MATHEMATICS OF OPERATIONS RESEARCH
Vol. 1, No. 2, May, 1976
Printed in U.S.A.

AUGMENTED LAGRANGIANS AND APPLICATIONS OF THE PROXIMAL POINT ALGORITHM IN CONVEX PROGRAMMING*

R. T. ROCKAFELLAR†

University of Washington

The theory of the proximal point algorithm for maximal monotone operators is applied to three algorithms for solving convex programs, one of which has not previously been formulated. Rate-of-convergence results for the "method of multipliers," of the strong sort already known, are derived in a generalized form relevant also to problems beyond the compass of the standard second-order conditions for optimality. The new algorithm, the "proximal method of multipliers," is shown to have much the same convergence properties, but with some potential advantages.

1. Introduction. Let C be a nonempty closed convex subset of R^n, and for $i = 0, 1, \ldots, m$ let $f_i : C \to R$ be a lower semicontinuous convex function. We consider the convex programming problem

$$\text{minimize } f_0(x) \text{ over all } x \in C \text{ satisfying } f_1(x) \leqslant 0, \ldots, f_m(x) \leqslant 0. \tag{P}$$

Our purpose is to present theoretical results on the convergence of three approaches to solving (P), all of which are shown to be realizations of the general "proximal point algorithm" for maximal monotone operators (multifunctions), studied in [1]. Each algorithm replaces (P) by a sequence of "better" minimization problems.

The first method is the primal application of the proximal point algorithm: the *proximal minimization* algorithm for (P). It is not the main focus of our attention, but it is nevertheless of some theoretical interest, and its properties have not before been analyzed in detail. There is a sequence of numbers c_k,

$$0 < c_k \uparrow c_\infty \leqslant +\infty \tag{1.1}$$

(which is given, or constructed in some manner not here specified) and an initial vector x^0. A sequence $\{x^k\}$ is generated by letting x^{k+1} be an "approximate" solution to the modified version of (P) in which $f_0(x)$ is replaced by

$$f_0^k(x) = f_0(x) + (1/2c_k)|x - x^k|^2, \tag{1.2}$$

where $|\cdot|$ denotes the Euclidean norm. (The exact meaning of "approximate" depends on the particular stopping criterion which is used in the minimization; this will be discussed later.)

An attractive feature of this approach (which must of course, in a given case, be weighed against the fact that a single problem is converted into a sequence of problems) is that the function in (1.2), in contrast probably to f_0 itself, is *strongly convex with modulus* $1/c_k$ (at least):

$$f_0^k((1-\lambda)x + \lambda x') \leqslant (1-\lambda)f_0^k(x) + \lambda f_0^k(x') - (\lambda(1-\lambda)/2c_k)|x - x'|^2$$

* Received November 24, 1975.
AMS subject classification. Primary: 90C25.
IAOR 1973 subject classification. Main: Programming: convex.
Key words. Convex programming, augmented Lagrangians, optimization algorithms, monotone operators, convergence rates.
† Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under AFOSR grant number 72-2269 at the University of Washington, Seattle.

for all x and x' in C and $\lambda \in (0, 1)$. In many algorithms, strong convexity is a boon to good convergence and makes possible more convenient stopping criteria, including estimates of how far one is from a minimum point. One is also assured that all the level sets of the objective function are bounded, and (assuming feasibility) that a unique optimal solution exists.

The possible benefits of strong convexity are not limited to direct algorithms for constrained minimization but are noteworthy too for indirect methods based on duality. Such are especially significant, for example, in the decomposition of large-scale problems. The ordinary dual problem associated with (P) is

$$\text{maximize } g_0(y) \text{ over all } y \geqslant 0, \tag{D}$$

where g_0 is the concave function defined by

$$g_0(y) = \inf_{x \in C} \{ f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x) \}. \tag{1.3}$$

Dual methods, which solve (P) as a by-product of trying to solve (D), typically involve executing, repeatedly for different choices of y, the minimization in (1.3). Even if there is no danger of the infimum being $-\infty$, so that implicit constraints besides $y \geqslant 0$ must be dealt with, troubles can arise because the infimum may not be attained at all or not attained uniquely. This may make it harder to generate simultaneously an asymptotically minimizing sequence for (P) itself.

These difficulties are avoided if f_0 is replaced by f_0^k as in the proximal minimization algorithm, since then the minimand in (1.3) is for any $y \geqslant 0$ strongly convex with modulus $1/c_k$ (entailing for the subproblem the advantages already mentioned), and furthermore g_0 is differentiable on R_+^m (cf. Falk [2]). Observe too that separability of the kind essential to decomposition methods is preserved: if

$$f_0(x) = f_{01}(x_1) + \cdots + f_{0N}(x_N) \quad \text{for } x_j \in R^{n_j},$$

then

$$f_0^k(x) = f_{01}^k(x_1) + \cdots + f_{0N}^k(x_N) \quad \text{with } f_{0j}^k(x_j) = f_{0j}(x_j) + (1/2c_k)|x_j - x_j^k|^2.$$

A crucial question in deciding when the passage from f_0 to the sequence $\{f_0^k\}$ is worthwhile, is whether the outer algorithm (generating approximate solutions x^k to the corresponding modified versions of (P)) has an adequate rate of convergence. This is some of the motivation for the results presented here concerning the proximal minimization algorithm.

One unusual property is that the sequence $\{x^k\}$ typically converges to a particular optimal solution to (P), even though this may not be the only optimal solution.

The second method we treat in this paper is the dual application of the proximal point algorithm, i.e., the application of the above idea to (D) instead of (P). Seen in the framework of its effects on (P), this is the important *method of multipliers*, originally suggested in basic form for equality-constrained nonlinear programs by Hestenes [3] and Powell [4]. This algorithm is expressed in terms of an "augmented Lagrangian function," which depends on a positive parameter c.

For the inequality-constrained problem (P), the *augmented Lagrangian* is

$$L(x, y, c) = f_0(x) + \sum_{i=1}^{m} \psi(f_i(x), y_i, c) \quad \text{for all } x \in C, y \in R^m, c > 0, \tag{1.4}$$

where

$$\begin{aligned} \psi(f_i(x), y_i, c) &= y_i f_i(x) + (c/2) f_i(x)^2 && \text{if } f_i(x) \geqslant -y_i/c, \\ &= -(1/2c) y_i^2 && \text{if } f_i(x) \leqslant -y_i/c. \end{aligned} \tag{1.5}$$

The algorithm in question again depends on a nondecreasing sequence of numbers c_k satisfying (1.1), and there is an initial multiplier vector y^0. The sequences $\{y^k\} \subset R^m$ and $\{x^k\} \subset C$ are generated as follows:

$$x^{k+1} \approx \arg\min_{x \in C} L(x, y^k, c_k), \tag{1.6}$$

$$y^{k+1} = y^k + c_k \nabla_y L(x^{k+1}, y^k, c_k) = Y(x^{k+1}, y^k, c_k), \tag{1.7}$$

where $Y(x, y, c)$ is the vector in R^m whose coordinates are

$$Y_i(x, y, c) = \max\{0, y_i + cf_i(x)\}. \tag{1.8}$$

The method of multipliers has recently attracted much attention for its evident advantages over ordinary penalty techniques, and many authors have made contributions, too numerous to describe here (see the recent survey article of Bertsekas [5], treating a bibliography of 66 items). The main result of this theory, embellished by many details and variations, is that when the "strong second-order sufficient conditions" for optimality hold in (P) and an appropriate criterion is chosen for the approximation in (1.6), the sequence (x^k, y^k) converges to an optimal pair $(\bar{x}, \bar{y})$ at a linear rate whose ratio is, roughly speaking, inversely proportional to c_∞ if c_∞ is large. (The convergence is superlinear if $c_\infty = +\infty$.) This is true in a local sense even without convexity: cf. Polyak and Tretyakov [6], Bertsekas [7], [8].

Global convergence in the convex case with equality in (1.6) was established by Tretyakov [9]. Rockafellar [10] obtained a similar result under weaker assumptions, allowing in particular for inexact minimization in (1.6), and showed that $\{y^k\}$ converges to a particular dual optimal solution y^∞, even though there may be more than one such. The fact that, for convex programming, the method of multipliers amounts to an application of the proximal point algorithm was noted and used in [10] in obtaining the latter fact, although the general theory of the proximal point algorithm in [1] was not then available.

Here we demonstrate that the rate-of-convergence results mentioned above can be obtained in the convex case as corollaries of results in this general theory, moreover in a form which is broader in some respects. Notably, we are able to replace the second-order conditions for optimality by a certain property (the Lipschitz continuity of a fundamental mapping T_l^{-1} just at the origin) which can be expected to be satisfied "usually" in more general classes of problems than those for which the second-order conditions can sensibly be formulated (see the remark at the end of §2). We also obtain new estimates which promise to be useful in terminating the computation.

Such results do not appear to carry over to "nonquadratic" variants of the method of multipliers, such as have been introduced by Kort and Bertsekas [11] and Golshtein and Tretyakov [12], and they thus serve to emphasize the special importance of the "quadratic" augmented Lagrangian (1.4)–(1.5).

One aspect of the method of multipliers which is not entirely satisfactory is how to formulate a really convenient stopping criterion for the inexact minimization that still ensures global convergence. The rule utilized in [10], [13] requires verifying that the value of the minimand in (1.6) is within a specified tolerance of the true minimum. This is practicable in some cases (see [14, §2]), and in particular, as has been utilized by Kort and Bertsekas [11], it can be translated into a verification of the smallness of the gradient of f_0, if $C = R^n$ and f_0 is differentiable and strongly convex with known modulus. (The latter result can be extended to cases with $C \neq R^n$; see §4 below.)

Since the desired strong convexity of f_0 is not always present, the idea arises of manufacturing it, as above, by replacing f_0 by the function f_0^k in (1.2) at each iteration.

The resulting algorithm, which may be dubbed the *proximal method of multipliers*, is identical to the method of multipliers, except that instead of (1.6) one has

$$x^{k+1} \approx \arg\min_{x \in C} \{ L(x, y^k, c_k) + (1/2c_k)|x - x^k|^2 \}. \tag{1.9}$$

It is remarkable that such a seemingly heuristic device can again be construed as a realization of the general proximal point algorithm. Indeed, it corresponds to the monotone operator associated with the ordinary Lagrangian function for (P).

For this reason, we are able to derive convergence results for the proximal method of multipliers along the same lines as those for the usual method of multipliers. It turns out that the new algorithm has essentially the same rate of convergence, but with notable advantages at hardly any apparent extra cost. Not only is there the desirable numerical feature that at each iteration one minimizes a strongly convex function with known modulus $1/c_k$, but also the sequences $\{x^k\}$ and $\{y^k\}$ are *both* convergent (even though primal and dual optimal solutions may not be unique).

The plan of the paper is to treat the three applications of the proximal point algorithm consecutively in §§3, 4, 5 after some preliminaries about monotone operators in §2. Remarks about various extensions, such as to equality constraints and infinite-dimensional problems, are made in §6.

2. Representation in terms of monotone operators. A *monotone operator* from R^N to R^N is a multifunction T with the property that $(z - z')\cdot(w - w') \geq 0$ whenever $w \in T(z)$, $w' \in T(z')$. It is *maximal* if its graph is not properly contained in the graph of any other monotone operator. Important examples are the subdifferential mappings associated with closed proper convex functions [15], [16, Cor. 31.5.2] and, more generally, modified forms of the subdifferential mappings of closed proper saddle-functions [17], [16, Cor. 37.5.2].

Many problems can be reduced to the calculation of an element z satisfying $0 \in T(z)$, where $T : R^N \to R^N$ is a maximal monotone operator. The general *proximal point algorithm* for T accomplishes this by associating with a nondecreasing sequence of scalars $c_k > 0$ and an initial vector z^0 a sequence $\{z^k\} \subset R^N$ generated by

$$z^{k+1} \approx P_k(z^k), \quad \text{where } P_k = (I + c_k T)^{-1}. \tag{2.1}$$

(It is a fact that P_k is always single-valued everywhere; cf. [1].)

Two general criteria for the approximate equality in (2.1) were studied in [1], namely

$$|z^{k+1} - P_k(z^k)| \leq \epsilon_k, \qquad \sum_{k=0}^{\infty} \epsilon_k < \infty, \quad \text{and} \tag{A}$$

$$|z^{k+1} - P_k(z^k)| \leq \delta_k |z^{k+1} - z^k|, \qquad \sum_{k=0}^{\infty} \delta_k < \infty, \tag{B}$$

where one has the estimate

$$|z^{k+1} - P_k(z^k)| \leq c_k \operatorname{dist}(0, T(z^{k+1}) + c_k^{-1}(z^{k+1} - z^k)). \tag{2.2}$$

To summarize the results, we showed very broadly that (A) guarantees the convergence of $\{z^k\}$ to a particular solution z^∞ to $0 \in T(z)$, even though there may be other solutions. If (B) is used with (A) and T^{-1} is "Lipschitz continuous at the origin," then the convergence is at least at a linear rate whose modulus tends to zero as c_k increases (thus providing superlinear convergence if $c_\infty = \infty$).

Lipschitz continuity of T^{-1} at the origin (with modulus $a \geq 0$) means that $T^{-1}(0)$ consists of a single element $\bar{z}$, and for some $\tau > 0$ we have $|z - \bar{z}| \leq a|w|$ whenever $w \in T(z)$ and $|w| \leq \tau$.

As explained in the introduction, our aim is to put these results in a more concrete form in the context of convex programming, showing that (A) and (B) can then be expressed in terms that are convenient and implementable. This requires that we look at problems (P) and (D) in a more abstract fashion.

Let us denote by l the ordinary Lagrangian function for (P) in extended form:

$$\begin{aligned} l(x, y) &= f_0(x) + \sum_{i=1}^{m} y_i f_i(x) && \text{if } x \in C \text{ and } y \in R^m_+, \\ &= -\infty && \text{if } x \in C \text{ and } y \notin R^m_+, \\ &= +\infty && \text{if } x \notin C. \end{aligned} \tag{2.3}$$

The essential objective function in (P) is

$$\begin{aligned} f(x) = \sup_{y \in R^m} l(x, y) &= f_0(x) && \text{if } x \text{ is feasible in (P)}, \\ &= +\infty && \text{if } x \text{ is not feasible in (P)}, \end{aligned} \tag{2.4}$$

while the essential objective function in (D) is

$$\begin{aligned} g(y) = \inf_{x \in R^n} l(x, y) &= g_0(y) && \text{if } y \geqslant 0, \\ &= -\infty && \text{if } y \ngeqslant 0. \end{aligned} \tag{2.5}$$

We observe that l is a closed saddle-function in the sense of [16, p. 363], due to our assumptions of convexity and continuity (and the closedness of C), and the mapping $T_l : (x, y) \to \{(v, u) \mid (v, -u) \in \partial l(x, y)\}$ is therefore a maximal monotone operator in R^{n+m} [16, Cor. 37.5.2]. A solution to $(0, 0) \in T_l(x, y)$ is a saddle point of l. At the same time, f is a closed convex function on R^n, while g is a closed concave function on R^m. Let $T_f = \partial f$ and $T_g = -\partial g$. If $f \not\equiv +\infty$, then f is proper, and T_f is a maximal monotone operator in R^n such that the solutions to $0 \in T_f(x)$ are the optimal solutions to (P) [16, §§23–24]. Similarly, if $g \not\equiv -\infty$, then g is proper, and T_g is a maximal monotone operator in R^m such that the solutions to $0 \in T_g(y)$ are the optimal solutions to (D).

The three algorithms described in the introduction will be shown to correspond respectively to the monotone operators T_f, T_g and T_l. The relationship between these operators and their various inverses is therefore of some importance.

For each $v \in R^n$ and $u \in R^m$, consider the following perturbed form of problem (P):

$$\begin{aligned} &\text{minimize } f_0(x) - x \cdot v \text{ over all } x \in C \\ &\quad \text{satisfying } f_i(x) + u_i \leqslant 0 \text{ for } i = 1, \ldots, m. \end{aligned} \tag{$P(v, u)$}$$

(Note that the variables u_i are introduced here with the opposite sign from that in other treatments such as [16]. This simplifies certain relationships below.) From the definition, we have

$$T_f^{-1}(v) = \arg\min_{x \in R^n} \{ f(x) - x \cdot v \}, \tag{2.6}$$

and hence, assuming the constraints $x \in C$ and $f_i(x) \leqslant 0$, $i = 1, \ldots, m$, can be satisfied,

$$T_f^{-1}(v) = \text{set of all optimal solutions to } (P(v, 0)). \tag{2.7}$$

By the same token, we have

$$\begin{aligned} T_g^{-1}(u) &= \arg\max_{y \in R^m} \{ g(y) + y \cdot u \} \\ &= \text{set of all optimal solutions to } (D(0, u)), \end{aligned} \tag{2.8}$$

where $D(v, u)$ is the ordinary dual of $(P(v, u))$, assuming $g \not\equiv -\infty$. If $\min(P(0, u)) = \sup(D(0, u))$ for all u, as is true for instance if C is bounded, or more generally if for some $\alpha \in R$ the set of all $x \in C$ satisfying $f_0(x) \leqslant \alpha$ and $f_i(x) \leqslant 0$ for $i = 1, \ldots, m$ is nonempty and bounded [16, Corollary 8.7.1 and Theorem 30.4(g)], [18, Theorems 17′, 18′], then

$$T_g^{-1}(u) = \partial p(u) \quad \text{where } p(u) = \inf(P(0, u)) \tag{2.9}$$

[16, Theorems 23.5, 30.2]. Finally,

$$T_l^{-1}(v, u) = \arg \operatorname*{minimax}_{x \in R^n, y \in R^m} \{l(x, y) - x \cdot v + y \cdot u\}, \tag{2.10}$$

where the "minimaximand" is none other than the ordinary Lagrangian for $(P(v, u))$. (See [16, p. 386].) Therefore

$$T_l^{-1}(v, u) = \text{set of all } (x, y) \text{ satisfying the Kuhn-Tucker saddle-point condition for } (P(v, u)). \tag{2.11}$$

Since a saddle point of the Lagrangian always yields primal and dual optimal solutions, it is clear that in general $T_f^{-1}(v)$ contains the projection of $T_l^{-1}(v, 0)$ on R^n, while $T_g^{-1}(u)$ contains the projection of $T_l^{-1}(0, u)$ on R^m. The inclusions may be strict without the imposition of further conditions, however.

PROPOSITION 1. *If T_l^{-1} is Lipschitz continuous at the origin with modulus a, then so are T_f^{-1} and T_g^{-1}. In fact, then T_f and T_g are maximal monotone, and*

$$T_f^{-1}(v) = \textit{projection of } T_l^{-1}(v, 0) \textit{ on } R^n \textit{ for each } v \in R^n, \tag{2.12}$$

$$T_g^{-1}(u) = \textit{projection of } T_l^{-1}(0, u) \textit{ on } R^m \textit{ for each } u \in R^m. \tag{2.13}$$

PROOF. Lipschitz continuity of T_l^{-1} at the origin entails the uniform boundedness of the sets $T_l^{-1}(v, u)$ in a neighborhood of $(0, 0)$, or in other words the local boundedness of T_l^{-1} at $(0, 0)$. But T_l^{-1}, as the inverse of a maximal monotone operator, is itself maximal monotone (by symmetry in the definition), and hence the local boundedness implies T_l^{-1} is also nonempty-valued on a neighborhood of $(0, 0)$ [19]. In particular, since $T_l^{-1}(0, 0) \neq \varnothing$, l has a saddle point $(\bar{x}, \bar{y})$; then $\bar{x}$ is an optimal solution to (P) and $\bar{y}$ is an optimal solution to (D) (so that $f \not\equiv +\infty$ and $g \not\equiv -\infty$), implying T_f and T_g are maximal monotone. In view of the fact noted above that the inclusion $\supset$ always holds in (2.12) and (2.13), we can complete the proof merely by showing that the operators in R^n and R^m corresponding to the right sides of (2.12) and (2.13), respectively, are maximal monotone under our hypothesis. Let M be the subspace of $R^n \times R^m$ corresponding to the R^n component; thus $M = R^n \times \{0\}$ and $M^\perp = \{0\} \times R^m$. The maximal monotonicity of the multifunction $v \to$ projection of $T_l^{-1}(v, 0)$ on R^n, as an operator in R^n, is equivalent to that of

$$\begin{aligned} S : (v, u) \to S(v, u) &= T_l^{-1}(v, u) + M^\perp \quad && \text{if } (v, u) \in M, \\ &= \varnothing && \text{if } (v, u) \notin M. \end{aligned}$$

Actually, $S = T_l^{-1} + \partial\delta_M$, where δ_M is the indicator of M (i.e., the closed proper convex function which vanishes on M and is $+\infty$ everywhere else). Thus S is the sum of two maximal monotone operators, T_l^{-1} and $\partial\delta_M$. Since T_l^{-1} is nonempty-valued on a neighborhood of a point where $\partial\delta_M$ is nonempty-valued (namely the point $(0, 0)$), we may conclude via [20, Theorem 1] that S is maximal monotone, as desired. The maximal monotonicity of $u \to$ projection of $T_l^{-1}(0, u)$ on R^m is argued in the same way, only with the roles of M and $M^\perp$ reversed in the definition of S.

To obtain a plausible condition for the Lipschitz continuity of T_l^{-1} at the origin, we appeal to the *strong second-order conditions for optimality in* (P). These are comprised of the following properties (cf. Hestenes [21, Chap. 1]).

(a) There is a saddle point $(\bar{x}, \bar{y})$ of l such that $\bar{x} \in \operatorname{int} C$. Moreover, the functions f_i for $i = 0, 1, \ldots, m$ are twice continuously differentiable on a neighborhood of $\bar{x}$.

(b) Let I be the set of active constraint indices at the point $\bar{x}$: $I = \{i \in [1, m] \mid f_i(\bar{x}) = 0\}$. Then $\bar{y}_i > 0$ for every $i \in I$, and the gradients $\nabla f_i(\bar{x})$ for $i \in I$ form a linearly independent set.

(c) The Hessian matrix $H = \nabla_x^2 l(\bar{x}, \bar{y})$ satisfies $w \cdot Hw > 0$ for every $w \neq 0$ such that $w \cdot \nabla f_i(\bar{x}) = 0$ for all $i \in I$.

As is well known, these conditions (along with convexity) imply that $\bar{x}$ and $\bar{y}$ are the unique optimal solutions to (P) and (D).

PROPOSITION 2. *If the strong second-order conditions* (a), (b), (c) *are satisfied, then* T_l^{-1} *is actually single-valued and continuously differentiable on a neighborhood of the origin, and so are* T_f^{-1} *and* T_g^{-1}. *Thus in particular, these mappings are all Lipschitz continuous at the origin.*

PROOF. Let Ω denote the neighborhood of $\bar{x}$ mentioned in condition (a), and suppose for notational simplicity that the active constraint set is $I = \{1, \ldots, r\}$. Then (a) and the complementary slackness in (b) can be expressed as

$$\nabla f_0(\bar{x}) + \sum_{i=1}^{r} \bar{y}_i \nabla f_i(\bar{x}) = 0, \tag{2.14}$$

$$f_i(\bar{x}) = 0 \text{ and } \bar{y}_i > 0 \quad \text{for } i = 1, \ldots, r, \tag{2.15}$$

$$f_i(\bar{x}) < 0 \text{ and } \bar{y}_i = 0 \quad \text{for } i = r+1, \ldots, m. \tag{2.16}$$

Let $G : \Omega \times R^r \to R^n \times R^r$ be the mapping defined by

$$G(x, y_1, \ldots, y_r) = \left(\nabla f_0(x) + \sum_{i=1}^{r} y_i \nabla f_i(x), \; -f_1(x), \ldots, -f_r(x) \right). \tag{2.17}$$

Then G is continuously differentiable. At $(\bar{x}, \bar{y}_1, \ldots, \bar{y}_r)$, G vanishes and its derivative is the linear transformation

$$G'(\bar{x}; \bar{y}_1, \ldots, \bar{y}_r) : (w; z_1, \ldots, z_r) \to$$
$$\left(Hw + \sum_{i=1}^{r} z_i \nabla f_i(\bar{x}), \; -w \cdot \nabla f_i(\bar{x}), \ldots, -w \cdot \nabla f_r(\bar{x}) \right). \tag{2.18}$$

This transformation is nonsingular by virtue of (c) and the linear independence in (b). (If $(w, z_1, \ldots, z_r)$ is such that the image in (2.18) is the zero vector, then $w \cdot \nabla f_i(\bar{x}) = 0$ for $i = 1, \ldots, r$ and $Hw + \sum_{i=1}^{r} z_i \nabla f_i(\bar{x}) = 0$. Hence $0 = w \cdot [Hw + \sum_{i=1}^{r} z_i \nabla f_i(\bar{x})] = w \cdot Hw$, implying by (c) that $w = 0$. It follows that $\sum_{i=1}^{r} z_i \nabla f_i(\bar{x}) = 0$, and hence from the linear independence in (b) that $z_i = 0$ for $i = 1, \ldots, r$.) The nonsingularity of $G'(\bar{x}; \bar{y}_1, \ldots, \bar{y}_m)$ guarantees via the implicit function theorem that there is a continuously differentiable function γ, defined on a neighborhood Ω' of the origin in $R^n \times R^r$, such that

$$G(\gamma(v; u_1, \ldots, u_r)) \equiv (v; u_1, \ldots, u_r) \quad \text{and } \gamma(0; 0, \ldots, 0) = (\bar{x}; \bar{y}_1, \ldots, \bar{y}_r). \tag{2.19}$$

Now consider the continuously differentiable mapping $S : \Omega' \times R^m \to (R^n \times R^r) \times R^{m-r} = R^n \times R^m$ defined by $S(v, u) = (\gamma(v, u_1, \ldots, u_r), 0)$. If v, u is sufficiently near the origin, then for $(x, y) = S(v, u)$ we have $x \in \operatorname{int} C$, $\{\nabla f_i(x) \mid i = 1, \ldots, r\}$

still linearly independent,

$$[\nabla f_0(x) - v] + \sum_{i=1}^{r} y_i \nabla f_i(x) = 0,$$

$$f_i(x) + u_i = 0 \text{ and } y_i > 0 \quad \text{for } i = 1, \ldots, r,$$

$$f_i(x) + u_i < 0 \text{ and } y_i = 0 \quad \text{for } i = r+1, \ldots, m,$$

as well as $w \cdot \nabla_x^2 l(x, y) w > 0$ for every $w = 0$ such that $w \cdot \nabla f_i(x) = 0$ for $i = 1, \ldots, r$. But this means that the strong second-order conditions for the perturbed problem $(P(v, u))$ are satisfied by (x, y), and hence of course that (x, y) is the unique pair satisfying the Kuhn-Tucker saddle-point condition for $(P(v, u))$. Therefore, in accordance with (2.10), T_l^{-1} agrees with the continuously differentiable mapping S in a neighborhood of the origin. The fact that then T_f^{-1} and T_g^{-1} are also continuously differentiable in a neighborhood of the origin follows from Proposition 1.

Next we give direct conditions for the Lipschitz continuity of T_f^{-1} and T_g^{-1}.

PROPOSITION 3. (a) *T_f^{-1} is Lipschitz continuous at the origin if and only if* (P) *has a unique optimal solution $\bar{x}$, and there exist $\lambda > 0$ and $\epsilon > 0$ such that*

$$f_0(x) \geqslant f_0(\bar{x}) + \lambda|x - \bar{x}|^2 \quad \textit{for all feasible } x \textit{ satisfying } |x - \bar{x}| \leqslant \epsilon. \tag{2.20}$$

In this event, for every $\alpha \in R$ the set of all feasible solutions x to (P) *satisfying $f_0(x) \leqslant \alpha$ is bounded.*

(b) *T_g^{-1} is Lipschitz continuous at the origin if and only if the convex function $p(u) = \inf(P(0, u))$ is finite and differentiable at $u = 0$, and there exist $\lambda > 0$ and $\epsilon > 0$ such that*

$$p(u) \leqslant p(0) + u \cdot \nabla p(0) + \lambda|u|^2 \quad \textit{for all } u \textit{ satisfying } |u| \leqslant \epsilon. \tag{2.21}$$

In this event, (P) *satisfies the Slater condition.*

PROOF. Part (a) is immediate from [1, Proposition 7(b)] applied to the present function f. (If $f \equiv +\infty$, then $T_f(x) = R^n$ for all x, so that Lipschitz continuity at the origin, which entails single-valuedness at the origin, is impossible.) The uniqueness of the optimal solution to (P) implies that the level set $\{x \mid f(x) \leqslant \alpha\}$ is nonempty and bounded for $\alpha = \min(P)$, and hence it is bounded for every $\alpha \in R$ [16, Cor. 8.7.1].

In Part (b) we use the conjugate relation $p^* = -g$ [16, Theorem 30.2]. We have $T_g = \partial p^*$, so by applying [1, Proposition 8(c)] to p^* instead of f we get that T_g^{-1} is Lipschitz continuous at the origin if and only if the condition in (b) is satisfied by the function $p^{**} = \text{cl}\, p$. However, this condition on $\text{cl}\, p$ is equivalent to the same condition on p, because it concerns an open set where the function in question must be finite, while p and $\text{cl}\, p$ agree on any such set [16, Theorem 7.4]. Of course, (2.21) implies $p(u) < +\infty$ for some $u = (u_1, \ldots, u_m)$ with every $u_i < 0$, so that for some $\tilde{x} \in C$ we have $f_i(\tilde{x}) < 0$ for $i = 1, \ldots, m$ (Slater condition).

Finally, we provide more clarification of the condition that T_l^{-1} be Lipschitz continuous at the origin.

PROPOSITION 4. *The function $\pi(v, u) = \inf(P(v, u))$ is concave in v and convex in u. It is finite on a neighborhood of* (0, 0) *if and only if* (P) *satisfies the Slater condition, and for some $\alpha \in R$ the set of all feasible solutions x to* (P) *with $f_0(x) \leqslant \alpha$ is nonempty and bounded.*

A sufficient (*but not necessary*) *condition for T_l^{-1} to be Lipschitz continuous at* (0, 0) *is that π be twice differentiable at* (0, 0). *On the other hand, the strong second-order conditions* (a), (b), (c) *imply the twice differentiability of π on an entire neighborhood of* (0, 0).

PROOF. The function

$$l^*(v, u) = -\inf(P(v, -u)) = \sup_x \inf_y \{v \cdot x + u \cdot y - l(x, y)\} \tag{2.22}$$

is conjugate to l, and in particular "closed" convex-concave [16, p. 389]; hence π is "closed" concave-convex. The finiteness of $\pi(v, u)$ on a neighborhood of the origin is thus (by the structure of such functions; [16, Theorem 34.2]) equivalent to that of the functions $p(u) = \pi(0, u)$ and $\pi(v, 0) = \inf\{f(x) - v \cdot x\} = -f^*(v)$. But p is finite on a neighborhood of 0 if and only if (P) satisfies the Slater condition, while $f^*(v)$ is finite on a neighborhood of 0 if and only if some set of the form $\{x \mid f(x) \leqslant \alpha\}$ is nonempty and bounded [16, Theorem 27.1(d)(f)]. This verifies the second assertion of the proposition. Recalling next that $(v, u) \in \partial l(x, y) \Leftrightarrow (x, y) \in \partial l^*(v, u)$ [16, Theorem 37.5], we see that $T_l^{-1}(v, u) = \partial l^*(v, -u)$. The differentiability of T_l^{-1} at (v, u) is thus equivalent to the twice differentiability of π at (v, u). But differentiability of T_l^{-1} at (v, u) implies Lipschitz continuity at (v, u) [1, Proposition 4]. This proves the rest, in view of Proposition 2.

REMARK. Proposition 4 leads, at least heuristically, to the conclusion that for "most" convex programs (P) it will be true that T_l^{-1} is Lipschitz continuous at the origin. The reasoning is that convex functions (and concave functions) are twice differentiable almost everywhere on any open set where they are finite (cf. [22]), and the same might therefore be expected to be true of concave-convex functions like π. (However, no one has in fact tackled this question.) The twice differentiability of π on an entire neighborhood of (0, 0) seems, by contrast, to be a rather special property. The Lipschitz condition on T_l^{-1} is thus more appealing as a hypothesis in the results below than the strong second-order conditions for optimality in (P), and of course it has the further advantage that it makes sense for a much wider class of problems.

Similar observations can be made about T_f^{-1} and T_g^{-1}: since these are just the subdifferentials of certain convex functions, they are single-valued and differentiable (hence Lipschitz continuous) at almost all interior points of the sets where the functions in question (namely f^* and cl p) are finite. We have seen for T_g^{-1} that the set has the origin in its interior if and only if (P) satisfies the Slater condition, while for T_f^{-1} this is the case if and only if for some $\alpha \in R$ the set of all feasible solutions to (P) satisfying $f_0(x) \leqslant \alpha$ is nonempty and bounded.

3. Primal application: proximal minimization algorithm. The proximal minimization algorithm for (P) can be expressed by

$$x^{k+1} \approx \arg\min_{x \in R^n} \phi_k(x), \tag{3.1}$$

where ϕ_k is the closed convex function on R^n defined by

$$\phi_k(x) = f(x) + (1/2c_k)|x - x^k|^2 \tag{3.2}$$

(f being the essential objective function for (P) as in (2.4)). The sequence $\{c_k\}$ satisfying (1.1) and the initial vector x^0 are given. Minimizing ϕ_k on R^n is the same as minimizing f_0^k (see (1.2)) over the feasible set

$$D = \{x \in C \mid f_i(x) \leqslant 0 \text{ for } i = 1, \ldots, m\}. \tag{3.3}$$

Assuming that $D \neq \varnothing$ (so that f and ϕ_k are proper), we have

$$\partial\phi_k(x) = \partial f(x) + c_k^{-1}(x - x^k) = T_f(x) + c_k^{-1}(x - x^k) \quad \text{for all } x \tag{3.4}$$

[16, Theorem 23.8]. This can be used in the estimate (2.2) to get the following criteria for the approximate relation in (3.1) that respectively imply (A) and (B) of §2 for T_f:

$$\operatorname{dist}(0, \partial\phi_k(x^{k+1})) \leqslant \epsilon_k/c_k, \qquad \sum_{k=0}^{\infty} \epsilon_k < \infty, \tag{A$'$}$$

$$\operatorname{dist}(0, \partial\phi_k(x^{k+1})) \leqslant (\delta_k/c_k)|x^{k+1} - x^k|, \qquad \sum_{k=0}^{\infty} \delta_k < \infty. \tag{B$'$}$$

Of course, exact minimization of ϕ_k corresponds to dist$(0, \partial\phi_k(x^{k+1})) = 0$, in which event (A′) and (B′) are both satisfied. By convention, dist$(0, \partial\phi_k(x^{k+1})) = +\infty$ if $\partial\phi_k(x^{k+1}) = \varnothing$; thus (A′) and (B′) entail $\partial\phi_k(x^{k+1}) \neq \varnothing$ and hence $x^{k+1} \in D$ (i.e., x^{k+1} must be a feasible solution to (P)).

The next proposition gives further information relevant to estimating dist$(0, \partial\phi_k(x^{k+1}))$. Here for a closed convex set $K \neq \varnothing$ and point v we let prox$(v \mid K)$ = unique point of K nearest to v. The closed tangent cone to the convex set D at a point x is denoted by $K_D(x)$; similarly $K_C(x)$ for $x \in C$. Thus prox$(-\nabla f_0^k(x) \mid K_D(x))$ in the next result gives the "direction of steepest descent" of f_0^k at x relative to the feasible set D.

PROPOSITION 5. *Suppose the functions f_i are all differentiable relative to C (with* int $C \neq \varnothing$). *Then for ϕ_k in* (3.2) *and any $x \in D$, one has*

$$\begin{aligned}\operatorname{dist}(0, \partial\phi_k(x)) &= |\operatorname{prox}(-\nabla f_0^k(x) \mid K_D(x))| \\ &\leqslant \min_{\substack{y_i \geqslant 0;\\ y_i f_i(x) = 0}} \left|\operatorname{prox}\left(-\nabla f_0^k(x) - \sum_{i=1}^m y_i \nabla f_i(x) \mid K_C(x)\right)\right|, \end{aligned} \tag{3.5}$$

where equality holds for example if (P) *satisfies the Slater condition.*

PROOF. These relations follow from the calculus of directional derivatives and subgradients in [16, §23]. The differentiability of f_0 relative to C (and hence that of f_0^k) gives us for any $x \in D$

$$\phi_k'(x; w) = w \cdot \nabla f_0^k(x) + \delta_D'(x; w), \tag{3.6}$$

where δ_D is the indicator of D. Passing on both sides to the conjugate function with respect to w, we see [16, Theorem 23.2 and p. 215] that

$$\partial\phi_k(x) = \nabla f_0^k(x) + N_D(x), \tag{3.7}$$

where $N_D(x)$ is the normal cone to D at x, the polar of $K_D(x)$. Moreover

$$N_D(x) \supset N_C(x) + \left\{ \sum_{i=1}^m y_i \nabla f_i(x) \mid y_i \geqslant 0, y_i f_i(x) = 0 \right\}, \tag{3.8}$$

with equality if the Slater condition holds [16, p. 283]. From (3.7) and [16, pp. 339–340] we have

$$\begin{aligned}\operatorname{dist}(0, \partial\phi_k(x)) &= \operatorname{dist}(-\nabla f_0^k(x), N_D(x)) \\ &= |-\nabla f_0^k(x) - \operatorname{prox}(-\nabla f_0^k(x) | N_D(x))| = |\operatorname{prox}(-\nabla f_0^k(x) | K_D(x))| \end{aligned} \tag{3.9}$$

as claimed in the first part of (3.5). On the other hand, for any multipliers $y_i \geqslant 0$ with $y_i f_i(x) = 0$, $i = 1, \ldots, m$, we have from (3.7) and (3.8) that

$$\partial\phi_k(x) \supset \nabla f_0^k(x) + \sum_{i=1}^m y_i \nabla f_i(x) + N_C(x),$$

and hence by the same reasoning as in (3.9)

$$\operatorname{dist}(0, \partial\phi_k(x)) \leqslant \left|\operatorname{prox}\left(-\nabla f_0^k(x) - \sum_{i=1}^m y_i \nabla f_i(x) | N_C(x)\right)\right|.$$

Thus the inequality in (3.5) is also valid (the infimum corresponding to a quadratic programming problem and therefore being attained). The argument shows that the inequality becomes equality if (3.8) holds with equality.

The following theorems are obtained simply by specializing to the essential objective function f the results of [1, §4] for a general closed proper convex function f on a Hilbert space. This amounts to taking $T = T_f$ in the general proximal point algorithm.

THEOREM 1. *Suppose* $\inf(\mathrm{P}) < +\infty$, *and let the proximal minimization algorithm be executed with stopping criterion* (A′) *applied to* ϕ_k *in* (3.2). *If the generated sequence* $\{x^k\}$ *of feasible solutions to* (P) *is bounded, then* $x^k \to x^\infty$, *where* x^∞ *is some optimal solution to* (P), *and moreover*

$$0 \leq f_0(x^{k+1}) - \min(\mathrm{P}) \leq c_k^{-1}|x^{k+1} - x^\infty|(\epsilon_k + |x^{k+1} - x^k|) \to 0. \tag{3.10}$$

The boundedness of x^k *under* (A′) *is in fact equivalent to the existence of an optimal solution to* (P). *Thus it is certain to hold if for some* $\alpha \in R$ *the set of feasible* x *satisfying* $f_0(x) \leq \alpha$ *is nonempty and bounded, or more particularly if* T_f^{-1} *is Lipschitz continuous at the origin* (*cf. Propositions* 1, 2, 3).

PROOF. This corresponds to [1, Theorem 4 (and Theorem 1)].

THEOREM 2. *Suppose* $\inf(\mathrm{P}) < +\infty$, *and let the proximal minimization algorithm be executed with stopping criterion* (B′) *applied to* ϕ_k *in* (3.2). *If* T_f^{-1} *is Lipschitz continuous at the origin with modulus* a_f (*cf. Propositions* 1, 2, 3) *and* $\{x^k\}$ *is bounded* (*cf. Theorem* 1), *then* $x^k \to \bar{x}$, *where* $\bar{x}$ *is the unique optimal solution to* (P), *and*

$$|x^{k+1} - \bar{x}| \leq \theta_k |x^k - \bar{x}| \quad \text{for all } k \text{ sufficiently large}, \tag{3.11}$$

where

$$\theta_k = \left[a_f(a_f^2 + c_k^2)^{-1/2} + \delta_k\right](1 - \delta_k)^{-1}$$
$$\to \theta_\infty = a_f(a_f^2 + c_\infty^2)^{-1/2} < 1 \qquad (\theta_\infty = 0 \text{ if } c_\infty = +\infty). \tag{3.12}$$

PROOF. Everything is covered by corresponding assertions of the general rate-of-convergence theorem for the proximal point algorithm [1, Theorem 2]. (See the remarks made there for some further refinements.)

Note that $\theta_\infty \approx a_f/c_\infty$ in Theorem 2 for large values of c_∞.

THEOREM 3. *Suppose* C *and the functions* f_i *are polyhedral convex* (*or affine*), *so that* (P) *is a polyhedral convex* (*or linear*) *program. If* $\inf(\mathrm{P})$ *is finite and the proximal minimization algorithm is executed with exact minimization at each step* (*i.e.*, $\epsilon_k = 0$ *in* (A′)), *one has convergence in finitely many iterations: there is an optimal solution* x^∞ *to* (P) *such that* $x^k = x^\infty$ *for all* k *sufficiently large.*

PROOF. This is a special case of [1, Proposition 8].

4. Dual application: method of multipliers. The augmented Lagrangian has been defined in (1.4)–(1.5) on $C \times R^m \times (0, +\infty)$, but it is convenient to extend it by setting $L(x, y, c) = +\infty$ whenever $x \notin C$. Observe that then (by direct calculation)

$$L(x, y, c) = \max_{\eta} \{l(x, \eta) - (1/2c)|\eta - y|^2\} \quad \text{for all } x, y, c > 0, \tag{4.1}$$

where if $x \in C$ the maximum is attained uniquely for $\eta = Y(x, y, c)$, the simple mapping Y being the one defined in (1.8).

The method of multipliers can be expressed as

$$x^{k+1} \approx \arg\min_{x \in R^n} \phi_k(x) \quad \text{and } y^{k+1} = Y(x^{k+1}, y^k, c_k), \tag{4.2}$$

where ϕ_k is the closed proper convex function defined by

$$\phi_k(x) = L(x, y^k, c_k). \tag{4.3}$$

The initial vector y^0 and the sequence $\{c_k\}$ satisfying (1.1) are given.

We shall investigate this procedure under three stopping criteria:

$$\phi_k(x^{k+1}) - \inf \phi_k \leq \epsilon_k^2/2c_k, \qquad \sum_{k=0}^{\infty} \epsilon_k < \infty, \tag{A''}$$

$$\phi_k(x^{k+1}) - \inf \phi_k \leq (\delta_k^2/2c_k)|y^{k+1} - y^k|^2, \qquad \sum_{k=0}^{\infty} \delta_k < \infty, \tag{B$_1''$}$$

$$\operatorname{dist}(0, \partial\phi_k(x^{k+1})) \leq (\delta_k'/c_k)|y^{k+1} - y^k|, \qquad 0 \leq \delta_k' \to 0, \tag{B$_2''$}$$

where of course

$$|y^{k+1} - y^k|^2 = |Y(x^{k+1}, y^k, c_k) - y^k|^2 = \sum_{i=1}^{m} \max{}^2\{-y_i^k, c_k f_i(x^{k+1})\}. \tag{4.4}$$

If f happens to be strongly convex with modulus b, then the same is true of ϕ_k, and one has the estimate

$$\phi_k(x) - \inf \phi_k \leq (1/2b)\operatorname{dist}^2(0, \partial\phi_k(x)), \tag{4.5}$$

a fact effectively used by Kort and Bertsekas [11] (see also Bertsekas [23]). (The proof of (4.5) is obtained from the fact that for any x, x' and $w \in \partial\phi_k(x)$. one has $\phi_k(x') \geq \phi_k(x) + (x' - x)\cdot w + (1/2b)|x' - x|^2$. Minimize both sides in x' and then maximize the right side over all $w \in \partial\phi_k(x)$.) For more on how to estimate $\phi_k(x) - \inf \phi_k$, see [14, discussion following Theorem 4].

Parallel to Proposition 5, one has (by the same argument) that for the present ϕ_k,

$$\operatorname{dist}(0, \partial\phi_k(x)) = |\operatorname{prox}(-\nabla_x L(x, y^k, c_k) \mid K_C(x))|, \tag{4.6}$$

if the functions f_i are differentiable relative to C (with int $C \neq \emptyset$).

The following result shows the relation between such estimates and the general proximal point algorithm of §2 in the case of $T = T_g$, or in other words

$$P_k(y) = (I + c_k T_g)^{-1}(y) = \arg\max_{y \in R^m} \{g(y) - (1/2c_k)|y - y^k|^2\}. \tag{4.7}$$

(It is assumed here that $g \not\equiv -\infty$, i.e., sup(D) $> -\infty$, so that T_g is indeed maximal monotone.)

PROPOSITION 6. *For P_k as in* (4.7), *ϕ_k as in* (4.3) *and $y^{k+1} = Y(x^{k+1}, y^k, c_k)$, one has*

$$|y^{k+1} - P_k(y^k)|^2/2c_k \leq \phi_k(x^{k+1}) - \inf \phi_k. \tag{4.8}$$

PROOF. The argument is essentially the same as [24, Lemma 4.3, p. 365], but we furnish a direct proof of this estimate in the present notation. Observing from (1.7) that

$$\nabla_y L(x^{k+1}, y^k, c_k) = c_k^{-1}(y^{k+1} - y^k), \tag{4.9}$$

we obtain from the concavity of $L(x, y, c)$ in y the following inequality for arbitrary $\eta \in R^m$:

$$\begin{aligned} L(x^{k+1}, y^k, c_k) &+ c_k^{-1}(y^{k+1} - y^k)\cdot(\eta - y^k) \geq L(x^{k+1}, \eta, c_k) \\ &\geq \inf_x L(x, \eta, c_k) = \inf_x \max_y\{l(x, y) - (1/2c_k)|y - \eta|^2\} \\ &= \max_y \inf_x\{l(x, y) - (1/2c_k)|y - \eta|^2\} \\ &= \max_y\{g(y) - (1/2c_k)|y - \eta|^2\} \geq g(P_k(y^k)) - (1/2c_k)|P_k(y^k) - \eta|^2. \end{aligned} \tag{4.10}$$

The interchange of $\inf_x$ and $\max_y$ in (4.10) is justified by the growth properties in y of the "minimaximand" in question [16, Theorem 37.3]. The same reasoning also yields

$$\inf_x L(x, y^k, c_k) = \max_y\{ g(y) - (1/2c_k)|y - y^k|^2\}$$
$$= g(P_k(y^k)) - (1/2c_k)|P_k(y^k) - y^k|^2. \quad (4.11)$$

Combining (4.10) and (4.11), we get

$$\phi_k(x^{k+1}) - \inf \phi_k = L(x^{k+1}, y^k, c_k) - \inf_x L(x, y^k, c_k)$$
$$\geqslant [|P_k(y^k) - y^k|^2 - |P_k(y^k) - \eta|^2 - 2(y^{k+1} - y^k)\cdot(\eta - y^k)]/2c_k$$
$$= [2(P_k(y^k) - y^{k+1})\cdot(\eta - y^k) - |\eta - y^k|^2]/2c_k.$$

As mentioned, this holds for all $\eta \in R^m$, so that we can take the maximum of the last expression in η to get a sharper inequality. The maximum is attained for $\eta = P_k(y^k) - y^{k+1} + y^k$ and equals the left side of (4.8), proving the assertion.

In stating our main results about the method of multipliers, we recall that a sequence $\{x^k\}$ in C is *asymptotically minimizing* for (P) if

$$\limsup_{k\to\infty} f_i(x^k) \leqslant 0 \quad \text{for } i = 1, \ldots, m, \quad (4.12)$$

and the quantity $\limsup_{k\to\infty} f_0(x^k)$ has the lowest value possible relative to sequences in C satisfying (4.12). The latter value is called the *asymptotic infimum* in (P) and is denoted by asym inf(P).

THEOREM 4. *Suppose* sup(D) $> -\infty$, *and let the method of multipliers be executed with stopping criterion* (A″) *applied to* ϕ_k *in* (4.3). *If the generated sequence* $\{y^k\} \subset R^m_+$ *is bounded, then* $y^k \to y^\infty$, *where* y^∞ *is some optimal solution to* (D), *and* $\{x^k\}$ *is asymptotically minimizing for* (P), *in fact with*

$$f_i(x^{k+1}) \leqslant c_k^{-1}|y^{k+1} - y^k| \to 0 \quad \textit{for } i = 1, \ldots, m, \quad (4.13)$$
$$f_0(x^{k+1}) - \text{asym inf(P)} \leqslant (1/2c_k)[\epsilon_k^2 + |y^k|^2 - |y^{k+1}|^2]. \quad (4.14)$$

The boundedness of $\{y^k\}$ *under* (A″) *is actually equivalent to the existence of an optimal solution to* (D). *Thus it is certain to hold if* (P) *satisfies the Slater condition, or more particularly if* T_g^{-1} *is Lipschitz continuous at the origin* (*cf. Propositions* 1, 2, 3); *in these cases one has* max(D) = inf(P) = asym inf(P).

If $\{y^k\}$ *is bounded and there is an* $\alpha \in R$ *such that the set of all feasible* x *in* (P) *satisfying* $f_0(x) \leqslant \alpha$ *is nonempty and bounded, then the sequence* $\{x^k\}$ *is also bounded, and all of its cluster points are optimal solutions to* (P).

REMARKS. Criterion (A″) with merely $\epsilon_k \to 0$ was in our general convergence analysis in [24] in demonstrating that $\{x^k\}$ would be asymptotically minimizing for (P) if $\{y^k\}$ was bounded and maximizing for (D). The latter property, on the other hand, was shown in [10] to follow if $\sum_{k=0}^\infty \epsilon_k < \infty$. Tretyakov [9] treated only exact minimization in establishing global convergence. The estimate (4.14) is entirely new.

PROOF OF THEOREM 4. Proposition 6 shows that (A″) implies the more general criterion (A) for $T = T_g$. We appeal to [1, Theorem 1] to obtain from this the convergence of $\{y^k\}$ to a solution y^∞ to $0 \in T(y^\infty)$, in other words a particular optimal solution to (D). Relation (4.13) is a consequence of (4.4) and $y^{k+1} - y^k \to 0$. Now we derive (4.14). Observing that

$$L(x, y, c) = f_0(x) + (1/2c)[|Y(x, y, c)|^2 - |y^k|^2] \quad \text{for } x \in C, \quad (4.15)$$

we see

$$\phi_k(x^{k+1}) - f_0(x^{k+1}) = (1/2c_k)[|y^{k+1}|^2 - |y^k|^2]. \tag{4.16}$$

At the same time, equation (4.11) established in the proof of Proposition 6 yields

$$\inf \phi_k \leqslant g(P_k(y^k)) \leqslant \max(\mathrm{D}). \tag{4.17}$$

Combining (4.16) and (4.17) we get

$$f_0(x^{k+1}) - \max(\mathrm{D}) \leqslant \phi_k(x^{k+1}) - \inf \phi_k + (1/2c_k)[|y^k|^2 - |y^{k+1}|^2]$$
$$\leqslant (1/2c_k)[\epsilon_k^2 + |y^k|^2 - |y^{k+1}|^2]. \tag{4.18}$$

But every $x \in C$ satisfies

$$f_0(x) + \sum_{i=1}^{m} y_i^\infty f_i(x) \geqslant \inf_x l(x, y^\infty) = g(y^\infty) = \max(\mathrm{D}),$$

so that $\max(\mathrm{D}) \leqslant$ asym $\inf(\mathrm{P})$. Therefore (4.18) implies (4.14).

Since (A″) implies criterion (A) for T_g, we know from [1, Theorem 1] that the boundedness of the generated sequence $\{y^k\}$ is equivalent to the existence of some y satisfying $0 \in T_g(y)$ i.e., an optimal solution to (D). It is well known that the Slater condition implies the existence of such a solution and $\max(\mathrm{D}) = \inf(\mathrm{P}) =$ asym $\inf(\mathrm{P})$. (The Slater condition is actually equivalent to the boundedness of all the level sets $\{y \mid g(y) \geqslant \beta\}$, $\beta \in R$.) That the Slater condition follows from the Lipschitz continuity of T_g^{-1} at the origin is asserted in Proposition 3(b).

The final statment of Theorem 4 is obtained from (4.13), (4.14) and the fact that, if the set $\{x \in C \mid f_0(x) \leqslant \alpha_0, f_1(x) \leqslant \alpha_1, \ldots, f_m(x) \leqslant \alpha_m\}$ is nonempty and bounded for one choice of $(\alpha_0, \alpha_1, \ldots, \alpha_m) \in R^{m+1}$, then by convexity and lower semicontinuity (and the closedness of C) it is bounded for every $(\alpha_0, \alpha_1, \ldots, \alpha_m) \in R^{m+1}$ [16; 8.3.3, 8.4, 8.7].

The next theorem requires a preliminary result of a fundamental nature.

PROPOSITION 7. *Let $x \in C$, $y \in R^m$, $c > 0$, and let $u = c^{-1}(y - y')$ where $y' = Y(x, y, c)$. Then for any $v \in \partial_x L(x, y, c)$ one has $(v, u) \in T_l(x, y')$ and consequently the estimate*

$$\operatorname{dist}(T_l(x, y'), (0, 0)) \leqslant |(v, u)|. \tag{4.19}$$

PROOF. It is elementary from the definitions of l and L and the calculus of subgradients [16, §23] that

$$\partial_x L(x, y, c) = \partial f_0(x) + \sum_{i=1}^{m} y_i' \partial f_i(x) = \partial_x l(x, y') \tag{4.20}$$

(where for the convenience of the moment $f_i(x)$ has been interpreted for $x \notin C$ as $+\infty$). (The fact that $\partial_x \psi(f_i(x), y_i, c) = y_i' \partial f_i(x)$ is obtained from $\partial\psi(t, y_i, c)/\partial t = \max\{0, y_i + ct\}$ by calculating the one-sided directional derivatives of $\psi(f_i(x), y_i, c)$ with respect to x.) Thus we have $v \in \partial_x l(x, y')$. On the other hand, as noted at the beginning of this section, the maximum in (4.1) is found at $\eta = Y(x, y, c) = y'$. Hence

$$0 \in \partial_y (l(x, \eta) - (1/2c)|\eta - y|^2)\big|_{\eta = y'} = \partial_y l(x, y') - c^{-1}(y' - y).$$

Therefore $-u \in \partial_y l(x, y')$, so that $(v, u) \in T_l(x, y')$ as claimed.

COROLLARY. *In the method of multipliers, one has the estimate*

$$\operatorname{dist}(T_l(x^{k+1}, y^{k+1}), (0, 0)) \leqslant \left[\operatorname{dist}^2(0, \partial\phi_k(x^{k+1})) + c_k^{-2}|y^{k+1} - y^k|^2\right]^{1/2}. \tag{4.21}$$

THEOREM 5. *Suppose* $\sup(\mathrm{D}) > -\infty$ *and let the method of multipliers be executed with stopping criterion* (B_1'') *applied to* ϕ_k *in* (4.3). *If* T_g^{-1} *is Lipschitz continuous at the* *origin with modulus* a_g (*cf. Propositions* 1, 2, 3) *and* $\{y^k\}$ *is bounded* (*cf. Theorem* 4), *then* $y^k \to \bar{y}$, *where* $\bar{y}$ *is the unique optimal solution to* (D) (*in fact, the unique Kuhn-Tucker vector for* (P), *because* max(D) = inf(P) = asym inf(P)), *and*

$$|y^{k+1} - \bar{y}| \leq \theta_k |y^k - \bar{y}| \quad \text{for all } k \text{ sufficiently large,} \tag{4.22}$$

where

$$\theta_k = \left[a_g \left(a_g^2 + c_k^2 \right)^{-1/2} + \delta_k \right] (1 - \delta_k)^{-1} \to \theta_\infty = a_g \left(a_g^2 + c_\infty^2 \right)^{-1/2} < 1.$$

Moreover, the conclusions of Theorem 4 *about* $\{x^k\}$ *are valid with* $\epsilon_k = \delta_k |y^{k+1} - y^k|$ *in* (4.14).

If in addition to (B_1'') *and the condition on* T_g^{-1} *one has* (B_2'') *and the stronger condition that* T_l^{-1} *is Lipschitz continuous at the origin with modulus* a_l ($\geq a_g$), *then* $x^k \to \bar{x}$, *where* $\bar{x}$ *is the unique optimal solution to* (P), *and one has*

$$|x^{k+1} - \bar{x}| \leq \theta_k' |y^{k+1} - y^k| \quad \text{for all } k \text{ sufficiently large} \tag{4.23}$$

where $\theta_k' = a_l(1 + \delta_k')/c_k \to \theta_\infty' = a_l/c_\infty$.

REMARK. Criterion (B_2'') is essentially what has been used by other authors in obtaining convergence results of the caliber of (4.22) and (4.23) for the method of multipliers (local convergence in the nonconvex case) (cf. Polyak and Tretyakov [6], Bertsekas [7], [8]), but under the more restrictive assumption that the strong second-order conditions for optimality hold in (P) (cf. Proposition 2), which precludes any real role for the constraint $x \in C$. Of course, (B_2'') with $\sum_{k=0}^{\infty} \delta_k' < \infty$ implies (B_1'') via (4.5) if f_0 is strongly convex with modulus b, or simply from the strong second-order conditions if (x^k, y^k) stays sufficiently near $(\bar{x}, \bar{y})$ and c_k is large enough, due to the well-known strong convexity of ϕ_k near $\bar{x}$ in the latter case.

PROOF OF THEOREM 5. The first part of the theorem is obvious from Theorem 4 and the general result [1, Theorem 2], since in view of Proposition 6 we are, in effect, executing the proximal point algorithm for T_g under (B). To establish the second part, we recall that if T_l^{-1} is Lipschitz continuous at (0, 0), then $T_l^{-1}(0, 0) = \{(\bar{x}, \bar{y})\}$, where $\bar{x}$ and $\bar{y}$ are the unique optimal solutions to (P) and (D), respectively. Therefore

$$|(x^{k+1}, y^{k+1}) - (\bar{x}, \bar{y})| \leq a_l \operatorname{dist}(T(x^{k+1}, y^{k+1}), (0, 0)),$$

if the distance in question is sufficiently small. Putting (B_2'') into the estimate of the corollary above, we obtain

$$|(x^{k+1} - \bar{x}, y^{k+1} - \bar{y})| \leq a_l (\delta_k'^2 + 1)^{1/2} c_k^{-1} |y^{k+1} - y^k|,$$

and this yields (4.23).

THEOREM 6. *Suppose* C *and the functions* f_i *are polyhedral convex* (*or affine*). *If* inf(P) *is finite and the method of multipliers is executed with exact minimization at each step* (*i.e.*, $\epsilon_k = 0$ *in* (A'')), *one has convergence in finitely many iterations*: *there is an optimal solution* y^∞ *to* (D) *such that for all* k *sufficiently large one has* $y^k = y^\infty$, *and* x^k *is optimal for* (P).

PROOF. As regards the sequence $\{y^k\}$, this result is, like Theorem 3, an application of [1, Proposition 8], namely to $-g$. The latter function is polyhedral convex when C and the functions f_i have this property, and then max(D) = min(P) [16, Theorems 29.2, 30.4(e) (f)]. The ultimate optimality of $\{x^k\}$ is seen from the estimates (4.13) and (4.14).

REMARK. The finite convergence in Theorem 6 for linear programming problems was discovered by Polyak and Tretyakov [25] and independently later in somewhat more general terms by Bertsekas [26].

5. Minimax application: the proximal method of multipliers. Like the ordinary method of multipliers, the proximal method of multipliers can be described by (4.2); however, instead of (4.3) we have

$$\phi_k(x) \equiv L(x, y^k, c_k) + (1/2c_k)|x - x^k|^2, \tag{5.1}$$

so that ϕ_k is strongly convex with modulus $1/c_k$. The initial pair (x^0, y^0) and the sequence $\{c_k\}$ satisfying (1.1) are given. We treat as stopping criteria:

$$\operatorname{dist}(0, \partial\phi_k(x^{k+1})) \leq \epsilon_k/c_k, \qquad \sum_{k=0}^{\infty} \epsilon_k < \infty, \tag{A'''}$$

$$\operatorname{dist}(0, \partial\phi_k(x^{k+1})) \leq (\delta_k/c_k)|(x^{k+1}, y^{k+1}) - (x^k, y^k)|, \qquad \sum_{k=0}^{\infty} \delta_k < \infty. \tag{B'''}$$

If the functions f_i are differentiable relative to C (with int $C \neq \varnothing$), we have

$$\operatorname{dist}(0, \partial\phi_k(x)) = \left|\operatorname{prox}\left(-\nabla f_0^k(x) - \sum_{i=1}^{m} y_i^k \nabla f_i(x) | K_C(x)\right)\right|, \tag{5.2}$$

corresponding to a simplified form of Proposition 5.

We apply the general convergence theory of [1] to the algorithm of §2 for

$$\begin{aligned} P_k(x, y) &= (I + c_k T_l)^{-1}(x, y) \\ &= \arg \underset{x \in R^n\; y \in R^m}{\operatorname{minimax}} \{l(x, y) + (1/2c_k)|x - x^k|^2 - (1/2c_k)|y - y^k|^2\}. \end{aligned} \tag{5.3}$$

The key to this application is the following.

PROPOSITION 8. *For P_k as in* (5.3), *ϕ_k as in* (5.1) *and $y^{k+1} = Y(x^{k+1}, y^k, c_k)$, one has*

$$|(x^{k+1}, y^{k+1}) - P_k(x^k, y^k)| \leq c_k \operatorname{dist}(0, \partial\phi_k(x^{k+1})). \tag{5.4}$$

PROOF. Formula (5.1) yields the subdifferential relation

$$\partial\phi_k(x^{k+1}) = \partial_x L(x^{k+1}, y^k, c_k) + c_k^{-1}(x^{k+1} - x^k).$$

For any $w \in \partial\phi_k(x^{k+1})$, we therefore have $w + c_k^{-1}(x^k - x^{k+1}) \in \partial_x L(x^{k+1}, y^k, c_k)$, and we can conclude from Proposition 7 that

$$(w + c_k^{-1}(x^k - x^{k+1}), c_k^{-1}(y^k - y^{k+1})) \in T_l(x^{k+1}, y^{k+1}), \tag{5.5}$$

or in other words $(c_k w + x^k, y^k) \in (I + c_k T_l)(x^{k+1}, y^{k+1})$. Therefore, $(x^{k+1}, y^{k+1}) = P_k(c_k w + x^k, y^k)$. But P_k, as the proximal mapping associated with the maximal monotone operator $c_k T_l$, is nonexpansive [1, §2]. Thus

$$|(x^{k+1}, y^{k+1}) - P_k(x^k, y^k)| \leq |(c_k w + x^k, y^k) - (x^k, y^k)| \leq c_k|w|.$$

Since this holds for all $w \in \partial\phi_k(x^{k+1})$, we have (5.4).

THEOREM 7. *Let the proximal method of multipliers be executed with stopping criterion* (A''') *applied to ϕ_k in* (5.1). *If the generated sequence $\{(x^k, y^k)\}$ in $C \times R_+^m$ is bounded, then $(x^k, y^k) \to (x^\infty, y^\infty)$, where x^∞ is a particular optimal solution to* (P) *and y^∞ is a particular optimal solution to* (D); min(P) = max(D). *Furthermore*, (4.13) *holds* *with* $y_i^{k+1}f_i(x^{k+1}) \to 0$ *for* $i = 1, \ldots, m$, *and*

$$f_0(x^{k+1}) - \min(\mathrm{P}) \leqslant c_k^{-1}|x^{k+1} - x^\infty|(\epsilon_k + |x^{k+1} - x^k|) - \sum_{i=1}^m y_i^{k+1}f_i(x^{k+1}) \to 0, \tag{5.6}$$

$$f_0(x^{k+1}) - \min(\mathrm{P}) \geqslant -c_k^{-1}|y^\infty| \cdot |y^{k+1} - y^k| \to 0. \tag{5.7}$$

The boundedness of $\{(x^k, y^k)\}$ *under* (A‴) *is actually equivalent to the existence of optimal solutions to* (P) *and* (D) *with* min(P) = max(D), *and thus it holds in particular if* T_l^{-1} *is Lipschitz continuous at* (0, 0) (*cf. Propositions* 1, 2, 3).

PROOF. From Proposition 8 and [1, Theorem 1] we know that (x^k, y^k) converges to some (x^∞, y^∞) satisfying $(0, 0) \in T_l(x^\infty, y^\infty)$. The latter relation means that (x^∞, y^∞) is a saddle point of l, and hence x^∞ is optimal in (P), y^∞ is optimal in (D), and min(P) = max(D). Since $y^k \to y^\infty$, we have $c_k^{-1}(y^{k+1} - y^k) \to 0$. But $y_i^{k+1} \geqslant 0$ and $f_i(x^{k+1}) \leqslant c_k^{-1}(y_i^{k+1} - y_i^k)$ by the definition of y^{k+1}, so (4.13) holds as claimed and

$$\limsup_{k\to\infty} y_i^{k+1}f_i(x^{k+1}) \leqslant y_i^\infty \lim_{k\to\infty} c_k^{-1}(y_i^{k+1} - y_i^k) = 0,$$

$$\liminf_{k\to\infty} y_i^{k+1}f_i(x^{k+1}) = y_i^\infty \liminf_{k\to\infty} f_i(x^{k+1}) \geqslant y_i^\infty f_i(x^\infty) = 0,$$

the second inequality holding by the lower semicontinuity of f_i and the fact that (x^∞, y^∞) is a saddle point of l. Thus $y_i^{k+1}f_i(x^{k+1}) \to 0$.

We derive (5.6) and (5.7) by taking any $w \in \partial\phi_k(x^{k+1})$ such that $|w| \leqslant \epsilon_k/c_k$, as exists under (A‴), and working with the corresponding relation (5.5) established in the proof of Proposition 8. This can be written as the two subdifferential inequalities:

$$l(x, y^{k+1}) \geqslant l(x^{k+1}, y^{k+1}) + (x - x^{k+1})\cdot(w + c_k^{-1}(x^k - x^{k+1})) \quad \text{for all } x, \tag{5.8}$$

$$l(x^{k+1}, y) \leqslant l(x^{k+1}, y^{k+1}) + c_k^{-1}(y - y^{k+1})\cdot(y^{k+1} - y^k) \quad \text{for all } y \tag{5.9}$$

(where $x^{k+1} \in C, y^{k+1} \in R_+^m$). Applying (5.8) to $x = x^\infty$ and using the saddle-point property of (x^∞, y^∞), we get

$$\begin{aligned} l(x^{k+1}, y^{k+1}) - l(x^\infty, y^\infty) &\leqslant l(x^{k+1}, y^{k+1}) - l(x^\infty, y^{k+1}) \\ &\leqslant |x^{k+1} - x^\infty|(|w| + c_k^{-1}|x^k - x^{k+1}|), \end{aligned}$$

which translates into (5.6) on the observation that

$$l(x^{k+1}, y^{k+1}) - l(x^\infty, y^\infty) = f_0(x^{k+1}) + \sum_{i=1}^m y_i^{k+1}f_i(x^{k+1}) - \min(\mathrm{P}).$$

Similarly, from (5.9) we get

$$\begin{aligned} l(x^\infty, y^\infty) &\leqslant l(x^{k+1}, y^\infty) \leqslant l(x^{k+1}, y^{k+1}) + c_k^{-1}(y^\infty - y^{k+1})\cdot(y^{k+1} - y^k) \\ &= f_0(x^{k+1}) + c_k^{-1}\left[y^\infty \cdot (y^{k+1} - y^k) - \sum_{i=1}^m y_i^{k+1}(y_i^{k+1} - y_i^k - c_k f_i(x^{k+1}))\right] \\ &\leqslant f_0(x^{k+1}) + c_k^{-1} y^\infty \cdot (y^{k+1} - y^k), \end{aligned}$$

whence (5.7). (These calculations make use of (1.7) again.)

The necessary and sufficient condition for boundedness of $\{(x^k, y^k)\}$ furnished by

[1, Theorem 1] is the existence of $(\bar{x}, \bar{y})$ satisfying $(0, 0) \in T_l(\bar{x}, \bar{y})$, and this is equivalent, as seen in §2, to the relation min(P) = max(D).

REMARK. The estimate (5.6) is of practical import, for example, if C is bounded with known diameter, since then $|x^{k+1} - x^{\infty}| \leqslant \operatorname{diam} C$. Also, if $\tilde{x} \in C$ is such that $f_i(\tilde{x}) < 0$ for $i = 1, \ldots, m$, and α is any real number such that $\alpha \leqslant$ max(D) = min(P) (as might be obtained via (5.6)) one has

$$y^{\infty} \in \{ y \mid g(y) \geqslant \alpha \} \subset \left\{ y \geqslant 0 \mid f_0(\tilde{x}) + \sum_{i=1}^{m} y_i f_i(\tilde{x}) \geqslant \alpha \right\}$$

and consequently

$$|y^{\infty}| \leqslant \max_{i=1}^{m} (f_0(\tilde{x}) - \alpha / |f_i(\tilde{x})|).$$

THEOREM 8. *Let the proximal method of multipliers be executed with stopping criterion* (B‴) *applied to* ϕ_k *in* (5.1). *If* T_l^{-1} *is Lipschitz continuous at the origin with modulus* a_l (*cf. Proposition* 2) *and* $\{(x^k, y^k)\}$ *is bounded* (*cf. Theorem* 7), *then* $(x^k, y^k) \to (\bar{x}, \bar{y})$ *where* $\bar{x}$ *is the unique optimal solution to* (P) *and* $\bar{y}$ *is the unique optimal solution to* (D); min(P) = max(D). *Furthermore, one has*

$$|(x^{k+1}, y^{k+1}) - (\bar{x}, \bar{y})| \leqslant \theta_k |(x^k, y^k) - (\bar{x}, \bar{y})| \quad \textit{for all } k \textit{ sufficiently large}, \tag{5.10}$$

where

$$\theta_k = \left[a_l (a_l^2 + c_k^2)^{-1/2} + \delta_k \right] (1 - \delta_k)^{-1} \to \theta_{\infty} = a_l (a_l^2 + c_{\infty}^2)^{-1/2} < 1,$$

and the conclusions of Theorem 7 *are valid with* $\epsilon_k = \delta_k |(x^{k+1}, y^{k+1}) - (x^k, y^k)|$.

PROOF. Again, this is just a matter of applying [1, Theorem 2] to the present context of Proposition 8 and Theorem 7.

6. Concluding comments. We have only dealt with inequality constraints, but equations can be treated in the same way. For an equality constraint, the function f_i is affine, and in place of $\psi(f_i(x), y_i, c)$ in the augmented Lagrangian (1.4) one simply has the term $y_i f_i(x) + (c/2) f_i(x)^2$. Correspondingly, $Y_i(x, y, c)$ in (1.8) is just $y_i + c f_i(x)$. The Slater condition must be reinterpreted as asserting the existence of a feasible solution $\tilde{x} \in \operatorname{int} C$ which satisfies all the inequality constraints in the problem strictly. In the strong second-order conditions for optimality, the multipliers $\bar{y}_i$ corresponding to the equality constraints are not required to be positive. *With these extensions, all the theorems and propositions in this paper remain valid for problems with mixed equality and inequality constraints.* (Incidentally, the closedness of C could be relaxed by working in the framework of [16, §28].)

Generalization to infinite dimensions is also easy, since the fundamental theory of the proximal point algorithm in [1] is cast in the framework of Hilbert spaces. If the primal vector x ranges over such a space X, but still there are only finitely many constraints (so that R^m is still the multiplier space), only minor changes are required. The extended interpretation just given of the Slater condition suffices, while the assumption that for some α the set

$$\{ x \in C \mid f_0(x) \leqslant \alpha \quad \text{and } x \text{ is feasible for (P)} \} \tag{6.1}$$

is nonempty and bounded, must be replaced, where it occurs, by the assumption that for some α and β the set

$$\bigcup_{|u| < \beta} \{ x \in C \mid f_0(x) \leqslant \alpha \quad \text{and } x \text{ is feasible for } (P(0, u)) \}, \tag{6.2}$$

is nonempty and bounded (cf. [18, Theorem 18′(e)]). The inequality in part (c) of the strong second-order conditions must be sharpened to $w \cdot Hw \geqslant \lambda |w|^2$ for some $\lambda > 0$ as in the article of Wierzbicki and Kurcyusz [27]. *Then again all the theorems and propositions remain valid.* (One has weak convergence under criterion (A) and its realizations, but strong convergence under (B) and its realizations.) (But Theorems 3 and 6 are of course intrinsically finite dimensional.)

Models with infinitely many constraints present a more serious challenge. The multiplier space R^n must be replaced by a Hilbert space U (rather than some other kind of Banach space), and the perturbed problem $(P(v, u))$ in §2 must therefore be generalized in terms of a perturbation vector u in this same space U. The Slater condition can be taken as the property that for every $u \in U$ there exists $\epsilon > 0$ such that $\inf(P(0, \epsilon u)) < +\infty$, i.e., $(P(0, \epsilon u))$ has a feasible solution (cf. [18, Theorem 18(c)]). The nonemptiness and boundedness of (6.2) can be substituted, as above, for that of (6.1). Although no suitable version of the second-order conditions for optimality is known for this case, so that Proposition 2 (like Theorems 3 and 6) must be left out, the rest of the theorems and propositions remain in force under these interpretations.

The real difficulty in this context, however, is that the Slater condition is too stringent for most applications, and indeed the existence of multipliers belonging to an infinite-dimensional Hilbert space is very hard to guarantee in terms of any reasonable assumption on the given problem.

The results can also be extended to generalized convex programming problems in the sense of [16], [18]. In this case, l is any "closed" proper convex-concave function on $X \times U$ (product of Hilbert spaces), and the augmented Lagrangian L is defined by (4.1), the maximum in (4.1) being attained uniquely at $Y(x, y, c)$. The functions f and g are defined by $f(x) = \sup_y l(x, y)$ and $g(y) = \inf_x l(x, y)$. Problem (P) consists of minimizing f over X, while (D) consists of maximizing g over U. The property in [18, Theorem 18(c)] can serve as the Slater condition, while the one in [18, Theorem 18′(e)] can serve in place of assuming the boundedness of (6.2). The generalized forms of all the theorems and propositions are valid in this context, except for Proposition 2 (no second-order conditions being available) and the estimates (4.13), (4.14), (5.6) and (5.7) (which would have to be reformulated somehow). (In (3.10), one can put f in place of f_0. In Theorem 7, it is nevertheless true that $l(x^k, y^k) \to l(x^\infty, y^\infty)$; see [1, Theorem 5]. The hypothesis of Theorem 3 is that X is finite dimensional and f is polyhedral convex, while in Theorem 6 it is that U is finite dimensional and g is polyhedral concave.)

For the method of multipliers and the proximal method of multipliers to be of interest for computation for generalized convex programs, it is essential that l be such that the maximization in (4.1) can be carried out in closed form (rather than just numerically), giving a manageable expression for the augmented Lagrangian L.

References

[1] Rockafellar, R. T. (1976). Monotone Operators and the Proximal Point Algorithm. *SIAM J. Control Opt.*, **14** No. 5.

[2] Falk, J. E. (1967). Lagrange Multipliers and Nonlinear Programming. *J. Math. Anal. Appl.* **19** 141–159.

[3] Hestenes, M. R. (1969). Multiplier and Gradient Methods. *J. Optimization Theory Appl.* **4** 303–320.

[4] Powell, M. J. D. (1969). A Method for Nonlinear Constraints in Minimization Problems. *Optimization*, R. Fletcher (ed.), Academic Press, 283–298.

[5] Bertsekas, D. P. (1975). Multiplier Methods: A Survey. *Proceedings of IFAC 6th Triennial World Congress, Part 1B.*

[6] Polyak, B. T. and Tretyakov, N. V. (1973). The Method of Penalty Bounds for Constrained Extremum Problems. *Zh. Vych. Mat i Mat. Fiz* **13** 34–46 = *U.S.S.R. Computational Math. and Math. Phys.* **13** 42–58.

[7] Bertsekas, D. P. (1975) On Penalty and Multiplier Methods for Constrained Minimization. *SIAM J. Control* **13**.
[8] Berteskas, D. P. (1973). Convergence Rate of Penalty and Multiplier Methods. *Proc. IEEE Conf. on Decision and Control*, San Diego, California, 260–264.
[9] Tretyakov, N. V. (1973). The Method of Penalty Estimates for Problems of Convex Programming. *Economics and Math. Methods* **9** 525–540 (Russian).
[10] Rockafellar, R. T. (1973). The Multiplier Method of Hestenes and Powell Applied to Convex Programming. *J. Optimization Theory Appl.* **12** 555–562.
[11] Kort, B. W. and Bertsekas, D. P. (1975). Combined Primal-Dual and Penalty Methods for Convex Programming. *SIAM J. Control Opt.* **13**.
[12] Golshtein, E. G. and Tretyakov, N. V. (1974). Modified Lagrangian Functions. *Economics and Math. Methods* **10** 568–591 (Russian).
[13] Rockafellar, R. T. (1974). Augmented Lagrange Multiplier Functions and Duality in Nonlinear Programming. *SIAM J. Control* **12** 268–285.
[14] ———, (1976). Solving a Nonlinear Programming Problem by Way of a Dual Problem. *Symposia Mathematica. XXVII*
[15] ———, (1970). On the Maximal Monotonicity of Subdifferential Mappings. *Pacific J. Math.* **33** 209–216.
[16] ———, (1970). *Convex Analysis*. Princeton University Press.
[17] ———, (1970). Monotone Operators Associated with Saddle Functions and Minimax Problems. In *Nonlinear Functional Analysis, Part 1* (F. E. Browder, ed.), Symposia in Pure Math. **18**, Amer. Math. Soc. 397–407.
[18] ———, (1974). *Conjugate Duality and Optimization*. Regional Conference Series in Applied Math. no. 16, SIAM Publications.
[19] ———, (1969). Local Boundedness of Nonlinear Monotone Operators. *Michigan Math. J.* **16** 397–407.
[20] ———, (1970). On the Maximality of Sums of Nonlinear Monotone Operators. *Trans. Amer. Math. Soc.* **149** 75–88.
[21] Hestenes, M. R. (1966). *Calculus of Variations and Optimal Control Theory*. Wiley.
[22] Reshetnyak, Yu. G. (1968). Generalized Derivatives and Differentiability Almost Everywhere. *Mat. Sbornik* **75** 323–334 = *Math. USSR-Sb.* **4** 293–302.
[23] Bertsekas, D. P. (1975). On the Method of Multipliers for Convex Programming. *IEEE Transactions on Automatic Control*. 385–388.
[24] Rockafellar, R. T. (1973). A Dual Approach to Solving Nonlinear Programming Problems by Unconstrained Optimization. *Math. Programming* **5** 354–373.
[25] Polyak, B. T. and Tretyakov, N. V. (1972). On an Iterative Method of Linear Programming and Its Economic Interpretation. *Economics and Math. Methods* **8** 740–751 (Russian).
[26] Bertsekas, D. P., (1975). Necessary and Sufficient Conditions for a Penalty Method to Be Exact, *Math. Programming* **9** 87–99.
[27] Wierzbicki, A. P. and Kurcyusz, S. (to appear). Projection on a Cone: Generalized Penalty Functionals and Duality Theory. *SIAM J. Control. Opt.*

DEPARTMENT OF MATHEMATICS, GN-50, UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON 98195