SIAM J. CONTROL
Vol. 13, No. 3, May 1975

COMBINED PRIMAL-DUAL AND PENALTY METHODS FOR CONSTRAINED MINIMIZATION*

DIMITRI P. BERTSEKAS†

Abstract. In this paper we consider a class of combined primal-dual and penalty methods often called methods of multipliers. The analysis focuses mainly on the rate of convergence of these methods. It is shown that this rate is considerably more favorable than the corresponding rate for penalty function methods. Some efficient versions of multiplier methods are also considered whereby the intermediate unconstrained minimizations involved are approximate and only asymptotically exact. It is shown that such approximation schemes may lead to a substantial deterioration of the convergence rate, and a special approximation scheme is proposed which exhibits the same rate as the method with exact minimization. Finally, we analyze the properties of the step size rule of the multiplier method in relation to other possible step sizes, and we consider a modified step size rule for the case of the convex programming problem.

1. Introduction. During recent years, penalty function methods (see, e.g., [7]) have gained recognition as one of the most effective class of methods for solving constrained minimization problems. Characteristic of such methods is that they require the solution of a sequence of unconstrained minimizations of the objective function of the problem to which an increasingly high penalty term is added. It is well known that these unconstrained minimization problems have increasingly unfavorable structure due to ill-conditioning [7], [17], a fact which often leads to slow convergence despite the use of efficient unconstrained minimization algorithms. Another important class of methods for constrained minimization is the so-called class of primal-dual methods (see, e.g., [17]). Such methods are, in effect, iterative ascent algorithms for solving the dual problem (defined under suitable local convexity assumptions [17]). Similar to penalty methods, they involve the solution of a sequence of unconstrained minimizations of a Lagrangian function, each of which yields the value and the gradient of the dual functional at the current value of the Lagrange multiplier. At the end of each minimization, the Lagrange multiplier is updated by means of an ascent iteration. Primal-dual methods are known to have serious disadvantages. First, the problem must have a locally convex structure in order for the dual functional to be defined. Second, it is usually necessary to solve a large number of unconstrained minimization problems since the ascent iteration converges only moderately fast. Thus primal-dual methods have found application only in the limited class of problems where the unconstrained minimizations can be carried out very efficiently due to special structure.

In the last few years, a number of researchers have proposed a new class of methods, often 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 but rather to the Lagrangian function which is ordinarily

* Received by the editors February 12, 1973, and in final revised form March 8, 1974.

† This work was carried out at the Department of Engineering-Economic Systems, Stanford University, Stanford, California, and supported by the National Science Foundation under Grant GK-29237. Now at Department of Electrical Engineering, University of Illinois, Urbana, Illinois 61801.

minimized in primal-dual methods. Again, a sequence of unconstrained minimization problems is solved; however, each minimization is followed by an ascent iteration on the Lagrange multiplier which is aimed at solving the dual problem. In contrast with penalty methods, the penalty term need not be increased to infinity, thus avoiding the associated extreme ill-conditioning. In addition, the ascent iteration converges fast, thus necessitating only a few unconstrained minimization cycles. By moderating the disadvantages of both penalty and primal-dual methods, multiplier methods have emerged as a very attractive class of algorithms for constrained minimization, a fact substantiated by the limited computational experience presently available. This paper provides an analysis of some aspects of these methods mostly related to their convergence rate and their efficient implementation.

The methods that we consider were initially proposed by Hestenes [12] and Powell [32], and somewhat later by Haarhoff and Buys [11]. Hestenes gave no interpretation or convergence proof of his method of multipliers, and Powell was motivated by a penalty function viewpoint. The primal-dual interpretation was given later by Luenberger [17], who in addition gave an argument indicating the fast convergence of the method, and by Buys [6], who in his recent thesis provided an extensive and well written analysis of multiplier methods. Buys [6] also proved local convergence of the method of multipliers both for the case of exact and approximate unconstrained minimization under the assumption that the penalty parameter is constant but sufficiently large. A similar convergence result for exact minimization was proved by Rupp [41], [42]. Global convergence results for nonconvex problems were proved recently by the author in [3] and [5]. For quadratic problems with linear constraints, global convergence was also proved by Martensson [19] who in addition proposed some variations on the multiplier method. The method of multipliers has been applied to the solution of some infinite-dimensional problems by Rupp [39], [40], [41]. Some variations of the method of multipliers were proposed by Miele, et al. [20], [21], and Tripathi and Narendra [44]. In these particular variations, the Lagrange multiplier is updated at the end of every gradient step or every conjugate gradient cycle in the unconstrained minimization problem. The convergence properties and the precise motivation for such methods is not as yet quite well understood. They seem to be somewhat related to multiplier methods with asymptotically exact unconstrained minimization, as will be explained later on in this paper. They are also related to the Lagrangian algorithms of Arrow, Hurwicz and Uzawa [2] (particularly the chapter by Arrow and Solow) as applied to the "penalized" problem (5) of the next section. Finally we note that multiplier methods as proposed in the above references are mainly applicable to problems with equality constraints. More recently, considerable attention has been directed towards extension of the method to treat inequality constraints. At the same time, the properties of the method when applied to convex programming problems have been analyzed in detail. In this connection, we mention the excellent papers by Rockafellar [34]–[37], the groundwork for which was laid in his early paper [33], and the work of B. Kort and the author [14]–[16], [4]. Generally speaking, methods of multipliers, as adapted to treat inequality constraints, exhibit similar behavior as for the case of equality constraints. However, for convex programming problems, the methods have some

very attractive properties, namely that they converge globally for any positive value of the penalty parameter [14], [35], [15], [4]. We mention also that there is a very interesting duality theory associated with multiplier methods primarily developed by Rockafellar [33], [34], [36] (see also [1], [3], [15], [18], [28]). Aside from its intrinsic value, this theory can form the basis for the development of efficient large-step Lagrangian methods. For one such algorithm based on Newton's method, see Mangasarian [18]. We note that Lagrangian methods utilizing the penalty idea have been proposed by Fletcher [8], [10], Fletcher and Lill [9], and by Miele and his associates [22], [23]. The precise connection of these methods with methods of multipliers is as yet unclear. Finally we mention that some work related to the method of multipliers has been reported recently in [26] and [43].

The present paper is organized as follows. In the next section we describe the basic method of multipliers in a framework which is suitable for analysis of its convergence rate. Subsequently in § 3 we obtain a useful expression for the rate of convergence of the method. It is shown in particular that as the penalty parameter is increased, the rate of convergence of the dual iteration approaches a superlinear rate. Furthermore, it is shown that we can expect multiplier methods to converge considerably faster than penalty methods which are operated sequentially. In § 4 we consider some efficient variants of the method of multipliers whereby the unconstrained minimizations are only asymptotically exact. We show that such approximate minimizations may lead, in general, to a substanial deterioration of the convergence rate, and we propose a particular approximation scheme which exhibits the same asymptotic convergence rate as the method with exact minimization. In § 5 we compare the step size of the multiplier method with other possible step size rules. We show that for certain problems which are not locally convex, the multiplier method step size is nearly optimal. For locally convex problems, we explain that this is not necessarily true and we propose an alternative step size rule which exhibits an improved convergence rate over the ordinary method. Finally in § 6 we present results of numerical experiments which generally support the conclusions of the theoretical analysis.

2. The method of multipliers. Consider the following constrained minimization problem:

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

where $f: R^n \to R$ is a given twice continuously differentiable function and $h: R^n \to R^m, m \leqq n$, is a given twice continuously differentiable mapping.

Let x^* be an optimal solution of problem (1). We shall assume that x^* satisfies the second order sufficiency conditions for an isolated local minimum, i.e., the matrix $\nabla h(x^*)$ has full rank and there exists a unique Lagrange multiplier (row) vector λ^* such that

$$\nabla l(x^*, \lambda^*) = \nabla f(x^*) + \lambda^* \nabla h(x^*) = 0 \tag{2}$$

and

$$y' L(x^*, \lambda^*) y > 0 \tag{3}$$

for all $y \in R^n$ such that $\nabla h(x^*)y = 0, y \neq 0$. In the above relations, $\nabla l(x^*, \lambda^*)$ and $L(x^*, \lambda^*)$ denote the gradient relative to x and the Hessian matrix relative to x, respectively, of the Lagrangian function

$$l(x, \lambda) = f(x) + \lambda h(x) \tag{4}$$

evaluated at (x^*, λ^*). The $m \times n$ matrix $\nabla h(x)$ denotes the matrix having as rows the gradients $\nabla h_i(x), i = 1, \cdots, m$, and a prime denotes transposition.

It is clear that problem (1) is equivalent to the following problem obtained from problem (1) by adding a penalty term to the objective function:

$$\text{minimize } f(x) + \tfrac{1}{2}c\|h(x)\|^2 \quad \text{subject to} \quad h(x) = 0, \tag{5}$$

where c is a positive scalar.

Consider now the Lagrangian function corresponding to problem (5):

$$l(x, \lambda, c) = f(x) + \lambda h(x) + \tfrac{1}{2}c\|h(x)\|^2, \tag{6}$$

and its Hessian evaluated at (x^*, λ^*):

$$L(x^*, \lambda^*, c) = L(x^*, \lambda^*) + c\nabla h(x^*)'\nabla h(x^*). \tag{7}$$

It follows from (3) that

$$y'L(x^*, \lambda^*, c)y > 0 \quad \forall\, y \in R^n, \quad y \neq 0 \tag{8}$$

if $c \geqq c^* > 0$, where c^* is sufficiently large to guarantee that the matrix $L(x^*, \lambda^*, c^*)$ is positive definite. As a result, for every c with $c \geqq c^*$, problem (5) has locally convex structure according to the definition of [17], and thus we can define for each $c \geqq c^*$ the dual functional

$$g_c(\lambda) = \min_x l(x, \lambda, c).$$

In the above equation, the dual functional $g_c(\lambda)$ is defined in a neighborhood of λ^* and the minimization is understood to be local in a neighborhood of x^*. The implicit function theorem and our assumptions guarantee that such neighborhoods exist for every $c \geqq c^*$. Since, however, in the algorithm which we shall describe, the scalar c may vary from one iteration to the next, it is necessary to provide a uniform definition of the dual functional over neighborhoods which do not depend on c. We shall restrict, however, the scalar c to take values in an interval $[c^*, \bar{c}]$, where $\bar{c}$ is an arbitrarily large constant. For practical purposes, this restriction results in no great loss of generality.

For any element z of a finite-dimensional space with the usual Euclidean norm and for any scalar $s > 0$, we denote by $B(z; s)$ the open ball centered at z and having radius s. We denote by $\bar{B}(z; s)$ the corresponding closed ball. We now have the following proposition.

PROPOSITION 1. *There exist positive scalars ε^* and δ^* such that for all $\lambda \in B(\lambda^*; \delta^*)$ and all $c \in [c^*, \bar{c}]$, the problem*

$$\textit{minimize } l(x, \lambda, c) = f(x) + \lambda h(x) + \tfrac{1}{2}c\|h(x)\|^2 \quad \textit{subject to} \quad x \in B(x^*; \varepsilon^*)$$

has a unique solution $x(\lambda, c)$. Furthermore, for every ε with $0 < \varepsilon \leqq \varepsilon^*$, there exists a δ with $0 < \delta \leqq \delta^*$ such that

$$x(\lambda, c) \in B(x^*; \varepsilon) \quad \forall\, \lambda \in B(\lambda^*; \delta), \quad c \in [c^*, \bar{c}].$$

Proof. The proof is based on a fixed-point argument similar to one used for the proof of the implicit function theorem (see, e.g., [13], [25]).

For each $\lambda \in R^m$ and $c \in [c^*, \bar{c}]$, consider the mapping $Q^{\lambda,c}: R^n \to R^n$ defined by

$$Q^{\lambda,c}(x) = x - [L(x^*, \lambda^*, c)]^{-1} \nabla l(x, \lambda, c),$$

where L, and ∇l denote the Hessian and gradient of the augmented Lagrangian given by (6) and (7). Taking the gradient of $Q^{\lambda,c}$ with respect to x, we have

$$\nabla Q^{\lambda,c}(x) = [L(x^*, \lambda^*, c)]^{-1}[L(x^*, \lambda^*, c) - L(x, \lambda, c)]$$

and

$$\|\nabla Q^{\lambda,c}(x)\| \leqq \|[L(x^*, \lambda^*, c)]^{-1}\| \, \|L(x^*, \lambda^*, c) - L(x, \lambda, c)\|.$$

Now given any $a \in (0, 1)$, there exist an $\varepsilon > 0$ and $\delta > 0$ such that $\|\nabla Q^{\lambda,c}(x)\| \leqq a < 1$ for $x \in \bar{B}(x^*; \varepsilon), \lambda \in \bar{B}(\lambda^*; \delta), c \in [c^*, \bar{c}]$.

On the other hand, we have

$$\begin{aligned} \|Q^{\lambda,c}(x^*) - x^*\| &\leqq \|[L(x^*, \lambda^*, c)]^{-1}\| \, \|\nabla l(x^*, \lambda, c)\| \\ &= \|[L(x^*, \lambda^*, c)]^{-1}\| \, \|\nabla l(x^*, \lambda, c) - \nabla l(x^*, \lambda^*, c)\|, \end{aligned}$$

and by letting δ be sufficiently small, we can assert that

$$\|Q^{\lambda,c}(x^*) - x^*\| \leqq \varepsilon(1 - a) \quad \forall\, c \in [c^*, \bar{c}].$$

Now we have

$$\begin{aligned} \|Q^{\lambda,c}(x) - x^*\| &\leqq \|Q^{\lambda,c}(x^*) - x^*\| + \|Q^{\lambda,c}(x) - Q^{\lambda,c}(x^*)\| \\ &\leqq \varepsilon(1 - a) + \sup_{0 < t < 1} \|\nabla Q^{\lambda,c}[x^* + t(x - x^*)]\| \, \|x - x^*\| \leqq \varepsilon(1 - a) + a\varepsilon = \varepsilon \end{aligned}$$

for all $x \in \bar{B}(x^*, \varepsilon), \lambda \in \bar{B}(x^*; \delta), c \in [c^*, \bar{c}]$.

Thus we have $Q^{\lambda,c}: \bar{B}(x^*; \varepsilon) \to \bar{B}(x^*; \varepsilon)$ and

$$\|\nabla Q^{\lambda,c}(x)\| \leqq a < 1 \quad \text{for each } \lambda \in \bar{B}(x^*; \delta), \quad c \in [c^*, \bar{c}].$$

Hence $Q^{\lambda,c}$ has a unique fixed point $x(\lambda, c)$, i.e.,

$$x(\lambda, c) = Q^{\lambda,c}[x(\lambda, c)] = x(\lambda, c) - [L(x^*, \lambda^*, c)]^{-1} \nabla l[x(\lambda, c), \lambda, c],$$

from which

$$\nabla l[x(\lambda, c), \lambda, c] = 0.$$

If we take, in addition, ε and δ sufficiently small so that $L(x, \lambda, c)$ is positive definite for all $(x, \lambda) \in \bar{B}(x^*; \varepsilon) \times \bar{B}(\lambda^*; \delta)$ and $c \in [c^*, \bar{c}]$, we have that the corresponding $x(\lambda, c)$ is a unique unconstrained minimum of $l(x, \lambda, c)$ within $B(x^*; \varepsilon^*)$ for ε^* sufficiently small, and the result of the proposition follows easily. Q.E.D.

The following corollary is an easy consequence of Proposition 1.

COROLLARY 1.1. *Let L be such that*

$$\|h(x) - h(y)\| \leqq L\|x - y\| \quad \forall\, x, y \in \bar{B}(x^*; \varepsilon^*),$$

and ε^*, δ^* *be as in Proposition* 1. *Then for every* ε *with* $0 < \varepsilon \leqq \varepsilon^*$, *there exists a* δ *with* $0 < \delta \leqq \delta^*$ *such that*

$$x(\lambda, c) \in B(x^*; \varepsilon), \qquad \lambda + ch[x(\lambda, c)] \in B(\lambda^*; \delta + \bar{c}L\varepsilon)$$

for all $\lambda \in B(\lambda^*; \delta), c\,[c^*, \bar{c}]$.

Proof. If δ corresponds to ε as in Proposition 1, then $x(\lambda, c) \in B(x^*; \varepsilon)$ and

$$\begin{aligned}\|\lambda + ch(x(\lambda, c)] - \lambda^*\| &\leqq \|\lambda - \lambda^*\| + \bar{c}\|h[x(\lambda, c)]\| \\ &= \|\lambda - \lambda^*\| + \bar{c}\|h[x(\lambda, c)] - h(x^*)\| < \delta + \bar{c}L\varepsilon. \quad \text{Q.E.D.}\end{aligned}$$

Proposition 1 essentially says that by locally minimizing the augmented Lagrangian, one obtains points which are arbitrarily and uniformly close to x^* provided λ is sufficiently close to λ^*. Furthermore, the proposition provides a means for defining the dual functional over a domain which is common for all $c \in [c^*, \bar{c}]$. We define, for all $\lambda \in B(\lambda^*; \delta^*)$ and all $c \in [c^*, \bar{c}]$, the dual functional as

$$g_c(\lambda) = \min_{x \in B(x^*; \varepsilon^*)} \{f(x) + \lambda h(x) + \tfrac{1}{2}c\|h(x)\|^2\}, \tag{9}$$

where the minimum over the open ball $B(x^*; \varepsilon^*)$ is attained by Proposition 1. It can be easily shown (see also [6], [17]) that the scalars ε^* and δ^* in Proposition 1 and Corollary 1.1 can be chosen so that the dual functional $g_c(\lambda)$ is twice continuously differentiable and concave in $B(\lambda^*; \delta^*)$ for all $c \in [c^*, \bar{c}]$. We shall assume that ε^* and δ^* have been so chosen. The gradient ∇g_c and Hessian matrix G_c are given by

$$\nabla g_c(\lambda) = h[x(\lambda, c)]', \tag{10}$$

$$G_c(\lambda) = -\nabla h[x(\lambda, c)]\{L[x(\lambda, c), \lambda, c]\}^{-1}\nabla h[x(\lambda, c)]'. \tag{11}$$

Furthermore, the dual functionals $g_c(\lambda), c \in [c^*, \bar{c}]$ have a common maximizing point, the Lagrange multiplier λ^*, and a common optimal value f^* which is equal to the optimal value $f(x^*)$ of problem (1):

$$g_c(\lambda^*) = \max_{\lambda} g_c(\lambda) = f^* \quad \forall\, c \in [c^*, \bar{c}].$$

The method of multipliers is simply a gradient method for maximizing the dual functional by means of the iteration

$$\lambda_{k+1} = \lambda_k + c\nabla g_c(\lambda_k). \tag{12}$$

The gradient $\nabla g_c(\lambda_k)$ is given by (10), where $x(\lambda_k, c)$ is an unconstrained minimum (within $B(x^*; \varepsilon^*)$) of the augmented Lagrangian

$$l(x, \lambda_k, c) = f(x) + \lambda_k h(x) + \tfrac{1}{2}c\|h(x)\|^2. \tag{13}$$

The iteration (12) is a fixed step size gradient method for solving the dual problem which can be shown [6], [41] to converge to λ^* provided the constant c is sufficiently large. This fact will also be proved in the next section in a more general setting where c may vary from one iteration to the next. It should be noted that in order for the method to converge, it is not necessary that the initial Lagrange multiplier estimate λ_0 is in $B(\lambda^*; \delta^*)$. Since the method can also be viewed as a penalty function method, it can be shown [3] that if the initial penalty parameter c is sufficiently large and the corresponding minimization problem yields a solution close to x^*, then the next point λ_1 will be arbitrarily close to λ^*. Thus in the initial

iterations, the penalty nature of the method is dominant and provides points sufficiently close to λ^*, and in subsequent iterations, the gradient nature of the algorithm becomes more pronounced.

It is important to realize that it is not necessary to keep the penalty parameter c fixed during the computation. Each constant c defines a dual functional $g_c(\lambda)$ via (9). The collection of all these dual functionals has the same local maximum λ^*. Thus when a different c (say c_k) is used at the kth unconstrained minimization,

$$\text{minimize } f(x) + \lambda_k h(x) + \tfrac{1}{2}c_k\|h(x)\|^2, \tag{14}$$

the iteration

$$\lambda_{k+1} = \lambda_k + c_k h[x(\lambda_k, c_k)]' \tag{15}$$

can be viewed as a gradient step for maximizing the corresponding dual functional $g_{c_k}(\lambda)$, which attains its maximum at λ^*. Furthermore, it is possible to let the sequence c_k increase to infinity. While the intermediate unconstrained minimization problems become increasingly ill-conditioned, the dual iteration (15) has increasingly faster convergence rate, as will be shown in the next section, and on balance, the method performs well. A reasonable method to update c suggested by Powell [32] and Buys [6], is to multiply c by a constant greater than one (say 5–10) at the end of each unconstrained miminization for which the resulting constraint violation as measured by $\|h(x)\|$ is not decreased by a certain factor. An alternative method for updating c has been suggested by Miele, et al. [20] in a somewhat different setting.

The method of multipliers can be easily extended to handle inequality constraints. As shown by Rockafellar [33]–[36], one may use slack variables to convert inequality constraints into equality constraints. However, the minimization with respect to the slack variables can be carried out explicitly, and as a result, the dimension of the unconstrained minimization problem is not increased. We do not further discuss inequality constraints in this paper, and we refer to [3], [15] and [16] for a discussion of the related rate of convergence aspects. Among other things, one may show that the approximate Lagrange multipliers corresponding to inactive constraints converge to zero in a finite number of steps. As a result, inactive constraints do not enter in any rate of convergence estimates, and the results of this paper under a strict complementarity assumption carry over to the inequality case in a straightforward manner.

We mention finally that the method of multipliers has an economic interpretation similar to the one given by Arrow and Solow [2] for their combined Lagrangian and penalty method. In this interpretation, the iterations of λ_k are viewed as market price adjustments to excess demand or supply, and the iterations of x_k are viewed as production vector changes in response to extrapolated market price changes.

3. Convergence rate of the method of multipliers. As mentioned in the previous section, the method of multipliers can be viewed as a gradient method for solving the dual problem. Thus one can obtain its convergence rate by using a corresponding result on gradient methods (see, e.g., [29], [30]). This result, however, is rather

uninformative, since it involves the eigenvalues of the Hessian $G_c(\lambda)$, which strongly depend on c. The following proposition is obtained by a modification of this result and provides an expression for the convergence rate which is more amenable to proper interpretation.

Let us consider the matrix

$$D(x, \lambda) = -\nabla h(x)[L(x, \lambda)]^{-1}\nabla h(x)', \tag{16}$$

where $L(x, \lambda)$ is the Hessian relative to x of the Lagrangian (4). Notice that $D(x^*, \lambda^*)$ would be the Hessian at λ^* of the ordinary dual functional g_0 if the problem had a locally convex structure [17]. Assume that $D(x, \lambda)$ is defined and is invertible in a set $\bar{B}(x^*; \varepsilon) \times \bar{B}(\lambda^*; \delta + \bar{c}L\varepsilon)$, where ε and δ are positive scalars such that $x(\lambda, c) \in B(x^*; \varepsilon)$, $\lambda + ch[x(\lambda, c)] \in B(\lambda^*; \delta + \bar{c}L\varepsilon)$ for all $\lambda \in B(\lambda^*; \delta)$ and all $c \in [c^*, \bar{c}]$ in accordance with Proposition 1 and Corollary 1.1. Assume also that the algorithm of (14), (15) yields a sequence of vectors (x_k, λ_k) converging to (x^*, λ^*) and that after some index $\bar{k}$, the vectors (x_k, λ_k) are contained in $B(x^*; \varepsilon) \times B(\lambda^*; \delta)$. Then we have the following proposition.

PROPOSITION 2. *Under the preceding assumptions, we have*

$$\|\lambda_{k+1} - \lambda^*\| \leqq r_k\|\lambda_k - \lambda^*\| \quad \forall\, k \geqq \bar{k}, \tag{17}$$

with

$$r_k = \max_{\substack{(x,\lambda)\in\bar{B}(x^*;\varepsilon)\times\bar{B}(\lambda^*;\delta+\bar{c}L\varepsilon) \\ i=1,\cdots,m}} \left|\frac{1}{1 - c_k e_i[D(x, \lambda)]}\right|, \tag{18}$$

where $e_i[D(x, \lambda)]$ denotes the i-th eigenvalue of $D(x, \lambda)$.

Proof. Consider the Hessian matrix (11). We have

$$G_c(\lambda) = -\nabla h[x(\lambda, c)]\{L[x(\lambda, c), \bar{\lambda}] + c\nabla h[x(\lambda, c)]'\nabla h[x(\lambda, c)]\}^{-1}\nabla h[x(\lambda, c)]',$$

where $\bar{\lambda} = \lambda + ch[x(\lambda, c)]'$. From a well-known matrix identity, we have

$$[I - cD[x(\lambda, c), \bar{\lambda}]]^{-1} = I + cG_c(\lambda),$$

and hence for the corresponding eigenvalues of $G_c(\lambda)$ and $D[x(\lambda, c), \bar{\lambda}]$, we have

$$\frac{1}{1 - ce_i[D[x(\lambda, c), \bar{\lambda}]]} = 1 + ce_i[G_c(\lambda)]. \tag{19}$$

Now by using the iteration (15), we have

$$\|\lambda_{k+1} - \lambda^*\| = \|\lambda_k - \lambda^* + c_k h(x_k)'\| = \left\|\lambda_k - \lambda^* + c_k \int_0^1 (\lambda_k - \lambda^*)G_{c_k}(\lambda)\, dt\right\|,$$

where $\lambda = \lambda^* + t(\lambda_k - \lambda^*)$. Hence

$$\|\lambda_{k+1} - \lambda^*\| \leqq \|\lambda_k - \lambda^*\| \left\|\int_0^1 [I + c_k G_{c_k}(\lambda)]\, dt\right\|$$

$$\leqq \|\lambda_k - \lambda^*\| \max_{\substack{\lambda=\lambda^*+t(\lambda_k-\lambda^*) \\ t\in[0,1] \\ i=1,\cdots,m}} |1 + c_k e_i[G_{c_k}(\lambda)]|.$$

By using (19) and Corollary 1.1, it follows that

$$\|\lambda_{k+1} - \lambda^*\| \leqq \|\lambda_k - \lambda^*\| \max_{\substack{(x,\lambda)\in \bar{B}(x;\varepsilon)\times \bar{B}(\lambda^*;\delta+\bar{c}L\varepsilon) \\ i=1,\cdots,m}} \left| \frac{1}{1 - c_k e_i[D(x,\lambda)]} \right|$$

$$= r_k \|\lambda_k - \lambda^*\|. \qquad \text{Q.E.D.}$$

Some important observations can be made from the result of Proposition 2. First of all, a trivial modification of its proof yields the following local convergence result.

COROLLARY 2.1. *Let ε and δ be positive scalars such that $x(\lambda, c) \in B(x^*; \varepsilon)$ and $\lambda + ch[x(\lambda, c)]' \in B(\lambda^*; \delta + \bar{c}L\varepsilon)$ for all $\lambda \in B(\lambda^*; \delta)$ and $c \in [c^*, \bar{c}]$ in accordance with Proposition* 1 *and Corollary* 1.1. *Assume that ε and δ are sufficiently small and c_k is sufficiently large so that for some constant μ,*

$$c_k \geqq \mu > \max\left\{0, \frac{2}{e_i[D(x,\lambda)]}\right\} \quad \forall k > 0 \tag{20}$$

for all eigenvalues $e_i[D(x, \lambda)]$ of $D(x, \lambda)$ over $\bar{B}(x^*; \varepsilon) \times \bar{B}(\lambda^*; \delta + \bar{c}L\varepsilon)$. Assume also that $\lambda_0 \in B(\lambda^*; \delta)$. Then the sequence $\{\lambda_k\}$ generated by the iteration (15) *remains in $B(\lambda^*; \delta)$ and converges to λ^*.*

Proof. By exact repetition of the argument in the proof of Proposition 1, we have $\|\lambda_1 - \lambda^*\| \leqq r_0 \|\lambda_0 - \lambda^*\|$, where r_0 is given by (18). By our assumption (20),

$$r_0 < \max \left| \frac{1}{1 - \mu e_i[D(x,\lambda)]} \right| = \rho < 1.$$

Hence $\|\lambda_1 - \lambda^*\| \leqq \rho \|\lambda_0 - \lambda^*\|$ and $\lambda_1 \in B(\lambda^*; \delta)$. Proceeding similarly, we prove for all k that $\|\lambda_k - \lambda^*\| \leqq \rho^k \|\lambda_0 - \lambda^*\|$, and the result follows immediately. Q.E.D.

The most important observation from Proposition 2 is that the sequence $\|\lambda_k - \lambda^*\|$ converges linearly with stepwise convergence ratio r_k. Furthermore, r_k decreases to zero as c_k is increased. Thus a superlinear rate is approached as c_k tends to infinity. This is consistent with the argument of Luenberger [17], who observed that as c_k increases, the gradient iteration (15) approaches a Newton step for solving the dual. If the sequence c_k converges to a finite value c, then we have

$$\limsup_{k\to\infty} \frac{\|\lambda_{k+1} - \lambda^*\|}{\|\lambda_k - \lambda^*\|} \leqq \max_i \left| \frac{1}{1 - c e_i[D(x^*, \lambda^*)]} \right| = \bar{r}. \tag{21}$$

It can easily be shown that at least for quadratic problems with linear constraints and a constant sequence c_k, $\bar{r}$ is a sharp bound in the sense that there exist starting points λ_0 for which (21) holds with equality.

It is of interest to compare the convergence rate of the multiplier method with the convergence rate of penalty function methods which are based on sequential unconstrained minimization of the function

$$f(x) + \tfrac{1}{2} c_k \|h(x)\|^2 \tag{22}$$

for a sequence $c_k \to \infty$. If the sequence $\{x_k\}$ of minimizing points of (22) converges to the point x^*, then the sequence $\{\lambda_k\}$, where $\lambda_k = c_k h(x_k)'$, converges to λ^*. It has

been shown [31], [24] that such penalty function methods generally exhibit a convergence rate governed by the relation

$$\|\lambda_k - \lambda^*\| \leqq q/c_k \quad \forall\, k > \bar{k}, \tag{23}$$

where $\bar{k}$ is some index and q is a constant depending on the problem. By comparing (17), (18) and (23), it can be seen that the sequence $\{\lambda_k\}$ can be expected to converge considerably faster in the multiplier method than in the quadratic penalty function method. This fact has been substantiated by numerical experiments. Given that the two methods involve a comparable amount of computation at each unconstrained minimization and share the advantage of simplicity, it appears that multiplier methods should be generally considered preferable to penalty function methods. For further elaboration on the comparison between penalty methods and multiplier methods we refer to [3] and [5].

4. Efficient implementations of the multiplier method. The multiplier method described in the previous section has the drawback that the unconstrained minimization of the augmented Lagrangian must be carried out exactly in order to update the Lagrange multiplier via the gradient iteration (15). This requires an unreasonably high amount of computation for the unconstrained minimizations. It appears that a more efficient scheme results if only moderate accuracy is demanded in the initial minimizations, and the accuracy is increased at later iterations. Such a procedure has been suggested by Buys [6] in a similar vein as in corresponding penalty function methods [24], [27], [31]. In this procedure, the minimization process in the problem

$$\text{minimize } l(x, \lambda_k, c_k) = f(x) + \lambda_k h(x) + \tfrac{1}{2} c_k \|h(x)\|^2 \tag{24}$$

is terminated at a point x_k such that

$$\|\nabla l(x_k, \lambda_k, c_k)\| \leqq \varepsilon_k, \tag{25}$$

where $\{\varepsilon_k\}$ is a preselected decreasing sequence tending to zero. The corresponding dual iteration can take several alternate forms. One possibility is to use the iteration of the previous section

$$\lambda_{k+1} = \lambda_k + c_k h(x_k)'. \tag{26}$$

Other possible methods of updating include the iteration

$$\lambda_{k+1} = \lambda_k + \beta_k h(x_k)', \tag{27}$$

where

$$\beta_k = c_k - \frac{h(x_k)'\nabla h(x_k)\nabla l(x_k, \lambda_k, c_k)}{h(x_k)'\nabla h(x_k)\nabla h(x_k)' h(x_k)}, \tag{28}$$

proposed by Miele, et al. [18] in a somewhat different setting, and the iteration

$$\lambda_{k+1} = -\nabla f(x_k)\nabla h(x_k)'[\nabla h(x_k)\nabla h(x_k)']^{-1} \tag{29}$$

suggested by Haarhoff and Buys [11], Buys [6], and Miele, et al. [22].

One way of justifying the iteration (27) is by observing that β_k as given by (28) minimizes the quantity $\|\nabla l(x_k, \lambda_k, \beta)\|$ over β [22]. Hence, lacking further

information, the vector $h(x_k)'$ can be considered as a more accurate approximation to the gradient $\nabla g_{\beta_k}(\lambda_k)$ of the dual functional $g_{\beta_k}(\lambda)$ than to the gradient $\nabla g_{c_k}(\lambda_k)$. A similar interpretation can be given for the iteration (29). It should be mentioned that both iterations (27) and (29) reduce to the basic iteration (26) if the unconstrained minimization (24) is carried out exactly.

First let us consider the algorithm with the termination criterion (25) and the updating rule (26) (call it Algorithm A1). Let us consider again the matrix

$$D(x, \lambda) = -\nabla h(x)[L(x, \lambda)]^{-1}\nabla h(x)' \tag{30}$$

defined over $\bar{B}(x^*; \varepsilon) \times \bar{B}(\lambda^*; \delta + \bar{c}L\varepsilon)$, where ε and δ are as in Proposition 1, and assume that the algorithm generates a sequence (x_k, λ_k) converging to (x^*, λ^*) and that after some index $\bar{k}$, the vectors (x_k, λ_k) are contained in $B(x^*; \varepsilon) \times B(\lambda^*; \delta)$. By Proposition 1, the exact minimizing point $x(\lambda_k, c_k)$ of $l(x, \lambda_k, c_k)$ belongs to $B(x^*; \varepsilon)$.

Let $L > 0$ be as in Corollary 1.1, i.e.,

$$\|h(x) - h(y)\| \leqq L\|x - y\| \quad \forall\, x, y \in B(x^*; \varepsilon^*), \tag{31}$$

and let M denote the minimum of the eigenvalues of the Hessian $L(x, \lambda, c)$ for $(x, \lambda) \in \bar{B}(x^*; \varepsilon) \times \bar{B}(\lambda^*; \delta), c \in [c^*, \bar{c}]$, i.e.,

$$M = \min_{\substack{(x,\lambda)\in\bar{B}(x^*;\varepsilon)\times\bar{B}(\lambda^*;\delta) \\ c\in[c^*,\bar{c}] \\ i=1,\cdots,n}} e_i[L(x, \lambda, c)]. \tag{32}$$

We assume that ε and δ are sufficiently small to guarantee that $M > 0$. We have the following proposition.

PROPOSITION 3. *Under the preceding assumptions, we have, for Algorithm* A1,

$$\|\lambda_{k+1} - \lambda^*\| \leqq r_k\|\lambda_k - \lambda^*\| + \varepsilon_k c_k(L/M) \quad \forall\, k \geqq \bar{k}, \tag{33}$$

where

$$r_k = \max_{(x,\lambda)\in\bar{B}(x^*;\varepsilon)\times\bar{B}(\lambda^*;\delta+\bar{c}L\varepsilon)} \left|\frac{1}{1 - c_k e_i[D(x, \lambda)]}\right|. \tag{34}$$

Proof. We have, by using (26),

$$\begin{aligned}\|\lambda_{k+1} - \lambda^*\| &= \|\lambda_k - \lambda^* + c_k h(x_k)'\| \\ &\leqq \|\lambda_k - \lambda^* + c_k h[x(\lambda_k, c_k)]'\| + c_k\|h(x_k) - h[x(\lambda_k, c_k)]\|.\end{aligned}$$

Now, similarly as in Proposition 1, we have

$$\|\lambda_k - \lambda^* + c_k h[x(\lambda_k, c_k)]'\| \leqq r_k\|\lambda_k - \lambda^*\|.$$

On the other hand, we have

$$\begin{aligned}\|h(x_k) - h[x(\lambda_k, c_k)]\| &\leqq L\|x_k - x(\lambda_k, c_k)\| \\ &\leqq \frac{L}{M}\|\nabla l(x_k, \lambda_k, c_k) - \nabla l[x(\lambda_k, c_k), \lambda_k, c_k]\| = \frac{L}{M}\|\nabla l(x_k, \lambda_k, c_k)\| \leqq \frac{L\varepsilon_k}{M}\end{aligned}$$

and the result follows. Q.E.D.

The result of Proposition 3 indicates that the convergence rate of Algorithm A1 may be different from the convergence rate of the multiplier method of the previous section. Indeed, if the sequence ε_k does not converge as fast as $\|\lambda_k - \lambda^*\|$, the convergence of the sequence $\|\lambda_k - \lambda^*\|$ may not even be linear. To illustrate this fact consider the following example.

Example. Consider the problem

$$\text{minimize } \tfrac{1}{2}x^2 \quad \text{subject to} \quad x = 0.$$

Take $c = 1$, and let the accuracy of the unconstrained minimization be determined by

$$\|\nabla l(x, \lambda_k, 1)\| \leqq \varepsilon_k = \frac{k-1}{k(k+1)}, \qquad k \geqq 2,$$

where the augmented Lagrangian $l(x, \lambda, 1)$ is given by

$$l(x, \lambda, 1) = \tfrac{1}{2}x^2 + \lambda x + \tfrac{1}{2}x^2.$$

Then by direct computation, it can be seen that a possible sequence $\{\lambda_k\}$ generated by the algorithm is the sequence

$$\lambda_k = 1/k, \qquad k \geqq 2,$$

if the starting point is $\lambda_2 = \frac{1}{2}$. Since $\lambda^* = 0$ for this problem, we have

$$|\lambda_{k+1} - \lambda^*|/|\lambda_k - \lambda^*| = k/k + 1, \qquad k \geqq 2,$$

showing that the convergence of the sequence $\{\lambda_k\}$ is not linear.

In order to preserve the convergence rate of the multiplier method, it is necessary to use an approximation scheme which guarantees that the minimization is sufficiently accurate at least when we are close to the solution. Such a scheme is obtained by using, instead of the termination criterion (25), the following termination criterion:

$$\|\nabla l(x_k, \lambda_k, c)\| \leqq \eta_k \|h(x_k)\|, \tag{35}$$

where $\{\eta_k\}$ is a decreasing sequence converging to zero. We shall call the algorithm resulting from use of the criterion (35) and the dual iteration (26) Algorithm A2. We can now prove the following proposition, under the assumptions of Proposition 3 and the additional assumption that $M - \eta_k L > 0$ for all $k \geqq \bar{k}$.

PROPOSITION 4. *Under the preceding assumptions, we have, for Algorithm* A2,

$$\|\lambda_{k+1} - \lambda^*\| \leqq \left(r_k + \frac{\eta_k L}{M - \eta_k L} p_k\right) \|\lambda_k - \lambda^*\| \quad \forall\, k \geqq \bar{k}, \tag{36}$$

where

$$r_k = \max_{\substack{(x,\lambda) \in \bar{B}(x^*;\varepsilon) \times \bar{B}(\lambda^*;\delta + \bar{c}L\varepsilon) \\ i = 1, \cdots, m}} \left| \frac{1}{1 - c_k e_i[D(x, \lambda)]} \right|, \tag{37}$$

$$p_k = \max_{\substack{(x,\lambda) \in \bar{B}(x^*;\varepsilon) \times \bar{B}(\lambda^*;\delta + \bar{c}L\varepsilon) \\ i = 1, \cdots, m}} \left| \frac{c_k e_i[D(x, \lambda)]}{1 - c_k e_i[D(x, \lambda)]} \right|. \tag{38}$$

Proof. We have

$$\begin{aligned}\|\lambda_{k+1} - \lambda^*\| &= \|\lambda_k - \lambda^* + c_k h(x_k)'\| \\ &\leqq \|\lambda_k - \lambda^* + c_k h[x(\lambda_k, c_k)]'\| + c_k \|h(x_k) - h[x(\lambda_k, c_k)]\|.\end{aligned} \tag{39}$$

Similarly as in Propositions 2 and 3, we have

$$\|\lambda_k - \lambda^* + c_k h[x(\lambda_k, c_k)]'\| \leqq r_k \|\lambda_k - \lambda^*\|. \tag{40}$$

Also we have

$$\begin{aligned}\|h(x_k) - h[x(\lambda_k, c_k)]\| &\leqq L\|x_k - x(\lambda_k, c_k)\| \\ &\leqq \frac{L}{M}\|\nabla l(x_k, \lambda_k, c_k)\| \leqq \frac{\eta_k L}{M}\|h(x_k)\| \\ &\leqq \frac{\eta_k L}{M}(\|h(x_k) - h[x(\lambda_k, c_k)]\| + \|h[x(\lambda_k, c_k)]\|),\end{aligned}$$

from which

$$c_k \|h(x_k) - h[x(\lambda_k, c_k)]\| \leqq \frac{\eta_k L c_k}{M - \eta_k L}\|h[x(\lambda_k, c_k)]\|. \tag{41}$$

Since $h[x(\lambda_k, c_k)]$ is the gradient of the dual functional $g_{c_k}(\lambda)$ at λ_k and $\nabla g_{c_k}(\lambda^*) = 0$, we have

$$c_k \|h[x(\lambda_k, c_k)]\| \leqq c_k \max_{\substack{\lambda \in \bar{B}(\lambda^*; \delta) \\ i = 1, \cdots, m}} |e_i[G_{c_k}(\lambda)]| \, \|\lambda_k - \lambda^*\|,$$

and by using (19) and (38),

$$c_k \|h[x(\lambda_k, c_k)]\| \leqq p_k \|\lambda_k - \lambda^*\|. \tag{42}$$

By combining now (39), (40), (41) and (42), the result follows. Q.E.D.

It is to be noted that p_k is bounded and tends to unity as c_k increases, so that for large c_k and small η_k, (36) becomes approximately

$$\|\lambda_{k+1} - \lambda^*\| \leqq (r_k + \eta_k L/M)\|\lambda_k - \lambda^*\|.$$

A comparison of Propositions 2 and 4 reveals now that Algorithm A2 has identical asymptotic convergence ratio with the method of multipliers.

It is easy to see that when the updating rule (27) is used instead of (26), the estimate of Proposition 4 becomes

$$\|\lambda_{k+1} - \lambda^*\| \leqq \left(\tilde{r}_k + \frac{\eta_k L}{M - \eta_k L}\tilde{p}_k\right)\|\lambda_k - \lambda^*\|, \tag{43}$$

where

$$\tilde{r}_k = \max_{\substack{(x,\lambda) \in \bar{B}(x^*; \varepsilon) \times \bar{B}(\lambda^*; \delta + \bar{c}L\varepsilon) \\ i = 1, \cdots, m}} \left|\frac{1}{1 - \beta_k e_i[D(x, \lambda)]}\right|,$$

$$\tilde{p}_k = \max_{\substack{(x,\lambda) \in \bar{B}(x^*; \varepsilon) \times \bar{B}(\lambda^*; \delta + \bar{c}L\varepsilon) \\ i = 1, \cdots, m}} \left|\frac{\beta_k e_i[D(x, \lambda)]}{1 - \beta_k e_i[D(x, \lambda)]}\right|.$$

When the sequence c_k converges to a finite value c, then in view of (28) and (35), we have $\beta_k \to c$, and the relation (42) yields

$$\limsup_{k\to\infty} \frac{\|\lambda_{k+1} - \lambda^*\|}{\|\lambda_k - \lambda^*\|} \leqq \bar{r},$$

where

$$\bar{r} = \max_{i=1,\cdots,m} \left| \frac{1}{1 - ce_i[D(x^*, \lambda^*)]} \right|.$$

In other words, Algorithm A2 with the iteration (27), (28) instead of (26) has the same asymptotic convergence ratio as the multiplier method of the previous section, which requires exact unconstrained minimization.

It should be noted that all the results of this and the previous section can be generalized for the more general algorithm which involves the (exact or approximate) unconstrained minimization of

$$l(x, \lambda_k, M_k) = f(x) + \lambda_k h(x) + \tfrac{1}{2} h(x)' M_k h(x),$$

where M_k are positive definite symmetric matrices. The dual iteration becomes, in this case, $\lambda_{k+1} = \lambda_k + h(x_k)' M_k$. When $M_k = c_k I$, we obtain the method of multipliers discussed earlier. The use of the matrices M_k has the effect of scaling the constraint equation. If $M_k \to M$, the convergence rate (21) becomes

$$\limsup_{k\to\infty} \frac{\|M_k^{-1/2}(\lambda_{k+1} - \lambda^*)'\|}{\|M_k^{-1/2}(\lambda_k - \lambda^*)'\|} \leqq \max_i \left| \frac{1}{1 - e_i[M^{1/2} D(x^*, \lambda^*) M^{1/2}]} \right|,$$

and similar results as those of Propositions 2, 3 and 4 can be obtained. The updating rule (29) can be justified in the context of this more general algorithm in that λ_{k+1} as given by (29) can be written as $\lambda_{k+1} = \lambda_k + h(x_k)' M_k$ for some diagonal matrix M_k, and furthermore, M_k minimizes $\|\nabla l(x_k, \lambda_k, M)\|$ over all diagonal matrices M [22].

We finally mention that it is easy to establish local convergence results, similar to Corollary 2.1, for Algorithms A1 and A2 by making use of the arguments in the proofs of Propositions 3 and 4. The additional assumption required is that the sequences $\{\varepsilon_k\}$ and $\{\eta_k\}$ are bounded above by sufficiently small positive numbers. This assumption is necessary in order to guarantee that the generated sequence $\{\lambda_k\}$ satisfies $\|\lambda_k - \lambda^*\| \leqq \|\lambda_0 - \lambda^*\| < \delta$ and hence the sequence $\{\lambda_k\}$ remains in the neighborhood $B(\lambda^*; \delta)$. Stronger global convergence results for algorithms similar to A1 and A2 have been obtained recently in [3] and [5].

5. Alternative step size choices for the method of multipliers. As mentioned in § 2, the method of multipliers can be interpreted as a fixed step size gradient method of maximizing the "penalized" dual functional $g_c(\lambda)$. It is well known [29] that in such gradient methods, the choice of step size parameter is crucial both in terms of the convergence and in terms of the rate of convergence of the method. It is a rather remarkable fact that the particular step size parameter c which is used in the multiplier method works so well from the point of view of both convergence and rate of convergence. Nonetheless, it is of interest to try to compare the step

size c with other possible step sizes and in particular with the optimal step size. This is what we attempt to do in this section. As it turns out for certain problems which are not locally convex, the step size c is close to the optimal and can hardly be improved upon. However, for the locally convex case (e.g., a convex programming problem), the analysis indicates the possibility of a significant improvement by modification of the step size. In what follows, we suggest a modified step size rule which has worked well in numerical experiments.

In order to simplify the analysis, we initially restrict attention to the case in which the objective function f is quadratic (with not necessarily positive definite or even semidefinite Hessian matrix) and the constraint functions h_i are linear. Since we shall be using results which have been proved in generality [29], it is a routine matter to extend our analysis to the general case.

Consider a multiplier method where c is held fixed for the purpose of unconstrained minimization. The step size now, however, is taken to be α rather than c, i.e., the iteration

$$\lambda_{k+1} = \lambda_k + \alpha \nabla g_c(\lambda_k) \quad \forall k \tag{44}$$

is used. Then by [29, Thm. 6], the iteration above converges for

$$0 < \alpha < 2/E_c, \tag{45}$$

where E_c is the largest eigenvalue of the negative Hessian $-G_c(\lambda^*)$, provided that $G_c(\lambda^*)$ is a negative definite matrix. Furthermore, the rate of convergence is linear and governed by

$$\frac{\|\lambda_{k+1} - \lambda^*\|}{\|\lambda_k - \lambda^*\|} \leqq r(\alpha) \quad \forall k, \tag{46}$$

where

$$r(\alpha) = \max \{|1 - \alpha E_c|, |1 - \alpha e_c|\}, \tag{47}$$

with E_c and e_c denoting the largest and smallest eigenvalues of $-G_c(\lambda^*)$. The optimal convergence ratio is attained for the step size α^* minimizing $r(\alpha)$ over α

$$\alpha^* = 2/(E_c + e_c) \tag{48}$$

and is given by

$$r(\alpha^*) = \frac{E_c - e_c}{E_c + e_c}. \tag{49}$$

In general, it is quite difficult to find the optimal step size, since this requires knowledge of the eigenvalues E_c and e_c.

Now by equation (19), we have

$$e_i[-G_c(\lambda^*)] = \frac{1}{\dfrac{1}{e_i[-D(x^*, \lambda^*)]} + c}$$

for all eigenvalues of the matrices $-G_c(\lambda^*)$ and $-D(x^*, \lambda^*)$. Let E_0 and e_0 denote

the eigenvalues of $-D(x^*, \lambda^*)$ corresponding to E_c and e_c in accordance with the above relation:

$$E_c = \frac{1}{(1/E_0) + c}, \qquad e_c = \frac{1}{(1/e_0) + c}. \tag{50}$$

If $-D(x^*, \lambda^*)$ is positive definite, then E_0 and e_0 are its largest and smallest eigenvalues. If, however, $-D(x^*, \lambda^*)$ is neither positive nor negative definite, then E_0 is its largest negative eigenvalue, e_0 is its smallest positive eigenvalue, and $E_0 < 0 < e_0$. In view of (45) and (50), we have that convergence occurs for all step sizes α satisfying

$$0 < \alpha < (2/E_0) + 2c. \tag{51}$$

It follows that for c much larger than $1/E_0$, the step size $\alpha = c$ of the method of multipliers is approximately in the middle of the interval of convergence, a fact which explains, to some extent, the excellent numerical stability of the method. It may also be observed that as $c \to \infty$, we have $E_c \to 1/c$, $e_c \to 1/c$, $G_c(x^*) \to -(1/c)I$, and the multiplier method iteration approaches a Newton step as noted by Luenberger [17].

The convergence ratio corresponding to the step size c is given by (cf. (47), (50))

$$r(c) = \max\left\{\left|\frac{1}{1 + cE_0}\right|, \left|\frac{1}{1 + ce_0}\right|\right\}. \tag{52}$$

By (48), (49) and (50), the convergence ratio corresponding to the optimal step size

$$\alpha^* = \frac{2(1 + cE_0)(1 + ce_0)}{E_0 + e_0 + 2cE_0e_0} \tag{53}$$

is given by

$$r(\alpha^*) = \frac{E_0 - e_0}{E_0 + e_0 + 2cE_0e_0}. \tag{54}$$

We now distinguish two cases of interest.

Case (a) ($E_0 < 0 < e_0$). Here we assume that the matrix $D(x^*, \lambda^*)$ is neither positive semidefinite nor negative semidefinite. In this case, by (51) we must have $(-1/E_0) < c$ in order to guarantee local convexity ($0 < e_c$), in which case there exist some step sizes α which achieve convergence ($r(\alpha) < 1$). However, the particular step size $\alpha = c$ guarantees convergence only if $(-2/E_0) < c$, in which case $r(c) < 1$ (cf. Corollary 2.1). For values of c close to $-2/E_0$, equation (52) shows that the convergence ratio $r(c)$ is poor (close to one). However, as c increases, not only does the convergence ratio $r(c)$ improve, but also the ratio $r(c)/(r(\alpha^*))$ decreases, and in fact from (52) and (54),

$$\lim_{c\to\infty} \frac{r(c)}{r(\alpha^*)} = \max\left\{\frac{2e_0}{e_0 - E_0}, \frac{2E_0}{E_0 - e_0}\right\} < 2.$$

Furthermore, it may be shown by direct calculation from (52) and (54) that if $c > (e_0 - 3E_0)/2E_0^2$, then $r(c)/(r(\alpha^*)) < 2$.

Thus for the case $E_0 < 0 < e_0$, not only is the convergence ratio $r(c)$ small for large c, but also $r(c)$ is close to being optimal and can be improved only by a factor of at most 2 by optimal step size choice. Given that $r(c)$ is already low for large c, it appears that for $E_0 < 0 < e_0$, there is rather little room for improvement of the performance of the multiplier method by alternative step size choice. This is particularly so since there are no simple ways for finding or approximating the optimal step size without explicit knowledge of the eigenvalues E_c, e_c of $-G_c$.

Case (b) $(0 < e_0 \leqq E_0)$. This is the locally convex case, which includes convex programming problems. For this case, the ordinary dual functional

$$g_0(\lambda) = \min_x \{f(x) + \lambda h(x)\}$$

is well-defined as a concave quadratic function. For any given $c > 0$, any step size α with $0 < \alpha < 2c$ satisfies $r(\alpha) < 1$ by (47) and (51), and hence achieves convergence. However, by direct calculation from (52) and (53), we have (assuming $e_0 \neq E_0$)

$$\frac{r(c)}{r(\alpha^*)} = \frac{1 + (e_0/E_0) + 2ce_0}{1 + ce_0} \cdot \frac{1}{1 - e_0/E_0}$$

and

$$\lim_{c \to \infty} \frac{r(c)}{r(\alpha^*)} = \frac{2}{1 - e_0/E_0} > 2.$$

The relations above show that, contrary to the previous case, there may be a substantial improvement of the convergence ratio if the optimal step size α^* can be found or approximated. The potential gain is increased as e_0 is close to E_0, i.e., the ordinary dual problem is well-conditioned.

While the exact optimal step size α^* cannot be found except by a complete eigenvalue analysis of the matrix $D(x, \lambda)$, one may devise simple means for improving the convergence ratio by alternative step size choice. For example, if an upper bound E is known for E_0, then the step size $\alpha = c + 1/E$ can readily be shown to yield a better convergence ratio.

In what follows, we describe a step size rule which is based on approximation of the minimum of the ordinary dual functional $g_0[\lambda_k + \alpha \nabla g_c(\lambda_k)]$ over α by means of a quadratic or cubic fit. The approximation is used every second iteration. We present the algorithm for a variable value of penalty parameter c_k.

Given λ_{2k}, and $c_{2k}, k = 0, 1, \cdots$, we obtain x_{2k} and $h(x_{2k})$ by unconstrained minimization of the augmented Lagrangian, and we set

$$\lambda_{2k+1} = \lambda_{2k} + c_{2k} h(x_{2k})'. \tag{55}$$

Similarly we obtain $x_{2k+1}, h(x_{2k+1})$ by means of unconstrained minimization of the augmented Lagrangian. However, now we set

$$\lambda_{2k+2} = \lambda_{2k+1} + \alpha_{2k+1} h(x_{2k+1})', \tag{56}$$

where

$$\alpha_{2k+1} = c_{2k+1} \frac{h(x_{2k+1})' h(x_{2k})}{h(x_{2k+1})' h(x_{2k}) - \|h(x_{2k+1})\|^2}. \tag{57}$$

This step size rule is obtained by observing that $h(x_{2k})$ is equal to the gradient $\nabla g_0(\lambda_{2k+1})$ and $h(x_{2k+1})$ is equal to the gradient $\nabla g_0[\lambda_{2k+1} + c_{2k+1}h(x_{2k+1})]$. Thus a quadratic approximation of $g_0[\lambda_{2k+1} + \alpha h(x_{2k+1})]$ can be made based on the two gradients and the difference $c_{2k+1}h(x_{2k+1})$ between the two points. The step size α_{2k+1} of (57) maximizes the quadratic approximation over α.

Another possibility is to determine the step size α_{2k+1} by means of a cubic fit based on the gradients $\nabla g_0(\lambda_{2k+1}), \nabla g_0[\lambda_{2k+1} + c_{2k+1}h(x_{2k+1})]$ and the values of the dual functional g_0:

$$g_0(\lambda_{2k+1}) = f(x_{2k}) + \lambda_{2k+1}h(x_{2k}), \tag{58}$$

$$g_0[\lambda_{2k+1} + c_{2k+1}h(x_{2k+1})] = f(x_{2k+1}) + \lambda_{2k+1}h(x_{2k+1}) + c_{2k+1}\|h(x_{2k+1})\|^2. \tag{59}$$

The corresponding formulas for α_{2k+1} are somewhat more complicated (see [17]), but the cubic fit is more accurate than the quadratic and can be expected to yield better results for nonquadratic problems. Also, alternate quadratic fits are possible by using the values (58), (59) and one of the two gradients.

It may be shown that the sequence $\{\lambda_{2k}\}$ generated by the modified multiplier method described above satisfies, for the case of a quadratic problem,

$$\frac{\|\lambda_{2k+2} - \lambda^*\|}{\|\lambda_{2k} - \lambda^*\|} \leqq \frac{1 - e_0/E_0}{(1 + c_{2k}e_0)(1 + c_{2k+1}e_0)}.$$

The bound above, though not sharp, compares favorably with the corresponding result

$$\frac{\|\lambda_{2k+2} - \lambda^*\|}{\|\lambda_{2k} - \lambda^*\|} \leqq \frac{1}{(1 + c_{2k}e_0)(1 + c_{2k+1}e_0)}$$

associated with the ordinary method.

Consider now a general locally convex problem with nonquadratic objective function or nonlinear constraints. In this case it is necessary to restrict the step size α_{2k+1} of (56) to the interval $[c_{2k+1}, 2c_{2k+1}]$ in order to prove local convergence. This choice of interval is guided by (51) and by the fact that (47) and (50) yield $r(\alpha) \geqq r(c)$ for all $0 < \alpha < c$ when $E_0 \geqq e_0 > 0$. Thus (56) is modified to take the form

$$\lambda_{2k+2} = \lambda_{2k+1} + \bar{\alpha}_{2k+1}h(x_{2k+1})', \tag{60}$$

where

$$\bar{\alpha}_{2k+1} = \begin{cases} 2c_{2k+1} & \text{if } 2c_{2k+1} < \alpha_{2k+1}, \\ \alpha_{2k+1} & \text{if } c_{2k+1} \leqq \alpha_{2k+1} \leqq 2c_{2k+1}, \\ c_{2k+1} & \text{if } \alpha_{2k+1} < c_{2k+1}, \end{cases} \tag{61}$$

where α_{2k+1} is given by (57) or is obtained by means of the cubic fit mentioned earlier. We shall prove local convergence of the dual iteration (55), (60), (61) by viewing it as a special case of a more general algorithm which will be shown to be locally convergent both for the case of exact and approximate mimization of the augmented Lagrangian.

Referring to the problem of § 2, we consider the special case in which the eigenvalues of the matrix $D(x, \lambda)$ of (16) satisfy

$$e_i[D(x, \lambda)] < 0, \qquad i = 1, \cdots, m, \tag{62}$$

for all (x, λ) in a set $\bar{B}(x^*, \varepsilon) \times \bar{B}(\lambda^*; \delta + \bar{c}L\varepsilon)$. The positive scalars ε and δ are as in Proposition 1 and Corollary 1.1. Let the sequence $\{c_k\}$ satisfy $0 < c^* \leqq c_k \leqq (\bar{c}/2)$ (it is assumed that $2c^* \leqq \bar{c}$) and consider the iteration

$$\lambda_{k+1} = \lambda_k + s_k h[x(\lambda_k, c_k)]', \tag{63}$$

where s_k satisfies, for all k,

$$c_k \leqq s_k \leqq 2c_k. \tag{64}$$

Then we have the following local convergence result, which parallels Proposition 2 and Corollary 2.1.

PROPOSITION 5. *Assume that the initial point λ_0 belongs to $B(\lambda^*; \delta)$. Then the sequence $\{\lambda_k\}$ generated by any iteration of the form* (63), (64) *remains in $B(\lambda^*; \delta)$ and converges to λ^*. Furthermore, we have*

$$\|\lambda_{k+1} - \lambda^*\| \leqq \bar{r}_k \|\lambda_k - \lambda^*\| \quad \forall k, \tag{65}$$

where

$$\bar{r}_k = \max_{\substack{(x,\lambda) \in \bar{B}(x^*;\varepsilon) \times \bar{B}(\lambda^*;\delta+\bar{c}L\varepsilon) \\ i = 1, \cdots, m}} \left| \frac{1 + (s_k - c_k) e_i[D(x, \lambda)]}{1 - c_k e_i[D(x, \lambda)]} \right|. \tag{66}$$

Proof. First, by using the facts $c_k \leqq s_k \leqq 2c_k$ and $e_i[D(x, \lambda)] < 0$, we have for every k,

$$\bar{r}_k \leqq \max_{\substack{(x,\lambda) \in \bar{B}(x^*;\varepsilon) \times \bar{B}(\lambda^*;\delta+\bar{c}L\varepsilon) \\ 0 < c^* \leqq c_k \leqq \bar{c}/2 \\ c_k \leqq s_k \leqq 2c_k \\ i = 1, \cdots, m}} \left| \frac{1 + (s_k - c_k) e_i[D(x, \lambda)]}{1 - c_k e_i[D(x, \lambda)]} \right| = p < 1.$$

Now by substituting the step size s_0 in place of c in the proofs of Propositions 1 and 2, and by using (19) and (63), we have

$$\|\lambda_1 - \lambda^*\| \leqq \bar{r}_0 \|\lambda_0 - \lambda^*\| \leqq p\|\lambda_0 - \lambda^*\|,$$

showing that $\lambda_1 \in B(\lambda^*; \delta)$. Proceeding similarly, we have for all k, $\lambda_k \in B(\lambda^*; \delta)$ and $\|\lambda_k - \lambda^*\| \leqq p^k \|\lambda_0 - \lambda^*\|$. Hence $\lambda_k \to \lambda^*$. Q.E.D.

One may also prove propositions similar to Propositions 3 and 4 for the algorithm (63), (64) for the case of inexact minimization with either one of the termination criteria (25) or (35). For the criterion (25), we have the estimate (cf. (33))

$$\|\lambda_{k+1} - \lambda^*\| \leqq \bar{r}_k \|\lambda_k - \lambda^*\| + \varepsilon_k s_k L/M,$$

and for the criterion (35) the estimate (cf. (36))

$$\|\lambda_{k+1} - \lambda^*\| \leqq \left(\bar{r}_k + \frac{\eta_k L}{M - \eta_k L}\bar{p}_k\right)\|\lambda_k - \lambda^*\|,$$

where

$$\bar{p}_k = \max_{\substack{(x,\lambda)\in \bar{B}(x^*;\varepsilon)\times \bar{B}(\lambda^*;\delta+cL\varepsilon)\\ i=1,\cdots,m}} \left|\frac{s_k e_i[D(x,\lambda)]}{1 - c_k e_i[D(x,\lambda)]}\right|$$

and $\bar{r}_k$ is given by (66). Local convergence results similar to Proposition 5 may also be proved assuming the sequences $\{\varepsilon_k\}$ and $\{\eta_k\}$ are bounded above by sufficiently small positive numbers.

Now the local convergence results obtained clearly apply (cf. (55), (61), (64)) to the iteration given by (55), (60), (61). As shown in the next section, this iteration worked very well in numerical experiments. The iteration (63), (64) (and hence also the iterations (55), (60), (61)) can be easily extended to the case of inequality constraints by using slack variables and therefore is fully applicable to the solution of convex programming problems. In fact, for such problems, the iteration can be shown to converge globally, i.e., for an arbitrary starting point λ_0 [4].

6. Computational experience. A limited number of numerical experiments were performed to test the analysis of this paper. As a general rule, the method of multipliers performed considerably better than the corresponding quadratic penalty function method ($\lambda_k = 0$ for all k). This was true for both exact and approximate unconstrained minimization. The schemes based on approximate minimization performed considerably better than the schemes based on exact minimization both for the penalty method and the multiplier method. The modified step size rule of the previous section performed better than the regular step size rule of the multiplier method in all runs except one. It was generally found that it is better to increase the penalty parameter c at each iteration rather than to keep it at a fixed value. It is interesting to note that for the approximate minimization schemes, the unconstrained minimizations typically required one cycle of the variable metric method after the first dual iteration. Thus the approximate minimization schemes were, in effect, similar to the conjugate gradient scheme proposed by Miele, et al. [20]. We present below some detailed results for the Rosen–Suzuki problem [38]:

$$\text{Minimize } f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4$$

subject to

$$h_1(x) = 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5 \leqq 0,$$

$$h_2(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8 \leqq 0,$$

$$h_3(x) = x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10 \leqq 0.$$

The optimal solution is $x^* = (0, 1, 2, -1)'$, and the Lagrange multiplier is $\lambda^* = (2, 1, 0)$. The optimal value of the objective is $f^* = -44.0$. The eigenvalues of the negative Hessian of the ordinary dual at λ^* are $e_0 \cong .29$ and $E_0 \cong 6.95$. Here only the active constraints $h_1(x) \leqq 0$, $h_2(x) \leqq 0$ are considered.

The inequality constraints were converted to equality constraints by using a vector of slack variables $z = (z_1, z_2, z_3)'$:

$$\bar{h}_i(x, z) = h_i(x) + z_i^2 = 0, \qquad i = 1, 2, 3.$$

The resulting augmented Lagrangian is given by

$$l(x, z, \lambda, c) = f(x) + \sum_{i=1}^{3} \lambda^i [h_i(x) + z_i^2] + \frac{1}{2} c \sum_{i=1}^{3} [h_i(x) + z_i^2]^2.$$

However, rather than minimizing $l(x, z, \lambda, c)$ jointly with respect to x and z, the minimization was first done explicitly with respect to z to yield

$$\text{(67)} \quad \bar{l}(x, \lambda, c) = \min_z l(x, z, \lambda, c) = f(x) + \frac{1}{2c} \sum_{i=1}^{3} \{\max [0, \lambda^i + c h_i(x)]\}^2 - (\lambda^i)^2.$$

Subsequently, $\bar{l}(x, \lambda, c)$ was minimized with respect to x by using the Fletcher–Powell method (available on the IBM-360 as the FMFP Scientific Subroutine). The iteration for λ_k in the method of multipliers takes the form

$$\lambda_{k+1}^i = \max [0, \lambda_k^i + c h_i(x_k)], \qquad i = 1, 2, 3,$$

where x_k is the minimizing point. This updating formula is obtained from the ordinary iteration of the method of multipliers

$$\lambda_{k+1}^i = \lambda_k^i + c \bar{h}_i(x_k, z_k), \qquad i = 1, 2, 3,$$

after substitution of the minimizing value z_k obtained from (67). Table 1 shows the number of function evaluations required by the multiplier method with and without quadratic fit, and by the pure penalty method. Each function evaluation corresponds to a calculation of the values of the objective and constraint functions and their gradients.

In runs 1–5 in Table 1, accuracy to 7 significant digits of the optimal value of the objective function was attained. In runs 6–8, the accuracy was to 4 significant digits. For the runs with approximate minimization, the termination criterion (25) was used.

TABLE 1

				number of function evaluations		
run no.	c_k	ε_k	λ_0	multiplier	multiplier with quadratic fit	penalty
1	10^k	1×10^{-k}	(1, 1, 1)	110	107	221
2	5^k	1×5^{-k}	(0, 0, 0)	96	92	260
3	4^k	$.1 \times 4^{-k}$	(1, 1, 1)	112	119	282
4	2^k	10^{-5}	(0, 0, 0)	174	126	555
5	8^k	$.25 \times 8^{-k}$	(0, 0, 0)	93	92	192
6	1	$.1 \times 10^{-k}$	(1, 1, 1)	201	118	
7	1	$.1 \times 10^{-k}$	(0, 0, 0)	216	119	
8	1	10^{-5}	(1, 1, 1)	279	186	

7. Conclusions. This paper provided an analysis of the convergence rate of multiplier methods with exact and approximate unconstrained minimization. The results show that such methods can be expected to converge considerably faster than conventional penalty function methods. Furthermore, it appears that the approximate minimization schemes result in more efficient computation than schemes with exact minimization. The modified step size rule considered in § 5 appears to be promising for convex programming problems. While both theoretical and experimental evidence strongly indicate the faster convergence property of multiplier methods over penalty methods, it does not seem appropriate to predict that penalty methods will be totally replaced in the future by multiplier methods. In many problems where solution accuracy is not of paramount importance, penalty methods are not operated sequentially, but rather a single unconstrained minimization problem is solved with what is considered to be a sufficiently high value of penalty parameter. The solution of this problem is then taken as the final answer. When such a philosophy is adopted, multiplier methods can offer no advantage over penalty methods.

Acknowledgment. The author would like to acknowledge his interactions with Daniel Gabay, Barry Kort and David Luenberger, which were helpful in shaping the results of this paper. Many thanks also are due to one of the referees for his careful review and numerous helpful comments.

REFERENCES

[1] K. J. ARROW, F. J. GOULD AND S. M. HOWE, *A general saddle point result for constrained otimization.* Inst. of Statistics Mimeo Ser. no. 774, Univ. of North Carolina, Chapel Hill, N.C., 1971.

[2] K. J. ARROW, L. HURWICZ AND H. UZAWA, *Studies in Linear and Nonlinear Programming*, Stanford University Press, Calif., 1958.

[3] D. P. BERTSEKAS, *On penalty and multiplier methods*, Dept. of Engrg.-Economic Syst. Working Paper, Stanford Univ., Stanford, Calif., 1973; this Journal, to appear.

[4] ———, *On the method of multipliers for convex programming.* Dept. of Engrg.-Economic Syst. Working Paper, Stanford Univ., Stanford, Calif., 1973.

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

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

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

[8] 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.

[9] R. FLETCHER AND S. LILL, *A class of methods for nonlinear programming. II: Computational experience*, Nonlinear Programming, J. B. Rosen, O. L. Mangasarian and K. Ritter, eds., Academic Press, New York, 1971.

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

[11] 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.

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

[13] L. V. KANTOROVICH AND G. P. AKILOV, Functional Analysis in Normed Spaces, Pergamon Press, New York, 1964.

[14] 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, La., pp. 162–166.
[15] ———, *Combined primal dual and penalty methods for convex programming*, Dept. of Engrg.-Economic Syst. Working Paper, Stanford Univ., Stanford, Calif., 1973.
[16] ———, *Multiplier methods for convex programming*, Proc. 1973 IEEE Conference on Decision and Control, San Diego, Calif., pp. 428–432.
[17] D. G. LUENBERGER, *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, Mass., 1973.
[18] O. L. MANGASARIAN, *Unconstrained Lagrangians in nonlinear programming*, Computer Sciences Tech. Rep. 174, Univ. of Wisconsin, Madison, 1973; this Journal, 13 (1975), pp. 772–791.
[19] K. MARTSENSSON, *New approaches to the numerical solution of optimal control problems*, Rep. 7206, Lund Inst. of Tech., Division of Automatic Control, Lund, Sweden, 1972.
[20] A. MIELE, P. E. MOSELEY, A. V. LEVY AND G. M. COGGINS, *On the method of multipliers for mathematical programming problems*, J. Optimization Theory Appl., 10 (1972), pp. 1–33.
[21] A. MIELE, P. E. MOSELEY AND E. E. CRAGG, *A modification of the method of multipliers for mathematical programming problems*, Techniques of Optimization, A. V. Balakrishnan, ed., Academic Press, New York, 1972.
[22] 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. Optimization Theory Appl., 8 (1971), pp. 115–130.
[23] A. MIELE, E. E. CRAGG AND A. V. LEVY, *Use of the augmented penalty function in mathematical programming problems, Part II*, Ibid., 8 (1971), pp. 131–153.
[24] R. MIFFLIN, *Convergence bounds for nonlinear programming algorithms*, Tech. Rep. 57, Dept. of Administrative Sciences, Yale Univ., New Haven, Conn., 1972.
[25] J. M. ORTEGA AND W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1969.
[26] D. A. PIERRE, *Multiplier algorithms for nonlinear programming*, 7th Internat. Symp. on Math. Programming, Stanford Univ., Stanford, Calif., 1973.
[27] E. POLAK, *Computational Methods in Optimization*, Academic Press, New York, 1971.
[28] M. A. POLLATSCHEK, *Generalized duality theory in nonlinear programming*, Operations Res., Statistics and Economics, Mimeo Ser. 122, Technion, Haifa, Israel, 1973.
[29] B. T. POLYAK, *Gradient methods for the minimization of functionals*, Ž. Vyčisl. Mat. i Mat. Fiz., 3 (1963), pp. 643–653.
[30] ———, *Iterative methods using Lagrange multipliers for solving extremal problems with constraints of the equation type*, Ibid., 10 (1970), pp. 1098–1106.
[31] ———, *The convergence rate of the penalty function method*, Ibid., 11 (1971), pp. 3–11.
[32] M. J. D. POWELL, *A method for nonlinear constraints in minimization problems*, Optimization, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298.
[33] R. T. ROCKAFELLAR, *New applications of duality in convex programming*, 7th Internat. Symp. on Math. Programming, the Hague, 1970; published in Proc. 4th Conf. on Probability, Brasov, Romania, 1971.
[34] ———, *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Math. Programming, 5 (1973), pp. 354–373.
[35] ———, *The multiplier method of Hestenes and Powell applied to convex programming*, J. Optimization Theory Appl., 12 (1973), pp. 555–562.
[36] ———, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, this Journal, 12 (1974), pp. 268–285.
[37] ———, *Penalty methods and augmented Lagrangians in nonlinear programming*, Proc. 5th IFIP Conf. on Optimization Techniques, Rome, 1973.
[38] J. B. ROSEN AND S. SUZUKI, *Construction on nonlinear programming test problems*, Comm. ACM, 8 (1965), p. 113.
[39] R. D. RUPP, *Approximation of the classical isoperimetric problem*, J. Optimization Theory Appl., 9 (1972), pp. 251–264.
[40] ———, *A method for solving a quadratic optimal control problem*, Ibid., 9 (1972), pp. 238–250.
[41] ———, *A nonlinear optimal control minimization technique*, Trans. Amer. Math. Soc., 178 (1973), pp. 357–381.

[42] ———, *On the combination of the multiplier method of Hestenes and Powell with Newton's method*, S.U.N.Y. Rep., Albany, N.Y., 1973.

[43] H. SAYAMA, Y. KAMEYAMA, H. NAKAYAMA AND Y. SAWARAGI, *The generalized Lagrangian functions for mathematical programming*, 8th Internat. Symp. on Math. Programming, Stanford Univ., Stanford, Calif., 1973.

[44] S. S. TRIPATHI AND K. S. NARENDRA, *Constrained optimization problems using multiplier methods*, J. Optimization Theory Appl., 9 (1972), pp. 59–70.