Control
and Cybernetics

VOL. 17 (1988) No. 2–3

Optimality and sensitivity of discrete time processes

by

R. B. VINTER

Department of Electrical Engineering
Imperial College
London SW7 2BT

A Pontryagin type maximum principle governing optimal discrete time processes has been known for many years. We give a new proof of recent refinements of the maximum principle for such processes: the data is permitted to be merely locally Lipschitz continuous in the state variable and the directional convexity hypothesis on the velocity set, invoked in earlier proofs, is weakened. Our approach is to study proximal normals to the epigraph of a value function. A byproduct of our methods is new sensitivity information regarding the dependence of the minimum cost on certain parameter values in the data.

1. Introduction

Discrete time optimal control concerns a class of optimization problems whose common feature is constraints involving difference equations. We shall study such problems, giving special emphasis to first order optimality conditions in the form of a maximum principle, within the following framework of the following problem, labelled P_0.

$$\text{Minimize } J(x_0, \ldots, x_N, u_0, \ldots, u_{N-1}) := \sum_{i=0}^{N-1} l_i(x_i, u_i) + h(x_0, x_N),$$

subject to

$$x_{i+1} = f_i(x_i, u_i), \quad i = 0, \ldots, N-1, \tag{1.1}$$

$$u_i \in U_i, \tag{1.2}$$

$$g_i(x_i) \leqslant 0, \quad i = 0, \ldots, N-1, \tag{1.3}$$

$$x_0, x_N \in A. \tag{1.4}$$

Here N, n, m and r are given positive integers,

$$l_i: R^n \times R^m \to R, f_i: R^n \times R^m \to R^n, \quad i = 0, \ldots, N-1,$$

$$g_i: R^n \to R^r, \ i = 0, \ldots, N-1, \ h: R^n \times R^n \to R,$$

are given functions, and

$$U_i \subset R^m, \ i = 0, \ldots, N-1, \ A \subset R^n \times R^n,$$

are given sets. The inequality (1.3) simply means the components of $g_i(x)$ must be non-negative.

A vector $(x_0, \ldots, x_N, u_0, \ldots, u_{N-1}) \in R^{n(N+1)} \times R^{mN}$ which satisfies the dynamical, control, state and endpoint constraints ((1.1)–(1.4) respectively) will be termed an **admissible process** (for P_0). We seek then an admissible process at which the value of J is a minimum. Such an admissible process is an **optimal process.**

The history of discrete time optimal control goes back virtually to inception of optimal control theory itself. There has been a resurgence of interest in discrete time problems in recent years, however, as digital control strategies gain ascendance over traditional analogue controllers throughout control engineering, and because of attention currently accorded to robotic control, an area where a good nonlinear, deterministic model is typically available and optimization issues are significant.

The evolution of the maximum principle mirrors that of other branches of optimization theory in many respects, notably continuous time optimal control. Early, direct approaches to deriving optimality conditions (see, e.g. Halkin's paper [5]) gave way to general theories of first order necessary conditions (such as those of Neustadt [11] and Ioffe and Tihomirov [8]) which treat discrete time optimal control as a special case. We refer also to the influential book by Boltyanskii [2]. In recent years advances in nonsmooth optimization have been absorbed into discrete time optimal control; they make possible streamlined treatment of implicit constraints (via the notion of the normal cone to a general closed set) and, of course permit consideration of nonsmooth data.

However, one of the most interesting recent developments in optimization theory has been new insights gained from the work of Aubin, Clarke, Gauvin, Rockafellar, Loewen and others into the relationship between Lagrange multipliers (or their equivalents in the maximum principle) and the sensitivity of the minimum cost ([1], [3], [4], [9], [10] and [12]). The implications for broad classes of problems in mathematical programming and continuous time optimal control have, in particular, received considerable attention. But these developments have not yet impinged on discrete time optimal control, to the author's knowledge; it is the main purpose of this paper to make good this omission.

It has long been appreciated that implicit in the maximum principle

is local information about the value function, which summarizes the effects on the minimum cost of changing certain parameter values. The recent developments we refer to is a program based on an inherently nonsmooth technique, proximal normal analysis, in which the relationship is used advantageously in both directions; examination of normal cones to the epigraph of the value function leads on the one hand to new proofs of first order optimality conditions, and on the other to formulae estimating the gradient of the value function. Since it is not reasonable to suppose that the value function is differentiable in a traditional sense, the gradients here are generalized ones.

This then is the program we pursue here, now with reference to discrete time optimal control. We will give a new proof of a maximum principle due to Ioffe ([7] and [8]), in which the hypotheses on the data are very mild, and also new estimates for generalized gradients to the value function relative to perturbations of the dynamical and state constraints.

Our final comment in this introduction concerns hypotheses on the velocity sets $Q_i(x) \in R^n$, $i = 0, 1, \ldots, N-1$:

$$Q_i(x) := \left\{ \begin{bmatrix} l_i(x, u) \\ f_i(x, u) \end{bmatrix} : u \in U_i \right\}.$$

It is well known fact about discrete time problems, supported by counter--examples (see, e.g., [5]), that some hypothesis akin to convexity of Q_i is required for the maximum principle to be valid. Emphasis has been given in the literature to the tasks of identifying and refining the hypothesis. This was in part for historical reasons: early attempts at derivations, in which convexity hypotheses did not feature, contained errors, and this needed to be stressed. The fact that the much publicized continuous time maximum principle was known to apply in the absence of such a hypothesis, has no doubt acted as a spur to weakening it in a discrete time setting as much as possible. This preoccupation is warranted also because, for digitally controlled nonlinear systems constrained to operate at low sampling rates, there is no reason to expect mathematical models with convex velocity sets will be accurate ones.

For purposes of simple description we drop the x dependence of Q_i. Let $\bar{v}(\in Q_i)$ be the value of $\begin{bmatrix} l_i \\ f_i \end{bmatrix}$ at the optimal process under consideration.

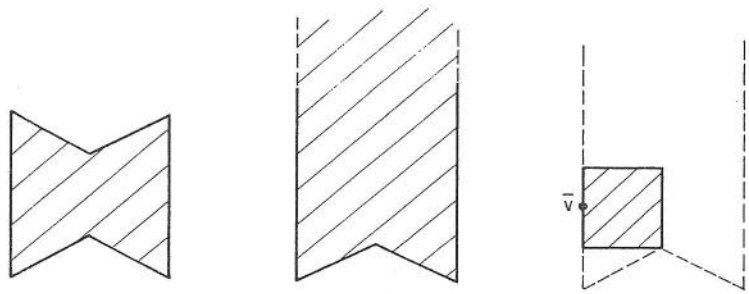
The discrete-time maximum principle was proved by Halkin [5] under the hypothesis that Q_i is convex. Convexity was subsequently weakened to **directional convexity** by Holtzman [6]. This is the hypothesis that Q^0 is convex where

$$Q_i^0 = Q_i + R^+ \times \{0\} \times \ldots \{0\}.$$

The hypothesis adopted here (which is implicit in Ioffe's work) is slightly weaker than Holtzman's. It is essentially: there exists $\alpha > 0$ such that

$$\bar{v} + \alpha \operatorname{conv}[Q_i - \bar{v}] \subset Q_i^0, \tag{1.5}$$

where "conv" denotes convex hull. (Convexity of Q_i^0 is equivalent to (1.5) if $\alpha = 1$). The hypothesis says that it must be possible to absorb the convex hull of Q_i^0 into Q^0 by shrinking it radially with respect to the point $\bar{v}$. The diagram illustrates a situation in which this last hypothesis is operative but the others fail.



The set Q_i (Q_i is not convex) The set Q° (Q_i° is not convex) The set $\bar{v} + \frac{1}{2}\operatorname{conv}[Q_i - \bar{v}]$ ($\subset Q_i^\circ$)

2. Generalized gradients

In later sections we make extensive reference to the calculus of generalized gradients. We briefly summarize here relevant aspects of the theory. (We refer to [3] for a full account).

Let C be a closed set in R^n and let $x \in C$.

We say that a non-zero vector $v \in R^n$ is a **proximal normal** to C at x, in symbols "$v \perp C$ at x", if there exists a positive number K such that

$$v \cdot (c - x) \leqslant K \|c - x\|^2, \quad \text{for all} \quad c \in C.$$

The geometric interpretation of v is that there exists $\alpha > 0$ such that the closed ball of radius $\alpha \|v\|$ and centre $x + \alpha v$ meets the set C at the single point x.

If v_i is a proximal normal to C at x_i for $i = 1, 2, \ldots$, such that $x_i \to x_i$ and $v_i \to v$, we say that v is a limiting normal to C at x.

The normal cone to C at x is the closed convex cone generated by the set of limiting proximal normals to C at x, i.e.

$$N_C(x) := \overline{\text{co}}\,\{\lim_{i\to\infty} v_i : v_i \perp C \text{ at } x_i,\ x_i \to x\}.$$

($\overline{\text{co}}$ denotes closed convex cone).

If x is a boundary point of C then there exist non-zero limiting proximal normals to C at x, and consequently $N_C(x)$ contains non-zero points.

Consider now a function $f: R^n \to R \cup \{+\infty\}$. We suppose that $f(x)$ is finite and the epigraph of f (written epif) is locally closed at $(f(x), x)$ (that is, the intersection of the set with some closed ball about $(f(x), x)$ is closed). The **generalized gradient** of f at x is the set

$$\partial f(x) = \{\xi \in R^m : (-1, \xi) \in N_{\text{epif}}(f(x), x)\}. \tag{2.3}$$

Information about non-Lipschitz aspects of f near x is embodied in the **asymptotic generalized gradient** of f at x:

$$\partial^\infty f(x) = \{\xi \in R^m : (0, \xi) \in N_{\text{epif}}(f(x), x)\}. \tag{2.4}$$

We remark that either $\partial f(x) \neq \emptyset$ or $\partial^\infty f(x) \neq \{0\}$, $\partial f(x)$ is a closed convex set and $\partial^\infty f(x)$ is a closed convex cone. The property "$\partial^\infty f(x) = \{0\}$" is a necessary and sufficient condition for f to be Lipschitz continuous in a neighbourhood of x.

A useful relationship between ∂f, ∂f^∞ and epif is

$$N_{\text{epif}}(f(x), x) = \{\lambda[-1, \xi] : \lambda > 0 \text{ and } \xi \in \partial V(x)\} \cup \{[0, \xi] : \xi \in \partial^\infty V(x)\}. \tag{2.5}$$

There is an important representation of $N_C(x)$ in terms of the Euclidean distance function $d_C(y)$ $(:= \text{Min}\,\{\|y-c\| : c \in C\})$, namely

$$N_C(x) = \overline{\bigcup_{\lambda>0}} \lambda\, \partial d_C(x).$$

Let $g: R^m \to \mathbf{R}$ be a function which is Lipschitz continuous on a neighbourhood of a point $x \in R^n$. The **generalized Jacobian** ∂_g of g at x is the set

$$\partial g(x) := \text{conv}\,\{\lim_{i\to\infty} \nabla g(x_i) : g \text{ is differentiable at } x_i,\ x_i \to x\}.$$

$\partial g(x)$ is a non-empty, closed set whose elements are bounded in norm by the Lipschitz rank of g in a neighbourhood of x. Our notation regarding generalized gradients and Jacobians is consistent since for scalar valued functions which are Lipschitz continuous in a neighbourhood of the base point the two concepts coincide.

We list several properties of generalized gradients and Jacobians for future use. (Throughout x is a point in R^n, g, $g': R^n \to R^m$ and $f: R^n \to R$ are functions which are Lipschitz continuous on a neighbourhood of x and C is a subset of C which is locally closed at x).

(i) (Regularity of the generalized Jacobian, treated as a multifunction of the base point): the graph of $y \to \partial g(y)$ is locally closed at x,

(ii) (A simple chain rule): for any m-vector d

$$\partial\,[d\cdot g\,(x)] = d\cdot\partial g\,(x)\,\big(:=\{d\cdot e : e\in\partial g\,(x)\}\big),$$

(iii) (Estimate of the generalized Jacobian of a sum of functions):

$$\partial\,(g+g')\,(x)\subset\partial g\,(x)+\partial g'(x),$$

and finally

(iv) (A result summarizing an exact penalization technique which is of great significance in nonsmooth optimization): suppose f is Lipschitz continuous on $x+\varepsilon B$ with rank k and

$$f\,(x)\leqslant f\,(y)\quad\text{for all}\quad y\in(x+\varepsilon B)\cap C.$$

Then

$$f\,(x)+kd_C\,(x)\leqslant f\,(y)+kd_C\,(y)\quad\text{for all}\quad y\in x+\varepsilon B.$$

Here B denotes the open unit ball.

3. A general optimization problem

The discrete time maximum principle is in essence a hybrid multiplier rule which incorporates the effects of local variations of certain variables and global variations of others. In this section we state such a multiplier rule. It relates to the optimization problem G_0:

$$\text{Minimize } F\,(x) \text{ over } (x,u)\in R^N\times R^M,$$

Subject to

$$G\,(x,u)=0,$$
$$x\in C,$$
$$u\in\Omega,$$

in which N, M and K are given integers, $F:R^N\to R$ and $G:R^N\times\mathbf{R}^M\to R^K$ are given functions, and $C\subset R^N$ and $\Omega\subset R^M$ are given sets. This is followed by a related sensitivity analysis.

As we shall see, the discrete time problem P_0 is merely problem G_0 in disguise and drawing conclusions about the discrete time problem will turn out to be largely a matter of transcription.

Define the function $L:R^N\times R^M\times R\times R^K\to R$ to be

$$L(x,u,\lambda,\mu):=\lambda F\,(x)+\mu\cdot G\,(x,u).$$

THEOREM 3.1. *Let $(\bar{x},\bar{v})$ solve problem G_0. Assume that*

H1: *C is a closed set,*

H2: *F is Lipschitz continuous on a neighbourhood of $\bar{x}$,*

H3: *for each $u\in\Omega$, $G\,(\cdot,u)$ is Lipschitz continuous on a neighbourhood of $\bar{x}$,* and

H4: *given any finite set* $\{u_1, \ldots, u_l\} \subset \Omega$ *there exists a positive number* α *such that, for all points* x *in some neighbourhood of* $\bar{x}$,

$$G(x, \bar{u}) + \alpha \Big(\sum_{j=1}^{l} \lambda_i \left[G(x, u_j) - G(x, \bar{u}) \right] \subset G(x, \Omega),$$

whenever $\{\lambda_1, \ldots, \lambda_l\} \in P^l$.

Then there exists a non-negative number λ and a vector $\mu \in R^K$ such that $(\lambda, \mu) \neq 0$,

$$0 \in \partial_x L(\bar{x}, \bar{u}, \lambda, \mu) + N_C(\bar{x}), \tag{3.1}$$

and

$$L(\bar{x}, \bar{u}, \lambda, \mu) \leqslant L(\bar{x}, u, \lambda, \mu) \quad \text{for all} \quad u \in \Omega. \tag{3.2}$$

The set P^l in hypothesis H4 is

$$P^l := \{(\lambda_1, \ldots, \lambda_l) \in R^l : \lambda_i \geqslant 0, 1, \ldots, l \quad \text{and} \quad \textstyle\sum_i \lambda_i \leqslant 1\}.$$

The set $\partial_x L$ is the generalized gradient with respect to the first vector coordinate.

This theorem was proved by Ioffe [7] (in fact under weaker hypotheses where x and u are permitted to belong to infinite dimensional linear spaces and where the convexity hypothesis H4 is required to hold only in an approximate sense). A new proof of Theorem 3.2, which yields as a byproduct sensitivity information about a minimum cost, is given in Section 6. Consider now the following hypotheses, which are slightly stronger than those invoked in Theorem 3.1. Let $\tilde{G}: R^N \times R^M \to R^{K+1}$ be the function

$$\tilde{G}(x, u) = \begin{bmatrix} F(x) \\ G(x, u) \end{bmatrix}.$$

S1: $\tilde{G}$ is continuous,

S2: there exists $\delta > 0$ such that, for any number γ, the set $\{(x, u): : \|G(x, u)\| < \delta, x \in C, u \in Q\} \cap \{(x, u): F(x) < \lambda\}$ is bounded,

S3: C and Ω are closed sets

and

S4: there exists a constant k such that, corresponding to any solution $(\bar{x}, \bar{u})$ to (G_0), $\varepsilon > 0$ can be chosen with the properties:

(a) $\|\bar{G}(x, u) - \bar{G}(y, u)\| \leqslant k \|x - y\|$, for all $x, y \in \bar{x} + \varepsilon B$ and $u \in Q$,

(b) $G(x, \Omega)$ is convex for all $x \in \bar{x} + \varepsilon B$,

(c) the graph of $(x, u) \to \partial_x \tilde{G}(x, u)$ is locally closed at $(\bar{x}, \bar{u})$.

We shall use the fact (see Corollary 6.1 below) that if $(\bar{x}, \bar{u})$ solves G_0 and the stronger hypotheses (S1)–(S4) are satisfied, then the conclusions of Theorem 3.1 can be strengthened as follows: "there exists $\lambda \geqslant 0$ and $\mu \in R^K$, such that $(\lambda, \mu) \neq 0$,

$$0 \in \partial_x L(\bar{x}, \bar{u}, \lambda, \mu) + k \cdot \|(\lambda, \mu)\| \cdot \partial d_C(\bar{x}), \tag{3.3}$$

and

$$L(\bar{x}, \bar{u}, \lambda, \mu) \leqslant L(\bar{x}, u, \lambda, \mu) \quad \text{for all} \quad u \in \Omega. \tag{3.4}$$

(The constant k is that of hypotheses S4). The difference here is that $N_C(\bar{x})$ has been replaced by the strict subset $k \cdot \|(\lambda, \mu)\| \, \partial d_C(\bar{x})$.

It is convenient to introduce notation to summarize the statement of the multiplier rule. Given $\lambda \geqslant 0$ and $(x, u) \in R^N \times R^M$ we define the index λ multiplier set at (x, u), $M^\lambda(x, u)$, to be

$$M^\lambda(x, u) := \{\mu \in R^K : (3.3) \text{ and } (3.4) \text{ are satisfied}\}.$$

We denote by Y the solution set for G_0

$$Y := \{(x, u) \in R^N \times R^M : (x, u) \text{ solves } G_0\},$$

and by $M^\lambda(Y)$ the set

$$M^\lambda(Y) := \cup \{M^1(x, u) : (x, u) \in Y\}.$$

In terms of our new notation, and under hypotheses S1–S4, the stronger statement of Theorem 3.1 is that, if $(\bar{x}, \bar{u})$ solves G_0, there exists $\lambda \geqslant 0$ and $\mu \in M^\lambda(x, u)$ such that $(\lambda, \mu) \neq 0$.

We direct attention now at sensitivity of the minimum cost to perturbaions of the equality constraint.

Given $\alpha \in R^K$, problem G_α is taken to be the new problem which results when, in problem G_0, the constraint "$G(x, u) = 0$" is replaced by "$G(x, u) = \alpha$". The value function of interest then is $W(\alpha)$,

$$W(\alpha) := \operatorname{Inf}\{G_\alpha\}.$$

(The infimum cost, $\inf\{G_\alpha\}$, is taken to have value to $+\infty$ if there are no vectors (x, u) which satisfy the constraints).

Under our strengthened hypotheses, we relate generalized gradients of W at $\alpha = 0$ to the multiplier sets.

THEOREM 3.2. *Suppose hypotheses* S1–S2 *are satisfied. Assume also that* $W(0) < +\infty$. *Then* W *is lower semi continuous in a neighbourhood of* 0 *and*

$$\partial W(0) = \overline{\text{co}}\,\{M^1(Y) \cap \partial W(0) + M^0(Y) \cap \partial^\infty W(0)\}. \tag{3.5}$$

If the cone $M^0(Y)$ *is pointed, i.e. contains no lines, then we can omit the closure operation in this identity and furthermore* $\partial^\infty W(0)$ *satisfies*

$$\partial^\infty W(0) = \text{co}\,\{M^0(Y) \cap \partial^\infty W(0)\}. \tag{3.6}$$

Identity (3.5) tells us a little bit more than

$$\partial W(0) \subset \overline{\text{co}}\,\{M^1(Y) + M^0(Y)\}. \tag{3.7}$$

This last inclusion is significant because often solution of the nominal

problem, by either analytical or numerical means, generates the multiplier sets $M^1(Y)$ or $M^0(Y)$, or approximations to these sets, along the way. In these circumstances we can estimate via (3.7) the effects on the minimum cost of small parameter changes from the nominal value $\alpha = 0$. Such information is required when we examine the implications of parameter drift in the model associated with the optimization problem, and also when we need to consider what small changes in the specifications of an optimal design problem will most enhance performance.

A further benefit we derive from the subgradient formulae is criteria for W to be finite-valued and regular, in some sense, in a neighbourhood of the nominal parameter value (and so, in particular for the constraints to remain consistent, even if they are subjected to small perturbations).

COROLLARY 3.3. *Let hypotheses* S1–S4 *hold and suppose that* $W(0) < +\infty$. *We have*

(i) *if* $M^0(Y) = \{0\}$ *then* W *is Lipschitz continuous in a neighbourhood of* 0 *and*

(ii) *if* $M^0(Y) = \{0\}$ *and* $M^1(Y)$ *consists of a single point then* W *is strictly differentiable at* 0.

"Strict differentiability", a property defined in [3], is intermediate in strength between Lipschitz continuity and continuous differentiability. The corollary is a direct consequence of Theorem 3.2 and the following facts. Firstly, an extended valued function on R^K, which is finite and whose graph is locally closed at a point in its domain, is Lipschitz continuous in a neighbourhood of that point if and only if the asymptotic gradient is just $\{0\}$ there. Secondly, if we are given a function on R^K, which is Lipschitz continuous on a neighbourhood of a point x in its domain, then the function is strictly differentiable at x if and only if the generalized gradient at x contains a single element. ([3, pp. 30 and 102]).

4. A discrete time maximum principle

We revert to the discrete time optimal control problem P_0 of Section 1. Let $(\{\bar{u}_i\}, \{\bar{x}_i\})$ be an optimal process.

In the following hypotheses the functions $\tilde{f}_i: R^n \times R^m \to R^{1+n}$, $i = 0, 1, \ldots$ $\ldots, N-1$, are taken to be

$$\tilde{f}_i(x, u) := \begin{bmatrix} l_i(x, u) \\ f_i(x, u) \end{bmatrix}, \quad \text{for} \quad (x, u) \in R^n \times R^m.$$

D1: h is Lipschitz continuous on a neighbourhood of $(\bar{x}_0, \bar{x}_N)$.

D2: $\tilde{f}_i(\cdot, u)$ is Lipschitz continuous on a neighbourhood of $\bar{x}_i$ for each $u \in U_i$ and for $i = 0, 1, \ldots, N-1$.

D3: g_i is Lipschitz continuous on a neighbourhood of $\bar{x}_i$ for $i = 0, 1, \ldots, N-1$.
D4: A is closed.
D5: Given any finite set $V_i \subset U_i$, there exists $\alpha > 0$ such that

$$\tilde{f}_i(x, \bar{u}_i) + \alpha\,[\mathrm{co}\,\tilde{f}_i(x, V_i) - \tilde{f}_i(x, V_i)] \subset \tilde{f}_i(x, V_i) + R^+ \times \{o\},$$

for all points x in some neighbourhood of $\bar{x}_i$ and $i = 0, 1, \ldots, N-1$.
We define the functions $H_i : R^n \times R^m \times R \times R^n \times R^r \to R$, $i = 0, 1, \ldots, N-1$, to be

$$H_i(x, u, \lambda, p, r) := -\lambda l_i(x, u) + p_i \cdot f_i(x, u) - r_i \cdot g(x).$$

THEOREM 4.1. *Suppose that hypotheses* D1–D5 *are satisfied. Then there exists a non-negative number* λ, *n-vectors* $p_0, \ldots, p_N$ *and r-vectors* $r_0, \ldots, r_{N-1}$, *not all zero*, such that

$$p_i \in \partial_{x_i} H_i(\bar{x}_i, \bar{u}_i, \lambda, p_{i+1}, r_i) \quad \text{for} \quad i = 0, 1, \ldots, N-1,$$

$$H_i(\bar{x}_i, \bar{u}_i, \lambda_i, p_i, r_i) = \operatorname*{Max}_{u \in U_i} H_i(\bar{x}_i, u, \lambda, p_i, r_i) \quad \text{for} \quad i = 0, 1, \ldots, N-1,$$

$$(p_0, -p_N) \in \lambda\, \partial h(\bar{x}_0, \bar{x}_N) + N_A((\bar{x}_0, \bar{x}_N)),$$

$$r_i \geqslant 0 \quad \text{for} \quad i = 0, 1, \ldots, N-1,$$

and

$$r_i \cdot g_i(\bar{x}_i) = 0 \quad \text{for} \quad i = 0, 1, \ldots, N-1.$$

We comment on some variants of Theorem 4.1.
Consider the discrete time optimal control problem which arises when we delete the unilateral state constraints "$g_i(x_i) \leqslant 0$" in P_0. For this problem a set of necessary conditions are obtained from those for P_0 by setting the multipliers $r_0, \ldots, r_{N-1}$ to zero. (Now $(\lambda, \{p_i\}) \neq 0$). To show this, set $g \equiv -1$ and apply Theorem 4.1.

Problem P_0 is just one example of a discrete optimal control problem, necessary conditions for which may be simply deduced from the general multiplier rule, Theorem 3.1. There is no difficulty in accomodating, for example, problems in which the unilateral state constraints are replaced by mixed uni- and bilateral constraints of the form

$$\psi_i(x_i, u_i) \leqslant 0 \quad \text{and} \quad \varphi_i(x_i, u_i) = 0 \quad \text{for} \quad i = 0, 1, \ldots, N-1.$$

Such problems admit simple reformulations (such as that given below) as special cases of problem P_0, and necessary conditions are obtained by applying Theorem 3.1.

We now describe a suitable reformulation of the discrete time optimal control problem P_0 as a special case of G_0. The partitioned vector variables $(x_0, \ldots, x_N, z_0, \ldots, z_N)$ and $(u_0, \ldots, u_{N-1}, v_0, \ldots, v_{N-1}, w_0, \ldots, w_{N-1})$ in P_0 assume the roles of x and u respectively in G_0. We set

$$G\left(\left(\{x_i\}, \{z_i\}\right), \left(\{u_i\}, \{v_i\}, \{w_i\}\right)\right) := \\ := \begin{bmatrix} x_1 - f_0(x_0, u_0) \\ \vdots \\ x_N - f_{N-1}(x_{N-1}, u_{N-1}) \\ l_0(x_0, u_0) + v_0 + z_0 - z_1 \\ \vdots \\ l_{N-1}(x_{N-1}, u_{N-1}) + v_{N-1} + z_{N-1} - z_N \\ g_0(x_0) + w_0 \\ \vdots \\ g_{N-1}(x_{N-1}) + w_{N-1} \end{bmatrix},$$

$$F\left(\left(\{x_i\}, \{z_i\}\right)\right) := h(x_0, x_N) + z_N,$$

$$C := \left\{\left(\{x_i\}, \{z_i\}\right) : (x_0, x_N) \in A,\ z_0 = 0\right\},$$

and

$$\Omega := \left\{\left(\{u_i\}, \{v_i\}, \{w_i\}\right) : u_i \in U_i,\ v_i \geqslant 0,\ w_i \geqslant 0 \quad \text{for all } i\right\}.$$

Proof of Theorem 4.1. When the above identifications are made we find that $\left(\left(\{\bar{x}_i\}, (z_i = 0)\right), \left(\{\bar{u}_i\}, \{v_i = 0\}, \{w_i = 0\}\right)\right)$ solves G_0, and the hypotheses of Theorem 3.1 are satisfied with reference to this solution. The multipliers in this case comprise a number $\lambda \geqslant 0$ and vectors $p_1, \ldots, p_N, \ldots, q_{N-1}, r_0, \ldots, r_{N-1}$ (not all of which are zero). The stationarity condition (3.1) yields

$$0 \in p_i - \partial_{x_i}\left[p_{i+1} \cdot f_i(\bar{x}_i, \bar{u}_i) - q_i\, l_i(\bar{x}_i, \bar{u}_i) - r_i \cdot g_i(x_i)\right], \quad \text{for} \quad i = 1, \ldots, N-1,$$

$$0 \in \partial_{x_0, x_N}\left[\lambda h(\bar{x}_0, \bar{x}_N) + p_N \cdot \bar{x}_N - p_1 \cdot f_0(\bar{x}_0, \bar{u}_0) + \right. \\ \left. + q_0\, l_0(\bar{x}_0, \bar{u}_0) + r_0 \cdot g_0(\bar{x}_0)\right] + N_A\left((\bar{x}_0, \bar{x}_N)\right), \tag{4.1}$$

and

$$q_0 = q_1 = \ldots = q_{N-1} = \lambda,$$

while from the minimization of the Hamiltonian condition (3.2) we deduce

$$-p_{i+1} \cdot f_i(\bar{x}_i, \bar{u}_i) + q_{i+1}\, l_1(\bar{x}_i, \bar{u}_i) \leqslant -p_{i+1} \cdot f(\bar{x}_i, \bar{u}_i) + q_{i+1}\, l_i(\bar{x}_i, \bar{u}_i) \\ \text{for all} \quad v \in U_i,\ i = 0, 1, \ldots, N-1,$$

$$r_i \geqslant 0, \quad i = 0, 1, \ldots, N-1,$$

and

$$g_i(\bar{x}_i) = 0 \quad \text{if} \quad r_i \neq 0.$$

Condition (4.1) together with property (iii) of Section 3 yield

$$(0, -p_N) \in \lambda\, \partial h(\bar{x}_0, \bar{x}_N) - \partial_{x_0}\left[p_1 \cdot f_0(\bar{x}_0, \bar{u}_0) - \right. \\ \left. - q_0\, l_0(\bar{x}_0, \bar{u}_0) - r_0 \cdot g_0(\bar{x}_0)\right] \times \{0\} + N_A\left((\bar{x}_0, \bar{x}_N)\right).$$

It follows that

$$(p_0, -p_N) \in \lambda\, \partial h\,(\bar{x}_0, \bar{x}_N) + N_A\,(\bar{x}_0, \bar{x}_N),$$

for some element

$$p_0 \in \partial_{x_0}\,[p_1 \cdot f_0\,(\bar{x}_0, \bar{u}_0) - q_0\, l_0\,(\bar{x}_0, \bar{u}_0) - r_0 \cdot g_0\,(\bar{x}_0)].$$

Scanning these relationships, we see that the assertions of the theorem are verified. (Note in particular that $(\lambda\,\{p_i\}, \{r_i\}) \neq 0$ since $\{q_i\} = 0$ if $\lambda = 0$). ■

5. Sensitivity of the minimum cost for discrete time problems

Theorem 3.2 provides information about the generalized gradient of the value function, corresponding to perturbations of the equality constraints. This yields, in turn, like results for the discrete time optimal control problem, via the reformulation of the previous section.

We label $P_{(\alpha,\beta)}$ the perturbed discrete time optimal control problem:

$$\text{Minimize } h\,(x_0, x_N) + \sum_{i=0}^{N-1} l_i\,(x_i, u_i),$$

subject to

$$x_{i+1} = f_i\,(x_i, u_i) + \alpha_i, \quad i = 0, \ldots, N-1,$$
$$g_i\,(x_i) \leqslant \beta_i \quad i = 0, \ldots, N-1,$$
$$(x_0, x_N) \in A,$$

and

$$u_i \in U_i, \quad i = 0, \ldots, N-1.$$

Here $(\alpha, \beta)(= (\{\alpha_i\}, \{\beta_i\}))$ is a point in $R^{nN} \times R^{mN}$. The nominal problem, P_0, is obtained by setting $(\alpha, \beta) = 0$.

We define $V: R^{nN} \times R^{mN} \to R$ to be the corresponding value function,

$$V\,(\alpha, \beta) := \operatorname{Inf}\,\{P_{(\alpha,\beta)}\}.$$

The following hypotheses are invoked:

E1: h is locally Lipschitz continuous,

E2: corresponding to any number σ there exists $\varepsilon > 0$ such that the set $\{\text{admissible processes for } P_{(\alpha,\beta)}: (\alpha, \beta) \in \varepsilon B \text{ and } h + \sum_i l_i < \sigma\}$ is bounded,

E3: C and Ω are closed,

E4: $\tilde{f}_i$ is continuous for $i = 0, \ldots, N-1$,

E5: there exists a number k such that, corresponding to any minimizing process $(\{\bar{x}_i\}, \{\bar{u}_i\})$, $\varepsilon > 0$ can be chosen with the following properties

(a) $\|\tilde{f}_i(x,u)-\tilde{f}_i(y,u)\| \leqslant k\,\|x-y\|$ for $(x,y)\in\bar{x}+\varepsilon B$ and $u\in U_i$, $i=0,\dots,N-1$,

(b) the set

$$(\alpha, f_i(x,u)):\alpha \geqslant l_i(x,u),\ u\in U_i,$$

is convex for all $x\in\bar{x}_i+\varepsilon B$, $i=0,\dots,N-1$,

and

(c) the graph of the multifunction

$$(x,u)\to\partial_x\tilde{f}_i(x,u),$$

is locally closed at $(\bar{x}_i,\bar{u}_i)$, $i=0,\dots,N-1$.

Under these hypotheses, the assertions of Theorem 4.1 are valid in strengthened form, where the transversality condition is expressed in terms of ∂d_A instead of N_A. (This follows from Corollary 6.1, via the reformulation of the discrete time optimal control problem described in Section 4). Our definition of multiplier sets R^λ for the discrete time problem is inspired by these stronger optimality conditions. Given $\lambda\geqslant 0$ and an admissible process $(\{x_i\},\{u_i\})$ we define

$$R^\lambda(\{x_i\},u_i\}):=$$
$$:=\{\{p_i\}_{i=1}^N,\{r_i\}_{i=0}^N:\text{conditions (a)–(d) are met, for some vector } p_0\}.$$

The conditions referred to are

(a) $$p_i\in\partial_x H_i(x_i,u_i,\lambda,p_i,r_i),\quad i=0,\dots,N-1,$$

(b) $$H_i(x_i,u_i,\lambda,p_i,r_i)=\sup_{u\in U_i}\{H_i(x_i,u,\lambda,p_i,r_i)\}\quad i=0,\dots,N-1,$$

(c) $$(p_0,-p_N)\in\lambda\partial h(x_0,x_N)+\bar{k}\partial d_A(x_0,x_N),$$

and

(d) $$r_i\geqslant 0,\quad i=0,\dots,N-1,$$

and

$$r_i\cdot g_i(x_i)=0,\quad i=0,\dots,N-1.$$

Let Z be the subset of $R^{n(N+1)}\times R^{mN}$ comprising solutions to the nominal problem P_0. In condition (c) $\bar{k}$ is a constant whose magnitude is determined by the Lipschitz rank of the data of the discrete time optimal control problem on some ball about Z.

THEOREM 5.1. *Suppose that* $V(0,0)<+\infty$. *Then under hypotheses* E1–E5 *we have*

$$\partial V(0,0)=\overline{\text{co}}\,\{R^1(Z)\cap\partial V(0,0)+R^0(Z)\cap\partial^\infty V(0,0)\}.$$

If the cone $R^0(z)$ *is pointed then we can omit the closure operation from this identity and, furthermore,* $\partial^\infty V(0,0)$ *satisfies*

$$\partial^\infty V(0,0) = \overline{\text{co}}\,\{R^0(Z) \cap \partial^\infty V(0)\}.$$

We refer back to Section 3 for comments on the significance of subgradient formulae such as these.

Proof. We begin by noting a minor variant of Theorem 3.2, namely "suppose that the hypotheses of Theorem 3.2 are satisfied in all respects, except that in hypothesis S4, part (c) the function G is inserted in place of $\tilde{G}\left(=\begin{bmatrix} F \\ G \end{bmatrix}\right)$". Then the assertions of the theorem remain valid provided the multiplier sets M^λ in (3.5) and (3.6) are replaced by

$$\tilde{M}^\lambda(x,u) = \{\mu \in R^K : 0 \in \lambda\, \partial F(x) + \mu \cdot \partial_x G(x,u) + k\, \partial d_C(x)\},$$

and

$$H(x,u,\lambda,\mu) \leqslant H(x,v,\lambda,\mu) \qquad \text{for all} \qquad v \in \Omega,$$

The difference here is that the multiplier rule implicit in $\tilde{M}^\lambda$ is expressed in separated form.

To prove this we use the same arguments as before except, before passing to the limit in the step summarized by Lemma 6.1, we replace inclusion (6.1) by the weaker inclusion

$$0 \in \lambda_i\, \partial_x F(x_i) + \mu \cdot \partial_x G(x_i, u_i) + \cdot \|(\lambda_i, \mu_i)\| \cdot \partial d_C(x_i).$$

We now apply the modified version of Theorem 3.2 to P_0, reformulated as a special case of problem G_0 in the manner described in Section 4.

There result subgradient formulae for a value function $\tilde{V}(\alpha,\sigma,\beta)$ expressed in terms of index λ multiplier sets $\tilde{R}^\lambda$. This is the value function associated with the optimization problems $\tilde{P}_{(\alpha,\sigma,\beta)}$, $(\alpha,\sigma,\beta) = (\{\alpha_i\},\{\sigma_i\},\{\beta_i\}) \subset R^{nM} \times \times R^{rN} \times R^N$:

Minimize $h(x_0, x_N) + z_N$
subject to

$$\begin{aligned} x_{i+1} &= f_i + \alpha_{i+1}, \quad i = 0, \ldots, N-1, \\ g_i + w_i &= \sigma_i \quad i = 0, \ldots, N-1, \\ l_i &= v_i + z_i - z_{i+1} - \beta_i \quad i = 0, \ldots, N-1, \\ ((x_0, x_N), Z_0) &\in A \times \{0\}, \quad i = 0, \ldots, N-1. \end{aligned}$$

It remains to translate these formulae into statements about the value function $V(\alpha,\sigma)$. This is easily accomplished since the extra perturbation vector $\{\beta_i\}$ associated with $\tilde{V}$ affects the minimization problem in only a trivial way. In fact it is easy to check that the set of solutions to $\tilde{P}_{(0,0,0)}$ coincide with that of solutions to $P_{(0,0)}$ and that for arbitrary values of $\{\alpha_i\}$ $\{\sigma_i\}$, $\{\beta_i\}$ and $\lambda \geqslant 0$ and an arbitrary solution $(\{x_i\},\{u_i\})$ to $P_{(0,0)}$ we have

$$\tilde{V}(\alpha, \sigma, \beta) = V(\{\alpha_i\}, \{\sigma_i\}) + \sum_{i=0}^{N-1} \beta_i,$$

and

$$\tilde{R}^{\lambda}(\{x_i\}, \{u_i\}) = R^{\lambda}(\{x_i\}, \{u_i\}) \times \{\lambda, \lambda, \ldots, \lambda\}).$$

We deduce from (5.1) that

$$\partial \tilde{V}(0, 0, 0) = \partial V(0, 0) \times \{1, \ldots, 1\},$$

and

$$\partial^{\infty} \tilde{V}(0, 0) = \partial^{\infty} V(0, 0) \times \{0, \ldots, 0\}.$$

The theorem is proved by appealing to these relationships and by projecting both sides of the subgradient formulae already obtained onto the subspace $R^{nN} \times R^{rN} \times \{0\}$. ■

6. Proof of theorems 3.1 and 3.2

The following lemma, which estimates limiting proximal normals to the epigraph of the value function W introduced in Section 2, is the key element in our proof of the general multiplier rule (Theorem 3.1) and of associated sensitivity results. Throughout this section k is the constant of hypothesis S4.

LEMMA 6.1. *Consider the family of problems* $\{G_\alpha\}$ *and the associated value function* W. *Suppose that hypotheses* S1–S4 *of Section 3 are satisfied and that* $W(0) < +\infty$. *Then epi* W *is locally closed at* $(W(0), 0)$. *Now let* $(-\bar{\lambda}, -\bar{\mu})$ *be an arbitrary limiting proximal normal to epi* W *at* $(W(0), 0)$. *Then* $\bar{\lambda} \geqslant 0$ *and there exists a solution* $(\bar{x}, \bar{u})$ *to* G_0 *such that*

$$0 \in \partial_x L(\bar{x}, \bar{u}, \bar{\lambda}, \bar{\mu}) + k \cdot \|(\bar{\lambda}, \bar{\mu})\| \cdot \partial d_C(\bar{x}),$$

and

$$L(\bar{x}, \bar{u}, \bar{\lambda}, \bar{\mu}) = \min_{u \in \Omega} L(\bar{x}, \bar{u}, \bar{\lambda}, \bar{\mu}).$$

Proof. Standard compactness arguments coupled with hypotheses S1–S4 lead us to the conclusion that epi W is locally closed at $(W(0), 0)$ and also that G_α has a solution whenever $W(\alpha) < +\infty$ and $\|\alpha\|$ is sufficiently small. It makes sense then to speak of proximal normals to epi W at points in a neighbourhood of $(W(0), 0)$. We assume that the limiting proximal normal $(-\bar{\lambda}, -\bar{\mu})$ under consideration is non-zero, for otherwise the assertions of the lemma are trivial. By definition of the limiting proximal normal there exist a sequence $\{\alpha_i\}$ of points in R^K, a sequence $\{\delta_i\}$ of non-negative numbers and a sequence of vectors $\{-\lambda_i, -\mu_i\}$ such that

$$(-\lambda_i, -\mu_i) \perp \text{epi } W \quad \text{at } (W(\alpha_i) + \delta_i, \alpha_i), \qquad i = 1, 2, \ldots$$

$$(W(\alpha_i)+\delta_i, \alpha_i) \to (W(0), 0) \quad \text{as} \quad i \to \infty,$$

and

$$(\lambda_i, \mu_i) \to (\bar{\lambda}, \bar{\mu}) \quad \text{as} \quad i \to \infty.$$

We can arrange (by elimination of initial elements from the sequences) that G_{α_i} has a solution, written (x_i, u_i), for $i = 1, 2, \ldots$. Since the points (x_i, u_i), $i = 1, 2, \ldots$ are confined to a bounded subset of the closed set $C \times \Omega$, we can also arrange (this time by subsequence extraction) that

$$x_i \to \bar{x} \quad \text{and} \quad u_i \to \bar{u} \quad \text{as} \quad i \to \infty,$$

for some point $(\bar{x}, \bar{\mu}) \in C \times \Omega$. From the continuity of F and G we deduce that $(\bar{x}, \bar{\mu})$ satisfies the constraints of problem G_0 whence

$$W(0) \leqslant F(\bar{x}).$$

However $F(x_i) = W(\alpha_i)$ for $i = 1, 2, \ldots$, so

$$W(0) \leqslant F(\bar{x}) = \lim_i F(x_i) \leqslant \lim_i F(x_i) + \delta_i = W(0).$$

It follows that $(\bar{x}, \bar{\mu})$ solves G_0.

Now according to the proximal normal inequality (2.1) applied to epi W at $(W(\alpha_i)+\delta_i, \alpha_i)$, the functional J_i is minimized over $C \times \Omega \times R^+$ at $(\bar{x}_i, \bar{u}, \bar{\delta}_i)$ for $i = 1, 2, \ldots$ where

$$J_i(x, u, \delta) := \lambda_i (F(x)+\delta) + \mu_i \cdot G(x, u) + \\ + K_i [|F(x)+\delta - F(x_i) - \delta_i|^2 + \|G(x, u) - G(x_i, u_i)\|^2].$$

(The K_i's are appropriately chosen constants).

Fix $(x, u) = (x_i, u_i)$ and consider minimization with respect to the δ variable. If the minimizer $\delta_i > 0$ then λ_i must be zero. On the other hand, if $\delta_i = 0$ then $\lambda_i \geqslant 0$. In either case we have $\lambda_i \geqslant 0$.

Now fix $u = u_i$ and $\delta = \delta_i$, and consider minimization with respect to the x variable. By eliminating initial terms in our sequences we can arrange that $\|x_i - \bar{x}\| < \varepsilon$ for all i. (Here ε is the constant in hypothesis S4 associated with the solution $(\bar{x}, \bar{u})$ to G_0). By property (iv) listed in Section 2,

$$0 \in \partial_x L(x_i, u_i, \lambda_i, \mu_i) + k \cdot \|(\lambda_i, \mu_i)\| \cdot \partial d_C(x_i). \tag{6.1}$$

Next fix $x = x_i$ and $\delta = \delta_i$, and consider minimization with respect to the u variable. We see that the function

$$\omega \to \mu_i \cdot \omega + K_i \|\omega - G(x_i, u_i)\|^2,$$

is minimized over $G(x_i, \Omega)$ at $G(x_i, u_i)$. Let $u \in \Omega$ and $\sigma > 0$ be arbitrarily chosen. Since $G(x_i, \Omega)$ is convex we have

$$(1-\sigma) G(x_i, u_i) + \sigma G(x_i, u) \in G(x_i, \Omega).$$

It follows that

$$\sigma\mu_i \cdot [G(x_i, u) - G(x_i, u_i)] + \sigma^2 K_i \|G(x_i, u) - G(x_i, u_i)\|^2 \geqslant 0.$$

Dividing across by σ and passing to the limit as $\sigma \to 0$ we obtain

$$\mu_i \cdot G(x_i, u) \geqslant \mu_i \cdot G(x_i, u_i). \tag{6.2}$$

Here, we recall, u is an arbitraty element in Ω.

We now pass to the limit $i \to \infty$. We obtain $\bar{\lambda} \geqslant 0$ since $\bar{\lambda}_i \geqslant 0$ for $i = 1, 2, \ldots$. Inequality (6.2) is preserved in the limit since G is continuous.

The inclusion (6.1) too is preserved in the limit. This follows from properties (i) and (ii) listed in Section 2, together with hypothesis S4. The lemma is proved.

Since $(W(0), 0)$ is a boundary point of epi W and epi W is locally closed at $(W(0), 0)$, the set of non-zero limiting proximal normals at $(W(0), 0)$ is non-empty. This observation together with Lemma 6.1 yields the following preliminary multiplier rule:

COROLLARY 6.2. *Suppose that hypotheses* S1–S4 *of Section* 2 *are satisfied and that there is a unique solution* $(\bar{x}, \bar{u})$ *to* G_0. *Then there exists a non-negative number* $\lambda \geqslant 0$ *and a vector* $\mu \in R^K$ *such that* $(\lambda, \mu) \neq 0$,

$$0 \in \partial_x L(\bar{x}, \bar{u}, \lambda, \mu) + k \cdot \|(\lambda, \mu)\| \cdot \partial d_C(\bar{x}),$$

and

$$L(\bar{x}, \bar{u}, \lambda, \mu) = \operatorname*{Max}_{u \in \Omega} L(\bar{x}, u, \lambda, \mu).$$

We are ready to prove a multiplier rule under merely the hypotheses of Theorem 3.1. This is achieved by applying the preliminary multiplier rule along a sequence of optimization problems, the data in each of which satisfies the more stringent hypotheses of Corollary 6.2.

Proof of Theorem 3.1. Choose a monotone sequence of positive numbers $\{k_i\}$, $k_i \to \infty$. Let S_i, $i = 1, 2, \ldots$ be an increasing sequence of finite subsets of the open unit ball B in R^k such that

$$B \subset S_i + k_i^{-2} B, \quad i = 1, 2, \ldots.$$

Define

$$\tilde{\Omega}_i := \{v \in \Omega : \|G(\bar{x}, v)\| < k_i\}.$$

By choosing the k_i's large enough we arrange that the set $\tilde{\Omega}_i$ is non-empty for each i. Corresponding to each point $s \in S_i$ we can choose $u_s \in \tilde{\Omega}_i$ such that

$$s \cdot G(\bar{x}, u_s) \leqslant s \cdot G(\bar{x}, v) + k^{-1} \quad \text{for all} \quad v \in \tilde{\Omega}_i.$$

We now define

$$\Omega_i := \{u_s : s \in S_i\}.$$

For each i let ε_i be a positive number such that $\bar{x} + 2\varepsilon_i B$ lies in the

neighbourhoods about $\bar{x}$ of hypotheses H2–H4 (with reference to the finite subset $\Omega_i \subset \Omega$). Again with reference to Ω_i, we write α_i in place of the positive number α of hypothesis H4.
Now fix i and label as $\{u_1, u_2, \ldots, u_l\}$ the elements in Ω_i. Consider the optimization problem (Q)

$$\text{Minimize } \{\tilde{F}(\tilde{x}) : \tilde{G}(\tilde{x}) = 0, \tilde{x} \in \tilde{C}\},$$

in which

$$\begin{aligned}
&\tilde{x} = (x, (\beta_1, \ldots, \beta_l)) \\
&\vdots \\
&\tilde{F}(x, (\beta_1, \ldots, \beta_l)) = F(x) \\
&\tilde{G}(x, (\beta_1, \ldots, \beta_l)) = G(x, \bar{u}) + \alpha_i \left[\sum_j \beta_j (G(x, u_j) - G(x, \bar{u}))\right]
\end{aligned}$$

and

$$\tilde{C} = (C \cap \tilde{x} + \varepsilon_i B) \times P^l.$$

(Q) is a particularly simple instance of problem (P). If $\tilde{x} = (x, (\beta_1, \ldots, \beta_l)) \in \tilde{C}$ and $\tilde{G}(\tilde{x}) = 0$ then $x \in C$ and there exists $u \in \Omega$ such that $G(x, u) = 0$. (This follows from hypothesis H4). Clearly then, $(\bar{x}, 0, 0, \ldots, 0)$ solves problem (Q). By appealing to well known optimality conditions in the mathematical programming literature (see, e.g., [3]), or Lemma 6.1 (the hypotheses involved are satisfied) we conclude existence of a point (λ_i, μ_i) such that $\lambda_i \geqslant 0$ and $\|(\lambda_i, \mu_i)\| = 1$, with the following properties:

$$0 \in \partial_x [\lambda_i F(\bar{x}) + \mu_i \cdot G(\bar{x}, \bar{u})] + N_C(\bar{x}), \tag{6.3}$$

and

$$0 \in \alpha_i (\{\mu \cdot [G(\bar{x}, u_j) - G(\bar{x}, \bar{u})]\}_{j=1}^{l}) + N_{Pl}(0).$$

But $N_{Pl}(0)$ is just the convex set, the negative orthant. Since the definition of normal cones which we adopt is consistent with the standard definition in the theory of convex sets, we deduce from this last inclusion that

$$\mu \cdot G(\bar{x}, \bar{u}) \leqslant \mu \cdot G(\bar{x}, u) \quad \text{for all} \quad u \in \Omega_i.$$

We arrange by extraction of subsequences that

$$\lambda_i \to \lambda, \quad \mu_i \to \mu, \quad \text{as} \quad i \to \infty, \tag{6.4}$$

for some points λ and μ. Clearly $\lambda \geqslant 0$ and $\|(\lambda, \mu)\| = 1$.
Let $u \in \Omega$ be an arbitrary vector. For all values of i sufficiently large, $u \in \tilde{\Omega}_i$ and there exists some $s \in S_i$ such that

$$\|s - \mu_i\| \leqslant k_i^{-2},$$

$$s \cdot G(\bar{x}, u_s) \leqslant s \cdot G(\bar{x}, u) + k_i^{-1},$$

and

$$\max \{\|G(\bar{x}, u_s)\|, \|G(\bar{x}, u)\|\} \leqslant k_i.$$

Consequently

$$\mu_i \cdot G(\bar{x}, u) \geqslant -\|s-\mu_i\| \, \|G(\bar{x}, u)\| + s \cdot G(\bar{x}, u) \geqslant$$
$$\geqslant -k_i^{-1} - k_i^{-1} + s \cdot G(\bar{x}, u_s) \geqslant -2k_i^{-1} - k_i^{-2} k_i + \mu_i \cdot G(\bar{x}, u_s).$$

Since $u_s \in \Omega_i$, we deduce from this last inequality and (6.4) that

$$\mu_i \cdot G(\bar{x}, \bar{u}) \leqslant \mu_i \cdot G(\bar{x}, u) + 3k_i^{-1}, \quad \text{for} \quad i = 1, 2, \ldots. \tag{6.5}$$

We now pass to the limit as $i \to \infty$ in (6.3) and (6.5). There results

$$0 \in \partial_x L(\bar{x}, \bar{u}, \lambda, \mu) + N_C(\bar{x}),$$

and

$$\mu \cdot G(\bar{x}, \bar{u}) \leqslant \mu \cdot G(\bar{x}, u).$$

Bearing in mind that u is an arbitrary element in Ω, we see that proof of the theorem is complete. Theorem 3.2 is proved by bringing together the characterization of the normal cone in the terms of proximal normals (2.2), Lemma 6.1 which relates proximal normals to multiplier sets and the following proposition due to Rockafellar [12, Prop. 15].

PROPOSITION 6.3. Let D and D^0 be closed subsets of R^K and suppose that D^0 is a cone which contains the origin and the recession cone of D. Define the closed cone

$$N := \{\lambda(-1, \mu) : \lambda > 0, \mu \in D\} \cup \{(0, \mu) : \mu \in D^0\},$$

in R^{k+1}. Then

$$\{\mu : (-1, \mu) \in \overline{\text{co}}\, N\} = \overline{\text{co}}\,(D + D^0). \tag{6.6}$$

If the cone D^0 is pointed, then the identity (6.6) is valid when the closure operation is dropped from the right hand side and furthermore

$$\{\mu : (0, \mu) \in \overline{\text{co}}\, N\} = \text{co}\, D^0.$$

Proof of Theorem 3.2. Take the sets D and D^0 of Proposition 6.3 to be $M^1(Y) \cap \partial V(0)$ and $M^0(Y) \cap \partial V(0)$ respectively. It is a straightforward matter to check that D is a closed set and D^0 is a closed cone containing the origin, as is required for application of the proposition.

We shall show shortly that

$$N_{\text{epi}\, V}(V(0), 0) = \overline{\text{co}}\, N, \tag{6.7}$$

where

$$N = \{\lambda(-1, \mu) : \lambda > 0, \mu \in D\} \cup \{(0, \mu) : \mu \in D^0\}.$$

We conclude from this identity, Proposition 6.3 and the definition of $\partial V(0)$ and $\partial^\infty V(0)$ that

$$\partial V(0) = \{\mu : (-1, \mu) \in \overline{\text{co}}\, N\} = \overline{\text{co}}\,(D + D^0),$$

and, in the event D^0 is pointed,

$$\partial V(0) = \text{co}\,(D + D^0) \quad \text{and} \quad \partial^\infty V(0) = \text{co}\,(D^0).$$

This will be recognized as the properties we set out to prove. It remains then to verify (6.7). The inclusion

$$N_{\text{epi}\,V}\,(V(0), 0) \supset \overline{\text{co}}\,N,$$

follows immediately from identity (2.4). To prove the reverse inclusion, we take an arbitrary limiting proximal normal (λ, μ) to epi V at $(V(0), 0)$. If $\lambda \neq 0$, it follows from definition (2.3) and Lemma 6.1 that $(1/\lambda)$ $\mu \in \partial V(0) \cup M^1(Y)$. If on the other hand $\lambda = 0$, definition (2.4) and Lemma 6.1 give $\mu \in \partial^\infty V(0) \cup M^0(Y)$. We conclude that

$$P \subset \{\lambda(-1, \mu) : \lambda > 0, \mu \in Y\} \cup \{(0, \mu) : \mu \in D^0\},$$

where P is the set of limiting proximal normal to epi V at $(V(0), 0)$. Examining the closed convex cones generalized by the sets on either side of this inclusion and using (2.2), we obtain

$$N_{\text{epi}\,V}\,(V(0), 0) = \overline{\text{co}}\,P \subset \overline{\text{co}}\,N.$$

References

[1] Aubin J. P., Clarke F. H. Shadow Prices and Duality for a Class of Optimal Control Problems. *SIAM J. Control Optim.*, **17** (1979), 567–587.

[2] Boltyanskii V. G. Optimal'noe Upravlenie Diskretnymi Sistemami. Moscow, Nauka, 1973.

[3] Clarke F. H. Optimization and Nonsmooth Analysis. New York, Wiley, 1983.

[4] Gauvin J. The Generalized Gradient of a Marginal Function in Mathematical Programming. *Math. Oper. Res.*, **4** (1979), 458–463.

[5] Halkin H. A Maximum Principle of the Pontryagin Type for Systems Described by Nonlinear Difference Equations. *SIAM J. Control*, **4** (1966), 90–111.

[6] Holtzman J. M. Convexity and the Maximum Principle for Discrete Systems. *I.E.E.E. Trans. Automatic Control*, **AC**-11 (1966), 30–35.

[7] Ioffe A. D. Necessary Conditions in Nonsmooth Optimization. *Math. Oper. Res.* **9** (1984), 159–189.

[8] Ioffe A. D., Tihomirov V. M. Theory of Extremal Problems. Moscow, Nauka, 1974, English trans. North Holland, Amsterdam, 1979.

[9] Clarke F. H., Loewen P. D. The Value Function in Optimal Control, Sensitivity, Controllability and Time Optimality. *SIAM J. Control Optim.*, **24** (1986), 243–263.

[10] Clarke F. H. Loewen P. D. State Constraints in Optimal Control: A Case Study in Proximal Normal Analysis. *SIAM J. Control Optim.*, to appear.

[11] Neustadt L. W. Optimization. A Theory of Necessary Conditions. Princeton, Princeton University Press, 1976.

[12] Rockefellar R. T. Lagrange Multipliers and Subderivatives of Optimal Value Functions in Nonlinear Programming. *Math. Programming Stud.*, **17** (1982), 28–66.

Received, March 1987.

Optymalność i wrażliwość zadań sterowania z czasem dyskretnym

Zasada maksimum Pontriagina dla zadań sterowania z czasem dyskretnym jest znana od wielu lat. W pracy przedstawiono nowy dowód tej zasady przy słabszych założeniach: funkcje występujące w sformułowaniu zadania mogą być tylko lipschitzowsko ciągłe względem zmiennych stanu, a kierunkową wypukłość zbioru rozszerzonych stanów, wymaganą we wcześniejszych dowodach, zamieniono warunkiem słabszym. Prezentowane podejście polega na zastosowaniu przybliżonych wektorów normalnych do epigrafu funkcji Bellmana. Dodatkowym wynikiem rozważań są współczynniki wrażliwości minimalnego kosztu względem parametrów zadania.

Оптимальность и чувствительность дискретных задач управления

Принцип максимума Понтрягина для дискретных задач известен уже много лет. В работе представлено новое доказательство этого принципа при более слабых предпосылках: функции участвующие при формулировке задачи могут быть лишь по Липшицу непрерывны по отношению к состояниям, а существенная выпуклость множества расширенных состояний, требуемая в более ранних доказательствах, заменяется более слабым условием. Представленный подход состоит в использовании векторов приближенно нормальных к эпиграфу функции Беллмана. Дополнительным результатом являлись коэффициенты чувствительности минимальных затрат по отношению к параметрам задачи.