Journal of the Operations Research
Society of Japan
Vol. 30, No. 4, December 1987

© 1987 The Operations Research Society of Japan

OPTIMIZATION OF NONLINEAR PROGRAMMING PROBLEMS USING PENALTY FUNCTIONS AND COMPLEX METHOD

Mohammed Abdul Mazeed
Kyushu University

Teruo Sunaga
Kyushu University

Eiji Kondo
Kyushu University

(Received September 26, 1986; Revised February 2, 1987)

Abstract Here the validity of a no-derivative Complex Method for the optimization of constrained nonlinear programming (NLP) problems is discussed. This method, starting with N (N≥n+1, where n is the dimension of the problem) feasible points, determines the optimum by a typical descent method. Though this method is capable of determining a nearly optimal feasible solution, its convergence in the general case is not guaranteed. However, this method has good convergence properties for unconstrained problems; hence transformation of the constrained NLP problem to a series of smooth unconstrained problems by the use of a penalty function and the application of the complex method to these functions is proposed for the determination of the optimum. A number of problems are solved by the complex method as well as by the use of a penalty function and the complex method, and the results are compared.

1. INTRODUCTION

Perhaps the most commonly used descent method for the minimization of a nonlinear function is the Quasi-Newton Method [2]. However, this method requires derivatives, and also there are cases for which it is difficult to determine the derivative information. There are two well known no-derivative methods: Powell's Conjugate Direction Method [6] for unconstrained problems and Box's Complex Method [1] for constrained problems. Though the algorithm for the latter is simple, its convergence in the general case is not guaranteed. Hence in order to ensure convergence in the general case, we propose the use of a penalty function and the application of a modified version of the complex method to this penalty function for the determination of the optimum.

2. THE COMPLEX METHOD

Here we treat the following NLP problem:

(1) $\text{Min.} f(x) \qquad x \in R^n$

s.t.

(2) $g_p(x) \leq 0 \qquad (p=1,\ldots,P).$

Next, we will explain Box's Complex Method briefly. Let N ($N \geq n+1$) feasible points be set as the vertices of the complex (polyhedron). (When $N=n+1$, this method is called the simplex method). Now the point having the highest objective value is rejected, and the centroid for the rest of the points is determined. Then the new vertex is located along the line joining the rejected point and the centroid at a distance equal to or greater than the distance from the rejected point to the centroid. If the new vertex gives a value higher than that of the rejected point, it is replaced by another vertex located half the distance from the new vertex to the centroid. If a constraint is violated, the new vertex is also moved halfway in towards the centroid. Repetition of this procedure for the new complex is known as the Complex Method.

The above procedure performs well when the centroid is a feasible point; this happens when all of the functions are convex. Since we do not confine the search to convex functions, we adopt the following procedure. That is, when the centroid is infeasible or when the objective value at the centroid is higher than the highest value corresponding to the rejected point of the complex, the new vertex is sought on a line joining the rejected point and the point that gives the minimum objective value.

The algorithm for the modified complex method described above is as follows:

Step 0: Let N feasible points $P_1,\ldots,P_N$ be generated using random numbers. Let P_1 be

(3) $x_1 = (x_{1j}) \qquad (j=1,\ldots,n)$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

and the rest of the points $x_s = (x_{sj})$ be generated as

(4) $$x_{sj} = x_{1j} + D_j r_{sj} \qquad (s=2,\dots,N \quad j=1,\dots,n)$$

Here D_j represents interval width corresponding to the variable x_j, and r_{sj} is a pseudo-random number uniformly distributed over [-0.5,0.5].

Step I: Vertices having the maximum and minimum values and the centroid are determined.

Let y_s denote the objective value corresponding to the point P_s. Let

(5) $$y_L = \min_{1\leq s\leq N} y_s, \qquad y_H = \max_{1\leq s\leq N} y_s$$

Let $\overline{P}$ be the centroid of $(N-1)$ points obtained by rejecting P_H and let $\overline{y}$ be the corresponding value. When $\overline{P}$ is feasible and $\overline{y} < y_H$, then go to **Step II** else go to **Step III**.

Step II: The new vertex P^* is sought on a line joining the rejected point and the centroid

(6) $$P^* = \overline{P} + \alpha(\overline{P} - P_H) \qquad (\alpha\geq 1)$$

Let y^* represent the objective value corresponding to the new vertex P^*. When P^* is feasible and $y^* < y_H$, then replace P_H by P^* and go to **Step I** else replace α by $\alpha\beta$ $(0<\beta<1)$ and go to **Step II.**

Step III: The new vertex P^{**} is sought on a line joining the rejected point and the point that gives the minimum objective value.

(7) $$P^{**} = P_L + \alpha(P_L - P_H) \qquad (\alpha\geq 1)$$

Let y^{**} represent the objective value corresponding to the vertex P^{**}. If P^{**} is feasible and $y^* < y_H$, then replace P_H by P^{**} and go to **Step I** else replace α by $\alpha\beta$ $(0<\beta<1)$ and go to **Step III.**

In the above algorithm, $N=2n, \alpha=1.3, \beta=0.5$ are recommended [1]. The

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

convergence criterion of the algorithm is to test if

(8) $y_h - y_L \leqq \varepsilon$

where ε is a positive and sufficiently small number.

The method described here will be referred to as *Method A* in the following.

3. OPTIMIZATION USING PENALTY FUNCTIONS AND COMPLEX METHOD

The complex method was originally designed for NLP problems with linear or nonlinear inequality constraints. It was developed to overcome difficulties encountered in *simplex method* of Spendley et al. [8] and *simplex-like method* of Nelder and Mead [6]. The difficulties with the simplex method are that it does not adopt any method for the accelaration of the search and that the search process is difficult in curving valleys. Hence Nelder and Mead proposed a simplex-like method which adapts itself to the topography of the objective function -- elongating along inclined planes, changing directions in curving valleys and contracting in the neighbourhood of the minimum. But a difficulty with these methods is that when the search process encounters a constraint, the infeasible vertex is withdrawn until it becomes feasible. After many withdrawls, the simplex collapses into $(n-1)$ or fewer dimensions, and the search becomes slow. Furthermore, if the constraint ceases to be active, the collapsed simplex cannot easily expand back to the full n-dimensions. The complex search overcomes these difficulties by using a polyhedron with more than n vertices.

As described above, the complex method was originally designed for constrained NLP problems. However, for a problem for which the optimum is a corner point or the intersection of two or more constraints, the complex method is not very much effective; also it fails to determine the optimum for problems of higher dimensions. This is illustrated through an example in section 4. Here we propose the use of a penalty function and the complex method for the determination of the optimum. The reason for this proposition is that when a function is smooth it can be locally regarded as a hyperplane; then the descent by the complex method is likely to converge to the right solution. Hence it can be expected that the constrained problem (1)-(2) can be conveniently transformed to a smooth unconstrained problem, which in turn

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

allows the application of the complex method for the determination of the optimum.

Next we consider the general NLP problem:

(9) $\text{Min.} f(x)$

s.t.

(10) $g_p(x) \leq 0 \qquad (p=1,\ldots,P)$

(11) $h_q(x) = 0 \qquad (q=1,\ldots,Q)$

For the transformation of the constrained NLP problem (9)-(11) to a smooth unconstrained problem, the use of penalty functions has already been proposed.

Let us consider the following exterior penalty function:

(12) $$F(x,w^{(k)}) = f(x) + w^{(k)}\sum_{p=1}^{P} g_p(x)_+^2 + w^{(k)}\sum_{q=1}^{Q} h_q^2(x)$$

where

(13) $$a_+ = \begin{cases} a & \text{if} \quad a \geq 0 \\ 0 & \text{if} \quad a < 0 \end{cases}$$

For the penalty function (12), convergence is guaranteed only when the penalty parameter $w^{(k)}$ tends to infinity [3]. A penalty function which does not demand this condition has already been reported by Morrison [5] and is stated here.

(14) $$F(x,f^{(k)},w) = \{f(x)-f^{(k)}\}_+^2 + w\sum_{p=1}^{P} g_p(x)_+^2 + w\sum_{q=1}^{Q} h_q^2(x)$$

where $f^{(k)}$ is the estimate of the objective value at any iteration k. The optimal solution to (14) is determined by updating $f^{(k)}$ successively.

Let the solution x^* to problem (9)-(11) be unique and let $f^* = f(x^*)$. Then the following convergence theorem holds.

Convergence Theorem: Let $f^{(1)}$ be the initial estimate of the optimal value f^* such that

(15) $$f^{(1)} \leq f^*$$

Let

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

(16) $\min_x F(x, f^{(k)}, w) = F(x^{(k)}, f^{(k)}, w) \equiv F^{(k)}$

Let $f^{(k)}$ be updated as

(17) $f^{(k+1)} = f^{(k)} + \sqrt{F^{(k)}}$

then $x^{(k)} \rightarrow x^*$, $f^{(k)} \rightarrow f^*$ and $F^{(k)} \rightarrow 0$.

A proof of this theorem can be found in Morrison [5].

Practically the algorithm is terminated when

(18) $F^{(k)} \leq \delta$

for sufficientyly small $\delta>0$.

For the optimization of the smooth unconstrained penalty function (14), here we propose the use of the complex method. This method will be referred to as *Method B* in the following.

REMARK: Since the use of the penalty function (14) is an exterior method, a feasible solution will not be obtained in a finite number of iterations. A feasible, actually optimal, solution may be attained only as a limit of the generated (infinite) sequence $\{ x^{(k)} \}$.

4. NUMERICAL EXAMPLES

In this section the results of some of the well known problems solved by Method A and Method B are presented and compared. The comparison of the results is based on the number of point evaluations and the precision in the solution.

As for Method B, the following convergence criterion

(19) $|1-y_L/y_H| \leq \varepsilon$

is used and the solution $x^{(k-1)}$ determined at the stage k-1 is used as the starting point for the stage k.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

EXAMPLE 1: Rosen-Suzuki's problem

$$\text{Min.} f(x) = x_1^2+x_2^2+2x_3^2+5x_1-5x_2-21x_3+7x_4$$
s.t.
$$g_1(x)=x_1^2+x_2^2+x_3^2+x_4^2+x_1-x_2+x_3-x_4-8 \leq 0$$
$$g_2(x)=x_1^2+2x_2^2+x_3^2+2x_4^2-x_1-x_4-10 \leq 0$$
$$g_3(x)=2x_1^2+x_2^2+x_3^2+2x_1-x_2-x_4-5 \leq 0$$

with a starting point $P_1 = (0,0,0,0)^T$. The optimum for this problem is at $x^* = (0,1,2,-1)^T$ and $f(x^*) = -44$.

Table 1-1 gives the solution determined by Method A at the end of 1600 point evaluations in five trials. For each trial the starting point P_1 is the same, whereas the rest of the vertices of the complex are generated by a different random seed. It can be seen from this table that the precision of the solution obtained even after 1600 point evaluations is not uniform.

Table 1-1: Example 1:Results Obtained by Method A

Number of Evaluations	Trial No. 1	2	3	4	5
200	-44+0.86	-44+0.06	-44+0.06	-44+0.03	-44+0.11
400	-44+0.26	-44+0.004	-44+0.006	$-44+3\times10^{-4}$	-44+0.003
800	-44+0.20	$-44+4\times10^{-5}$	$-44+1.5\times10^{-4}$	$-44+4\times10^{-5}$	$-44+3\times10^{-4}$
1600	-44+0.20	$-44+3\times10^{-5}$	$-44+1.1\times10^{-4}$	$-44+6\times10^{-5}$	$-44+3\times10^{-4}$

Table 1-2 gives the solution obtained by Method A with the restarting procedure. The solution obtained at the end of first 800 point evaluations is used as the starting point for the next 800 point evaluations. The solution obtained by this procedure is better than that observed in Table 1-1. However, the additional 800 point evaluations by the restarting procedure does not increase the precision in the solution very much.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

Table 1-2: Results Obtained Method A with Restarting Procedure

Number of Evaluations	Trial No. 1	2	3	4	5
800	$-44+0.20$	$-44+4\times10^{-5}$	$-44+1\times10^{-4}$	$-44+4\times10^{-5}$	$-44+3\times10^{-4}$
+800	$-44+4\times10^{-10}$	$-44+1\times10^{-5}$	$-44+3\times10^{-5}$	$-44+1\times10^{-6}$	$-44+9\times10^{-5}$

This problem is also solved by Method B. The parameters are taken as $f^{(1)}=-100$, $w=1000$, $\varepsilon=10^{-3}$. The estimate $f^{(k+1)}$ of the optimal value f^* and the total number of point evaluations at the end of each stage k are given in Table 1-3. A comparison of Table 1-1 and Table 1-3 makes it clear that Method A attains a fairly good objective value at the end of 400 point evaluations, whereas the objective value obtained by Method B at the end of the first stage (in which the number of point evaluations are nearly 400) is not so good. The final solution obtained by Method B, however, is precise. This confirms that the descent by the complex method is rapid in the beginning and then becomes very slow, and that the descent by Method B is slow in the beginning however it converges to a precise solution eventually.

Table 1-3: Results Obtained by Method B

Stage k	Trial No. 1	2	3	4	5
1	-44.13* 513	-44.13 423	-44.13 415	-44.13 402	-44.13 415
2	$-44-3\times10^{-4}$ 1068	$-44-3\times10^{-4}$ 1106	$-44-3\times10^{-4}$ 989	$-44-3\times10^{-4}$ 1024	$-44-3\times10^{-4}$ 929
3	$-44-8\times10^{-7}$ 1763	$-44-8\times10^{-7}$ 2100	$-44-8\times10^{-7}$ 1715	$-44-7\times10^{-7}$ 1605	$-44-8\times10^{-7}$ 1810

* indicates $f^{(k+1)}$/No. of point evaluations

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

EXAMPLE 2: Box's Example

The problem posed is as follows:

$$\text{Max.} f(x) = (a_2y_1+a_3y_2+a_4y_3+a_5y_4+7480a_6-100000a_0-50800ba_7 +k_{31}+k_{32}x_2+k_{33}x_3+k_{34}x_4+k_{35}x_5)x_1-24345+a_1x_6$$

where

$$b = x_2+0.01x_3$$
$$x_6 = (k_1+k_2x_2+k_3x_3+k_4x_4+k_5x_5)x_1$$
$$y_1 = k_6+k_7x_2+k_8x_3+k_9x_4+k_{10}x_5$$
$$y_2 = k_{11}+k_{12}x_2+k_{13}x_3+k_{14}x_4+k_{15}x_5$$
$$y_3 = k_{16}+k_{17}x_2+k_{18}x_3+k_{19}x_4+k_{20}x_5$$
$$y_4 = k_{21}+k_{22}x_2+k_{23}x_3+k_{24}x_4+k_{25}x_5$$
$$x_7 = (y_1+y_2+y_3)x_1$$
$$x_8 = (k_{26}+k_{27}x_2+k_{28}x_3+k_{29}x_4+k_{30}x_5)x_1+x_6+x_7$$

Here x_1, x_2, x_3, x_4 and x_5 are independent variables, and the optimum is required to satisfy the following constraints.

$$0 \leq x_1$$
$$1.2 \leq x_2 \leq 2.4$$
$$20 \leq x_3 \leq 60$$
$$9 \leq x_4 \leq 9.3$$
$$6.5 \leq x_5 \leq 7$$
$$0 \leq x_6 \leq 294000$$
$$0 \leq x_7 \leq 294000$$
$$0 \leq x_8 \leq 277200$$

For the values of the constants a_i and the coefficients k_i, please refer to [1].

The initial point is given as $P_1 = (2.52, 2, 37.5, 9.25, 6.8)^T$. The optimum for this problem is at $x^* = (4.53743, 2.4, 60, 9.3, 7.0)^T$ and $f(x^*) = 5280334$.

This problem is attempted in a similar way as in Example 1. Table 2-1 gives the solution obtained by Method A at the end of 4000 point evaluations in five trials. This solution varies to a great degree and is not precise.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

Table 2-1: Example 2:Results Obtained by Method A

Number of Evaluations	Trial No. 1	2	3	4	5
500	5.214×10^6	4.951×10^6	5.245×10^6	5.179×10^6	5.181×10^6
1000	5.244×10^6	5.069×10^6	5.253×10^6	5.189×10^6	5.212×10^6
2000	5.248×10^6	5.235×10^6	5.257×10^6	5.226×10^6	5.221×10^6
4000	5.250×10^6	5.248×10^6	5.266×10^6	5.231×10^6	5.248×10^6

Table 2-2 gives the solution obtained by Method A with the restarting procedure. The solution obtained after 3000 point evaluations is used as a starting point for the next 3000 point evaluations. The solution obtained by this procedure is found to be not so much better than that observed in Table 2-1.

Table 2-2: Results Obtained Method A with Restarting Procedure

Number of Evalua-tions	Trial No. 1	2	3	4	5
3000	5.249×10^6	5.246×10^6	5.266×10^6	5.228×10^6	5.224×10^6
+3000	5.250×10^6	5.275×10^6	5.279×10^6	5.230×10^6	5.278×10^6

This problem is also attempted by Method B. The objective function and the constraints are scaled so that each term of the penalty function (14) is of the same order. The parameters are taken as $f^{(1)}=10^7$, $w=100$, $\varepsilon=10^{-5}$ The estimate $f^{(k+1)}$ of the optimal value f^* and the total number of point evaluations at the end of each stage k are given in Table 2-3. From the results of Table 2-1 and Table 2-3 it is clear that in case of Method A the increase in the objective value is small and the solution obtained after 4000 point evaluations is not precise, and that in case of Method B the increase in the objective value is considerable in each stage and attains a very good objective value at the end of the fourth stage.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

Table 2-3: Results Obtained by Method B

Stage k	Trial No. 1	2	3	4	5
1	5380000 716	5380000 1193	5380000 1047	5380000 912	5380000 1011
2	5283000 1916	5282000 2560	5283000 2280	5283000 2601	5283000 2118
3	5280385 3448	5280382 4548	5280385 4499	5280384 4458	5280385 3841
4	5280335.5 5727	5280335.4 6098	5280336.5 6821	5280335.5 6679	5280335.5 6192

Example 3: Pearson's Example

This is one of the test problems given in the appendix of Himmelblau [4]. It is a problem of 9 variables, 13 nonlinear constraints and 1 lower bound. The objective is to maximize the area of a hexagon in which the maximum diameter is unity.

$$\text{Max.} f(x)=0.5(x_1x_4 - x_2x_3 + x_3x_9 - x_5x_9 + x_5x_8 - x_6x_7)$$

s.t.

$$1 - x_3^2 - x_4^2 \geqq 0$$
$$1 - x_9^2 \geqq 0$$
$$1 - x_5^2 - x_6^2 \geqq 0$$
$$1 - x_1^2 - (x_2-x_9)^2 \geqq 0$$
$$1 - (x_1 - x_5)^2 - (x_2 - x_6)^2 \geqq 0$$
$$1 - (x_1 - x_7)^2 - (x_2 - x_8)^2 \geqq 0$$
$$1 - (x_3 - x_5)^2 - (x_4 - x_6)^2 \geqq 0$$
$$1 - (x_3 - x_7)^2 - (x_4 - x_8)^2 \geqq 0$$
$$1 - x_7^2 - (x_8 - x_9)^2 \geqq 0$$
$$x_1x_4 - x_2x_3 \geqq 0$$
$$x_3x_9 \geqq 0$$
$$- x_5x_9 \geqq 0$$
$$x_5x_8 - x_6x_7 \geqq 0$$
$$x_9 \geqq 0$$

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

with a starting point $P_i = 0$, $i = 1,\ldots,9$. This problem has several local minima, but $f(x^*) = 0.8660$ for all minima.

Table 3-1 gives the solution obtained by Method A at the end of 4000 point evaluations in five trials. In all the five trials, the solution obtained at the end of 4000 point evaluations varies to a great degree and is quite far from the optimum.

Table 3-1: Results Obtained by Method A

Number of Evaluations	Trial No. 1	2	3	4	5
1000	0.50355	0.63819	0.23383	0.43815	0.47121
2000	0.50648	0.65647	0.47245	0.44655	0.48828
3000	0.50718	0.65745	0.47643	0.45016	0.48886
4000	0.50721	0.65958	0.47693	0.45116	0.48892

Table 3-2 gives the solution obtained by Method A with the restarting procedure. The solution obtained at the end of 3000 point evaluations is used as a starting point for the next 3000 point evaluations. The solution obtained by this method is found to be not so much better than that observed in Table 3-1.

Table 3-2: Results Obtained Method A with Restarting Procedure

Number of Evaluations	Trial No. 1	2	3	4	5
3000	0.50718	0.65745	0.47638	0.45016	0.48886
+3000	0.50949	0.65763	0.49585	0.46571	0.77615

Table 3-3 gives the solution obtained by Method B. The parameters are taken as $f^{(1)} = 2$, $w = 1$, $\epsilon = 10^{-3}$. The estimate $f^{(k+1)}$ of the optimal value f^* and the total number of point evaluations at the end of each stage k are given in Table 3-3. A comparison of Table 3-1, Table 3-2 and Table 3-3 makes it clear that Method B determines the optimum precisely whereas Method A even

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

with the restarting procedure fails to do so.

Table 3-3: Results Obtained by Method B

Stage	Trial No.				
k	1	2	3	4	5
1	0.93019 1098	0.93003 1509	0.93023 1481	0.93054 1418	0.93041 1096
2	0.86938 2375	0.86966 2893	0.86966 2694	0.86970 2893	0.86967 2353
3	0.86621 4264	0.86623 5003	0.86622 4234	0.86623 4663	0.86622 3790
4	0.86603 7032	0.86603 7011	0.86603 6164	0.86601 6737	0.86603 5669

5. CONCLUSIONS

The following are observed from the results of our test problems. The complex method (Method A) performs good descent initially, but from there the search process becomes slow and may terminate at a point remote from the optimum. This method, even with the restarting procedure, generally does not increase precision in the solution very much. On the other hand, in the case of the complex method using the penalty function (Method B), though the descent is slow in early stages, it determines the optimum precisely.

It can be concluded here that Box's complex method, originally designed for constrained NLP problems, associates some skepticism with its search process for the constrained problems. Hence the complex method together with the use of a penalty function can be proposed as a good alternative method for the constrained NLP problems. The proposed method seems to be effective for NLP problems for which the derivative information seems difficult to obtain.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

Acknowledgements:

The authors wish to express their thanks to the referees for their valuable comments made on this paper.

REFERENCES

[1] Box, M.J.: A New Method of Constrained Optimization and Comparison with Other Methods, *Computer Journal*, Vol.8, No.42, 1965, pp 42-52.

[2] Fletcher, R.: *Practical Methods of Optimization*, Vol.1, John Wiley and Sons, 1980.

[3] Fletcher, R.: *Practical Methods of Optimization*, Vol.2, John Wiley and Sons, 1981.

[4] Himmelblau, D.M.: *Applied Nonlinear Programming*, McGraw-Hill Inc, 1972

[5] Morrison, D.D.: Optimization by Least Squares, *SIAM J. NUMER. ANAL.*, Vol.5, No.1, March 1968, pp 83-88.

[6] Nelder, J.A. and Mead, R.: A Simplex Method for Function Minimization, *Computer Journal*, Vol.7, 1964, pp 308-313.

[7] Powell, M.J.D.: An efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives, *Computer Journal*, No.7, 1964, pp 155-162.

[8] Spendley, W., Hext, G.R. and Himsworth, F.R.: Sequential Application of Simple Designs in Optimization and Evolutionary Operation, *Technometrics*, Vol.4, 1962, pp 441-461.

Teruo SUNAGA: Department of
Mechanical Engineering,
Faculty of Engineering,
Kyushu University,
6-10-1 Hakozaki, Higashi Ku
Fukuoka 812, Japan.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.