Automatica, Vol. 12, pp. 133–145. Pergamon Press, 1976. Printed in Great Britain

Multiplier Methods: A Survey*†

DIMITRI P. BERTSEKAS‡

An analysis of the convergence properties of multiplier methods demonstrates their superiority over ordinary penalty methods for constrained minimization.

Summary—The purpose of this paper is to provide a survey of convergence and rate of convergence aspects of a class of recently proposed methods for constrained minimization—the, so-called, multiplier methods. The results discussed highlight the operational aspects of multiplier methods and demonstrate their significant advantages over ordinary penalty methods.

1. *Introduction*

DURING recent years, penalty function methods as described in [F1] have been widely accepted in practice as an effective class of methods for constrained optimization. Let us briefly describe penalty methods for the equality constrained problem

$$\left.\begin{array}{ll}\text{minimize} & f(x),\\ \text{subject to} & x\in X,\quad h_1(x)=h_2(x)=\ldots=h_m(x)=0,\end{array}\right\}\quad (1)$$

where $f, h_1, h_2, \ldots, h_m$ are real-valued functions on R^n, n-dimensional Euclidean space, and X is a given subset of R^n. The penalty function method consists of sequential minimizations of the form

$$\underset{x\in X}{\text{minimize}}\, f(x)+c_k\sum_{i=1}^{m}\phi[h_i(x)] \qquad (2)$$

for a scalar sequence $\{c_k\}$ such that $c_k \leq c_{k+1}$ for all k and $c_k \to \infty$. The scalar penalty function $\phi: R \to [0, +\infty]$ is such that

$$\phi(t)\geqslant 0,\ \forall\ t,\ \phi(t)=0 \quad \text{if and only if } t=0. \qquad (3)$$

The most common penalty function is the quadratic $(\phi(t) = \frac{1}{2}t^2)$; however, on some occasions it may be desirable to use other penalty functions. The sequential minimization process yields

$$\lim_{c_k\to\infty}\ \inf_{x\in X}\left\{f(x)+c_k\sum_{i=1}^{m}\phi[h_i(x)]\right\}. \qquad (4)$$

In view of property (3) of the function ϕ, the optimal value of problem (1) can be written as

$$\inf_{x\in X}\ \lim_{c_k\to\infty}\left\{f(x)+c_k\sum_{i=1}^{m}\phi[h_i(x)]\right\} \qquad (5)$$

and hence the success of the penalty method hinges on the equality of expressions (4) and (5), i.e. the validity of the interchange of 'lim' and 'inf'. One may show under relatively mild assumptions that this interchange is valid for wide classes of problems as explained for example in [F1, L1, P1, Z1]. Basically these assumptions require continuity of the functions f, h_i and ϕ, at least near a solution, and guarantee that a solution of problem (2) exists. Simultaneously with the generation of the minimizing points x_k of problem (2), penalty methods generate the sequence $\{\tilde{y}_k\}$ where $\tilde{y}_k = (c_k\,\phi'[h_1(x_k)], \ldots, c_k\,\phi'[h_m(x_k)])$ where ϕ' denotes the first derivative of ϕ—assumed to exist. The sequence $\{\tilde{y}_k\}$ under appropriate assumptions is known to converge to a Lagrange multiplier of the problem (see e.g. [F1, L1, Z1]).

Penalty methods are simple to implement, are applicable to a broad class of problems and take advantage of the very powerful unconstrained minimization methods that have been developed in recent years, for solving problem (2), in the case where $X = R^n$. These are the main reasons for their wide acceptance. On the negative side, penalty methods are hampered by slow convergence and numerical instabilities associated with ill-conditioning in problem (2) induced by large values of the penalty parameter c_k.

Another important class of methods for solving problem (1) is based on sequential minimizations of the Lagrangian function defined for every $x \in R^n$ and $y = (y^1, \ldots, y^m) \in R^m$ by

$$L(x,y)=f(x)+\sum_{i=1}^{m}y^i\,h_i(x). \qquad (6)$$

In the simplest and most widely known such method, as discussed, for example, in [L1], one minimizes $L(x, y_k)$ over $x \in X$ for a sequence of multiplier vectors $\{y_k\}$. This sequence is generated by iterations of the form

$$y_{k+1}{}^i = y_k{}^i+\alpha_k\,h_i(x_k),\quad i=1,\ldots,m, \qquad (7)$$

where x_k is a minimizing point of $L(x, y_k)$ over $x \in X$, and α_k is a stepsize (scalar) parameter. The iteration above may be viewed as a steepest ascent iteration aimed at finding an optimal solution of an associated dual problem. For this reason the corresponding algorithm is called a primal-dual method. Methods such as the one described above are known to have serious disadvantages. First, problem (1) must have a locally convex structure in order for the dual problem to be well defined and iteration (7) to be meaningful as discussed in [L1]. Second, it is usually necessary to minimize the Lagrangian function (6) a large number of times since the ascent iteration (7) converges only moderately fast. Thus primal-dual methods of the type described above have found application only in the limited class of problems where minimization of the Lagrangian (6) can be carried out very efficiently due to special structure as shown in [L1, L2].

* Received 13 May 1975; revised 6 October 1975. The original version of this paper was presented at the 6th IFAC Congress which was held in Boston/Cambridge, MA, U.S.A., during August 1975. The published proceedings of this IFAC meeting may be ordered from the ISA (Instrument Society of America), 400 Stanwix Street, Pittsburgh, PA 15222, U.S.A. It was recommended for publication in revised form by associate editor S. J. Kahne.

† This work was supported in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy and U.S. Air Force) under Contract DAAB-07-72-C-0259, by the U.S. Air Force under Grant AFOSR-73-2570 and by the National Science Foundation under Grant ENG-74-19332.

‡ Department of Electrical Engineering and Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801, U.S.A.

In the last few years, a number of researchers have proposed a new class of methods, called methods of multipliers, in which the penalty idea is merged with the primal-dual philosophy. In these methods, the penalty term is added not to the objective function f but rather to the Lagrangian function L of (6) thus forming the *Augmented Lagrangian* function

$$L_c(x, y) = f(x) + \sum_{i=1}^{m} y^i h_i(x) + c \sum_{i=1}^{m} \phi[h_i(x)]. \quad (8)$$

A sequence of minimizations of the form

$$\underset{x \in X}{\text{minimize}}\, L_{c_k}(x, y_k) = f(x) + \sum_{i=1}^{m} y_k{}^i h_i(x) + c_k \sum_{i=1}^{m} \phi[h_i(x)] \quad (9)$$

is performed where $\{c_k\}$ is a sequence of positive penalty parameters. The multiplier sequence $\{y_k\}$ is generated according to the iteration

$$y_{k+1}{}^i = y_k{}^i + c_k \phi'[h_i(x_k)], \quad i = 1, \ldots, m, \quad (10)$$

where ϕ' is the first derivative of ϕ, assumed to exist, and x_k is a point minimizing over $x \in X$ the Augmented Lagrangian $L_{c_k}(x, y_k)$. The initial multiplier vector y_0 is selected *a priori* and the sequence $\{c_k\}$ may be either preselected or generated during the computation according to some scheme. Initially the method was proposed for a quadratic penalty ($\phi(t) = \frac{1}{2}t^2$) in which case iteration (10) is written as

$$y_{k+1}{}^i = y_k{}^i + c_k h_i(x_k), \quad i = 1, \ldots, m$$

and is a special case of iteration (7) considered earlier.

Now if we select a penalty parameter sequence $\{c_k\}$ with $c_k \to \infty$ and the generated sequence $\{y_k\}$ turns out to be bounded, then the method is guaranteed to yield in the limit the optimal value of problem (1) provided sufficient assumptions are satisfied which guarantee the validity of interchange of 'lim' and 'inf' in the expression

$$\lim_{c_k \to \infty} \inf_{x \in X} \left\{ f(x) + \sum_{i=1}^{m} y_k{}^i h_i(x) + c_k \sum_{i=1}^{m} \phi[h_i(x)] \right\}$$

similarly as for the penalty method considered earlier. The important aspect of multiplier methods, however, is that *convergence may occur without the need to increase c_k to infinity*, i.e. convergence may be induced not merely by ever increasing values of the penalty parameter but also by the multiplier iteration (10). Thus the ill-conditioning associated with penalty methods can be avoided. In addition, *iteration* (10) *converges fast to a Lagrange multiplier vector of problem* (1), under relatively mild assumptions, much faster than in primal-dual methods considered earlier. Furthermore, there is no need for problem (1) to have a locally convex structure in order for the method to be applicable.

By moderating the disadvantages of both penalty and primal-dual methods, multiplier algorithms have emerged as a most attractive class of methods for constrained optimization. Since their original proposal in 1968, a considerable amount of research has been directed towards their analysis. The aim of this paper is to provide a survey of the convergence and rate of convergence aspects of multiplier methods with some emphasis placed on demonstrating the important role of the inherent structure in problem (1) as well as the form of the penalty function employed. There is an important and aesthetically pleasing duality theory associated with multiplier methods which, in contrast with past duality formulations, is applicable to convex as well as non-convex programming problems. For an excellent account of these developments the reader may consult the survey paper of Rockafellar [R6].

The paper is organized as follows. In Sections 2 and 3 we describe the convergence and rate of convergence properties of multiplier methods with quadratic-like penalty function under second-order sufficiency assumptions on problem (1). In Section 2 we provide an interpretation of multiplier methods as generalized penalty methods while in Section 3 we view the multiplier iteration (10) as a gradient iteration for solving a certain dual problem. The results and the simple computational example provided illustrate the significant advantages of multiplier methods over penalty methods in terms of reliability and speed of convergence. In Section 4 we describe the form of multiplier methods as applied to problems with one-sided and two-sided inequality constraints. In Section 5 we provide convergence and rate of convergence results for the case of a convex programming problem. We also demonstrate how the choice of penalty function can affect significantly the performance of multiplier methods. In Section 6 we briefly describe a number of variations of multiplier methods and point out connections with other related methods. Finally, in Section 7 we survey the literature relating to multiplier methods for infinite dimensional problems and particularly optimal control problems.

2. *Multiplier methods from a penalty viewpoint*

Given problem (1), consider the *augmented Lagrangian* function L_c: $R^n \times R^m \to (-\infty, +\infty]$

$$L_c(x, y) = f(x) + y' h(x) + c \sum_{i=1}^{m} \phi[h_i(x)], \quad (11)$$

where ϕ: $R \to [0, +\infty]$ is a penalty function satisfying (3) and $c \geqslant 0$ is a penalty parameter. In the above equation $h(x)$ denotes the m-vector $\{h_1(x), \ldots, h_m(x)\}$ viewed as a column vector and the prime denotes transposition. We shall utilize the following two assumptions related to problem (1) and the penalty function ϕ.

Assumption (S): There exists a local minimizing point $\bar{x}$ of problem (1) which is an interior point of X and satisfies the standard second-order sufficiency conditions for an isolated local minimum ([L1], p. 226), i.e. f, h_i are twice continuously differentiable in a neighborhood of $\bar{x}$, the gradients $\nabla h_i(\bar{x})$, $i = 1, \ldots, m$, are linearly independent and there exists a Lagrange multiplier vector $\bar{y} \in R^m$ such that $\nabla L_0(\bar{x}, \bar{y}) = 0$ and $z'\nabla^2 L_0(\bar{x}, \bar{y}) z > 0$ for all $z \in R^n$ with $z \neq 0$, $\nabla h_i(\bar{x})'z = 0$, $i = 1, \ldots, m$, where ∇L_0, $\nabla^2 L_0$ denote the gradient and Hessian matrix with respect to x of $L_0(x, y) = f(x) + y' h(x)$.

Assumption (Q): The penalty function ϕ: $R \to [0, +\infty]$ is twice continuously differentiable in an open interval containing zero and $\phi''(0) = 1$, where ϕ'' denotes the second derivative of ϕ.

The penalty function considered in original studies of multiplier methods was the quadratic $\phi(t) = \frac{1}{2}t^2$ which of course satisfies (Q). Since functions satisfying (Q) behave similarly as $\phi(t) = \frac{1}{2}t^2$ we refer to such penalty functions as *essentially quadratic*. Notice that property (3) of the penalty function implies that $\phi'(0) = 0$ when ϕ is differentiable in a neighborhood of zero. Notice also that, in view of the presence of the penalty parameter c in (11), there is no loss of generality in assuming $\phi''(0) = 1$ rather than $\phi''(0) > 0$.

The following proposition yields estimates related to minimizing points of $L_c(x, y)$ of (11) and forms the basis for establishing the validity and convergence of multiplier methods as well as ordinary penalty methods.

Proposition 1: Let (S), (Q) hold and assume that the Hessian matrices $\nabla^2 f(x)$, $\nabla^2 h_i(x)$, $i = 1, \ldots, m$ and the second derivative $\phi''(t)$ are Lipschitz continuous in neighborhoods of $\bar{x}$ and zero respectively. Then for any given bounded set $Y \subset R^m$ there exists a scalar $c^* \geqslant 0$ such that for every $c > c^*$ and every $y \in Y$ the function $L_c(x, y)$ has a unique minimizing point $x(y, c)$ within some open ball centered at $\bar{x}$. Furthermore, for some scalar $M > 0$ we have

$$\| x(y, c) - \bar{x} \| \leqslant \frac{M \| y - \bar{y} \|}{c}, \quad \forall c > c^*, \quad y \in Y, \quad (12)$$

$$\| \tilde{y}(y, c) - \bar{y} \| \leqslant \frac{M \| y - \bar{y} \|}{c}, \quad \forall c > c^*, \quad y \in Y, \quad (13)$$

where the vector $\bar{y}(y, c) \in R^m$ has coordinates given by

$$\bar{y}(y, c)^i = y^i + c\phi'[h_i[x(y, c)]], \quad i = 1, \ldots, m \qquad (14)$$

ϕ' is the first derivative of ϕ, and $\|\cdot\|$ is the usual Euclidean norm.

The proposition above can yield both a convergence and a rate of convergence result for a method of multipliers which updates the current multiplier vector y_k by means of the iteration

$$y_{k+1}{}^i = y_k{}^i + c_k\, \phi'[h_i[x(y_k, c_k)]], \quad i = 1, \ldots, m$$

and constitutes the strongest convergence result available for nonconvex problems. It shows that if Y contains $\bar{y}$ in its interior, the generated sequence $\{y_k\}$ remains in Y (this can be enforced if necessary by, for example, leaving y_k unchanged if $y_{k+1} \notin Y$), the penalty parameter c_k is sufficiently large after a certain index, and the minimization of $L_{c_k}(x, y_k)$ yields the local minimum $x(y_k, c_k)$ which is closest to $\bar{y}$, then we obtain $x(y_k, c_k) \to \bar{x}$, $y_k \to \bar{y}$. We make this result precise in the following proposition:

Proposition 2: Under the assumptions of Proposition 1, let Y be an arbitrary sphere centered at $\bar{y}$. Consider also a non-decreasing positive penalty parameter sequence $\{c_k\}$ such that for some integer $\bar{k} \geqslant 0$ we have $c_{\bar{k}} \geqslant \max\{M, c^*\}$ where M and c^* correspond to Y as in Proposition 1. Let $\{y_k\}$ be a sequence such that $y_{\bar{k}} \in Y$ and generated for all $k \geqslant \bar{k}$ by

$$y_{k+1}{}^i = y_k{}^i + c_k\, \phi'[h_i(x_k)], \quad i = 1, \ldots, m \qquad (15)$$

where x_k is the local minimizing point of $L_{c_k}(\cdot, y_k)$ closest to $\bar{x}$ in the sense of Proposition 1. Then $x_k \to \bar{x}$, $y_k \to \bar{y}$. Furthermore, if $c_k \to \bar{c} < \infty$ and $y_k \neq \bar{y}$ for all k we have

$$\limsup_{k\to\infty} \frac{\| y_{k+1} - \bar{y} \|}{\| y_k - \bar{y} \|} \leqslant \frac{M}{\bar{c}} \quad \text{(linear convergence)} \qquad (16)$$

while if $c_k \to \infty$ and $y_k \neq \bar{y}$ for all k

$$\lim_{k\to\infty} \frac{\| y_{k+1} - \bar{y} \|}{\| y_k - \bar{y} \|} = 0 \quad \text{(superlinear convergence).} \qquad (17)$$

Interpretation of results—computational aspects. Since Proposition 1 applies to both multiplier methods and pure penalty methods ($y_k = 0$ for all k) it provides a natural vehicle for comparison of these methods. From estimates (12) and (13) it follows that in the penalty method ($y_k = 0$) it is ordinarily necessary to increase the penalty parameter c_k to infinity. It follows also from Proposition 2 that *it is not necessary to increase c_k to infinity in order for the method of multipliers to converge.* Thus the ill-conditioning effects associated with large penalty parameters can be eliminated or at least moderated in multiplier methods. The second important advantage of multiplier methods may be inferred by comparing the convergence rate of the two methods as given by the estimates (12) and (13). While in multiplier methods the convergence rate is linear or superlinear (c.f. (16) or (17)), the convergence rate of penalty methods is much worse. This advantage in speed of convergence has been verified in many computational studies where a consistent reduction in computation time ranging roughly from 80 to 30% has been reported when the multiplier iteration (15) was employed over the case where y_k was kept constant. For illustration purposes we provide the following example which is, of course, trivial in terms of computational complexity, but nonetheless is representative of the computational savings resulting from employment of the multiplier iteration (15).

Computational example: Consider the two-dimensional problem

$$\min_{x^1+x^2=1} \tfrac{1}{2}[(x^1)^2 + \tfrac{1}{3}(x^2)^2].$$

The Augmented Lagrangian for a quadratic penalty function is

$$\tfrac{1}{2}[(x^1)^2 + \tfrac{1}{3}(x^2)^2] + y_k(x^1 + x^2 - 1) + \tfrac{1}{2}c_k\,(x^1 + x^2 - 1)^2.$$

Minimization of the Augmented Lagrangian yields

$$x_k{}^1 = \frac{c_k - y_k}{1 + 4c_k}, \quad x_k{}^2 = \frac{3(c_k - y_k)}{1 + 4c_k}.$$

The optimal solution is $\bar{x}^1 = 0{\cdot}25$, $\bar{x}^2 = 0{\cdot}75$ and the corresponding Lagrange multiplier is $\bar{y} = -0{\cdot}25$. In Table 1 we show the results of the computation for the penalty method where $y_k = 0$ for all k and for the multiplier method where

$$y_{k+1} = y_k + c_k(x_k{}^1 + x_k{}^2 - 1)$$

and $y_0 = 0$. Notice that the multiplier method requires a smaller number of minimizations to obtain the solution within four significant digits of accuracy than the penalty method. The number of minimizations required for both methods decreases when the penalty parameter is increased at a faster rate. However, the effects of ill-conditioning are felt more under these circumstances when the unconstrained minimization is carried out numerically.

It is perhaps worthwhile to note the fact that the global convergence property of the method of multipliers concluded in Proposition 2 is contingent upon the generation

TABLE 1

	$c_k = 0{\cdot}1 \times 2^k$				$c_k = 0{\cdot}1 \times 4^k$				$c_k = 0{\cdot}1 \times 8^k$			
	Penalty		Multiplier		Penalty		Multiplier		Penalty		Multiplier	
k	$x_k{}^1$	$x_k{}^2$	$x_k{}^1$	$x_k{}^2$	$x_k{}^1$	$x_k{}^2$	$x_k{}^1$	$x_k{}^2$	$x_k{}^1$	$x_k{}^2$	$x_k{}^1$	$x_k{}^2$
0	0·0714	0·2142	0·0714	0·2142	0·0714	0·2142	0·0714	0·2142	0·0714	0·2142	0·0714	0·2142
1	0·1111	0·3333	0·1507	0·4523	0·1538	0·4615	0·1813	0·5439	0·1904	0·5714	0·2074	0·6224
2	0·1538	0·4615	0·2118	0·6355	0·2162	0·6486	0·2407	0·7221	0·2406	0·7218	0·2484	0·7452
3	0·1904	0·5714	0·2409	0·7227	0·2406	0·7218	0·2496	0·7489	0·2487	0·7463	0·2499	0·2499
4	0·2162	0·6486	0·2487	0·7463	0·2475	0·7427	0·2499	0·7499	0·2498	0·7495		
5	0·2318	0·6956	0·2499	0·7497	0·2493	0·7481			0·2499	0·7499		
6	0·2406	0·7218	0·2499	0·7499	0·2498	0·7495						
7	0·2452	0·7356			0·2499	0·7498						
8	0·2475	0·7427			0·2499	0·7499						
9	0·2487	0·7463										
10	0·2493	0·7481										
11	0·2496	0·7490										
12	0·2498	0·7495										
13	0·2499	0·2497										
14	0·2499	0·2498										
15	0·2499	0·2499										

of the points $x(y_k, c_k)$, by the unconstrained minimization method employed, at least for all k after a certain index. These points are, by Proposition 1, well defined as local minimizing points of $L_{c_k}(x, y_k)$ which are closest to $\bar{x}$. Naturally $L_{c_k}(x, y_k)$ may have other local minimizing points to which the unconstrained minimization method may be attracted, and unless the unconstrained minimization method stays after some index in the neighborhood of the same minimizing point of $L_{c_k}(x, y_k)$ 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 $(k+1)$th minimization is helpful in producing sequences $\{x_k\}$ which are close to one and the same local minimizing point of $L_{c_k}(x, y_k)$.

Another point concerns the fact that the estimates (12) and (13) of Proposition 1 are valid only for c greater than the threshold value c^* which depends on the set Y as well as on the data of the problem. In convex programming problems one may take $c^* = 0$, as will be seen in Section 5, but in general the constant c^* is unavailable and for this reason it is impossible to know *a priori* the range of values of penalty parameter c for which the estimates (12) and (13) are in effect and induce fast convergence. This situation, however, is not really troublesome if one adopts a penalty parameter adjustment scheme whereby c_k is monotonically increased with each minimization; for example, by multiplication by a constant factor $\beta > 1$, i.e. $c_{k+1} = \beta c_k$ for all k. Then, since $c_k \to \infty$, eventually (12) and (13) will become effective. It is to be noted that large values of c induce an ill-conditioning effect in the unconstrained minimization procedure which tends to make the problem $\min_x L_{c_k}(x, y_k)$ harder to solve. On the other hand, (12) and (13) indicate that for large values of c the convergence of $\{y_k\}$ to $\bar{y}$ is faster. On balance a procedure of continuously increasing c usually works well and the author strongly recommends it provided:

(a) the penalty parameter is not increased too fast, β is not much larger than one, in which case too much ill-conditioning is forced upon the unconstrained minimization routine too early;

(b) the last point $x(y_k, c_k)$ of the kth minimization is used as the starting point in the $(k+1)$st minimization—this policy tends to reduce the effect of ill-conditioning since $x(y_k, c_k)$ and $x(y_{k+1}, c_{k+1})$ tend to be close to each other.

Another possible penalty parameter adjustment scheme, recommended by Powell [P3], is to increase c_k by multiplying it by a certain factor $\beta > 1$ only if the constraint violation as measured by $\| h[x(y_k, c_k)] \|$ is not decreased by a certain factor over the previous minimization, i.e. $c_{k+1} = \beta c_k$ if $\| h[x(y_k, c_k)] \| > \gamma \| h[x(y_{k-1}, c_{k-1})] \|$ and $c_{k+1} = c_k$ otherwise where $\beta > 1$, $\gamma < 1$ are some prespecified scalars. This is also a very satisfactory scheme. One may prove under our assumptions that for some constant M' and all c_k, c_{k-1} which are sufficiently large and satisfy $c_k \geq c_{k-1}$ there holds

$$\| h[x(y_k, c_k)] \| \leq \frac{M'}{c_k} \| h[x(y_{k-1}, c_{k-1})] \| .$$

As a result the scheme described above will generate a penalty parameter sequence that will be constant after a certain index, i.e. $c_{k+1} = c_k$ for all k sufficiently large, while it will achieve convergence by virtue of enforcement of asymptotic feasibility of the constraints, i.e.

$$\lim_{k\to\infty} \| h[x(y_k, c_k)] \| = 0.$$

Still another possibility along the same lines is to use a different penalty parameter for each constraint equation $h_i(x) = 0$, and to increase by a certain factor only the penalty parameters which correspond to those constraint equations for which the constraint violation as measured by $| h_i[x(y_k, c_k)] |$ is not decreased by a certain factor over the previous minimization. It is to be noted that the case where a separate penalty parameter is used for each constraint corresponds to merely scaling these constraints. It is easy to prove a simple modification of Proposition 1 to cover this case.

Geometrical interpretation. A transparent geometrical interpretation of the method of multipliers which demonstrates the basic conclusions of Propositions 1 and 2 may be obtained by considering the primal functional (or perturbation function) p of problem (1) defined for values of u in a neighborhood of the origin by

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

where the minimization is understood to be local in a neighborhood of $\bar{x}$. Clearly $p(0)$ is equal to the value of problem (1) corresponding to $\bar{x}$, i.e. $p(0) = f(\bar{x})$.

Furthermore, under (S) one may show that p is twice continuously differentiable in a neighborhood of the origin and

$$\nabla p(0) = -\bar{y}.$$

Now one may write

$$\begin{aligned}\min_x L_c(x, y) &= \min_u \min_{h_i(x)=u^i} \left\{ f(x) + \sum_{i=1}^m y^i h_i(x) + c \sum_{i=1}^m \phi[h_i(x)] \right\} \\ &= \min_u \left\{ p(u) + \sum_{i=1}^m y^i u^i + c \sum_{i=1}^m \phi(u^i) \right\},\end{aligned}$$

where the minimization above is understood to be local in a neighborhood of $u = 0$. This minimization may be interpreted as shown in Fig. 1. It can be seen from the figure that, if c is sufficiently large so that $p(u) + c \sum_{i=1}^m \phi(u^i)$ is convex in a neighborhood of zero, the value $\min_x L_c(x, y)$ is close to $p(0)$ for values of y close to $\bar{y}$ and large values of c, as indicated by (12) and (13). The multiplier iteration (15) is shown in Fig. 2. A closer examination of this figure

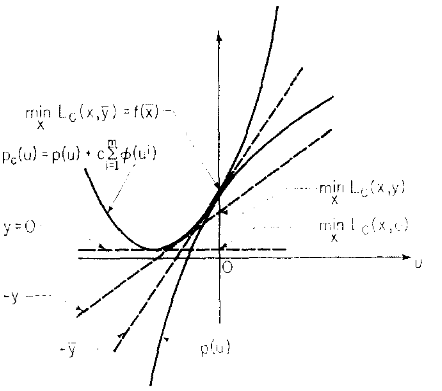

FIG. 1. Geometric interpretation of minimization of the Augmented Lagrangian.

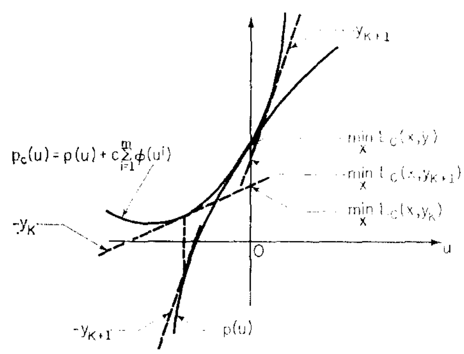

FIG. 2. Geometric interpretation of the multiplier iteration.

suggests that this iteration converges to $\bar{y}$ even if c_k is constant but sufficiently large after a certain index and the convergence is faster for large values of c_k as in (13). It is interesting also to note that the two figures demonstrate the crucial ingredients for the success of iterations of the form (15). It is necessary among other things that $p(u)$ be either differentiable or convex near zero and, furthermore, $p(u)+c\sum_{i=1}^{m}\phi(u_i)$ must be convex with sufficiently large 'curvature' near zero. These conditions are satisfied under (S) and (Q) but they are also satisfied under different sets of assumptions some of which will be presented in Section 5.

Multiplier methods with partial elimination of constraints. In the multiplier methods described above each one of the equality constraints of the problem has been eliminated by means of a generalized penalty function. In some problems it is convenient to eliminate only part of the constraints while retaining explicitly the remaining constraints during sequential minimization of the Augmented Lagrangian. Thus a possible multiplier algorithm for the problem

$$\text{minimize} \quad f(x),$$
$$\text{subject to} \quad h_i(x) = 0, \quad l_j(x) = 0, \quad i = 1, \ldots, m,$$
$$j = 1, \ldots, r,$$

is based on sequential minimizations of the form

$$\text{minimize} \quad f(x) + \sum_{i=1}^{m} \{y_k{}^i\, h_i(x) + c_k\, \phi[h_i(x)]\},$$
$$\text{subject to} \quad l_j(x) = 0, \quad j = 1, \ldots, r.$$

The multiplier iteration is given by

$$y_{k+1}{}^k = y_k{}^i + c_k\, \phi'[h_i(x_k)], \quad i = 1, \ldots, m,$$

where x_k solves the minimization problem above.

The convergence properties of such algorithms are very similar to those for multiplier methods with full elimination of constraints, and one may prove direct analogs of Propositions 1 and 2 for these algorithms.

Notes and references. The method of multipliers (12) with $\phi(t) = \frac{1}{2}t^2$ was originally proposed by Hestenes [H1] and independently by Powell [P3] in a different but equivalent form. It was also proposed a year later by Haarhoff and Buys [H2].* Of these authors only Powell discussed the convergence properties of the method. Powell's paper predicted the superiority of the method over the ordinary penalty method but while it provided some analysis of asymptotic behavior it actually stopped short of proving local convergence of the method for a bounded penalty parameter. Such local convergence results, based on the assumption that the starting multiplier vector y_0 is sufficiently close to $\bar{y}$, were given later and apparently independently by Buys [B7] and Rupp [R8], (see also Wierzbicki [W1]). These authors showed also that the rate of convergence is linear but did not provide sharp estimates of the convergence ratio. Such estimates will be given in the next section. Propositions 1 and 2 have been proved for the case of a quadratic penalty function ($\phi(t) = \frac{1}{2}t^2$) by the author [B1–3] and independently by Polyak and Tret'yakov [P2]. The extension provided here admits essentially the same proof as for a quadratic penalty. An analysis of multiplier methods with partial elimination of constraints is provided in [B8, 9].

* Polyak and Tret'yakov [P2] give independent credit for the proposal of the multiplier method to Syrov and Churkveidze (see [P2]). It is to be noted that there has been considerable interest and significant recent work on multiplier methods in the Soviet Union. Professor Rockafellar kindly pointed out some related papers of Polyak, Gol'shtein and Tret'yakov [G1, P7, T2]. There is no English translation of these papers and the author is not familiar with their precise contents.

3. *Multiplier methods from a primal-dual viewpoint*

Gradient interpretation of the multiplier iteration. Consider problem (1) under assumption (S) and a penalty function ϕ satisfying assumption (Q). It is clear that in a neighborhood of $\bar{x}$ problem (1) is equivalent for every $c \geq 0$ to the problem

$$\min_{h(x)=0} \left\{ f(x) + c \sum_{i=1}^{m} \phi[h_i(x)] \right\} \tag{18}$$

in the sense that both problems have $\bar{x}$ as local minimum and $\bar{y}$ as associated Lagrange multiplier vector. The Hessian with respect to x of the Augmented Lagrangian $L_c(x, y)$ of (11) evaluated at $\bar{x}$, $\bar{y}$ is given by

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

One may show easily using (S) that for $c \geq c^*$, where c^* is a suitable non-negative scalar, the matrix $\nabla^2 L_c(\bar{x}, \bar{y})$ is positive definite, a fact pointed out and utilized in an algorithmic context by Arrow and Solow in 1958 [A2]. As a result, for $c \geq c^*$, problem (18) has a locally convex structure as defined in [L1] and one may define for $c \geq c^*$ the *dual functional*

$$q_c(y) = \min_x L_c(x, y). \tag{19}$$

In the above equation the dual functional is defined in a neighborhood of $\bar{y}$ and the minimization is understood to be local in a neighborhood of $\bar{x}$. By using the implicit function theorem and our assumptions one may show that such neighborhoods exist for each $c \geq c^*$. In fact one may show that if $x(y, c)$ is the locally minimizing point in (19) then q_c is a twice continuously differentiable concave function in its domain of definition with gradient given by

$$\nabla q_c(y) = h[x(y, c)] \tag{20}$$

and Hessian matrix evaluated at y given by

$$\nabla^2 q_c(y) = -\nabla h[x(y, c)]\, [\nabla^2 L_c[x(y, c), y]]^{-1} \times \nabla h[x(y, c)]', \tag{21}$$

where $\nabla h(x)$ is the $m \times n$ matrix with rows $\nabla h_i(x)$, $i = 1, \ldots, m$. Furthermore, $q_c(y)$ is maximized at $\bar{y}$.

Now in view of (20) the iteration of the method of multipliers

$$y_{k+1}{}^i = y_k{}^i + c\phi'[h_i[x(y_k, c)]], \quad i = 1, \ldots, m \tag{22}$$

may be written as

$$y_{k+1} = y_k + c\Phi[x(y_k, c)]\, \nabla q_c(y_k), \tag{23}$$

where Φ is the diagonal $m \times m$ matrix having as its ith diagonal element the scalar $\int_0^1 \phi''[\lambda h_i[x(y_k, c)]]\, d\lambda$. This expression is obtained by Taylor's theorem using the fact $\phi'(0) = 0$.

From (23) one can see that the multiplier iteration (22) may be viewed as an iteration of the ascent type for finding the maximizing point $\bar{y}$ of the dual functional q_c. Since $\Phi[x(y, c)]$ tends to the identity matrix as $x(y, c) \to \bar{x}$ (in view of $\phi''(0) = 1$), iteration (23) becomes a fixed stepsize steepest ascent iteration in the limit as $y_k \to \bar{y}$. In fact if $\phi(t) = \frac{1}{2}t^2$ then (23) is equivalent to the steepest ascent iteration.

A tight bound on the convergence rate of multiplier methods. Based on the interpretation of the multiplier iteration as a gradient iteration one may obtain a sharp rate of convergence result for iteration (23) by using a simple variation of a known result [P4] for the steepest ascent method. This result involves, however, the eigenvalues of $\nabla^2 q_c(\bar{y})$ of (21) which strongly depend on c. A modification of this result which is more amenable to proper interpretation is provided by the following proposition.

Proposition 3: Let (S), (Q) hold and assume that y_0 is sufficiently close to $\bar{y}$. Then the sequence $\{y_k\}$ generated by

(22) converges to $\bar{y}$. Furthermore, assuming $y_k \neq \bar{y}$ for all k

$$\limsup_{k\to\infty} \frac{\| y_{k+1}-\bar{y} \|}{\| y_k-\bar{y} \|} \leqslant \max_{i=1,\ldots,m} \left| \frac{1}{1-ce_i(D)} \right|, \quad (24)$$

where $e_i(D)$ denotes the ith eigenvalue of the matrix

$$D = -\nabla h(\bar{x}) [\nabla^2 L_0(\bar{x}, \bar{y})]^{-1} \nabla h(\bar{x})'$$

and it is assumed that the inverse of $\nabla^2 L_0$ exists. In addition the bound (24) is sharp in the sense that, if ϕ and f are quadratic functions and $h_1, \ldots, h_m$ are affine functions, then there exist starting points y_0 for which (24) is satisfied with equality.

Dependence of convergence rate upon assumptions (Q) and (S). The proposition above confirms the fact that the convergence rate of the multiplier iteration (22) is linear with convergence ratio essentially inversely proportional to the penalty parameter c. This fact, however, is *strongly dependent upon the assumptions (Q) and (S)*. If either of the assumptions is relaxed the convergence rate may become sublinear or superlinear as the following examples show.

Example 1: Consider the scalar problem

$$\min \{\tfrac{1}{2} x^2 \mid x = 0\}$$

with optimal solution $\bar{x} = 0$ and Lagrange multiplier $\bar{y} = 0$. For $y \leqslant 0$, $\phi(t) = \frac{1}{3} | t |^3$, $c = 1$ the minimizing point of $L_1(x, y)$ is $x(y, 1) = [-1+\sqrt{(1-4y)}]/2$. For a starting point $y_0 < 0$ the iteration (22) yields

$$y_{k+1} = [1-\sqrt{(1-4y_k)}]/2$$

and $\lim_{k\to\infty}(y_{k+1}/y_k) = 1$, i.e. we have sublinear convergence rate.

Example 2: Consider the same problem as in Example 1. For $y \leqslant 0$, $\phi(t) = \frac{2}{3} | t |^{\frac{3}{2}}$, $c = 1$ we obtain $x(y, 1) = [-1+\sqrt{(1-4y)}]^2/4$. For a starting point $y_0 < 0$ the iteration (18) yields

$$y_{k+1} = y_k + [-1+\sqrt{(1-4y)}]/2$$

One may show that

$$\lim_{k\to\infty} \frac{| y_{k+1} |}{y_k^2} = 1$$

and hence we have superlinear convergence (order 2).

Example 3: Consider the problem $\min (\frac{1}{3} | x |^3 | x = 0)$. Again here $\bar{x} = \bar{y} = 0$ but (S) is not satisfied. For $y < 0$, $\phi(t) = \frac{1}{2} t^2$, $c = 1$ the minimizing point of $L_1(x, y)$ is $x(y, 1) = [-1+\sqrt{(1-4y)}]/2$. For a starting point $y_0 < 0$ the iteration (18) yields

$$y_{k+1} = y_k + [-1+\sqrt{(1-4y_k)}]/2.$$

Again

$$\lim_{k\to\infty} \frac{| y_{k+1} |}{y_k^2} = 1$$

and we have superlinear convergence.

The convergence behavior exhibited in the above examples may be explained by close examination of the geometrical constructions of Figs. 1 and 2. It may be seen from these figures that the convergence rate is influenced substantially by the rates of change of the derivatives of the primal functional $p(u)$ and the 'penalized' primal functional $p(u) + c \sum_{i=1}^m \phi(u_i)$ near $u = 0$. *The convergence is faster as the rate of change of $\nabla p(u)$ is small and the rate of change of $c\, \phi'(u_i)$ is large near $u = 0$.*

In Example 1 the rate of change of $\phi'(u_i)$ is small near $u_i = 0$ and convergence is slow, while in Example 2 it is large and convergence is fast. In Example 3 the rate of change of $\nabla p(u)$ is small near $u = 0$ thus explaining the fast convergence. An extreme case of small rate of change of $\nabla p(u)$ is when $p(u)$ is a linear or affine function in which case Fig. 2 shows that the method of multipliers converges in a single iteration. A general convergence rate result which establishes the behavior described above will be given in Section 5.

The following example shows also that in the absence of (S) iteration (22) may not lead to convergence for any $c > 0$ when ϕ is essentially quadratic.

Example 4: Consider the problem $\min \{- | x |^\rho \mid x = 0\}$ where $1 < \rho < 2$. Then for any $c > 0$ one can find a neighborhood of $\bar{x} = 0$ within which the Augmented Lagrangian $L_c(x, y)$ does not have a local minimum for any value of $y \in R$ when ϕ is essentially quadratic. This situation can be corrected by using $\phi(t) = | t |^{\rho'}$ or $\phi(t) = | t |^{\rho'} + \frac{1}{2} t^2$ where ρ' satisfies $1 < \rho' < \rho$.

Notes and references. The primal dual framework adopted here for viewing the multiplier iteration (22), with $\phi(t) = \frac{1}{2} t^2$, was suggested by Luenberger [L1] and by Buys [B11] in his well-written disertation which is devoted to multiplier methods. Proposition 3 was proved by the author [B4] for the case $\phi(t) = \frac{1}{2} t^2$ and in a more general setting where the parameter c may change from one iteration to the next. For related duality frameworks for viewing the method of multipliers see [R1, R2, R4, R6, B1–3, P2, P5, A1, M1, W2]. The duality frameworks of [R4, B1–3, P2] are global in nature in the sense that the dual functional is everywhere defined. The construction in [R4] is carried out under very general assumptions. In [B1–3] and [P2] the assumptions are more restrictive; however, the dual functional constructed has strong first and second differentiability properties.

4. *Treatment of inequality constraints*

The original papers on the method of multipliers [H1, P3, H2] do not deal with or express inability to handle inequality constraints. It turns out, however, that one may handle inequality constraints trivially by converting them to equality constraints by using additional variables but without loss of computational efficiency due to increased dimensionality.

One-sided inequality constraints. Consider the following problem involving one-sided inequality constraints

$$\min_{\substack{g_j(x) \leqslant 0 \\ j=1,\ldots,r}} f(x). \quad (25)$$

The problem above is equivalent to the equality constrained problem

$$\min_{\substack{g_j(x)+z_j^2=0 \\ j=1,\ldots,r}} f(x), \quad (26)$$

where $z_1, \ldots, z_r$ are additional variables. Thus one may use a method of multipliers to solve problem (26) in place of (25) and if (S) holds for problem (26) the results of the past two sections are applicable. One may prove that if $\{\bar{x}, \bar{y}\}$ are an optimal solution–Lagrange multiplier pair satisfying the standard second-order sufficiency conditions for optimality [L1] (including strict complementarity, i.e. $\bar{y}^j > 0$ if and only if $g_j(\bar{x}) = 0$) for problem (25) then (S) holds for problem (26) in connection with the pair $\{(\bar{x}, \bar{z}_1, \ldots, \bar{z}_r), \bar{y}\}$ where $\bar{z}_j = | g_j(\bar{x}) |^{\frac{1}{2}}, j = 1, \ldots, r$.

Turning to the multiplier iteration for problem (26) the Augmented Lagrangian is

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

Let us make the following assumption on ϕ.

Assumption (G): ϕ is real valued, strictly convex and continuously differentiable on R. Furthermore, $\phi(0) = 0$, $\phi'(0) = 0$, $\lim_{t\to-\infty} \phi'(t) = -\infty$, $\lim_{t\to\infty} \phi'(t) = +\infty$.

Then minimization of $L_c(x, z, y)$ can be carried out first with respect to $z_1, \ldots, z_r$ yielding after some calculation

$$L_c(x, y) = \min_z L_c(x, z, y) = f(x) + \sum_{j=1}^r \bar{p}[g_j(x), y^j, c], \quad (27)$$

where $\bar{p}$ is defined by

$$\bar{p}[g_j(x), y^j, c] = \begin{cases} y^j g_j(x) + c\phi[g_j(x)] & \text{if } y^j + c\,\phi'[g_j(x)] \geq 0, \\ \min\limits_{\tau \in R} \{y^j \tau + c\phi(\tau)\} & \text{if } y^j + c\phi'[g_j(x)] < 0. \end{cases}$$

The form of the function above is shown in Fig. 3. An equivalent form of (27) which is somewhat more convenient for analytical purposes is given by

$$L_c(x, y) = f(x) + \frac{1}{c} \sum_{j=1}^{r} p[cg_j(x), y^j], \qquad (28)$$

where p: $R \times R \to R$ is defined by

$$p(t, \lambda) = \begin{cases} \lambda t + \tilde{\phi}(t) & \text{if } \lambda + \tilde{\phi}'(t) \geq 0, \\ \min\limits_{\tau \in R} \{\lambda\tau + \tilde{\phi}(\tau)\} & \text{if } \lambda + \tilde{\phi}'(t) < 0, \end{cases}$$

with

$$\tilde{\phi}(t) = c^2\,\phi(t/c).$$

If ϕ satisfies assumption (*G*) so does $\tilde{\phi}$ and vice versa. For a quadratic penalty function $\phi(t) = \tilde{\phi}(t) = \frac{1}{2}t^2$ we obtain

$$\frac{1}{c} p(ct, \lambda) = \frac{1}{2c} [(\max\{0, \lambda + ct\})^2 - \lambda^2]$$

which is the original function considered by Rockafellar [R1–3]. Thus minimization of $L_c(x, z, y)$ with respect to (x, z) is equivalent to minimization of $L_c(x, y)$ with respect to x which does not involve the variables $z_1, \ldots, z_r$.

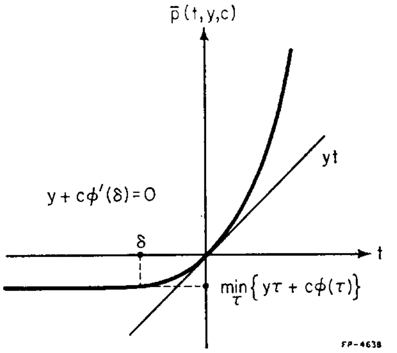

FIG. 3. Penalty function for one-sided inequality constraints.

The minimization in (27) yields the minimizing values of $z_1, \ldots, z_r$ as functions of x, y, c and once these values are substituted in the multiplier iteration (15) for problem (26) one obtains by straightforward calculation the iteration

$$\begin{aligned} y_{k+1}{}^j &= \max\{0, y_k{}^j + c_k\,\phi'[g_j[x(y_k, c_k)]]\} \\ &= \max\{0, y_k{}^j + \tilde{\phi}'[c_k\,g_j[x(y_k, c_k)]]\}, \quad j = 1, \ldots, r. \end{aligned} \qquad (29)$$

where $x(y_k, c_k)$ minimizes (locally) $L_{c_k}(\cdot, y_k)$ given by (27) or (28).

Thus even though additional variables are used to convert problem (25) to the equality constrained problem (26), the multiplier iteration for problem (26) takes the form (29) which does not involve these additional variables.

Two-sided inequality constraints. Consider the following problem involving two-sided inequality constraints

$$\min_{\substack{\alpha_j \leq g_j(x) \leq \beta_j \\ j=1, \ldots, r}} f(x), \qquad (30)$$

where f, g_j are real-valued functions on R^n and α_j, β_j are given scalars with $\alpha_j < \beta_j$ for all j. Each two-sided constraint in problem (30) could, of course, be separated into two one-sided constraints which could be treated as described earlier in this section. This would require, however, the assignment of two multipliers per two-sided constraint. A more efficient method for handling such constraints which requires only *one* multiplier per two-sided constraint is obtained by considering the following problem

$$\min_{\substack{\alpha_j \leq g_j(x) - u_j \leq \beta_j \\ u_j = 0, \; j=1, \ldots, r}} f(x) \qquad (31)$$

which involves additional variables $u_1, \ldots, u_r$ and is equivalent to problem (30). Now consider a multiplier method for problem (31) where only the constraints $u_j = 0$ are eliminated by means of a generalized penalty function, a partial elimination of constraints. This method consists of sequential minimizations over $x, u_1, \ldots, u_r$ of the form

$$\left.\begin{aligned} &\text{minimize} \quad f(x) + \sum_{j=1}^{r} [y_k{}^j u_j + c_k\,\phi(u^j)], \\ &\text{subject to} \quad \alpha_j \leq g_j(x) - u_j \leq \beta_j, \quad j = 1, \ldots, r. \end{aligned}\right\} \qquad (32)$$

The multipliers $y_k{}^j$ are updated by means of the iteration

$$y_{k+1}{}^j = y_k{}^j + c_k\,\phi'(u_j{}^k), \quad j = 1, \ldots, r, \qquad (33)$$

where $u_1{}^k, \ldots, u_r{}^k$ together with a vector x_k, solve problem (32). Now the minimization in problem (32) can be carried out first with respect to u_j yielding the equivalent problem

$$\min_x \left\{ f(x) + \sum_{i=1}^{r} p_j[g_j(x), y_k{}^j, c_k] \right\}, \qquad (34)$$

where the function p_j is defined by

$$p_j[g_j(x), y^j, c] = \begin{cases} y^j[g_j(x) - \beta_j] + c\,\phi[g_j(x) - \beta_j] \\ \qquad \text{if } y^j + c\,\phi'[g_j(x) - \beta_j] \geq 0, \\ y^j[g_j(x) - \alpha_j] + c\,\phi[g_j(x) - \alpha_j] \\ \qquad \text{if } y^j + c\,\phi'[g_j(x) - \alpha_j] \leq 0, \\ \min\limits_{\tau \in R} \{y^j\,\tau + c\,\phi(\tau)\} \\ \qquad \text{otherwise.} \end{cases}$$

The form of the function above is shown in Fig. 4. Notice that this function is continuously differentiable in x whenever the functions g_j are continuously differentiable.

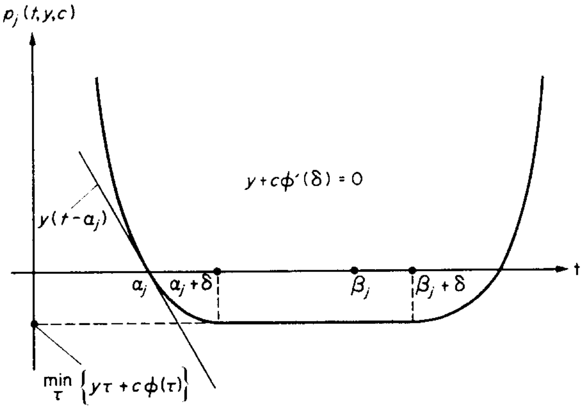

FIG. 4. Penalty function for two-sided inequality constraints.

The minimization in (32) yields the minimizing values of $u_1, \ldots, u_r$ as functions of x, y_k and c_k and once these values are substituted in (33) one obtains after straightforward calculation the multiplier iteration

$$y_{k+1}{}^j = \begin{cases} y_k{}^j + c_k\,\phi'[g_j(x_k) - \beta_j] \\ \qquad \text{if } y_k{}^j + c_k\,\phi'[g_j(x_k) - \beta_j] \geq 0, \\ y_k{}^j + c_k\,\phi'[g_j(x_k) - \alpha_j] \\ \qquad \text{if } y_k{}^j + c_k\,\phi'[g_j(x_k) - \alpha_j] \leq 0, \\ 0 \qquad \text{otherwise,} \end{cases}$$

where x_k solves the problem (34). Thus again the additional variables $z_1, \ldots, z_r$ need not enter explicitly into the computations.

Notes and references. The extension of multiplier methods to cover the case of one-sided inequality constraints is due to Rockafellar [R2, R1]. The method for treating two-sided inequality constraints by using a single multiplier per constraint was given recently by the author [B8, B9]. A similar approach has been used to construct approximation algorithms for non-differentiable or ill-conditioned optimization problems [B7–10, G2]. A number of approaches other than the one presented here have been suggested for treating inequality constraints [K1–5, M1, M2, P5, S1, A1]. In these approaches, Augmented Lagrangians suitable for inequality constraints are introduced directly. An alternative approach which attempts to identify those inequality constraints which are active at the optimal solution and subsequently treats them as equality constraints does not seem to be very satisfactory.

5. *Multiplier methods for convex programming*

The developments of Sections 2 and 3 require second-order sufficiency assumptions for problem (1) but do not require convexity of the objective or the constraint functions. For the case of a convex programming problem much stronger convergence and rate of convergence results may be obtained. Consider the problem

$$\left.\begin{array}{ll}\text{minimize} & f(x),\\ \text{subject to} & x \in X \subset R^m, \quad g_1(x) \leqslant 0, \ldots, g_r(x) \leqslant 0.\end{array}\right\} \quad (35)$$

For simplicity we do not consider (affine) equality or two-sided constraints. The results to be presented have direct extensions to cases where such constraints are present.

Assumption (C): The functions $f, g_1, \ldots, g_r$ are real-valued convex functions and the set X is closed and convex. Furthermore, problem (35) has a non-empty and compact set of optimal solutions X^* and a non-empty and compact set of Lagrange multiplier vectors Y^*, or Kuhn–Tucker vectors according to the definition of [R7].

Consider in accordance with the previous section the Augmented Lagrangian as in (28).

$$L_c(x, y) = f(x) + \frac{1}{c} \sum_{j=1}^{r} p[cg_j(x), y^j], \quad (36)$$

where $p: R \times R \to R$ is defined by

$$p(t, \lambda) = \begin{cases} \lambda t + \tilde{\phi}(t) & \text{if } \lambda + \tilde{\phi}'(t) \geqslant 0,\\ \min_{\tau \in R}\{\lambda \tau + \tilde{\phi}(t)\} & \text{if } \lambda + \tilde{\phi}'(t) < 0,\end{cases}$$

and $\tilde{\phi}$ satisfies assumption (G). The multiplier iteration is given by

$$y_{k+1}{}^j = \max\{0, y_k{}^j + \tilde{\phi}'[c_k g_j(x_k)]\}, \quad j = 1, \ldots, r, \quad (37)$$

where x_k is any minimizing point of $L_{r_k}(\cdot, y_k)$ over the constraint set X. One may show that under (C) for any $y \in R^m$ with $y^j \geqslant 0$, $j = 1, \ldots, r$, and $c > 0$ there exist such minimizing points of $L_c(\cdot, y)$ [K2, K3]. The following result has been proved by Kort and the author [K2–5]. A related result was proved independently by Rockafellar [R3] for the case $\tilde{\phi}(t) = \frac{1}{2}t^2$ under somewhat different assumptions.

Proposition 4: Let (C), (G) hold. Then for any starting point $y_0 \in R^m$ with $y_0{}^j \geqslant 0$, $j = 1, \ldots, r$, and any $\{c_k\}$ bounded below by a positive number, a sequence $\{y_k\}$ generated by iteration (37) satisfies

$$\lim_{k\to\infty} |||\, y_k - Y^* \,||| = 0,$$

where Y^* is the set of Lagrange multiplier vectors of problem (35) and $||| \cdot |||$ is given for all $y \in R^m$ by

$$|||\, y - Y^* \,||| = \min_{y^* \in Y^*} \| y - y^* \|.$$

Notice that there is no requirement that c_k is sufficiently large. Furthermore, convergence is attained for an arbitrary starting point y_0. One may also prove that any limit point of a generated sequence $\{x_k\}$ is an optimal solution of problem (35). Rockafellar [R3] has also proved the remarkable fact that for $\tilde{\phi}(t) = \frac{1}{2}t^2$ the sequence $\{y_k\}$ converges to some Lagrange multiplier vector $\bar{y}$, even though Y^* may contain more than one point. A similar result is not available for $\tilde{\phi}$ nonquadratic.

We also note that the iteration (37) has an interesting geometrical interpretation in dual space as a proximation algorithm as shown in [R3] for $\tilde{\phi}(t) = \frac{1}{2}t^2$ and [K3–5] for the more general case.

Concerning the rate of convergence of iteration (37) the following result is given in [K2] and a weaker version is given in [K3, K4].

Proposition 5: Let (C), (G) hold and assume:

(a) there exist scalars $M_2 \geqslant M_1 > 0$ and $\rho > 1$ such that for some neighborhood N_0 of the origin

$$M_1 \,|\, t \,|^{\rho-1} \leqslant |\, \tilde{\phi}'(t) \,| \leqslant M_2 \,|\, t \,|^{\rho-1}, \quad \forall t \in N_0; \quad (38)$$

(b) there exist scalars $\delta > 0$, $\gamma > 0$, $\sigma > 1$ such that the dual functional defined by

$$q_0(y) = \begin{cases} \inf_{x \in X}\left\{f(x) + \sum_{j=1}^{r} y^j g_j(x)\right\} & \text{if } y^j \geqslant 0,\\ & j = 1, \ldots, m,\\ -\infty & \text{otherwise}\end{cases} \quad (39)$$

satisfies

$$q_0(y) - \max_y q_0(y) \leqslant -\gamma \,|||\, y - Y^* \,|||^\sigma,$$
$$\forall y \in \{\bar{y} \mid |||\, y - Y^* \,||| \leqslant \delta\}. \quad (40)$$

Then if $\alpha = (\rho - 1)^{-1}(\sigma - 1)^{-1} > 1$ there holds

$$\limsup_{k\to\infty} \frac{|||\, y_{k+1} - Y^* \,|||}{|||\, y_k - Y^* \,|||^\alpha} < +\infty,$$

i.e. the sequence $\{|||\, y_k - Y^* \,|||\}$ converges to zero super-linearly with order of convergence at least α.

Some remarks may be helpful in explaining the assumptions and the conclusion of the above proposition. Assumption (a) is a growth assumption on $\tilde{\phi}$. Roughly speaking it states that, in a neighborhood of zero, $\tilde{\phi}(t)$ behaves like $|\, t \,|^\rho$. The assumption is satisfied, for example, for

$$\tilde{\phi}(t) = |\, t \,|^{\rho_1} + |\, t \,|^{\rho_2} + \ldots + |\, t \,|^{\rho_s}$$

with $\rho = \min\{\rho_1, \rho_2, \ldots, \rho_s\}$. Similarly assumption (b) is a growth assumption on the dual functional q_0. It says that in a neighborhood of the maximum set Y^*, $q_0(y)$ grows (downward) at least as fast as $\gamma \,|||\, y - Y^* \,|||^\sigma$. When q_0 has a unique maximizing point $\bar{y}$ and is twice continuously differentiable in a neighborhood around that point, as for example under assumption (S), then σ may be taken equal to 2. When q_0 is a polyhedral concave functional, as for example in a linear programming problem, then σ may be taken arbitrarily large.

Proposition 5 shows that the rate of convergence of iteration (37) depends crucially on the form of the penalty function $\tilde{\phi}$ (via ρ) and on the form of the dual functional q_0 (via σ). Thus by using the penalty function

$$\tilde{\phi}(t) = |\, t \,|^\rho$$

or the computationally more efficient (see [K5]), penalty function

$$\tilde{\phi}(t) = |\, t \,|^\rho + \tfrac{1}{2} t^2$$

and ρ sufficiently close to unity, *an arbitrarily high order of convergence may be achieved.* In Example 2 of Section 3 we have, by considering the equivalent convex programming problem $\min\{\frac{1}{2}x^2 \mid x \geq 0\}$, $\sigma = 2$ and $\rho = \frac{3}{2}$ and the order 2 convergence is explained by the above proposition. In Example 2 we have $\sigma = \frac{3}{2}$ and $\rho = 2$ and the superlinear convergence is again explained. In linear programming problems, where σ may be taken arbitrarily large, the order of convergence is infinity. This is consistent with a result of [B6] which shows that for polyhedral convex programs and when $\phi(t) = \frac{1}{2}t^2$ the method of multipliers converges in a finite number of iterations.

One may show that if $\rho = \sigma = 2$ in (38) and (40), as for example when (*S*) and (*Q*) hold, the convergence rate of $\{|||y_k - Y^*|||\}$ is at least linear as shown in [K2]. When $(\rho-1)^{-1}(\sigma-1)^{-1} < 1$ the convergence rate may be sublinear as shown in Example 1 of Section 3. In any case, however, the convergence rate of methods of multipliers is much more favorable than the one of the corresponding penalty methods where $L_{c_k}(x, 0)$ is minimized for a sequence $c_k \to \infty$, i.e. $y_k = 0$, $\forall k$. This is shown in [K2] where an estimate on the convergence rate of such penalty methods is derived. Given that multiplier methods under (*C*) are convergent for an arbitrary non-decreasing penalty parameter sequence, their advantages over penalty methods are overwhelming when the problem is convex.

6. *Inexact minimization—variations of multiplier methods*

The methods described in previous sections may be viewed as the basic multiplier iterations. As is frequently the case in constrained minimization methods it is possible to introduce a number of variations and modifications which are aimed at improving computational efficiency. There is a plethora of such modifications of multiplier methods as well as other related methods and in this section we make an attempt to classify them in three broad categories.

(a) *Methods with inexact minimizations.* These methods are characterized by the fact that the minimization of the Augmented Lagrangian (11) is not carried out exactly but only approximately. The unconstrained minimization process is terminated as soon as some stopping criterion is satisfied. The stopping criterion becomes more stringent after every multiplier iteration so that minimization is asymptotically exact.

Consider for convenience the case of equality constraints only and a quadratic penalty $\phi(t) = \frac{1}{2}t^2$. The most natural possibility is to terminate the kth unconstrained minimization when a point x_k is found satisfying

$$\| \nabla L_{c_k}(x_k, y_k) \| \leq \varepsilon_k \tag{41}$$

where $\{\varepsilon_k\}$ is a decreasing sequence with $\varepsilon_k \to 0$. This was suggested by Buys [B7] as a natural extension of related procedures for penalty methods. Another possibility is to use the stopping criterion.

$$\| \nabla L_{c_k}(x_k, y_k) \| \leq \eta_k \| h(x_k) \| \tag{42}$$

or

$$\| \nabla L_{c_k}(x_k, y_k) \| \leq \min\{\varepsilon_k, \eta_k \| h(x_k) \|\} \tag{43}$$

with $\{\varepsilon_k\}$ and $\{\eta_k\}$ decreasing sequences with $\varepsilon_k \to 0$, $\eta_k \to 0$. These and other closely related stopping criteria have been considered in [B1–4, P2, K2–5]. Rockafellar in his convergence analysis of [R3] considered terminating the minimization at an ε_k-optimal solution of the Augmented Lagrangian (with $\varepsilon_k \to 0$) and showed later [R6] the relation of such termination procedures with implementable gradient-based stopping criteria such as those mentioned above.

When inexact minimization is employed the multiplier iteration may take several alternate forms. One possibility is to use the same iteration as with exact minimization

$$y_{k+1} = y_k + c_k h(x_k). \tag{44}$$

Other possibilities include the iteration

$$y_{k+1} = y_k + \beta_k h(x_k), \tag{45}$$

where

$$\beta_k = c_k - \frac{h(x_k)' \nabla h(x_k) \nabla L_{c_k}(x_k, y_k)}{h(x_k)' \nabla h(x_k) \nabla h(x_k)' h(x_k)} \tag{46}$$

suggested by Miele *et al.* [21] in a somewhat different setting and the iteration

$$y_{k+1} = -[\nabla h(x_k) \nabla h(x_k)']^{-1} \nabla h(x_k) \nabla f(x_k) \tag{47}$$

suggested by Haarhoff and Buys [H2], Buys [B7] and Miele *et al.* [21]. Both iterations (45), (47) reduce to the basic iteration (44) when the unconstrained minimization is carried out exactly.

Computational experience thus far indicates that considerable savings may be realized by making use of inexact minimization. It is unclear whether from the practical point of view any particular combination of the stopping criteria (41), (42) or (43) and of the iterations (44), (45) or (47) is superior than the others. From the theoretical point of view it has been established [B5] that the stopping criterion (41) may lead to a substantial deterioration of the convergence rate of the method to the point where the rate of convergence is not any more linear. By contrast the stopping criteria (42) and (43) are characterized by an asymptotic rate of convergence which is identical to the one associated with exact unconstrained minimization [B5]. Additional convergence rate analysis relating to methods employing inexact minimization may be found in [B1–3, P2] for non-convex problems and in [K2–5] for convex programming. For convergence analysis see [B1–5, K2–5, P2, R3].

(b) *Modified multiplier iterations.* Methods in this category are characterized by multiplier iterations which are different from the basic iteration (44) even for the case when minimization of the Augmented Lagrangian is exact. Consider again the case of equality constraints only. Based on the analysis of Section 3 the method of multipliers is a steepest ascent method for maximizing the dual functional q_c of (19). It is possible to consider, in place of the steepest ascent iteration, a Newton or Quasi-Newton iteration of the form

$$y_{k+1} = y_k - G_k \nabla g_c(y_k) \tag{48}$$

with G_c equal to either the inverse Hessian matrix $[\nabla^2 q_c(y_k)]^{-1}$ where

$$\nabla^2 q_c(y_k) = -\nabla h(x_k) [\nabla^2 L_c(x_k, y_k)]^{-1} \nabla h(x_k)' \tag{49}$$

or some approximation to $[\nabla^2 q_c(y_k)]^{-1}$. If the minimization of $L_c(x, y_k)$ is carried out by a method utilizing second derivatives, such as Newton's method, then $\nabla^2 q_c$ is available at the end of the minimization cycle and little additional effort is required to implement (48). This possibility has been suggested by Buys [B11]. Another possibility suggested by Fletcher [F5] (see also Brusch [B12]) rests on the fact that when $L_c(x, y_k)$ is minimized by using a quasi-Newton method such as the Davidon–Fletcher–Powell method one obtains usually, but not always, an approximation of $[\nabla^2 L_c(x_k, y_k)]^{-1}$ which in turn could be used to generate an approximation of $\nabla^2 q_c$ via (49). This last possibility does not require the availability and computation of second derivatives—a major advantage. Other modifications suggested for convex programming problems are based on alteration of the stepsize of the basic method of multipliers by means of extrapolation [B4, B5].

It is difficult to evaluate the utility of modifications such as the ones described above since there is not much related computational experience available at the present time. Undoubtedly iterations such as (48), (49) or Quasi-Newton versions of them become less desirable as the dimension of the problem increases. It is also unclear whether the potential computational savings are worth the extra programming effort involved. This is particularly so since the simple steepest ascent iteration of the method of multipliers already has good convergence rate. Regarding convergence of iterations such as (48) there is available a general local convergence result due to Buys ([B11], Theorem 4·9), which guarantees convergence provided the starting point y_0 is sufficiently close to $\bar{y}$. Some rate of convergence analysis for a Quasi-Newton iteration of the form (48) was also given recently in [H3]. We offer below a new global convergence result for the case of Newton's method where $G_k = [\nabla^2 q_c(y_k)]^{-1}$ in (48) and the Hessian matrix $\nabla^2 q_c$ is given by (49). The result is given for $\phi(t) = \frac{1}{2}t^2$. A related result may be proved for a general essentially quadratic penalty function. The proof follows in a simple manner from the analysis of [B1] and [B3].

Proposition 6: Let the assumptions of Proposition 1 hold, and assume $\phi(t) = \frac{1}{2}t^2$. For any given bounded set $Y \subset R^m$ let c^* and $x(y, c)$ be as in Proposition 1. Then there exists a scalar $\hat{M}$ such that

$$\| \hat{y}(y, c) - \bar{y} \| \leqslant \frac{\hat{M} \| y - \bar{y} \|^2}{c^2}, \quad \forall c > c^*, \quad y \in Y,$$

where

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

Proof: Consider the duality framework of [B1] (Section 4) or [B3] (Section 3) and the functions p, p_c and d_c defined there. We have in terms of the notation of [B1, B3]

$$\begin{aligned}\hat{y}(y, c) - \bar{y} &= y - \bar{y} - [\nabla^2 d_c(y)]^{-1} h[x(y, c)]\\ &= \nabla p_c(0) - \nabla p_c[h[x(y, c)]] + \nabla^2 p_c[h[x(y, c)]]\\ &\quad \times h[x(y, c)].\end{aligned}$$

Now under our assumptions h is Lipschitz continuous in the region of interest with Lipschitz constant denoted by L, and $\nabla^2 p_c$ is Lipschitz continuous with Lipschitz constant, denoted by K, which is independent of c. By using these facts and Taylor's theorem in the expression above we obtain

$$\| \hat{y}(y, c) - \bar{y} \| \leqslant K \| h[x(y, c)] \|^2/2 \leqslant KL^2 \| x(y, c) - \bar{x} \|^2/2.$$

Using the result of Proposition 1 (c.f. (12)) we obtain

$$\| \hat{y}(y, c) - \bar{y} \| \leqslant \frac{KL^2 M^2 \| y - \bar{y} \|^2}{2c^2}$$

and by setting $\hat{M} = KL^2 M^2/2$ the desired relation follows. Q.E.D.

It is to be noted that one may easily modify the proof of the proposition above to show that if $[D_c(y)]^{-1}$ is replaced in the definition of $\hat{y}(y, c)$ by a matrix $G_c(y)$ where

$$\| G_c(y) - [D_c(y)]^{-1} \| \leqslant \varepsilon$$

and ε is some constant, then

$$\| \hat{y}(y, c) - \bar{y} \| \leqslant \frac{\hat{M} \| y - \bar{y} \|^2}{c^2} + \frac{\varepsilon \hat{M}_1 \| y - \bar{y} \|}{c}$$

where $\hat{M}$ is the constant in Proposition 6 and $\hat{M}_1$ is some other constant. If $G_c(y)$ satisfies

$$\| G_c(y) - [D_c(y)]^{-1} \| \leqslant \varepsilon \| h[x(y, c)] \|$$

then the estimate becomes

$$\| \hat{y}(y, c) - \bar{y} \| \leqslant \frac{(\hat{M} + \varepsilon \hat{M}_2) \| y - \bar{y} \|^2}{c^2},$$

where $\hat{M}$ is the constant in Proposition 6 and $\hat{M}_2$ is some other constant. The estimates above show the extent of degradation of speed of convergence when approximations to the inverse Hessian of the dual functional are used in a Newton iteration. Such approximations, for example, may be obtained through equation (49) when x_k is an approximate minimizing point of the Augmented Lagrangian.

Based on Proposition 6 one may establish a convergence result similar to Proposition 2 for the Newton iteration (48) where G_k is equal to $[\nabla^2 q_{c_k}(y_k)]^{-1}$ as given by (49) with x_k being a minimizing point of $L_{c_k}(\cdot, y_k)$. Convergence is obtained for *any starting point* provided the initial penalty parameter is sufficiently large. This is perhaps surprising since usually one can guarantee only local convergence for Newton's method. The convergence rate is at least second order and is governed by

$$\| y_{k+1} - \bar{y} \| \leqslant \frac{\hat{M} \| y_k - \bar{y} \|^2}{c_k^2}$$

which indicates extremely fast convergence. Notice that the error $\| y_{k+1} - \bar{y} \|$ is inversely proportional to the *square* of c_k rather than c_k as in the case of the steepest ascent iteration. From the practical point of view, however, the potential computational savings associated with the Newton iteration are not spectacular since the ascent iteration usually takes very few iterations to converge to within a satisfactory degree of accuracy. In fact, 3–6 is a typical range.

For the case of the inequality constrained problem

$$\min_{\substack{g_j(x) \leqslant 0 \\ j=1, \dots, r}} f(x)$$

the analysis given above is applicable once the problem is reformulated to one involving exclusively equality constraints by introducing additional squared variables. The form of the Hessian matrix of the dual functional q_c can be obtained by differentiation of the gradient ∇q_c which can be calculated to be

$$\frac{\partial q_c(y)}{\partial y^j} = \max \left\{ g_j[x(y, c)], -\frac{y^j}{c} \right\}, \quad j = 1, \dots, r.$$

Let us reorder the indices of y and g so that for some integer p with $0 \leqslant p \leqslant r$

$$g_j[x(y, c)] > -\frac{y^j}{c} \quad \text{for } j = 1, \dots, p,$$

$$g_j[x(y, c)] < -\frac{y^j}{c} \quad \text{for } j = p+1, \dots, r.$$

It is possible to prove under our assumptions that for c sufficiently large we have $g_j[x(y, c)] \neq -y^j/c$ for all j. Then we can write

$$\nabla^2 q_c(y) = \left[\begin{array}{c|c} D_c(y) & 0 \\ \hline 0 & -\frac{1}{c} I \end{array} \right]$$

with I is the $(r-p) \times (r-p)$ identity matrix and

$$D_c(y) = -G(y, c) [\nabla^2 L_c[x(y, c), y]]^{-1} G(y, c)',$$

where $G(y, c)$ is the $p \times n$ matrix having as rows the gradients $\nabla g_1[x(y, c)], \dots, \nabla g_1[x(y, c)]$, and L_c is given by (27). In view of the above relations the Newton iteration for the dual problem

$$y_{k+1} = y_k - [\nabla^2 q_c(y_k)]^{-1} \nabla q_c(y_k)$$

can be written in a convenient form. Similar Newton iterations may be obtained for the case where there are two-sided inequality constraints present.

(c) *Related methods*. There have been several methods proposed recently which utilize the penalty function idea as well as multiplier adjustment formulas. Many of these methods do not involve exact or even asymptotically exact minimization of an Augmented Lagrangian although they do include intermediate iterations on the primal variables, while others involve 'continuous' adjustment schemes for the multipliers. Such methods cannot be viewed as variations of multiplier methods but rather as independent methods. Their relation, however, with multiplier methods, though not very clear as yet, is undoubtedly strong and thus it would be an omission if we did not mention them.

Several methods of the Lagrangian type have been proposed by Miele and his associates [M5–8] for the case of equality constraints only. These methods are seeking saddle points of a Lagrangian function to which a quadratic penalty term has been added. For none of these methods there exists yet a convergence or rate of convergence analysis although some local convergence results can be inferred using known results for Lagrangian methods, see e.g. [P6]. A wide class of Lagrangian methods has also been proposed and rigorously analyzed by Mangasarian [M1, 2]. These methods are applicable to inequality constraints as well, and do not require positivity constraints on the corresponding Lagrange multipliers. Furthermore, the related Lagrangian functions have strong differentiability properties.

A different type of method has been proposed by Fletcher [F2] based on the observation that the problem

$$\min_{h(x)=0} f(x) \tag{50}$$

may be solved by solving

$$\min_{x} f(x) - h(x)'N^{+}\nabla f(x) + \tfrac{1}{2}\, c h(x)'Qh(x), \tag{51}$$

where the $m \times n$ matrix N^{+} is given by

$$N^{+} = [\nabla h(x)\, \nabla h(x)']^{-1}\, \nabla h(x). \tag{52}$$

The penalty parameter c is sufficiently large and Q is a positive definite matrix, perhaps depending on x. In this way a single minimization is required to solve the problem as opposed to a sequence of minimizations required in the multiplier method. The vector

$$-N^{+}\nabla f(x)$$

may be viewed as an estimate of the Lagrange multiplier of the problem, as in (47), and thus (51) may be viewed as minimization of an Augmented Lagrangian where the multiplier vector is not constant but rather continuously depends on the value of x. Fletcher has proposed several possible implementations of his ideas [F2–4] and has provided related convergence and rate of convergence analysis. In particular he has provided [F4] superlinearly convergent algorithms for solving problem (51) which employ only second derivatives of the functions f and h, as opposed to third derivatives that Newton's method requires. A related algorithm which incorporates an automatic scheme for adjusting the penalty parameter has been proposed and analyzed recently by Mukai and Polak [M9]. Two drawbacks of methods based on Fletcher's ideas are the lack of *a priori* knowledge of a suitable range for the penalty parameter, although the implementation of [M9] alleviates somewhat this disadvantage, and the need to compute the matrix N^{+} of (52) at each function evaluation—an expensive operation when the number of constraint equations is large. In addition a satisfactory way to handle inequality constraints directly has not as yet been established.

7. *Infinite dimensional problems*

The minimization problems considered thus far are defined over a finite dimensional space and there is a finite number of equality and inequality constraints. For such problems the convergence analysis has reached to date a fairly satisfying stage of development. This is not as yet the case for infinite dimensional problems such as those appearing in continuous time optimal control. There have been, however, some advances in this area. Consider three, possibly infinite dimensional, linear spaces S, S_1, S_2 and the problem

$$\left.\begin{aligned} &\text{maximize} \quad f(x), \\ &\text{subject to} \quad x \in X, \quad h(x) = 0, \quad g(x) \leqslant 0, \end{aligned}\right\} \tag{53}$$

where $f\colon S \to R$, $X \subset S$, $h\colon S \to S_1$, $g\colon S \to S_2$ and the inequality $g(x) \leqslant 0$ is with respect to a cone $C \subset S_2$ [L3], i.e. we write $g(x) \leqslant 0$ if $g(x) \in C$.

The present paper has dealt so far with the case where S, S_1, S_2 are finite dimensional spaces. Infinite dimensional special cases of problem (53) and multiplier methods based on quadratic-like penalty functions have been considered by Rockafellar [R3], Rupp [R8–10], Polyak and Tret'yakov [P2] and Wierzbicki and Kurcyusz [W2]. Reference [R3] considers a convex programming problem with no equality constraints and S_2 finite dimensional and provides global convergence analysis for the multiplier iteration (29) with $\phi(t) = \frac{1}{2} t^2$. References [R8–10] consider various optimal control problems where there are either a finite number of isoperimetric constraints or the dynamic system equation constraint is eliminated by means of a generalized penalty function. A result similar to Proposition 1 of this paper has been given in [P2] for the case where S, S_1 are Hilbert spaces, $X = S$ and there are no inequality constraints. Contrary to the finite dimensional case the treatment of inequality constraints is not anymore a simple extension of the equality case. Nonetheless inequality constraints have been treated in an interesting manner in [W2] for the case where S_2 is a Hilbert space and some convergence analysis has been given. There are no results so far dealing with the case where the constraint spaces S_1 or S_2 are infinite-dimensional Banach spaces. Such results would be most useful since they would cover important cases of optimal control problems with state constraints.

It is to be noted that from the point of view of practical computation it is important to take into account the fact that solution of an infinite dimensional problem in a digital computer requires some type of 'finite dimensional' approximation procedure. There are two general computational approaches for such problems. The first is to approximate the infinite dimensional problem by a finite dimensional version and then apply a particular algorithm to the latter. There has been considerable outstanding work by Jane Cullum [C1–4], and others [B13, D1] relating to approximation procedures for optimal control problems which are stable in the sense that the original problem may be solved to within an arbitrary degree of accuracy by solving a finite dimensional approximation of it. There is of course no difficulty in combining multiplier methods with such an approach. The second possibility is to establish the validity of a certain 'theoretical' algorithm for solution of an infinite dimensional problem and subsequently provide an implementable version, i.e. a finite dimensional approximation, of this algorithm. There has been very little analysis in the literature addressing the related theoretical and practical issues even for classical methods, a notable exception being the adaptive integration procedure of Klessig and Polak [K6] for unconstrained optimal control problems. No attention has been given so far to approximation procedures in connection with multiplier methods as applied to infinite dimensional problems and this appears to be an interesting subject for investigation.

8. *Conclusions*

The presentation of the convergence properties of multiplier methods given in this paper should demonstrate to the reader that multiplier methods are superior to ordinary penalty methods for the great majority of practical problems. On the basis of this fact multiplier methods currently occupy a prominent position among methods for constrained optimization. In the opinion of most researchers in the field they are the best methods available for problems with nonlinear constraints in the absence of special structure. They are also very suitable for multi-dimensional problems with many constraints, such as constrained optimal control problems, where gradient projection methods and the reduced gradient method, or Newton and quasi-Newton versions of them, may encounter difficulties due to large dimensionality. At the present time the properties of multiplier methods as applied to finite dimensional problems seem to be fairly well understood. This is not as yet the case for infinite dimensional problems. A considerable amount of research remains to be done in this area. Furthermore, future research will undoubtedly provide new insights and will broaden the scope of application of primal-dual iterations in constrained optimization and possibly in other areas such as, for example, approximation.

References

[A1] K. J. ARROW, F. J. GOULD and S. M. HOWE: A general saddle point result for constrained optimization. *Inst. of Statistics Mimeo Ser. No.* 774, Univ. of North Carolina, Chapel Hill, NC (1971). Also *Mathematical Programming* **5**, 225–234 (1973).

[A2] K. J. ARROW, L. HURWICZ and H. UZAWA: *Studies in Linear and Nonlinear Programming.* Stanford University Press, CA (1958).

[B1] D. P. BERTSEKAS: On penalty and multiplier methods. Dept. of Engng—Economic Systems Working Paper, Stanford Univ., Standford, CA (1973) *SIAM J. Control* **14** (1976).

[B2] D. P. BERTSEKAS: Convergence rate of penalty and multiplier methods. *Proc. 1973 IEEE Conf. on Decision and Control,* pp. 260–264, San Diego, CA.

[B3] D. P. BERTSEKAS: On penalty and multiplier methods for constrained minimization. In *Nonlinear Programming 2* (O. MANGASARIAN, R. MEYER and S. ROBINSON, eds.), pp. 165–191. Academic Press, New York (1975).

[B4] D. P. BERTSEKAS: On the method of multipliers for convex programming. *IEEE Trans. Aut. Control* **AC-20**, 385–388 (1975).

[B5] D. P. BERTSEKAS: Combined primal-dual and penalty methods for constrained minimization. *SIAM J. Control* **13**, 521–544 (1975).

[B6] D. P. BERTSEKAS: Necessary and sufficient conditions for a penalty method to be exact. Dept. of Engng—Economic Systems Working Paper, Stanford Univ., Stanford, CA (1973) *Math. Programming* **9**, 87–99 (1975).

[B7] D. P. BERTSEKAS: Nondifferentiable optimization via approximation. *Proc. of Twelfth Annual Allerton Conf. on Circuit and System Theory.* pp. 41–52. Allerton Park, Illinois, October (1974). Also in *Mathematical Programming Study 3* (M. BALINSKI and P. WOLFE, eds.). North-Holland, 1975.

[B8] D. P. BERTSEKAS: Approximation procedures based on the method of multipliers. Coordinated Science Lab. Working Paper, University of Illinois, Urbana, IL, January (1976).

[B9] D. P. BERTSEKAS: A general method for approximation based on the method of multipliers. *Proc. of Thirteenth Annual Allerton Conf. on Circuit and System Theory.* Allerton Park, IL, October (1975).

[B10] D. P. BERTSEKAS: A new algorithm for analysis of nonlinear resistive networks. *Proc. of Thirteenth Annual Conf. on Circuit and System Theory, Allerton* Park, IL, October (1975).

[B11] J. D. BUYS: Dual algorithms for constrained optimization. Ph.D. Thesis, Rijksuniversiteit de Leiden (1972).

[B12] R. B. BRUSCH: A rapidly convergent method for equality constrained function minimization. *Proc. of 1973 IEEE Conf. on Decision and Control,* 80–81, (1973).

[B13] B. M. BUDAK, E. M. BERKOVICH and E. N. SOLOV'EVA: Difference approximations in optimal control problems. *SIAM J. Control* 7, 18–31 (1969).

[C1] J. CULLUM: Perturbations of optimal control problems. *SIAM J. Control* 4, 473–487 (1966).

[C2] J. CULLUM: Discrete approximations to continuous optimal control problems. *SIAM J. Control* 7, 32–49 (1969).

[C3] J. CULLUM: An explicit method for discretizing continuous optimal control problems. *J. Opt. Theory Appl.* **8**, 15–34 (1971).

[C4] J. CULLUM: Finite-dimensional approximations of state-constrained continuous optimal control problems. *SIAM J. Control* **10**, 649–670 (1972).

[D1] J. W. DANIEL: On the convergence of a numerical method for optimal control problems. *J. Opt. Theory Appl.* 4, 330–342 (1969).

[F1] A. V. FIACCO and G. P. MCCORMICK: *Nonlinear Programming: Sequential Unconstrained minimization Techniques.* Wiley, New York (1968).

[F2] R. FLETCHER: A class of methods for nonlinear programming with termination and convergence properties. *Integer and Nonlinear Programming* (J. ABADIE, ed.). North-Holland, Amsterdam (1970).

[F3] R. FLETCHER and S. LILL: A class of methods for nonlinear programming: II. Computational experience. In *Nonlinear Programming* (J. B. ROSEN, O. L. MANGASARIAN and K. RITTER, eds.), Academic Press, New York (1971).

[F4] R. FLETCHER: A class of methods for nonlinear programming: III. Rates of convergence. In *Numerical Methods for Nonlinear Optimization* (F. A. LOOTSMA, ed.). Academic Press, New York (1973).

[F5] R. FLETCHER: An ideal penalty function for constrained optimization. In *Nonlinear Programming 2* (O. MANGASARIAN, R. MEYER and S. ROBINSON, eds.), pp. 121–163. Academic Press, New York (1975).

[G1] E. G. GOL'SHTEIN and N. V. TRET'YAKOV: Modified Lagrangian functions. *Economics Math. Methods* **10**, (3), 568–591 (In Russian.) (1974).

[G2] D. GABAY and B. MERCIER: A dual algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation. *IRIA-LABORIA Research Report No.* 126 (1975).

[H1] M. R. HESTENES: Multiplier and gradient methods. *J. Opt. Theory Appl.* 4, 303–320 (1969).

[H2] P. C. HAARHOFF and J. D. BUYS: A new method for the optimization of a nonlinear function subject to nonlinear constraints. *Comput. J.* **13**, 178–184 (1970).

[H3] S. P. HAN: Penalty Lagrangian Methods via a Quasi-Newton Approach. Dept. of Computer Science, TR 75-252, Cornell University, Ithaca, New York, July (1975).

[K1] B. W. KORT and D. P. BERTSEKAS: A new penalty function method for constrained minimization. *Proc. 1972 IEEE Conf. on Decision and Control,* pp. 162–166, New Orleans, LA.

[K2] B. W. KORT and D. P. BERTSEKAS: Combined Primal Dual and Penalty Methods for Convex Programming. Dept. of Engng—Economic Syst. Working Paper, Stanford Univ., Stanford, CA (1973), revised September (1974). *SIAM J. Control.* **14** (1976).

[K3] B. W. KORT and D. P. BERTSEKAS: Multiplier methods for convex programming. *Proc. 1973 IEEE Conf. on Decision and Control,* pp. 428–432, San Diego, CA.

[K4] B. W. KORT: Rate of convergence of the method of multipliers with inexact minimization. In *Nonlinear Programming 2* (O. MANGASARIAN, R. MEYER and S. ROBINSON, eds.), pp. 193–214. Academic Press, New York (1975).

[K5] B. W. KORT: Combined Primal-Dual and Penalty Function Algorithms for Nonlinear Programming. Ph.D. Thesis, Stanford Univ., Palo Alto, CA (1975).

[K6] R. KLESSIG and E. POLAK: An adaptive precision gradient method for optimal control. *SIAM J. Control* **11**, 80–93 (1973).

[L1] D. G. LUENBERGER: *Introduction to Linear and Nonlinear Programming.* Addison-Wesley, Reading, MA (1973).

[L2] L. LASDON: *Optimization Theory for Large Systems.* Macmillan, New York (1970).

[L3] D. G. LUENBERGER: *Optimization by Vector Space Methods.* Wiley, New York (1969).

[M1] O. L. MANGASARIAN: Unconstrained Lagrangians in nonlinear programming. *SIAM J. Control* **13**, 772–791 (1975).

[M2] O. L. MANGASARIAN: Unconstrained methods in optimization. *Proc. of Twelfth Allerton Conf. on Circuit and System Theory,* pp. 153–160, Univ. of Illinois, Urbana, Ill., October (1974).

[M3] K. MARTENSSON: New Approaches to the Numerical Solution of Optimal Control Problems. Rep. 1206, Lund Inst. of Tech., Division of Automatic Control, Lund, Sweden (1972).

[M4] K. MARTENSSON: A new approach to constrained function optimization. *J. Opt. Theory Appl.* **12**, 531–554. (1973).

[M5] A. MIELE, P. E. MOSELEY, A. V. LEVY and G. M. COGGINS: On the method of multipliers for mathematical programming problems. *J. Opt. Theory Appl.* **10**, 1–33 (1972).

[M6] A. MIELE, P. E. MOSELEY and E. E. CRAGG: A modification of the method of multipliers for mathematical programming problems. In *Techniques of Optimization* (A. V. BALAKRISHNAN, ed.). Academic Press, New York (1972).

[M7] A. MIELE, E. E. CRAGG, R. R. IYER and A. V. LEVY: Use of the augmented penalty function in mathematical programming problems, Part I. *J. Opt. Theory Appl.* **8**, 115–130 (1971).

[M8] A. MIELE, E. E. CRAGG and A. V. LEVY: Use of the augmented penalty function in mathematical programming problems, Part II. *J. Opt. Theory Appl.* **8**, 131–153 (1971).

[M9] H. MUKAI and E. POLAK: A Quadratically Convergent primal-dual algorithm with global convergence properties for solving optimization problems with equality constraints. *Mathematical Programming* **9**, 336–349 (1975).

[N1] J. S. NEWELL and D. M. HIMMELBLAU: Nonlinear programming via a new penalty function. Presented at Nonlinear Programming Symposium, Madison, Wisc. (1974).

[N2] H. NAKAYAMA, H. SAYAMA and Y. SAWARAGI: Multiplier method and optimal control problems with terminal state constraints. *International Journal of Systems Science* **6**, 465–477 (1975).

[N3] J. E. NAHRA: Balance function for the optimal control problems. *J. Opt. Theory Appl.* **8**, 35–48 (1971).

[P1] E. POLAK: *Computational Methods in Optimization.* Academic Press, New York (1971).

[P2] V. T. POLYAK and N. V. TRET'YAKOV: The method of penalty estimates for conditional extremum problems. *USSR Computational Mathematics and Mathematical Physics* **13**, 42–58 (1974).

[P3] M. J. D. POWELL: A method for nonliner constraints in minimization problems. In *Optimization* (R. FLETCHER, ed.), pp. 283–298. Academic Press, New York (1969).

[P4] B. T. POLYAK: Gradient methods for the minimization of functionals. *Z. Vycisl. Mat. i Mat. Fiz.* **3**, pp. 643–653 (1963). (Translated in *USSR Computational Mathematics and Mathematical Physics.*)

[P5] M. A. POLLATSCHEK: Generalized duality theory in nonlinear programming. Operations Res. Statistics and Economics, Mimeo Ser. 122, Technion, Haifa, Istael (1973).

[P6] B. T. POLYAK: Iterative methods using Lagrange multipliers for solving extremal problems with constraints of the equation type. *Z. Vycisl. Mat. i Mat. Fiz.* **10**, pp. 1098–1106 (1970). (Translated in *USSR Computational Mathematics Mathematical Physics.*)

[P7] B. T. POLYAK and N. V. TRET'YAKOV: On an iterative method of linear programming and its economic interpretation. *Economics Math. Methods* **8**, 740–751 (1972). (In Russian.)

[R1] R. T. ROCKAFELLAR: New applications of duality in convex programming. 7th International Symposium on Math. Programming, The Hague (1970). Published in *Proc. 4th Conf. on Probability*, Brasov, Romania (1971).

[R2] R. T. ROCKAFELLAR: A Dual Approach to Solving Nonlinear Programming Problems by Unconstrained Optimization. *Math. Programming* **5**, 354–373 (1973).

[R3] R. T. ROCKAFELLAR: The multiplier method of Hestenes and Powell applied to convex programming. *J. Opt. Theory Appl.* **12**, 555–562 (1973).

[R4] R. T. ROCKAFELLAR: Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM J. Control* **12**, 268–285 (1974).

[R5] R. T. ROCKAFELLAR: Penalty methods and augmented Lagrangians in nonlinear programming. *Proc. 5th IFIP Conf. on Optimization Techniques*, Rome (1973). Springer-Verlag, (1974).

[R6] R. T. ROCKAFELLAR: Solving a nonlinear programming problem by way of a dual problem. *Symposia Mathematica*, to appear.

[R7] R. T. ROCKAFELLAR: *Convex Analysis.* Princeton University Press, Princeton, N.J. (1970).

[R8] R. D. RUPP: Approximation of the classical isoperimetric problem. *J. Opt. Theory Appl.* **9**, 251–264 (1972).

[R9] R. D. RUPP: A method for solving a quadratic optimal control problem. *J. Opt. Theory Appl.* **9**, pp. 238–250 (1972).

[R10] R. D. RUPP: A nonlinear optimal control minimization technique. *Trans. Am. Math. Soc.* **178**, 357–381 (1973).

[S1] H. SAYAMA, Y. KAMEYAMA, H. NAKAYAMA and Y. SAWARAGI: The Generalized Lagrangian Functions for Mathematical Programming Problems. Kansas State University Report, February (1974).

[T1] S. S. TRIPATHI and K. S. NARENDRA: Constrained Optimization Problems Using Multiplier Methods. *J. Opt. Theory and Appl.* **9**, 59–70 (1972).

[T2] N. V. TRET'YAKOV: The method of penalty estimates for problems of convex programming. *Economics Math. Methods* **9**, 525–540 (1973). (In Russian.)

[W1] A. P. WIERZBICKI: A penalty function shifting method in constrained static optimization and its convergence properties. *Archiwum Automatyki i Telemechaniki* **16**, 395–416 (1971).

[W2] A. P. WIERZBICKI and S. KURCYUSZ: Projection on a Cone, Generalized Penalty Functionals and Duality Theory. Institute of Automatic Control, Technical Univ. of Warsaw, Report No. 1/1974, February (1974).

[Z1] W. ZANGWILL: *Nonlinear Programming: A Unified Approach.* Prentice-Hall New York (1969).