Chapter 20

Lagrange Multipliers and Variational Inequalities

R. T. Rockafellar

1. INTRODUCTION

Variational inequalities have been used to characterize the solutions to many problems involving partial differential equations with unilateral constraints. The complementarity problem in mathematical programming concerns a special type of variational inequality in finite dimensions that has been the focus of important algorithmic developments. General variational inequalities can be reduced to this special type through discretization and the introduction of Lagrange multipliers, and this provides an approach to computation.

Other approaches are suggested by analogies with convex programming. Many variational inequalities actually express the condition for the minimum of a convex functional relative to a certain convex set which, in the course of 'discretization', is represented by a finite system of convex (or linear) inequalities. A broader class of variational inequalities, covering perhaps the majority of applications, is obtained by replacing the gradient mapping associated with the convex minimand by a mapping that is 'monotone' in the general sense due to Minty. To the extent that algorithms for convex programming can be formulated entirely in terms of the gradient of the minimand, rather than the minimand itself (including such numerical considerations as stopping criteria), one can get computational procedures for variational inequalities that may offer advantages in some cases over complementarity.

For example, penalty methods have been used in the solution of variational inequalities, since they reduce a constrained problem to a sequence of 'unconstrained' problems to which classical numerical techniques can be applied. Nowadays in mathematical programming, penalty methods in pure form are in disrepute because of their inherent numerical instabilities. They have been supplanted by methods that are based on augmented Lagrangian functions and include varying Lagrange multiplier values (dual variables) as well as penalty parameters.

Some efforts have been made to apply such penalty-duality methods to variational inequalities, but only, it seems, in the case of equality constraints, and even then in terms of what is analogous to exact minimization in each unconstrained problem. For greater effectiveness, it is important to have procedures that are capable of handling inequality constraints and can be shown to converge under more practical criteria (tolerance levels for certain accessible quantities).

The purpose of this chapter, besides explaining some of the background to such matters, is to draw attention to a new penalty-duality method that has been designed with these requirements in mind. It converts a variational inequality for a monotone mapping (and a convex set defined by a finite system of differentiable convex constraint functions) into a sequence of unconstrained subproblems, in each of which one calculates an 'approximate' root of a non-linear equation for a certain strongly monotone mapping. The overall convergence rate is generically linear, with a ratio that approaches zero as the penalty parameter is increased and therefore yields superlinear convergence if the penalty parameter goes to infinity. From the close parallel with earlier methods, which are known to be highly effective in non-linear programming but do not carry over so easily in their formulation to general variational inequalities, one can hope for very good results. However, the verdict must await more testing. It is also clear that for applications to variational inequalities it would be helpful to incorporate additional features and flexibility beyond what has seemed possible within the present theoretical framework.

2. THE GENERAL PROBLEM

Let K be a non-empty closed convex set in a real Hilbert space V (finite or infinite-dimensional), and let $A : V \to V$ be a mapping (single-valued) that may be linear or non-linear. The *variational inequality problem* for K, A and an element $a \in V$ is to determine an element u satisfying the conditions

(1) $$u \in K, \qquad \langle A(u) - a, v - u\rangle \geq 0 \quad \text{for all } v \in K$$

Here $\langle \cdot, \cdot \rangle$ denotes the inner product in V.

If K is all of V, (1) reduces to the equation $A(u) = a$. More generally it expresses a normality condition very familiar to everyone who has studied optimization theory. The normal cone to K at a point u consists by definition of all the normal vectors to half-spaces that support K at u:

(2) $$N_K(u) = \{w : \langle w, v - u\rangle \leq 0 \quad \text{for all} \quad v \in K\}$$

This is a closed convex cone containing the origin, and in terms of it we write (1) in the form

(3) $$u \in K \quad \text{and} \quad a - A(u) \in N_K(u)$$

If there is a differentiable function F on V whose gradient mapping satisfies

(4) $$A(v) - a = \nabla F(v) \quad \forall\, v \in V$$

the variational inequality expresses the fundamental first-order necessary condition for u to be a local solution to the optimization problem

(5) $$\text{minimize } F(v) \quad \text{over all } v \in K$$

The condition is sufficient for global optimality when F is convex. Note that when A is linear, the existence of F satisfying (4) is equivalent to A being symmetric, in which event one has

(6) $$F(v) = \tfrac{1}{2}\langle A(v), v\rangle + \langle a, v\rangle + \text{constant}$$

An important concept in this context is that of monotonicity in the sense of Minty. The mapping A is *monotone* if

(7) $$\langle A(v) - A(\bar{v}), v - \bar{v}\rangle \geq 0 \quad \forall v, \forall \bar{v}$$

and *strongly monotone* (with modulus $\alpha > 0$) if

(8) $$\langle A(v) - A(\bar{v}), v - \bar{v}\rangle \geq \alpha\, \|v - \bar{v}\|^2 \quad \forall v, \forall \bar{v}$$

If A is linear, the expression on the left reduces to

$$\langle A(v - \bar{v}), v - \bar{v}\rangle = \langle A_s(v - \bar{v}), v - \bar{v}\rangle$$

where $A_s = \frac{1}{2}(A + A^*)$ is the symmetric part of A. Monotonicity then means that A is positive semidefinite, while strong monotonicity corresponds to positive definiteness (these terms being employed whether or not A is itself symmetric).

If F is a functional such that (4) holds (with A not necessarily linear), the monotonicity condition (7) can be written as

$$\langle \nabla F(\bar{v} + z) - \nabla F(\bar{v}), z\rangle \geq 0 \quad \forall z, \forall \bar{v}$$

It is not difficult to show that this is true if and only if F is *convex*. Thus the optimization problems (5), where a diffentiable convex function is minimized over a convex set, correspond to a special class of variational inequalities where A is monotone. However, not every variational inequality with A monotone can be interpreted in this way, since, for example, if A is linear, but not symmetric, a gradient representation (4) is impossible. It will be seen below that variational inequalities, where the mapping is monotone and not the gradient of any functional, can nevertheless give the optimality conditions for some minimization problems of convex type, when Lagrange multipliers are brought into the picture. No one knows whether by some extension of these ideas, all 'monotone' variational inequalities can be interpreted as arising from 'optimization'. What is clear, though, is that monotone mappings not of gradient type do arise in a number of ways, in particular in certain physical problems involving friction.

In applications to partial differential equations, V is usually a space of real-valued functions, such as a Sobolev space defined on a region Ω in $\mathbb{R}^N$. The mapping A represents a differential operator, and K incorporates various boundary conditions. Functional analytic considerations are then crucial, for instance in making sure that V, K, and A are well chosen for the problem one has in mind. The theory of distributions is used in studying the existence of weak and strong solutions, the regularity of solutions and other aspects. These aspects are important in the discussion of schemes of discretization, but for someone in mathematical programming who is interested mainly in solving problems that are already discretized, they are not essential.

Incidentally, if variational inequalities are viewed in the form (3), there is a very natural generalization to the case where K is non-convex: interpret $N_K(u)$ as the normal cone in the sense of Clarke [1] (see also Rockafellar [2, 3]). Clarke's work with the calculus of variations (for example, [4, 5]) and mathematical programming [6] indicates that first-order necessary conditions for many non-convex problems of optimization can be written as variational inequalities in this sense. As far as the theory of variational inequalities is concerned, such generalizations have not yet been explored. Most of the existing results are based on the consequences of convexity and the monotonicity idea.

3. DISCRETIZATION

For the purpose of computation, a variational inequality (1) is often 'approximated' by one involving a *finite* constraint system in a *finite*-dimensional space. Thus K is replaced by a set K_0 that lies in a finite-dimensional subspace V_0 of V and has the form

$$K_0 = \{v \in V_0 : F_1(v) \leq 0, \ldots, F_m(v) \leq 0\} \tag{9}$$

where each F_i is a finite *convex* function on V_0. (Equality constraints for affine functions are also permissible but will be kept out of the discussion temporarily for notational simplicity.) The variational inequality is reduced accordingly to

$$\begin{aligned} &u \in K_0 \\ &\langle A(u) - a, v - u\rangle \geq 0 \qquad \text{for all} \quad v \in K_0 \end{aligned} \tag{10}$$

This condition really depends only on the orthogonal projections of $A(u)$ and a on V_0:

$$A_0(u) = \text{proj } A(u) \qquad a_0 = \text{proj } a \tag{11}$$

We therefore get another variational inequality which is entirely in V_0, namely the one for K_0, the mapping $A_0 : V_0 \to V_0$, and the element $a_0 \in V_0$. In terms of normal cones, we have discretized condition (3) to obtain

$$u \in K_0 \quad \text{and} \quad a_0 - A_0(u) \in N_{K_0}(u) \tag{12}$$

In the gradient case (4), the projection step (11) in this procedure amounts to restricting the functional F to V_0. Denoting the restriction by F_0, one has

$$A_0(v) - a_0 = \nabla F_0(v) \qquad \forall\, v \in V_0 \tag{13}$$

where the gradient is taken in the sense of the space V_0. The corresponding optimization problem (5) is thereby discretized to

$$\begin{aligned} &\text{minimize} \qquad F_0(v) \qquad \text{over all} \quad v \in V_0 \\ &\text{satisfying} \qquad F_1(v) \leq 0, \ldots, F_m(v) \leq 0 \end{aligned} \tag{14}$$

It is important to recall that when A is monotone, the functional F is convex, and hence F_0 too is convex. The discretized problem then falls into the classical pattern of *convex programming*.

Optimization problems and their extensions typically involve not only primal variables, but also dual variables that lend themselves to interpretation as Lagrange multipliers associated with various constraints. This is why a process of discretization (in the sense used in this context — mathematical programmers are accustomed to speaking of discretization only in the extreme case where all 'continuous' structure is abandoned) must aim at a finite system of constraint functions, as well as a primal space that is finite-dimensional.

Dual variables in the case of (12) arise in representing the normal cone $N_{K_0}(u)$. Let us assume that $F_1, \ldots, F_m$ are not only convex but differentiable and that

$$\exists\, v \in V_0 \quad \text{with} \quad F_1(v) < 0, \ldots, F_m(v) < 0 \tag{15}$$

Then as is well known, $N_{K_0}(u)$ for a point $u \in K_0$ is the convex cone (containing 0) generated by the gradients $\nabla F_i(u)$ of the *active* constraints at u (i. e. those having $F_i(u) = 0$ rather than $F_i(u) < 0$). In other words,

$$N_{K_0}(u) = \left\{ \sum_{i=1}^{m} y_i \nabla F_i(u) \; : \; y_i \geq 0,\; y_i F_i(u) = 0 \right\} \tag{16}$$

The variational inequality (12) then can be rewritten as a condition in both $u \in V_0$ and $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$:

$$\begin{aligned} &F_i(u) \leq 0 \quad y_i \geq 0 \quad \text{and} \quad y_i F_i(u) = 0 \quad \text{for } i = 1, \ldots, m \\ &A_0(u) - a_0 + \sum_{i=1}^{m} y_i \nabla F_i(u) = 0 \end{aligned} \tag{17}$$

In the gradient case (13) these are known as the Kuhn-Tucker conditions for problem (14), and they play an enormous role in computational techniques. The conclusion we wish to emphasize is that *monotone* variational inequalities, after discretization, correspond to an extended form of the classical optimality conditions for convex programming, in the sense of involving a monotone term $A_0(u) - a_0$ that does not have to be the gradient of anything. One can attempt to generalize methods for solving (17) in the case of

$$A_0(u) - a_0 = \nabla F_0(u)$$

to this broader framework. Since monotone mappings have many powerful properties that seem to place them within the realm of convex analysis, it is natural therefore to regard the solution of monotone variational inequalities as a kind of 'extended convex programming'.

When equality constraints are of interest (the functions in question being affine), one can, of course, express an equation $F_i(u) = 0$ by a pair of inequalities $F_i(u) \leq 0$ and $-F_i(u) \leq 0$, at least for the sake of theoretical uniformity. A familiar refinement of assumption (15) can then be invoked in passing to the representation (17): strict inequality can be relaxed to weak inequality for each constraint function that is affine.

4. COMPLEMENTARITY

Many situations lead to the following model. Given a mapping $M : \mathbb{R}^N \to \mathbb{R}^n$ and vector $q \in \mathbb{R}^N$, determine $z \in \mathbb{R}^N$ such that

$$z \geq 0 \qquad M(z) + q \geq 0 \qquad z \cdot [M(z) + q] = 0 \tag{18}$$

This is the *complementarity problem* for M and q. The notation $z \geq 0$ means that z belongs to the non-negative orthant $\mathbb{R}^N_+$. Thus in terms of $z = (z_1, \ldots, z_N)$ and $M(z) + q = w = (w_1, \ldots, w_N)$, the condition says that

$$\begin{aligned} &\text{for each} \quad j = 1, \ldots, N, \quad \text{one has} \quad z_j \geq 0, \quad w_j \geq 0 \\ &\text{and either} \quad z_j = 0 \quad \text{or} \quad w_j = 0 \end{aligned} \tag{19}$$

Another version of the condition can be stated in terms of the normal cones associated with $\mathbb{R}^N_+$. Observing that

$$N_{\mathbb{R}^N_+}(z) = \{ -w : w \geq 0,\; z \cdot w = 0 \} \quad \text{for } z \geq 0$$

we see that (18) is the same as

$$z \in \mathbb{R}^N_+ \quad \text{and} \quad -q - M(z) \in N_{\mathbb{R}^N_+}(z) \tag{20}$$

The complementarity problem is thus simply the variational inequality problem for $K = \mathbb{R}^N_+$. Its simple form has lent itself to the development of a number of algorithms, mainly for linear mappings M (see other chapters in this volume). Again in this context the monotonicity of the mapping plays an important role.

Not so obvious is the fact that *every* monotone variational inequality in the discretized form above can be reformulated as a monotone complementarity problem. Thus the special case is not so special after all.

To demonstrate the truth of this assertion, let us first look at a somewhat simpler, but very common case where $K_0 \subset \mathbb{R}^n_+$ (the space V_0 being identified now notationally with $\mathbb{R}^n$). Specifically, let us suppose that in the constraints in (9) one has $m = n + p$ and for $v = (v_1, \dots, v_n)$

$$F_{p+j}(v) = -v_j \quad \text{for} \quad j = 1, \dots, n$$

Then in terms of

$$\hat{G}(v) = (F_1(v), \dots, F_p(v)) \tag{21}$$

and $\hat{y} = (y_1, \dots, y_p)$, $\tilde{y} = (y_{p+1}, \dots, y_m)$, one can write (17) as

$$\hat{y} \geq 0 \quad \hat{G}(u) \leq 0 \quad \hat{y} \cdot \hat{G}(u) = 0$$

$$u \geq 0 \quad \tilde{y} \geq 0 \quad \tilde{y} \cdot u = 0 \tag{22}$$

$$A_0(u) - a_0 + \sum_{i=1}^{p} \hat{y}_i \nabla F_i(u) = -\tilde{y}$$

Setting $q = (-a_0, 0)$ and

$$M(u, \hat{y}) = \left(A_0(u) + \sum_{i=1}^{p} \hat{y}_i \nabla F_i(u), \ -\hat{G}(u)\right) \tag{23}$$

we get (22) into the complementarity form

$$(u, \hat{y}) \geq 0 \quad M(u, \hat{y}) + q \geq 0 \quad (u, \hat{y}) \cdot [M(u, \hat{y}) + q] = 0 \tag{24}$$

For the general case of (17) the reformulation as a complementarity problem is similar but requires some elementary tricks well known in optimization theory. The equation must be written as a pair of inequalities, and the vector u must be expressed as the difference of vectors u^+ and u^- that are constrained to be non-negative. Thus one sets $q = (-a_0, a_0, 0)$ and

$$M(u^+, u^-, y) = (H(u^+ - u^-, y), -H(u^+ - u^-, y), -G(u^+ - u^-)) \tag{25}$$

where

$$G(u) = (F_1(u), \dots, F_m(u)) \tag{26}$$

$$H(u, y) = A_0(u) + \sum_{i=1}^{m} y_i \nabla F_i(u) \tag{27}$$

and this puts (17) in the form

$$z \geq 0 \quad M(z) + q \geq 0 \quad z \cdot [M(z) + q] = 0 \quad \text{for} \quad z = (u^+, u^-, y) \tag{28}$$

Proposition 1

If the mapping A_0 in the expanded variational inequality (17) is monotone and continuous (and each F_i is convex), then the mapping M in the corresponding complementarity problem (28) (or (24)) is monotone and continuous relative to the non-negative orthant. However, M cannot be strongly monotone (even if A_0 is strongly monotone), when the constraint system (9) is non-vacuous.

To prove the first assertion, with M given by (25), we observe that the form

$$[M(z) - M(\bar{z})] \cdot [z - \bar{z}] \quad \text{for} \quad z = (u^+, u^-, y), \bar{z} = (\bar{u}^+, \bar{u}^-, \bar{y})$$

reduces to

$$[H(u, y) - H(\bar{u}, \bar{y})] \cdot [u - \bar{u}] - [G(u) - G(\bar{u})] [y - \bar{y}] = \tag{29}$$

$$= [A_0(u) - A_0(\bar{u})] \cdot [u - \bar{u}] +$$

$$+ \sum_{i=1}^{m} [(y_i \nabla F_i(u) - \bar{y}_i \nabla F_i(\bar{u}))(u - \bar{u}) + (y_i - \bar{y}_i)(F_i(u) - F_i(\bar{u}))]$$

where $u = u^+ - u^-$ and $\bar{u} = \bar{u}^+ - \bar{u}^-$. If A_0 is monotone, we have

$$[A_0(u) - A_0(\bar{u})] \cdot [u - \bar{u}] \geq 0$$

On the other hand, the convexity of each F_i yields

$$(y_i \nabla F_i(u) - \bar{y}_i \nabla F_i(\bar{u}))(u - \bar{u}) + (y_i - \bar{y}_i)(F_i(u) - F_i(\bar{u})) \geq 0$$

through combination of the two inequalities

$$\bar{y}_i F_i(u) \geq \bar{y}_i F_i(\bar{u}) + \bar{y}_i \nabla F_i(\bar{u}) \cdot (u - \bar{u})$$

$$y_i F_i(\bar{u}) \geq y_i F_i(u) + y_i \nabla F_i(u) \cdot (\bar{u} - u)$$

(Non-negativity of y_i and $\bar{y}_i$ is needed here.) The expression in (29) is then non-negative (when $y \geq 0$, $\bar{y} \geq 0$), and the monotonicity of M with respect to the non-negative orthant, when A_0 is monotone, is established. The continuity of M follows from the continuity of A_0 and the fact that a differentiable convex function F_i is always continuously differentiable. Strong monotonicity is impossible because (19) vanishes when $u = \bar{u}$ but $y \neq \bar{y}$.

Incidentally, it can also be proved that a mapping M of form (25) (or (23)) *cannot* be the gradient of any function, even if A_0 is of gradient type. Thus, through the process of reformulation using Lagrange multipliers, we see that monotone mappings that are not strongly monotone nor of gradient type have an essential role to play in the theory. Even when we start from a problem of optimization, we may be led to a complementarity problem, and hence a particular kind of variational inequality, involving such a mapping.

5. LINEARITY VERSUS NON-LINEARITY

At the present stage of development, most of the techniques for solving the complementarity problem, at least the ones that might take advantage of monotonicity or other such structure rather than just reducing everything to the location of a general fixed point, concern only the case of a linear mapping. The following elementary fact is therefore central in any discussion of solving variational inequalities by way of such techniques.

Proposition 2
If A_0 in the expanded 'discretized' variational inequality (17) is linear and each F_i is affine, then M is linear in the corresponding complementarity problem (28) (or 24)). The converse is also true.

Of course, in discretizing a general variational inequality (1), we obtain A_0 as a restriction and projection of A, so A_0 is linear if A is. On the other hand, the closed convex set K is the intersection of all the closed half-spaces containing it, so in principle we may regard K as defined by a possibly infinite system of linear inequalities. Discretization could be achieved by passing to a finite-dimensional subsystem.

Thus in a certain sense, the infinite-dimensional variational inequality problems that can be approximated by *linear* complementarity problems can be identified simply as the ones in which the mapping A is linear. In the optimization context, they are the ones that correspond to infinite-dimensional quadratic/linear programming.

The linear complementarity approach to computation could be pushed a bit further in considering a scheme in which a non-linear A is linearized iteratively to get subproblems that fall within the guidelines just laid down. However, no such scheme has been shown to be attractive for convex programming, until recently, and then in forms that may not readily extend to general variational inequalities.

Many important variational inequalities do involve linear monotone mappings and constraints which in passing to a finite-dimensional subspace V_0 (based on a triangularization of the domain Ω over which the function space V is defined) do reduce neatly to finite systems of linear inequalities. Lest one take this sort of behaviour for granted, though, it is well to look at a classical case where the constraints cannot be handled in such a simple fashion. This will also illustrate how there may be an advantage in using convex (non-linear) inequalities in the discretization process, although the original variational inequality appears to be just 'linear'.

A good example is the potential problem studied by Stampacchia in the early days of the theory of variational inequalities. In this problem there is a bounded open domain Ω in $\mathbb{R}^N$ and a non-empty closed set $E \subset \Omega$. The Sobolev space $H_0^1(\Omega)$ serves as V, while K consists of the closure of the set of 'test functions' (C^∞ functions) v in $H_0^1(\Omega)$ that satisfy

$$(30)\qquad \begin{aligned} v(x) &\geq 0 \quad \text{for all } x \in \Omega\setminus E \\ v(x) &\geq 1 \quad \text{for all } x \in E \end{aligned}$$

Note that (30) has the appearance of an infinite system of linear inequalities indexed by E and $\Omega\setminus E$:

$$(31)\qquad \begin{aligned} f_1^x(v) &\leq 0 \quad \text{for all } x \in E \\ f_2^x(v) &\leq 0 \quad \text{for all } x \in \Omega\setminus E \end{aligned}$$

where

$$(32)\qquad f_1^x(v) = 1 - v(x) \qquad f_2^x(v) = -v(x)$$

The trouble is that the functionals f_1^x and f_2^x are not continuous or even defined everywhere on $V = H_0^1(\Omega)$. Indeed, the elements v of this space are really just equivalence classes of functions that differ only on a set of measure zero. Evaluation at a point x therefore does not make sense unless an equivalence class can be identified with a distinguished member that is continuous, say (or in the case of the 'test functions', infinitely differentiable).

An idea that might at first look tempting in this situation would be to represent the constraint by an abstract inequality in a partially ordered space. Thus if $H_0^1(\Omega)$ is given the natural ordering induced by the cone of functions that are non-negative almost everywhere, one could hope to write (30) as

$$(33)\qquad v \geq \chi_E$$

where χ_E is the characteristic function of E. But χ_E does not belong to the space $H_0^1(\Omega)$, so this formulation falls short. In any event, like the preceding formulation, it fails to take into account the fact that K is merely the *closure* of the set of *test functions* satisfying the constraints.

The lack of continuity of the functionals (32) also bodes ill for schemes of discretization where the infinite system (31) is replaced by a finite subsystem. Although the finite-dimensional subspace V_0 that is chosen may be such that these functionals are well defined, there is little to guarantee that the discretized problem really 'approximates' the given one.

One way around the impasse is to forgo the linearity of the constraint representation. For 'test functions' ν in $H_0^1(\Omega)$, define

$$(34)\qquad \begin{aligned} F_1(\nu) &= \max\ \{1 - \nu(x) : x \in E\} \\ F_2(\nu) &= \max\ \{-\nu(x) : x \in \mathrm{c}\ell\,[\Omega\backslash E]\} \end{aligned}$$

(see equations (32)). Then define $G : H_0^1(\Omega) \to \mathbb{R} \cup \{+\infty\}$ by

$$G(\nu) = \liminf_{\nu' \to \nu}\ [\max\{F_1(\nu'), F_2(\nu')\}]$$

where the limit is taken over all 'test function' sequences converging to ν in the norm topology of $H_0^1(\Omega)$. The set we are interested in is, by definition,

$$K = \{\nu \in H_0^1(\Omega) : G(\nu) \leq 0\}$$

Note that G is a lower semi-continuous *convex* functional. On certain subspaces V_0, in particular those whose elements are all test functions, it will be true that

$$(35)\qquad G(\nu) = \max\ \{F_1(\nu), F_2(\nu)\}$$

with F_1 and F_2 still well defined by (34). Then

$$(36)\qquad K \cap V_0 = \{\nu \in V_0 : F_1(\nu) \leq 0, F_2(\nu) \leq 0\}$$

Taking the latter to be K_0, we have a discretization where the constraint system is expressed by a pair of finite *convex* functions. As a means of approximate the original problem, this approach is much more stable.

Obviously the convex functions F_1 and F_2 will not usually be *differentiable* on V_0, but this need not be a serious obstacle. In the theory of convex programming, much attention has recently been given to non-differentiable functions. In part, this is due precisely to their role in reformulating problems that might otherwise involve more constraints than may be handled conveniently at one time. A kind of 'aggregation of dual variables' is involved. If the infinite system (31) were approximated by a large finite subsystem, there would be a correspondingly large number of Lagrange multipliers that would have to be kept track of. In the representation (36) there are only two multipliers. In evaluating F_1 or F_2 or one of their subgradients at a point ν, as may be required by an algorithm, we need only invoke a subroutine for solving the maximization problems in x that are embedded in the formulas (34). Thus, in effect, we can generate, as we go along, the affine functions f_1^x or f_2^x that turn out to be important, rather than having to treat all of them individually.

6. GENERALIZATION TO MULTIFUNCTIONS

If non-differentiable functions are to be treated, the gradient mappings that appear in (17) (and (13)) must be replaced by something more general. For a convex functional $F : V \to \mathbb{R} \cup \{+\infty\}$, it will be recalled that the set of *subgradients* of F at ν is

$$(37)\qquad \partial F(\nu) = \{w \in V : F(\nu') \geq F(\nu) + \langle w, \nu' - \nu\rangle,\ \forall \nu' \in V\}$$

The multifunction ∂F associates, in other words, with each $\nu \in V$ a certain closed convex (possibly empty) set in V. (For the appropriate definition of ∂F when F is not convex, see [1,3].)

The connection between convexity and monotonicity remains strong under this generalization. A multifunction (set-valued mapping) $T : V \rightrightarrows V$ is said to be *monotone* if

$$(38)\qquad \langle w - \bar{w}, \nu - \bar{\nu}\rangle \geq 0 \quad \text{for all } \nu, \bar{\nu} \text{ in } V \text{ and all } w \in T(\nu), \bar{w} \in T(\bar{\nu})$$

It is *maximal* monotone if, in addition, there does not exist any monotone multifunction $T' : V \rightrightarrows V$ with $T'(\nu) \supset T(\nu)$ for all ν and $T'(\nu) \neq T(\nu)$ for at least one ν.

Moreau [7] proved in the Hilbert space case that ∂F is maximal monotone if F is a convex function on V that is lower semicontinuous and not identically $+\infty$. (For the generalization to Banach spaces and the precise characterization of the class of maximal monotone multifunctions that are obtained in this fashion, see Rockafellar [8]).

The general variational inequality (1) in the case of a *multifunction* A takes the form

$$(39)\qquad \begin{aligned} &u \in K \quad w \in A(u) \\ &\langle w - a, \nu - u\rangle \leq 0 \quad \text{for all } \nu \in K \end{aligned}$$

or in other words

$$(40)\qquad u \in K \quad \text{and} \quad a \in A(u) + N_K(u)$$

The complementarity problem for a multifunction M has (18) replaced by

$$(41)\qquad z \geq 0 \quad w \geq 0 \quad z \cdot w = 0 \quad w \in M(z) + q$$

so that the condition is

$$(42)\qquad -q \in M(z) + N_{\mathbb{R}_+^N}(z)$$

In the reformulations discussed earlier, (17) becomes

$$F_i(u) \leq 0 \quad y_i \geq 0 \quad \text{and } y_i F_i(u) = 0 \quad \text{for } i = 1, \ldots, m \tag{43}$$
$$a_0 \in A_0(u) + \sum_{i=1}^{m} y_i \, \partial F_i(u)$$

while in the definition (25) of M one takes

$$H(u,y) = A_0(u) + \sum_{i=1}^{m} y_i \, \partial F_i(u) \tag{44}$$

instead of (27).

Proposition 3
Suppose A_0 is maximal in (43) (and each F_i is convex), and that (15) holds for an element v such that $A_0(v) \neq \emptyset$. Then the multifunction M in the corresponding complementarity problem (defined by (25)and (44)) is maximal monotone.

Proof: This follows from results in [9] when the sum of two maximal monotone mappings is again maximal monotone – for the argument, see [10, proposition 5].

The assumption about (15) holding with $A_0(v) \neq \emptyset$ can be replaced by the following. There exists $v \in K_0 \cap \text{int } D(A_0)$ such that $F_i(v) < 0$ for all non-affine functions F_i, where $D(A_0)$ is the set of points where A_0 is non-empty-valued. This modification is needed in treating the case of equality constraints.

The connection between Proposition 3 and Proposition 2 is this: a monotone mapping $A : V \to V$ that is weakly continuous relative to each line segment in V is maximal monotone when viewed as a multifunction. Furthermore, if $A : V \rightrightarrows V$ is a maximal monotone multifunction and $V_0 \cap \text{int } D(A) \neq \emptyset$, then the projection $A_0 : V_0 \to V_0$ is maximal monotone. This is a consequence of [9, theorem 1].)

Whether the extended problems in terms of multifunctions can be solved effectively, remains a largely unexplored question. But in view of the trend in convex programming, there is certainly hope in this direction.

7. PENALTY-DUALITY METHODS

Variational inequality problems that occur in applications are very often

generalizations of classical boundary-value problems for partial differential equations. Therefore, it is very natural to look for ways of solving them in terms of reducing them to a succession of classical problems for which numerical tchniques are already highly developed. Sometimes this can be accomplished by the introduction of penalties to force the satisfaction of constraints. However, there now appears to be a possibility of using more sophisticated methods based on ideas of recent years in non-linear programming.

To see these ideas in their original and simplest form, let us consider first the case of a non-linear programming problem with *equality* constraints:

$$\text{minimize } F_0(v) \quad \text{subject to} \quad F_1(v) = 0, \ldots, F_m(v) = 0 \tag{45}$$

The *augmented Lagrangian* function for this problem is

$$L(v,y,r) = F_0(v) + \sum_{i=1}^{m} [\, y_i F_i(v) + \frac{r}{2} F_i(v)^2 \,] \tag{46}$$

where r is a non-negative variable that serves as a penalty parameter.

A fundamental class of penalty-duality algorithms can be described as follows. At iteration k we have a vector $y^k = (y_1^k, \ldots, y_m^k)$ in $\mathbb{R}^m$ and a value $r_k > 0$. We determine u^{k+1} as an 'approximate' solution to the (unconstrained!) problem

$$\text{minimize } L(v, y^k, r_k) \quad \text{over all } v \tag{47}$$

Then, by means of some rule that may involve information gleaned during this process, we generate y^{k+1} and r_{k+1} and repeat the step. The aim is to get a sequence $\{u^k\}$ that in some sense yields in the limit an optimal solution to the constrained problem (15).

Taking $y^k \equiv 0$ and $r_k \nearrow \infty$, we obtain the classical quadratic penalty method. Pure duality methods have non-trivial sequences $\{y^k\}$ but $r_k \equiv 0$. In 1968, Hestenes [11] and Powell [12] independently proposed the very easy rule

$$y_i^{k+1} = y_i^k + r_k F_i(u^{k+1}) \tag{48}$$

The sequence $\{r_k\}$ was non-decreasing but did *not* have to tend to infinity in order to ensure that $\{u^k\}$ approached a solution to the constrained problem. (Powell actually allowed a different value of r_k for each F_i.) This was an important breakthrough, since faster convergence was demonstrated than with pure penalty methods, and some of the intrinsic numerical instability in the latter was sidestepped.

In the case of an inequality constraint $F_i(v) \leq 0$ in (45), the expression

$$y_i F_i(v) + \frac{r}{2} F_i(v)^2$$

in the augmented Lagrangian should be replaced by

$$(49)\qquad \begin{cases} y_i F_i(v) + \frac{r}{2} F_i(v)^2 & \text{if } F_i(v) \geq -\frac{y_i}{r} \\ -\frac{y_i^2}{2r} & \text{if } F_i(v) \leq -\frac{y_i}{r} \end{cases}$$

The corresponding form of rule (48) is

$$(50)\qquad y_i^{k+1} = \max\{0, y_i^k + r_k F_i(u^{k+1})\}$$

For the theory of this case and a discussion of the criteria that can be used in solving (47) 'approximately', see [13, 14, 15, 16]. The theory makes heavy use of properties of maximal monotone multifunctions.

The Hestenes-Powell method has already been applied to certain variational inequalities that correspond to boundary-value problems of (convex) optimization type with equality constraints only; see Glowinski/Marrocco [17] and Mercier [18]. This work supposes u^{k+1} to be an *exact* solution to (47) at each iteration. Of course, solving (47) exactly is equivalent to solving a certain equation in v, namely

$$(51)\qquad \begin{aligned} 0 &= \nabla_v L(v, y^k, r_k) \\ &= \nabla F_0(v) + \sum_{i=1}^{m} y_i(v, y_i^k, r_k)\, \nabla F_i(v) \end{aligned}$$

where

$$(52)\qquad Y_i(v, y_i, r) = \begin{cases} y_i + rF_i(v) & \text{for equalities} \\ \max\{0, y_i + rF_i(v)\} & \text{for inequalities} \end{cases}$$

This could easily be generalized to variational inequalities not of optimization type by replacing the term $\nabla F_0(v)$ by $A_0(v) - a_0$.

Unfortunately, exact solution at each iteration cannot be obtained except in very special cases. Yet the kind of criterion for 'approximate' minimization in (47) that has been used in proofs of global convergence is

$$L(u^{k+1}, y^k, r_k) \leq \inf_v L(v, y^k, r_k) + \delta_k$$

and this has no analogue for variational inequalities not of optimization type.

8. PROXIMAL METHOD OF MULTIPLIERS

We have proposed [10] a modified version of the Hestenes-Powell algorithm that does carry over to generalized variational inequalities as represented in the form (17). There is no space here to explain the natural motivation; suffice it to say that the theory of maximal monotone multifunctions is deeply involved.

To apply the method one needs to specify a parameter value $s > 0$, and positive sequences $\{r_k\}$ and $\{\epsilon_k\}$ satisfying

$$(53)\qquad r_k \nearrow r_\infty \leq \infty \quad \text{and} \quad \sum_{i=0}^{\infty} \epsilon_k < \infty$$

One also makes an 'initial guess' (u^0, y^0) of the solution to (17). (In fact u^0 and y^0 can be chosen arbitrarily, for instance both zero). In iteration k, one forms the mapping

$$(54)\qquad T_k(v) = \frac{1}{s}(v - u^k) + A_0(v) - a_0 + \sum_{i=1}^{m} Y_i(v, y_i^k, r_k) \nabla F_i(v)$$

and looks for an approximate solution u^{k+1} to the equation $T_k(v) = 0$. Specifically, u^{k+1} is taken to be any vector satisfying

$$(55)\qquad \| T_k(u^{k+1}) \| \leq \frac{\epsilon_k}{r_k} \max\{1, \|(u^{k+1}, y^{k+1}) - (u^k, y^k)\|_s\}$$

where

$$(56)\qquad \|(u, y)\|_s = [s^{-2}\|u\|^2 + \|y\|^2]^{1/2}$$

Then y^{k+1} is defined by

$$(57)\qquad y_i^{k+1} = Y_i(u^{k+1}, y_i^k, r_k) \quad \text{for} \quad i = 1, \ldots, m$$

(Here Y_i is given by (52); this formulation covers any mixture of equality and inequality constraints. In the case of an equality constraint, the corresponding conditions in the first line of (17) are replaced simply by $F_i(u) = 0$.)

In problems of optimization type, the task of solving $T_k(v) = 0$ approximately reduces to that of minimizing $L(v, y^k, r_k) + (1/2s)\| v - u^k \|^2$ approximately.

Theorem

(see [10]) If A_0 is a continuous monotone mapping and each F_i is convex and differentiable (or in the case of an equality constraint, affine), then the mapping T_k in each iteration is strongly monotone with modulus $(1/s)$.

Assuming also that the expanded variational inequality (17) has at least one solution (u, y), it will be true that

(58) $$(u^k, y^k) \to (u^\infty, y^\infty)$$

where (u^∞, y^∞) is some particular solution (even though there may be more than one solution!)

Moreover there is a constant $q_s \in [0, \infty]$ such that

(59) $$\limsup_{k \to \infty} \frac{\|(u^{k+1}, y^{k+1}) - (u^\infty, y^\infty)\|_s}{\|(u^k, y^k) - (u^\infty, y^\infty)\|_s} < \theta\left(\frac{q_s}{r_\infty}\right)$$

where

$$\theta\left(\frac{q_s}{r_\infty}\right) = \begin{cases} \left(\frac{q_s}{r_\infty}\right) / \left[1 + \left(\frac{q_s}{r_\infty}\right)^2\right]^{1/2} < 1 & \text{if } q_s < \infty, r_\infty < \infty \\ 0 & \text{if } q_s < \infty, r_\infty = \infty \\ 1 & \text{if } q_s = \infty \end{cases}$$

If q_s is finite, and more will be said about this in a moment, the last part of the theorem guarantees linear convergence at a rate that can be controlled by how high the penalty parameter values r_k are allowed to go. The case where $r_k \nearrow \infty$ yields superlinear convergence. This is much superior to what would happen with a pure penalty method. In practice, the improved rate of convergence means that a satisfactory termination can be reached before r_k gets so high as to cause numerical instabilities. As $s \nearrow \infty$, q_s decreases to the constant that would appear in the corresponding convergence rate formula for the Hestenes-Powell algorithm, in the case of convex programming in $\mathbb{R}^N$.

If $q_s = \infty$, one still has global convergence (58) from any starting point (u^0, y^0), but it is no longer possible to establish a linear rate. However, it can be shown that this case is generically rare.

Proposition 4

Suppose $F_i(v) = G_i(v) - b_i$ for $i = 1,\ldots,m$. Then for almost every choice of $a_0 \in V_0$ (finite-dimensional) and $b = (b_1,\ldots,b_m) \in \mathbb{R}^m$ such that the variational inequality (17) has a solution,the corresponding constant q_s is finite.

This follows from a theorem of Mignot on the almost-everywhere differentiability of a maximal monotone multifunction on the interior of its effective domain in $\mathbb{R}^n$. For the argument, see [10].

The choice of s in the proximal point algorithm is somewhat problematical. It s is too low, q_s may be high, and the convergence ratio $\theta(q_s/r_\infty)$ will suffer. While this could be compensated for by choosing r_k high, there might be a price paid in numerical stability, After all, the whole point of the augmented Lagrangian approach is to succeed without r_k getting too high. It would be an improvement if instead of a fixed value for s the argorithm could be expressed in terms of a non-decreasing sequence $s_k \nearrow s_\infty < \infty$. This would allow some control over the phenomenon, just as in the case of the penalty parameter. Better still, in order to take account of second-order information, the factor $1/s$ in the definition (54) of T_k might be generalized to S_k^{-1}, where S_k is a positive definite matrix. The use of separate values of r_k in (57) for each constraint function, as proposed by Powell, could also be restored. Undoubtedly such developments are possible and desirable, but a significant enlargement of the theoretical apparatus (in terms of maximal monotone multifunctions) would be required.

Another direction of generalization, as mentioned earlier, would be to allow A_0 to be a multifunction and F_i to be non-differentiable. For this, the existing theory carries over with $w^{k+1} = T_k(u^{k+1})$ replaced by an element $w^{k+1} \in T_k(u^{k+1})$ in (55) at each interation. But effective numerical techniques would need to be developed for determining such u^{k+1} and w^{k+1}.

REFERENCES

[1] F.H. Clarke. 'Generalized gradients and applications'. *Trans. Am. Math. Soc.*, **205** (1975), 247-62.

[2] R.T. Rockafellar. 'Clarke's tangent cones and the boundaries of closed sets in $\mathbb{R}^n$'. *Non-linear Anal.* **3** (1979), 145-54.

[3] R.T, Rockafellar. 'Generalized derivatives and subgradients of nonconvex functions'. *Can. J. Math.* To appear.

[4] F.H. Clarke. 'The generalized problem of Bolza'. *SIAM J. Control & Optimiz.* **14** (1976), 682-99.

[5] F.H. Clarke. 'Extremal arcs and extended Hamiltonian systems'. *Trans. Am. Math. Soc.*, **231** (1977), 349-67.

[6] F.H. Clarke. 'A new approach to Lagrange multipliers'. *Math. Oper. Res.*, **1** (1976), 165-74.

[7] J.J. Moreau. 'Proximité et dualité dans un espace hilbertien'. *Bull. Soc. Math. Fr.*, **93** (1965), 273-99.

[8] R.T. Rockafellar. 'On the maximal monotonicity of subdifferential mappings'. *Pacific J. Math.*, **33** (1970), 209-16.

[9] R.T. Rockafellar. 'Maximality of sums of nonlinear monotone operators'. *Trans. Am. Math. Soc.*, **149** (1970), 75-88.

[10] R.T. Rockafellar. 'Monotone operators and augmented Lagrangian methods in non-linear programming'. In *Nonlinerar Programming* vol. 3, eds. O.L. Mangasarian et al., Academic Press, New York (1978), 1-25.

[11] M.R. Hestenes. 'Multiplier and gradient methods'. *J. Optimiz. Theor. & Appl.*, **4** (1969), 303-20.

[12] M.J.D. Powell. 'A method for nonlinear constraints in minimization problems'. In *Optimization* ed. R. Fletcher, Academic Press, New York (1969), 283-98.

[13] R.T. Rockafellar. 'The multiplier method of Hestenes and Powell applied to convex programming'. *J. Optimiz. Theor. & Appl.*, **12** (1973), 555-62.

[14] R.T. Rockafellar. 'Solving a nonlinear programming problem by way of a dual problem'. *Symposia Mathematica* (Istituto Nazionale Alta Matematica) vol. XIX, Academic Press, New York (1976), 135-60.

[15] R.T. Rockafellar. 'Augumented Lagrangians and applications of the proximal point algorithm in convex programming'. *Math. Oper. Res.*, **1** (1976), 97-116.

[16] D.P. Bertsekas. 'Multiplier methods: a survey'. *Proc. IFAC 6th Triennial World Congress*, Part 1B.

[17] R.Glowinski and A. Marocco. 'Sur l'approximation, par éléments finis d'ordre un, et la resolution, par pénalisation-dualité, d'une classe de problems de Dirichlet non linéaires'. *C.R. Acad. Sci., Paris,* **278** (1974), 1649.

[18] B. Mercier. 'Approximation par éléments finis et resolution, par un algorithme de penalisation-dualité, d'un problème d'élasto-plaslicité'. *C.R. Acad. Sci., Paris,* **280** (1975), 287-90.