JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 25, No. 3, JULY 1978

TECHNICAL NOTE

On the Convergence Properties of Second-Order Multiplier Methods[1]

D. P. BERTSEKAS[2]

Communicated by O. L. Mangasarian

Abstract. The purpose of this note is to provide some estimates relating to Newton-type methods of multipliers. These estimates can be used to infer that convergence in such methods can be achieved for an arbitrary choice of the initial multiplier vector by selecting the penalty parameter sufficiently large.

Key Words. Multiplier methods, Newton's method, quadratic convergence.

1. Problem Formulation and Main Result

Consider the problem

$$\text{minimize } f(x), \qquad \text{subject to } h(x)=0, \tag{1}$$

where

$$f: R^n \to R, \qquad h: R^n \to R^m, \qquad h=(h_1, h_2, \ldots, h_m)'.$$

Let x^* be a local minimizer and assume the following.

Assumption 1.1. The functions f and h are twice continuously differentiable with Lipschitz continuous Hessians in a neighborhood of x^*. The $n \times m$ matrix $\nabla h(x^*)$ having as columns the gradients $\nabla h_i(x^*)$, $i = 1, \ldots, m$, has full rank, and hence there exists a unique Lagrange multiplier vector

$$y^* = (y^{1*}, \ldots, y^{m*})'$$

[1] This work was supported by Grant No. NSF ENG 74-19332.

[2] Associate Professor, Department of Electrical Engineering and Coordinated Science Laboratory, University of Illinois, Urbana, Illinois.

0022-3239/78/0700-0443$05.00/0

satisfying

$$\nabla f(x^*)+\nabla h(x^*)y^*=0.$$

Furthermore, there holds

$$z'\left[\nabla^2 f(x^*)+\sum_{i=1}^{m} y^{i*}\nabla^2 h_i(x^*)\right]z>0 \quad \text{for all} \quad z\neq 0 \quad \text{with} \quad \nabla h(x^*)'z=0.$$

In the above relations and in the sequel, all vectors are considered to be column vectors. A prime denotes transposition. The usual Euclidean norm in R^n is denoted $|\cdot|$. All derivatives of various functions are with respect to the argument x.

We shall restrict ourselves to the case of equality constraints. A straightforward extension of our analysis to inequality constraints can be obtained in the manner described in Refs. 1–2.

For any scalar c, consider the augmented Lagrangian function

$$L_c(x, y)=f(x)+y'h(x)+\tfrac{1}{2}c|h(x)|^2. \tag{2}$$

We will obtain a result relating to second-order multiplier methods of the form

$$y_{k+1}=y_k+(N_k'B_k^{-1}N_k)^{-1}[h(x_k)-N_k'B_k^{-1}\nabla L_{c_k}(x_k, y_k)], \tag{3}$$

where y_0 is given, $\{c_k\}$ is a penalty parameter sequence with

$$c_{k+1}\geq c_k>0,$$

x_k satisfies

$$|\nabla L_{c_k}(x_k, y_k)|\leq \min\{\gamma_k/c_k, \delta_k|h(x_k)|\}, \tag{4}$$

$\{\gamma_k\}$, $\{\delta_k\}$ are bounded sequences with

$$0\leq\gamma_k, \qquad 0\leq\delta_k,$$

and N_k, B_k are defined by

$$N_k=\nabla h(x_k), \qquad B_k=\nabla^2 L_{c_k}(x_k, y_k). \tag{5}$$

The Newton-type iteration (3) appears in Tapia (Ref. 3) for $c_k=0$ and in Tapia (Ref. 4) and Han (Ref. 5) for $c_k\neq 0$. When

$$c_k\equiv c, \qquad \gamma_k\equiv\delta_k\equiv 0,$$

then (3) reduces to Newton's method applied to maximization of the dual functional

$$q_c(y)=\min_x L_c(x, y),$$

where minimization with respect to x is understood to be local and c is sufficiently large (see, e.g., Ref. 1 and the references quoted therein). As is well known, in the latter case, when y_0 is sufficiently close to y^* and c is sufficiently large, the method converges to y^* with a convergence rate which is at least quadratic. However, the requirements that y_0 be close to y^*, c be constant, and

$$\nabla L_{c_k}(x_k, y_k) = 0$$

for all k represent severe restrictions from the practical point of view. One would like to guarantee convergence even when a good initial choice y_0 is unknown, while, for computational efficiency reasons, it is desirable to allow for inexact minimization and variability of the penalty parameter c. The analysis of this paper is motivated by these concerns.

The main result of the paper is Proposition 1.1 below. It provides some estimates which can be used in the analysis of first-order and second-order multiplier methods. It shows, in particular, that one can compensate for a poor initial estimate y_0 by choosing the penalty parameter sufficiently large. The proposition, except for the estimate (8), appeared in 1973 in Bertsekas (Ref. 6, also see Ref. 2) and in Polyak and Tretyakov (Ref. 7). The special case of the estimate (8) where

$$\gamma = \delta = 0$$

was given in the author's survey paper (Ref. 1, Proposition 6). The proof in that paper is not readily generalizable. The line of argument given here is based on an interesting relation of multiplier methods with Newton-type Lagrangian methods (Lemma 2.1).

Proposition 1.1. Let Assumption 1.1 hold, and let $Y \subset R^m$ be a given bounded set, and γ, δ be given scalars with

$$0 \le \gamma, \qquad 0 \le \delta.$$

Then there exist nonnegative scalars c^*, M, $\hat{M}$ (depending on Y, γ, δ, f, h, and x^*) such that:

(a) For every

$$c > c^*, \qquad y \in Y,$$

and every vector

$$a \in R^n \qquad \text{with} \qquad |a| \le \gamma/c,$$

there exists a unique vector denoted $x_a(y, c)$ within some open sphere centered at x^* that satisfies

$$\nabla L_c[x_a(y, c), y] = a,$$

and is such that $\nabla h[x_a(y, c)]$ has full rank and $\nabla^2 L_c[x_a(y, c), y]$ is positive definite.

(b) For every

$$c > c^*, \qquad y \in Y,$$

and every vector

$$a \in R^n \qquad \text{with} \qquad |a| \leq \gamma/c$$

for which the vector $x_a(y, c)$ defined in (a) above satisfies

$$|a| \leq \delta |h[x_a(y, c)]|,$$

we have

$$|x_a(y, c) - x^*| \leq M(2\delta + 1)|y - y^*|/c, \tag{6}$$

$$|y_a(y, c) - y^*| \leq M(2\delta + 1)|y - y^*|/c, \tag{7}$$

$$|\hat{y}_a(y, c) - y^*| \leq \hat{M}(2\delta + 1)^2 |y - y^*|^2/c^2, \tag{8}$$

where $y_a(y, c)$, $\hat{y}_a(y, c)$ are defined by

$$y_a(y, c) = y + ch[x_a(y, c)], \tag{9}$$

$$\hat{y}_a(y, c) = y + \{\nabla h[x_a(y, c)]'[\nabla^2 L_c[x_a(y, c), y]]^{-1} \nabla h[x_a(y, c)]\}^{-1} \{h[x_a(y, c)] \\ - \nabla h[x_a(y, c)]'[\nabla^2 L_c[x_a(y, c), y]]^{-1} \nabla L_c[x_a(y, c), y]\}. \tag{10}$$

The proof of Proposition 1.1 is given in the next section. The proposition is not in itself a convergence or rate-of-convergence result for any specific algorithm. Rather, it may be viewed as an aid for stating and analyzing algorithms of the multiplier type similarly as in Refs. 1, 2, 6, and 7.

2. Proof of Proposition 1.1

As mentioned in the previous section, all the statements of Proposition 1.1 have been established earlier, with the exception of the estimate (8). We use these statements in the proof of (8).

For a given triple

$$(x, y, c) \in R^n \times R^m \times R,$$

consider the system of equations in $(\hat{x}, \hat{y})$

$$\begin{bmatrix} \nabla^2 L_c(x, y) & \nabla h(x) \\ \nabla h(x)' & 0 \end{bmatrix} \begin{bmatrix} \hat{x} - x \\ \hat{y} - y \end{bmatrix} = -\begin{bmatrix} \nabla L_c(x, y) \\ h(x) \end{bmatrix}. \tag{11}$$

Note that a system of this type is solved at each iteration of Newton's method applied to the system of necessary conditions

$$\nabla L_c(x, y) = 0, \qquad h(x) = 0.$$

Notation. For a triple (x, y, c) for which the matrix on the left-hand side of (11) is invertible, we denote by $\hat{x}(x, y, c)$, $\hat{y}(x, y, c)$ the unique solution of (11) in $(\hat{x}, \hat{y})$ and say that $\hat{x}(x, y, x)$, $\hat{y}(x, y, c)$ are *well defined.*

Note that, if for a triple (x, y, c) the matrices

$$\nabla^2 L_c(x, y) \qquad \text{and} \qquad \nabla h(x)'[\nabla^2 L_c(x, y)]^{-1}\nabla h(x)$$

are invertible, then the vectors $\hat{x}(x, y, c)$, $\hat{y}(x, y, c)$ are well defined and in fact they are given by (Refs. 3–4)

$$\hat{y}(x, y, c) = y + [\nabla h(x)'[\nabla^2 L_c(x, y)]^{-1}\nabla h(x)]^{-1}[h(x) - \nabla h(x)'[\nabla^2 L_c(x, y)]^{-1}\nabla L_c(x, y)], \tag{12}$$

$$\hat{x}(x, y, c) = x - [\nabla^2 L_c(x, y)]^{-1}\nabla L_c[x, \hat{y}(x, y, c)]. \tag{13}$$

Our proof of Proposition 1.1 rests on the following lemma, the straightforward proof of which may be found in Bertsekas (Ref. 8).

Lemma 2.1. For a triple (x, y, c), the vectors $\hat{x}(x, y, c)$, $\hat{y}(x, y, c)$ are well defined iff the vectors

$$\hat{x}[x, y + ch(x), 0], \qquad \hat{y}[x, y + ch(x), 0]$$

are well defined. Furthermore, there holds

$$\hat{x}(x, y, c) = \hat{x}[x, y + ch(x), 0], \tag{14}$$

$$\hat{y}(x, y, c) = \hat{y}[x, y + ch(x), 0]. \tag{15}$$

We now show (8). We have, for

$$y \in Y, \qquad c > c^*,$$

and $a \in R^n$ for which

$$|a| \leq \min\{\gamma/c, \delta|h[x_a(y, c)]|\},$$

that $\nabla^2 L_c[x_a(y, c), y]$ is positive definite and $\nabla h[x_a(y, c)]$ has full rank. Hence,

$$\hat{x}[x_a(y, c), y, c], \qquad \hat{y}[x_a(y, c), y, c]$$

are well defined; and, from (9), (10), (12), and Lemma 2.1, we obtain

$$\hat{y}_a(y, c) = \hat{y}[x_a(y, c), y, c] = \hat{y}[x_a(y, c), y_a(y, c), 0]. \tag{16}$$

In addition,

$$\hat{x}[x_a(y, c), y_a(y, c), 0], \qquad \hat{y}[x_a(y, c), y_a(y, c), 0]$$

are well defined.

Take now c^* sufficiently high to ensure [see (6), (7)] that $x_a(y, c)$, $y_a(y, c)$ lie within a sufficiently small sphere centered at (x^*, y^*) within which quadratic convergence of Newton's method for the system of equations

$$\nabla L_0(x, y) = 0, \qquad h(x) = 0$$

holds. Then, there is constant K such that, for all

$$c > c^*, \qquad y \in Y,$$

and a with

$$|a| \le \min\{\gamma/c, \delta\, |h[x_a(y, c)]|\},$$

there holds

$$\begin{aligned}(|\hat{x}[x_a(y, c), y_a(y, c), 0] - x^*|^2 + |\hat{y}[x_a(y, c), y_a(y, c), 0] - y^*|^2)^{1/2} \\ \le K\{|x_a(y, c) - x^*|^2 + |y_a(y, c) - y^*|^2\}.\end{aligned} \tag{17}$$

From (6), (7), (16), and (17), we obtain

$$|\hat{y}_a(y, c) - y^*| \le 2KM^2(2\delta + 1)^2 |y - y^*|^2/c^2;$$

and, by setting

$$\hat{M} = 2KM^2,$$

(8) is proved. □

References

1. BERTSEKAS, D. P., *Multiplier Methods: A Survey*, Automatica, Vol. 12, pp. 133–145, 1976.
2. BERTSEKAS, D. P., *On Penalty and Multiplier Methods for Constrained Minimization*, SIAM Journal on Control and Optimization, Vol. 14, pp. 216–235, 1976.
3. TAPIA, R. A., *Newton's Method for Optimization Problems with Equality Constraints*, SIAM Journal on Numerical Analysis, Vol. 11, pp. 874–886, 1974.
4. TAPIA, R. A., *Diagonalized Multiplier Methods and Quasi-Newton Methods for Constrained Optimization*, Journal of Optimization Theory and Applications, Vol. 22, pp. 135–194, 1977.

5. HAN, S. P., *Dual Variable Metric Algorithms for Constrained Optimization*, SIAM Journal on Control and Optimization, Vol. 15, pp. 546–565, 1977.
6. BERTSEKAS, D. P., *Convergence Rate of Penalty and Multiplier Methods*, Proceedings of 1973 IEEE Conference on Decision and Control, San Diego, California, pp. 260–264, 1973.
7. POLYAK, V. T., and TRETYAKOV, N. V., *The Method of Penalty Estimates for Conditional Extremum Problems*, USSR Computational Mathematics and Mathematical Physics, Vol. 13, pp. 42–58, 1974.
8. BERTSEKAS, D. P., *Combined Penalty and Lagrangian Methods for Constrained Optimization*, University of Illinois, Urbana, Illinois, Coordinated Science Laboratory, Working Paper, 1977.