Bull. Korean Math. Soc. **26**(1989), No. 2, pp. 127-134

A METHOD USING PARAMETRIC APPROACH WITH QUASI-NEWTON METHOD FOR CONSTRAINED OPTIMIZATION

YONG-JOON RYANG AND WON-SERK KIM

1. Introduction

This paper proposes a deformation method for solving practical nonlinear programming problems. Utilizing the nonlinear parametric programming technique with Quasi-Newton method [6, 7], the method solves the problem by imbedding it into a suitable one-parameter family of problems.

The approach discussed in this paper was originally developed with the aim of solving a system of structural optimization problems which frequently appears in various kind of engineering design. It is assumed that we have to solve more than one structural problem of the same type.

If an optimal solution of one of these problems is available, then the optimal solutions of the other problems can be easily obtained by using this known problem and its optimal solution as the initial problem of our parametric method.

The method of nonlinear programming does not generally converge to the optimal solution from an arbitrary starting point if the initial estimate is not sufficiently close to the solution. On the other hand, the deformation method described in this paper is advantageous in that it is likely to obtain the optimal solution even if the initial point is not necessarily in a small neighborhood of the solution. The Jacobian matrix of the iteration formula has the special structural features [2, 3].

Section 2 describes nonlinear parametric programming problem imbedded into a one-parameter family of problems. In Section 3 the iteration formulas for one-parameter are developed. Section 4 discusses parametric approach for Quasi-Newton method and gives algorithm for finding the

Received September 29, 1988.
This research was supported by Institute for Basic Science, Inha University.

optimal solution.

2. The parametric problem

The nonlinear programming problem to be solved is assumed to have the form:

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to } g(x)=0, \end{aligned} \tag{1}$$

where x is an n-dimensional vector with components $x_1, x_2, \ldots, x_n$. The functions $f: R^n \to R$ and $g=(g_1, g_2, \ldots g_m): R^n \to R^m$ iare twice continuously differentiable in x.

We suppose that the vector x° is known to be an optimal solution of the problem:

$$\begin{aligned} &\text{minimize} \quad (1/2)\|x-x^\circ\|^2 \\ &\text{subject to} \quad g(x)-g(x^\circ)=0 \end{aligned} \tag{2}$$

where the problem (2) has same as $f(x)$ and $g(x)$ in (1). In practice, the solution of (2) may be considerably easier to obtain than that of (1).

The parametric problem we consider here is

$$\begin{aligned} &\text{minimize} \quad (1-e^{-t})f(x)+(1/2)e^{-t}\|x-x^\circ\|^2 \\ &\text{subject to} \quad (1-e^{-t})g(x)+e^{-t}(g(x)-g(x^\circ))=0, \end{aligned} \tag{3}$$

where t is a scalar parameter. It is verified that (1) and (2) are identified with (3) for $t=\infty$ and $t=0$, respectively. Therefore, an optimal solution of (1) may be obtained by solving (3) parametrically as t increases until $t=\infty$ with the initial condition $x=x^\circ$ for $t=0$.

Consider the Lagrangian for problem (3), i.e.,

$$L(x, \lambda, t)=tf(x)+(1-t)(1/2)\|x-x^\circ\|^2+\lambda(g(x)-(1-t)g(x^\circ)), \tag{4}$$
$$L(x, \lambda, t)=(1-e^{-t})f(x)+e^{-t}(1/2)\|x-x^\circ\|^2+\lambda(g(x)-e^{-t}g(x^\circ)), \tag{5}$$

where λ is multiplier in R^m. The equation (4) and (5) are equivalent for $t\to 1$ and $t\to\infty$.

3. The iteration formula

The gradient with respect to x, λ can be expressed by

$$\nabla L(x, \lambda, t)=\begin{bmatrix}\dfrac{\partial L}{\partial x}\\ \dfrac{\partial L}{\partial \lambda}\end{bmatrix}=\begin{bmatrix}(1-e^{-t})\nabla f(x)+e^{-t}(x-x^\circ)+\nabla g(x)\lambda\\ g(x)-e^{-t}g(x^\circ)\end{bmatrix}. \quad (6)$$

The following Lemma [4] will be fundamental to our convergence analysis.

LEMMA 1. *Let the one-parameter imbedding be*

$$P(x, \lambda, t)=\begin{bmatrix}a(t)\nabla f(x)+b(t)G(x)+\nabla g(x)\lambda\\ g(x)+c(t)H(x)\end{bmatrix}, \quad 0\leq t\leq 1, \qquad (7)$$

where we assume that $a, b, c : [0,1]\to R_1$, $G, H : R^n \to R^n$,

$$P(x,\lambda,t)=\begin{bmatrix}\nabla f(x)+\nabla g(x)\lambda\\ g(x)\end{bmatrix}, \quad \textit{and } P(x^\circ, \lambda^\circ, 0)=0,$$

$x^\circ\in R^n$, $\lambda^\circ\in R^m$. *If* $x(t)$, $\lambda(t)$ *are solutions of the equation* $p(x,\lambda,t)=0$ *for all* $t\in[0,1]$, *then, by the implicit function theorem,* $x(t), \lambda(t)$ *satisfies*

$$P_{(x,\lambda)}(x(t),\lambda(t),t)\ \frac{d(x,\lambda)}{dt}+P_t(x(t),\lambda(t),t)=0 \qquad (8)$$

$$\frac{d(x,\lambda)}{dt}=-P_{(x,\lambda)}(x(t),\lambda(t),t)^{-1}P_t(x(t),\lambda(t),t) \qquad (9)$$

Conversely, If $x(t)$, $\lambda(t)$ *are solutions of* (8) *on* $[0,1]$, *then*

$$\frac{dp}{dt}(x(t),\lambda(t),t)=0 \ \textit{for } t\in[0,1].$$

Hence $P(x(1),\lambda(1),1)=P(x^\circ,\lambda^\circ,0)=0$ *and so* $x(1)$, $\lambda(1)$ *are solution of*

$$\begin{bmatrix}\nabla f(x)+\nabla g(x)\lambda\\ g(x)\end{bmatrix}=0$$

We consider the following equations

$$P(x,\lambda,t)=\begin{bmatrix}(1-t)(x-x^\circ)+t\nabla f(x)+\nabla g(x)\lambda\\ g(x)+(1-t)g(x^\circ)\end{bmatrix} \qquad (10)$$

$$P(x,\lambda,t)=\begin{bmatrix}e^{-t}(x-x^\circ)+(1-e^{-t})\nabla f(x)+\nabla g(x)\lambda\\ g(x)+e^{-t}g(x^\circ)\end{bmatrix} \qquad (11)$$

We calcuate

$$\frac{dp}{dt}=P_t+(P_x, P_\lambda)\cdot\begin{bmatrix}\dfrac{dx}{dt}\\ \dfrac{d\lambda}{dt}\end{bmatrix}$$

$$=\begin{bmatrix}-e^{-t}(x-x^\circ)+e^{-t}\nabla f(x)\\ -e^{-t}g(x^\circ)\end{bmatrix}+$$

$$\begin{bmatrix}e^{-t}I+(1-e^{-t})\nabla^2 f(x)+\nabla^2 g(x)\lambda & \nabla g(x)\\ \nabla g(x)^T & 0\end{bmatrix}\cdot\begin{bmatrix}\dfrac{dx}{dt}\\ \dfrac{d\lambda}{dt}\end{bmatrix} \tag{12}$$

$$=0$$

and obtain the differential equation

$$\begin{bmatrix}\dfrac{dx}{dt}\\ \dfrac{d\lambda}{dt}\end{bmatrix}=-\begin{bmatrix}e^{-t}I+(1-e^{-t})\nabla^2 f(x)+\nabla^2 g(x)\lambda & \nabla g(x)\\ \nabla g(x)^T & 0\end{bmatrix}^{-1}\cdot$$

$$\begin{bmatrix}\nabla f(x)+\nabla g(x)\lambda\\ g(x)\end{bmatrix} \tag{13}$$

Using Euler's rule, the differential equation (13) motivates the iteration formulas

$$\begin{bmatrix}x^{k+1}\\ \lambda^{k+1}\end{bmatrix}=\begin{bmatrix}x^k\\ \lambda^k\end{bmatrix}$$

$$-\begin{bmatrix}e^{-t^k}I+(1-e^{-t^k})\nabla^2 f(x^k)+\nabla^2 g(x^k)\lambda^k & \nabla g(x^k)\\ \nabla g(x^k)^T & 0\end{bmatrix}^{-1}\cdot$$

$$\begin{bmatrix}\nabla f(x^k)+\nabla g(x^k)\lambda^k\\ g(x^k)\end{bmatrix} \tag{14}$$

$$\begin{bmatrix}x^{k+1}\\ \lambda^{k+1}\end{bmatrix}=\begin{bmatrix}x^k\\ \lambda^k\end{bmatrix}-\begin{bmatrix}(1-t^k)I+t^k\nabla^2 f(x^k)+\nabla^2 g(x^k)\lambda^k & \nabla g(x_k)\\ \nabla g(x)^T & 0\end{bmatrix}^{-1}\cdot$$

$$\begin{bmatrix}\nabla f(x^k)+\nabla g(x^k)\lambda^k\\ g(x^k)\end{bmatrix} \tag{15}$$

Condition for existence and uniqueness of solution of the differential equation (13) on t are given by [1, 5]

t^k is chosen by appropriate selection rule, an example of which can be described as follows [1].

LEMMA 2. *Let $h: R\to R$ be a twice continuously differentiable function satisfying*

$$h(1)=1,\quad h'(1)=0.$$

If we can take the sequence $\{t^k\}$ *by*

$$t^{k+1}=h(t^k)=1-\alpha(t^k-1)^2,\ 0<\alpha<1 \text{ with } |t_o-1|\leqq 1, \qquad (16)$$

the iterates of (16) *converge quadratically to* 1.

Condition for existence and uniqueness of solution of the differential equation $P(x(t),\lambda(t),t)=0$ on t are given by [1, 7]

4. Parametric approach for Quasi-Newton method

A Quasi-Newton method on the formula (14), (15) must have the form $\begin{bmatrix} x^{k+1} \\ \lambda^{k+1} \end{bmatrix}=\begin{bmatrix} x^k \\ \lambda^k \end{bmatrix}+S$, where $BS=-\begin{bmatrix} \nabla f(x)+\nabla g(x)\lambda \\ g(x) \end{bmatrix}$.

If we set $K_k=(1-t^k)I+t^kV_k$, and partition B into

$$B_{k+1}=\begin{bmatrix} K_{k+1} & N_{k+1} \\ M_{k+1} & 0 \end{bmatrix},\quad B_k=\begin{bmatrix} K_k & N_k \\ M_k & 0 \end{bmatrix},$$

$$\begin{bmatrix} K_k & N_k \\ M_k & 0 \end{bmatrix}\begin{bmatrix} P_k \\ q_k \end{bmatrix}=\begin{bmatrix} d_k \\ C_k \end{bmatrix},\qquad B_{k+1}S_k=\begin{bmatrix} d_{k+1}-d_k \\ C_{k+1}-C_k \end{bmatrix}$$

then it is easy to calculate matrices (M,N,B) with respect to t, respectively. We now have

$$[(1-t^{k+1})I+t^{k+1}V_{k+1}]P_k+N_{k+1}q_k=d_{k+1}-d_k,\quad M_{k+1}P_k=C_{k+1}-C_l,$$

$$[(1-t^{k+1})I+t^{k+1}V_{k+1}]P_k=d_{k+1}-d_k-J_{k+1}{}^Tq_k,\quad M_{k+1}P_k=J_{k+1}P_k,$$

where $(1-t^k)I+t^kV_k$ and $(1-t^{k+1})I+t^{k+1}V_{k+1}$ are symmetric, $N_{k+1}=M_{k+1}{}^T$, $M_{k+1}=J_{k+1}$, and $N_{k+1}=J_{k+1}{}^T$. Thus it follows that

$$M_{k+1}=M_k+(Z_k-M_kP_k)\bar{P}_k{}^T,$$

where $Z_k=C_{k+1}-C_k$ and $\bar{P}_k=(1/P_k{}^TP_k)P_k$.

Also, we obtain

$$\begin{aligned}&(1-t^{k+1})I+t^{k+1}V_{k+1}\\ &=(1-t^k)I+t^kV_k+(y_k\bar{P}_k{}^T+\bar{P}_ky_k{}^T)\\ &\quad-(1-t^k)(P_k\bar{P}_k{}^T+P_k\bar{P}_kP_k{}^T)\\ &\quad-t^k(V_kP_k\bar{P}_k{}^T+\bar{P}_kP_k{}^TV_k)-2\theta\bar{P}_k\bar{P}_k{}^T,\end{aligned}$$

where

$$\begin{aligned}2\theta&=P_k{}^Ty_k-P_k{}^T((1-t^k)I+t^kV_k)P_k\\ &=P_k{}^Ty_k-(1-t^k)P_k{}^TP_k-t^kP_k{}^TV_kP_k\end{aligned}$$

and

$$y_k=d_{k+1}-d_k-M_{k+1}{}^Tq_k.$$

Since B_{k+1} now leads to the following formulation

$$B_{k+1}=B_k+\alpha_k\beta_k^T+\beta_k\alpha_k^T,$$

where

$$\alpha_k=\begin{bmatrix} y_k-((1-t^k)I+t^kV_k)P_k-\theta\bar{P}_k \\ Z_k-M_kP_k \end{bmatrix}$$

$$\beta_k=\begin{bmatrix}\bar{P}_k \\ 0\end{bmatrix}.$$

On the basis of the iteration formula (14), (15), we state the following algorithm.

ALGORITHM.

step 1: Obtain an optimal solution (x°, λ°, t°) of problem (4) for $t=0$ by an appropriate method.
Set $k=k^\circ$ and compute $B_\circ$.

step 2: Find s_k such that $B_kS_k=-\begin{bmatrix} d_k \\ c_k\end{bmatrix}$

step 3: Determine

$$\begin{bmatrix} x^{k+1} \\ \lambda^{k+1}\end{bmatrix}=\begin{bmatrix} x^k \\ \lambda^k\end{bmatrix}+S^k$$

step 4: If $t=1$, terminate.
Otherwise, compute B_{k+1}, $t^{k+1}=h(t^k)$ and
set $k=k+1$,
go to step 2.

In the following we deal with convergence of sequences. The local convergence of the iteration (15) follows from results of Ortega and Rheinboldt as stated below.

THEOREM 1. *Let $x^*\in R^*$ be a solution of* (1) *with associated Lagrange multiplier λ^* and let N be a sufficiently small neighborhood of* (x^*, λ^*). *If* $(x^\circ, \lambda^\circ)\in N$ *and* $\{t^k\}\to 1$, *then the iteration* (15) *remains in N and converges to x^*. If* $\lim_{k\to\infty}\{t^k\}=1$, *then* (15) *is Q-superlinearly to x^*.*

proof. Ortega and Rheinboldt [8, p. 357]

THEOREM 2. *Suppose that h satisfies* (16) *and* $\{t^k\}$ *defined by* (16) *converges to* 1. *From* (15), (16), *the sequence* $\{x^k, \lambda^k, t^k\}$ *converges locally to* $\{x^*, \lambda^*, 1\}$ *and the convergence is Q-quadratic.*

Proof. Using a parameter selection function (16) we introduce the operator S defined by

$$S(x^k, \lambda^k, t^k) = \begin{bmatrix} x^k \\ \lambda^k \end{bmatrix} - \begin{bmatrix} (1-t^k)I + t^k\nabla^2 f(x^k) + \nabla^2 g(x^k) & \nabla g(x^k) \\ g(x^k)^T & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \nabla f(x^k) + \nabla g(x^k) \\ g^T(x^k) \end{bmatrix}.$$

If the operator T is defined by $T(x, \lambda, t) = (s(x, \lambda, t), h(t))$, where $T: R^{n+m-1} \to R^{n+m+1}$, then Taylor's theorem and the Schwarz inequality gives

$$\begin{aligned} &\|T((x^k, \lambda^k), t^k) - T((x^k, \lambda^k), 1)\| \\ &\leqq (1/2)\|\nabla^2 T((x^*, \lambda^*) + \theta(x^k - x^*, \lambda^k - \lambda^*)), 1 + (1-t^k)\| \cdot \\ &\qquad \|(x^k, \lambda^k, t^k) - (x^*, \lambda^*, 1)\|^2. \end{aligned}$$

Hence we have

$$\lim_{k\to\infty} \frac{\|(x^{k+1}, \lambda^{k+1}, t^{k+1}) - (x^*, \lambda^*, 1)\|}{\|(x^k, \lambda^k, t^k) - (x^*, \lambda^*, 1)\|^2} < \infty$$

for θ between 0 and 1. It means that the rate of convergence is Q-quadratic.

5. Concluding remarks

In this paper, we have proposed the deformation method for solving nonlinear programming problems. It has been assumed that the optimal solution $x(t)$ of problem (1) is continuous with respect to t. In fact, the continuity of $x(t)$ is an indispensable condition to be satisfied when Algorithm is applied. We would mention that the choice of successive parameter values is important in obtaining the optimal solutions. On the other hand, this method may converge quite slowly.

Furthermore, it shoud always be remarked that the nonsingularity of the Jacobian is a standard assumption in many iterative methods such as the method for solving a system nonlinear equations. In addition, it is noted that scaling of the variables of the optimization problems seems to improve the convergence properties of the Algorithm.

References

1. M. Avriel and J. P. Dauer, *A homotopy based approach to unconstrained optimization*, Applicable Analysis, Vol. 8, 1979, 319-336.
2. C. G. Broyden and W. E. Hart, *Quasi-Newton methods for constrained optimization.*

3. R. Fletcher and M. J. D. Powell, *A rapidly convergent descent method for minimization*, Computer Journal, Vol. **6**, 1963, 163-168.
4. G. H. Meyer, *On solving nonlinear equations wtih a one-parameter operator imbedding*, SIAM, J. Numer. Anal., Vol. **5**, 1968, No. 4.
5. H. Mine, M. Hukushima and Y. J. Ryang, *Methods of parametric nonlinear programming*, Int. J. System Sci., Vol. **12**, 1981, No. 1, 95-110.
6. R. A. Tapia, *Quasi-Newton methods for constrained optimization*, Rice University, Texas, Dep. of Math. Sci., Technical Report, 1975.
7. R. A. Tapia, *Diagonalized multiplier methods and Quasi-Newton methods for constrained optimization*, Journal of Optimization Theory and Applications, Vol. **22**, 1977, No. 2.
8. J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970.

Inha University
Incheon 402-751, Korea
and
Kang-nung University
Kangnung 210-320, Korea