数理解析研究所講究録
第747巻 1991年 172-182

Sequential Truncated QP Method for Constrained Optimization

東京理科大学・工学部　矢部　博 (Hiroshi Yabe)
システム計画研究所　八巻直一 (Naokazu Yamaki)
東京理科大学・理学部　高橋　悟 (Satoru Takahashi)

1 Introduction

Constrained nonlinear optimization has been the subject of a significant amount of research during the past two decades. As a result, a variety of different types of methods for solving nonlinear optimization problems have been proposed and developed, for example, the augmented Lagrangian method, the sequential quadratic programming (SQP) method, the penalty method, the reduced gradient method, the trust region method and so forth. Among these methods, those based on the SQP method have been found to be very efficient for solving small to medium-sized problems.

Recently, methods for solving large and sparse nonlinear constrained problems have been required. As yet, however, there have been relatively few methods to solve large-sized problems. One of such attempts is to use the successive linear programming (SLP) algorithms[10]. Unfortunately, the above mentioned methods for solving general nonlinear optimization problems can not necessarily applied to large and sparse problems.

Though applying the SQP method to large problems seems promising, the solution of a large-sized quadratic programming subproblem might be rather expensive and could be as costly as solving the original problem directly. So it is reasonable to solve the QP subproblems inexactly by using iterative methods, e.g., conjugate gradient like and SOR like methods [5], [6], [8]. This notion was suggested by Dembo and Tulowitzki[1] and Fontecilla[2], [3] and they proved the locally superlinear convergence of their methods. Further, sparse quasi-Newton updates for unconstrained optimization might be applied to the SQP method[9]. The SQP method for solving sparse problems has been studied

by Nickel and Tolle[7]. These types of SQP methods are called the sequential truncated QP method or the inexact quasi-Newton method.

In this paper, we propose a sequential truncated QP method and, by using the results of Han[4], we show the global convergence of the method.

The general form of the nonlinear optimization problems to be considered is:

(NLP)

$$\begin{array}{lll} \text{minimize} & f(x) & \text{with respect to } x \\ \text{subject to} & g_i(x) & \leq 0,\ i = 1, \ldots, m, \\ & h_j(x) & = 0,\ j = 1, \ldots, l, \end{array}$$

where

$$x \in R^n,\quad f : R^n \to R,\quad g_i : R^n \to R,\quad h_j : R^n \to R,$$
$$g(x) = (g_1(x), \ldots, g_m(x))^T,\quad h(x) = (h_1(x), \ldots, h_l(x))^T.$$

Throughout this paper, $\| \bullet \|$ and $\| \bullet \|_1$ denote the 2 and 1 norms, respectively. ∇f is a gradient vector of $f(x)$ and ∇g, ∇h mean Jacobian matrices of $g(x)$ and $h(x)$, respectively. Further, we recall that a directional derivative $D(q(x); d)$ of a real-valued function q at a point x in a direction d is the quantity defined by

$$D(q(x)\ ;\ d) = \lim_{t \to +0} \frac{q(x + td) - q(x)}{t}.$$

2 Sequential Truncated QP Method

The algorithm of the original SQP method is as follows:

(SQP method)

Step 0. Select an initial point x^1 and an $n \times n$ symmetric positive definite matrix B_1. Set $k = 1$.

Step 1. Having x^k and B_k, find the search direction d^k by solving the QP subproblem:

(QP subproblem)

$$\text{minimize } \frac{1}{2} d^T B_k d + \nabla f(x^k)^T d$$

$$\text{subject to} \quad g(x^k) + \nabla g(x^k)d \leq 0,$$
$$h(x^k) + \nabla h(x^k)d = 0,$$

and choose λ^{k+1} and μ^{k+1} to be the optimal multiplier vectors for this problem.

Step 2. If $(x^k, \lambda^{k+1}, \mu^{k+1})$ satisfy the Karush-Kuhn-Tucker(K-K-T) condition of Problem NLP, then stop; otherwise, go to Step 3.

Step 3. Determine a step size α_k by a suitable line search procedure.

Step 4. Set $x^{k+1} = x^k + \alpha_k d^k$.

Step 5. Update B_k giving a symmetric positive definite matrix B_{k+1} by a suitable quasi-Newton formula.

Step 6. Set $k = k + 1$ and go to Step 1.

Though the original SQP method requires the QP subproblem to be solved exactly, the sequential truncated QP (STQP) method relaxes the K-K-T condition of the QP subproblem. We propose the following relaxed K-K-T condition:

(The parameters ε_1, ε_2, ε_3 and ε_4 are given in Step 0 of the above algorithm.)

Step.1' Find the triple ($d^k, \lambda^{k+1}, \mu^{k+1}$) satisfying

$$\begin{aligned}
&(1) \quad \|B_k d + \nabla f(x^k) + \nabla g(x^k)^T\lambda + \nabla h(x^k)^T\mu\| \leq \varepsilon_1\|d\|,\\
&(2) \quad |\lambda^T(g(x^k) + \nabla g(x^k)d)| \leq \varepsilon_2\|d\|^2,\\
&(3) \quad (\ |\mu^T(h(x^k) + \nabla h(x^k)d)| \leq \varepsilon_3\|d\|^2 \ ),\\
&(4) \quad |h_i(x^k) + \nabla h_i(x^k)^T d| \leq \varepsilon_4\|d\|^2, \qquad i = 1, \ldots, l,\\
&(5) \quad g(x^k) + \nabla g(x^k)d \leq 0,\\
&(6) \quad \lambda \geq 0.
\end{aligned}$$

Setting $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = 0$, we have the exact K-K-T condition of the original QP subproblem. The expression (2) corresponds to the complementarity condition for

the inequality constraints, while the expression (3) for the equality constraints is redundant and can be obtained by combining the expression (4) and the boundedness of the Lagrangian multiplier vectors, which is assumed in the next section.

In Step 3 of the above algorithm, we use Han's merit function[4] as follows:

$$\theta_r(x) = f(x) + rp(x), \tag{7}$$

where r is a penalty parameter and

$$p(x) = \max(0, g_1(x), \ldots, g_m(x), |h_1(x)|, \ldots, |h_l(x)|). \tag{8}$$

Then we have the following STQP algorithm:

(STQP method)

Step 0. Select an initial point x^1 and an $n \times n$ symmetric positive definite matrix B_1. Set $k = 1$. Give the parameters ε_1, ε_2, ε_3, ε_4, $r > 0$, $0 < \nu < 1/2$ and $0 < \tau < 1$.

Step 1. Having x^k and B_k, find the triple ($d^k, \lambda^{k+1}, \mu^{k+1}$) satisfying the conditions

$$\begin{aligned}
\|B_k d + \nabla f(x^k) + \nabla g(x^k)^T \lambda + \nabla h(x^k)^T \mu\| &\leq \varepsilon_1 \|d\|, \\
|\lambda^T (g(x^k) + \nabla g(x^k) d)| &\leq \varepsilon_2 \|d\|^2, \\
(\ |\mu^T (h(x^k) + \nabla h(x^k) d)| &\leq \varepsilon_3 \|d\|^2 \ ), \\
|h_i(x^k) + \nabla h_i(x^k)^T d| &\leq \varepsilon_4 \|d\|^2, \qquad i = 1, \ldots, l, \\
g(x^k) + \nabla g(x^k) d &\leq 0, \\
\lambda &\geq 0.
\end{aligned}$$

Step 2. If $(x^k, \lambda^{k+1}, \mu^{k+1})$ satisfy the Karush-Kuhn-Tucker(K-K-T) condition of Problem NLP, then stop; otherwise, go to Step 3.

Step 3. Determine a step size α_k by Han's line search procedure;

Step 3.1 Set $\beta_{k,1} = 1$ and $i = 1$.

Step 3.2 If the generalized Armijo's criterion

$$\theta_r(x^k + \beta_{k,i} d^k) \leq \theta_r(x^k) - \nu \beta_{k,i} (d^k)^T B_k d^k \tag{9}$$

is satisfied, then set $\alpha_k = \beta_{k,i}$ and go to Step 4; otherwise, go to Step 3.3.

Step 3.3 Set $\beta_{k,i+1} = \tau\beta_{k,i}$, $i = i + 1$ and go to Step 3.2.

Step 4. Set $x^{k+1} = x^k + \alpha_k d^k$.

Step 5. Update B_k giving a symmetric positive definite matrix B_{k+1} by a suitable quasi-Newton formula.

Step 6. Set $k = k + 1$ and go to Step 1.

3 Global Convergence

In this section, following to Han[4], we show the global convergence property of the STQP method given in the previous section. Suppose the assumptions:

(A1) f, g_i and h_j are twice continuously differentiable.

(A2) The subproblem in the STQP method is solvable at each iteration.

(A3) For given r, the level set of the merit function

$$L_r(x^1) = \{x \in R^n | \, \theta_r(x) \leq \theta_r(x^1)\} \tag{10}$$

is compact.

(A4) There exist positive constants M_1 and M_2 such that

$$M_1\|v\|^2 \leq v^T B_k v \leq M_2\|v\|^2 \tag{11}$$

for all v and for each $k \geq 1$.

(A5) There holds $r \geq \|\lambda^k\|_1 + \|\mu^k\|_1$ for each $k \geq 1$.

(A6) The parameters are chosen so that $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4 < M_1(\frac{1}{2} - \nu)$.

Considering the relaxed K-K-T conditions (1) $\sim$ (6), we obtain directly the following result.

Theorem 1 *If $d^k = 0$ for some k, the triple $(x^k, \lambda^{k+1}, \mu^{k+1})$ is a K-K-T point of Problem NLP.*

The above implies the case where the stationary point is achieved in a finite number of iterations. In the below, we consider the case of $d^k \neq 0$. In which case, it is desirable that the vector d^k is a descent search direction of the merit function. We show below that the directional derivative $D(\theta_r(x^k)\ ;\ d^k)$ is negative. First, we estimate the directional derivative of $f(x)$ at a point x^k in a direction d^k.

Lemma 1

$$\nabla f(x^k)^T d^k \leq -(d^k)^T B_k d^k + (\|\lambda^{k+1}\|_1 + \|\mu^{k+1}\|_1)p(x^k) + (\sum_{i=1}^{3} \varepsilon_i)\|d^k\|^2. \tag{12}$$

Proof. It follows from the Cauchy-Schwarz inequality and (1) that

$$\begin{aligned}(d^k)^T(B_k d^k &+ \nabla f(x^k) + \nabla g(x^k)^T \lambda^{k+1} + \nabla h(x^k)^T \mu^{k+1}) \\ &\leq \|d^k\|\ \|B_k d^k + \nabla f(x^k) + \nabla g(x^k)^T \lambda^{k+1} + \nabla h(x^k)^T \mu^{k+1}\| \\ &\leq \varepsilon_1 \|d^k\|^2.\end{aligned}$$

So the expressions (2) and (3) yield

$$\begin{aligned}\nabla f(x^k)^T d^k &\leq -(d^k)^T B_k d^k + (\lambda^{k+1})^T g(x^k) + (\mu^{k+1})^T h(x^k) + (\sum_{i=1}^{3} \varepsilon_i)\|d^k\|^2 \\ &\leq -(d^k)^T B_k d^k + \sum_{i=1}^{m} \lambda_i^{k+1} g_i(x^k) + \sum_{j=1}^{l} |\mu_j^{k+1} h_j(x^k)| + (\sum_{i=1}^{3} \varepsilon_i)\|d^k\|^2 \\ &\leq -(d^k)^T B_k d^k + (\|\lambda^{k+1}\|_1 + \|\mu^{k+1}\|_1)p(x^k) + (\sum_{i=1}^{3} \varepsilon_i)\|d^k\|^2.\end{aligned}$$

∎

Secondly, the next lemma suggests that d^k is a descent search direction of $\theta_r(x)$.

Lemma 2

$$\begin{aligned}&D(\theta_r(x^k)\ ;\ d^k) \\ &\leq -(d^k)^T B_k d^k + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\|d^k\|^2 - \{r - (\|\lambda^{k+1}\|_1 + \|\mu^{k+1}\|_1)\}\ p(x^k) \\ &\leq -\{M_1 - (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\}\|d^k\|^2 - \{r - (\ \|\lambda^{k+1}\|_1 + \|\mu^{k+1}\|_1)\}\ p(x^k) \\ &< 0.\end{aligned} \tag{13}$$

Proof. Let

$$
\begin{aligned}
I_1 &= \{\, i \mid g_i(x^k) = p(x^k),\ \ i \in \{0, 1, \ldots, m\} \,\}, \qquad g_0(x) = 0 \\
I_2 &= \{\, j \mid h_j(x^k) = p(x^k),\ \ j \in \{1, \ldots, l\} \,\}, \\
I_3 &= \{\, j \mid -h_j(x^k) = p(x^k),\ \ j \in \{1, \ldots, l\} \,\}.
\end{aligned}
$$

Since the expressions (4) and (5) implies that

$$
\max_{i \in I_1} (\nabla g_i(x^k)^T d^k) \le \max_{i \in I_1}(-g_i(x^k)) = -p(x^k) \le \varepsilon_4 \|d^k\|^2 - p(x^k),
$$

$$
\max_{j \in I_2} (\nabla h_j(x^k)^T d^k) \le \max_{j \in I_2}(\varepsilon_4 \|d^k\|^2 - h_j(x^k)) = \varepsilon_4 \|d^k\|^2 - p(x^k),
$$

and

$$
\max_{j \in I_3} (-\nabla h_j(x^k)^T d^k) \le \max_{j \in I_3}(\varepsilon_4 \|d^k\|^2 + h_j(x^k)) = \varepsilon_4 \|d^k\|^2 - p(x^k),
$$

we have

$$
\begin{aligned}
D(p(x^k)\ ;\ d^k) &= \max\{\, \max_{i \in I_1}(\nabla g_i(x^k)^T d^k),\ \max_{j \in I_2}(\nabla h_j(x^k)^T d^k),\ \max_{j \in I_3}(-\nabla h_j(x^k)^T d^k) \} \\
&\le \varepsilon_4 \|d^k\|^2 - p(x^k).
\end{aligned}
$$

So, by the property of the directional derivative, we have

$$
\begin{aligned}
& D(\theta_r(x^k)\ ;\ d^k) \\
&= \nabla f(x^k)^T d^k + r\ D(p(x^k)\ ;\ d^k) \\
&\le -(d^k)^T B_k d^k + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\|d^k\|^2 + \Big(\sum_{i=1}^{m} \lambda_i^{k+1} + \sum_{j=1}^{l} |\mu_j^{k+1}| - r \Big)\ p(x^k) \\
&\le -(d^k)^T B_k d^k + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\|d^k\|^2 - \{r - (\|\lambda^{k+1}\|_1 + \|\mu^{k+1}\|_1)\}\ p(x^k) \\
&\le -\{M_1 - (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\}\|d^k\|^2 - \{r - (\ \|\lambda^{k+1}\|_1 + \|\mu^{k+1}\|_1)\}\ p(x^k).
\end{aligned}
$$

The proof is complete. ∎

In order to prove the global convergence property, it is required that the line search procedure terminates in a finite number of iterations at each k. We give a justification for the line search procedure given in the STQP method.

Lemma 3 *The step size α_k has a positive value and $\alpha_k = \beta_{k,t}$ for some finite t.*

Proof. Suppose that, for all $i \geq 1$,

$$\theta_r(x^k + \beta_{k,i}d^k) > \theta_r(x^k) - \nu\beta_{k,i}(d^k)^T B_k d^k. \tag{14}$$

Then we have

$$-\nu(d^k)^T B_k d^k < \frac{\theta_r(x^k + \beta_{k,i}d^k) - \theta_r(x^k)}{\beta_{k,i}}.$$

Since $i \to \infty$ implies $\beta_{k,i} \to +0$, it follows from Lemma 2 that

$$\begin{aligned} -\nu(d^k)^T B_k d^k &\leq \lim_{i\to\infty} \frac{\theta_r(x^k + \beta_{k,i}d^k) - \theta_r(x^k)}{\beta_{k,i}} \\ &= D(\theta_r(x^k); d^k) \\ &\leq -(d^k)^T B_k d^k + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\|d^k\|^2. \end{aligned}$$

Thus we have, by (A4) and (A6),

$$\begin{aligned} 0 &\leq -(1-\nu)(d^k)^T B_k d^k + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\|d^k\|^2 \\ &\leq -\left(\frac{1}{2} - \nu\right) M_1\|d^k\|^2 + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\|d^k\|^2 \\ &\leq -\left\{\left(\frac{1}{2} - \nu\right) M_1 - (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\right\}\|d^k\|^2 \\ &< 0, \end{aligned}$$

which is the contradiction. ∎

The above lemmas guarantee $x^k \in L_r(x^1)$ for any k. Further, since the level set of the merit function at the initial point is compact, there exist positive constants M_3, M_4 and M_5 such that

$$M_3 \geq \sup_{0<\rho<1} \|\nabla^2 f(x^k + \rho d^k)\|, \tag{15}$$

$$M_4 \geq \sup_{0<\rho<1} \|\nabla^2 g_i(x^k + \rho d^k)\|, \quad i = 1, \ldots, m, \tag{16}$$

$$M_5 \geq \sup_{0<\rho<1} \|\nabla^2 h_j(x^k + \rho d^k)\|, \quad j = 1, \ldots, l. \tag{17}$$

The following lemma suggests that the step size α_k is uniformly bounded below.

Lemma 4 *For any k,*

$$\alpha_k \geq \alpha^* = \tau \min\left(\frac{M_1}{M_3 + r(M_4 + M_5)}, 1\right). \tag{18}$$

Proof. Suppose that, for some k, $\alpha_k < \alpha^*$. Then there exists an integer t such that $\alpha_k = \beta_{k,t}$. Since $\alpha^* \leq \tau < 1$, we have $t \geq 2$. So $\beta_{k,t-1}$ satisfies

$$\theta_r(x^k + \beta_{k,t-1}d^k) > \theta_r(x^k) - \nu\beta_{k,t-1}(d^k)^T B_k d^k,$$

which implies

$$\frac{\theta_r(x^k + \beta_{k,t-1}d^k) - \theta_r(x^k)}{\beta_{k,t-1}} > -\nu(d^k)^T B_k d^k. \tag{19}$$

It follows from the mean value theorem that there exist ξ, $(\eta_1, \ldots, \eta_m)$ and $(\zeta_1, \ldots, \zeta_l)$ such that

$$f(x^k + \beta_{k,t-1}d^k) = f(x^k) + \beta_{k,t-1}\nabla f(x^k)^T d^k + \frac{1}{2}\beta_{k,t-1}^2 (d^k)^T \nabla^2 f(\xi) d^k$$

and

$$\begin{aligned} p(x^k + \beta_{k,t-1}d^k) &= \max(0, g_i(x^k + \beta_{k,t-1}d^k), |h_j(x^k + \beta_{k,t-1}d^k)|) \\ &= \max(0, g_i(x^k) + \beta_{k,t-1}\nabla g_i(x^k)^T d^k + \frac{1}{2}\beta_{k,t-1}^2 (d^k)^T \nabla^2 g_i(\eta_i) d^k, \\ &\quad |h_j(x^k) + \beta_{k,t-1}\nabla h_j(x^k)^T d^k + \frac{1}{2}\beta_{k,t-1}^2 (d^k)^T \nabla^2 h_j(\zeta_j) d^k|) \\ &\leq (1 - \beta_{k,t-1})p(x^k) + \frac{1}{2}\beta_{k,t-1}^2 (M_4 + M_5)\|d^k\|^2 + \beta_{k,t-1}\varepsilon_4\|d^k\|^2. \end{aligned}$$

Noting that $\alpha_k = \beta_{k,t} = \tau\beta_{k,t-1} < \alpha^*$, we have

$$\beta_{k,t-1} < \frac{M_1}{M_3 + r(M_4 + M_5)}, \tag{20}$$

which yields

$$\begin{aligned} &\frac{\theta_r(x^k + \beta_{k,t-1}d^k) - \theta_r(x^k)}{\beta_{k,t-1}} \\ &\leq \frac{(f(x^k + \beta_{k,t-1}d^k) - f(x^k)) + r(p(x^k + \beta_{k,t-1}d^k) - p(x^k))}{\beta_{k,t-1}} \\ &\leq \left(\nabla f(x^k)^T d^k + \frac{1}{2}\beta_{k,t-1}(d^k)^T \nabla^2 f(\xi) d^k\right) \\ &\quad + r\left(-p(x^k) + \frac{1}{2}\beta_{k,t-1}(M_4 + M_5)\|d^k\|^2 + \varepsilon_4\|d^k\|^2\right). \end{aligned}$$

Thus, by Lemma 1 and (20), we have

$$\begin{aligned} &\frac{\theta_r(x^k + \beta_{k,t-1}d^k) - \theta_r(x^k)}{\beta_{k,t-1}} \\ &\leq -\frac{1}{2}(d^k)^T B_k d^k + (\|\lambda^{k+1}\|_1 + \|\mu^{k+1}\|_1 - r)p(x^k) + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\|d^k\|^2. \end{aligned} \tag{21}$$

Therefore, by (19) and (21), we have

$$\begin{aligned} 0 &< -\{r - (\|\lambda^{k+1}\|_1 + \|\mu^{k+1}\|_1)\}p(x^k) \\ &\quad - \left(\frac{1}{2} - \nu\right)(d^k)^T B^k d^k + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\|d^k\|^2 \\ &\leq -\{r - (\|\lambda^{k+1}\|_1 + \|\mu^{k+1}\|_1)\}p(x^k) \\ &\quad - \left\{\left(\frac{1}{2} - \nu\right)M_1 - (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + r\varepsilon_4)\right\}\|d^k\|^2 \\ &< 0, \end{aligned}$$

which is the contradiction. Hence the proof is complete. ∎

Using the above lemmas, we obtain the following convergence theorem.

Theorem 2

(22) $$\lim_{k\to\infty} \|d^k\| = 0.$$

Proof. By the choice of the step size α_k, we have

$$\theta_r(x^{k+1}) \leq \theta_r(x^k) - \nu\alpha_k(d^k)^T B_k d^k.$$

Taking the sum of both sides of the above inequality for $k = 1, \ldots, N$,

$$\begin{aligned} \theta_r(x^1) &\geq \theta_r(x^{N+1}) + \nu\sum_{k=1}^{N} \alpha_k(d^k)^T B_k d^k \\ &\geq f(x^{N+1}) + \nu\alpha^* M_1 \sum_{k=1}^{N} \|d^k\|^2. \end{aligned}$$

Since the level set of the merit function is compact and f is continuously differentiable, the function f is bounded below and $f(x^{N+1}) \geq f^{\natural}$ for some $f^{\natural}$. Then we have

$$0 < \nu\alpha^* M_1 \sum_{k=1}^{N} \|d^k\|^2 \leq \theta_r(x^1) - f^{\natural},$$

which implies that the series

$$\sum_{k=1}^{\infty} \|d^k\|^2$$

converges. Thus theorem is proven. ∎

The above theorem guarantees that any accumulation point of the sequence generated by the STQP method is a K-K-T point of Problem NLP. With the additional assumption that all the functions to be treated are convex, any accumulation point of $\{x^k\}$ is actually an optimal solution of Problem NLP.

References

[1] R.S.Dembo and U.Tulowitzki: Sequential truncated quadratic programming methods, in "*Numerical Optimization 1984* (P.T.Boggs, R.H.Byrd and R.B.Schnabel eds.)", SIAM, pp.83-101 (1985).

[2] R.Fontecilla: On inexact quasi-Newton methods for constrained optimization, *ibid*, SIAM, pp.102-118 (1985).

[3] R.Fontecilla: Inexact secant methods for nonlinear constrained optimization, *SIAM J. Numerical Analysis*, Vol.27, pp.154-165 (1990).

[4] S.P.Han: Variable metric methods for minimizing a class of nondifferentiable functions, *Mathematical Programming*, Vol.20, pp.1-13 (1981).

[5] S.P.Han and O.L.Mangasarian: A dual differentiable exact penalty function, *Mathematical Programming*, Vol.25, pp.293-306 (1983).

[6] Y.Y.Lin and J.S.Pang: Iterative methods for large convex quadratic programs - A survey, *SIAM J. Control and Optimization*, Vol.25, No.2, pp.383-411 (1987).

[7] R.H.Nickel and J.W.Tolle: A sparse sequential quadratic programming algorithm, *Journal of Optimization Theory and Applications*, Vol.60, No.3, pp.453-473 (1989).

[8] D.P.O'Leary: A generalized conjugate gradient algorithm for solving a class of quadratic programming problems, *Linear Algebra and its Applications*, Vol.34, pp.371-399 (1980).

[9] H.Yabe, S.Takahashi and N.Yamaki: Secant updates of quasi-Newton methods for solving sparse nonlinear equations, *TRU Mathematics*, Vol.19, pp.75-87 (1983).

[10] J.Zhang, N.H.Kim and L.Lasdon: An improved successive linear programming algorithm, *Management Science*, Vol.31, pp.1312-1331 (1985).