ON SOLVING EQUATIONS ARISING FROM OPTIMIZATION PROBLEMS BY SOME GENERALIZED NEWTON METHOD

Seiji SAITO

Received March 29, 2003; revised April 1, 2003

Abstract. In this paper we consider equations arising from an optimization problem by the Lagragian method and show a new algorithm for solving the equations by a generalized Newton method. In a finite number of iterations we can find matrices which are closed to the inverse of the Jacobian matrix by the generalied Newton method. Moreover we get the superlinear convergence to the optimal solution of the optimization problem from our theorem.

1 Introduction Consider the following nonlinear optimization problem:

$$\text{minimize } f(x) \quad \text{subject to } g(x) = 0, \quad x \geq 0, \quad x \in \mathbf{R}^n \tag{1}$$

where $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are twice times continuously differentiable. We get some equations arising from the above problem (1). Further we deal with the Karush-Kuhn-Tucker (K-K-T) condition to (1) (*e.g.* see [5]).

The authors ([1], [2], [5] and so on) assume that the second order sufficiency condition for the optimality in order to have the K-K-T point, but it seems to be difficult that we apply the sufficiency condition to optimization problems. We assume that the inverse of the Jacobian matrix at the K-K-T point is non-singular. In [5] the norm of the inverse of the Jacobian matrix, whose norm seems to be not easily evaluated. We use a well-known theorem in linear algebra and we get an upper bound for the norm of the inverse.

By applying the Newton method , which is due to [3] and [4] without finding the inverse of Jacobian matrices to the equations, we show a new algorithm for solving the equations to the K-K-T condition. In Theorem 2 we prove superlinear convergence to the K-K-T point by our algorithm in the same result as in [3] and [4]. Theorem 3 ensures that the iteration of our algorithm shows the monotone convergence to the optimal solution.

2 Prelinimaries and Notations Let $w = (x, y, z)^T \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n$, where $z \geq 0$ and y are multiplier vectors. In order to find the K-K-T condition, we consider the following Lagrangian function

$$K(w) = f(x) + y^T g(x) - z^T x,$$

where $x = (x_1, x_2, \cdots, x_n)^T$. Denote a norm

$$\| x \| = \sum_{i=1}^{n} |x_i|$$

and

$$\| A \| = \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|$$

2000 *Mathematics Subject Classification.* Primary 65K05; Secondary 90C30.

Key words and phrases. optimization problem, Lagrangian, K-K-T point, generalized Newton method, superlinear convergence.

for $n \times n$-matrix $A = (a_{ij})$ where $a_{ij} \in \mathbf{R}$.

Consider the following K-K-T Condition:

$$\begin{aligned} \nabla_x K(w) &= 0; \\ g(x) &= 0; \\ ZXe &= 0. \end{aligned}$$

Here $X = diag(x_1, x_2, \cdots, x_n),\ e = (1, 1, \cdots, 1)^T,\ Z = diag(z_1, z_2, \cdots, z_n)$ and

$$\nabla_x K(w) = \nabla f(x) + y^T \nabla g(x) - z.$$

In what follows we deal with equations arising from Problem (1):

$$r(w) = 0, \tag{2}$$

where r is an into mapping on $\mathbf{R}^{2n+m}$ such that,

$$r(w) = \begin{pmatrix} \nabla_x K(w) \\ g(x) \\ ZXe \end{pmatrix}.$$

Let $w^* = (x^*, y^*, z^*)^T$ be a K-K-T point of (1). We deal with finding solutions of (2) under the following assumptions.

Assumption. Let $I^* = \{i \in \mathbf{N} : (x^*)_i = 0\}$ and assume that the following conditions (A1)-(A5) hold:

(A1) Functions f and $g = (g_1, g_2, \cdots, g_m)^T$ have continuous twice derivertives. Functions $\nabla^2 f$ and $\nabla_x(y^T g)$ are uniformly Lipschitzian on a neighborhood D of w^*, *i.e.*, there exist $L_1 > 0$ and $L_2 > 0$ satisfying

$$\begin{aligned} \| \nabla^2 f(x_1) - \nabla^2 f(x_2) \| &\leq L_1 \| x_1 - x_2 \| \\ \| \nabla_x^2(y_1^T g(x_1)) - \nabla_x^2(y_2^T g(x_2)) \| &\leq L_2(\| x_1 - x_2 \| + \| y_1 - y_2 \|) \\ \text{for } (x_1, y_1, z),\ (x_2, y_2, z) \in D, & \end{aligned}$$

respectively;

(A2) If $i \in I^*$, then we have the i-th element $(z^*)_i > 0$ of z^*;

(A3) It follows that

$$\nabla_x^2 K(w^*) > 0,$$

where

$$\nabla_x^2 K(w^*) = \nabla^2 f(x^*) + \nabla_x^2((y^*)^T g(x^*));$$

(A4) Let the set

$$\{\frac{\partial g_i}{\partial x}^T (x^*) : i = 1, 2, \cdots, m\} \cup \{e_i = (0, \cdots, 0, \overset{i}{1}, 0, \cdots, 0)^T : i \in I^*\}$$

be linearly independent;

(A5) Let $m \leq |I^*|$.

The following lemma can be proved by applying results in linear algebra.

Lemma 1 *Under Assumption* (A1)-(A5) *the following* (i) *and* (ii) *hold;*

(i) *Let*

$$\mathcal{M} = \begin{pmatrix} \nabla_x^2 K & \nabla g^T \\ \nabla g & O \end{pmatrix},$$

where $\nabla_x^2 K = \nabla_x^2 K(w^*), \nabla g^T = \nabla g^T(x^*)$. *Then there exists the inverse matrix of* $\mathcal{M}$ *such that*

$$\begin{aligned}\mathcal{M}^{-1} &= \begin{pmatrix} I & -(\nabla_x^2 K)^{-1} \nabla g^T \\ O & I \end{pmatrix} \times \begin{pmatrix} (\nabla_x^2 K)^{-1} & O \\ O & (-\nabla g(\nabla_x^2 K)^{-1} \nabla g^T)^{-1} \end{pmatrix} \\ &\times \begin{pmatrix} I & O \\ -\nabla g(\nabla_x^2 K)^{-1} & I \end{pmatrix}.\end{aligned}$$

(ii) *Let*

$$\nabla r(w^*) = \begin{pmatrix} \mathcal{M} & Z_{12} \\ Z_{21} & X^* \end{pmatrix},$$

where

$$Z_{12} = \begin{pmatrix} -I \\ O \end{pmatrix} \quad \textit{and} \quad Z_{21} = (Z^* \; O).$$

There exists the inverse matrix of $\nabla r(w^*)$ *such that*

$$\begin{aligned}(\nabla r(w^*))^{-1} &= \begin{pmatrix} I & -\mathcal{M}^{-1} Z_{12} \\ O & I \end{pmatrix} \times \begin{pmatrix} \mathcal{M}^{-1} & O \\ O & (X^* - Z_{21}\mathcal{M}^{-1} Z_{12})^{-1} \end{pmatrix} \\ &\times \begin{pmatrix} I & O \\ -Z_{21}\mathcal{M}^{-1} & I \end{pmatrix}.\end{aligned}$$

Proof. (i) Since $\nabla^2 K = (\dfrac{\partial^2 f}{\partial x_i \partial x_j})_{ij} > 0$, it follows that $\nabla^2 K$ is symmetric and there exists an $n \times n-$ matrix $\mathcal{U}$ such that $\mathcal{U}^{-1} = \mathcal{U}^T$ and $\nabla^2 K = \mathcal{U}^T \Lambda \mathcal{U}$, where $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$ and $\lambda_i > 0$ for $i = 1, \cdots, n$ are eigenvalues of $\nabla^2 K$. Then we have

$$(\nabla^2 K)^{-1} = \mathcal{U}^T \Lambda^{-1} \mathcal{U}$$

as well as $(\nabla^2 K)^{-1}$ is symmetric. From (A3), we get $\nabla g(\nabla^2 K)^{-1} \nabla g^T > 0$, as

$$y^T \nabla g(\nabla^2 K)^{-1} \nabla g^T y = (\mathcal{U} \nabla g^T y)^T \Lambda^{-1} \mathcal{U} \nabla g^T y > 0$$

with $y \neq 0 \in \mathbf{R}^m$. Thus there exists the inverse of $\nabla g(\nabla_x^2 K)^{-1} \nabla g^T$ and we get the inverse of $\mathcal{M}$ as above.
(ii) By putting

$$\mathcal{K}^* = (\nabla^2 K)^{-1} + (\nabla^2 K)^{-1} \nabla g^T [-\nabla g(\nabla^2 K)^{-1} \nabla g^T]^{-1} \nabla g(\nabla^2 K)^{-1},$$

it follows that $\mathcal{K}^*$ is symmetric, because of the following equalities:

$$(\bigtriangledown g(\bigtriangledown^2 K)^{-1} \bigtriangledown g^T)^T = \bigtriangledown g(\bigtriangledown^2 K)^{-1} \bigtriangledown g^T;$$
$$((\bigtriangledown g(\bigtriangledown^2 K)^{-1} \bigtriangledown g^T)^{-1})^T = (\bigtriangledown g(\bigtriangledown^2 K)^{-1} \bigtriangledown g^T)^{-1}.$$

Since $-Z_{21}\mathcal{M}^{-1}Z_{12} = Z^*\mathcal{K}^*$, $\bigtriangledown r(w^*)$ is nonsingular if and only if $X^* + Z^*\mathcal{K}^*$ is nonsingular. Without loss of generality we consider

$$X^* \quad = \quad diag(x_1^*, \cdots, x_a^*, x_{a+1}^*, \cdots, x_n^*)$$

and

$$I^* = \{i : x_i^* = 0, z_i^* > 0, \text{for } i = 1, \cdots, a\},$$
$$Q^* = \{q : x_q^* > 0, z_q^* = 0, \text{for } q = a+1, \cdots, n\}.$$

Here $a = |I^*|$. In order to show $det(\bigtriangledown r(w^*)) \neq 0$ we will prove $det(X^* + Z^*\mathcal{K}^*) \neq 0$. Suppose that $(X^* + Z^*\mathcal{K}^*)x = 0$, where $x = (x_1, x_2, \cdots, x_n)^T$. In $q \in Q^*$ such that $z_q^* = 0$ we have $x_q = 0$. By putting

$$\mathcal{Z}^{-1} = diag(\frac{1}{z_1^*}, \frac{1}{z_2^*}, \cdots, \frac{1}{z_a^*}, 0, \cdots, 0)$$

we get

$$(\mathcal{Z}^{-1}X^* + \mathcal{K}^*)(x_1, \cdots, x_i, \cdots, x_a, 0, \cdots, 0)^T = 0,$$

so that we have $\mathcal{K}^*(x_1, \cdots, x_i, \cdots, x_a, 0, \cdots, 0)^T = 0$ for $i \in I^*$.

Suppose that some $\tilde{x} = (x_1, x_2, \cdots, x_a, 0, \cdots, 0)^T \neq 0$ and $\mathcal{K}^*\tilde{x} = 0$. From (A5) and $rank(\bigtriangledown g(w^*)) \geq m$, it follows that, for some $c_i \in \mathbf{R}$,

$$\sum_{i=1}^{m} c_i \frac{\partial g_i}{\partial x}(w^*) = -\sum_{j=1}^{a} x_j e_j.$$

From (A4), we have $c_i = x_j = 0$ for any i, j. This contradicts with $\tilde{x} \neq 0$. Thus it follows that $x_i = 0$ for $i \in I^*$. Since $det(X^* + Z^*\mathcal{K}^*) \neq 0$, there exists the inverse of $\bigtriangledown(r(w^*))$ as above. This completes the proof.

From Lemma 1 it follows that there exist positive numbers $\varepsilon, \rho, M, N, B_1, B_2, B_3, L_3, L_4, L_5$ and a subset $D \subset \mathbf{R}^{2n+m}$ satisfying the following assumtion:

(A6) Let $0 < \rho < 1$ and $0 < \varepsilon < 1$ satisfy

$$B_3\rho + \varepsilon L_5 < \sqrt{2} - 1 \text{ and } 1 - M\varepsilon B_2 > 0$$

and an integer $N \geq 3$ be

$$\Big(\frac{B_3\rho + \varepsilon L_5 + 1}{(B_3\rho + \varepsilon L_5)^{-1} - 1}\Big)^{N-2} \leq \frac{\rho - \rho^2}{B_3(\rho^2+1)\varepsilon\Big[L_3 + \sqrt{L_3{}^2 + \frac{4L_4(\rho-\rho^2)}{B_3(\rho^2+1)}}\Big]} \quad (= B_4(\rho)).$$

Here

$$\begin{aligned}
D &= \{w \in \mathbf{R}^{2n+m} : \| w - w^* \| \leq \varepsilon\},\\
M &= L_1 + L_2(1 + 2n\varepsilon) + 1,\\
B_1 &= \max_{w \in D} \| \bigtriangledown r(w) \|,\\
B_2 &= \max_{w \in D} \| (\bigtriangledown r(w))^{-1} \|,\\
B_3 &= \frac{B_2(M\varepsilon + B_1)}{1 - M\varepsilon B_2},\\
L_3 &= L_1 + L_2 + 1,\\
L_4 &= 2nL_2,\\
L_5 &= \max_{w \in D} \| \bigtriangledown_x (y^T g(x)) \| \frac{B_2(1+\rho^2)}{1 - M\varepsilon B_2}.
\end{aligned}$$

If sufficiently small $\varepsilon, \rho > 0$ are fixed, then we get a small neighborhood D of w^* and the other positive numbers $B_i, i = 1, 2, 3$, etc.

Remarks. Since $B_3\rho + \varepsilon L_5 < \sqrt{2} - 1$, it follows that

$$(B_3\rho + \varepsilon L_5)^{-1} > 1 \text{ and } \frac{B_3\rho + \varepsilon L_5 + 1}{(B_3\rho + \varepsilon L_5)^{-1} - 1} < 1.$$

3 Algorithm by the Generalized Newton Method

For $k = 1, 2, \cdots$, we show an algorithm for

$$\{w^{(k)} = (x^{(k)}, y^{(k)}, z^{(k)})^T \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n\},$$

such that $\lim_{k \to \infty} w^{(k)} = w^*$, by applying the generalized Newton method. Denote as follows:

$$\begin{aligned}
J^{(k)} &= \bigtriangledown r(w^{(k)});\\
&= \begin{pmatrix} \bigtriangledown_x^2 K(w^{(k)}) & \bigtriangledown g(x^{(k)})^T & -I \\ \bigtriangledown g(x^{(k)}) & O & O \\ Z^{(k)} & O & X^{(k)} \end{pmatrix};
\end{aligned}$$

$$\begin{aligned}
\bigtriangledown_x^2 K(w) &= \bigtriangledown^2 f(x) + \bigtriangledown_x^2 (y^T g(x));\\
X^{(k)} &= diag(x_1^{(k)}, x_2^{(k)}, \cdots, x_n^{(k)});\\
Z^{(k)} &= diag(z_1^{(k)}, z_2^{(k)}, \cdots, z_n^{(k)}).
\end{aligned}$$

For $k = 1, 2, \cdots$, we construct the following sequences of $l \times l$-matrices $\{U(k)\}$ and $\{V^{(p)}(k)\}$ for $p = 1, 2, \cdots, k$. In order to solve (2), we show the following algorithm for computing the above $V^{(k)}(k)$, which is sufficiently close to $[\bigtriangledown r(w^*)]^{-1}$ rather than finding the inverse matrices of $\bigtriangledown r(w^{(k)})$.

Algorithm. Let ρ and $N \in \mathbf{N}$ satisfy (A6). We choose $w^{(1)} \in D$ and put $J^{(1)} = \bigtriangledown r(w^{(1)})$. For $k = 1, 2, \cdots$, do the following steps.

Step 1. For $k = 1, 2, \cdots, N$, find a matrix $U(k)$ such that

$$\| J^{(k)} U(k) - I \| \leq \rho. \quad (cf. \quad Theorem\ 1)$$

Put $V^{(0)}(k) = U(k)$ and go to Step 2.

For $k = N+1, N+2, \cdots$, put

$$V^{(0)}(k) = V^{(k-1)}(k-1). \quad (cf. \quad Theorem\ 3)$$

Go to Step 2.

Step 2. For $p = 1, 2, \cdots, k$, compute a sequence $\{V^{(p)}(k)\}$ such that

$$V^{(p)}(k) = V^{(p-1)}(k)[2I - J^{(k)}V^{(p-1)}(k)].$$

Step 3. Compute

$$w^{(k+1)} = w^{(k)} - V^{(k)}(k)r(w^{(k)}). \quad (cf. \quad Theorem\ 2)$$

Go to Step 1.

4 Superlinear Convergence

We prove the closeness of $J^{(k)}$ to $\bigtriangledown r(w^*)$ as follows.

Lemma 2 *For $k = 1, 2, \cdots$, we get*

$$\| J^{(k)} - \bigtriangledown r(w^*) \| \leq (L_3 + L_4 \| x^{(k)} - x^* \|) \| w^{(k)} - w^* \| .$$

Proof. From (A1) we get for $k \in \mathbf{N}$

$$\begin{aligned}
&\| J^{(k)} - \bigtriangledown r(w^*) \| \\
&\leq \| \bigtriangledown_x^2 K(w^{(k)}) - \bigtriangledown_x^2 K(w^*) \| + 2 \| \bigtriangledown g(x^{(k)}) - \bigtriangledown g(x^*) \| \\
&\quad + \| Z^{(k)} - Z^* \| + \| X^{(k)} - X^* \| \\
&\leq \| \bigtriangledown^2 f(x^{(k)}) - \bigtriangledown^2 f(x^*) \| + \| \bigtriangledown_x^2((y^{(k)})^T g(x^{(k)})) - \bigtriangledown_x^2((y^*)^T g(x^*)) \| \\
&\quad + 2 \sum_i \sum_j | \int_0^1 \frac{\partial}{\partial x} \Big[\frac{\partial g_i(x^* + \theta(x^{(k)} - x^*))}{\partial x_j}) \Big]^T d\theta (x^{(k)} - x^*) | \\
&\quad + \| w^{(k)} - w^* \| \\
&\leq \| \bigtriangledown^2 f(x^{(k)}) - \bigtriangledown^2 f(x^*) \| + \| \bigtriangledown_x^2((y^{(k)})^T g(x^{(k)})) - \bigtriangledown_x^2((y^*)^T g(x^*)) \| \\
&\quad + 2 \Big[\sum_i \sum_j \sum_q \int_0^1 | \frac{\partial}{\partial x_q} \Big(\frac{\partial g_i(x^*)}{\partial x_j} \Big) | d\theta | x_q^{(k)} - x_q^* | \Big] + \| w^{(k)} - w^* \| \\
&\leq L_1 \| w^{(k)} - w^* \| + L_2 \| w^{(k)} - w^* \| + 2nL_2 \| x^{(k)} - x^* \|^2 + \| w^{(k)} - w^* \| \\
&\leq (L_3 + L_4 \| x^{(k)} - x^* \|) \| w^{(k)} - w^* \| .
\end{aligned}$$

This completes the proof.

We prove the convergence of $\{V^{(k)}(k)\}$ approaching to the inverse matrix $[\bigtriangledown r(w^*)]^{-1}$ as $k \to \infty$.

Theorem 1 *Under assumptions* (A1) - (A6) *the following statements* (i)-(iv) *hold for* $k = 1, 2, \cdots,$ *and* $p = 1, 2, \cdots, k,$*:*

(i) *It follows that*

$$\| J^{(k)} V^{(p)}(k) - I \| \leq \rho^{2^{(p-1)}} \quad and \quad \lim_{k\to\infty} J^{(k)} V^{(k)}(k) = I;$$

(ii) *As long as* $w^{(k)} \in D$, *we have*

$$\begin{aligned}\| J^{(k)} \| &\leq M\varepsilon + B_1,\\ \| [J^{(k)}]^{-1} \| &\leq \frac{B_2}{1 - M\varepsilon B_2};\end{aligned}$$

(iii) *We get*

$$\| V^{(k)}(k) - [J^{(k)}]^{-1} \| \leq \frac{B_2 \rho^{2^{(k-1)}}}{1 - M\varepsilon B_2};$$

(iv) *There exists a positive number* L *such that*

$$\| V^{(k)}(k) - [\bigtriangledown r(w^*)]^{-1} \| \leq L \| w^{(k)} - w^* \| + \frac{B_2 \rho^{2^{(k-1)}}}{1 - M\varepsilon B_2}.$$

Proof. (i) By the mathematical induction it can be proved for $k = 1, 2, \cdots$. If $p = 1$, we put $C_k = J^{(k)} V^{(0)}(k) - I$ and $\| C_k \| \leq \rho$. Assume that $J^{(k)} V^{(q)}(k) = I + C_k^{2^{(q-1)}}$ for $q = 1, 2, \cdots, k-1$. From Step 2 in Algorithm, it follows that

$$\begin{aligned}J^{(k)} V^{(q+1)}(k) &= [I + C_k^{2^{(q-1)}}][I - C_k^{2^{(q-1)}}]\\ &= I - C_k^{2^q}.\end{aligned}$$

And also we have

$$\lim_{k \to \infty} J^{(k)} V^{(k)}(k) = I.$$

(ii) From Lemma 2 we have

$$\| J^{(k)} \| \leq M\varepsilon + B_1.$$

Since

$$\begin{aligned}\| I - J^{(k)}[\bigtriangledown r(w^*)]^{-1} \| &= \| (\bigtriangledown r(w^*) - J^{(k)})(\bigtriangledown r(w^*))^{-1} \|\\ &\leq \varepsilon B_2 M < 1,\end{aligned}$$

it follows that

$$\begin{aligned}\| [J^{(k)}]^{-1} \| &= \| [\bigtriangledown r(w^*)]^{-1} (J^{(k)}[\bigtriangledown r(w^*)]^{-1})^{-1} \|\\ &\leq \| [\bigtriangledown r(w^*)]^{-1} \| \| \sum_{i=0}^{\infty} (I - J^{(k)}[\bigtriangledown r(w^*)]^{-1})^i \|\\ &\leq B_2 \sum_{i=0}^{\infty} (M\varepsilon B_2)^i\\ &= \frac{B_2}{1 - M\varepsilon B_2}.\end{aligned}$$

(iii)From (ii) we have

$$\begin{aligned}\| V^{(k)}(k) - [J^{(k)}]^{-1} \| &= \| [J^{(k)}]^{-1} [J^{(k)} V^{(k)}(k) - I]) \|\\ &\leq \| [J^{(k)}]^{-1} \| \| J^{(k)} V^{(k)}(k) - I \|\\ &\leq \frac{B_2 \rho^{2^{(k-1)}}}{1 - M\varepsilon B_2}.\end{aligned}$$

(iv) From (iii) we have

$$\begin{aligned}\| V^{(k)}(k)-[\bigtriangledown r(w^*)]^{-1} \| &\le \| V^{(k)}(k)-[J^{(k)}]^{-1} \| + \| [J^{(k)}]^{-1}-[\bigtriangledown r(w^*)]^{-1} \| \\ &\le \frac{B_2\rho^{2^{(k-1)}}}{1-M\varepsilon B_2} + \| [\bigtriangledown r(w^*)]^{-1} \| \| \bigtriangledown r(w^*) - J^{(k)} \| \| [J^{(k)}]^{-1} \| \\ &\le \frac{B_2\rho^{2^{(k-1)}}}{1-M\varepsilon B_2} + \frac{B_2^2 M}{1-M\varepsilon B_2} \| w^{(k)}-w^* \| .\end{aligned}$$

This completes the proof.

We have the superlinear convergence for $w^{(k)}$ to w^* as $k \to \infty$ by the generalized Newton method as well as we get $w^{(k)} \in D$ for any $k = 1, 2, \cdots$.

Theorem 2 *Under assumptions* (A1) - (A6) *and* $w^{(1)} \in D$ *it follows that*

$$\lim_{k\to\infty} \frac{\| w^{(k+1)} - w^* \|}{\| w^{(k)} - w^* \|} = 0$$

and $w^{(k)} \in D$ *for* $k = 2, 3, \cdots$.

Proof. Since $r(w^*) = 0$, we have, as $k \longrightarrow \infty$

$$\begin{aligned} w^{(k+1)} - w^* &= w^{(k)} - w^* - V^{(k)}(k)[r(w^{(k)}) - r(w^*)] \\ &= w^{(k)} - w^* + V^{(k)}(k)[J^{(k)}(w^* - w^{(k)}) + O(\| w^{(k)} - w^* \|^2)] \\ &= ([J^{(k)}]^{-1} - V^{(k)}(k))J^{(k)}(w^{(k)} - w^*) + V^{(k)}(k)O(\| w^{(k)} - w^* \|^2).\end{aligned}$$

From Theorem 1 we get

$$\| w^{(k+1)} - w^* \| \le \frac{B_2\rho^{2^{(k-1)}}(M\varepsilon + B_1)}{1 - M\varepsilon B_2} \| w^{(k)} - w^* \| + L_5 \| w^{(k)} - w^* \|^2,$$

where

$$L_5 = \max_{w\in D} \| \bigtriangledown_x^2(y^T g(x)) \| \frac{B_2(1+\rho^2)}{1 - M\varepsilon B_2}.$$

From (A6), we have, as long as $w^{(k)} \in D$,

$$\| w^{(k+1)} - w^* \| \le (B_3\rho + \varepsilon L_5)\varepsilon < \varepsilon.$$

This completes the proof.

Under Assumption we have $\| w^{(k+1)} - w^* \| < \| w^{(k)} - w^* \|$ and $\| w^{(k+1)} - w^{(k)} \| < \| w^{(k)} - w^{(k-1)} \|$ for $k = 1, 2, \cdots$. The number N is given in Assupmtion (A6).

Theorem 3 *Under assumptions* (A1) - (A6) *we get the following statements* (i) *and* (ii).
(i) *For a* $\delta > 0$ *such that*

$$\delta \ge \frac{B_3\rho^{2^{k-1}} + \varepsilon L_5 + 1}{(B_3\rho^{2^{k-2}} + \varepsilon L_5)^{-1} - 1} \quad (= B_5(k)) \text{ for } k = 1, 2, \cdots ,$$

we get

$$\frac{\| w^{(k+1)} - w^{(k)} \|}{\| w^{(k)} - w^{(k-1)} \|} \le \delta.$$

(ii) *Moreover if* $\delta^{N-2} \leq B_4(\rho)$, *then it follows that*

$$\| J^{(N+1)} V^{(N)}(N) - I \| \leq \rho$$

and $\| w^{(k+1)} - w^{(k)} \|$ *is decreasing for* $k = 1, 2, \cdots$.

Proof. (i) By putting $d_k = \| w^{(k)} - w^* \|$, we have

$$\begin{aligned} \frac{\| w^{(k+1)} - w^{(k)} \|}{\| w^{(k)} - w^{(k-1)} \|} &\leq \frac{d_{k+1} + d_k}{|d_k - d_{k-1}|} \\ &= \frac{d_{k+1}/d_k + 1}{|(d_k/d_{k-1})^{-1} - 1|}. \end{aligned}$$

From the proof of Theorem 2, it follows that

$$\begin{aligned} \frac{d_{k+1}}{d_k} &\leq B_3 \rho^{2^{k-1}} + L_5 d_k \\ &\leq B_3 \rho^{2^{k-1}} + L_5 \varepsilon \\ &\leq B_3 \rho + \varepsilon L_5 < 1 \end{aligned}$$

for $k = 2, 3, \cdots$.Then we get

$$\frac{1}{d_{k-1}/d_k - 1} \leq \frac{1}{(B_3 \rho^{2^{k-2}} + \varepsilon L_5)^{-1} - 1},$$

and also

$$\frac{\| w^{(k+1)} - w^{(k)} \|}{\| w^{(k)} - w^{(k-1)} \|} \leq B_5(k) \leq \delta.$$

(ii) From (i), we have

$$\begin{aligned} \| w^{(k+1)} - w^{(k)} \| &\leq \delta \| w^{(k)} - w^{(k-1)} \| \\ &\leq \delta^{k-1} \| w^{(2)} - w^{(1)} \| \\ &\leq 2\varepsilon \delta^{k-1}. \end{aligned}$$

By the same proof of Lemma 2 it follows that

$$\begin{aligned} \| J^{(N+1)} - J^{(N)} \| &\leq (L_3 + L_4 \| x^{(N)} - x^{(N-1)} \|) \| w^{(N)} - w^{(N-1)} \| \\ &\leq 2(L_3 + 2L_4 \varepsilon \delta^{N-2}) \varepsilon \delta^{N-2}, \end{aligned}$$

as well as

$$\begin{aligned} \| J^{(N+1)} V^{(N)}(N) - I \| &\leq \| J^{(N)} V^{(N)}(N) - I \| + \| J^{(N)} - J^{(N+1)} \| \| V^{(N)}(N) \| \\ &\leq \rho^{2^{N-1}} + 2(L_3 + 2L_4 \varepsilon \delta^{N-2}) \varepsilon \delta^{N-2} B_3(\rho^{2^{N-1}} + 1). \end{aligned}$$

When $N \geq 3$, we get

$$\frac{\rho - \rho^{2^{N-1}}}{B_3(\rho^{2^{N-1}} + 1)\varepsilon \left[L_3 + \sqrt{L_3^2 + \frac{4L_4(\rho - \rho^{2^{N-1}})}{B_3(\rho^{2^{N-1}} + 1)}} \right]} \geq B_4(\rho).$$

From $\delta^{N-2} \leq B_4(\rho)$, then we have

$$4\varepsilon^2 L_4 \delta^{2(N-2)} + 2\varepsilon L_3 \delta^{N-2} + \frac{\rho^{2^{N-1}} - \rho}{B_3(\rho^{2^{N-1}} + 1)} \leq 0,$$

so that

$$\| J^{(N+1)} V^{(N)}(N) - I \| \leq \rho$$

and the other conclusion holds. This completes the proof.

Remarks. (i) From $(A6)$, it follows that $B_5(k) < 1$. Therefore we can have $0 < \delta < 1$ such that $B_5(k) \leq \delta$ and $\delta^{N-2} \leq B_4(\rho)$.

(ii) If $B_4(\rho) < 1$, then we can set

$$N = \left\lceil \frac{\log B_4(\rho)}{\log \delta} \right\rceil + 2.$$

References

[1] P.T.Boggs, J.W.Tolle and P.Wang: On the local convergence of quasi-Newton methos for constrained optimization, *SIAM. J. on Control and Optimization*, **20** (1982), 161-171.

[2] C.G. Broyden, J.E. Dennis, Jr. and J.J. Moré: On the local and superlinear convergence of quasi-linear Newton method, *J. Intitute of Mathematics and its Applications*, **12** (1973), 223-245.

[3] T.Noda and T.Fijita: A general Newton method for systems of nonlinear Equations I, *Mathematica Japonica*, **44** (1998), 541-547.

[4] T.Noda and H.Ishii: A general Newton method for systems of nonlinear Equations II, *Mathematica Japonica*, **48** (1998), 447-452.

[5] H.Yabe and H.Yamashita: Q-superlinear convergence of primal-dual interior point quasi-Newton methods for constrained optimization, *J. Operations Reseasrch Society of Japan*, **40** (1997), 415-436.

Graduate School of Information Science and Technology, Osaka University, 2-1 Yamada-oka, Suita, 565-0871, Japan

e-mail: saito@ist.osaka-u.ac.jp