August 1986 LIDS-P-1596

# THE NONLINEAR PROXIMAL POINT ALGORITHM AND MULTIPLIER METHODS

JAVIER LUQUE

Department of Mathematical Sciences
University of North Florida
4567 St. Johns Bluff Rd. So.
Jacksonville, FL 32216

August 13, 1986

Submitted to Mathematics of Operations Research

# THE NONLINEAR PROXIMAL POINT ALGORITHM AND MULTIPLIER METHODS*

Abstract. A new Augmented Lagrangian is introduced. The corresponding method of multipliers is equivalent to the application of the proximal point algorithm (NPA) to the subdifferential map of the ordinary dual function. Assuming exact minimization of the Augmented Lagrangian, results on the global and asymptotic convergence of the method of multipliers follow from the theory of the NPA. The sequence of dual variables monotonically ascends the ordinary dual function and converges to some element of the (not necessarily compact) Lagrange multiplier set. The sequence of primal variables is asymptotically optimal, and all of its cluster points are optimal solutions. The growth properties of the subdifferential maps of both, the ordinary dual function, and of the dual of the penalty function, characterize the asymptotic convergence. If those growths are bounded by power functions with exponents $t$, $s > 0$, respectively, with $st \geq 1$, convergence is linear, superlinear, or in finitely many steps --which can be reduced to one-- as $st = 1$, $st > 1$, or $t = \infty$. If $st = 1$ and the penalty parameter sequence grows to $\infty$, superlinear convergence obtains. Conditions implying sublinear convergence are given. The speed of approach to the common optimal value of both primal and ordinary dual programs is estimated.

Key words. Convex program, multiplier method, nonlinear proximality algorithm, global convergence, asymptotic convergence.

*This research is part of the author's dissertation written under the supervision of Prof. Dimitri P. Bertsekas of the Massachusetts Institute of Technology. It was supported by the ITP Foundation, Madrid, Spain, and the National Science Foundation under grant ECS 8217668.

1. Introduction. This paper deals with an application of the nonlinear proximal point algorithm (NPA) in convex programming. (For more details on the NPA, see Luque (1984a, 1986a), for the "linear" proximal point algorithm (PPA), see Rockafellar (1976a) and Luque (1984b).) Let us consider the convex program

(P) minimize $f_0(x)$, subject to $f_i(x) \leqq 0$ $(i = 1,\ldots,m)$,

where $f_0, f_1,\ldots,f_m : \mathbb{R}^n \to \mathbb{R}$ are convex functions. Its (ordinary) Lagrangian and dual functions are defined, respectively, by

$$L(x,y) = f_0(x) + (f(x),y) \qquad \text{for all } y \in \mathbb{R}^m_+,$$

$$g(y) = \inf_{x\in\mathbb{R}^n} L(x,y) \qquad \text{for all } y \in \mathbb{R}^m_+$$

where $f(x) = (f_1(x),\ldots,f_m(x))$, and $(\cdot,\cdot)$ is the inner product of $\mathbb{R}^m$.

The Lagrange multipliers of (P), the "primal" program, can be characterized as the optimal solutions of the dual program

(D) maximize $g(y)$, subject to $y \in \mathbb{R}^m_+$.

The method of multipliers of Hestenes and Powell for the solution of (P), consists in the sequential constrained minimization of an Augmented Lagrangian

$$L_c(x,y) = f_0(x) + P_c(x,y),$$

where $c > 0$ is the penalty parameter, and $P_c$ is the penalizing or augmenting function. At each cycle, one is given $c_k > 0$ and $y^k$. By (approximately) minimizing $L_{c_k}(\cdot,y^k)$ over $\mathbb{R}^n$ one then obtains $x^k$. Then

$y^k$ is updated by means of some explicit formula depending on $c_k$ and $x^k$, a new $c_k$ is selected and the process is repeated. The limit points of $\{x^k\}$ and $\{y^k\}$ are optimal solutions of (P) and (D) respectively.

Rockafellar (1973, 1976b) has shown that the method of multipliers with quadratic-like penalty function $P_c$, is the application of the PPA to the solution of (D). Thus one attempts to solve $0 \in \partial(-g)$, where $\partial(-g)$ is the subdifferential of the convex function $-g$. This equivalence arises through the fact that the identity map I is the gradient map of the convex function $\frac{1}{2}|\cdot|^2$, $|\cdot|$ being the euclidean norm of $\mathbb{R}^m$.

Kort and Bertsekas (1972, 1973, 1976) have shown that by choosing suitable nonquadratic penalty functions $P_c$, any (Q-) order of convergence (Ortega and Rheinboldt 1970) can be attained. These methods were the original motivation for the development of the NPA, the objective being to develop an algorithm that would stand in the same relation to nonquadratic methods of multipliers, as the PPA does with respect to the quadratic method of multipliers.

In section 2, Rockafellar's (1970, 1974, 1976c) perturbational theory of duality is quickly reviewed. We begin by stating the basic facts about ordinary duality, and then proceed to introduce another duality scheme. This latter scheme is at the core of the method of multipliers developed here.

The relation between the method of multipliers and the NPA is analyzed in section 3. It is shown that the method of multipliers --based on the Augmented Lagrangian introduced in section 2-- is equivalent to the NPA applied to the subdifferential map of the ordinary dual function. This is in complete analogy with the parallelism between the

PPA and the method of multipliers for a quadratically augmented Lagrangian.

The last section studies the global and asymptotic convergence of the method of multipliers introduced in sections 2 and 3. Every result in this section depends on the assumption of exact minimization of the Augmented Lagrangian. The reasons for this are twofold. First, using nonquadratic penalty functions introduces fundamental difficulties when obtaining error estimates, if one wants to use monotonicity methods. Second, if one gives up the use of monotonicity and instead uses the fact that the objects under study are numerical functions, then compactness of their level sets could be used to bypass those difficulties. But this would imply a replication of the analysis performed by Kort and Bertsekas (1972, 1973, 1976).

Under the exact minimization assumption and without assuming that the Lagrange multiplier set is compact, we prove that the sequence of dual variables produced by our method of multipliers converges to a (not necessarily unique) Lagrange multiplier. It is also shown that the corresponding sequence of primal variables is asymptotically optimal and each of its cluster points is in fact an optimal solution of the convex program. This extends the already known results on the exact method of multipliers with nonquadratic penalties (ibid.) in two directions. First, the compactness assumption on the set of Lagrange multipliers is dispensed with. Second, it is shown that the sequence of dual variables is convergent to a vector, rather than to the Lagrange multiplier set.

The remainder of the section studies the asymptotic convergence properties of the sequence of dual variables. Using the general theory of the NPA (Luque 1984a, 1986a), sufficient conditions for the linear and superlinear convergence to zero of $d(y^k,\bar{Y}) = \min\{|y^k - y| \; |y \in \bar{Y}\}$, where $\{y^k\}$ is the sequence of dual variables generated by the algorithm and $\bar{Y}$ the Lagrange multiplier set, are given. These results corroborate those already obtained by Kort and Bertsekas (1976). We also study conditions resulting in convergence in finitely many steps (cf. Bertsekas 1975, 1982). Finally we study two subjects that to the best of the author's knowledge have no counterpart in the literature. The first of these is a sufficient condition for the sublinear convergence of $\{d(y^k,\bar{Y})\}$. The second is an estimate of the speed at which $\{g_0(y^k)\}$ approaches the common optimal value of both the primal and dual problems.

2. Duality relations. Let C be a nonempty convex subset of $\mathbb{R}^n$, and let $f_i : \mathbb{R}^n \to (-\infty,\infty]$ be a proper closed convex function for $i = 0,1,\ldots,m$. Let us consider the convex program

(P) $\min \{f_0(x) \mid x \in C,\ f_i(x) \leqq 0 \text{ for } i = 1,2,\ldots,m\}$.

In formulating such a problem, we will adhere to the conventions set forth in Rockafellar (1970 §28)

(a) $\operatorname{dom} f_0 = C$,

(b) $\operatorname{dom} f_i \supseteqq C$, $\operatorname{ri}(\operatorname{dom} f_i) \supseteqq \operatorname{ri} C$, $i = 1,\ldots,m$,

where ri means relative interior (see Rockafellar 1970 §6). A vector $x \in \mathbb{R}^n$ is a feasible solution of (P) iff $x \in C$ and $f_i(x) \leqq 0$, $i = 1,\ldots,m$. The set of such feasible solutions is

$$C_0 = \{x \in C \mid f_i(x) \leqq 0,\ i = 1,\ldots,m\}.$$

Let $\psi(\cdot;C_0)$ denote the convex indicator function of the set $C_0$, i.e., $\psi(x;C_0) = 0$ iff $x \in C_0$, otherwise $\psi(x;C_0) = \infty$. The convex function $f_0 + \psi(\cdot;C_0)$ is the objective function of (P). It is closed because so are $f_0, f_1, \ldots, f_m$. It never takes the value $-\infty$, thus it is proper iff $C_0 \neq \emptyset$. Minimizing $f_0 + \psi(\cdot;C_0)$ over $\mathbb{R}^n$ is equivalent to minimizing $f_0$ over $C_0$, i.e., (P). The infimum of $f_0 + \psi(\cdot;C_0)$, which belongs to $\overline{\mathbb{R}}$, is the optimal value of (P). If $f_0 + \psi(\cdot;C_0) \not\equiv \infty$, i.e., if $C_0 \neq \emptyset$, the vectors where such infimum is attained are the optimal solutions of (P).

We will now quickly review Rockafellar's (1970 ,1974,1976c) perturbational theory of duality. Let U,Y be two real vector spaces in duality via the nondegenerate bilinear form $(\cdot,\cdot) : U \times Y \to \mathbb{R}$ (Bourbaki 1981) Rockafellar's starting point is the bifunction (i.e., a function of two arguments) $F:\mathbb{R}^n \times U \to \overline{\mathbb{R}}$ such that $F(\cdot;0) = f_0 + \psi(\cdot;C_0)$. For $u \neq 0$, $F(\cdot,u)$ can be thought of as the objective function of a (generalized) convex program obtained by perturbing (P), the perturbations being parameterized by the vector $u \in U$.

The Lagrangian function associated with (P) via F, $L:\mathbb{R}^n \times Y \to \overline{\mathbb{R}}$ is defined by

$$L(x,y) = \inf \{F(x,u) - (u,y) | u \in U\},$$

Since $L(x,\cdot)$ is the pointwise infimum of a collection of affine functions, it is closed and concave.

Let X, V be two real vector spaces paired by a nondegenerate bilinear function $(\cdot,\cdot):X \times V \to \mathbb{R}$. Let $f:X \to \overline{\mathbb{R}}$, its Fenchel conjugate $f^*:V \to \overline{\mathbb{R}}$ is defined by

$$f^*(v) = \sup \{(x,v) - f(x) | x \in X\}.$$

$f^*$ is always a closed convex function (Rockafellar 1974, th 5). Similarly, the function $f_*: V \to \overline{\mathbb{R}}$ defined by (ibid. p 18)

$$f_*(v) = \inf \{(x,v) - f(x) | x \in X\}$$

is always a closed concave function. The functions $f^*, f_*$ will be called, respectively, the convex and concave Fenchel conjugates of f. If f is convex then $f_* \equiv -\infty$, similarly if f is concave $f^* \equiv \infty$, thus for convex or concave functions it will not be necessary to specify which form of the Fenchel conjugacy we are using as one of them will always yield a trivial result.

With these definitions, the Lagrangian function can be expressed as

$$L(x,y) = (-F)(x,_*y) = -F(x,^*-y),$$

where $(-F)(x,_*\cdot)$ denotes the (partial) concave Fenchel conjugate of $-F(x,\cdot)$, and $F(x,^*\cdot)$ the (partial) convex Fenchel conjugate of $F(x,\cdot)$.

Two more functions are of interest. The perturbation function $q:U \to \overline{\mathbb{R}}$, defined by

$$q(u) = \inf \{F(x,u) \mid x \in \mathbb{R}^n\},$$

and the dual function $g:Y \to \overline{\mathbb{R}}$, given by

$$g(y) = \inf \{L(x,y) \mid x \in \mathbb{R}^n\}.$$

Note that $q(0)$ is the optimal value of (P), and that since g is the pointwise infimum of a family of closed concave functions $L(x,\cdot)$, it is always closed concave. The dual problem associated with (P) is

(D) $\max \{g(y) \mid y \in Y\}.$

Replacing $L(x,y)$ by its expression in terms of $F(x,u)$ in the definition of g one easily gets

$$\begin{aligned} g(y) &= \inf \{q(u) + (u,y) \mid u \in U\} \\ &= (-q)_*(y) = -q^*(-y), \end{aligned}$$

The optimal value of the dual problem, is $(-g)_*(0)$ which by the above formula equals $-(-q)_{**}(0) = q^{**}(0)$. It is well known that for any function $h:U \to \overline{\mathbb{R}}$, $h \leqq h_{**}$ or $h \geqq h^{**}$, therefore

$$\sup_{y \in Y} g(y) = -g_*(0) = -(-q)_{**}(0)$$
$$\leqq q(0) = \inf \{f_0(x) + \psi(x;C_0) \mid x \in \mathbb{R}^n\},$$

and the optimal value of the dual problem (D) is never larger than the optimal value of (P), the "primal" problem. The difference between the two values is the duality gap.

In what follows the following will be assumed

(A1) $F$ is a closed proper convex function on $\mathbb{R}^n \times U$.

(A2) $L$ is a closed proper saddle function on $\mathbb{R}^n \times Y$ convex in the first argument and concave in the second.

(A3) $q$ is a closed proper convex function on $U$.

(A4) $g$ is a closed proper concave function on $Y$.

Some of these assumptions are clearly redundant. The joint convexity of $F$ on $(x,u)$ implies that $L(\cdot,y)$ is convex, since it is the pointwise infimum, with respect to $u$, of the functions $F(x,u) + (u,y)$ which are convex on $(x,u)$. Also as we saw above $L(x,\cdot)$ is always closed concave. Similar arguments show that under (A1) $q$ will be convex and $g$ closed concave. Conditions on (P) and $F$ that ensure that (A1) - (A4) hold are discussed in Rockafellar (1970a,1974,1976c).

Under these assumptions, the formulae giving $L$ and $g$ as partial or total conjugates of $F$ and $q$ respectively, can be inverted yielding

$$F(x,u) = \sup \{L(x,y) - (u,y) \mid y \in Y\}$$
$$= -L(x,_*u) = (-L)(x,{}^*-u),$$

$$q(u) = \sup \{g(y) - (u,y) | y \in Y\}$$

$$= -g_*(u) = (-g)^*(-u),$$

The first of these expressions implies that under the above assumptions a suitable saddle function L can be taken as the departure point for constructing F,q,g. In particular one sees that

$$\sup_{y \in Y} L(x,y) = -L(x,{}^*0) = F(x,0) = f_0 + \psi(\cdot;C_0),$$

and being q proper closed convex $-(-q)_{**} = q$, thus the optimal values of the primal and dual problems are equal.

Let $U = \mathbb{R}^m = Y$, and let

$$F_0(x,u) = f_0(x) + \psi(x;C_u)$$

where $C_u = \{x \in C | f(x) \leqq u\}$, $f = (f_1,\ldots,f_m)$, and $f(x) \leqq u$ is understood componentwise. The associated Lagrangian is

$$L_0(x,y) = \inf \{F_0(x,u) + (u,y) | u \in \mathbb{R}^m\}$$

$$= \inf \{f_0(x) + (u,y) | f(x) \leqq u\}$$

$$= \begin{cases} f_0(x) + (y,f(x)) & \text{if } x \in C,\ y \geqq 0 \\ -\infty & \text{if } x \in C,\ y \not\geqq 0 \\ \infty & \text{if } x \notin C, \end{cases}$$

Clearly

$$\sup_{y \in \mathbb{R}^m} L_0(x,y) = \sup_{y \in \mathbb{R}^m_+} L(x,y) = f_0(x) + \psi(x;C_0).$$

Furthermore

$$q_0(u) = \inf_{x\in\mathbb{R}^n} F_0(x,u) = \inf_{x\in C_u} f_0(x),$$

$$g_0(y) = \inf_{x\in\mathbb{R}^n} L_0(x,y) = \inf_{x\in C} L_0(x,y)$$

$$= \inf_{x\in C} [f_0(x) + (y,f(x))] - \psi(y;\mathbb{R}^m_+).$$

The ordinary dual problem is

$$(D_0)\ \max\{g_0(y)\,|\,y \in \mathbb{R}^m\} = \max\{g_0(y)\,|\,y \in \mathbb{R}^m_+\}.$$

Let $\phi:\mathbb{R} \to \mathbb{R}$ be a closed proper convex nonnegative function everywhere differentiable, and such that $\phi(x) = 0$ iff $x = 0$. Suppose, in addition, that $\phi^*$ also has these properties. Let $\Phi = \phi\circ|\cdot|$, $|\cdot|$ being the usual norm in $\mathbb{R}^m$. For any $c > 0$, let $\Phi_c = \frac{1}{c}\Phi(c\cdot)$, in other words, $\Phi_c$ is $\Phi$ multiplied on the right by $1/c$. Then $\Phi^*_c = (\Phi_c)^* = \frac{1}{c}\Phi^*$.

Using the above ordinary duality for (P), and the function $\Phi$, we will construct another duality scheme for (P) which will allow us to use the Nonlinear Proximal Point Algorithm (NPA) to solve (P). In order to do so, one starts by introducing the primal convex bifunction

$$F_c(x,u) = F_0(x,u) + \Phi_c(u).$$

From $F_c$ one can derive all the other elements of the duality scheme. In particular one has $q_c = q_0 + \Phi_c$.

Under the usual componentwise partial order $\mathbb{R}^n$ becomes a vector lattice. Given $x,y \in \mathbb{R}^n$, $\sup[x,y]$, $\inf[x,y]$ denote, respectively, the supremum and the infimum of $\{x,y\}$ with respect to this partial order.

For any $x \in \mathbb{R}^n$, $x^+$ will denote $\sup[x,0]$.

Given a function $h: \mathbb{R}^n \to \mathbb{R}$, it is inf-compact if the level sets

$$h^{\leqq}(\rho) = \{x \in \mathbb{R}^n \mid h(x) \leqq \rho\}$$

are compact for every $\rho \in \mathbb{R}$.

<u>Theorem 2.1.</u> Let $F_c$ be as above, and let $L_c$ be the corresponding Lagrangian. For all $x \in C$, $y \in \mathbb{R}^n$, $c > 0$

$$\begin{aligned} L_c(x,y) &= \inf\{F_c(x,u) + (u,y) \mid u \in \mathbb{R}^m\} \\ &= \sup\{L_0(x,S) - \Phi_c^*(s-y) \mid s \in \mathbb{R}^m\}. \end{aligned}$$

Both extrema are finite and uniquely attained, respectively, at

$$u_L(x,y,c) = \sup[f(x), \nabla\Phi_c^*(-y)]$$

$$s_L(x,y,c) = [\nabla\Phi_c(f(x)) + y]^+.$$

For all $x \in C$, $L_c(x,\cdot)$ is a continuously differentiable concave function on $\mathbb{R}^m$ which has the same set of maximizers as $L_0(x,\cdot)$. Furthermore

$$\partial_1 L_0(x, s_L(x,y,c)) \subseteqq \partial_1 L_c(x,y),$$

$$\begin{aligned} u_L(x,y,c) &= \nabla\Phi_c^*(s_L(x,y,c) - y) \\ &= \nabla_2 L_c(x,y) \in \partial_2 L_0(x, s_L(x,y,c)), \end{aligned}$$

$$\begin{aligned} -y &\in \partial_2 F_c(x, u_L(x,y,c)) \\ &= \partial_2 F_0(x, u_L(x,y,c)) + \nabla\Phi_c(u_L(x,y,c)). \end{aligned}$$

Proof. By definition

$$L_c(x,y) = \inf \{F_c(x,u) + (u,y) \mid u \in \mathbb{R}^m\}$$

$$= \inf \{F_0(x,u) + \Phi_c(u) + (u,y) \mid u \in \mathbb{R}^m\}.$$

The properties of $F_0(x,\cdot)$, $\Phi_c$ imply that the minimand is a proper closed inf-compact strictly convex function. Let us denote by $u_L(x,y,c)$ the unique vector where the above infimum is attained. Since $\Phi_c$ is finite, Fenchel's duality theorem implies that

$$L_c(x,y) = \sup \{L_0(x,s) - \Phi_c^*(s-y) \mid s \in \mathbb{R}^m\},$$

where the supremum is uniquely attained, by the strict convexity of $\Phi_c^*$, at $s_L(x,y,c)$. As

$$-\Phi_c^*(s-y) = (-\Phi_c)_*(y-s),$$

it follows that if $\Box$ denotes the sup-convolution of concave functions,

$$L_c(x,y) = (L_0(x,\cdot)\Box(-\Phi_c)_*)(y).$$

Let $\Box$ denote also, the convolution of multifunctions. $(-\Phi_c)_*$ is everywhere differentiable thus (Luque 1984a, th II.4.1(2), 1986b, th 4.1(2))

$$\partial_2 L_c(x,y) = (\partial_2 L_0(x,\cdot)\Box\,\partial(-\Phi)_*)(y)$$

$$= \{\nabla(-\Phi_c)_*(y - s_L(x,y,c))\},$$

which implies (by concavity) that $L_c(x,\cdot)$ is continuously differentiable with gradient

$$\nabla_2 L_c(x,y) = \nabla\Phi_c^*(s_L(x,y,c) - y) \in \partial_2 L_0(x,s_L(x,y,c)).$$

As $x \in C$

$$\inf \{F_0(x,u) + \Phi_c(u) + (u,y) \mid u \in \mathbb{R}^m\}$$

$$= \inf \{f_0(x) + \Phi_c(u) + (u,y) \mid u \geqq f(x)\}$$

from which the value of $u_L(x,y,c)$ follows. Similarly, $s_L(x,y,c)$ follows from

$$\sup \{L_0(x,s) - \Phi_c^*(s-y) \mid s \in \mathbb{R}^m\}$$

$$= \sup \{f_0(x) + (s,f(x)) - \Phi_c^*(s-y) \mid s \in \mathbb{R}_+^m\}$$

The above developments imply that

$$L_c(x,y) - F_c(x,u_L(x,y,c)) + (y,u_L(x,y,c))$$

$$= F_0(x,u_L(x,y,c)) + \Phi_c(u_L(x,y,c)) + (y,u_L(x,y,c))$$

$$= L_0(x,s_L(x,y,c)) - \Phi_c^*(s_L(x,y,c) - y),$$

which can be broken down into

$$L_c(x,y) - F_c(x,u_L(x,y,c)) = (y,u_L(x,y,c)),$$

$$L_0(x,s_L(x,y,c)) - F_0(x,u_L(x,y,c))$$

$$= \Phi_c(u_L(x,y,c)) + \Phi_c^*(s_L(x,y,c) - y) + (y,u_L(x,y,c)).$$

Given the conjugacy relations between primal bifunctions and their corresponding Lagrangians plus Young's inequality, it follows that

$$u_L(x,y,c) = \nabla_2 L_c(x,y),$$

$$-y \in \partial_2 L_c(x,u_L(x,y,c)),$$

$$u_L(x,y,c) = \nabla\Phi_c^*(s_L(x,y,c) - y),$$

$$s_L(x,y,c) - y = \nabla\Phi_c(u_L(x,y,c)),$$

$$u_L(x,y,c) \in \partial_2 L_0(x,s_L(x,y,c)),$$

$$-s_L(x,y,c) \in \partial_2 F_0(x,u_L(x,y,c)).$$

QED

The next theorem studies the properties of the penalty function corresponding to $L_c$.

<u>Theorem 2.2.</u> Let $c > 0$, for all $(x,y) \in C \times \mathbb{R}^m$

$$L_c(x,y) = f_0(x) + P_c(f(x),y),$$

where $P_c : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is a closed proper convex-concave saddle function, nondecreasing with respect to the first argument, and continuously differentiable in its two arguments. The expressions of $P_c$ and of its two partial derivatives are

$$P_c(t,y) = \Phi_c(\sup[t,\nabla\Phi_c^*(-y)]) + (y,\sup[t,\nabla\Phi_c^*(-y)])$$

$$= (t,[\nabla\Phi_c(t)+y]^+) - \Phi_c^*(\sup[\nabla\Phi_c(t),-y]),$$

$$\nabla_1 P_c(t,y) = [\nabla\Phi_c(t)+y]^+,$$

$$\nabla_2 P_c(t,y) = \sup\,[t, \nabla\Phi_c^*(-y)],$$

Proof. By the preceding theorem, taking into account the forms of $F_0(x,u)$ and of $L_0(x,s)$ on $C \times \mathbb{R}^m$, the expression of $L_c(x,y)$ on $C \times \mathbb{R}^m$ is

$$L_c(x,y) = f_0 + \inf_{u \geqq f(x)} [\Phi_c(u) + (u,y)]$$

$$= f_0(x) + \sup_{s \geqq 0} [(s,f(x)) - \Phi_c^*(s-y)].$$

Therefore, for all $t,y \in \mathbb{R}^m$

$$P_c(t,y) = \inf_{u \geqq t} [\Phi_c(u) + (u,y)] = \sup_{s \geqq 0} [(s,t) - \Phi_c^*(s-y)].$$

Since $P_c(\cdot,y)$ is the pointwise supremum of a family of affine functions, it is closed convex. Similarly, $P_c(t,\cdot)$ is closed concave. The inf-compactness and strict convexity of $\Phi_c$ and $\Phi_c^*$ imply that both extrema are uniquely attained at a common finite value. This ensures the finiteness of $P_c$ on $\mathbb{R}^m \times \mathbb{R}^m$. As before, the optimal solutions to the above optimization problems are

$$u_p(t,y,c) = \sup\,[t, \nabla\Phi_c^*(-y)],$$

$$s_p(t,y,c) = [\nabla\Phi_c(t) + y]^+.$$

In addition, $P_c(\cdot,y)$ is the pointwise supremum of the family of affine functions $\{(s,\cdot) - \Phi_c^*(s-y) \mid s \geqq 0\}$, only one of which is exact, the one corresponding to $s_p(\cdot,y,c,)$. Thus $\nabla_1 P_c(t,y)$ equals the slope of the affine function exact at $t$, i.e., $s_P(t,y,c)$. Similarly

$$\nabla_2 P_c(t,y) = u_P(t,y,c).$$

QED

The properties of $g_c$ are analyzed in the next

<u>Theorem 2.3.</u> Let $g_c$ be the dual function associated with $L_c$. Then for all $y \in \mathbb{R}^m$, $c > 0$

$$g_c(y) = \inf_{x \in \mathbb{R}^n} L_c(x,y) = \inf_{x \in C} L_c(x,y)$$

$$= \inf_{u \geqq 0} [q_c(u) + (u,y)]$$

$$= \sup_{s \geqq 0} [g_0(s) - \Phi_c^*(s-y)].$$

The last two finite extrema are uniquely attained at $u_g(y,c)$, $s_g(y,c)$ respectively. $g_c$ is an everywhere continuously differentiable concave function on $\mathbb{R}^m$ which has the same sets of maximizers as $g_0$. For all $y \in \mathbb{R}^m$, $c > 0$, one has

$$u_g(y,c) = \nabla\Phi_c^*(s_g(y,c) - y)$$

$$= \nabla g_c(y) \in \partial g_0(s_g(y,c)),$$

$$-y \in \partial q_c(u_g(y,c)) = \partial q_0(u_g(y,c)) + \nabla\Phi_c(u_g(y,c)).$$

If the infimum of $L_c(\cdot,y)$ is reached at $x_L(y,c)$, then

$$s_g(y,c) = s_L(x_L(y,c),y,c) = [\nabla\Phi_c(f(x_L(y,c))) + y]^+,$$

$$u_g(y,c) = u_L(x_L(y,c),y,c) = \sup\ [f(x_L(y,c)),\nabla\Phi_c^*(-y)],$$

$$\nabla_2 L_c(x_L(y,c),y) = \nabla g_c(y),$$

$$\partial_2 L_0(x_L(y,c),s_g(y,c)) \subseteqq \partial g_0(s_g(y,c)),$$

$$g_0(s_g(y,c)) = L_0(x_L(y,c),s_g(y,c)).$$

The last equation is equivalent to the fact that $L_0(\cdot,s_g(y,c))$ reaches its infimum at $x_L(y,c)$.

Proof. In the expression of $g_c$, the first equality is by definition, as well as the second, since dom $L_c(\cdot,y) = C$. The third equality is standard, it can be readily obtained upon replacing $L_c$ by its definition in terms of $F_c$ and interchange of the order of minimizations. The fourth is obtained from $q_c = q_0 + \Phi_c$. Since by assumption $q_0$ is proper closed convex and $\Phi_c$ is finite everywhere, Fenchel's duality theorem yields not only the last equality but also that the supremum with respect to s is attained. By assumption $g_c$ is proper, in addition to always being closed concave, and $\Phi_c^*$ is everywhere finite, thus Fenchel's duality theorem implies that the infimum with respect to u is reached. The fact that these extrema are attained can also be deduced from the inf-compactness of both $\Phi_c,\Phi_c^*$. Their strict convexity implies that the extrema are uniquely attained. Fenchel's duality theorem, or properness plus existence of extremizers, implies that their common optimal value is finite. As was done for $L_c$ in theorem 2.1,

$$\partial g_c(y) = (\partial g_0 \,\square\, \partial(-\Phi_c)_*)(y) = \{\nabla(-\Phi_c)_*(y - s_g(y,c))\},$$

which implies that $g_c$ is everywhere continuously (by concavity) differentiable with gradient

$$\nabla g_c(y) = \nabla\Phi_c^*(s_g(y,c) - y) \in \partial g_0(s_g(y,c)).$$

Also

$$\sup_{y\in\mathbb{R}^m} g_c(y) = \sup_{y\in\mathbb{R}^m} \sup_{s\in\mathbb{R}^m} [g_0(s) - \Phi_c^*(s-y)]$$

$$= \sup_{s\in\mathbb{R}^m} \sup_{y\in\mathbb{R}^m} [g_0(s) - \Phi_c^*(s-y)]$$

$$= \sup_{s\in\mathbb{R}^m} g_0(s).$$

From the expression of $g_c$ in terms of $g_0$, it follows that $g_c \geqq g_0$, thus taking into account the above inequality, it follows that

$$\text{Arg max } g_0 \subseteqq \text{Arg max } g_c.$$

To prove the reverse inclusion, Let us suppose that $\bar{y} \in$ Arg max $g_c$, then

$$\sup_{\mathbb{R}^m} g_c = g_c(\bar{y}) = g_0(s_g(\bar{y},c)) - \Phi_c^*(s_g(\bar{y},c) - \bar{y})$$

$$\leqq g_0(s_g(\bar{y},c)) \leqq \sup_{y\in\mathbb{R}^m} g_0(y),$$

where the first inequality follows from the fact that $\Phi_c^*$ is nonnegative. Since the suprema over $\mathbb{R}^m$ of $g_0$ and $g_c$ are equal, it follows that $\Phi_c^*(s_g(\bar{y},c) - \bar{y}) = 0$, thus $s_g(\bar{y},c) = \bar{y}$, and $\bar{y} \in$ Arg max $g_c$.

From these developments

$$g_c(y) = q_c(u_g(y,c)) + (y,u_g(y,c))$$

$$= q_0 \ (u_g(y,c)) + \Phi_c(u_g(y,c) + (y,u_g(y,c))$$

$$= g_0(s_g(y,c)) - \Phi_c^*(s_g(y,c) - y),$$

which can be broken down into

$$g_0(y) - q_c(u_g(y,c)) = (y,u_g(y,c))$$

$$g_0(s_g(y,c)) - q_0(u_g(y,c)) = \Phi_c(u_g(y,c)) + \Phi_c^*(s_g(y,c) - y) + (y,u_g(y,c)).$$

Given that $g_0 = (-q_0)_*$, $g_c = (-q_c)_*$, Young's inequality yields

$$u_g(y,c) = \nabla g_c(y),$$

$$-y \in \partial q_c(u_g(y,c)),$$

$$u_g(y,c) = \nabla\Phi_c^*(s_g(y,c) - y),$$

$$s_g(y,c) - y = \nabla\Phi_c(u_g(y,c)),$$

$$u_g(y,c) \in \partial g_0(s_g(y,c)),$$

$$-s_g(y,c) \in \partial q_0(u_g(y,c)).$$

Now, let us assume that the infimum of $L_c(\cdot,y)$ is reached at $x_L(y,c)$. The subgradient inequality for $L_c$ with respect to its second argument yields for all $y' \in \mathbb{R}^m$

$$L_c(x_L(y,c),y') \leqq L_c(x_L(y,c),y) + (y'-y,\nabla_2 L_c(x_L(y,c),y)).$$

Thus $L_c(x_L(y,c),y') \geqq g_c(y'y)$ and $g_c(y) = L_c(x_L(y,c),y)$ imply, via the above inequality, that

$$\nabla g_c(y) = \nabla_2 L_c(x_L(y,c),y).$$

By above and theorem 2.1

$$\nabla g_c(y) = u_g(y,c) = \nabla\Phi_c^*(s_g(y,c) - y),$$

$$\nabla_2 L_c(x_L(y,c),y) = u_L(x_L(y,c),y,c)$$

$$= \nabla\Phi_c^*(s_L(x_L(y,c),y,c) - y),$$

from which

$$u_g(y,c) = u_L(x_L(y,c),y,c) = \sup\,[f(x_L(y,c)),\nabla\Phi_c^*(-y)],$$

$$s_g(y,c) = s_L(x_L(y,c),y,c) = [\nabla\Phi_c(f(x_L(y,c))) + y]^+.$$

Also

$$L_c(x_L(y,c),y) = g_c(y) = g_0(s_g(y,c)) - \Phi_c^*(s_g(y,c) - y)$$

$$= L_o(x_L(y,c),s_L(x_L(y,c),y,c)) - \Phi_c^*(s_L(x_L(y,c),y,c) - y).$$

Since $s_L(x_L(y,c),y,c) = s_g(y,c)$, it follows that $g_0(s_g(y,c)) = L_0(x_L(y,c),s_g(y,c))$, and thus $x_L(y,c)$ minimizes $L_0(\cdot,s_g(y,c))$. Let $w \in \partial_2 L_0(x_L(y,c),s_g(y,c))$, for all $y' \in \mathbb{R}^m$

$$L_0(x_L(y,c),y') \leqq L_0(x_L(y,c),s_g(y,c)) + (y' - y,w).$$

But $L_0(x_L(y,c),s_g(y,c)) = g_0(s_g(y,c))$ and $L_0(x_L(y,c),y') \geqq g_0(y')$ imply that $w \in \partial g_0(s_g(y,c))$.

QED

3. The method of multipliers. The method of multipliers is a sequential unconstrained minimization technique applied to the Lagrangian function $L_c$. After selecting a penalty function $\phi$ of the type described in section 2, the augmenting function $P_c$ of theorem 2.2 is used to construct the Augmented Lagrangian $L_c$. For some value of dual variable y and of the penalty parameter c, $L_c(\cdot,y)$ is minimized, and the dual variable y, is updated as suggested by theorem 2.3.

Formally

(1) Select $c_0 > 0$, and $y^0 \in \mathbb{R}^m$.

(2) Given $c_k$, $y^k$, find $x^k$ such that

$$L_{c_k}(x^k,y^k) = \min_{y\in \mathbb{R}^m} L_{c_k}(x,y^k).$$

(3) Compute $y^{k+1} = \nabla_1 P_{c_k}(f(x^k),y^k)$. If $y^{k+1} = y^k$, stop, otherwise select $c_{k+1} > 0$, set $k = k + 1$, and go to 2.

As will be seen below this algorithm is equivalent to the application of the NPA to the maximization of the ordinary dual function $g_0$. Thus the optimal solution set of (D), $\bar{Y}$, must be nonempty (Luque 1984a th III.2.1, 1986a th 2.1). Conditions that insure that $\bar{Y}$ is not empty are given in Rockafellar (1970 , §28).

In order to be able to carry out step (2), the set of global minimizers of $L_{c_k}(\cdot,y^k)$ must be nonempty for each $k \geq 0$. Again sufficient conditions to insure this fact can be given. Typically, one assumes that the set of optimal solutions of (P), $\bar{X}$, is compact (Bertsekas 1982, ch 5).

The problem with the sufficient conditions alluded to in the above two paragraphs is that they are stronger than needed. For example, a typical byproduct of them is that the set of optimal solutions of (D) is compact. Thus rather than take those sufficient conditions as our primary assumptions, we will assume the following. Problem (P) is such that

(A5) For every $c > 0$, $y \in \mathbb{R}^m$, $L_c(\cdot, y)$ has a nonempty set of minimizers.

(A6) The set of optimal solutions of (D) is nonempty.

Henceforth, we will assume that these two assumptions, in addition to (A1)-(A4) of section 2, are in force. Note that (A1)-(A6) is not the minimal such set that would imply the well-possedness of our developments.

The termination criterion in (3) can be justified directly by an appeal to results of the Kuhn-Tucker saddle point theorem variety (see Bertsekas, 1982, ch 5), or else to the theory of the NPA (Luque 1984a, 1986a). In the latter case, it is justified because $y^{k+1} = y^k$ implies that $y^k$ is a fixed point of the proximal map of $\partial g_0$ with respect to a well-behaved enough monotone map to be described below. Then $y^k$ is an optimal solution of (D) (ibid., th III.2.4, th 2.4, respectively). And $x^k$, the minimizer of $L_{c_k}(\cdot, y^k)$, can be shown to be an optimal solution of (P) by the same argument as in Bertsekas (1982, prop 5(a)).

Finally let us note that by theorem 2.2, the sequence of dual variables $\{y^k\}_{k \geq 1}$ is feasible, i.e., $y^k \geq 0$ for all $k \geq 1$

We now show that the method of multipliers described above is just the NPA applied to $-\partial g_0$.

Theorem 3.1. The method of multipliers for the solution of (P) is a realization of the NPA in which $T = -\partial g_0 = \partial(-g_0)$, and $S = \nabla\Phi^*$.

Proof. From theorem 2.2

$$y^{k+1} = \nabla_1 P_{c_k}(f(x^k),y^k) = [\nabla\Phi_{c_k}(f(x^k)) + y^k]^+.$$

But $x^k$ minimizes $L_{c_k}(\cdot,y^k)$, thus $x^k = x_L(y^k,c_k)$ and by theorem 2.3, $y^{k+1} = s_g(y^k,c_k)$. Therefore $y^{k+1}$ satisfies

$$g_{c_k}(y^k) = \sup_{s\in\mathbb{R}^m} [g_0(s) - \Phi^*_{c_k}(s - y^k)]$$

$$= g_0(y^{k+1}) - \Phi^*(y^{k+1} - y^k).$$

Reversing signs throughout

$$((-g_0)\,\Box\,\Phi^*_{c_k}(-\cdot))(y^k) = (-g_0)(y^{k+1}) + \Phi^*_{c_k}(y^{k+1} - y^k).$$

Let $p(\cdot,\cdot)$, $P(\cdot,\cdot)$ denote respectively, proximal maps for functions, and for multifunctions (Luque 1984a, 1986b,c). Strict convexity of $\Phi^*$ implies the first equality of

$$\{y^{k+1}\} = p(-g_0,\Phi^*_{c_k}(-\cdot))(y^k) = P(\partial(-g_0),-\nabla\Phi^*_{c_k}(-\cdot))(y^k).$$

dom $\Phi^* = \mathbb{R}^m$ and Luque (1984a th II.4.1, 1986b th 4.1) imply the second one.

follows from the fact that $\Phi^*$ is everywhere finite and theorem II.4.1.

The function $\Phi = \phi\circ|\cdot|$ is clearly even, thus so is $\Phi^*$, and $-\nabla\Phi^*_{c_k}(-\cdot) = \nabla\Phi^*_{c_k} = c_k^{-1}\nabla\Phi^*$. Finally

$$\{y^{k+1}\} = P(\partial(-g_0),c_k^{-1}\nabla\Phi^*)y^k = P(c_k\partial(-g_0),\nabla\Phi^*)y^k. \qquad \text{QED}$$

4. Convergence of the method of multipliers. The fact that the method of multipliers is a realization of the NPA (theorem 3.1), allows us to prove results about the global and symptotic convergence properties of the method of multipliers.

Theorem 4.1. Let the penalty sequence $\{c_k\}$ be bounded away from zero. Let $\{x^k\}$, $\{y^k\}$ be sequences generated by the method of multipliers as above. Then for all $k \geqq 1$, $y^k \geqq 0$, the sequence $\{y^k\}$ ascends the dual functional to the optimal value $\bar{g}$ of the dual program (D), and converges towards a (not necessarily unique) maximizer of $g_0$. The sequence $\{x^k\}$ is asymptotically feasible, i.e., $\limsup_{k \to \infty} f(x^k) \leqq 0$, and asymptotically optimal, i.e., $\lim_{k\to\infty} f_0(x^k) = \bar{f}$, the optimal value of (P). Furthermore, every cluster point of $\{x^k\}$ is an optimal solution of (P).

Proof. By theorem 2.2

$$y^{k+1} = \nabla_1 P_{c_k}(f(x^k),y^k)) = [\nabla\Phi_{c_k}(f(x^k)) + y^k]^+$$

thus for all $k \geqq 1$, $y^k \geqq 0$. Since by theorem 3.1, the method of multipliers is but a realization of the NPA, it follows that $y^k \to y^\infty$ a maximizer of $g_0$ (Luque 1984a th III.2.12, 1986a th 2.12).

By its definition $x^k = x_L(y^k,c_k)$, which substituted in the above expression of $y^{k+1}$ yields

$$\begin{aligned} y^{k+1} &= [\nabla\Phi_{c_k}(f(x_L(y^k,c_k))) + y^k]^+ \\ &= s_L(x_L(y^k,c_k),y^k,c_k) = s_g(y^k,c_k), \end{aligned}$$

where the last two equalities follow from theorem 2.3. From the same theorem it follows that

$$\nabla\Phi^*_{c_k}(y^{k+1} - y^k) = \nabla\Phi^*_{c_k}(s_g(y^k, c_k) - y^k)$$

$$= u_g(y^k, c_k) = u_L(x_L(y^k, c_k), y^k, c_k)$$

$$= \sup\,[f(x_L(y^k, c_k)),\ \nabla\Phi^*_{c_k}(-y^k)]$$

$$= \sup\,[f(x^k),\ \nabla\Phi^*_{c_k}(-y^k)] \geqq f(x^k).$$

and

$$\limsup_{k\to\infty} f(x^k) \leqq \limsup_{k\to\infty} \nabla\Phi^*_{c_k}(y^{k+1} - y^k).$$

Furthermore (ibid.) $y^{k+1} - y^k \to 0$. Also $\nabla\Phi^*_{c_k} = c_k^{-1}\nabla\Phi^*$, $c_k$ is bounded away from zero, and $\nabla\Phi^*$ is continuous with $\nabla\Phi^* = (0) = 0$. thus

$$\lim_{k\to\infty} \nabla\Phi^*_{c_k}(y^{k+1} - y^k) = 0$$

from which the asymptotic feasibility of $\{x^k\}$ follows.

By theorem 2.1

$$L_{c_k}(x^k, y^k) = F_0(x^k, u_L(x^k, y^k, c_k))$$

$$+ \Phi_{c_k}(u_L(x^k, y^k, c_k)) + (y^k, u_L(x^k, y^k, c_k)).$$

We have seen above that $f(x^k) \leqq u_L(x^k, y^k, c_k)$, thus $F_0(x^k, u_L(x^k, y^k, c_k))$ $= f_0(x^k)$. Since $x^k = x_L(y^k, c_k)$ it follows that $L_{c_k}(x^k, y^k) = g_{c_k}(y^k)$. Therefore

$$g_{c_k}(y^k) = f_0(x^k) + \Phi_{c_k}(u_L(x^k,y^k,c_k)) + (y^k, u_L(x^u,y^k,c_k)).$$

From the expression of $g_{c_k}$ in theorem 2.3, one obtains

$$g_{c_k}(y^k) \geqq g_0(y^k) - \Phi^*_{c_k}(y^k - y^k) = g_0(y^k)$$

and

$$g_{c_k}(y^k) = g_0(s_g(y^k,c_k)) - \Phi^*_{c_k}(s_g(y^k,c_k) - y^k)$$

$$= g_0(y^{k+1}) - \Phi^*_{c_k}(y^{k+1} - y^k) \leqq g_0(y^{k+1}).$$

Thus for all $k \geqq 0$

$$g_0(y^k) \leqq g_{c_k}(y^k) \leqq g_0(y^{k+1}),$$

and the sequence $\{g_0(y^k)\}$ is nondecreasing. It is also bounded above by the optimal value of the dual problem (D), thus it converges. Using again Luque (1984a th III.2.12, 1986a th 2.12)

$$y^k \to y \in \text{Arg max } g_0.$$

Since $g_0$ is upper semicontinuous by closedness, it follows that

$$\lim_{k\to\infty} g_0(y^k) = \limsup_{k \to \infty} g_0(y^k) = g_0(y^\infty) = \bar{g},$$

and $\{y^k\}$ ascends $g_0$ to the optimal value of (D).

The interleavedness of $\{g_{c_k}(y^k)\}$ and $\{g_0(y^k)\}$ implies that

$$\lim_{k\to\infty} g_{c_k}(y^k) = \bar{g}.$$

Since $\{y^k\}$ is convergent, it is bounded. Also we have seen that

$$u_L(x^k,y^k,c_k) = u_L(x_L(y^k,c_k),y^k,c_k)$$

$$= \nabla\Phi^*_{c_k}(y^{k+1} - y^k) \to 0,$$

and by theorem 2.1

$$\Phi c_k(u_L(x^k,y^k,c_k)) = c_k^{-1}\Phi(c_k u_L(x^k,y^k,c_k))$$

$$= c_k^{-1}\Phi(c_k\nabla\Phi^*_{c_k}(s_L(x^k,y^k,c_k) - y^k))$$

$$= c_k^{-1}\Phi(\nabla\Phi^*(s_L(x^k,y^k,c_k) - y^k))$$

from which it is clear that $\Phi_{c_k}(u_L(x^k,y^k,c_k)) \to 0$.

This is so because $c_k$ is bounded away from zero, $\Phi$ is continuous with $\Phi(0) = 0$, and

$$\nabla\Phi^*(s_L(x^k,y^k,c_k) - y^k) = \nabla\Phi^*(y^{k+1} - y^k) \to 0$$

as discussed above. Putting all these facts together, it follows that

$$\lim_{k\to\infty} f_0(x^k) = \lim_{k\to\infty} g_{c_k}(y^k) = \bar{g} \leqq \bar{f},$$

where we actually have equality by assumptions (1) - (4) of section 2.

Let $\bar{x}$ be a cluster point of $\{x^k\}$, thus $\{x^k\}_K \to \bar{x}$ for some subsequence indexed by $K \subseteq \mathbb{N}$. Since the functions $f_i, i=0,1,\ldots,m$ are lower semicontinuous one has

$$f(\bar{x}) = \liminf_{k \in K} f(x^k) \leqq \limsup_{k \in K} f(x^k) \leqq \limsup_{k \to \infty} f(x^k) \leqq 0$$

and $\bar{x}$ is feasible. Also

$$f_0(\bar{x}) = \liminf_{k \in K} f_0(x^k) \leqq \limsup_{k \in K} f_0(x^k) \leqq \limsup_{k \to \infty} f_0(x^k)$$
$$= \lim_{k\to\infty} f_0(x^k) = \bar{f},$$

from which $f_0(\bar{x}) \leqq \bar{f}$. By the feasibility of $\bar{x}$, it follows that $\bar{x}$ is an optimal solution of (P)

QED.

We now turn our attention to the study of the asymptotic convergence of the method of multipliers. Let $\bar{Y}$ be the set of optimal solutions of the dual problem (D). Since $g_0$ is closed concave, $\bar{Y}$ is always closed and convex. Furthermore, by assumption (A6) (see section 3), it is nonempty. Thus for any sequence of dual variables $\{y^k\}$ generated by the algorithm, the sequence $\{d(y^k,\bar{Y})\}$ is well defined.

We have seen (theorem 4.1) that $y^k \to y$ a (not necessarily unique) maximizer of $g_0$, thus $d(y^k,\bar{Y}) \to 0$. In our asymptotic convergence analysis of the method of multipliers we focus our attention on the speed of convergence of $d(y^k,\bar{Y})$. Via theorem 3.1 we are able to apply the results developed in Luque (1984a, 1986a) for the general NPA.

*Theorem 4.2.* Let $g_0$ satisfy the following estimate for some numbers $\eta, a, t > 0$

$$\forall y^* \in \eta B, \ (y^* \in \partial g_0(y) \implies d(y,\bar{Y}) \leqq a|y^*|^t).$$

Let $\Phi^*$ satisfy the following estimate for some numbers $\delta, b, s > 0$ such that $st \geqq 1$

$$\forall y \in \delta B, \ |\nabla\Phi^*(y)| \leqq b|y|^s.$$

If $st = 1$, the method of multipliers converges linearly at a rate bounded above by

$$\frac{a}{(a^2 + (\underline{c}/b)^{2t})^{\frac{1}{2}}},$$

where $\underline{c} = \liminf_{k \to \infty} c_k$, and thus superlinearly if $\underline{c} = \infty$. In any case the (Q-) order of convergence is at least $st \geqq 1$.

*Proof.* Everything follows directly from the equivalence between the method of multipliers and the NPA shown in theorem 3.1, plus a theorem of Luque (1984a th III.3.1, 1986a th 3.1) QED

*Theorem 4.3.* Let everything be as in theorem 4.2, except that now $\partial g_0$ satisfies the following estimate

$$\forall y^* \in \eta B, \ (y^* \in \partial g_0(y) \implies y \in \bar{Y}).$$

Then, the method of multipliers converges in finitely many steps which can be reduced to one if $\delta \geqq d(y^0,\bar{Y})$ and $c_0 \geqq bd(y^0,\bar{Y})^s/\eta$.

Proof. Luque (1984a th III.3.2, 1986a th 3.2). QED

Theorem 4.4. Let $a, t, \delta$ be positive numbers, and let $\tau: \mathbb{R}_+ \to \mathbb{R}_+$, be continuous at zero with $\tau(0) = 0$. Let $\partial g_0$ satisfy the following estimate.

$$\forall y^* \in \delta B, \; (y^* \in \partial g_0(y) \;\Rightarrow\; a|y^*|^t \leqq d(y,\bar{Y}) \leqq \tau(|y^*|)).$$

Let $b, s, \eta$ be positive numbers, and let $\sigma: \mathbb{R}_+ \to \mathbb{R}_+$ be continuous at zero with $\sigma(0) = 0$. Let $\nabla\Phi^*$ satisfy

$$\forall y \in \eta B, \; b|y|^s \leqq |\nabla\Phi^*(y)| \leqq \sigma(|y|).$$

If $st < 1$, then

$$\liminf_{k \to \infty} \frac{d(y^{k+1},\bar{Y})}{d(y^k,\bar{Y})} = 1,$$

and the convergence cannot be faster than sublinear. If $st = 1$, then

$$\liminf_{k \to \infty} \frac{d(y^{k+1},\bar{Y})}{d(y^k,\bar{Y})} = \frac{a}{ab^t + \bar{c}} \quad ,$$

where $\bar{c} = \limsup_{k \to \infty} c_k$.

Proof. Luque (1984a th III.3.3, 1986a th 3.3). QED

We can also give an estimate of the speed of convergence of $g_0(y^k)$.

Theorem 4.5. Under the assumptions of theorem 4.2, for all k large enough

$$\frac{\bar{g} - g_0(y^{k+1})}{|y^{\infty} - y^{k+1}|} \leqq c_k^{-1} b d(y^k, \bar{Y}).$$

Proof. Luque (1984a th III.5.2, 1986a th 5.2). QED

The results of theorems 4.2 and 4.3 are similar to those reported in the literature (Bertsekas 1982, props 5.20, 5.21, 5.22). On the other hand, to the best of the author's knowledge, theorems 4.4 and 4.5 are completely new.

REFERENCES

[1] Bertsekas D.P. (1975) Necessary and sufficient conditions for a penalty method to be exact. Math. Programming 9, 87-99.

[2] ----- (1982) Constrained optimization and Lagrange multiplier methods. Academic Press, New York.

[3] Bourbaki N. (1981) Espaces vectoriels topologiques. Masson, Paris.

[4] Kort B.W., and Bertsekas D.P. (1972) A new penalty function method for constrained optimization. Proceedings IEEE Conference on Decision and Control, New Orleans LA, 162-166.

[5] ----- (1973) Multiplier methods for convex programming. Proceedings IEEE Conference on Decision and Control, San Diego CA, 428-432.

[6] ----- (1976) Combined primal-dual and penalty methods for convex programming. SIAM J. Control and Optimization 14: 2, 268-294.

[7] Luque J. (1984a) Nonlinear proximal point algorithms. Dissertation, Operations Research Center, Massachusetts Institute of Technology, May 1984.

[8] ----- (1984b) Asymptotic convergence analysis of the proximal point algorithm. SIAM J. Control and Optimization 22: 2, 277-293.

[9] ----- (1986a) The nonlinear proximal point algorithm. Submitted.

[10] ----- (1986b) Convolution of maximal monotone mappings. Submitted.

[11] ----- (1986c) Convolutions of maximal monotone mappings, abstract. Abstracts of papers presented to the Amer. Math. Soc. 7: 1, 108, no. *825-49-541.

[12] Ortega J.M. and Rheinboldt W.C. (1970) Iterative solution of nonlinear equations in several variables. Academic Press, New York.

[13] Rockafellar R.T. (1970) Convex analysis. Princeton University Press.

[14] ----- (1973) The multiplier method of Hestenes and Powell applied to convex programming. J.O.T.A. 12: 6, 555-562.

[15] ----- (1974) Conjugate duality and optimization. Regional Conference Series in Applied Mathematics #16, SIAM, Philadelphia.

[16] ----- (1976a) Monotone operators and the proximal point algorithm. SIAM J. Control and Optimization 14: 5, 877-898.

[17] ----- (1976b) Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Operations Research 1: 2, 97-116.

[18] ----- (1976c) Solving a nonlinear programming problem by way of a dual problem. Symposia Mathematica 19, 135-160.