SIAM J. CONTROL AND OPTIMIZATION
Vol. 14, No. 2, February 1976

ON PENALTY AND MULTIPLIER METHODS FOR CONSTRAINED MINIMIZATION*

DIMITRI P. BERTSEKAS†

Abstract. In this paper we consider a generalized class of quadratic penalty function methods for the solution of nonconvex nonlinear programming problems. This class contains as special cases both the usual quadratic penalty function method and the recently proposed multiplier method. We obtain convergence and rate of convergence results for the sequences of primal and dual variables generated. The convergence results for the multiplier method are global in nature and constitute a substantial improvement over existing local convergence results. The rate of convergence results show that the multiplier method should be expected to converge considerably faster than the pure penalty method. At the same time, we construct a global duality framework for nonconvex optimization problems. The dual functional is concave, everywhere finite, and has strong differentiability properties. Furthermore, its value, gradient and Hessian matrix within an arbitrary bounded set can be obtained by unconstrained minimization of a certain augmented Lagrangian.

1. Introduction. One of the most effective methods for solving the constrained optimization problem

$$(1)\qquad \begin{aligned}&\text{minimize } f(x)\\ &\text{subject to } h_i(x)=0, \qquad i=1,\cdots,m,\end{aligned}$$

is the quadratic penalty function method (see, e.g., [6], [12], [13]). This method consists of sequential unconstrained minimization of the function

$$(2)\qquad f(x)+\frac{c_k}{2}\sum_{i=1}^{m}[h_i(x)]^2$$

for an increasing unbounded scalar sequence $\{c_k\}$. The properties of the method are well known, and we refer to [6] for an extensive discussion.

Recently a method, often referred to as the multiplier method, has been proposed and investigated by a number of authors [2]–[5], [7]–[11], [15]–[18] (see [2] and the survey papers [20], [21] for a more detailed account). In this method, the function

$$(3)\qquad f(x)+\sum_{i=1}^{m} y_k^i h_i(x)+\frac{c_k}{2}\sum_{i=1}^{m}[h_i(x)]^2$$

is minimized over x for a sequence of vectors $y_k=(y_k^1,\cdots,y_k^m)'$, and scalars c_k. The function above can be interpreted as a Lagrangian function to which a penalty term has been added. A number of ways of updating of the scalar c_k have been proposed. One possibility is to let c_k increase to infinity in a predetermined fashion. It is also possible to keep c_k fixed after a certain index. The distinctive feature of the method is that after each unconstrained minimization, yielding a

* Received by the editors September 14, 1973, and in revised form February 2, 1975.

† Department of Electrical Engineering and Coordinated Science Laboratory of the University of Illinois, Urbana, Illinois 61801. This research was carried out partly at the Department of Engineering-Economic Systems, Stanford University, Stanford, California, and supported by the National Science Foundation under Grant GK 32870, and partly at University of Illinois, Coordinated Science Laboratory and supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAB-07-72-C-0259.

minimizing point x_k, the vector y_k is updated by means of the iteration

(4) $$y_{k+1} = y_k + c_k h(x_k), \qquad i = 1, \cdots, m,$$

where $h(x_k)$ denotes the column vector $(h_1(x_k), \cdots, h_m(x_k))'$, (prime throughout this paper denotes transposition).

The convergence of iteration (4) to a Lagrange multiplier $\bar{y}$ of the problem has been shown under various assumptions. Global convergence results (i.e., results where the starting point y_0 is not required to be sufficiently close to $\bar{y}$) have been given for convex programming problems in [17], and in [3], [10], [11]. For nonconvex problems, the results available [2], [5], assume boundedness of the penalty parameter sequence $\{c_k\}$ and are local in nature; i.e., convergence has been shown under the assumption that the initial point y_0 is within a sufficiently small neighborhood of $\bar{y}$. Existing rate of convergence results [2] also assume boundedness of the sequence $\{c_k\}$.

All the results mentioned above have been obtained by interpreting the multiplier method as a primal-dual method. In this paper we adopt instead a penalty function viewpoint. Both the quadratic penalty method and the multiplier method are imbedded in a more general penalty function algorithm. In this algorithm, the augmented Lagrangian (3) is minimized for sequences of scalars $\{c_k\}$ and vectors $\{y_k\}$. The only requirement imposed on the sequence $\{y_k\}$ is that it remains within an arbitrary given bounded set S. Thus the quadratic penalty method is obtained as a special case by taking

$$c_k \to \infty \quad \text{and} \quad y_k = 0, \qquad \forall k.$$

The multiplier method is obtained by updating y_k via iteration (4), whenever $y_k + c_k h(x_k) \in S$.

Under assumptions which are specified in the next section, we show that for the general penalty method described above there exist nonnegative scalars c^* and M such that for all $c_k > c^*$ and $y_k \in S$, we have

(5) $$\|x_k - \bar{x}\| \leqq M \|y_k - \bar{y}\| / c_k$$

and

(6) $$\|y_{k+1} - \bar{y}\| \leqq M \|y_k - \bar{y}\| / c_k$$

where $\bar{x}$, $\bar{y}$ are the optimal solution and Lagrange multiplier vector for problem (1), x_k is a point locally minimizing the augmented Lagrangian (3) in a neighborhood of $\bar{x}$, and y_{k+1} is given in terms of c_k, y_k, and x_k by (4). The result mentioned above can be used to establish global convergence of the multiplier method, when S is, for example, an open sphere centered at $\bar{y}$, under the assumption that $c_k > M$, $c_k > c^*$ for all k greater than some index. Furthermore, the result shows that in the multiplier method, the sequence $\{\|y_k - \bar{y}\|\}$ converges at least linearly if c_k is bounded above and superlinearly if $c_k \to \infty$, while in the quadratic penalty method $(y_k = 0)$, the convergence rate is much less favorable. A similar (but sharper) rate of convergence result has been shown in [2] under the assumption that c_k is bounded above.

From the computational point of view, it appears advantageous to carry out the minimization of the augmented Lagrangian only approximately while increasing the accuracy of the approximation with each minimization. We consider this case as well, and we obtain estimates similar to (5), (6), for two different gradient-based termination criteria. The estimates obtained are used in turn to establish global convergence and rate of convergence results for the corresponding algorithms.

In § 4 we use the results obtained to construct a global duality theory much in the spirit of the one recently proposed by Rockafellar [18]. However, our dual functional is continuously differentiable, and its value and gradient can be calculated by unconstrained minimization of the augmented Lagrangian (3) in a manner similar to that for convex programming problems. In this way we are able to interpret multiplier methods as primal-dual methods in a global sense.

For simplicity of presentation, we consider equality constraints only. The analysis, however, applies in its entirety to inequality constraints as well, since such constraints can be converted to equality constraints by using (squared) slack variables. This device, due to Rockafellar [16], results in no loss of computational efficiency and is discussed in § 5.

2. A generalized penalty function algorithm. Consider the nonlinear programming problem

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } h_i(x)=0, \qquad i=1,\cdots,m. \end{aligned} \tag{7}$$

The functions f and h_i for all i are real-valued functions on R^n (n-dimensional Euclidean space). Let $\bar{x}$ be an optimal solution of problem (7). We make the following assumptions concerning the nature of f and h_i in an open ball $B(\bar{x}, \varepsilon)$ of radius $\varepsilon > 0$ centered at $\bar{x}$.

A. The point $\bar{x}$ together with a unique Lagrange multiplier vector $\bar{y}$ satisfies the standard second order sufficiency conditions for $\bar{x}$ to be a local minimum [12, p. 226], i.e.,

A.1. The functions f, h_i, $i=1,\cdots,m$, are twice continuously differentiable within the open ball $B(\bar{x}, \varepsilon)$.

A.2. The gradients $\nabla h_i(\bar{x})$, $i=1,\cdots,m$, are linearly independent, and there exists a unique Lagrange multiplier vector $\bar{y}=(\bar{y}^1,\cdots,\bar{y}^m)'$ such that

$$\nabla f(\bar{x})+\sum_{i=1}^{m} \bar{y}^i \nabla h_i(\bar{x})=0.$$

A.3. The Hessian matrix of the Lagrangian $L_0(x, y)=f(x)+\sum_{i=1}^m y^i h_i(x)$,

$$\nabla^2 L_0(\bar{x}, \bar{y})=\nabla^2 f(\bar{x})+\sum_{i=1}^{m} \bar{y}^i \nabla^2 h_i(\bar{x}),$$

is positive definite on the tangent plane corresponding to the constraints, i.e.,

$$w'\nabla^2 L_0(\bar{x}, \bar{y})w>0$$

for all $w \in R^n$ such that

$$w \neq 0, \qquad w'\nabla h_i(\bar{x}) = 0, \qquad i = 1, \cdots, m.$$

B. The Hessian matrices $\nabla^2 f$, $\nabla^2 h_i$ are Lipschitz continuous within the open ball $B(\bar{x}, \varepsilon)$, i.e., for some $K > 0$, we have for all $x, x' \in B(\bar{x}, \varepsilon)$

$$\|\nabla^2 f(x) - \nabla^2 f(x')\| \leqq K\|x - x'\|$$

and

$$\|\nabla^2 h_i(x) - \nabla^2 h_i(x')\| \leqq K\|x - x'\|, \qquad i = 1, \cdots, m,$$

where $\|\cdot\|$ corresponds to the usual Euclidean norm.

Now let S be an arbitrary bounded subset of R^m. Consider also for any scalar $c > 0$ and any vector $y \in S$ the augmented Lagrangian function

$$L(x, y, c) = f(x) + y'h(x) + \frac{c}{2}\|h(x)\|^2. \tag{8}$$

We shall be interested in algorithms of the following general (and imprecise) form:

Step 1. Given $c_k > 0$, $y_k \in S$, find a (perhaps approximate) minimizing point x_k of the function $L(x, y_k, c_k)$ defined by (8).

Step 2. Determine $c_{k+1} > 0$, $y_{k+1} \in S$ on the basis of x_k, y_k, c_k according to some procedure and return to Step 1.

It is easy to verify that for every $x \in R^n$ we have

$$L(x, y_k, c_k) \geqq f(x) + \frac{c_k}{4}\|h(x)\|^2 - \frac{1}{c}\|y_k\|^2.$$

Hence, as $c_k \to \infty$, we have $L(x, y_k, c_k) \to \infty$ for all sequences $\{y_k\} \in S$, and all infeasible vectors x. It is, thus, evident that one may devise a penalty function method based on sequential unconstrained minimization of $L(x, y_k, c_k)$ for any sequences $\{c_k\} \to \infty$, $\{y_k\} \subset S$. This method exhibits the same convergence properties as the usual quadratic penalty function method [6]. Thus there is no difficulty in showing convergence of some sort for the general alogrithm described earlier whenever $c_k \to \infty$. The question which is most interesting, however, is to determine methods of updating y_k which result in desirable behavior such as accelerated convergence. Before proceeding to a detailed analysis, let us consider a heuristic geometric argument which shows that *it is advantageous to select y_k as close as possible to the Lagrange multiplier $\bar{y}$.*

Let p be the primal functional or perturbation function [19] corresponding to problem (7)

$$p(u) = \min_{h(x)=u.} f(x)$$

In the above equation, the minimization is understood to be local within an appropriate neighborhood of $\bar{x}$. Also p is defined locally on a neighborhood of

$u=0$. It is known that

$$p(0)=f(\bar{x})=\text{optimal value of problem (7)},$$
$$\nabla p(0)=-\bar{y}.$$

Now we can write

$$\min_x L(x,y,c)=\min_u \min_{h(x)=u}\left\{f(x)+y'h(x)+\frac{c}{2}\|h(x)\|^2\right\}$$

or

$$\min_x L(x,y,c)=\min_u\left\{p(u)+y'u+\frac{c}{2}\|u\|^2\right\}.$$

The above equation can be interpreted geometrically as shown in Fig. 1. Notice that the addition of $(c/2)\|u\|^2$ to $p(u)$ has an important convexification effect. It can be seen from Fig. 1 that as y is closer to the Lagrange multiplier $\bar{y}$, the corresponding value $\min_x L(x,y,c)$ is closer to the optimal value of the problem. This fact will also be brought out by the analysis that follows.

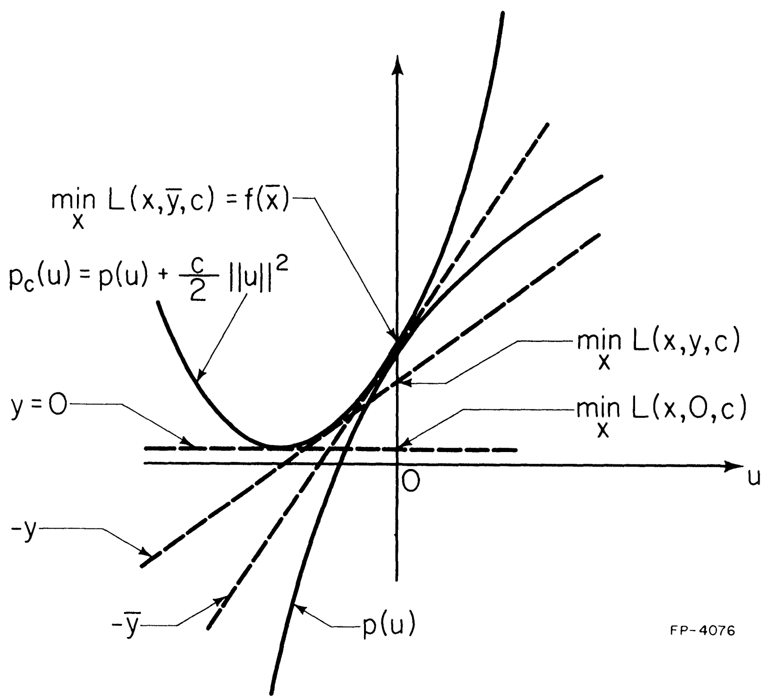

FIG. 1

The preceding argument leads us to the conclusion that the convergence of the generalized penalty algorithm could be accelerated if, at the kth minimization, a vector y_k, where $y_k \to \bar{y}$, were to be used in the augmented Lagrangian (8) in place of $y=0$ which is used in the ordinary penalty function method. Such vectors are readily available. It is easily shown that if $x(y, c)$ minimizes $L(x, y, c)$, then the vector

$$\tilde{y} = y + ch[x(y, c)]$$

is an approximation to the Lagrange multiplier $\bar{y}$ in the sense that $\lim_{c\to\infty} \tilde{y} = \bar{y}$. Thus we are led to a particular scheme whereby at the end of each minimization, the vector y is updated by means of the equation above. This iteration is identical to the one used in the method of multipliers. In the analysis that follows, it is shown that this iteration leads to a much faster convergence rate than the one of the ordinary penalty method. Furthermore, in order for the iteration to converge to $\bar{y}$, it is not necessary to increase c_k to infinity.

PROPOSITION 1. *There exists a scalar $c_1^* \geqq 0$ such that for every $c > c_1^*$, and $y \in S$, the augmented Lagrangian $L(x, y, c)$ of* (8) *has a unique minimizing point $x(y, c)$ with respect to x within some open ball centered at $\bar{x}$. Furthermore, for some scalar $M_1 > 0$ we have*

$$\|x(y, c) - \bar{x}\| \leqq \frac{M_1\|y - \bar{y}\|}{c} \quad \forall c > c_1^* \text{ and } y \in S \tag{9}$$

and

$$\|\tilde{y}(y, c) - \bar{y}\| \leqq \frac{M_1\|y - \bar{y}\|}{c} \quad \forall c > c_1^* \text{ and } y \in S, \tag{10}$$

where the vector $\tilde{y}(y, c) \in R^m$ is given by

$$\tilde{y}(y, c) = y + ch[x(y, c)]. \tag{11}$$

The proof of the above proposition is given in the next section. The result of the proposition has been proved for the case of the pure quadratic penalty method ($y = 0$) by Polyak [14] under the additional assumption that the Hessian matrix $\nabla^2 L_0$ in assumption A.3 is positive definite over the whole space, i.e., local strong convexity holds. Our proof is based in part on Polyak's analysis.

Some important conclusions can now be obtained from the result of Proposition 1. Assuming that $0 \in S$, we have that in the quadratic penalty method ($y_k = 0$), we obtain convergence if $c_k \to \infty$ and, furthermore, the sequences $\{x(0, c_k)\}$, $\{\tilde{y}(0, c_k)\}$ converge to $\bar{x}$, $\bar{y}$, respectively, at least as fast as $M_1\|\bar{y}\|/c_k$. It is evident, however, from the proposition that a great deal can be gained if the vector y_k is not held fixed but rather is updated by means of the iteration of the multiplier method

$$y_{k+1} = \tilde{y}(y_k, c_k) = y_k + c_k h[x(y_k, c_k)]. \tag{12}$$

In order to guarantee that the sequence $\{y_k\}$ remains bounded, we require that the updating takes place provided the resulting vector y_{k+1} belongs to the set S. Otherwise $y_{k+1} = y_k$, i.e., y_k is left unchanged. Of course, the choice of S is arbitrary, and in particular, we can assume that S contains $\bar{y}$ as an interior point. Under these circumstances, we have that if $c_k \to \infty$, then

$$\lim_{k\to\infty} \frac{\|y_{k+1} - \bar{y}\|}{\|y_k - \bar{y}\|} = 0,$$

i.e., the sequence $\{y_k\}$ converges to $\bar{y}$ superlinearly. If $c_k \to c < \infty$, where c is sufficiently large (large enough to ensure that $c > M_1$, $c > c_1^*$ and that $y_k + ch(x_k)$ belongs to an open sphere centered at $\bar{y}$ and contained in S), then

$$\limsup_{k\to\infty} \frac{\|y_{k+1} - \bar{y}\|}{\|y_k - \bar{y}\|} \leqq \frac{M_1}{c},$$

i.e., $\{y_k\}$ converges to $\bar{y}$ at least linearly with a convergence ratio inversely proportional to c

In conclusion, the method of multipliers defined by (12) converges from an arbitrary starting point within the bounded set S provided c_k is sufficiently large after some index $\bar{k}$, $\bar{y}$ is an interior point of S, and the unconstrained minimizations yield the points $x(y_k, c_k)$ for all $k \geqq \bar{k}$. In addition, the multiplier method offers distinct advantages over the quadratic penalty method in that it avoids the necessity of increasing c_k to infinity, and furthermore, the estimate of its convergence rate is much more favorable. For example, if $c_k = s^k$, $s > 1$, then for the penalty method, we have

$$\|x(0, c_k) - \bar{x}\| \leqq M_1 \|\bar{y}\| s^{-k},$$

while in the multiplier method with $y_0 = 0$,

$$\|x(y_k, c_k) - \bar{x}\| \leqq M_1^{k+1} \|\bar{y}\| s^{-(1+2+\cdots+k)}.$$

The ratio of the two bounds in the above inequalities is

$$\prod_{i=0}^{k-1} \frac{s^i}{M_1}$$

and tends to infinity as $k \to \infty$.

In order to avoid creating false impressions, it is perhaps worthwhile to emphasize the fact that the global convergence property of the method of multipliers concluded above is contingent upon the generation of the points $x(y_k, c_k)$, $k \geqq \bar{k}$, by the unconstrained minimization method employed. These points are, by Proposition 1, well-defined as local minimizing points of $L(x, y_k, c_k)$ which are closest to $\bar{x}$. Naturally $L(x, y_k, c_k)$ may have other local minimizing points to which the unconstrained minimization method may be attracted, and unless after some index the unconstrained minimization method stays in the neighborhood of the same local minimizing point of problem (7), our convergence analysis is invalid and there is no reason to believe that the method of multipliers should do better (or worse) than the penalty method. On the other hand, it should be noted that the usual practice of using the last point x_k of the kth minimization as the starting point of the $(k+1)$-st minimization is helpful in producing sequences $\{x_k\}$ which are close to one and the same local minimizing point of problem (7).

We now turn our attention to a generalized penalty method where, given c_k and y_k the augmented Lagrangian $L(x, y_k, c_k)$ of (8) is not minimized exactly, but rather the minimization process is terminated when a certain stopping criterion is satisfied. We consider two different stopping criteria. Similar criteria have been considered in the past in the context of penalty [14] and multiplier methods [2], [3], [5], [10]. According to the first criterion, minimization of $L(x, y_k, c_k)$ is terminated at a point x_k satisfying

$$\|\nabla L(x_k, y_k, c_k)\| \leqq \gamma_k/c_k, \tag{13}$$

where $\{\gamma_k\}$ is a bounded sequence with $\gamma_k \geqq 0$. According to the second criterion minimization is terminated at a point x_k satisfying

$$\|\nabla L(x_k, y_k, c_k)\| \leqq \min\{\gamma_k/c_k, \gamma_k'\|h(x_k)\|\}, \tag{14}$$

where $\{\gamma_k\}, \{\gamma_k'\}$ are bounded with $\gamma_k \geqq 0$, $\gamma_k' \geqq 0$.

We have the following proposition, the proof of which is given in the next section.

PROPOSITION 2. *There exists a scalar $c_2^* \geqq 0$ such that for every $c > c_2^*$ and $y \in S$ and every vector $a \in R^n$ with $\|a\| \leqq \gamma_k/c$, there exists a unique point $x_a(y, c)$ within some open ball centered at $\bar{x}$ satisfying*

$$\nabla L[x_a(y, c), y, c] = a. \tag{15}$$

Furthermore, for some scalar $M_2 > 0$ we have

$$\|x_a(y, c) - \bar{x}\| \leqq \frac{M_2(\|y-\bar{y}\|^2 + \gamma_k^2)^{1/2}}{c} \quad \forall c > c_2^*,\ y \in S \text{ and } \|a\| \leqq \frac{\gamma_k}{c} \tag{16}$$

and

$$\|\tilde{y}_a(y, c) - \bar{y}\| \leqq \frac{M_2(\|y-\bar{y}\|^2 + \gamma_k^2)^{1/2}}{c} \quad \forall c > c_2^*,\ y \in S \text{ and } \|a\| \leqq \frac{\gamma_k}{c}, \tag{17}$$

where $\tilde{y}_a$ is given by

$$\tilde{y}_a(y, c) = y + ch[x_a(y, c)]. \tag{18}$$

If, in addition, a and $x_a(y, c)$ satisfy

$$\|a\| \leqq \gamma_k'\|h[x_a(y, c)]\| \tag{19}$$

then we have

$$\|x_a(y, c) - \bar{x}\| \leqq \frac{M_2(4(\gamma_k')^2 + 1)^{1/2}\|y-\bar{y}\|}{c} \quad \forall c > c_2^* \text{ and } y \in S, \tag{20}$$

and

$$\|\tilde{y}_a(y, c) - \bar{y}\| \leqq \frac{M_2(4(\gamma_k')^2 + 1)^{1/2}\|y-\bar{y}\|}{c} \quad \forall c > c_2^* \text{ and } y \in S. \tag{21}$$

The proposition above may now be used to establish convergence and rate of convergence results for the iteration

$$y_{k+1} = y_k + c_k h(x_k). \tag{22}$$

This iteration takes place if $y_k + c_k h(x_k) \in S$. Otherwise $y_{k+1} = y_k$, i.e., y_k is left unchanged. The point x_k satisfies either the criterion

$$\|\nabla L(x_k, y_k, c_k)\| \leqq \gamma_k/c_k \tag{23}$$

or the criterion

$$\|\nabla L(x_k, y_k, c_k)\| \leqq \min\{\gamma_k/c_k, \gamma_k'\|h(x_k)\|\}. \tag{24}$$

Furthermore, x_k is the unique point $x_a(y_k, c_k)$ corresponding to $a = \nabla L(x_k, y_k, c_k)$ and closest to $\bar{x}$ in accordance with Proposition 2. It is assumed that the unconstrained minimization algorithm yields such points after a certain index.

It is clear from Proposition 2 that any sequence $\{x_k, y_k\}$ generated by the iteration (22) and the termination criterion (24) converges to $(\bar{x}, \bar{y})$, provided c_k is sufficiently large after a certain index and $\bar{y}$ is an interior point of S. Furthermore, $\{y_k\}$ converges to $\bar{y}$ at least linearly when $c_k \to c < \infty$, and superlinearly when $c_k \to \infty$. However, for the termination criterion (23), linear convergence cannot be guaranteed, and in fact an example given in [2] shows that convergence may not be linear. In addition, for this termination criterion it is necessary to increase c_k to infinity in order to achieve global convergence unless $\{\gamma_k\}$ is a sequence converging to zero.

3. Proofs of Propositions 1 and 2.

Proof of Proposition 1. The proof proceeds in two parts. We first prove the proposition under the following condition:

C. The Hessian matrix of the ordinary Lagrangian function

$$\nabla^2 L_0(\bar{x}, \bar{y}) = \nabla^2 f(\bar{x}) + \sum_{i=1}^{m} \bar{y}^i \nabla^2 h_i(\bar{x})$$

is a positive definite matrix; i.e., local strong convexity holds.

Subsequently, we extend the result to the general case.

Let C hold. For all $x \in B(\bar{x}, \varepsilon)$ and any fixed $y \in S$, $c > 0$, consider the auxiliary variables

$$p = x - \bar{x}, \qquad q = y + ch(x) - \bar{y}, \tag{25}$$

where $h(x)$ is the m-vector with coordinates $h_1(x)$, $i = 1, \cdots, m$. For every $x \in B(\bar{x}, \varepsilon)$ we have

$$\nabla f(x) = \nabla f(\bar{x}) + \nabla^2 f(\bar{x})p + r_1(p) \tag{26}$$

$$\nabla h_i(x) = \nabla h_i(\bar{x}) + \nabla^2 h_i(\bar{x})p + r_2^i(p), \qquad i = 1, \cdots, m, \tag{27}$$

where r_1 and r_2^i are n-vector valued functions of p satisfying

$$\begin{aligned} &r_1(0) = r_2^i(0) = 0, \qquad i = 1, \cdots, m, \\ &\nabla r_1(p) = \nabla^2 f(x) - \nabla^2 f(\bar{x}), \\ &\nabla r_2^i(p) = \nabla^2 h_i(x) - \nabla^2 h_i(\bar{x}), \qquad i = 1, \cdots, m. \end{aligned} \tag{28}$$

By the Lipschitz condition assumption B, we have for all $\|p\| < \varepsilon$,

$$\|\nabla r_1(p)\| \leqq K\|p\| \tag{29}$$

and

$$\|\nabla r_2^i(p)\| \leqq K\|p\| \quad \forall i = 1, \cdots, m. \tag{30}$$

Consider now the augmented Lagrangian $L(x, y, c)$ of (8). We have, by (25), (26) and (27),

$$\nabla L(x, y, c) = \nabla f(x) + \nabla h(x)[y + ch(x)] = \nabla f(\bar{x}) + \nabla^2 f(\bar{x})p + r_1(p) + \sum_{i=1}^{m} (q^i + \bar{y}^i)[\nabla h_i(\bar{x}) + \nabla^2 h_i(\bar{x})p + r_2^i(p)],$$

or equivalently,

$$\nabla L(x, y, c) = \nabla^2 L_0(\bar{x}, \bar{y})p + \nabla h(\bar{x})q + r_3(p, q), \tag{31}$$

where $\nabla h(\bar{x})$ is the $n \times m$ matrix with columns $\nabla h_i(\bar{x})$, and $r_3(p, q) \in R^n$ is the vector defined by

$$r_3(p, q) = r_1(p) + \sum_{i=1}^{m} (q^i + \bar{y}^i) r_2^i(p) + \sum_{i=1}^{m} q^i \nabla^2 h_i(\bar{x})p. \tag{32}$$

We also have from (25),

$$\frac{q + \bar{y} - y}{c} = h(x) = h(\bar{x}) + \nabla h(\bar{x})'p + r_4(p),$$

or equivalently,

$$\nabla h(\bar{x})'p - \frac{1}{c}q = \frac{1}{c}(\bar{y} - y) - r_4(p), \tag{33}$$

where the function $r_4: B(0, \varepsilon) \to R^m$ satisfies

$$r_4(0) = 0, \qquad \nabla r_4^i(p) = \nabla h_i(x) - \nabla h_i(\bar{x}), \qquad i = 1, \cdots, m, \tag{34}$$

and using the Lipschitz condition assumption B,

$$\|\nabla r_4^i(p)\| \leqq (K\|p\| + \|\nabla^2 h_i(\bar{x})\|)\|p\| \leqq (K\varepsilon + \|\nabla^2 h_i(\bar{x})\|)\|p\|. \tag{35}$$

Combining now (31) and (33), we have that in order for a point $x \in B(\bar{x}, \varepsilon)$ to satisfy $\nabla L(x, y, c) = 0$ it is necessary and sufficient that the corresponding point $s = [p', q']'$, as given by (25), solves the equation

$$As = t + r(s), \tag{36}$$

where we use the notation

$$A = \begin{bmatrix} \nabla^2 L_0(\bar{x}, \bar{y}) & \nabla h(\bar{x}) \\ \nabla h(\bar{x})' & -\dfrac{I}{c} \end{bmatrix}, \quad s = \begin{bmatrix} p \\ q \end{bmatrix}, \quad t = \begin{bmatrix} 0 \\ \dfrac{\bar{y} - y}{c} \end{bmatrix}, \quad r(s) = \begin{bmatrix} -r_3(p, q) \\ -r_4(p) \end{bmatrix} \tag{37}$$

and I is the $m \times m$ identity matrix. Concerning $r(s)$, we have, from (28), (32), (34) and (35),

$$r(0) = 0. \tag{38}$$

Furthermore, for any s corresponding to an $x \in B(\bar{x}, \varepsilon)$ we have, by straightforward calculation from (29), (30), (32), (35) and (37),

$$\|\nabla r(s)\| \leqq \alpha \|s\|, \tag{39}$$

where $\alpha > 0$ is a constant depending only on ε.

The proof now follows the pattern of [14] by showing that (36) has a unique solution within the domain of definition of s for any $y \in S$ and $c > c_1^*$, where c_1^* is a sufficiently large constant. We make use of the following two lemmas due to Polyak [14].

LEMMA 1. *The matrix* A *of* (37) *has an inverse for every* $c > 0$. *Furthermore, the inverse is uniformly bounded, i.e., for some* $M_1 > 0$ *and all* $c > 0$,

$$\|A^{-1}\| \leqq 2M_1. \tag{40}$$

LEMMA 2. *The equation* (36), $As = t + r(s)$, *has a unique solution* s^* *within the open ball* $B(0, 8M_1\|t\|) \subset B(0, \varepsilon)$ *for every* $y \in S$ *and every* c *sufficiently large to guarantee that*

$$\|t\| \leqq \min\left\{\frac{1}{16M_1\alpha}, \frac{\varepsilon}{8M_1}\right\},$$

where α, M_1 *are as in* (39), (40). *The solution* s^* *satisfies* $\|s^*\| \leqq M_1\|t\|$.

Now from Lemma 2 and the definition (25), (36), (37), it follows immediately that for every $y \in S$ and $c > c_1^*$, where c_1^* is a sufficiently large constant, the equation

$$\nabla L(x, y, c) = 0 \tag{41}$$

has a unique solution $x(y, c)$ within an open ball centered at $\bar{x}$ satisfying

$$\|x(y, c) - \bar{x}\| \leqq \frac{M_1\|y - \bar{y}\|}{c}, \tag{42}$$

$$\|y + ch[x(y, c)] - \bar{y}\| \leqq \frac{M_1\|y - \bar{y}\|}{c}. \tag{43}$$

Hence, in order to complete the proof of Proposition 1, we only need to show that for c sufficiently large the point $x(y, c)$ is a local minimum of $L(x, y, c)$. To this end, it is sufficient to show that $\nabla^2 L[x(y, c), y, c]$ is positive definite for all $y \in S$ and c sufficiently large. Indeed, we have

$$\begin{aligned}\nabla^2 L[x(y, c); y, c] = \nabla^2 f[x(y, c)] &+ \sum_{i=1}^{m} (y^i + ch_i[x(y, c)])\nabla^2 h_i[x(y, c)] \\ &+ c\nabla h[x(y, c)]\nabla h[x(y, c)]'.\end{aligned}$$

Now the third term in the above expression is a positive semidefinite matrix. The sum of the first two terms, in view of (42), (43), is arbitrarily close to the positive matrix $\nabla^2 L_0(\bar{x}, \bar{y})$ for sufficiently large c. Hence, for all c greater than some c_1^*, $\nabla^2 L[x(y, c), y, c]$ is positive definite and $x(y, c)$ is a local minimum of $L(x, y, c)$. Thus Proposition 1 has been proved under condition C.

In order to extend the proof of Proposition 1 to the general case where condition C is not satisfied, we convert the general nonlinear programming problem (7) to an equivalent locally convex problem for which condition C is satisfied. We achieve local convexity by adding a sufficiently high penalty term to the objective function as first indicated by Arrow and Solow [1].

It is evident that problem (7) is equivalent for every $\mu \geqq 0$ to the following problem

$$\text{minimize } f(x)+\frac{\mu}{2}\|h(x)\|^2 \tag{44}$$
$$\text{subject to } h_i(x)=0, \qquad i=1,\cdots,m.$$

Problem (44) has $\bar{x}$ as an optimal solution and $\bar{y}$ as Lagrange multiplier vector. Now consider the Hessian with respect to x of the ordinary Lagrangian of problem (44). We have

$$\nabla^2 L_\mu(\bar{x},\bar{y})=\nabla^2 L_0(\bar{x},\bar{y})+\mu[\nabla h(\bar{x})\nabla h(\bar{x})'], \tag{45}$$

where $\nabla^2 L_0(\bar{x},\bar{y})$ is the Hessian of the ordinary Lagrangian of problem (7). Using assumption A.3, we have the following easily proved lemma.

LEMMA 3. *There exists a scalar* $\mu^*>0$ *such that for every* $\mu \geqq \mu^*$, *the matrix* $\nabla^2 L_\mu(\bar{x},\bar{y})$ *of* (45) *is positive definite.*

The immediate consequence of the above lemma is that problem (44) satisfies the local convexity condition C for all $\mu \geqq \mu^*$. We apply now the result of Proposition 1 as proved under C and with c replaced by $c-\mu^*$, to problem (44) with $\mu=\mu^*$. We have the following:

There exists $c^* \geqq 0$ such that for all $c-\mu^*>c^*$, $y \in S$ the augmented Lagrangian

$$L(x,y,c)=f(x)+y'h(x)+\frac{c}{2}\|h(x)\|^2 \tag{46}$$

has a unique unconstrained minimum $x(y,c)$ within some open ball centered at $\bar{x}$. From the estimates (9), (10), we obtain that for some constant $M>0$,

$$\|x(y,c)-\bar{x}\| \leqq \frac{M\|y-\bar{y}\|}{c-\mu^*}, \tag{47}$$

$$\|\tilde{y}(y,c)-\bar{y}-\delta(y,c)\| \leqq \frac{M\|y-\bar{y}\|}{c-\mu^*}, \tag{48}$$

where $\tilde{y}$ is given by (11) and the vector $\delta(y,c) \in R^m$ is given by

$$\delta(y,c)=\mu^* h[x(y,c)]. \tag{49}$$

From the equation above and (47), it follows that for some constant $B>0$

$$\|\delta(y,c)\| \leqq \mu^* B\|x(y,c)-\bar{x}\| \leqq \frac{\mu^* BM\|y-\bar{y}\|}{c-\mu^*}. \tag{50}$$

Combining the inequalities (47), (48) and (50), we have

$$\|x(y,c)-\bar{x}\| \leqq \frac{M\|y-\bar{y}\|}{c-\mu^*} \tag{51}$$

and

$$\begin{aligned} \|\tilde{y}(y, c) - \bar{y}\| &\leqq \frac{M\|y - \bar{y}\|}{c - \mu^*} + \|\delta(y, c)\| \\ &\leqq \frac{(M + \mu^* BM)\|y - \bar{y}\|}{c - \mu^*}. \end{aligned} \tag{52}$$

Let $M_1 > 0$ and $c_1^* \geqq c^* + \mu^* > 0$ be any constants such that

$$\frac{M}{c - \mu^*} \leqq \frac{M + \mu^* BM}{c - \mu^*} \leqq \frac{M_1}{c} \quad \forall c \geqq c_1^*. \tag{53}$$

Then the desired estimates (9) and (10) follow immediately from (51), (52) and (53), and the proposition is completely proved. Q.E.D.

Proof of Proposition 2. The proof of Proposition 2 follows similar lines as the proof of Proposition 1. Again we assume first that C holds. For an y $a \in R^n$ with $\|a\| \leqq \gamma_k/c$, we have in place of (36) the equation

$$As = t_a + r(s), \tag{54}$$

where A, s, $r(s)$ are as in (37) and t_a is given by

$$t_a = \begin{bmatrix} a \\ \dfrac{\bar{y} - y}{c} \end{bmatrix}. \tag{55}$$

Now from Lemma 2 with t_a in place of t and using the fact that $\{\gamma_k\}$ is bounded and $\|a\| \leqq \gamma_k/c$, we obtain for some $M_2 > 0$ and all $y \in S$, $c \geqq c_2^*$, where c_2^* is a sufficiently large positive scalar (cf. (42) and (43)),

$$\|x_a(y, c) - \bar{x}\| \leqq M_2 \left(\frac{\|y - \bar{y}\|^2}{c^2} + \|a\|^2 \right)^{1/2} \leqq \frac{M_2(\|y - \bar{y}\|^2 + \gamma_k^2)^{1/2}}{c} \tag{56}$$

and

$$\|\tilde{y}_a(y, c) - \bar{y}\| \leqq M_2 \left(\frac{\|y - \bar{y}\|^2}{c^2} + \|a\|^2 \right)^{1/2} \leqq \frac{M_2(\|y - \bar{y}\|^2 + \gamma_k^2)^{1/2}}{c}, \tag{57}$$

which are the relations to be proved.

Now assume, in addition, that a and $x_a(y, c)$ satisfy

$$\|a\| \leqq \gamma_k' \|h[x_a(y, c)]\|. \tag{58}$$

Then we have, from (56), (57) and (58),

$$\|x_a(y, c) - \bar{x}\| \leqq M_2 \left(\frac{\|y - \bar{y}\|^2}{c^2} + (\gamma_k')^2 \|h[x_a(y, c)]\|^2 \right)^{1/2} \tag{59}$$

and

$$\|\tilde{y}_a(y, c) - \bar{y}\| \leqq M_2 \left(\frac{\|y - \bar{y}\|^2}{c^2} + (\gamma_k')^2 \|h[x_a(y, c)]\|^2 \right)^{1/2}. \tag{60}$$

Using (18), the last relation is written as

$$\|y-\bar{y}+ch[x_a(y,c)]\| \leqq M_2\left(\frac{\|y-\bar{y}\|^2}{c^2}+(\gamma_k')^2\|h[x_a(y,c)]\|^2\right)^{1/2},$$

from which

$$\begin{aligned} c\|h[x_a(y,c)]\| &\leqq M_2\left(\frac{\|y-\bar{y}\|^2}{c^2}+(\gamma_k')^2\|h[x_a(y,c)]\|^2\right)^{1/2}+\|y-\bar{y}\| \\ &\leqq \left(\frac{M_2}{c}+1\right)\|y-\bar{y}\|+M_2\gamma_k'\|h[x_a(y,c)]\|. \end{aligned}$$

Thus, finally, we have

$$\|h[x_a(y,c)]\| \leqq \frac{c+M_2}{c(c-M_2\gamma_k')}\|y-\bar{y}\|.$$

For $c \geqq (1+2\gamma_k')M_2$, the inequality above yields

$$\|h[x_a(y,c)]\| \leqq \frac{2\|y-\bar{y}\|}{c}. \tag{61}$$

Substitution of (61) in (59) and (60) yields

$$\|x_a(y,c)-\bar{x}\| \leqq \frac{M_2(1+4(\gamma_k')^2)^{1/2}\|y-\bar{y}\|}{c}$$

and

$$\|\bar{y}_a(y,c)-\bar{y}\| \leqq \frac{M_2(1+4(\gamma_k')^2)^{1/2}\|y-\bar{y}\|}{c},$$

which are the desired estimates. Thus Proposition 2 is proved under condition C. The extension to the general case is entirely similar to the corresponding extension in Proposition 1 and is omitted. Q.E.D.

4. A global duality framework for the method of multipliers. In this section we utilize the results of § 2 to construct a duality framework for problem (7). In contrast with past formulations for nonconvex problems (see, e.g., [5], [12]), the framework is *global* in nature (at least in as much as the dual variables are concerned). By this we mean that the dual functional is an everywhere defined real-valued concave function. The theory is similar in spirit with the one recently proposed by Rockafellar [18] under weaker assumptions, and the one of Buys [5] which is local in nature. Our construction, however, is more suitable to the analysis of algorithms since in our case *the dual functional has strong differentiability properties.* Furthermore, its value and derivatives within an arbitrary open bounded set may be computed by unconstrained local minimization of the augmented Lagrangian. In this way the iteration of the multiplier method can be interpreted as a gradient iteration in a global sense.

For any vector $u \in R^m$, consider the minimization problem

$$\min_{h(x)=u} f(x). \tag{62}$$

Now by applying the implicit function theorem to the system of equations

$$\nabla f(x)+\sum_{i=1}^{m} y_i \nabla h_i(x)=0, \qquad h_i(x)=u_i, \quad i=1,\cdots,m,$$

and using assumption A, we have the following lemma.

LEMMA 4. *Under assumption* A, *there exist positive scalars β and δ such that for every u with $\|u\|<\beta$, problem* (62) *has a unique solution $x(u)$ within the open ball $B(\bar{x},\delta)$ with a Lagrange multiplier $y(u)$ satisfying $\|y(u)-\bar{y}\|<\delta$. Furthermore, the functions $x(u)$, $y(u)$ are continuously differentiable within the open ball $B(0,\beta)$ and satisfy $x(0)=\bar{x}$, $y(0)=\bar{y}$.*

We define now the *primal functional* $p: B(0,\beta)\to R$ by means of

$$p(u)=\min_{\substack{h(x)=u\\ x\in B(x,\delta)}} f(x)=f[x(u)]. \tag{63}$$

It follows from the implicit function theorem that

$$\nabla p(u)=-y(u), \qquad u\in B(0,\beta), \tag{64}$$

and, since $y(u)$ is continuously differentiable, we have that p is twice continuously differentiable on $B(0,\beta)$. Without loss of generality, we assume that the Hessian matrix of p is uniformly bounded on $B(0,\beta)$.

Now for any $c\geqq 0$, consider the function

$$p_c(u)=p(u)+\frac{c}{2}\|u\|^2.$$

It is clear that for c sufficiently large, the Hessian matrix of p_c is positive definite on $B(0,\beta)$, and hence p_c is strictly convex on $B(0,\beta)$. We define for such c the *dual functional* $d_c: R^m\to R$ by means of

$$d_c(y)=\inf_{u\in B(0,\beta)}\left\{p(u)+\frac{c}{2}\|u\|^2+y'u\right\}=\inf_{u\in B(0,\beta)}\{p_c(u)+y'u\}. \tag{65}$$

We note that this way of defining the dual functional is not unusual, since it corresponds to a perturbation function taking the value $p_c(u)$ on $B(0,\beta)$ and $+\infty$ outside $B(0,\beta)$.

The function d_c of (65) has the following properties which we state as a proposition.

PROPOSITION 3. *Under assumption* A, *for every c for which the Hessian matrix of p_c is positive definite on $B(0,\beta)$, we have the following*:

(a) *The function d_c is a real-valued, everywhere continuously differentiable concave function. Furthermore, it is twice continuously differentiable on the open set $A=\{y\,|\,y=-\nabla p_c(u),\ u\in B(0,\beta)\}$.*

(b) *For any $y\in A$, the infimum in* (65) *is attained at a unique point $u_y\in B(0,\beta)$, and we have $\nabla d_c(y)=u_y$, $\nabla^2 d_c(y)=-[\nabla^2 p_c(u_y)]^{-1}$.*

(c) *The function d_c has a unique maximizing point, the Lagrange multiplier $\bar{y}$.*

Proof. (a) Consider the closure [19, § 7] of the convex function taking the value $p_c(u)$ on $B(0,\beta)$ and $+\infty$ outside $B(0,\beta)$. The closure is essentially strictly convex [19, § 26], and its effective domain is a compact set. Hence its conjugate convex function is real-valued and continuously differentiable [19, Thm. 26.3]. Since this conjugate is simply $-d_c(-y)$, we have that d_c is everywhere finite and continuously differentiable. Also by the conjugacy relation between p_c and d_c, we have

$$
\begin{aligned}
&\nabla d_c[-\nabla p_c(u)] = u \quad \forall u \in B(0, \beta),\\
&\nabla p_c[\nabla d_c(y)] = -y \quad \forall y \in A,
\end{aligned}
\tag{66}
$$

where

$$
A = \{y \mid y = -\nabla p_c(u),\ u \in B(0, \beta)\} = \{y \mid \nabla d_c(y) = u,\ u \in B(0, \beta)\}.
$$

The set on the right above is open by the continuity of ∇d_c, thus implying that the set A is open. Now let $\tilde{y}$ be any point in A and let $\tilde{u} = \nabla d_c(\tilde{y})$. We have that $\tilde{u} \in B(0, \beta)$ and

$$
\nabla p_c(\tilde{u}) = -\tilde{y}.
$$

Applying the implicit function theorem in the equation above, we have that there exists an open ball $B(\tilde{y}, \lambda) \subset A$ and a continuously differentiable function $u(\cdot): B(\tilde{y}, \lambda) \to B(0, \beta)$ such that $u(\tilde{y}) = \tilde{u}$ and

$$
\nabla p_c[u(y)] = -y.
$$

It follows from (66) that

$$
\nabla d_c(y) = u(y) \quad \forall y \in B(\tilde{y}, \lambda).
$$

Since $u(y)$ is continuously differentiable on $B(\tilde{y}, \lambda)$, so is $\nabla d_c(y)$. Hence d_c is twice continuously differentiable at $\tilde{y}$. Since $\tilde{y}$ is an arbitrary point in A, we have that d_c is twice continuously differentiable on A, which was to be proved.

(b) The fact that for $y \in A$ the infimum in (65) is attained at a unique point is evident from the argument above. The formula $\nabla^2 d_c(y) = -[\nabla^2 p_c(u_y)]^{-1}$ follows from (66).

(c) We have, by (66) and the fact that $\nabla p_c(0) = -\bar{y}$,

$$
\nabla d_c(\bar{y}) = 0,
$$

and hence $\bar{y}$ is a maximizing point of d_c. It is a unique maximizing point by the differentiability of p_c. Q.E.D.

We now proceed to show that the value and the derivatives of the dual functional d_c can be obtained by local minimization of the augmented Lagrangian $L(x, y, c)$ of (8) provided c is sufficiently large. Let S be any open bounded subset of R^m. Then for any $y \in S$, by Proposition 1, we have that for c sufficiently large,

$$
\|x(y, c) - \bar{x}\| = \frac{M_1\|y - \bar{y}\|}{c} < \delta,
$$

$$
\|\tilde{y}(y, c) - \bar{y}\| \leqq \frac{M_1\|y - \bar{y}\|}{c} < \delta, \qquad \|\tilde{u}\| < \beta,
$$

where

$$\tilde{y}(y, c) = y + ch[x(y, c)], \qquad \tilde{u} = h[x(y, c)].$$

Furthermore, we have

$$\nabla f[x(y, c)] + \sum_{i=1}^{m} \tilde{y}^i(y, c)\nabla h_i[x(y, c)] = 0.$$

It follows from the implicit function theorem and Lemma 4 that $x(y, c)$ is the unique minimizing point in problem (62) when $u = \tilde{u}$. This implies

$$p(\tilde{u}) = f[x(y, c)], \qquad \nabla p(\tilde{u}) = -\tilde{y}(y, c) = -y - c\tilde{u},$$

and therefore

$$\nabla p_c(\tilde{u}) + y = 0.$$

Hence $y \in A$, $\tilde{u}$ attains the infimum in the right-hand side of (65), and by part (b) of Proposition 3,

$$\nabla d_c(y) = \tilde{u} = h[x(y, c)], \qquad \nabla^2 d_c(y) = -[\nabla^2 p_c(\tilde{u})]^{-1}.$$

Furthermore,

$$\begin{aligned} d_c(y) &= p(\tilde{u}) + y'\tilde{u} + \frac{c}{2}\|\tilde{u}\|^2 \\ &= f[x(y, c)] + y'h[x(y, c)] + \frac{c}{2}\|h[x(y, c)]\|^2 = \min_x L(x, y, c), \end{aligned}$$

where the minimization above is understood to be local in the sense of Proposition 1. In addition, a straightforward calculation [5], [12] yields

$$D_c(y) = \nabla^2 d_c(y) = -\nabla h[x(y, c)]'\{\nabla^2 L[x(y, c), y, c]\}^{-1}\nabla h[x(y, c)], \tag{67}$$

where $\nabla h[x(y, c)]$ is the $n \times m$ matrix having as columns the gradients $\nabla h_i[x(y, c)]$, $i = 1, \cdots, m$, and $\nabla^2 L$ denotes the Hessian matrix of the augmented Lagrangian L with respect to x. Thus we have proved the following proposition.

PROPOSITION 4. *Let S be any open bounded subset of R^m, and let assumptions* A *and* B *hold. Then there exists a scalar $c^* \geqq 0$ such that for every $y \in S$ and every $c > c^*$, the dual functional d_c satisfies*

$$d_c(y) = f[x(y, c)] + y'h[x(y, c)] + \frac{c}{2}\|h[x(y, c)]\|^2 = \min_x L(x, y, c),$$

$$\nabla d_c(y) = h[x(y, c)],$$

where $x(y, c)$ is as in Proposition 1. *Furthermore, d_c is twice continuously differentiable on S and $\nabla^2 d_c(y)$ is given by* (67).

It is now clear that the iteration of the method of the multipliers can be written, for c sufficiently large,

$$y_{k+1} = y_k + c\nabla d_c(y_k),$$

and hence can be viewed as a *fixed step size gradient iteration* for maximizing the dual functional d_c. Thus one may obtain a tight rate of convergence result by utilizing a known result on gradient methods. This result, however, is rather uniformative since it involves the eigenvalues of the matrix D_c of (67) which strongly depend on c. A modified version of this result which is more amenable to proper interpretation is given in [2], together with an analysis of the convergence rate aspects of the method of multipliers in the presence of inexact minimization.

The primal-dual interpretation of the multiplier method suggests also several possibilities for modification of the basic iteration. One such modification was suggested in [2], [3]. Another interesting possibility rests on the fact that when second derivatives are calculated during the unconstrained minimization cycle, then one obtains the Hessian matrix D_c of (67) in addition to the gradient ∇d_c. Thus it is possible to carry out a Newton iteration aimed at maximizing d_c in place of the gradient iteration corresponding to the method of multipliers. It is also possible to use a variable metric method for maximization of d_c. Such possibilities have already been suggested by Buys [5], who in addition provided some local convergence results. It is to be noted, however, that for large scale problems arising, for example, in optimal control, where the number of primal and dual variables may easily reach several hundreds or even thousands, such modifications do not seem to be attractive. This is particularly so since the simple gradient iteration already has excellent convergence rate.

5. Treatment of inequality constraints. As pointed out in the Introduction, inequality constraints may be treated in a simple way by introducing slack variables. Indeed, the problem

$$\min_{g_j(x)\leqq 0,\quad j=1,\cdots,r} f(x) \tag{68}$$

is equivalent to the equality constrained problem

$$\min_{g_j(x)+z_j^2=0,\quad j=1,\cdots,r} f(x), \tag{69}$$

where $z_1, \cdots, z_r$ represent additional variables.

Now assume that $(\bar{x}, \bar{y})$ is an optimal solution–Lagrange multiplier pair for problem (68) satisfying the following second order sufficiency conditions for optimality (which include strict complementarity).

A′. The functions $f, g_j, j=1, \cdots, r$, are twice continuously differentiable within an open ball $B(\bar{x}, \varepsilon)$. The gradients $\nabla g_j(\bar{x}), j\in J(\bar{x})$, with $J(\bar{x}) = \{j | g_j(x)=0\}$, are linearly independent. We have $\nabla f(x) + \sum_{j=1}^r \bar{y}^j \nabla g_j(\bar{x}) = 0$ and $\bar{y}^j \geqq 0$ with $\bar{y}^j > 0$ if and only if $j \in J(\bar{x})$. Furthermore,

$$w'\left[\nabla^2 f(\bar{x}) + \sum_{j=1}^r \bar{y}^j \nabla^2 g_j(\bar{x})\right] w > 0$$

for all $w \neq 0$ such that $w'\nabla g_j(\bar{x}) = 0$ for all $j \in J(\bar{x})$.

Then it is easy to show that $(\bar{x}, |g_1(\bar{x})|^{1/2}, \cdots, |g_r(\bar{x})|^{1/2})$ is an optimal solution of problem (69) satisfying (together with $\bar{y}$) assumption A and hence it is covered by the theory of §§ 2 and 3 provided the Lipschitz assumption B is also satisfied. Thus one may use the multiplier method for solving problem (69) instead of problem (68). On the other hand, slack variables need not be present explicitly in the computations, since the minimization of the augmented Lagrangian,

$$L(x, z, y, c) = f(x) + \sum_{j=1}^{r} y^j[g_j(x) + z_j^2] + \frac{c}{2}\sum_{j=1}^{r} [g_j(x) + z_j^2]^2,$$

can be carried out first with respect to $z_1, \cdots, z_r$, yielding

$$\tilde{L}(x, y, c) = \min_z L(x, z, y, c)$$

$$= f(x) + \frac{1}{2c}\left\{\sum_{j=1}^{r} [\max(0, y^j + cg_j(x))]^2 - (y^j)^2\right\}.$$

The optimal values of z_j are given in terms of x, y, c by

(70) $$z_j^2(x, y, c) = \max[0, -y^j/c - g_j(x)], \qquad j = 1, \cdots, r.$$

Now minimization of $\tilde{L}(x, y, c)$ with respect to x yields a vector $x(y, c)$, and the multiplier method iteration in view of (70) takes the form

(71) $$\begin{aligned} y^j_{k+1} &= y^j_k + c[g_j[x(y, c)] + z_j^2[x(y, c), y, c]] \\ &= \max[0, y^j_k + cg_j[x(y, c)]], \qquad j = 1, \cdots, r. \end{aligned}$$

Also in view of (70), the stopping criterion (14) takes the form

$$\|\nabla\tilde{L}(x_k, y_k, c_k)\| \leqq \min\left\{\frac{\gamma_k}{\mathcal{S}_k}, \gamma'_k\left(\sum_{j=1}^{r}\left[\max\left\{-\frac{y^j_k}{c_k}, g_j(x_k)\right\}\right]^2\right)^{1/2}\right\}.$$

Thus there is no difference in treating equality or inequality constraints, at least within the second order sufficiency assumption framework of this paper.

6. Conclusions. In this paper we provided an analysis of multiplier methods by imbedding them within a general class of penalty function methods. The viewpoint adopted yields strong global convergence results. Furthermore, it provides a fair basis for comparison of multiplier methods with pure penalty function methods. This comparison conclusively demonstrates the superiority of multiplier over penalty methods. The global duality theory obtained has many similarities with the duality theory associated with multiplier methods for convex programming. In addition, it provides a framework within which multiplier methods can be viewed as primal-dual methods in a global sense.

Notes added in proof. The results of this paper have been presented at the 1973 IEEE Decision and Control Conference, San Diego, Calif., Dec. 1973 and at the SIGMAP Symposium on Nonlinear Programming, Madison, Wis., April 1974. They were reported without proofs in [4] and in *Nonlinear Programming 2*, O. L. Mangasarian, R. R. Meyer and S. M. Robinson, eds., Academic Press, New York, 1975, pp. 165–191.

While this paper was under review, results similar as those of Propositions 1 and 2 appeared in B. T. Polyak and N. V. Tret'yakov, *The Method of Penalty Estimates for Conditional Extremum Problems*, U.S.S.R. Comput. Math. and Mathematical Phys., 13(1974), pp. 42–58.

Generalized versions of Propositions 1 and 2, involving augmented Lagrangians with nonquadratic penalty functions and adjusting essentially the same proof as the one given here, are provided in D. P. Bertsekas, *Multiplier Methods: A Survey*, Preprints of IFAC 6th Triennial World Congress, Part IB, Boston, Mass., Aug. 1975.

REFERENCES

[1] K. J. ARROW AND R. M. SOLOW, *Gradient methods for constrained maxima with weakened assumptions*, Studies in Linear and Nonlinear Programming, K. Arrow, L. Hurwicz, and H. Uzawa, eds., Stanford University Press, Stanford, Calif., 1958.

[2] D. P. BERTSEKAS, *Combined primal-dual and penalty methods for constrained minimization*, this Journal, 13 (1975), pp. 521–544.

[3] ———, *On the method of multipliers for convex programming*, EES Dept. Working Paper, Stanford Univ., 1973; IEEE Trans. Automatic Control, AC-20 (1975), pp. 385–388.

[4] ———, *Convergence rate of penalty and multiplier methods*, Proc. 1973 IEEE Conf. on Decision and Control, San Diego, Calif., Dec. 1973, pp. 260–264.

[5] J. D. BUYS, *Dual algorithms for constrained optimization*, Ph.D. thesis, Rijksuniversiteit de Leiden, 1972.

[6] A. V. FIACCO AND G. P. MCCORMICK, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley, New York, 1968.

[7] P. C. HAARHOFF AND J. D. BUYS, *A new method for the optimization of a nonlinear function subject to nonlinear constraints*, Comput. J., 13 (1970), pp. 178–184.

[8] M. R. HESTENES, *Multiplier and gradient methods*, J. Optimization Theory Appl., 4 (1969), pp. 303–320.

[9] B. W. KORT AND D. P. BERTSEKAS, *A new penalty function method for constrained minimization*, Proc. 1972 IEEE Conf. on Decision and Control, New Orleans, Dec. 1972.

[10] ———, *Combined primal-dual and penalty methods for convex programming*, EES Dept. Working Paper, Stanford Univ., 1973; this Journal, 14 (1976), pp. 268–294.

[11] ———, *Multiplier methods for convex programming*, Proc. 1973 IEEE Conf. on Decision and Control, San Diego, Calif., Dec. 1973, pp. 428–432.

[12] D. G. LUENBERGER, *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, Mass., 1973.

[13] E. POLAK, *Computational Methods in Optimization*, Academic Press, New York, 1971.

[14] B. T. POLYAK, *The convergence rate of the penalty function method*, Ž. Vyčisl. Mat. i Mat. Fiz., 11 (1971), pp. 3–11.

[15] M. J. D. POWELL, *A method for nonlinear constraints in minimization problems*, Optimization, R. Fletcher, ed., Academic Press, New York, pp. 283–298.

[16] R. T. ROCKAFELLAR, *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Math. Programming, 5 (1973), pp. 354–373.

[17] ———, *The multiplier method of Hestenes and Powell applied to convex programming*, J. Optimization Theory Appl., 12 (1973), pp. 555–562.

[18] ———, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, this Journal, 12 (1974), pp. 268–285.

[19] ———, *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.

[20] ———, *Penalty methods and augmented Lagrangians in nonlinear programming*, Proc. 5th IFIP Conf. on Optimization Techniques, Rome, 1973; Springer-Verlag, 1974.

[21] ———, *Solving a nonlinear programming problem by way of a dual problem*, Symposia Matematica, to appear.