Can. J. Math., Vol. XXIX, No. 3, 1977, pp. 528–540

INEQUALITY CONSTRAINTS IN THE CALCULUS OF VARIATIONS

FRANK H. CLARKE

1. Introduction. The classical multiplier rule. The purpose of this section is to review the multiplier rule in order to place the results of this report in perspective. Let us begin by considering the following problem of Mayer in the calculus of variations: we seek to minimize

$$(1.1) \qquad \varphi(x(1))$$

over a class of functions $x : [0, 1] \rightarrow R^n$, subject to the boundary conditions

$$(1.2) \qquad x(0) \in C_0, \quad x(1) \in C_1$$

as well as the equality constraints

$$(1.3) \qquad f_i(x(t), \dot{x}(t)) = 0 \quad (i = 1, 2, \ldots, r; t \in [0, 1]).$$

In the above, the functions φ and f_i and the sets C_0 and C_1 are given; we leave unspecified for now the class of functions x admitted to competition, as well as other details. Let us mention the well-known fact that superficially different problems involving the minimization of integrals can be reshaped to fit the above mould (see [**11**, Chapter 6]).

Suppose now that the function z solves this problem. The "multiplier rule" is a theorem stating that, under suitable hypotheses, there exist functions λ_i $(i = 1, 2, \ldots, r)$ not all zero (these are the "Lagrange multipliers") such that z satisfies the Euler equation for the minimization of the integral

$$\int_0^1 \sum \lambda_i f_i(x, \dot{x}) dt$$

(summations are from 1 to r). That is, the following differential equation holds:

$$(1.4) \qquad \frac{d}{dt}\{\sum \lambda_i D_2 f_i(z, \dot{z})\} = \sum \lambda_i D_1 f_i(z, \dot{z}).$$

(D_1 and D_2 denote differentiation of $f(x, \dot{x})$ with respect to the x and $\dot{x}$ variables respectively.)

The proof of the multiplier rule was finally completed by Hilbert following the contributions of many mathematicians (see [**2**] for historical details). It turns out that the main requirement to assure its validity is the following:

(1.5) The vectors $D_2 f_i(z, \dot{z})$ in R^n are linearly independent for each t.

Received May 31, 1976.

https://doi.org/10.4153/CJM-1977-055-6 Published online by Cambridge University Press

Consider now a different problem, where instead of the equality constraints (1.3) being imposed, we have the inequality constraints

(1.6) $f_i(x, \dot{x}) \leqq 0 \quad (i = 1, 2, \ldots, r)$.

Forty years ago, F. A. Valentine [**12**] proposed a method (called that of "slack variables") whereby this problem could be treated by the existing theory for the case of equality constraints; ever since, it is this method that has been used in handling constraints of the form (1.6) (see for example [**1**; **7**; **8**]). When the multiplier rule is applied to the problem via Valentine's method, the analysis yields as before a nontrivial set of λ_i satisfying (1.4). Additionally, it follows that the λ_i are nonnegative, and that for any t such that $f_i(z, \dot{z}) < 0$ (the constraint $f_i \leqq 0$ is then said to be *inactive*), we have $\lambda_i(t) = 0$.

We stress that this approach to the multiplier rule for inequality constraints requires (as in the equality case) that hypothesis (1.5) be made (for the active indices).

The central thesis of this article is that the case of inequality constraints is best treated on its own. For example, we will show (Corollary 2) that in the example discussed above, hypothesis (1.5) can be replaced by the following weaker condition:

(1.7) The vectors $D_2 f_i(z, \dot{z})$ (active indices i) are convexly independent for each t,

by which we mean that no convex combination of these vectors is equal to zero. An immediate consequence of this is that we are now able to treat problems in which the number of (active) inequality constraints is greater than the dimension n (this would be precluded, of course, by condition (1.5)), and possibly infinite.

An equally important feature of the results is that no differentiability hypotheses intervene. We give an example in § 3 of a variant of a classical problem in which a nondifferentiable function appears quite naturally. The next section is devoted to the statement and elaboration of the main result, the proof of which is given in § 4.

2. A new multiplier rule. An *arc* is an absolutely continuous function $x : [0, 1] \to R^n$. We are given the functions $\varphi : R^n \to R$ and $f : R^n \times R^n \to R$, as well as two subsets C_0 and C_1 of R^n. The problem we consider is the following: to minimize

(2.1) $\varphi(x(1))$

over all arcs x which satisfy

(2.2) $x(0) \in C_0, \quad x(1) \in C_1$

as well as the inequality constraint

(2.3) $f(x, \dot{x}) \leqq 0$ a.e.

https://doi.org/10.4153/CJM-1977-055-6 Published online by Cambridge University Press

The notation "a.e." signifies "for almost all t in $[0, 1]$", in the sense of Lebesgue measure. The choice of the interval $[0, 1]$ is merely a convenient normalization.

The following hypotheses are made throughout: C_0 and C_1 are closed, and φ and f are locally Lipschitz. The requirement that φ (for example) be locally Lipschitz is equivalent to the following: for any bounded subset B of R^n, there is a scalar K (depending on B) such that for all x_1 and x_2 in B, we have

$$|\varphi(x_1) - \varphi(x_2)| \leqq K|x_1 - x_2|.$$

The classical multiplier rule is stated in terms of derivatives. Since differentiability is not being posited, a substitute for derivatives will be used. This is the "generalized gradient" introduced by the author in [**3**] (see [**6**] for the infinite- dimensional definition). In the case of a locally Lipschitz function $g : R^n \rightarrow R$, the generalized gradient of g at the point x, denoted $\partial g(x)$, may be defined as follows:

$$\partial g(x) = \operatorname{co}\left\{\zeta : \zeta = \lim_{i\to\infty} \nabla g(x_i), \lim_{i\to\infty} x_i = x\right\}. \tag{2.4}$$

That is, we consider all sequences x_i converging to x such that $\nabla g(x_i)$ exists for each i, and such that the indicated limit ζ exists. The convex hull of all the points ζ obtained in this way is $\partial g(x)$. It is evident that if g is C^1, then $\partial g(x) = \{\nabla g(x)\}$. Furthermore, it may be shown that when g is convex, $\partial g(x)$ is the subdifferential of convex analysis [**9**].

We now recall some terminology familiar from the calculus of variations. The arc z is a *weak local minimum* in the above problem if, for some positive ϵ, z solves the minimization problem (2.1)–(2.3) relative to the arcs x satisfying

$$|x(t) - z(t)| < \epsilon, \quad |\dot{x}(t) - \dot{z}(t)| < \epsilon \quad \text{a.e.}$$

The arc z is *piecewise-smooth* if there is a partition $0 = t_0 < t_1 \ldots < t_k = 1$ of $[0, 1]$ such that $\dot{z}$ exists and is continuous on $(t_{i-1}, t_i)(i = 1, 2, \ldots, k)$ and admits finite limits at both t_{i-1} (from the right) and t_i (from the left). These limits are denoted $\dot{z}(t_{i-1}+)$ and $\dot{z}(t_i-)$ respectively. When z fails to be differentiable at a point, z is said to have a *corner* there.

Definition. For a piecewise-smooth arc z, we say ∂f is *regular* along z if the following condition is satisfied for all t such that $f(z(t), \dot{z}(t)) = 0$:

$$(R^n \times \{0\}) \cap \partial f(z, \dot{z}) = \emptyset, \tag{2.5}$$

where for corner points t the condition is understood to hold with $\dot{z}(t)$ replaced by both $\dot{z}(t+)$ and $\dot{z}(t-)$. Thus ∂f is regular along z when the $\dot{x}$-component of any element of $\partial f(z, \dot{z})$ is nonzero, for any t such that $f(z, \dot{z}) = 0$.

THEOREM 1. *Let the piecewise-smooth arc z provide a weak local minimum for the problem* (2.1)–(2.3), *where ∂f is regular along z. Then there exist an arc p,*

https://doi.org/10.4153/CJM-1977-055-6 Published online by Cambridge University Press

a measurable function $\lambda : [0, 1] \to R$, *and a scalar* λ_0 *equal to* 0 *or* 1 *such that:*

(2.6) $(\dot{p}(t), p(t)) \in \lambda(t)\partial f(z(t), \dot{z}(t))$ *a.e.*,

(2.7) $\lambda(t) \geqq 0, \lambda(t) = 0$ *when* $f(z(t), \dot{z}(t)) < 0$,

(2.8) $p(0)$ *is normal to* C_0 *at* $z(0)$,

(2.9) *there is a vector* ζ *in* $\partial\varphi(z(1))$ *such that*
$-p(1) - \lambda_0\zeta$ *is normal to* C_1 *at* $z(1)$.

(2.10) $|p(t)| + \lambda_0$ *is never zero.*

Remark 1. The word "normal" appearing in the "transversality conditions" (2.8)–(2.9) is used in a generalized sense defined in [**3**]; this reduces to the usual concepts in the case of a C^1-manifold or a convex set. When there is no endpoint constraint (i.e. $C_1 = R^n$), it follows that $\lambda_0 = 1$, and (2.9) becomes

$$-p(1) \in \partial\varphi(z(1)).$$

The applicability of Theorem 1 may at first appear limited due to the fact that only the single inequality constraint (2.3) is considered, whereas most problems will incorporate multiple constraints. We shall see that in making the transition to such problems, the fact that f need not be differentiable is crucial. We indicate at the end of § 4 the modifications to be made in Theorem 1 when f has an explicit dependence on t.

Let us now consider the problem of minimizing (2.1) subject to (2.2) and the r inequality constraints

(2.11) $f_i(x, \dot{x}) \leqq 0 \quad (i = 1, 2, \ldots, r)$.

We shall suppose that each f_i is locally Lipschitz. Let us define f as follows:

(2.12) $f(s, v) = \max_{1 \leqq i \leqq r} f_i(s, v)$.

Then the system of inequalities (2.11) is equivalent to the single inequality (2.3).

COROLLARY 1. *Let the piecewise-smooth arc* z *provide a weak local minimum for the problem of minimizing* (2.1) *subject to* (2.2) *and* (2.11), *and suppose that for each* t, *for each point* (ζ_1, ζ_2) *in the common convex hull of the sets*

$$\partial f_i(z, \dot{z}), \quad i \text{ active},$$

we have $\zeta_2 \neq 0$. *Then there exist an arc* p, *measurable functions* $\lambda_i : [0, 1] \to R (i = 1, 2, \ldots, r)$, *and a scalar* λ_0 *equal to* 0 *or* 1 *such that* (2.8)–(2.10) *hold, and also:*

(2.13) $(\dot{p}, p) \in \sum \lambda_i(t)\partial f_i(z, \dot{z})$ *a.e.*,

(2.14) $\lambda_i \geqq 0, \lambda_i(t) = 0$ *when* $f_i(z, \dot{z}) < 0$.

Proof. When f is defined by (2.12), the set $\partial f(s, v)$ is contained in the common convex hull of the sets $\partial f_i(s, v)$ over the indices i for which the maximum in (2.12) is attained [**6**, Proposition 9]. It follows from this that ∂f is regular along z, so that Theorem 1 may be applied. Upon invoking a measurable selection theorem (see for example [**10**]), (2.6) yields: there exist nonnegative measurable functions γ_i such that

$$(\dot{p}, p) \in \lambda(t) \sum \gamma_i(t) \partial f_i(z, \dot{z}),$$

and if $f_i(z, \dot{z}) < 0$ then either $\lambda(t)$ or $\gamma_i(t)$ is zero. The required conclusions now follow upon setting $\lambda_i = \lambda \gamma_i$.

We now specialize to the classic case of continuous differentiability. As mentioned in § 1, hypothesis (1.5) is replaced by the less restrictive (1.7).

COROLLARY 2. *Let the piecewise-smooth arc z solve the problem of minimizing* (2.1) *subject to* (2.2) *and the r inequalities* (2.11), *where the functions f_i are C^1. Suppose that condition* (1.7) *holds. Then there exist measurable functions λ_i $(i = 1, 2, \ldots, r)$ and a scalar λ_0 equal to* 0 *or* 1 *such that*:

(2.15) $p(t) = \sum \lambda_i(t) D_2 f_i(z, \dot{z})$ *is an absolutely continuous function of t satisfying* (2.8)–(2.10),

(2.16) $\lambda_i \geqq 0, \lambda_i(t) = 0 \quad$ *when* $f_i(z, \dot{z}) < 0$,

(2.17) $$\frac{d}{dt}\left\{ \sum \lambda_i(t) D_2 f_i(z, \dot{z}) \right\} = \sum \lambda_i(t) D_1 f_i(z, \dot{z}).$$

Proof. It suffices to apply Corollary 1, noting that generalized gradients reduce here to derivatives.

Remark 2. In analogy to the classical case, the above allows us to assert that the λ_i are not all zero if no vector in $-\partial\varphi(z(1))$ is normal to C_1 at $z(1)$.

Remark 3. There is a theorem concerning the generalized gradient of the upper envelope of a family of functions [**3**, Theorem 2.1] that can be used to derive from Theorem 1 a version of the multiplier rule for an infinite number of constraints, in a manner completely analogous to that in which the above corollaries were obtained.

3. Example—Queen Dido and the badlands. Queen Dido is given a length of rope with which to enclose a region along the shore, the latter being represented by the line $x = 0$ in the $t - x$ plane (see Figure 1). In doing this, she seeks to join the point $(0, 0)$ to the point $(1, 0)$ by a curve of length L lying in the half-plane $x \geqq 0$ so as to maximize the area between the curve and the t-axis. The problem as described to this point is classical, but let us now suppose that for a given positive α, the terrain $x > \alpha$ is inferior, and worth only half as much as the terrain $x < \alpha$. The return corresponding to a choice

https://doi.org/10.4153/CJM-1977-055-6 Published online by Cambridge University Press

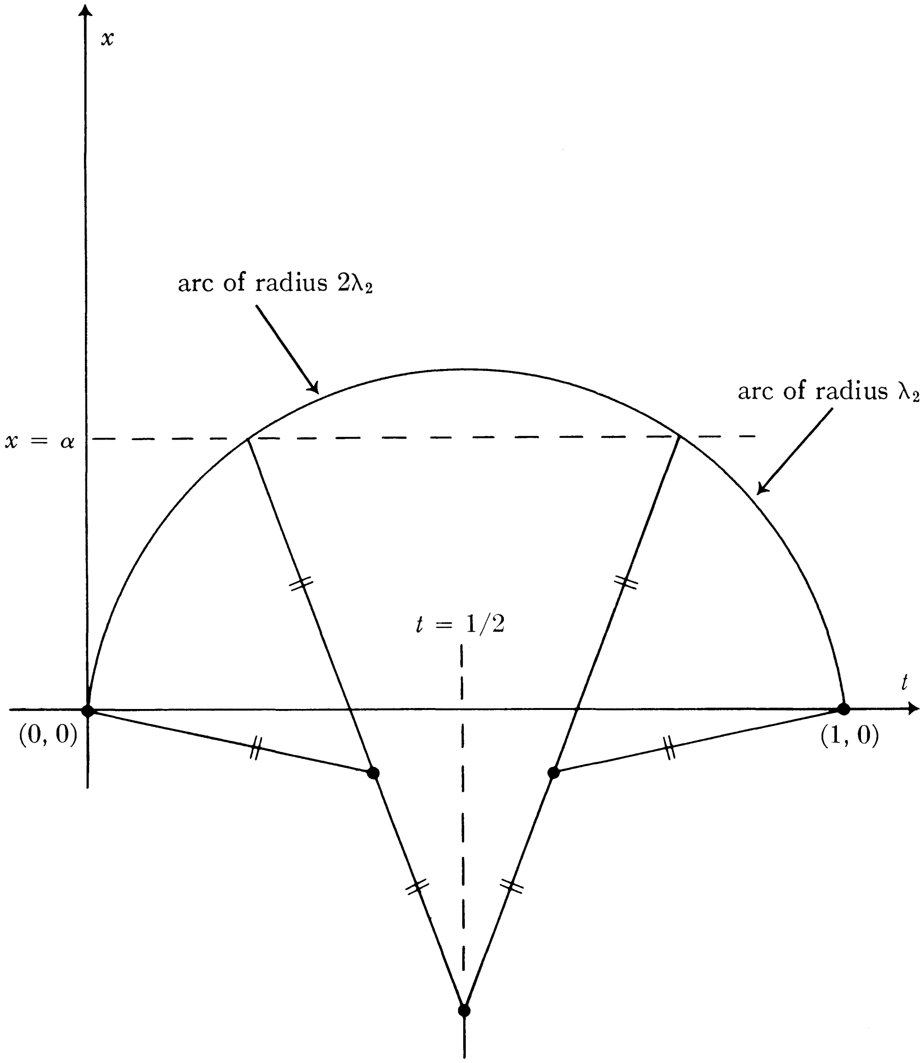

FIGURE 1

of border function $x(t)$ is then

$$(3.1) \qquad \int_0^1 g(x(t))dt,$$

where

$$g(x) = \begin{cases} x & \text{if } x \leqq \alpha \\ (x + \alpha)/2 & \text{if } x \geqq \alpha. \end{cases}$$

Her majesty is seeking to maximize (3.1) (or minimize its negative) subject to

$$x(0) = 0, \quad x(1) = 0, \tag{3.2}$$

$$\int_0^1 \sqrt{1 + \dot{x}^2}\, dt = L. \tag{3.3}$$

Note that g is Lipschitz and nondifferentiable.

We proceed to place this problem within the framework of § 2, Corollary 1. We consider the two additional variables y and z and the constraints

$$f_1(x, y, z, \dot{x}, \dot{y}, \dot{z}) = -\dot{y} - g(x) \leqq 0, \tag{3.4}$$

$$f_2(x, y, z, \dot{x}, \dot{y}, \dot{z}) = -\dot{z} + \sqrt{1 + \dot{x}^2} \leqq 0, \tag{3.5}$$

$$x(0) = 0, \quad y(0) = 0, \quad z(0) = 0, \quad x(1) = 0, \quad z(1) = L, \tag{3.6}$$

and we define

$$\varphi(x(1), y(1), z(1)) = y(1). \tag{3.7}$$

It is not difficult to see that the problem of minimizing (3.7) subject to (3.4)–(3.6) is equivalent to Queen Dido's. The equality (3.3) has been replaced by

$$\int_0^1 \sqrt{1 + \dot{x}^2}\, dt \leqq L$$

in this transition, which makes no difference in as much as all the available cord will be used. In fact, it is clear from the nature of the problem that both constraints (3.4) and (3.5) will be active at all times.

In applying Corollary 1, note that the vector x is here replaced by (x, y, z), that $n = 3$ and $r = 2$. The sets C_0 and C_1 are $\{(0, 0, 0)\}$ and $\{0\} \times R \times \{L\}$ respectively. The functions involved are Lipschitz as required, and the sets ∂f_1 and ∂f_2 are seen to be:

$$\partial f_1(x, y, z, \dot{x}, \dot{y}, \dot{z}) = \{(\zeta, 0, 0, 0, -1, 0) : -\zeta \in \partial g(x)\},$$
$$\partial f_2(x, y, z, \dot{x}, \dot{y}, \dot{z}) = \{(0, 0, 0, \dot{x}/\sqrt{1 + \dot{x}^2}, 0, -1)\},$$

from which we infer that the conclusions of Corollary 1 are available to us for any piecewise-smooth solution, which we shall denote (x, y, z). We deduce the existence of nonnegative functions λ_1 and λ_2 such that:

the function $p(t)$ defined by $p(t) = [\lambda_2 \dot{x}/\sqrt{1 + \dot{x}^2}, -\lambda_1, -\lambda_2]$ is absolutely continuous, and

$$\dot{p}(t) \in \{-\lambda_1(t)\partial g(x)\} \times \{0\} \times \{0\}. \tag{3.8}$$

It follows that λ_1 and λ_2 are constant. From (2.9) we obtain:

$$\lambda_1 - \lambda_0 = 0.$$

If λ_0 is zero, then λ_1 is zero also, and it follows from (2.10) that λ_2 must be

https://doi.org/10.4153/CJM-1977-055-6 Published online by Cambridge University Press

strictly positive. But then (3.8) implies that the sign of $\dot{x}$ is constant, which is not possible except in the degenerate case $L = 1$.

We may thus suppose $\lambda_0 = 1 = \lambda_1$. Now if λ_2 were zero, (3.8) would yield

$$0 \in \partial g(x),$$

which is not possible in view of (2.4). Thus λ_2 is positive.

We have arrived at the following conclusions: $\dot{x}$ is continuous and satisfies the equation

$$(3.9) \quad \begin{aligned} \frac{d}{dt}\{\dot{x}/\sqrt{1+\dot{x}^2}\} &= -1/\lambda_2 \quad \text{if } x < \alpha \\ &= -1/(2\lambda_2) \quad \text{if } x > \alpha. \end{aligned}$$

Note that $x(t)$ cannot equal α in any interval, since zero does not belong to $\partial g(\alpha)$.

The solutions to the two separate cases in (3.9) are well-known, since each case is the type of equation that arises in the classical version of Queen Dido's problem. We find with no difficulty that x describes an arc of a circle of radius λ_2 for $x < \alpha$, and an arc of a circle of radius $2\lambda_2$ for $x > \alpha$. The requirement that these arcs meet with a common tangent (at $x = \alpha$) assures that to each λ_2 there corresponds at most one such configuration (see Figure 1).

Consequently, the optimal arc x is uniquely specified once λ_2 is known; λ_2 is determined by the condition that x is of given length L. Once the nature of x is known to be as described above, it is an easy exercise to obtain (implicit) equations for λ_2 (and the other parameters of the solution). These relations could then be used to calculate explicitly the solution x.

It is interesting to determine the nature of the information contributed by the new multiplier rule. Based on the known classical solution, one might expect the solution to the present problem to consist of an amalgam of circular arcs on either side of the line $x = \alpha$ (as indeed it does). The multiplier rule has served to rule out the possibility that x lies along the line $x = \alpha$ for any length of time, and has yielded the crucial facts that the radii of the upper and lower arcs are in the ratio of two to one, and that these three pieces are smoothly joined. Thus the information obtained from its use has been essentially global.

4. Proof of Theorem 1. For ease of notation, we denote $f(z(t), \dot{z}(t))$ and $\partial f(z(t), \dot{z}(t))$ by $f(t)$ and $\partial f(t)$ respectively. When t is a corner point, there will be occasions when $f(t)$ is to be interpreted as $f(z(t), \dot{z}(t+))$ or $f(z(t), \dot{z}(t-))$, but the context will make this evident. The open unit ball in R^{2n} is denoted B.

LEMMA 1. *There is a constant M with the following property: given any t in $[0, 1]$ and (s, v) in $(z(t), \dot{z}(t)) + B$, then for all ζ in $\partial f(s, v)$ we have $|\zeta| \leqq M$.*

Proof. This follows from the hypothesis that f is Lipschitz on bounded sets, and from the definition (2.4) of generalized gradient.

LEMMA 2. *There exist positive numbers δ_1 and δ_2 such that, for any t in* $[0, 1]$, *for any* (s, v) *in* $(z(t), \dot{z}(t)) + \delta_1 B$, *for any* (α, β) *in* $\partial f(s, v)$, *we have* $|\beta| \geqq \delta_2$.

Proof. Suppose the lemma false. Then for each $i = 1, 2, \ldots$, there exist t_i in $[0, 1]$, (s_i, v_i) in $(z(t_i), \dot{z}(t_i)) + (1/i)B$ and (α_i, β_i) in $\partial f(s_i, v_i)$ such that $|\beta_i| < 1/i$. By taking subsequences we may assume that, for some t in $[0, 1]$, for some (s, v) and $(\alpha, 0)$ in R^{2n}, we have $t_i \to t$, $(s_i, v_i) \to (s, v)$, and $(\alpha_i, \beta_i) \to (\alpha, 0)$. It follows that $(s, v) = (z(t), \dot{z}(t))$. Furthermore, by the upper-semi-continuity of the generalized gradient [**3**], we know that $(\alpha, 0)$ belongs to $\partial f(t)$. This contradicts the regularity of ∂f along z.

Now let any positive integer K be given, and choose ϵ_K so that, for any t in $[0, 1]$, the inequality

$$|(s, v) - (z(t), \dot{z}(t))| < \epsilon_K$$

implies

$$f(s, v) \leqq f(z(t), \dot{z}(t)) + 1/K.$$

Such a choice is possible because f is uniformly continuous on compact sets. We may suppose that ϵ_K is less than $1/K$, and also less than the ϵ occurring in the definition of weak local minimum (§ 2).

Let us set

$$A_K(t) = \bigcup \{\zeta : \zeta \in \partial f(s, v), |(s, v) - (z(t), \dot{z}(t))| < 1/K\},$$

and define, for t such that $f(t) > -1/K$,

$$G_K(t) = A_K(t)^* = \{\gamma : \gamma \cdot \zeta \leqq 0 \text{ for all } \zeta \text{ in } A_K(t)\}.$$

For t such that $f(t) \leqq -1/K$, set $G_K(t) = R^{2n}$.

Now let K be larger than $1/\delta_1$. The following result then follows from Lemmas 1 and 2:

LEMMA 3. *There is a constant $N > 1$ such that the convex cone $G_K(t)$ has the following property for each t: given any s in R^n, there exists v in R^n such that $|v| \leqq N|s|$ and $(s, v) \in G_K(t)$.*

We now define a multifunction E_K from $[0, 1]$ to R^n as follows:

$$E_K(t, s) = \{v : |v| \leqq \epsilon_K/2, (s, v) \in G_K(t)\}.$$

In the terminology of [**5**], it follows that for $|s| < \epsilon_K/(2N)$, the multifunction $E_K(t, s)$ is nonempty, compact-valued, integrably bounded, measurable in t and Lipschitz in s with Lipschitz constant N.

LEMMA 4. *The arc $x(t) \equiv 0$ minimizes*

$$\varphi(z(1) + x(1))$$

over all arcs x satisfying $|x(t)| < \epsilon_K/(2N)$ and the constraints

$$(4.1) \qquad x(0) \in C_0 - z(0), \quad x(1) \in C_1 - z(1),$$
$$\dot{x}(t) \in E_K(t, x(t)) \quad a.e.$$

https://doi.org/10.4153/CJM-1977-055-6 Published online by Cambridge University Press

Proof. Let any such x be given. Notice that it suffices to prove the inequality

$$f(z + x, \dot{z} + \dot{x}) \leqq 0 \quad \text{a.e.,} \tag{4.2}$$

since then the fact that z is optimal for our original problem over a class of arcs including $z + x$ yields

$$\varphi(z(1)) \leqq \varphi(z(1) + x(1)).$$

In proving (4.2), consider first any t such that $f(t) \leqq -1/K$. Then (4.2) follows from the choice of ϵ_K, since we have

$$|(x(t), \dot{x}(t))| < \epsilon_K.$$

Now let us consider any t such that $f(t) > -1/K$. We have

$$f(z(t) + x(t), \dot{z}(t) + \dot{x}(t)) = f(t) + \int_0^1 Dg(\lambda)d\lambda, \tag{4.3}$$

where the Lipschitz function g is defined by

$$g(\lambda) = f(z(t) + \lambda x(t), \dot{z}(t) + \lambda\dot{x}(t)),$$

and $Dg(\lambda)$ exists a.e. It now suffices to prove that $Dg(\lambda)$ is nonpositive for λ in $[0, 1]$, since then (4.3) implies

$$f(z(t) + x(t), \dot{z}(t) + \dot{x}(t)) \leqq f(t) \leqq 0.$$

In turn, in order to prove the nonpositivity of $Dg(\lambda)$, it suffices to prove that $Dg(\lambda)$ belongs to the set (interval)

$$S = \partial f(z(t) + \lambda x(t), \dot{z}(t) + \lambda\dot{x}(t)) \cdot (x(t), \dot{x}(t)),$$

in view of the definition of $A_K(t)$ and the fact that $(x(t), \dot{x}(t))$ belongs to $G_K(t)$. We proceed now to prove this.

According to [**3**, Proposition 1.4] we have

$$\max\{\sigma : \sigma \in S\} = \limsup\, [f(z + \lambda x + h + \delta x, \dot{z} + \lambda\dot{x} + h' + \delta\dot{x}) \\ - f(z + \lambda x + h, \dot{z} + \lambda\dot{x} + h')]/\delta,$$

where the lim sup is taken as h and h' converge to 0 in R^n and δ decreases to 0. By definition, $Dg(\lambda)$ is equal to

$$\lim\, [f(z + \lambda x + \delta x, \dot{z} + \lambda\dot{x} + \delta\dot{x}) - f(z + \lambda x, \dot{z} + \lambda\dot{x})]/\delta$$

(limit as δ decreases to zero), whence

$$Dg(\lambda) \leqq \max\{\sigma : \sigma \in S\}.$$

A similar argument with $\min\{\sigma : \sigma \in S\}$ shows that $Dg(\lambda)$ belongs to the interval S. This completes the proof.

We now apply [**5**, Theorem 2] to the problem in the statement of Lemma 4. If the function $H : [0, 1] \times R^n \times R^n \to R$ is defined as follows:

$$H(t, s, p) = \max\{p \cdot v : v \in E_K(t, s)\},$$

https://doi.org/10.4153/CJM-1977-055-6 Published online by Cambridge University Press

we deduce that an arc p_K and a scalar λ_K equal to 0 or 1 exist such that:

(4.4) $(-\dot{p}_K, 0) \in \partial H(t, 0, p_K)$ a.e.,

(4.5) $p_K(0)$ is normal to $C_0 - z(0)$ at 0,

(4.6) for some vector ζ_K in $\partial\varphi(z(1))$,
$-p_K(1) - \lambda_K \zeta_K$ is normal to $C_1 - z(1)$ at 0,

(4.7) $|p_K(t)| + \lambda_K$ is never zero.

LEMMA 5. *For almost all* t,

(4.8) $\dot{p}_K \cdot s + p_K \cdot v \leqq 0$ *for all* $(s, v) \in G_K(t)$.

Proof. It suffices to show this for $|v|$ small, since $G_K(t)$ is a cone. Let t be such that (4.4) holds. Then we may suppose that v belongs to $E_K(t, s)$, and consequently

(4.9) $H(t, s, p_K) \geqq p_K \cdot v$.

It is elementary to verify that the function $H(t, x, p)$ is concave in x; along with (4.4), this implies that $-\dot{p}_K$ belongs to the superdifferential at 0 of the concave function $x \to H(t, x, p_K)$. From this we deduce:

(4.10) $H(t, s, p_K) - H(t, 0, p_K) \leqq -\dot{p}_K \cdot s$.

Since 0 belongs to $E(t, 0)$, it follows from the definition of H that we have

(4.11) $H(t, 0, p_K) \geqq 0$.

Now we combine (4.9)–(4.11) to obtain (4.8).

Remark. From Lemma 5 and the definition of $G_K(t)$ we deduce:

(4.12) $\dot{p}_K(t)$ and $p_K(t)$ are zero when $f(t) < -1/K$.

We shall now be considering all the above as the integer K increases to infinity. By taking subsequences, we may assume that the λ_K are either all 0 or all equal to 1, and that the ζ_K converge to a vector ζ. From the easily proven fact that the function $x \to H(t, x, p)$ is Lipschitz with constant $N|p|$, along with (4.4), we deduce:

(4.13) $|\dot{p}_K| \leqq N|p_K|$ a.e.,

where the constant N is independent of K (since G_K increases with K, N can only decrease as K increases).

LEMMA 6. *There exist an arc* p *and a scalar* λ_0 *equal to* 0 *or* 1 *satisfying* (2.8)–(2.10) *as well as*:

(4.14) $\dot{p} \cdot s + p \cdot v \leqq 0$ *for all* $(s, v) \in \partial f(t)^*$, *a.e.*,

(4.15) $\dot{p}$ *and* p *equal* 0 *when* $f(t) < 0$.

Proof. Case 1: The λ_K are all 0. By scaling, we may assume that all the p_K are nonvanishing and $\|p_K\| = 1$ ($\|\cdot\|$ denotes the supremum norm on $[0, 1]$), where the rescaled functions continue to satisfy (4.12), (4.13), (4.5) and (4.6) (with $\lambda_K = 0$). In view of (4.13), the Dunford-Pettis criterion implies that $\{\dot{p}_K\}$ admits a subsequence converging weakly in L^1 to $\dot{p}$ (say). It follows for suitable subsequences that $\dot{p}$ is the derivative of an arc p to which p_K converges uniformly (see [**4**, Lemma 5] for the details of the argument). Since p satisfies (4.13) and $\|p\| = 1$, (2.10) holds (with $\lambda_0 = 0$), as well as (2.8)–(2.9). Relation (4.15) is an immediate consequence of (4.12). In order to prove (4.14), note first that $G_K(t)$ increases to $\partial f(t)^*$ for any t such that $f(t) = 0$ (this uses the upper semicontinuity of ∂f [**3**]). Furthermore, weak convergence preserves linear inequalities such as (4.8); the result follows.

Case 2: The λ_K are all equal to 1, and $\|p_K\|$ is bounded. In this case the argument is unchanged, except that the need to rescale initially is eliminated. The conclusions (2.9)–(2.10) hold with $\lambda_0 = 1$.

Case 3: The λ_K are all equal to 1, and $\|p_K\|$ is unbounded. We may assume that $\|p_K\|$ increases to infinity. We rescale the arcs p_K by dividing by $\|p_K\|$ (which is certainly nonzero for K large). The argument then continues as in Case 1, and we get conditions (2.9) and (2.10) with $\lambda_0 = 0$, since $\lambda_K/\|p_K\|$ converges to 0. This proves the lemma.

In order to complete the proof of the theorem, it now suffices to infer (2.6) and (2.7) from (4.14) and (4.15). The condition (4.14) says that $(\dot{p}, p)$ belongs to $(\partial f(t)^*)^*$, which is the closed convex cone generated by $\partial f(t)$. This has the following characterization, for any t such that $f(t) = 0$:

$$(\partial f(t)^*)^* = \{\lambda\zeta : \lambda \geqq 0, \zeta \in \partial f(t)\},$$

because $\partial f(t)$ is a compact convex set not containing zero. Invoking a measurable selection theorem (see for example [**10**]), we obtain (2.6) when $f(t) = 0$, and (2.7) follows by simply setting $\lambda(t) = 0$ when $f(t) < 0$ and using (4.15).

Remark. The case in which f has an explicit dependence on t may be treated exactly as above with the additional hypotheses:

(a) $f(t, x, v)$ is a measurable function of t for each (x, v),

(b) $\partial f(t, x, v)$ is an upper semicontinuous multifunction (here, ∂f refers to the generalized gradient with respect to (x, v)).

Both these hypotheses are automatically satisfied when f is independent of t. In the case of t-dependence, (a) is required to ensure that the multifunction E_K constructed in the proof is measurable in t, while (b) is necessary for the conclusions of Lemmas 2 and 6.

References

1. L. D. Berkovitz, *Variational methods in problems of control and programming*, J. Math. Anal. Appl. *3* (1961), 145–169.

2. G. A. Bliss, *The problem of Lagrange in the calculus of variations*, Amer. J. Math. *52* (1930), 673–744.

3. F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc. *205* (1975), 247–262.

4. ——— *The maximum principle under minimal hypotheses*, SIAM J. Control and Optimization *14* (1976) 1078–1091.

5. ——— *Necessary conditions for a general control problem*, in Proceedings of the Symposium on the Calculus of Variations and Optimal Control, D. L. Russell, Editor (Mathematics Research Center, University of Wisconsin-Madison), Academic Press, N.Y. (1976).

6. ——— *A new approach to Lagrange multipliers*, Math. of Operations Res. *1* (1976) 165–174.

7. M. R. Hestenes, *Calculus of variations and optimal control theory* (Wiley, N.Y., 1966).

8. L. L. Pennisi, *An indirect sufficiency proof for the problem of Lagrange with differential inequalities as added side conditions*, Trans. Amer. Math. Soc. *74* (1953), 177–198.

9. R. T. Rockafellar, *Convex analysis* (Princeton Press, Princeton, N.J., 1970).

10. ——— *Measurable dependence of convex sets*, J. Math. Anal. Appl. *28* (1969), 4–25.

11. H. Sagan, *Introduction to the calculus of variations* (McGraw-Hill, N.Y., 1969).

12. F. A. Valentine, *The problem of Lagrange with differential inequalities as added side conditions*, in Contributions to the Calculus of Variations 1933–37, Department of Mathematics, University of Chicago (University of Chicago Press, Chicago).

University of British Columbia,
Vancouver, British Columbia