Journal of the Operations Reserch
Society of Japan
Vol. 19, No. 2, June, 1976

AN ITERATION METHOD FOR NONLINEAR PROGRAMMING PROBLEMS

HISASHI MINE and KATSUHISA OHNO, *Kyoto University*

TATSUO NODA, *Toyama College of Technology*

(Received February 26, 1975)

Abstract

The purpose of this paper is to propose a simple and practical iteration method for solving a nonlinear programming problem. It can be shown that the sequence of points generated by the iteration method converges to a local optimum solution of the nonlinear programming problem.

---

*This work was done while the third author was studying at Department of Applied Mathematics and Physics, Kyoto University as a visiting reseacher from October 1974 to March 1975.

1. Introduction

Let R^n be the n-dimensional Euclidean space, and let $h_i(x_1,x_2,\ldots,x_n)$ $(i=1,2,\ldots,m)$ and $f(x_1,x_2,\ldots,x_n)$ be nonlinear and real-valued functions defined on R^n. Consider the following nonlinear programming problem:

(1) Minimize $f(x_1,x_2,\ldots,x_n)$

subject to $(x_1,x_2,\ldots,x_n)\in D$,

where

$$D = \{(x_1,x_2,\ldots,x_n);\ h_i(x_1,x_2,\ldots,x_n) \leq 0,\ i=1,2,\ldots,m\}.$$

In general, it seems difficult to find the global minima of the problem (1). In some practical cases, however, local minima are no less important than global minima. In this paper, consequently, we deal with the problem of finding a local minimum $(\bar{x}_1,\bar{x}_2,\ldots,\bar{x}_n)\in D$ of nonlinear programming problem (1).

Let us introduce slack variables x_{n+i} $(i=1,2,\ldots,m)$ and define functions $g_i(x_1,x_2,\ldots,x_{n+m})$ $(i=1,2,\ldots,m)$ as

$$g_i(x_1,x_2,\ldots,x_{n+m}) = h_i(x_1,x_2,\ldots,x_n) + x_{n+i}^2.$$

Then problem (1) can be rewritten as:

(2) Minimize $f(x_1,x_2,\ldots,x_n)$

subject to

$g_i(x_1,x_2,\ldots,x_{n+m}) = 0$ $(i=1,2,\ldots,m)$,

$x_j\in R$ $(j=1,2,\ldots,n+m)$.

First, Lagrangian function $\phi(x_1,x_2,\ldots,x_{n+m};\ \lambda_1,\lambda_2,\ldots,\lambda_m)$ associated with problem (2) is introduced as

$$(3)\qquad \phi = f(x_1,x_2,\ldots,x_n) + \sum_{i=1}^{m} \lambda_i g_i(x_1,x_2,\ldots,x_{n+m}).$$

Define an (n+m)-dimensional vector x, an m-dimensional vector λ and an m-dimensional vector-valued function g(x) as follows:

$$x = (x_1,x_2,\ldots,x_{n+m}),$$

$$\lambda = (\lambda_1,\lambda_2,\ldots,\lambda_m),$$

and

$$g(x) = (g_1(x), g_2(x), \dots, g_m(x)).$$

Denote by * the transpose. Then, Lagrangian (3) is reduced to

$$\phi(x,\lambda) = f(x) + \lambda(g(x))^*.$$

2. The Main Theorem

It is assumed that $h_i (i=1,2,\dots,m)$ and f are three times continuously differentiable on R^n. Denote by $\partial g(x)/\partial x$ the $m\times(n+m)$ matrix with (i,j) component $\partial g_i(x)/\partial x_j$. Define an $(n+2m)$-dimensional vector $p(x,\lambda)$ and an $(n+2m)\times(n+2m)$ matrix $A(x,\lambda)$ as follows:

$$p(x,\lambda) = (\phi_x(x,\lambda),\ \phi_\lambda(x,\lambda)),$$

and

$$A(x,\lambda) = \begin{bmatrix} \phi_{xx}(x,\lambda) & (\partial g(x)/\partial x)^* \\ \partial g(x)/\partial x & 0 \end{bmatrix},$$

where ϕ_x and ϕ_λ are gradient vectors with components $\partial\phi/\partial x_j$ and $\partial\phi/\partial\lambda_i$ respectively, and ϕ_{xx} is the Hessian matrix with (j,k) component $\partial^2\phi/\partial x_j\partial x_k$. Then the following lemma is well known. (see, for example, Hadley [4, page 101]).

Lemma 1. Suppose that the following conditions (a)-(c) are satisfied:

(a) There exists at least one of solutions $(\bar{x},\bar{\lambda})$, $\bar{x}\in D\times R^m$, $\bar{\lambda}\in R^m$ satisfying $\phi_x(\bar{x},\bar{\lambda}) = 0$ and $\phi_\lambda(\bar{x},\bar{\lambda}) = 0$.

(b) The Hessian matrix $\phi_{xx}(\bar{x},\bar{\lambda})$ is positive definite.

(c) $\quad \text{rank } (\partial g(\bar{x})/\partial x) = m.$

Then for any $x\in W(\bar{x})-\{\bar{x}\}$, it holds that

$$f(\bar{x}) < f(x),$$

where $W(\bar{x})$ is a suitably chosen neighbourhood of $\bar{x}$.

The main theorem in this paper is:

Theorem. Suppose that the same conditions as in lemma 1 are satisfied.

For arbitrary $x^{(0)} \in R^{n+m}$ and $\lambda^{(0)} \in R^m$, define the sequence $\{(x^{(k)}, \lambda^{(k)});\ k=0, 1, \ldots\}$ by the following iteration method:

(4) $$(x^{(k+1)}, \lambda^{(k+1)}) = (x^{(k)}, \lambda^{(k)}) - \alpha ||A(x^{(k)}, \lambda^{(k)})||^{-2} p(x^{(k)}, \lambda^{(k)}) A(x^{(k)}, \lambda^{(k)})$$

($k=0,1,\ldots$), where α is a constant such that $0 < \alpha < 2$ and $||.||$ denotes the Euclidean norm.

Then the sequence $(x^{(k)}, \lambda^{(k)})$ starting from any initial vectors $x^{(0)} \in U_0(\bar{x})$ and $\lambda^{(0)} \in V_0(\bar{\lambda})$ converges to $(\bar{x}, \bar{\lambda})$ given in lemma 1, as k tends to infinity, where $U_0(\bar{x})$ and $V_0(\bar{\lambda})$ are suitably chosen neighbourhoods of $\bar{x}$ and $\bar{\lambda}$, respectively.

3. Preliminaries

Some preliminaries are required to prove the main theorem. Define $E(x,\lambda)$ as

$$E(x,\lambda) = ||p(x,\lambda)||^2 = ||\phi_x(x,\lambda)||^2 + ||\phi_\lambda(x,\lambda)||^2.$$

Lemma 2. If $(\bar{x}, \bar{\lambda})$ satisfies condition (a), then it follows that

$$\text{grad } E(\bar{x}, \bar{\lambda}) = 0.$$

Proof. From condition (a), it follows that

$$\partial E(\bar{x}, \bar{\lambda})/\partial x_j = 2\phi_x((\partial\phi/\partial x_j)_x)^* + 2\phi_\lambda((\partial\phi/\partial x_j)_\lambda)^* = 0 \quad (j=1,2,\ldots,n+m),$$

and

$$\partial E(\bar{x}, \bar{\lambda})/\partial \lambda_i = 2\phi_x((g_i)_x)^* = 0 \quad (i=1,2,\ldots,m).$$

This proves the lemma.

Lemma 3. Let G_j and H_i be the Hessian matrices of $\partial\phi/\partial x_j$ ($j=1,2,\ldots,n+m$) and $\partial\phi/\partial\lambda_i$ ($i=1,2,\ldots,m$) respectively. Define an $(n+2m)\times(n+2m)$ matrix $C(x,\lambda)$ by

$$C(x,\lambda) = \sum_{j=1}^{n+m} (\partial\phi(x,\lambda)/\partial x_j) G_j + \sum_{i=1}^{m} (\partial\phi(x,\lambda)/\partial\lambda_i) H_i.$$

Then it holds that

$$\min_{\|\rho\|=1} ||\rho A(\bar{x},\bar{\lambda})||^2 > ||C(\bar{x},\bar{\lambda})||.$$

Proof. It follows from conditions (b), (c) and the well-known fact of matrix rank (see, for example, Beltrami [1, page 144]) that

$$\text{rank } A(\bar{x},\bar{\lambda}) = n + 2m.$$

Thus

$$\min_{\|\rho\|=1} ||\rho A(\bar{x},\bar{\lambda})||^2 > 0.$$

On the other hand, condition (a) implies that $||C(\bar{x},\bar{\lambda})|| = 0$. This completes the proof.

4. Proof of the Main Theorem

Denote an (n+2m)-dimensional vector $h^{(k)}$ by

$$h^{(k)} = (x^{(k)},\lambda^{(k)}) - (\bar{x},\bar{\lambda}).$$

Define μ_k and an (n+2m)×(n+2m) matrix $H(x,\lambda)$ as follows:

$$\mu_k = ||A(x^{(k)},\lambda^{(k)})||^{-2},$$

and

$$H(x,\lambda) = (A(x,\lambda))^* A(x,\lambda) + C(x,\lambda).$$

Lemma 2 and 3 are useful in proving the main theorem in a similar way to Yamamoto [8]. Since

$$\text{grad } E(x,\lambda) = 2p(x,\lambda)A(x,\lambda),$$

it follows from (4) and lemma 2 that the recurrence relations

$$(5) \qquad h^{(k+1)} = h^{(k)} - \frac{1}{2}\alpha\mu_k\{\text{grad } E(x^{(k)},\lambda^{(k)}) - \text{grad } E(\bar{x},\bar{\lambda})\}$$

hold for k=0,1,... As is easily shown, $2H(x,\lambda)$ is the Hessian matrix of $E(x,\lambda)$. By Taylor's expansion, (5) is rewritten as the following form:

$$(6) \qquad h^{(k+1)} = h^{(k)} - \alpha\mu_k h^{(k)}\int_0^1 H(\bar{x}+t(x^{(k)}-\bar{x}),\bar{\lambda}+t(\lambda^{(k)}-\bar{\lambda}))dt$$

$$= h^{(k)}(I - J_k),$$

where I is the identity matrix and

$$J_k = \alpha\mu_k\int_0^1 H(\bar{x}+t(x^{(k)}-\bar{x}),\bar{\lambda}+t(\lambda^{(k)}-\bar{\lambda}))dt.$$

Let ε be an arbitrarily fixed positive constant such that

(7) $$0 < \varepsilon < \min\{d, e(\alpha)\},$$

where

$$d \equiv \min_{\|\rho\|=1} ||\rho A(\bar{x},\bar{\lambda})||^2,$$

and

$$e(\alpha) = \frac{2-\alpha}{2\alpha+2} ||A(\bar{x},\bar{\lambda})||^2.$$

Then lemma 3 implies that $d > 0$ and $e(\alpha) > 0$ for constant α such as $0 < \alpha < 2$. Thus for any ε satisfying (7), there exists a positive constant δ such that

$$z \in U_0(\bar{x}) \times V_0(\bar{\lambda}) \equiv \{(x,\lambda);\ ||(x,\lambda) - (\bar{x},\bar{\lambda})|| < \delta\}$$

implies

$$||C(z)|| \leqq \varepsilon,$$

$$||A(\bar{x},\bar{\lambda})||^2 - \varepsilon \leqq ||A(z)||^2 \leqq ||A(\bar{x},\bar{\lambda})||^2 + \varepsilon,$$

and

$$d - \varepsilon \leqq \min_{\|\rho\|=1} \rho H(z)\rho^*.$$

Then for $(x^{(k)},\lambda^{(k)}) \in U_0(\bar{x}) \times V_0(\bar{\lambda})$,

$$\mu \equiv \frac{\alpha(d-\varepsilon)}{||A(\bar{x},\bar{\lambda})||^2 + \varepsilon} \leqq \rho J_k \rho^* \leqq \frac{\alpha(||A(\bar{x},\bar{\lambda})||^2 + 2\varepsilon)}{||A(\bar{x},\bar{\lambda})||^2 - \varepsilon} \equiv M,$$

and $0 < \mu < M < 2$. Let $K = \max(|1-\mu|, |1-M|)$. Then it holds that $0 < K < 1$. Clearly

(8) $$1-M \leqq \rho(I - J_k)\rho^* \leqq 1-\mu.$$

Since $I - J_k$ is a symmetric matrix, it follows from a result of the matrix theory (see, for example, Varga [7, page 11]) that

(9) $$\max_{\|\rho\|=1} ||(I - J_k)\rho^*|| = \max_{\|\rho\|=1} |\rho(I - J_k)\rho^*|.$$

By (8) and (9),

(10) $$\max_{\|\rho\|=1} ||\rho(I - J_k)|| \leqq K.$$

Let $\xi^{(k)} = h^{(k)}/||h^{(k)}||$. Then, from (6) and (10), it holds that

$$||h^{(k+1)}|| = ||h^{(k)}||\ ||\xi^{(k)}(I - J_k)||$$
$$\leqq ||h^{(k)}||(\max_{\|\rho\|=1} ||\rho(I - J_k)||)$$
$$\leqq K\ ||h^{(k)}||.$$

This completes the proof of the main theorem.

5. Numerical Example

As a numerical example, let us consider the following Rosen-Suzuki Test Problem [6]:

Minimize

$$f = x_1^2+x_2^2+2x_3^2+x_4^2-5x_1-5x_2-21x_3+7x_4$$

subject to

$$x_1^2+x_2^2+x_3^2+x_4^2+x_1-x_2+x_3-x_4 -8 \leqq 0,$$

$$x_1^2+2x_2^2+x_3^2+2x_4^2-x_1 \quad -x_4-10 \leqq 0,$$

$$2x_1^2+x_2^2+x_3^2 \quad +2x_1-x_2-x_4 -5 \leqq 0.$$

Minimum point and corresponding value of f are given by

$$(\bar{x}_1,\bar{x}_2,\bar{x}_3,\bar{x}_4) = (0,\ 1,\ 2,\ -1)$$

and

$$f(\bar{x}_1,\bar{x}_2,\bar{x}_3,\bar{x}_4) = -44.$$

Let us stop the iteration process (4) if

$$|x_j^{(k+1)} - x_j^{(k)}| < \varepsilon \ (j=1,2,\ldots,7)$$

and

$$|\lambda_i^{(k+1)} - \lambda_i^{(k)}| < \varepsilon \ (i=1,2,3)$$

are satisfied, where ε is a suitably chosen positive constant. The above problem can be also solved by SUMT transformation and Davidon-Fletcher-Powell method [2], [3]. The numerical computations with $\varepsilon = 10^{-4}$ were carried out on a FACOM 230-75 computer of Kyoto University Computation Center. These results are shown in the following table.

Table. Numerical Solutions for Rosen-Suzuki Problem

Present method / SUMT transformation

Initial data (0.0, 0.0,...,0.0) / (0.0, 0.0, 0.0, 0.0)

	α = 0.9	1.3	1.9	SUMT
x_1	-0.04145	-0.04222	-0.04024	0.00735
x_2	1.16725	1.17805	1.09902	1.00012
x_3	1.94409	1.94038	1.98684	1.99927
x_4	-1.07554	-1.07123	-1.05832	-1.00726
f	-43.9042	-43.8589	-43.9141	-43.9983
CPU TIME (sec)	3.9	3.3	2.4	2.6

Initial data (1.0, 1.0,...,1.0) / (1.0, 1.0, 1.0, 1.0)

	α = 0.9	1.3	1.9	SUMT
x_1	0.01438	0.01254	0.00207	0.00104
x_2	0.99685	0.99625	0.99337	0.99886
x_3	2.01153	2.00789	2.00202	1.99924
x_4	-0.97818	-0.98709	-0.99339	-1.00085
f	-44.0120	-44.0076	-43.9834	-43.9927
CPU TIME (sec)	3.7	3.2	2.6	5.8

Initial data (1.1, 1.1,...,1.1) / (1.1, 1.1, 1.1, 1.1)

	α = 0.9	1.3	1.9	SUMT
x_1	0.00111	0.00643	-0.00213	0.00124
x_2	0.99227	0.99246	0.99004	0.99902
x_3	2.00754	2.00481	1.99968	1.99897
x_4	-0.97278	-0.98812	-1.00219	-1.00114
f	-43.9933	-43.9875	-43.9661	-43.9982
CPU TIME (sec)	3.9	3.5	2.9	48.0

6. Remark

The present iteration method (4) is based on the previous method [5], minimizing a sum of squares of nonlinear functions. The present method is simple in the sense that it need not compute the inverse of the Hessian matrix. Moreover, the computation results in §5 show that the present method is rather better than SUMT transformation , so far as computation time is concerned. It should be noted that the present method requires more storage capacity than SUMT transformation.

References

[1] Beltrami, E.J., An Algorithmic Approach to Nonlinear Analysis and Optimization, Academic Press, New York and London, 1970.

[2] Fiacco, A.V. and G.P.McCormic, "Computational Algorithm for the Sequential Unconstrained Minimization Technique for Nonlinear Programming," Management Sci., 10(1964), 601-617.

[3] Fletcher, R. and M.J.D.Powell, "A Rapidly Convergent Descent Method for Minimization," Computer J., 6(1963), 163-168.

[4] Hadley, G., Nonlinear and Dynamic Programming, Addison-Wesley, Reading, Mass., 1964.

[5] Noda, T., "On an Iteration Method for Finding a Local Minimum of $\sum_{i=1}^{m}(f_i(x_1,x_2,\ldots,x_n))^2$ (in Japanese)," Sūgaku (Edited by the Mathematical Society of Japan), 26(1974), 37-40.

[6] Rosen, J.B. and S.Suzuki, "Construction of Nonlinear Programming Test Problems," Comm. ACM, 8(1965), 113.

[7] Varga, R.S., Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.

[8] Yamamoto, T., "On Noda Method for Finding a Local Minimum of

$\sum_{i=1}^{m} (f_i(x_1, x_2, \ldots, x_n))^2$ (in Japanese)," Sūgaku (Edited by the Mathematical Society of Japan), 26(1974), 349-351.