SIAM J. CONTROL
Vol. 12, No. 2, May 1974

AUGMENTED LAGRANGE MULTIPLIER FUNCTIONS AND DUALITY IN NONCONVEX PROGRAMMING*

R. TYRRELL ROCKAFELLAR†

Abstract. If a nonlinear programming problem is analyzed in terms of its ordinary Lagrangian function, there is usually a duality gap, unless the objective and constraint functions are convex. It is shown here that the gap can be removed by passing to an augmented Lagrangian which involves quadratic penalty-like terms. The modified dual problem then consists of maximizing a concave function of the Lagrange multipliers and an additional variable, which is a penalty parameter. In contrast to the classical case, the multipliers corresponding to inequality constraints in the primal are not constrained a priori to be nonnegative in the dual. If the maximum in the dual problem is attained (and conditions implying this are given), optimal solutions to the primal can be represented in terms of global saddle points of the augmented Lagrangian. This suggests possible improvements of existing penalty methods for computing solutions.

1. Introduction. Let $f_0, f_1, \cdots f_m$ be real-valued functions defined on a set $S \subset R^n$. We shall be concerned with the nonlinear programming problem:

$$\text{(P)} \qquad \begin{aligned} &\text{minimize} \quad f_0(x) \quad \text{over all } x \in S \text{ satisfying} \\ &f_i(x) \leqq 0 \quad \text{for } i = 1, \cdots, m. \end{aligned}$$

The ordinary Lagrangian function associated with problem (P) is

$$L_0(x, y) = f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x) \quad \text{for } (x, y) \in S \times R^m_+, \tag{1.1}$$

and this corresponds to the dual problem:

$$(\mathrm{D}_0) \qquad \begin{aligned} &\text{maximize} \quad g_0(y) \quad \text{over all } y \in R^m_+, \quad \text{where} \\ &g_0(y) = \inf_{x \in S} L_0(x, y). \end{aligned}$$

It is well known that the optimal values in these two problems satisfy

$$\inf (\mathrm{P}) \geqq \sup (\mathrm{D}_0), \tag{1.2}$$

but equality cannot be expected to hold, aside from freakish cases, unless S and the functions f_i are convex. The discrepancy in (1.2) is termed a "duality gap".

In recent years a number of authors have addressed the question of whether this duality gap in nonconvex programming could be eliminated by changing the Lagrangian function. Such a change might also be of benefit computationally in some situations, even in convex programming, where the plurality of useful Lagrangians and dual problems has been known for some time. Computational considerations in nonconvex problems with equality constraints have led in particular to algorithms based on an augmented Lagrangian in which "penalty" terms of the form $rf_i(x)^2$, $i = 1, \cdots, m$, are added to $L_0(x, y)$; cf. Arrow and Solow [2], Bertsekas [3], Buys [4], Fletcher [6], [7], [8], Haarhoff and Buys [9], Hestenes [10], Kort and Bertsekas [11], Lill [12], Miele et al. [14], [15], [16], [17], Poljak [30],

* Received by the editors March 5, 1973, and in revised form August 30, 1973.

† Department of Mathematics, University of Washington, Seattle, Washington 98195. This work was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-72-2269.

Tripathi and Narendra [25], and Wierzbicki [27], [28], [29]. For the inequality-constrained problem (P), the simple terms $rf_i(x)^2$ are not suitable, and the analogous augmented Lagrangian (suggested in [21] and investigated by Buys in his thesis [4]) turns out to be

$$L(x, y, r) = f_0(x) + \sum_{i=1}^{m} [y_i \max\{f_i(x), -y_i/2r\} + r \max^2\{f_i(x), -y_i/2r\}]$$

(1.3)

$$= f_0(x) + r \sum_{i=1}^{m} \psi(f_i(x), y_i/r) \quad \text{for } x \in S, \quad (y, r) \in T,$$

where $T = R^m \times (0, +\infty)$ and

$$\psi(\alpha, \beta) = [\max^2\{0, 2\alpha + \beta\} - \beta^2]/4$$

(1.4)

$$= \begin{cases} \alpha\beta + \alpha^2 & \text{if } \alpha \geqq -\beta/2, \\ -\beta^2/4 & \text{if } \alpha \leqq -\beta/2. \end{cases}$$

We have demonstrated in [13] that in the convex case this augmented Lagrangian is not only a natural choice but has a number of strong properties not possessed by the ordinary Lagrangian L_0. In [14], we have derived some consequences of these properties for the multiplier method of Hestenes and Powell. It is the purpose of the present paper to develop general properties of L in the nonconvex case, especially with regard to duality.

Arrow, Gould and Howe [1, Thm. 2] have already shown that if $\bar{x}$ is an isolated local solution to (P) satisfying the standard second order sufficiency conditions for optimality with strict complementarity, the Lagrange multiplier vector being $\bar{y}$, and if $\bar{r}$ is sufficiently large, then there is a neighborhood N of $\bar{x}$ in S such that

(1.5) $$\min_{x \in N} L(x, \bar{y}, \bar{r}) = L(\bar{x}, \bar{y}, \bar{r}) = \max_{y \in R^m_+} L(\bar{x}, y, \bar{r}),$$

with the minimum in (1.5) attained uniquely at $\bar{x}$. This saddle-point theorem is strengthened below (Corollary 6.1) in three ways: by extending the maximum in (1.5) to the maximum of $L(\bar{x}, y, r)$ over all $(y, r) \in T$ (thus in particular removing the constraint $y \geqq 0$), by deleting the strict complementarity assumption, and (under the hypothesis that $\bar{x}$ is the unique globally optimal solution to (P) "in the strong sense") by extending the minimum in (1.5) to the minimum over all $x \in S$. Introducing the ordinary perturbations associated with (P), we also give necessary and sufficient conditions in terms of stability for the existence of a global saddle point $(\bar{x}, \bar{y}, \bar{r})$ of L with respect to $S \times T$ and more generally characterize the case where at least the global "inf sup" and "sup inf" of L are equal.

These results correspond to a detailed study of the following dual problem in place of (D_0):

(D)

$$\text{maximize} \quad g(y, r) \quad \text{over all } (y, r) \in T, \quad \text{where}$$

$$g(y, r) = \inf_{x \in S} L(x, y, r) < +\infty.$$

Of course, the optimal value in (D) is by definition

(1.6) $$\sup(D) = \sup_{(y,r) \in T} \inf_{x \in S} L(x, y, r).$$

On the other hand, the optimal value in (P) satisfies

$$\text{(1.7)} \qquad \inf (\mathrm{P}) = \inf_{x \in S} \sup_{(y,r) \in T} L(x, y, r),$$

inasmuch as

$$\text{(1.8)} \qquad \sup_{(y,r)\in T} L(x,y,r) = \begin{cases} f_0(x) & \text{if } x \text{ is feasible}, \\ +\infty & \text{if } x \text{ is not feasible}. \end{cases}$$

The latter is immediate from the fact that

$$\text{(1.9)} \qquad \sup_{\beta \in R} \psi(\alpha, \beta) = \begin{cases} 0 & \text{if } \alpha \leqq 0, \\ +\infty & \text{if } \alpha > 0. \end{cases}$$

Thus the relation

$$\text{(1.10)} \qquad \inf (\mathrm{P}) \geqq \sup (\mathrm{D})$$

holds, and minimax theorems for L are equivalent to duality theorems asserting the equality and attainment of the optimal values in (1.10). (For related work on duality since this paper was submitted for publication, see Mangasarian [13], Pollatschek [18] and Rockafellar [24].)

For notational simplicity, only inequality constraints are treated in this paper. However, the same results apply with only the obvious changes if explicit equality constraints are also allowed (the corresponding terms $r\psi(f_i(x), y_i/r)$ in (1.3) being replaced by $y_i f_i(x) + r f_i(x)^2$). The routine alterations in the proofs are left to the reader.

Except for Theorem 6, which requires second order differentiability of the functions f_i, the results remain valid if S is a subset of an arbitrary topological real vector space.

2. The nature of the dual problem. Let $p: R^m \to [-\infty, +\infty]$ be the ordinary perturbation function (min-value function) associated with (P), that is,

$$\text{(2.1)} \qquad p(u) = \inf_{x \in S} F(x, u),$$

where for each $(x, u) \in S \times R^m$:

$$\text{(2.2)} \qquad F(x,u) = \begin{cases} f_0(x) & \text{if } f_i(x) \leqq u_i \text{ for } i = 1, \cdots, m, \\ +\infty & \text{otherwise}. \end{cases}$$

Then

$$\text{(2.3)} \qquad \inf_{u \in R^m} \{F(x,u) + y \cdot u\} = \begin{cases} L_0(x,y) & \text{if } y \in R^m_+, \\ -\infty & \text{if } y \notin R^m_+, \end{cases}$$

$$\text{(2.4)} \qquad \inf_{u \in R^m} \{p(u) + y \cdot u\} = \begin{cases} g_0(y) & \text{if } y \in R^m_+, \\ -\infty & \text{if } y \notin R^m_+. \end{cases}$$

More generally, it is elementary to calculate that

$$L(x, y, r) = \inf_{u \in R^m} \{F(x, u) + y \cdot u + r|u|^2\} \quad \text{for all } (y, r) \in T, \tag{2.5}$$

$$g(y, r) = \inf_{u \in R^m} \{p(u) + y \cdot u + r|u|^2\} \quad \text{for all } (y, r) \in T. \tag{2.6}$$

In order that (2.3) and (2.4) can be regarded as instances of (2.5) and (2.6), we adopt the convention that

$$L(x, y, 0) = \begin{cases} L_0(x, y) & \text{if } y \in R^m_+, \\ -\infty & \text{if } y \notin R^m_+, \end{cases} \tag{2.7}$$

$$g(y, 0) = \begin{cases} g_0(y) & \text{if } y \in R^m_+, \\ -\infty & \text{if } y \notin R^m_+. \end{cases} \tag{2.8}$$

This extends the definition of $L(x, \cdot, \cdot)$ and g to cl T.

THEOREM 1. *The functions $L(x, y, r)$ and $g(y, r)$ are concave and upper semicontinuous in $(y, r) \in \text{cl } T = R^m \times R^1_+$ and nondecreasing in $r \in R^1_+$, nowhere $+\infty$. Furthermore, whenever $r > s \geqq 0$ one has*

$$g(y, r) \geqq \max_{z \in R^m} \{g(z, s) - |y - z|^2/4(r - s)\}. \tag{2.9}$$

Proof. The first assertion is implied by (2.5) and (2.6), since the pointwise infimum of a collection of affine functions of (y, r) which are nondecreasing in r is an upper semicontinuous, concave function which is nondecreasing in r. For any (y, r) and (z, s) satisfying $r > s \geqq 0$, we have from (2.6) that

$$\begin{aligned} g(y, r) &= \inf_{u \in R^m} \{p(u) + z \cdot u + s|u|^2 + (y - z) \cdot u + (r - s)|u|^2\} \\ &\geqq \inf_{u \in R^m} \{p(u) + z \cdot u + s|u|^2\} + \inf_{u \in R^m} \{(y - z) \cdot u + (r - s)|u|^2\} \\ &= g(z, s) - |y - z|^2/4(r - s), \end{aligned}$$

and this yields (2.9). The maximum (instead of supremum) in (2.9) is valid because $g(\cdot, s)$ is an upper semicontinuous concave function nowhere having the value $+\infty$ and hence in particular is majorized by at least one affine function. (Thus the function of z being maximized is upper semicontinuous; its level sets are bounded because it is majorized by a negative definite quadratic function of z.)

Remark. In the convex case (i.e., where S and the functions f_i are all convex), $L(x, y, r)$ is convex in x and relation (2.9) holds as an equation [13]. Then for every $r > 0$ the function $g(\cdot, r)$ has the same maximum and even the same maximizing set as $g(\cdot, 0)$, since in the formula

$$g(y, r) = \max_{z \in R^m} \{g(z, 0) - |y - z|^2/4r\}$$

the bracketed expression is maximized jointly in y and z if and only if y maximizes $g(\cdot, 0)$ and $z = y$. In other words, in the convex case a pair $(\bar{y}, \bar{r})$ with $\bar{r} > 0$ is an optimal solution to the dual problem (D) if and only if $\bar{y}$ is an optimal solution to the ordinary dual (D_0). In the nonconvex case this is no longer true, although the monotonicity of $g(y, r)$ in r still implies that if $(\bar{y}, \bar{r})$ is an optimal solution to (D) and $r > \bar{r}$, then $(\bar{y}, r)$ is also an optimal solution to (D).

COROLLARY 1.1. *There is an* r_0, $0 \leqq r_0 \leqq +\infty$, *such that* $g(y, r)$ *is finite for all* $y \in R^m$ *if* $r_0 < r < +\infty$, *whereas* $g(y, r) = -\infty$ *for all* $y \in R^m$ *if* $0 \leqq r < r_0$.

Proof. This is obvious from (2.9), according to which $g(y, r) > -\infty$ if there exists some $s \in [0, r)$ and $z \in R^m$ such that $g(z, s) > -\infty$.

In view of the fact that $g(y, r)$ is nondecreasing in r, Corollary 1.1 says there are no real constraints at all in (D), even implicit ones. This is in contrast to the situation for (D_0), where the feasibility condition $g_0(y) > -\infty$ requires the satisfaction of $y \geqq 0$, as well as other possible constraints. (It is not always possible a priori to specify for (D) an r such that $r > r_0$, although, for example, one has $r_0 = 0$ if f_0 is bounded below on S. In this connection, see the remarks preceding Theorem 2 in the next section.)

COROLLARY 1.2. *For every* $y \in R^m$, *one has*

$$\lim_{r \to +\infty} g(y, r) = \sup_T g = \sup (D). \tag{2.10}$$

Proof. Given any $(z, s) \in T$ and $\varepsilon > 0$, one has $g(y, r) \geqq g(z, s) - \varepsilon$ for all r sufficiently large by (2.9).

The last result brings out the close relationship between the dual (D) and penalty methods for solving (P). By definition, we have

$$L(x, 0, r) = f_0(x) + r \sum_{i=1}^{m} \max{}^2 \{0, f_i(x)\}, \tag{2.11}$$

and consequently

$$g(0, r) = \inf_{x \in S} \left\{ f_0(x) + r \sum_{i=1}^{m} \max{}^2 \{0, f_i(x)\} \right\}. \tag{2.12}$$

The limit of the infimum (2.12) as $r \to +\infty$ is the optimal value sup (D), according to Corollary 1.2. Thus the relationship between sup (D) and inf (P) is of fundamental importance for the penalty method in which (2.12) is calculated for a sequence of r values tending to $+\infty$. Note that if we fix any $y \in R^m$ and minimize $L(\cdot, y, r)$, instead of $L(\cdot, 0, r)$, for a sequence of r values tending to $+\infty$, the limit of the infima is still sup (D) by Corollary 1.2. This procedure can be regarded as a modified penalty method. Still more broadly, one can try to solve (P) by minimizing $L(\cdot, y, r)$ for a sequence of vectors $(y, r) \in T$ such that $g(y, r) \to \sup (D)$. If the sequence can be generated in such a manner that the r values remain *bounded*, there is the advantage that the numerical instabilities associated with minimizing (2.11) for ever-larger values of r could be avoided. The results below demarcate the region of validity and potential effectiveness of such algorithms, from a theoretical point of view. Theorem 6 indicates that indeed, penalty methods can be constructed which are capable of solving "most" problems *without* $r \to +\infty$.

3. Solving (P) in the asymptotic sense. We say that (P) satisfies the *quadratic growth condition* if there is an $r \geqq 0$ such that the expression (2.11) is bounded below as a function of $x \in S$. This certainly holds if f_0 is bounded below on S, and in particular if S is compact and f_0 lower semicontinuous. In general, since by (2.6) and the definition of g we have

$$\inf_{x \in S} L(x, 0, r) = g(0, r) = \inf_{u \in R^m} \{p(u) + r|u|^2\}, \tag{3.1}$$

the quadratic growth condition holds if and only if there exist real numbers $r \geqq 0$ and q such that

$$p(u) \geqq q - r|u|^2 \quad \text{for all } u \in R^m. \tag{3.2}$$

The condition is therefore equivalent also to the relation

$$\liminf_{|u| \to +\infty} p(u)/|u|^2 > -\infty.$$

Observe that the r_0 in Corollary 1.1 is the infimum of all the numbers $r \geqq 0$ for which the quadratic growth condition holds, since it is the infimum of all the numbers $r \geqq 0$ such that $g(0, r) > -\infty$. Thus (P) satisfies the condition if and only if g is not identically $-\infty$ on T, or, in other words, if and only if (D) has "feasible solutions". This also shows that the quadratic growth condition is equivalent to the seemingly more general condition that for some $y \in R^m$ (not necessarily $y = 0$) and some $r \geqq 0$, the infimum of $L(x, y, r)$ over all $x \in S$ is not $-\infty$.

THEOREM 2. *If* (P) *satisfies the quadratic growth condition, one has*

$$\begin{aligned} -\infty < \sup(\mathrm{D}) &= \liminf_{u \to 0} p(u) \\ &\leqq p(0) = \inf(\mathrm{P}). \end{aligned} \tag{3.3}$$

If (P) *does not satisfy the quadratic growth condition, one has* $\sup(\mathrm{D}) = -\infty$.

Proof. The preceding remark makes clear that $\sup(\mathrm{D}) = -\infty$ if and only if the quadratic growth condition fails to be satisfied. Assume henceforth that the condition is satisfied; thus (3.2) holds for a certain $\bar{q}$ and $\bar{r}$. From (3.1) we see that

$$g(0, r) \leqq \liminf_{u \to 0} p(u) \quad \text{for all } r \geqq 0.$$

Taking the limit as $r \to +\infty$ and invoking Corollary 1.2, we obtain

$$\sup(\mathrm{D}) \leqq \liminf_{u \to 0} p(u).$$

To establish the opposite inequality, and thereby complete the proof of the theorem, consider now an arbitrary real number q such that

$$q < \liminf_{u \to 0} p(u). \tag{3.4}$$

Choose ε sufficiently small that $p(u) \geqq q$ whenever $|u| < \varepsilon$. For r sufficiently large, we have

$$q - r|u|^2 \leqq \bar{q} - \bar{r}|u|^2 \quad \text{if } |u| \geqq \varepsilon$$

(with $\bar{q}$ and $\bar{r}$ as above), and therefore

$$q - r|u|^2 \leqq p(u) \quad \text{for all } u.$$

But then

$$q \leqq \inf_{u \in R^m} \{p(u) + r|u|^2\} = g(0, r) \leqq \sup(\mathrm{D}).$$

Since q was any real number satisfying (3.4), this shows that

$$\sup(\mathrm{D}) \geqq \liminf_{u \to 0} p(u),$$

and we are done.

The quantity

$$\liminf_{u \to 0} p(u) \tag{3.5}$$

in Theorem 2 is the *asymptotic optimal value* in (P). It can also be described as the minimum of

$$\limsup_{k \to \infty} f_0(x^k) \tag{3.6}$$

over all asymptotically feasible sequences $(x^k)_{k=1}^{\infty}$ for (P): that is, sequences in S satisfying

$$\limsup_{k \to \infty} f_i(x^k) \leqq 0 \quad \text{for } i = 1, \cdots, m. \tag{3.7}$$

Indeed, according to the definition of p, (3.5) is the lowest possible limit achievable by any sequence $(\alpha_k)_{k=1}^{\infty}$ such that there exist $u^k \in R^m$ and $x^k \in S$ with $u^k \to 0$, $f_i(x^k) \leqq u_i^k$ for $i = 1, \cdots, m$, and $f_0(x^k) \leqq \alpha_k$.

Let us call a sequence $(x^k)_{k=1}^{\infty}$ *asymptotically minimizing* for (P) if it is asymptotically feasible and yields the minimum possible value for (3.6). We can then obtain from Theorem 2 a result which shows how any procedure for solving (D) can be used to solve (P) in the sense of constructing an asymptotically minimizing sequence. (A similar result involving more detailed estimates in the convex case has been demonstrated in [22].)

THEOREM 3. *Let $(y^k, r_k)_{k=1}^{\infty}$ be a sequence such that for some $\delta > 0$ one has $(y^k, r_k - \delta) \in T$ and*

$$\lim_{k \to \infty} g(y^k, r_k - \delta) = \sup(\mathrm{D}) < +\infty. \tag{3.8}$$

Let $x^k \in S$ satisfy

$$L(x^k, y^k, r_k) \leqq \inf_{x \in S} L(x, y^k, r_k) + \alpha_k, \tag{3.9}$$

where $\alpha_k \to 0$. Then $(x^k)_{k=1}^{\infty}$ is asymptotically feasible and

$$\liminf_{k \to \infty} y_i^k / r_k \geqq 0 \quad \textit{for } i = 1, \cdots, m. \tag{3.10}$$

If in addition $(y^k)_{k=1}^{\infty}$ is bounded, then $(x^k)_{k=1}^{\infty}$ is an asymptotically minimizing sequence for (P).

Proof. From (3.9) and (3.8) we have

$$L(x^k, y^k, r_k) \leqq g(y^k, r_k) + \alpha_k \leqq \sup(\mathrm{D}) + \alpha_k < +\infty. \tag{3.11}$$

In particular, sup (D) is finite. On the other hand, (2.5) and (2.2) imply

$$L(x^k, y^k, r_k) = f_0(x^k) + y^k \cdot u^k + r_k |u^k|^2, \tag{3.12}$$

where

$$u_i^k = \max\{f_i(x^k), -y_i^k / 2r_k\} \quad \text{for } i = 1, \cdots, m. \tag{3.13}$$

Therefore, using (2.1) and (2.6),

$$(3.14)\qquad \begin{aligned} L(x^k, y^k, r_k) &\geqq p(u^k) + y^k \cdot u^k + (r_k - \delta)|u^k|^2 + \delta|u^k|^2 \\ &\geqq g(y^k, r_k - \delta) + \delta|u^k|^2. \end{aligned}$$

We combine (3.14) with (3.11) to obtain

$$(3.15)\qquad \delta|u^k|^2 \leqq \sup(\mathrm{D}) - g(y^k, r_k - \delta) + \alpha_k \to 0.$$

Thus $u^k \to 0$, and this establishes in view of (3.13) that (3.7) and (3.10) hold. Next we argue from (3.11) and (3.14) that

$$(3.16)\qquad \lim_{k\to\infty} L(x^k, y^k, r_k) = \sup(\mathrm{D}).$$

If the y^k sequence is bounded, then (3.12) and the fact that $u^k \to 0$ give us

$$\lim_{k\to\infty} f_0(x^k) = \sup(\mathrm{D}).$$

But sup (D), since it is finite, is the asymptotic optimal value in (P) by Theorem 2. This completes the proof.

The need for the boundedness of $(y^k)_{k=1}^{\infty}$ in Theorem 3, even in the convex case, is illustrated by the following counterexample.

Example 1. Define f_0, f_1, f_2 for $x = (x_1, x_2, x_3) \in R^3$ by $f_0(x) = x_3, f_1(x) = x_1$, $f_2(x) = x_2$. Let

$$S = \{x \in R^3 | x_1 y + x_2 y_2 - x_3 \leqq 0 \text{ for all } (y_1, y_2) \in C\},$$

where

$$C = \{y \in R^2 | y_1 \leqq 0, y_1^2 + 2y_2 \leqq 0\}.$$

Note that S is a closed convex cone which can also be expressed as

$$S = \{x \in R^3 | x_3 \geqq \phi(x_1, x_2)\},$$

where ϕ is the *support function* of C:

$$\begin{aligned} \phi(x_1, x_2) &= \sup\{x_1 y_1 + x_2 y_2 | (y_1, y_2) \in C\} \\ &= \begin{cases} x_1^2/2x_2 & \text{if } x_1 \leqq 0 \text{ and } x_2 > 0, \\ 0 & \text{if } x_1 \geqq 0, x_2 \geqq 0, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

The function ϕ is nonincreasing in x_1 and x_2, so obviously

$$p(u_1, u_2) = \phi(u_1, u_2) \quad \text{for all } u_1, u_2.$$

It can be shown, incidentally, from this fact and formula (2.6) by means of elementary results about conjugate functions, that

$$g(y, r) = -(1/4r)\,\mathrm{dist}^2(-y, C).$$

All we really need to know at the moment, however, is that $g(y, r) \leqq 0$ everywhere and

$$(3.17)\qquad g(y, r) = (1/r)g(y, 1) \quad \text{for } r > 0.$$

These relations follow from (2.6) because $p(0) = 0$ and $p(u/r) = p(u)/r$. Let

$$u^k = (-k^{-1}, k^{-3}), \quad x^k = (-k^{-1}, k^{-3}, k),$$

$$r_k \equiv 1, \quad y^k = -\nabla p(u^k) - 2u^k = (k^2 + 2k^{-1}, (1/2)k^4 - 2k^{-3}).$$

Then if y^k and r_k are substituted into (2.6), the minimum is attained uniquely at u^k, indicating that

$$g(y^k, r_k) = -|u^k|^2 \to 0 = \sup(\mathrm{D}).$$

Hence also $g(y^k, r_k - \delta) \to \sup(\mathrm{D})$ by (3.17), if $0 < \delta < 1$. On the other hand, the minimum in (2.1) for $u = u^k$ is attained uniquely at x^k. Thus x^k uniquely minimizes $L(\cdot, y^k, r_k)$ over S (cf. (2.5)), and all the assumptions in Theorem 3 are satisfied except for the boundedness of $(y^k)_{k=1}^{\infty}$. But $f_0(x^k) = k \to +\infty$, so that $(x^k)_{k=1}^{\infty}$ is certainly not an asymptotically minimizing sequence for (P).

Two corollaries of Theorem 3 may now be stated.

COROLLARY 3.1. *Assume the asymptotic optimal value in* (P) *is not* $+\infty$. *Fix any* $y \in R^m$. *Let* x^k *satisfy*

$$L(x^k, y, r_k) \leqq \inf_{x \in S} L(x, y, r_k) + \alpha_k, \tag{3.18}$$

where $r_k \to +\infty$ *and* $\alpha_k \to 0$. *Then* $(x^k)_{k=1}^{\infty}$ *is an asymptotically minimizing sequence for* (P).

Proof. With $y^k \equiv y$, we have (3.8) by Corollary 1.2 so that the conclusions of Theorem 3 are justified.

COROLLARY 3.2. *Let* $(\bar{y}, \bar{r})$ *be such that for some* $\delta > 0$ *one has* $(\bar{y}, \bar{r} - \delta) \in T$ *and*

$$-\infty < g(\bar{y}, \bar{r} - \delta) = \sup(\mathrm{D}). \tag{3.19}$$

Let $(x^k)_{k=1}^{\infty}$ *be a minimizing sequence in* S *for the function* $L(\cdot, \bar{y}, \bar{r})$. *Then* $\bar{y} \geqq 0$, *and* $(x^k)_{k=1}^{\infty}$ *is an asymptotically minimizing sequence for* (P). *Moreover, if* $\bar{x}$ *is a point at which the minimum of* $L(\cdot, \bar{y}, \bar{r})$ *over* S *is attained, then* $\bar{x}$ *is actually an optimal solution to* (P).

Proof. Take $(y^k, r^k) \equiv (\bar{y}, \bar{r})$ in Theorem 3. For the final assertion of the corollary, take $x^k \equiv \bar{x}$.

Theorem 3 makes clear the computational relevance of the questions of when sup (D) equals inf (P) and when sup (D) is attained. These questions are answered in the next section in terms of the stability of (P).

4. Duality theorems and stability. Problem (P) will be called (*lower*) *stable of degree k* (where k is a nonnegative integer) if there is an open neighborhood U of the origin in R^m and a function $\pi: U \to R$ of class C^k such that

$$p(u) \geqq \pi(u) \quad \text{for all } u \in U, \quad \text{with } p(0) = \pi(0). \tag{4.1}$$

This implies of course that inf (P) is finite.

Stability of degree 0 is equivalent to the property that

$$p(0) = \liminf_{u \to 0} p(u) \quad \text{(finite)}. \tag{4.2}$$

The necessity of (4.2) is evident. On the other hand, if (4.2) holds, then the nonincreasing function

$$\theta(s) = \inf_{|u| \leqq s} p(u), \qquad s \geqq 0,$$

satisfies $\theta(s) \to \theta(0)$ as $s \to 0$. Choose $\varepsilon > 0$ small enough that $\theta(\varepsilon) > -\infty$, and define the function θ_0 on $[0, \varepsilon/2]$ as follows: $\theta_0(0) = \theta(0)$, $\theta_0(\varepsilon/(j+1)) = \theta(\varepsilon/j)$ for positive integers j, θ_0 interpolated linearly over the intervals $[\varepsilon/(j+1), \varepsilon/j]$. Then θ_0 is continuous and $\theta_0 \leqq \theta$. The definition of stability of degree 0 is therefore satisfied by $\pi(u) = \theta_0(|u|)$.

Theorem 2 therefore gives us the following.

THEOREM 4. *Suppose that* (P) *satisfies the quadratic growth condition. In order that the duality relation*

$$\inf(\text{P}) = \sup(\text{D}) \tag{4.3}$$

hold, or equivalently

$$\inf_S \sup_T L(x, y, r) = \sup_T \inf_S L(x, y, r), \tag{4.4}$$

it is necessary and sufficient that (P) *be stable of degree* 0.

Various conditions are known which guarantee stability of degree 0, i.e., (4.2). The most basic perhaps is the following: S is closed, the functions f_i are all lower semicontinuous, and for some $u \in \operatorname{int} R^m_+$ and $\alpha > \inf(\text{P})$ the set

$$\{x \in S \mid f_0(x) \leqq \alpha, f_1(x) \leqq u_1, \cdots, f_m(x) \leqq u_m\} \tag{4.5}$$

is compact. (This is evident from the characterization of (3.5) in terms of asymptotically minimizing sequences.) In the convex case, the Slater condition and its variants suffice [20], [24].

Stability of degree 1 is a generalization of the stability condition in convex programming that p be subdifferentiable at $u = 0$. As a matter of fact, in the convex case stability of degree 1 implies stability of all higher orders. In the absence of convexity, however, stability of degree 2 plays an essential role.

THEOREM 5. *Suppose that* (P) *satisfies the quadratic growth condition. In order that the duality relation*

$$\inf(\text{P}) = \max(\text{D}) \tag{4.6}$$

hold, or equivalently,

$$\inf_S \sup_T L(x, y, r) = \max_T \inf_S L(x, y, r), \tag{4.7}$$

it is necessary and sufficient that (P) *be stable of degree* 2. *Indeed,* $(\bar{y}, \bar{r})$ *is an optimal solution to* (D) *for some* $\bar{r} > 0$ *if and only if* $\bar{y} = -\nabla\pi(0)$ *for some function* π *as in the definition of stability of degree* 2.

Proof. Clearly (4.6) is equivalent to the existence of $(\bar{y}, \bar{r}) \in T$ such that

$$\inf(\text{P}) \leqq g(\bar{y}, \bar{r}) > -\infty, \tag{4.8}$$

since $\inf(\text{P}) \geqq \sup(\text{D})$ in general, while $\sup(\text{D}) > -\infty$ by Theorem 2. Using (2.6), we can write (4.8) in the form

$$-\infty < p(0) \leqq p(u) + \bar{y} \cdot u + \bar{r}|u|^2 \quad \text{for all } u \in R^m. \tag{4.9}$$

If this is fulfilled, then $p(0)$ is finite and the condition for (P) to be stable of degree 2 is satisfied with

$$\pi(u) = p(0) - \bar{y}\cdot u - \bar{r}|u|^2, \qquad U = R^m.$$

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

Assume now conversely that the stability condition is satisfied for a certain π and U. Then $\pi(0) = p(0)$ (finite). Define $\bar{y} = -\nabla\pi(0)$, and choose $\varepsilon > 0$ small enough that $|u| \leqq \varepsilon$ implies $u \in U$. Since π is of class C^2, there is an $r_1 > 0$ such that

$$z\cdot\nabla^2\pi(u)z \geqq -2r_1|z|^2 \quad \text{for all } z \in R^m \text{ if } |u| \leqq \varepsilon. \tag{4.10}$$

Then

$$\pi(u) \geqq p(0) - \bar{y}\cdot u - r_1|u|^2 \quad \text{if } |u| \leqq \varepsilon. \tag{4.11}$$

This follows from the fact that for $h(t) = \pi(tu)$, $0 \leqq t \leqq 1$, one has

$$h(1) = h(0) + \int_0^1 \left[h'(0) + \int_0^t h''(\tau)\,d\tau \right] dt,$$

where

$$h''(\tau) = u\cdot\nabla^2\pi(tu)u.$$

Since (P) satisfies the quadratic growth condition, there exist numbers q and r such that (3.2) holds. We can choose $r_2 > 0$ so that

$$q - r|u|^2 \geqq p(0) - \bar{y}\cdot u - r_2|u|^2 \quad \text{if } |u| \geqq \varepsilon. \tag{4.12}$$

Then (4.12) and (3.2) imply

$$p(u) \geqq p(0) - \bar{y}\cdot u - r_2|u|^2 \quad \text{if } |u| \geqq \varepsilon,$$

while (4.11) and (4.1) imply

$$p(u) \geqq p(0) - \bar{y}\cdot u - r_1|u|^2 \quad \text{if } |u| \leqq \varepsilon.$$

Taking $\bar{r} = \max\{r_1, r_2\}$, we have (4.9), and hence equivalently (4.6) as already noted.

COROLLARY 5.1. *Suppose* (P) *satisfies the quadratic growth condition and is stable of degree* 0. *Then* (D) *has an optimal solution if and only if* (P) *is stable of degree* 2.

Proof. This is obtained by combining Theorem 5 with Theorem 2.

COROLLARY 5.2. *Suppose* (P) *satisfies the quadratic growth condition and is stable of degree* 2. *In order that* $\bar{x} \in S$ *be an optimal solution to* (P), *it is necessary and sufficient that there exist* $(\bar{y}, \bar{r}) \in T$ *such that*

$$L(x, \bar{y}, \bar{r}) \geqq L(\bar{x}, \bar{y}, \bar{r}) \geqq L(\bar{x}, y, r) \quad \text{for all } x \in S, \quad (y, r) \in T. \tag{4.13}$$

Moreover, this condition is satisfied by $(\bar{y}, \bar{r})$ *if and only if* $(\bar{y}, \bar{r})$ *is an optimal solution to* (D).

Proof. The saddle-point condition (4.13) is equivalent by virtue of (1.8) and (1.10) to $\bar{x}$ being a feasible solution to (P) such that

$$f_0(\bar{x}) = \min(\mathrm{P}) = \max(\mathrm{D}) = g(\bar{y}, \bar{r}), \tag{4.14}$$

in which case the common value in (4.14) is $L(\bar{x}, \bar{y}, \bar{r})$.

Remark. If there exist $\bar{x} \in S$ and $(\bar{y}, \bar{r}) \in T$ satisfying (4.13), and therefore (4.14), then (P) must satisfy the quadratic growth condition (cf. remark preceding Theorem 2) and hence be stable of degree 2 (Theorem 5). Compare also with Corollary 3.2.

Corollary 5.2 may be regarded as a generalization of the Kuhn–Tucker theorem in convex programming. Qualitatively, we may expect that most problems encountered in practice will be stable of degree 2, so that the result will be applicable. But, as in the case of "constraint qualifications" and other familiar conditions in the theory of nonlinear programming, it is hard to give verifiable criteria directly in terms of the constraint functions (rather than an unknown optimal solution) which imply such stability. Of course, convexity plus some form of the Slater condition is sufficient. In the next section we investigate the nonconvex case further in terms of the local conditions which are usually satisfied by optimal solutions to (P).

It should be emphasized that the saddle-point relation (4.13) does yield the usual differential Kuhn–Tucker conditions if $\bar{x} \in \operatorname{int} S$ and the functions f_i are differentiable at $\bar{x}$. Indeed, (4.13) implies

$$(4.15) \qquad 0 = \frac{\partial L}{\partial y_i}(\bar{x}, \bar{y}, \bar{r}) = \max\{f_i(\bar{x}), -\bar{y}_i/2\bar{r}\} \quad \text{for } i = 1, \cdots, m,$$

$$(4.16) \qquad \begin{aligned} 0 = \nabla_x L(\bar{x}, \bar{y}, \bar{r}) &= \nabla f_0(\bar{x}) + \sum_{i=1}^{m} \max\{0, \bar{y}_i + 2rf_i(\bar{x})\}\nabla f_i(\bar{x}) \\ &= \nabla f_0(\bar{x}) + \sum_{i=1}^{m} [\bar{y}_i + 2\bar{r}\max\{f_i(\bar{x}), -\bar{y}_i/2\bar{r}\}]\nabla f_i(\bar{x}), \end{aligned}$$

or in other words,

$$(4.17) \qquad f_i(\bar{x}) \leqq 0, \quad \bar{y}_i \geqq 0, \quad \bar{y}_i f_i(\bar{x}) = 0 \quad \text{for } i = 1, \cdots, m,$$

$$(4.18) \qquad \nabla f_0(\bar{x}) + \bar{y}_1 \nabla f_1(\bar{x}) + \cdots + \bar{y}_m \nabla f_m(\bar{x}) = 0.$$

At all events, the vectors $\bar{y}$ involved in Theorem 5 and its corollaries can be interpreted in terms of "equilibrium prices" for perturbations of (P). As seen at the beginning of the proof of Theorem 5, a pair $(\bar{y}, \bar{r}) \in T$ satisfies

$$\inf(\mathrm{P}) = \sup(\mathrm{D}) = g(\bar{y}, \bar{r})$$

if and only if

$$(4.19) \qquad p(u) + \bar{y} \cdot u + \bar{r}|u|^2$$

is minimized in u when $u = 0$. Let us imagine an "economic" situation where we are allowed to perturb (P) by replacing the constraint functions f_i by $f_i - u_i$, so as to obtain perhaps a lower minimum "cost" value $p(u)$, but the cost associated with the perturbation vector $u = (u_1, \cdots, u_m)$ is $\bar{y} \cdot u + \bar{r}|u|^2$. The expression in (4.19) gives the resulting total cost associated with the perturbed problem. Thus (4.19) describes the "equilibrium" where the costs are such that no advantage is to be gained from perturbation, and we are "content with (P) as it is." In particular, we would have (assuming $p(0)$ finite):

$$-\bar{y}\cdot u \leqq \liminf_{\lambda\downarrow 0}\frac{p(\lambda u)-p(0)}{\lambda} \quad \text{for all } u. \tag{4.20}$$

As is well known, such a global "equilibrium" cannot be achieved with $\bar{r} = 0$ unless, at the very least, p coincides at 0 with its convexification, a property which is very unlikely in nonconvex programming.

5. Local criterion for stability of degree 2. We consider now an $\bar{x}$ which is an optimal solution to (P) and show that, if certain conditions slightly stronger than those usually necessary for optimality are satisfied at $\bar{x}$, (P) must be stable of degree 2. In doing this, we extend a result of Arrow, Gould and Howe [1], as described in the Introduction.

The point $\bar{x}$ is said to be the *unique optimal solution to* (P) *in the strong sense* if every asymptotically minimizing sequence for (P) converges to $\bar{x}$. This condition is milder than it might seem. For example, assuming the functions f_i are lower semicontinuous, it is satisfied if S is replaced by any compact subset in which $\bar{x}$ is the only locally optimal solution to (P).

The following conditions are well known to be sufficient (and "almost necessary") for $\bar{x}$ to be an isolated locally optimal solution to (P) (cf. [5, p. 30]):

(a) S contains an open neighborhood N_0 of $\bar{x}$ on which the functions f_i are all of class C^2;

(b) there is a vector $\bar{y}\in R^m$ such that the Kuhn–Tucker conditions (4.17) and (4.18) hold;

(c) for the Hessian matrix

$$H = \nabla^2 f_0(\bar{x}) + \bar{y}_1\nabla^2 f_1(\bar{x}) + \cdots + \bar{y}_m\nabla^2 f_m(\bar{x}) = \nabla_x^2 L_0(\bar{x},\bar{y}) \tag{5.1}$$

and the index sets

$$I_0 = \{i \neq 0 \mid f_i(\bar{x}) = 0, \bar{y}_i > 0\}, \qquad I_1 = \{i \neq 0 \mid f_i(\bar{x}) = 0, \bar{y}_i = 0\}, \tag{5.2}$$

one has $z\cdot Hz > 0$ for every nonzero $z\in R^m$ such that

$$z\cdot\nabla f_i(\bar{x}) = 0 \quad \text{for all } i\in I_0 \quad \text{and} \quad z\cdot\nabla f_i(\bar{x}) \leqq 0 \quad \text{for all } i\in I_1. \tag{5.3}$$

These will be referred to as the *standard* (*second order*) *sufficiency conditions.*

THEOREM 6. *Suppose* (P) *satisfies the quadratic growth condition. Let* $\bar{x}$ *be the unique optimal solution to* (P) *in the strong sense, and assume that* $\bar{x}$ *satisfies the standard sufficiency conditions with* $\bar{y}$ *as the vector of multipliers. Then* (P) *is stable of degree* 2, *and for all* $\bar{r}$ *sufficiently large the pair* $(\bar{y},\bar{r})$ *is an optimal solution to* (D).

Proof. Let $N \subset S$ denote a neighborhood of $\bar{x}$, the nature of which will be specified later, and define

$$p^0(u) = \inf\{f_0(x) \mid x\in N \text{ and } f_i(x) \leqq u_i, i = 1,\cdots,m\}. \tag{5.4}$$

Since $\bar{x}$ is the unique optimal solution to (P) in the strong sense, there exists $\varepsilon > 0$ such that $x\in N$ whenever $x\in S$, $f_i(x) \leqq \varepsilon$ for $i = 1,\cdots,m$, and $f_0(x) \leqq f_0(\bar{x}) + \varepsilon$. Then

$$p(u) = p^0(u) \quad \text{for all } u\in U_0, \tag{5.5}$$

where

(5.6) $$U_0 = \{u \in R^m | u_i \leqq \varepsilon \text{ for } i = 1, \cdots, m \text{ and } p(u) < p(0) + \varepsilon\}.$$

Suppose we can construct a function π of class C^2 on an open neighborhood U_1 of the origin in R^m such that

(5.7) $$p^0(u) \geqq \pi(u) \quad \text{for all } u \in U_1, \quad \text{with } p^0(0) = \pi(0).$$

We will then have

(5.8) $$p(u) \geqq \pi(u) \quad \text{for all } u \in U_0 \cap U_1 \quad \text{with } p(0) = \pi(0),$$

so that the definition of stability of degree 2 will be satisfied with

(5.9) $$U = \{u \in U_1 | \pi(u) < \pi(0) + \varepsilon \text{ and } u_i < \varepsilon \text{ for } i = 1, \cdots, m\}.$$

(If $u \in U$ but $u \notin U_0$, we have $p(u) \geqq p(0) + \varepsilon = \pi(0) + \varepsilon$ and hence $p(u) > \pi(u)$.) If also $\nabla\pi(0) = -\bar{y}$, then $(\bar{y}, \bar{r})$ is an optimal solution to (D) for all $\bar{r}$ sufficiently large by Theorem 5 and the monotonicity of $g(y, r)$ in r. Thus the proof of the theorem is reduced to the construction of N, U_1 and π satisfying (5.7), such that π is of class C^2 on U_1 and $\nabla\pi(0) = -\bar{y}$.

It will be enough actually to show the existence of N such that, for some $\bar{r} > 0$,

(5.10) $$L(x, \bar{y}, \bar{r}) \geqq L(\bar{x}, \bar{y}, \bar{r}) = f_0(\bar{x}) \quad \text{for all } x \in N.$$

Indeed, this will imply from (2.5) that

$$f_0(\bar{x}) = \inf_{x \in N} L(x, \bar{y}, \bar{r}) = \inf_{x \in N} \inf_{u \in R^m} \{F(x, u) + \bar{y} \cdot u + \bar{r}|u|^2\}$$

(5.11) $$= \inf_{u \in R^m} \inf_{x \in N} \{F(x, u) + \bar{y} \cdot u + \bar{r}|u|^2\}$$

$$= \inf_{u \in R^m} \{p^0(u) + \bar{y} \cdot u + \bar{r}|u|^2\}.$$

Since $p^0(0) = f_0(\bar{x})$, we will then have

(5.12) $$p^0(u) \geqq p^0(0) - \bar{y} \cdot u - \bar{r}|u|^2 \quad \text{for all } u \in R^m.$$

In other words, the desired properties will hold for $\pi(u) = p^0(0) - \bar{y} \cdot u - \bar{r}|u|^2$ and $U_1 = R^m$.

Let I_0 and I_1 be the index sets in (5.2), and let

(5.13) $$I_2 = \{i \neq 0 | f_i(\bar{x}) < 0\}.$$

Let N_0 be the neighborhood of $\bar{x}$ in the standard sufficiency conditions. For all $r > 0$, define

(5.14) $$N_1(r) = N_0 \bigcap_{i \in I_0} \{x | f_i(x) > -\bar{y}_i/2r\} \bigcap_{i \in I_2} \{x | f_i(x) < 0\}.$$

Then $N_1(r)$ is an open neighborhood of $\bar{x}$, and for all $x \in N_1(r)$ we have

(5.15) $$L(x, \bar{y}, r) = f_0(x) + \sum_{i \in I_0} [\bar{y}_i f_i(x) + r f_i(x)^2] + r \sum_{i \in I_1} \theta(f_i(x))^2,$$

where

(5.16) $$\theta(\alpha) = \max\{\alpha, 0\}.$$

Observe that $L(\bar{x}, \bar{y}, r) = f_0(\bar{x})$, and by the Kuhn–Tucker conditions

$$\nabla_x L(\bar{x}, \bar{y}, r) = \nabla f_0(\bar{x}) + \sum_{i \in I_0} \bar{y}_i \nabla f_i(\bar{x}) = 0. \tag{5.17}$$

We shall show next that in fact

$$L(x, \bar{y}, r) = f_0(\bar{x}) + h(x - \bar{x}) + rk(x - \bar{x}) + o(|x - \bar{x}|^2), \tag{5.18}$$

where $h(z) = z \cdot Hz$ and

$$k(z) = \sum_{i \in I_0} (z \cdot \nabla f_i(\bar{x}))^2 + \sum_{i \in I_1} \theta(z \cdot \nabla f_i(\bar{x}))^2 \geqq 0. \tag{5.19}$$

Since by (5.17)

$$f_0(x) + \sum_{i \in I_0} \bar{y}_i f_i(x) = f_0(\bar{x}) + h(x - \bar{x}) + o(|x - \bar{x}|^2), \tag{5.20}$$

and since (for $f_i(\bar{x}) = 0$)

$$f_i(x)^2 = ((x - \bar{x}) \cdot \nabla f_i(\bar{x}))^2 + o(|x - \bar{x}|^2), \tag{5.21}$$

we need only prove that the expansion

$$\theta(f_i(x))^2 = \theta((x - \bar{x}) \cdot \nabla f_i(\bar{x}))^2 + o(|x - \bar{x}|^2) \tag{5.22}$$

is valid when $f_i(\bar{x}) = 0$. This amounts to establishing that

$$0 = \lim_{t \downarrow 0} \frac{\theta(f_i(\bar{x} + tz))^2 - \theta(tz \cdot \nabla f_i(\bar{x}))^2}{t^2} \tag{5.23}$$

uniformly in $z \in B$, where

$$B = \{z \in R^n \mid |z| = 1\}.$$

But the latter is obvious from the continuity of θ and the fact that the difference quotient in (5.23) can be rewritten as

$$\theta(f_i(\bar{x} + tz)/t)^2 - \theta(z \cdot \nabla f_i(\bar{x}))^2 = \theta(z \cdot \nabla f_i(\bar{x}) + w(tz))^2 - \theta(z \cdot \nabla f_i(\bar{x}))^2,$$

where $w(tz) \to 0$ uniformly in $z \in B$ as $t \downarrow 0$.

We now demonstrate the existence of $\bar{r} > 0$ and $\delta > 0$ with

$$h(z) + \bar{r}k(z) \geqq 2\delta \quad \text{for all } z \in B. \tag{5.24}$$

Let $B_0 = \{z \in B \mid h(z) \leqq 0\}$. According to part (c) of the sufficiency conditions, if $k(z) = 0$, i.e., (5.3) holds, we have $h(z) > 0$. Thus $k(z) > 0$ for all $z \in B_0$, implying that the quotient $-h(z)/k(z)$ is well-defined and bounded above as a function of $z \in B_0$. Choose any $\bar{r} > 0$ such that

$$\bar{r} > -h(z)/k(z) \quad \text{for all } z \in B_0.$$

Then $h(z) + \bar{r}k(z) > 0$ for all $z \in B_0$; the same inequality also holds trivially for $z \in B \setminus B_0$, because there $h(z) > 0$ and $k(z) \geqq 0$. Thus $h + \bar{r}k$ is a positive, continuous function on the compact set B, and (5.24) is valid for some $\delta > 0$ as claimed. Of course (5.24) implies

$$h(z) + \bar{r}k(z) \geqq 2\delta |z|^2 \quad \text{for all } z \in R^m, \tag{5.25}$$

because h and k are both positively homogeneous of degree 2.

It remains only to combine (5.25) with (5.18). There exists by (5.18) a neighborhood N of $\bar{x}$, $N \subset N_1(\bar{r})$, such that

(5.26) $$L(x, \bar{y}, \bar{r}) \geqq f_0(\bar{x}) + h(x - \bar{x}) + \bar{r}k(x - \bar{x}) - \delta|x - \bar{x}|^2$$

for all $x \in N$. Using (5.25), we obtain

(5.27) $$L(x, \bar{y}, \bar{r}) \geqq f_0(\bar{x}) + \delta(x - \bar{x})^2 \quad \text{for all } x \in N.$$

Thus (5.10) holds for N and $\bar{r}$, and the proof of Theorem 6 is complete.

COROLLARY 6.1. *Under the assumptions in Theorem* 6, *the global saddle-point condition* (4.13) *holds for all* $\bar{r}$ *sufficiently large.*

Proof. The proof is immediate from Corollary 5.2.

We conclude this section with a counterexample demonstrating the need for the second order condition in the hypothesis of Theorem 6.

Example 2. Here all the assumptions in Theorem 6 are satisfied, except for a slight weakening of part (c) of the sufficiency conditions, and (P) is stable of degree 1. But (P) is not stable of degree 2. The problem consists of minimizing

$$f_0(x_1, x_2) = 4x_1(x_2 - 1) + x_2^4$$

over $S = \{x = (x_1, x_2) \in R^2 | -1 \leqq x_1 \leqq 1\}$ subject to

$$0 \geqq f_1(x_1, x_2) = x_1.$$

The minimum of $f_0(x_1, x_2)$ in x_2 for fixed x_1 is $-4x_1 - 3x_1^{4/3}$, attained only at $x_2 = -x_1^{1/3}$, and this minimum is a strictly decreasing function of x_1 as long as $x_1 \geqq -1$. Thus $\bar{x} = (0, 0)$ is the unique optimal solution to (P) in the strong sense. The quadratic growth condition is satisfied, because f_0 is bounded below on S. Furthermore, the Kuhn–Tucker conditions hold at $\bar{x}$ with $\bar{y}_1 = 4$ and with the gradients $\nabla f_0(\bar{x})$ and $\nabla f_1(\bar{x})$ nonzero (thus one has "strict complementarity" in (4.17), and moreover "the gradients of the active constraints at $\bar{x}$ form a linearly independent set"). Although the Hessian matrix H of the function $l(x) = f_0(x) + \bar{y}_1 f_1(x)$ at $\bar{x}$ does not have the positive definiteness property required in (c) of the sufficiency conditions, it is true at least that $l(\bar{x} + z) > l(\bar{x})$ for every nonzero z such that (5.3) holds (i.e., $z \cdot \nabla f_1(\bar{x}) = 0$). However,

$$p(u_1) = -4u_1 - 3u_1^{4/3} \quad \text{for } u_1 \in [-1, 1].$$

The function p is continuously differentiable around $u_1 = 0$, but it does not majorize near 0 any function π of class C^2 such that $\pi(0) = p(0) = 0$. Thus (P) is stable of degree 1 but not of degree 2.

Remark. We have already noted towards the end of § 4 that, if $(\bar{x}, \bar{y}, \bar{r})$ is a saddle point of L and the functions f_i are differentiable at $\bar{x}$ (and $\bar{x} \in \text{int } S$), then $\bar{x}$ and $\bar{y}$ satisfy the Kuhn–Tucker conditions. In fact, if every f_i is twice-differentiable at $\bar{x}$, then the standard *second order* necessary conditions [5, p. 25] are satisfied, i.e., besides the Kuhn–Tucker conditions one has condition (c) at the beginning of this section, but with the inequality $z \cdot Hz > 0$ weakened to $z \cdot Hz \geqq 0$. This is true because (5.15) holds (with $\bar{r}$ in place of r) for all x in some neighborhood of $\bar{x}$, so that the right side of (5.15) must have a local minimum at $x = \bar{x}$. From (5.21) and (5.22), it is clear that the latter implies

$$(5.28) \qquad 0 \leqq \frac{d^2}{dt^2}[f_0(\bar{x} + tz) + \sum_{i \in I_0} \bar{y}_i f_i(\bar{x} + tz)]|_{t=0}$$

for all z satisfying (5.3), and this derivative equals $z \cdot Hz$.

Thus for twice-differentiable functions f_i and S open, the situation can be summarized as follows. *If $(\bar{x}, \bar{y}, \bar{r})$ is a saddle point of L for some $\bar{r} \geqq 0$, then $\bar{x}$ and $\bar{y}$ satisfy the standard second order necessary conditions for optimality, and $\bar{x}$ is (globally) optimal. On the other hand, if $\bar{x}$ and $\bar{y}$ satisfy the standard second order sufficient conditions and $\bar{x}$ is the unique (globally) optimal solution in the strong sense, and the quadratic growth condition is satisfied, then $(\bar{x}, \bar{y}, \bar{r})$ is a saddle point of L for some $\bar{r} \geqq 0$.*

REFERENCES

[1] K. J. Arrow, P. J. Gould and S. M. Howe, *A general saddle point result for constrained optimization*, Institute of Statistics Mimeo Series No. 774, Univ. of North Carolina, Chapel Hill, 1971. (Revised version: Nov. 1972.)

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

[3] D. P. Bertsekas, *Combined primal-dual and penalty methods for constrained minimization*, this Journal, 13 (1975), to appear.

[4] J. D. Buys, *Dual algorithms for constrained optimization*, Thesis, Leiden, 1972.

[5] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Optimization Techniques*, John Wiley, New York, 1968.

[6] R. Fletcher, *A class of methods for nonlinear programming with termination and convergence properties*, Integer and Nonlinear Programming, J. Abadie, ed., North-Holland, Amsterdam, 1970.

[7] ———, *A class of methods for non-linear programming. III: Rates of convergence*, Numerical Methods for Non-Linear Optimization, F. A. Lootsma, ed., Academic Press, New York, 1973.

[8] R. Fletcher and S. Lill, *A class of methods for nonlinear programming, II: Computational experience*, Nonlinear Programming, J. B. Rosen, O. L. Mangasarian and K. Ritter, eds., Academic Press, New York, 1970.

[9] P. C. Haarhoff and J. D. Buys, *A new method for the optimization of a nonlinear function subject to nonlinear constraints*, Comput. J., 13 (1970), pp. 178–184.

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

[11] B. W. Kort and D. P. Bertsekas, *A new penalty function method for constrained minimization*, Proc. IEEE Decision and Control Conference, New Orleans, 1972.

[12] S. A. Lill, *Generalization of an exact method for solving equality constrained problems to deal with inequality constraints*, Numerical Methods for Nonlinear Optimization, F. A. Lootsma, ed., Academic Press, New York, 1973.

[13] O. L. Mangasarian, *Unconstrained Lagrangians in nonlinear programming*, Computer Sciences Tech. Rep. 174, Univ. of Wisconsin, Madison, 1973.

[14] A. Miele, E. E. Cragg, R. R. Iver and A. V. Levy, *Use of the augmented penalty function in mathematical programming, part I*, J. Optimization Theory Appl., 8 (1971), pp. 115–130.

[15] A. Miele, E. E. Cragg and A. V. Levy, *Use of the augmented penalty function in mathematical programming problems, part II*, Ibid., 8 (1971), pp. 131–153.

[16] A. Miele, P. E. Moseley and E. E. Cragg, *A modification of the method of multipliers for mathematical programming problems*, Techniques of Optimization, A. V. Balakrishnan, ed., Academic Press, New York, 1972.

[17] A. Miele, P. E. Moseley, A. V. Levy and G. M. Coggins, *On the method of multipliers for mathematical programming problems*, J. Optimization Theory Appl., 10 (1972), pp. 1–33.

[18] M. A. POLLATSCHEK, *Generalized duality theory in nonlinear programming*, Operations Research Mimeograph Series No. 122, Technion, Haifa, Israel, 1973.

[19] M. J. D. POWELL, *A method for nonlinear optimization in minimization problems*, Optimization, R. Fletcher, ed., Academic Press, New York, 1969.

[20] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton Univ. Press, Princeton, N.J., 1970.

[21] ———, *New applications of duality in convex programming*, Proc. 4th Conference on Probability, Brasov, Romania, 1971. (This is the written version of a talk given at several conferences, including the 7th International Symposium on Mathematical Programming, the Hague, 1970.)

[22] ———, *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Math. Prog., to appear.

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

[24] ———, *Conjugate Duality and Optimization*, SIAM/CBMS lecture note series, Philadelphia, 1974.

[25] ———, *Penalty methods and augmented Lagrangians in nonlinear programming*, Proc. 5th IFIP Conference on Optimization Techniques, Rome, Springer-Verlag, Berlin, 1973.

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

[27] A. P. WIERZBICKI, *Convergence properties of a penalty shifting algorithm for nonlinear programming problems with inequality constraints*, Arch. Automat. i Telemech. (1970).

[28] ———, *A penalty function shifting method in constrained static optimization and its convergence properties*, Ibid., 16 (1971), pp. 395–416.

[29] A. P. WIERZBICKI AND A. HATKO, *Computational methods in Hilbert space for optimal control problems with delays*, Proc. 5th IFIP Conference on Optimization Techniques, Rome, Springer-Verlag, Berlin, 1973.

[30] B. T. POLJAK, *Iteration methods using Lagrange multipliers for the solution of extremal problems with constraints of the equality type*, USSR Comput. Math. and Math. Phys., 10 (1970), no. 5, pp. 1098–1106.