We see that, by conditions (A.1)–(A.3), $f(x)$ has a local minimum at \bar{x} . When Theorem 1 is applied to the problem of finding the extremum of functions, the value of α in (1.1) is required. Since \bar{x} is unknown, in general it is very difficult to choose α so that (A.4) holds.

Now, for computational purpose, we propose instead of (1.1) an iteration method

$$(2.1) \quad x^{(k+1)} = x^{(k)} - \frac{\beta}{\|H(x^{(k)})\|} \nabla f(x^{(k)})$$

with a constant β satisfying

$$(A.5) \quad 0 < \beta < 2.$$

Then our iteration method (2.1) leads to the following

Theorem 2. *Under conditions (A.1), (A.2), (A.3) and (A.5), there exists a neighbourhood $U(\bar{x}; \delta) \subset D$ such that, for arbitrary $x^{(0)} \in U(\bar{x}; \delta)$,*

$$x^{(k)} \longrightarrow \bar{x} \quad \text{as } k \longrightarrow \infty,$$

where the $x^{(k)}$ are generated by (2.1).

3. Proof of Theorem 2. We shall prove Theorem 2. First, define an n -dimensional vector $g(x) = (g_i(x))$ by

$$(3.1) \quad g(x) = x - \frac{\beta}{\|H(x)\|} \nabla f(x).$$

By (A.3),

$$0 < (\rho, H(\bar{x})\rho) \leq \|H(\bar{x})\|$$

for any $\rho \in R^n$ with $\|\rho\| = 1$. Since, by (A.1), $\|H(x)\|$ is continuous at every point $x \in D$, there exists a neighbourhood $U(\bar{x}, \delta_1) \subset D$ such that $x \in U(\bar{x}; \delta_1)$ implies $\|H(x)\| > 0$. Then, we see that $g(x)$ is continuously differentiable on $U(\bar{x}; \delta_1)$, and, from (3.1), by (A.2),

$$(3.2) \quad \bar{x} = g(\bar{x}).$$

The $n \times n$ symmetric matrix $H(\bar{x})$ being positive definite from (A.3), all its eigenvalues λ_i ($1 \leq i \leq n$) are positive. Next, define an $n \times n$ matrix $G(x)$ by

$$G(x) = (\partial g_i(x) / \partial x_j) \quad (1 \leq i, j \leq n).$$

As easily seen, it holds

$$(3.3) \quad G(\bar{x}) = I - \frac{\beta}{\|H(\bar{x})\|} H(\bar{x}).$$

We observe that the right hand side of (3.3) is the symmetric matrix, and so, we have

$$(3.4) \quad \|G(\bar{x})\|_s = \max_{1 \leq i \leq n} \left| 1 - \frac{\beta}{\|H(\bar{x})\|} \lambda_i \right|.$$

On the other hand,

$$0 < \lambda_i \leq \|H(\bar{x})\|_s \leq \|H(\bar{x})\|$$

always holds, so that

$$(3.5) \quad 0 < \frac{\beta \lambda_i}{\|H(\bar{x})\|} < 2 \quad (0 < \beta < 2)$$

for $i=1, 2, \dots, n$. Using (3.5), (3.4) implies $\|G(\bar{x})\|_s < 1$. Choose a constant M so as to satisfy $\|G(\bar{x})\|_s < M < 1$. By (A.1), there exists a constant $\delta \leq \delta_1$ such that $U(\bar{x}; \delta) \subset U(\bar{x}; \delta_1)$ and

$$(3.6) \quad \|G(x)\|_s < M \quad \text{for } x \in U(\bar{x}; \delta).$$

Since, by (2.1), (3.1) and (3.2),

$$x^{(k+1)} - \bar{x} = g(x^{(k)}) - g(\bar{x}),$$

this yields

$$x^{(k+1)} - \bar{x} = \int_0^1 G(\bar{x} + t(x^{(k)} - \bar{x}))(x^{(k)} - \bar{x}) dt.$$

We note that $\bar{x} + t(x^{(k)} - \bar{x}) \in U(\bar{x}; \delta)$ ($0 \leq t \leq 1$), provided $x^{(k)} \in U(\bar{x}; \delta)$. Therefore, by (3.6), we obtain

$$\begin{aligned} \|x^{(k+1)} - \bar{x}\| &\leq \left(\int_0^1 \|G(\bar{x} + t(x^{(k)} - \bar{x}))\|_s dt \right) \|x^{(k)} - \bar{x}\| \\ &\leq M \|x^{(k)} - \bar{x}\| \quad (k=0, 1, 2, \dots). \end{aligned}$$

This shows that there exists a neighbourhood $U(\bar{x}; \delta) \subset D$ such that, for arbitrary $x^{(0)} \in U(\bar{x}; \delta)$,

$$x^{(k)} \longrightarrow \bar{x} \quad \text{as } k \longrightarrow \infty.$$

In this way, we have proved Theorem 2, as desired.

4. Numerical example. Masuyama [1] deals with a function

$$y(x; a, b, c, d) = e^{ax}(c \cos bx + d \sin bx) \quad (a < 0),$$

which is called damped oscillation. This type of function is well known in the fields of engineering and physical science. In order to show the efficiency of the iteration method (2.1), we consider a system of nonlinear equations, Example 4.1. The solution of Example 4.1 using the iteration method (2.1) is presented in Table 4.1 below, together with the solution by the iteration method [2, (4.1)].

$$\text{Example 4.1 : } \begin{cases} y(0.0; a, b, c, d) = 1.50, \\ y(0.8; a, b, c, d) = -0.05, \\ y(1.6; a, b, c, d) = -0.12, \\ y(2.4; a, b, c, d) = 0.04. \end{cases}$$

The solution is $(a, b, c, d) = (-1.50, -2.50, 1.50, -0.50)$.

Table 4.1. Computation results for Example 4.1

Methods	Solutions
Iteration method [2, (4.1)] ($\alpha=0.99$)	$(-1.506440, -2.502051,$ $1.499880, -0.5017658)$
Iteration method (2.1) ($\beta=0.99$)	$(-1.506458, -2.501487,$ $1.499880, -0.5009617)$
$(\alpha^{(0)}, b^{(0)}, c^{(0)}, d^{(0)}) = (-1.0, -1.0, -1.0, -1.0)$	

The author would like to express his hearty thanks to H. Mine, Professor Emeritus of Kyoto University, for many valuable suggestions.

References

- [1] M. Masuyama: Zikkenkôsi no Motomekata. Takeuti Syoten (1962) (in Japanese).
- [2] T. Noda: An iteration method for finding a local minimum of $\sum_{i=1}^m (f_i(x_1, x_2, \dots, x_n))^2$.
Sûgaku, **26**, 37-40 (1974) (in Japanese).
- [3] D. L. Russell: Optimization Theory. Benjamin, New York (1970).