to the objective function. Thus, the problem can be represented by the Lagrangian

$$K_r(x, y) = f_0(x) + \sum_{i=1}^{m} [y_i f_i(x) + r f_i(x)^2], \tag{2}$$

where the constant $r > 0$ is the *penalty factor*. The penalty terms suggest the likelihood that, if $\bar{x}$ is a local optimal solution with corresponding multiplier vector $\bar{y}$ satisfying

$$\nabla f_0(x) + \sum_{i=1}^{m} y_i \nabla f_i(x) = 0, \tag{3}$$

and if r is sufficiently high, then $K_r(x, \bar{y})$ will have a local minimum at $\bar{x}$. In fact, it is easy to show that this is true if $\bar{x}$ satisfies the standard second-order conditions sufficient for local optimality in (1). Therefore, one might attempt to solve (1) by solving a sequence of unconstrained problems of the following form:

$$\text{determine } x^k \text{ minimizing } K_r(x, y^k). \tag{4}$$

Here, r could be increased from time to time if convergence did not seem fast enough. Since x^k satisfies

$$0 = \nabla_x K_r(x^k, y^k) = \nabla f_0(x^k) + \sum_{i=1}^{m} (y_i^k + 2r f_i(x^k)) \nabla f_i(x^k), \tag{5}$$

a simple and natural way to generate the sequence $\{y^k\}$ is to set

$$y_i^{k+1} = y_i^k + 2r f_i(x^k), \qquad i = 1,..., m. \tag{6}$$

Hestenes did not develop this idea theoretically beyond this stage, but Powell worked out an algorithm with a rule for when to increase r. He actually allowed separate penalty factors for each of the terms $f_i(x)^2$. He demonstrated in Ref. 3 that, if the second-order sufficient conditions for optimality were satisfied, the algorithm should converge locally at a linear rate, without the need for having $r \to +\infty$. The main advantage of the algorithm lies in the latter property, since it provides a numerical stability that is not found in the usual penalty methods. Powell did not address the question of to what extent the exact global minimum in (4) can be relinquished in favor of an approximate local minimum.

More recently, Miele *et al.* (Refs. 4–7) have tried out various modifications of this algorithm computationally. Fletcher (Refs. 8–9) has developed a related technique where the multiplier vector y is adjusted *continuously* as the minimization of $K_r(x, y)$ in x is carried out. An earlier method along these general lines may also be found in a paper of Arrow and Solow (Ref. 10), which treats the calculation of saddle points by means of solving differential equations.

Aside from an extension which Arrow and Solow described for their differential equation approach, no specific algorithm related to the method of Hestenes and Powell has been proposed for problems with inequality constraints. However, Rockafellar (Ref. 11) has studied the analogue of the Lagrangian K_r for inequality constraints and, in the convex case, has established some of its general computational properties. A saddle-point theorem for nonconvex, inequality-constrained problems has also been proved by Arrow, Gould, and Howe (Ref. 12).

If the constraints $f_i(x) = 0$ are replaced by $f_i(x) \leq 0$, the natural replacement for the Lagrangian K_r, as demonstrated in Ref. 11, is

$$L_r(x, y) = f_0(x) + (1/4)r \sum_{i=1}^{m} [\theta(y_i + 2r f_i(x))^2 - y_i^2], \tag{7}$$

where

$$\theta(t) = \max\{t, 0\}. \tag{8}$$

Of course,

$$(1/4r)[\theta(y_i + 2r f_i(x))^2 - y_i^2] = \begin{cases} r f_i(x)^2 + y_i f_i(x) & \text{if } f_i(x) \geq -y_i/2r, \\ -y_i^2/4r & \text{if } f_i(x) \leq -y_i/2r. \end{cases} \tag{9}$$

Note that the multipliers y_i are **not** restricted to be nonnegative, despite the inequality constraints in the problem.

It is easy to generalize the algorithmic approach of Hestenes and Powell to the inequality case in terms of L_r. The basic procedure is that, given y^k (and $r > 0$), we

$$\text{determine } x^k \text{ minimizing } L_r(x, y^k), \tag{10}$$

so that

$$0 = \nabla_x L_r(x^k, y^k) = \nabla f_0(x^k) + \sum_{i=1}^{m} \theta(y_i^k + 2r f_i(x^k)) \nabla f_i(x^k). \tag{11}$$

We then set

$$y_i^{k+1} = \theta(y_i^k + 2r f_i(x^k)) \geq 0 \qquad \text{for } i = 1,..., m, \tag{12}$$

or, in other words,

$$y^{k+1} = y^k + 2r \nabla_y L(x^k, y^k). \tag{13}$$

Steps (10) and (12)–(13) make sense, of course, even if the functions are not differentiable. The transformation of this procedure into a locally convergent algorithm patterned after Powell's, where r is increased in a systematic way if necessary, is straightforward and will not be carried out here. Our aim instead is to show using our earlier results in Ref. 11 that, in the case where the functions f_i are convex and not necessarily differentiable, the procedure given by (10) and (12) converges globally for arbitrary fixed r and with the minimization only approximate, at least if some form of the Slater condition is satisfied. By relating the procedure to Moreau's theory of proximal mappings (Ref. 13), we are able to demonstrate the unusual property that the sequence of multiplier vectors y^k then converges, even though there may be more than one possible limit, depending on the initial choice of y^0.

2. Convergence Theorem

Henceforth, let X denote a convex set in a real-vector space E (possibly infinite-dimensional), and let $f_0, f_1, ..., f_m$ be real-valued convex functions on X. We shall be concerned with the following Problem (P):

$$\text{minimize } f_0(x) \text{ over all } x \in X \text{ satisfying } f_i(x) \leq 0 \quad \text{for} \quad i = 1,..., m.$$

Rather than introducing (affine) equality constraints in the model, we suppose for notational simplicity that each such constraint is represented by a pair of inequalities. Equality constraints could also be treated explicitly, if so desired, and this would necessitate only routine and obvious changes in the statements of the results below. The convexity of the functions f_i implies that the Lagrangian $L_r(x, y)$ is convex in $x \in X$ and concave in $y \in R^m$ (Ref. 11, Theorem 1).

The algorithm to be investigated is the following. It depends on the initial choice of $r > 0$, $y^0 \in R^m$, and a sequence $\{\alpha_k\}$ with $0 \leq \alpha_k \to 0$.

Algorithm. Given $y^k \in R^m$, determine $x^k \in X$ such that

$$L_r(x^k, y^k) \leq \inf_{x \in X} L_r(x, y^k) + \alpha_k . \tag{14}$$

Then, define y^{k+1} by (12) or, equivalently, (13).

Let μ denote the optimal value in (P), that is, the infimum subject to the constraints. A *Kuhn–Tucker vector* for (P) is a vector $\bar{y} \in R_+^m$ with the property that

$$\inf_{x \in X} \{f_0(x) + \bar{y}_1 f_1(x) + \cdots + \bar{y}_m f_m(x)\} = \mu > -\infty. \tag{15}$$

Our main result can now be stated.

Theorem 2.1. Suppose that (P) possesses at least one Kuhn–Tucker vector and that

$$\sum_{k=1}^{\infty} \sqrt{\alpha_k} < +\infty. \tag{16}$$

Then, $\{y^k\}$ converges to some Kuhn–Tucker vector, while $\{x^k\}$ satisfies

$$\limsup_{k \to \infty} f_i(x^k) \leq 0 \quad \text{for} \quad i = 1,..., m, \tag{17}$$

$$\lim_{k \to \infty} f_0(x^k) = \mu = \inf(\text{P}). \tag{18}$$

Assuming that the optimal value μ is finite, it is well known (Ref. 14) that a Kuhn–Tucker vector does exist if the *Slater condition* is satisfied: there is a point $x \in X$ such that $f_i(x) < 0$ for $i = 1,..., m$. For problems in which equality constraints have been represented by pairs of (affine) inequalities, this criterion is not applicable, but it may be replaced by the *modified Slater condition*: there is a point $x \in \text{core } X$ such that $f_i(x) \leq 0$ for $i = 1,..., m$, with strict inequality for every f_i which is not affine. The core of X consists of the points x such that every ray emanating from x meets X in at least one point besides x. The argument given in Ref. 14, Theorem 28.2, for the sufficiency of the modified Slater condition carries over to the infinite-dimensional case. More generally, still assuming μ to be finite, a Kuhn–Tucker vector fails to exist iff there is a vector $u \in R^m$ such that

$$\lim_{\lambda \to 0} [p(\lambda u) - p(0)]/\lambda = -\infty, \tag{19}$$

where

$$p(u) = \inf\{f_0(x) \mid x \in X, f_1(x) \leq u_1, ..., f_m(x) \leq u_m\} \tag{20}$$

(see Ref. 15, Corollary 29.1.2). When a Kuhn–Tucker vector exists, it is not possible for any sequence $\{x^k\}$ in X satisfying (17) to satisfy

$$\lim_{k \to \infty} \inf f_0(x_k) < \mu. \tag{21}$$

The proof of Theorem 2.1 is based on results in Ref. 11 concerning the concave function

$$g_r(y) = \inf_{x \in X} L_r(x, y). \tag{22}$$

Under the hypothesis that a Kuhn–Tucker vector exists, g_r is everywhere finite and continuously differentiable on R^m, and the Kuhn–Tucker vectors form its maximizing set (Ref. 11, after Theorem 2). Furthermore,

$$g_r(y) = \max_{w \in R^m} \{g_0(w) - (1/4r)|w - y|^2\}, \tag{23}$$

where

$$g_0(y) = \begin{cases} \inf_{x \in X} \{f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x)\} & \text{if } y \geq 0, \\ -\infty & \text{if } y \not\geq 0, \end{cases} \tag{24}$$

(Ref. 11, Theorem 2); denoting by $M(y)$ the unique w for which the maximum in (23) is attained, we see that $\nabla g_r(y)$ coincides with the gradient of $y \to g_0(w) - (1/4r)| w - y |^2$ at $w = M(y)$, and hence,

$$M(y) = y + 2r \nabla g_r(y). \tag{25}$$

Proof of Theorem 2.1. To obtain the desired conclusion about $\{x^k\}$, it suffices, according to Theorem 4 of Ref. 15, to prove that $\{y^k\}$ is a bounded maximizing sequence for g_r. If, in addition, $\{y^k\}$ is convergent, the whole theorem is established.

As shown in Ref. 11, Lemma 2, condition (14) yields the estimate

$$r \,| \nabla L_r(x^k, y^k) - \nabla g_r(y^k)|^2 \leq \alpha_k . \tag{26}$$

Therefore, by (25) applied to y^k and by (13),

$$(1/r)| y^{k+1} - M(y^k)|^2 \leq \alpha_k ; \tag{27}$$

and, in particular,

$$\lim_{k \to \infty} [y^{k+1} - M(y^k)] = 0. \tag{28}$$

Now, M is by definition the proximal mapping in the sense of Moreau (Ref. 13) associated with the convex function $f = -2rg_0$. Thus, M is nonexpansive (Ref. 13), that is,

$$| M(y) - M(z)| \leq | y - z | \qquad \text{for all} \quad y, z. \tag{29}$$

The fixed points of M are precisely the points y such that $\nabla g_r(y) = 0$; in other words, they are the Kuhn–Tucker vectors for (P). Let $\bar{y}$ be an arbitrary Kuhn–Tucker vector. Then,

$$| M(y^k) - \bar{y} | = | M(y^k) - M(\bar{y})| \leq | y^k - \bar{y} |, \tag{30}$$

by (29), so that, by (27),

$$| y^{k+1} - \bar{y} | \leq | y^{k+1} - M(y^k)| + | M(y^k) - \bar{y} | \leq (r\alpha_k)^{1/2} + | y^k - \bar{y} |. \tag{31}$$

It follows from (16) that

$$| y^l - \bar{y} | \leq | y^k - \bar{y} | + \sum_{s=k}^{\infty} (r\alpha_s)^{1/2} < +\infty \qquad \text{whenever} \quad l > k. \tag{32}$$

In particular, $\{y^k\}$ is a bounded sequence. The uniform continuity of g_r on bounded sets then implies that

$$\lim_{k \to \infty} [g_r(y^{k+1}) - g_r(M(y^k))] = 0, \tag{33}$$

in view of (28). We observe next from (23) that

$$g_r(y^k) = g_0(M(y^k)) - (1/4r)| M(y^k) - y^k |^2;$$

or, in other words, using (25),

$$g_r(y^k) + r \,| \nabla g_r(y^k)|^2 = g_0(M(y^k)). \tag{34}$$

But (23) also yields

$$g_r(M(y^k)) \geq g_0(M(y^k)) - (1/4r)| M(y^k) - M(y^k)|^2 = g_0(M(y^k)). \tag{35}$$

Combining (35) with (34), we have

$$g_r(M(y^k)) \geq g_r(y^k) + r \,| \nabla g_r(y^k)|^2 \qquad \text{for all } k. \tag{36}$$

From (36), (33), and the fact that g_r is bounded above [since the maximizing set for g_r, consisting of the Kuhn–Tucker vectors for (P), is nonempty], we are able to conclude that

$$\lim_{k \to \infty} | \nabla g_r(y^k)| = 0.$$

Consequently, if $\bar{y}$ is an arbitrary cluster point of the bounded sequence $\{y^k\}$, we have $\nabla g_r(\bar{y}) = 0$. Since g_r is concave, this means that $\bar{y}$ maximizes g_r and, hence, is a Kuhn–Tucker vector. The inequality (32), which holds for any Kuhn–Tucker vector, now implies there can be only one such cluster point $\bar{y}$ of $\{y^k\}$. Thus, $\{y^k\}$ is a maximizing sequence for g_r which converges to some Kuhn–Tucker vector for (P), and the proof of the theorem is complete.

References

1. Hestenes, M. R., *Multiplier and Gradient Methods*, Computing Methods in Optimization Problems—2, Edited by L. A. Zadeh, L. W. Neustadt, and A. V. Balakrishnan, Academic Press, New York, New York, 1969.
2. Hestenes, M. R., *Multiplier and Gradient Methods*, Journal of Optimization Theory and Applications, Vol. 4, pp. 303–320, 1969.

3. Powell, M. J. D., *A Method for Nonlinear Constraints in Minimization Problems*, Optimization, Edited by R. Fletcher, Academic Press, New York, New York, 1972.
4. Miele, A., Cragg, E. E., Iyer, R. R., and Levy, A. V., *Use of the Augmented Penalty Function in Mathematical Programming Problems, Part 1*, Journal of Optimization Theory and Applications, Vol. 8, pp. 115–130, 1971.
5. Miele, A., Cragg, E. E., and Levy, A. V., *Use of the Augmented Penalty Function in Mathematical Programming Problems, Part 2*, Journal of Optimization Theory and Applications, Vol. 8, pp. 131–153, 1971.
6. Miele, A., Moseley, P. E., and Cragg, E. E., *A Modification of the Method of Multipliers for Mathematical Programming Problems*, Techniques of Optimization, Edited by A. V. Balakrishnan, Academic Press, New York, New York, 1972.
7. Miele, A., Moseley, P. E., Levy, A. V., and Coggins, G. M., *On the Method of Multipliers for Mathematical Programming Problems*, Journal of Optimization Theory and Applications, Vol. 10, pp. 1–33, 1972.
8. Fletcher, R., *A Class of Methods for Nonlinear Programming with Termination and Convergence Properties*, Integer and Nonlinear Programming, Edited by J. Abadie, North-Holland Publishing Company, Amsterdam, Holland, 1970.
9. Fletcher, R., and Lill, S. A., *A Class of Methods for Nonlinear Programming, II: Computational Experience*, Nonlinear Programming, Edited by J. B. Rosen, O. L. Mangasarian, and K. Ritter, Academic Press, New York, New York, 1971.
10. Arrow, K. J., and Solow, R. M., *Gradient Methods for Constrained Maxima, with Weakened Assumptions*, Studies in Linear and Nonlinear Programming, Edited by K. Arrow, L. Hurwicz, and H. Uzawa, Stanford University Press, Stanford, California, 1958.
11. Rockafellar, R. T., *A Dual Approach to Solving Nonlinear Programming Problems by Unconstrained Optimization*, Mathematical Programming (to appear).
12. Arrow, K. J., Gould, F. J., and Howe, S. M., *A General Saddle Point Result for Constrained Optimization*, Mathematical Programming (to appear).
13. Moreau, J. J., *Proximité et Dualité dans un Espace Hilbertien*, Bulletin de la Société Mathématique de France, Vol. 93, pp. 273–279, 1965.
14. Fan, K., Glicksberg, I., and Hoffman, A. J., *Systems of Inequalities Involving Convex Functions*, Proceedings of the American Mathematical Society, Vol. 8, pp. 617–622, 1957.
15. Rockafellar, R. T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.

Printed by the St Catherine Press Ltd., Tempelhof 37, Bruges, Belgium.

Reprinted from JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS Vol. 12, No. 6, December 1973
Printed in Belgium

The Multiplier Method of Hestenes and Powell Applied to Convex Programming[1]

R. T. ROCKAFELLAR[2]

Communicated by M. R. Hestenes

Abstract. For nonlinear programming problems with equality constraints, Hestenes and Powell have independently proposed a dual method of solution in which squares of the constraint functions are added as penalties to the Lagrangian, and a certain simple rule is used for updating the Lagrange multipliers after each cycle. Powell has essentially shown that the rate of convergence is linear if one starts with a sufficiently high penalty factor and sufficiently near to a local solution satisfying the usual second-order sufficient conditions for optimality. This paper furnishes the corresponding method for inequality-constrained problems. Global convergence to an optimal solution is established in the convex case for an arbitrary penalty factor and without the requirement that an exact minimum be calculated at each cycle. Furthermore, the Lagrange multipliers are shown to converge, even though the optimal multipliers may not be unique.

1. Introduction

The following idea, in effect, was raised independently by Hestenes (Refs. 1–2) and Powell (Ref. 3) at conferences in the spring of 1968. Consider the following nonlinear programming problem:

$$\text{minimize } f_0(x) \text{ subject to } f_i(x) = 0, \qquad i = 1,..., m, \tag{1}$$

where the functions $f_i : R^n \to R$ are differentiable. The solutions to this problem are not altered if terms of the form $rf_i(x)^2$, $i \geq 1$, are added

[1] This work was supported in part by the Air Force Office of Scientific Research under Grant No. AF-AFOSR-72-2269.
[2] Professor, Department of Mathematics, University of Washington, Seattle, Washington.