SIAM J. CONTROL OPTIM.
Vol. 44, No. 3, pp. 939–968

MINIMAX OPTIMAL CONTROL*

R. B. VINTER†

Abstract. This paper provides a framework for deriving necessary conditions, in the form of a maximum principle, for minimax optimal control problems. The distinguishing feature of these problems is that the data depends on a vector α of unknown parameters, and "optimality" is defined on a worst case basis, as α ranges over the parameter set $\mathcal{A}$. The centerpiece, a minimax maximum principle, is a set of optimality conditions for such problems. Here, the parameter set $\mathcal{A}$ is taken to be an arbitrary compact metric space and the hypotheses imposed on the dynamics and endpoint constraints are of an unrestrictive nature. The minimax maximum principle captures as special cases necessary conditions for optimal control problems with minimax costs, for problems involving "semi-infinite" endpoint constraints, and also a maximum principle for state constrained optimal control problems.

Key words. optimal control, minimax problems, nonsmooth analysis, robust control

AMS subject classifications. Primary, 49L20, 49N25; Secondary, 34A37, 49L99

DOI. 10.1137/S0363012902415244

1. Introduction. The purpose of this paper is to derive, in a unified fashion, necessary conditions of optimality for optimal control problems involving an unknown vector parameter. In these problems, "optimality" is typically defined in terms of worst case performance, i.e., the cost of a particular control strategy is that associated with the strategy and a system response corresponding to the least favorable value of the unknown parameter, and constraints are required to be satisfied for all values of the unknown parameter.

Fix a compact metric space $(\mathcal{A}, \rho_{\mathcal{A}}(.,.))$. Take functions $f : [0,1]\times R^n \times R^m \times \mathcal{A} \to R^n$ and $g : R^n \times \mathcal{A} \to R$, a vector $x_0 \in R^n$, a time dependent set $\Omega(t) \subset R^m$, $0 \le t \le 1$, and a family of closed sets $\{C(\alpha) \subset R^n \mid \alpha \in \mathcal{A}\}$.

A *control function* is a measurable function $u : [0,1] \to R^m$ satisfying $u(t) \in \Omega(t)$ *a.e.* The set of control functions is written $\mathcal{U}$. A *process* $(u, \{x(.;\alpha) \mid \alpha \in \mathcal{A}\})$ comprises a control function u and a family $\{x(.;\alpha) \in W^{1,1}([0,1];R^n) \mid \alpha \in \mathcal{A}\}$ of arcs satisfying, for each $\alpha \in \mathcal{A}$,

$$\begin{cases} \dot{x}(t;\alpha) = f(t, x(t;\alpha), u(t), \alpha) \quad a.e. \\ x(0;\alpha) = x_0. \end{cases}$$

The process is termed *feasible* if the $x(.;\alpha)$s satisfy the terminal constraints

$$x(1;\alpha) \in C(\alpha) \quad \text{for all } \alpha \in \mathcal{A}.$$

The optimization problem of interest in this paper, which will be referred to as the

*Received by the editors September 27, 2002; accepted for publication (in revised form) January 4, 2005; published electronically September 15, 2005.
http://www.siam.org/journals/sicon/44-3/41524.html
†Department of Electrical and Electronic Engineering, Imperial College of Science Technology and Medicine, Exhibition Road, London SW7 2BT, UK (r.vinter@imperial.ac.uk).

general minimax optimal control problem, is as follows:

$$
(\mathrm{P}) \begin{cases} \text{Minimize } \max_{\alpha\in\mathcal{A}} g(x(1;\alpha),\alpha) \\ \qquad \text{over measurable functions } u:[0,1]\to R^m \text{ such that} \\ u(t)\in\Omega(t) \quad a.e.\ t\in[0,1] \\ \qquad \text{and arcs } \{x(.;\alpha):[0,1]\to R^n \mid \alpha\in\mathcal{A}\} \text{ such that, for each } \alpha\in\mathcal{A}, \\ \dot{x}(t;\alpha)=f(t,x(t;\alpha),u(t),\alpha) \quad a.e.\ t\in[0,1], \\ x(0;\alpha)=x_0 \quad \text{ and } \quad x(1;\alpha)\in C(\alpha). \end{cases}
$$

Briefly stated, the problem is to minimize $\sup_{\alpha\in\mathcal{A}} g(x(1;\alpha),\alpha)$ over feasible processes $(u,\{x(.;\alpha)\mid\alpha\in\mathcal{A}\})$.

A feasible process $(\bar{u},\{\bar{x}(.;\alpha)\mid\alpha\in\mathcal{A}\})$ is said to be a strong local minimizer when there exists $\epsilon>0$ such that

$$
\sup_{\alpha\in\mathcal{A}} g(x(1;\alpha),\alpha) \;\geq\; \sup_{\alpha\in\mathcal{A}} g(\bar{x}(1;\alpha),\alpha)
$$

for all feasible processes $(u,\{x(.;\alpha),\alpha\in\mathcal{A}\})$ such that

$$
\|x(.;\alpha)-\bar{x}(.;\alpha)\|_C\le\epsilon \quad \text{for all } \alpha\in\mathcal{A}\,.
$$

The implications of our necessary conditions for various special cases of interest will also be investigated.

Our framework permits the set $\mathcal{A}$ of unknown parameter values to be an arbitrary compact metric space. It therefore covers minimax optimal control problems in which components of α comprise unknown gain values lying within specified bounds, magnitudes of step disturbances, etc., important cases that would be excluded by the requirement that $\mathcal{A}$ be a finite set.

The presence of, possibly, an infinite number of elements in $\mathcal{A}$ is the principal source of difficulty in the derivation of necessary conditions for minimax optimal control problems. In case $\mathcal{A}$ is a finite set $\{\alpha_1,\alpha_2,\dots,\alpha_N\}$, the minimax optimal control problems studied here can be reformulated as standard optimal control problems, for which necessary conditions are already known. (See section 2.)

We comment on related earlier research. The most extensively studied minimax optimal control problems are zero sum differential games, in which a minimizer is chosen from a class of closed loop controls, appropriately defined, and the parameter set $\mathcal{A}$, from which a "worst case" element is selected, comprises open loop control functions of an opposing player [1], [4]. The fact that differential games are posed over closed loop controls gives them a quite different character to the problems studied here, in which the choice variables are open loop controls. Analysis of solutions to differential games is almost exclusively of a global nature, centering on the relationship between the value of the differential game and the solutions to the Hamilton–Jacobi equation; variants on the pontryagin maximum principle, such as featured in this paper, have a limited role in the analysis of optimal feedback strategies.

Versions of the open loop minimax optimal control problem were previously investigated by Warga, in the context of "relaxed and hyper-relaxed adverse controls." Warga adopts a broader framework than ours, in which the parameter set can include open control functions of an opposing player as well as finite dimensional vector parameters. Furthermore, he addresses questions of existence of solutions to minimax optimal control problems and appropriate relaxation schemes as well as local optimality conditions. Our minimax maximum principle, involving a Hamiltonian averaged with respect to some measure, is implicit in the necessary conditions in ([11], Chapters IX and X). Warga's necessary conditions apply only in cases when the endpoint constraint sets are closed, convex sets with nonempty interiors and for smooth dynamics. The necessary conditions of this paper are proved by quite different methods and under significantly weaker hypotheses (for the minimax problems here considered). Furthermore, we give new insights into the limits of validity of the kinds of necessary conditions investigated here, by presenting some counterexamples where they no longer apply. Optimality conditions akin to those of section 2 below are featured also in [2], but only in the elementary case when the parameter set is a finite set and the endpoint constraint is specified by a functional inequality. The role of measures to estimate "gradients" of max functions is evident in the early Russian optimal control literature [5] and is widely exploited in nonsmooth analysis, for example, in applications of nonsmooth analysis to derive optimality conditions for state constrained optimal control problems [3].

Another point of contact with earlier work is semi-infinite programming. This is a branch of nonlinear programming that aims to provide efficient computational methods for optimization problems, in which constraints must be satisfied for a continuum of values of some parameter α. (See [8].) Minimax optimal control problems can be reformulated, by introduction of additional variables, as semi-infinite programming problems over function spaces with dynamic constraints.

One possible approach to the computation of solutions to a minimax optimal control problem is to approximate it by a (finite-dimensional) semi-infinite programming problem by means of time discretization and to apply semi-infinite programming algorithms. The emphasis in this paper is on structural properties of solutions to minimax optimal control problems. But the necessary conditions of optimality we provide may ultimately find application in convergence analysis of algorithms for minimax optimal control, based on semi-infinite programming or other approaches.

We allow nonsmooth data and express necessary conditions in terms of "limiting subdifferentials" and other constructs of nonsmooth analysis. We stress, however, that it is the unrestrictive nature of the conditions that we place on the parameter set $\mathcal{A}$, "$\mathcal{A}$ is an arbitrary compact metric space," which is the most significant feature of our analysis. The main optimality conditions supplied here (the maximum principle for the general minimax optimal control problem of section 3 and the implications explored in section 5) are new, even when specialized to the smooth case.

Finally, some notation. Throughout, $|\,.\,|$ denotes the Euclidean norm. We write B for the closed unit ball in Euclidean space. $B_{\mathcal{A}}(\alpha,\epsilon)$ denotes the set $\{\alpha' \in \mathcal{A} \mid \rho_{\mathcal{A}}(\alpha,\alpha') \le \epsilon\}$.

$W^{1,1}([0,1];R^n)$ is the space of absolutely continuous R^n-valued functions on $[0,1]$. Take a compact metric space A. $C(A)$ denotes the space of continuous real valued functions on A. We write $||.||_C$ for the supremum norm on this space. $C^*(A)$ denotes the topological dual of $C(A)$ with the norm topology. We use the fact that elements in $C^*(A)$ can be identified with the space of Radon measures on the Borel subsets of A. The dual norm of an element $\mu \in C^*(A)$ is written $||\mu||_{T.V.}$, a choice of notation that reflects the fact that the dual norm of μ coincides with the total variation of the Radon measure that represents μ.

The graph of a multifunction $D : A \rightsquigarrow R^k$ is denoted by $Gr\,D$,

$$Gr\,D \;:=\; \{(a,d) \in A \times R^k \mid d \in D(a)\}\,.$$

For a given set $E \subset R^d$, $d_E(.)$ denotes the Euclidean distance function

$$d_E(z) \;:=\; inf_{e\in E}|z-e|.$$

The *limiting normal cone* to a given closed set $C \subset R^k$ at $x \in R^k$ is the set

$$N_C(x) \;:=\; \big\{\xi \in R^n \mid \exists\, \xi_i \to \xi, x_i \xrightarrow{C} x \text{ and } \{M_i\} \subset R^+ \\ \text{such that, for each } i,\ \xi_i \cdot (x - x_i) \leq M_i |x - x_i|^2 \;\forall\, x \in C\big\}.$$

Here "$x_i \xrightarrow{C} x$" means "$x_i \to x$ and $x_i \in C$ for all i." Note that $N_C(x) = \emptyset$, in the case $x \notin C$.

Take a function $f : R^n \to R \cup \{+\infty\}$ and a point $x \in \mathit{dom}\, f$. Here, $\mathit{dom}\, f$ is taken to be the set

$$\mathit{dom}\, f = \{y \in R^n \mid f(y) < +\infty\}\,.$$

The epigraph set of f is the set

$$\mathit{epi}\, f := \{(x, \alpha) \in R^n \times R \mid \alpha \geq f(x)\}.$$

The *limiting subdifferential* $\partial f(x)$ of $f : R^n \to R \cup \{+\infty\}$ at a point $x \in \mathit{dom}\, f$ is the set

$$\partial f(x) := \big\{\eta \mid (\eta, -1) \in N_{\mathrm{epi} f}(x, f(x))\big\}.$$

The partial limiting subdifferential $\partial_x f(x, y)$ of an extended valued function f of two variables x and y is the limiting subdifferential of $x \to f(x, y)$ for fixed y.

$N_C(x)$ and $\partial f(x)$ are widely used constructs from nonsmooth analysis in optimal control, that generalize classical notions of the set of outward normal vectors to a set with smooth boundary and of the gradient of a continuously differentiable function. They are also referred to as the normal cone and the subdifferential, respectively. For a review of their properties (and historical comments), see, for example, [7], [9], [10].

2. The finite parameter set case.

Necessary conditions for minimax problems involving an arbitrary compact metric space parameter set $\mathcal{A}$ will be derived by approximating $\mathcal{A}$ by a finite set $\mathcal{A}_N$, by establishing properties of approximate minimizers for problems involving A_N, and passage to the limit. Necessary conditions for problems with finite parameter sets have an important intermediate role then in the proof of more general necessary conditions. This is one reason for attending to the finite parameter set case at this early stage. But studying this special case also gives insights into the necessary conditions we should expect to be valid in more general circumstances.

We shall invoke the following hypotheses on the data for the general minimax optimal control problem, in which $(\bar{u}, \{\bar{x}(.;\alpha) \mid \alpha \in \mathcal{A}\})$ is the strong local minimizer under consideration. For some $\delta > 0$,

(H1) The function $f(., x, ., \alpha)$ is $\mathcal{L} \times \mathcal{B}^m$ measurable for each $(x, \alpha) \in R^n \times \mathcal{A}$. ($\mathcal{L}$ denotes the Lebesgue subsets of $[0, 1]$ and $\mathcal{B}^m$ denotes the Borel subsets of R^m.) $t \rightsquigarrow \Omega(t)$ has a Borel measurable graph.

(H2) There exists a Borel measurable function $k_f : [0, 1] \times R^m$ such that $t \to k_f(t, \bar{u}(t))$ is integrable and, for each $\alpha \in \mathcal{A}$,

$$|f(t, x, u, \alpha) - f(t, x', u, \alpha)| \leq k_f(t, u)|x - x'|$$

for all $x, x' \in \bar{x}(t; \alpha) + \delta B$, $u \in \Omega(t)$, $a.e.\ t \in [0, 1]$.

(H3) The function $g(.,\alpha)$ is Lipschitz continuous on $\bar{x}(1;\alpha)+\delta B$ for all $\alpha \in \mathcal{A}$.
Define the Hamiltonian

$$H(t,x,p,u,\alpha) := p \cdot f(t,x,u,\alpha).$$

PROPOSITION 2.1. *Let* $(\bar{u},\{\bar{x}(.;\alpha) \mid \alpha \in \mathcal{A}\})$ *be a strong local minimizer for the general minimax optimal control problem* (P). *Assume that* $\mathcal{A}$ *is a finite set and that, for some* $\delta > 0$, *hypotheses* (H1)–(H3) *are satisfied.*

Then

$$\int H(t,\bar{x}(t;\alpha),\bar{u}(t),p(t;\alpha),\alpha)\ \Lambda(d\alpha)$$
$$= \max_{u\in\Omega(t)} \int H(t,\bar{x}(t;\alpha),u,p(t;\alpha),\alpha)\ \Lambda(d\alpha). \qquad a.e.\ t \in [0,1],$$

for some Radon probability measure $\Lambda \in C^*(\mathcal{A})$ *and some family of arcs* $\{p(.;\alpha) \in W^{1,1}([0,1];R^n) \mid \alpha \in \mathcal{A}\}$ *such that, for* Λ – *a.e.* $\alpha \in \mathcal{A}$,

$$\begin{aligned} -\dot{p}(t;\alpha) &\in co\,\partial_x H(t,\bar{x}(t;\alpha),\bar{u}(t),p(t;\alpha),\alpha) \quad a.e., \\ -p(1;\alpha) &\in \bigcup_{0\le r\le 1} \{rG_0(\bar{x}(1;\alpha),\alpha)+(1-r)N(\bar{x}(1;\alpha),\alpha)\} \end{aligned} \tag{2.1}$$

and

$$supp\,\Lambda \subset \{\alpha \mid either\ G_0(\bar{x}(1;\alpha),\alpha) \neq \emptyset\ or\ N(\bar{x}(1;\alpha),\alpha) \neq \emptyset\}.$$

Here,

$$G_0(x,\alpha) := \begin{cases} \partial_x g(x,\alpha) & \text{if } g(x,\alpha) = max_{\alpha'\in\mathcal{A}}\ g(x,\alpha') \\ \emptyset & \text{otherwise} \end{cases}$$

and

$$N(x,\alpha) := \{\xi \in N_{C(\alpha)}(x) \mid |\xi| = 1\}.$$

In condition (2.1), we allow the possibilities that (for some values of α) $G_0(x,\alpha) = \emptyset$ or $N(x,\alpha) = \emptyset$. If $G_0(x,\alpha) = \emptyset$, then $rG_0(x,\alpha)$ is defined only if $r = 0$; in this case $rG_0(x,\alpha) := \{0\}$. If $N(x,\alpha) = \emptyset$, then $(1-r)N(x,\alpha)$ is defined only if $r = 1$; in this case $(1-r)N(x,\alpha) := \{0\}$. Thus (2.1) implies that if $\Lambda(\{\alpha\}) > 0$, then the parameter α is "active" in either the endpoint constraint or in the objective, in the sense that

$$g(\bar{x}(1,\alpha),\alpha) = \max_{\alpha'\in\mathcal{A}} g(\bar{x}(1,\alpha'),\alpha') \quad \text{or} \quad \bar{x}(1:\alpha) \in bdy\,C(\alpha)\,.$$

($bdy\,C(\alpha)$ denotes the "boundary of the set $C(\alpha)$.")

Proof of Proposition 2.1. List the elements in the finite set $\mathcal{A}$ as

$$\mathcal{A} = \{\alpha_1,\alpha_2,\ldots,\alpha_N\}.$$

Denote by $\bar{x} = col\,\{\bar{x}(.;\alpha_1),\bar{x}(.;\alpha_2),\ldots,\bar{x}(.;\alpha_N)\}$ the collection of state trajectories corresponding to $\bar{u}$. Then $(\bar{u},\bar{x})$ is a strong local minimizer for the standard optimal control problem

$$(\tilde{\mathrm{P}}) \begin{cases} \text{Minimize } \tilde{g}(x(1)) \text{ over } u(.) \text{ satisfying} \\ \dot{x}(t) = \tilde{f}(t,x(t),u(t)) \quad a.e.\ t \in [0,1], \\ x(0) = \tilde{x}_0, \\ x(1) \in \tilde{C}, \\ u(t) \in \Omega(t) \quad a.e.\ t \in [0,1]\,, \end{cases}$$

in which the $N \times n$ dimensional state vector is partitioned as

$$\begin{aligned} x &= col\ \{x_1, x_2, \dots, x_N\}, \\ \tilde{f}(t,x,u) &= col\ \{f(t,x_i,u,\alpha_i)\}_{i=1}^{N}, \\ \tilde{C} &= C(\alpha_1) \times C(\alpha_2) \times \cdots \times C(\alpha_N), \\ \tilde{x}_0 &= col\, \{x_0, x_0, \dots, x_0\}, \\ \tilde{g}(x) &= \max_i\ g(x(.;\alpha_i),\alpha_i). \end{aligned}$$

Under the stated hypotheses, we deduce from the nonsmooth maximum principle (see, for example, [10], Theorem 6.2.1)), with the help of the max rule ([10], Theorem 5.5.2) to evaluate the limiting subdifferential of the cost function $\tilde{g}$, the following information. There exist numbers $\lambda_1, \dots, \lambda_N \geq 0$, arcs $q(.;\alpha_i) \in W^{1,1}$, and elements $\xi_i \in N_{C(\alpha_i)}(x(1;\alpha_i))$, $i = 1, 2, \dots, N$, such that

$$\text{(i)} \sum_{i=1}^{N} H(t, \bar{x}(t;\alpha_i), \bar{u}(t), q(t;\alpha_i), \alpha_i) = \max_{u \in \Omega(t)}\ \sum_{i=1}^{N} H(t, \bar{x}(t;\alpha_i), u, q(t;\alpha_i), \alpha_i) \quad a.e.$$

$$\text{(ii)} \sum_{i=1}^{N} (\lambda_i + |\xi_i|)\ =\ 1$$

and, for each i,

$$\begin{aligned} &\text{(iii)} -\dot{q}(t;\alpha_i)\ \in\ \mathrm{co}\, \partial_x H(t, \bar{x}(t;\alpha_i), \bar{u}(t), q(t;\alpha_i), \alpha_i) \quad a.e., \\ &\text{(iv)} -q(1;\alpha_i)\ \in\ \lambda_i \partial_x g(\bar{x}(t;\alpha_i), \alpha_i) + \xi_i, \\ &\text{(v)} \lambda_i = 0 \quad \text{if} \quad g(\bar{x}(1;\alpha_i),\alpha_i) < \max_j\ g(\bar{x}(1;\alpha_j),\alpha_j). \end{aligned}$$

Define Λ to be the discrete probability measure

$$\Lambda = \sum_{i=1}^{N} (\lambda_i + |\xi_i|)\ \delta_{\alpha_i},$$

in which δ_{α_i} denotes the unit measure concentrated at $\alpha = \alpha_i$. If $\alpha \in supp\ \{\Lambda\}$, in which case $\alpha = \alpha_i$ for some i such that $(\lambda_i + |\xi_i|) > 0$, define

$$p(t;\alpha_i) = \frac{1}{\lambda_i + |\xi_i|}\ q(t;\alpha_i).$$

If $\alpha \notin supp\ \{\Lambda\}$, choose the $W^{1,1}$ function $p(.;\alpha)$ arbitrarily.
All the assertions of the proposition can be confirmed for this choice of Λ and $\{p(.;\alpha) \mid \alpha \in \mathcal{A}\}$.

Note, in particular, that, if $\alpha_i \in supp\ \{\Lambda\}$, then

$$-p(1;\alpha_i) \in r_i \partial_x g(\bar{x}(1;\alpha_i),\alpha_i) + (1 - r_i)\{\xi \in N_{C(\alpha_i)}(\bar{x}(1;\alpha_i)) \mid |\xi| = 1\}\,.$$

Here, r_i, $0 \leq r_i \leq 1$, is the number

$$r_i = \frac{\lambda_i}{\lambda_i + |\xi_i|}.$$

We also observe that, for each $t \in [0,1]$ and $u \in \Omega(t)$,

$$\sum_{i=1}^{N} H(t, \bar{x}(t;\alpha_i), u, q(t;\alpha_i), \alpha_i) = \int_{\mathcal{A}} H(t, \bar{x}(t;\alpha), u, p(t;\alpha), \alpha)\ \Lambda(d\alpha),$$

i.e., the maximization of the "averaged" Hamiltonian condition is satisfied. Finally, note that for $\Lambda - a.e.\ \alpha \in \mathcal{A}$,

$$-\dot{p}(t;\alpha)\ \in\ \text{co}\,\partial_x H(t, \bar{x}(t;\alpha), \bar{u}(t), p(t;\alpha), \alpha),$$

by positive homogeneity. □

3. A maximum principle for the general minimax optimal control problem. This section provides necessary conditions of optimality for the general minimax optimal control problem (P) of section 1, when the parameter set $\mathcal{A}$ is an arbitrary compact metric space.

Take $(\bar{u}, \{\bar{x}(.;\alpha)\,|\,\alpha \in \mathcal{A}\})$ to be the local minimizer for problem (P) of interest. For $\alpha \in \mathcal{A}$, define the set

$$Q_0(\alpha) := \big\{p(.;\alpha) \in W^{1,1}\ |\ -\dot{p}(t;\alpha)\ \in\ \text{co}\,\partial_x H(t, \bar{x}(t;\alpha), \bar{u}(t), p(t;\alpha), \alpha) \quad a.e.$$
$$\text{and } -p(1;\alpha) \in \cup_{r\in[0,1]}\,(rG_0(\bar{x}(1;\alpha),\alpha) + (1-r)N(\bar{x}(1;\alpha),\alpha)),$$

in which, for $\epsilon \in [0,1]$,

$$G_\epsilon(x,\alpha) := \begin{cases} \partial_x g(x,\alpha) & \text{if } g(x,\alpha) \geq \max_{\alpha'\in\mathcal{A}}\ g(x,\alpha')\ -\ \epsilon \\ \emptyset & \text{otherwise} \end{cases} \tag{3.1}$$

and

$$N(x,\alpha) := \big\{\xi \in N_{C(\alpha)}(x)\ |\ |\xi| = 1\big\}. \tag{3.2}$$

(Only $G_{\epsilon=0}(x,0)$ is involved in the definition of $Q_0(\alpha)$. $G_\epsilon(x,\alpha)$, with $\epsilon > 0$, is required for later analysis.)

The assertions of Proposition 2.1 can be expressed in terms of the set $Q_0(\alpha)$ as follows. If $(\bar{u}, \{\bar{x}(.;\alpha)\ |\ \alpha \in \mathcal{A}\})$ is a strong local minimizer and $\mathcal{A}$ is a finite set, then

$$\int_{\mathcal{A}} H(t, \bar{x}(t;\alpha), \bar{u}(t), p(t;\alpha), \alpha)\ \Lambda(d\alpha) =$$
$$\max_{u\in\Omega(t)} \int_{\mathcal{A}} H(t, \bar{x}(t;\alpha), u, p(t;\alpha), \alpha)\ \Lambda(d\alpha) \quad a.e.\ t \in [0,1]$$

for some Radon probability measure $\Lambda \in C^*(\mathcal{A})$ and family of arcs $\{p(.;\alpha)\ |\ \alpha \in \mathcal{A}\}$ such that

$$p(.;\alpha) \in Q_0(\alpha)\ \text{ for } \Lambda - a.e.\ \alpha \in \mathcal{A}.$$

(Note that $Q_0(\alpha)$ may be empty unless α is "active" in the sense of our earlier remarks.) Unfortunately, the above optimality condition no longer remains valid in general, when we allow $\mathcal{A}$ to be an arbitrary compact metric space. Confirmation is provided by the counter examples of section 5. Indeed, standard variational techniques break down when $\mathcal{A}$ is an infinite set, because the multifunction

$$Q_0(.)\ :\ \mathcal{A} \to \{\text{subsets of } W^{1,1}\}$$

may lack the requisite convexity and closure properties for limit taking. To derive necessary conditions in this more general context, we need to replace $Q_0(\alpha)$ with a larger set, better matched to the limit taking operations involved.

Let $(\bar{u}, \{\bar{x}(.;\alpha) \mid \alpha \in \mathcal{A}\})$ be the process of interest. We embed $Q_0(.)$ in a family of multifunctions $\{Q_\epsilon(.)|\epsilon \geq 0\}$ defined as follows. For any $\epsilon \geq 0$ and $\alpha \in \mathcal{A}$ we define

$$Q_\epsilon(\alpha) \;:=\; \big\{p(.;\alpha) \in W^{1,1} \mid \text{conditions (a) and (b) below are satisfied}\big\}$$

in which

(a)

$$-\dot{p}(t;\alpha) \in \bigcup_{x \in \bar{x}(t;\alpha)+\epsilon B} \operatorname{co} \partial_x H(t, x, \bar{u}(t), p(t;\alpha), \alpha) \quad a.e.$$

(b)

$$-p(1;\alpha) \in \bigcup_{x \in \bar{x}(1;\alpha)+\epsilon B} \; \bigcup_{r \in [0,1]} (rG_\epsilon(x, \alpha) + (1-r)N(x, \alpha))$$

The sets $G_\epsilon(x, \alpha)$ and $N(x, \alpha)$ appearing in these conditions were defined in (3.1) and (3.2).

The defining properties of the "costate" arcs $p(.;\alpha)$ will now include the condition

$$p(.;\alpha) \;\in\; \overline{Q}_0(\alpha),$$

where

$$\overline{Q}_0(\alpha) \;:=\; \bigcap_{\epsilon>0} \overline{co} \left(\bigcup_{\alpha' \in B_{\mathcal{A}}(\alpha,\epsilon)} Q_\epsilon(\alpha') \right). \tag{3.3}$$

Here $\overline{co}$ denotes "convex closure" with respect to the strong $W^{1,1}$ topology. Note that the right side is a subset of $W^{1,1}([0,1]; R^n)$. This relationship involves a multifunction that is obtained from the multifunction $\alpha \rightsquigarrow Q_0(\alpha)$ by enlarging its graph. The enlargement is carried out in such a manner that the new multifunction has closed graph and convex values.

In certain cases, notably when the data is smooth and the right endpoint constraints are absent,

$$Q_0(\alpha) \;=\; \overline{Q}_0(\alpha).$$

But in many cases of interest, $Q_0(\alpha)$ is a strict subset of its "closed, convexified" counterpart. We discuss these points in section 5.

We now come to the main result of this paper, namely a maximum principle for the general minimax optimal control problem. Here, as usual, the Hamiltonian is

$$H(t, x, p, u, \alpha) := p \cdot f(t, x, u, \alpha).$$

The following hypotheses will be invoked, in which $(\bar{u}, \{\bar{x}(.;\alpha) \mid \alpha \in \mathcal{A}\})$ is the strong local minimizer for (P) of interest. For some $\delta > 0$,

(S1) The function $f(., x, ., .)$ is $\mathcal{L} \times \mathcal{B}^m \times \mathcal{B}_{\mathcal{A}}$ measurable for each $x \in R^n$. ($\mathcal{B}_{\mathcal{A}}$ denotes the Borel subsets of $\mathcal{A}$.) $t \rightsquigarrow \Omega(t)$ has a Borel measurable graph.

(S2) There exists $k_f \in L^1$ and $c_f > 0$ such that

$$|f(t,x,u,\alpha) - f(t,x',u,\alpha)| \leq k_f(t)|x-x'| \quad \text{and} \quad |f(t,x,u,\alpha)| \leq c_f$$

for all $x, x' \in \bar{x}(t;\alpha) + \delta B$, $u \in \Omega(t)$ and $\alpha \in \mathcal{A}$, *a.e.* $t \in [0,1]$.

(S3) g is continuous and there exists $k_g > 0$ such that

$$|g(x,\alpha) - g(x',\alpha)| \leq k_g|x-x'|$$

for all $x, x' \in \bar{x}(1;\alpha) + \delta B$ and $\alpha \in \mathcal{A}$.

(S4) There exists $\theta : [0,+\infty) \to [0,+\infty)$ such that $\lim_{s\downarrow 0}\theta(s) = 0$ and, for all $\alpha, \alpha' \in \mathcal{A}$,

$$\int_0^1 \sup_{x\in\bar{x}(t)+\delta B,\ u\in\Omega(t)} |f(t,x,u,\alpha) - f(t,x,u,\alpha')| \, dt \leq \theta(\rho_{\mathcal{A}}(\alpha,\alpha')).$$

(S5) $\alpha \to d_{C(\alpha)}(x)$ is continuous on $\mathcal{A}$ for each $x \in R^n$.

In the following theorem, $\mathcal{A}$ is an arbitrary compact metric space.

THEOREM 3.1. *Let* $(\bar{u}, \{\bar{x}(.;\alpha) \mid \alpha \in \mathcal{A}\})$ *be a strong local minimizer for* (P). *Assume that, for some* $\delta > 0$, *Hypotheses* (S1)–(S5) *are satisfied.*

Then

$$\int H(t,\bar{x}(t;\alpha),\bar{u}(t),p(t;\alpha),\alpha)\ \Lambda(d\alpha)$$
$$= \max_{u\in\Omega(t)} \int H(t,\bar{x}(t;\alpha),u,p(t;\alpha),\alpha)\ \Lambda(d\alpha) \quad a.e.\ t \in [0,1], \tag{3.4}$$

for some Radon probability measure $\Lambda \in C^*(\mathcal{A})$ *and family of arcs* $\{p(.;\alpha) \in W^{1,1} \mid \alpha \in \mathcal{A}\}$ *such that, for* Λ – *a.e.* $\alpha \in \mathcal{A}$,

$$p(.;\alpha) \ \in\ \overline{Q}_0(\alpha). \tag{3.5}$$

(Recall the definition of $\overline{Q}_0(\alpha)$ *in* (3.3).*)*

Note that the right side of (3.5) may be empty for certain values of α. The set is nonempty, however, on a set of full Λ measure.

Implicit in the optimality conditions is the assertion that the integrals in the maximization of the Hamiltonian condition (3.4) are well-defined, i.e., the function $\alpha \to H(t,\bar{x}(t;\alpha),u,p(t;\alpha),\alpha)$ is Λ-integrable for each $u \in \Omega(t)$, *a.e.* $t \in [0,1]$.

We might expect that necessary conditions of optimality are valid for a hybrid form of the minimax optimal control problem, in which the parameter set $\mathcal{A}$ separates into the union of a "discrete" and a "continuous" set, and which specializes to a version of Proposition 2.1 (valid under the stronger hypotheses of Theorem 3.1) and Theorem 3.1 in the extreme cases "$\mathcal{A}$ is purely discrete" and "$\mathcal{A}$ is purely continuous." The following theorem supplies such conditions. We explore some consequences in section 5.

THEOREM 3.2. *Let* $(\bar{u}, \{\bar{x}(.;\alpha) \mid \alpha \in \mathcal{A}\})$ *be a strong local minimizer for the general minimax optimal control problem* (P). *Assume that Hypotheses* (S1)–(S5) *are satisfied. Assume, furthermore, we can partition the compact metric space* $\mathcal{A}$ *into disjoint sets*

$$\mathcal{A} = \mathcal{A}^{(1)} \cup \mathcal{A}^{(2)},$$

in which $\mathcal{A}^{(1)}$ *is a compact metric space and* $\mathcal{A}^{(2)}$ *is a finite set.*

Then

$$\int H(t,\bar{x}(t;\alpha),\bar{u}(t),p(t;\alpha),\alpha)\ \Lambda(d\alpha) = \max_{u\in\Omega(t)} \int H(t,\bar{x}(t;\alpha),u,p(t;\alpha),\alpha)\ \Lambda(d\alpha) \quad a.e.\ t\in[0,1]$$

for some Radon probability measure $\Lambda \in C^*(\mathcal{A})$ *and family of arcs* $\{p(.;\alpha)\in W^{1,1} \mid \alpha \in \mathcal{A}\}$ *such that*

$$p(.;\alpha)\ \in\ \overline{Q}_0(\alpha) \quad \textit{for } \Lambda - a.e.\ \alpha \in \mathcal{A}^{(1)}$$

and

$$p(.;\alpha)\ \in\ Q_0(\alpha) \quad \textit{for } \Lambda - a.e.\ \alpha \in \mathcal{A}^{(2)}.$$

We conclude this section by stating a version of the foregoing theorems covering problems in which the endpoint constraints take the form of a finite collection of functional inequality constraints, namely problems for which each $C(\alpha)$ has the representation

$$C(\alpha)\ =\ \{x \in R^n \mid \psi(x,\alpha) \le 0\}, \tag{3.6}$$

for some function $\psi : R^n \times \mathcal{A} \to R^r$. The inequalities are interpreted in a "componentwise" sense. It will be assumed that, for some $\delta > 0$, ψ satisfies the following hypothesis:
(H) ψ is continuous and there exist k_ψ such that

$$|\psi(x,\alpha)-\psi(x',\alpha)| \le k_\psi |x-x'| \ \text{ for all } x,x' \in \bar{x}(1;\alpha)+\delta B,\ \alpha \in \mathcal{A}.$$

Minor modifications to the proof of Theorems 3.1 and 3.2 yield the following optimality condition for problems involving endpoint functional inequality constraints:

THEOREM 3.3. *Let* $(\bar{u},\{\bar{x}(.;\alpha) \mid \alpha \in \mathcal{A}\})$ *be a strong local minimizer for* (P). *Assume that the endpoint constraint sets* $\{C(\alpha) \mid \alpha \in \mathcal{A}\}$ *take the form of a collection of functional inequality constraints* (3.6) *which satisfy Hypothesis* (H). *Then*

$$\int H(t,\bar{x}(t;\alpha),\bar{u}(t),p(t;\alpha),\alpha)\ \Lambda(d\alpha) = \max_{u\in\Omega(t)} \int H(t,\bar{x}(t;\alpha),u,p(t;\alpha),\alpha)\ \Lambda(d\alpha) \quad a.e.\ t\in[0,1],$$

for some Radon probability measure $\Lambda \in C^*(\mathcal{A})$ *and family of arcs* $\{p(.;\alpha)\in W^{1,1} \mid \alpha \in \mathcal{A}\}$ *such that,*

(a) *if* $\mathcal{A}$ *is a finite set and Hypotheses* (H1)–(H3) *are satisfied, then*

$$p(.;\alpha)\ \in\ Q_0^{\psi}(\alpha) \quad \textit{for } \Lambda - a.e.\ \alpha \in \mathcal{A}.$$

(b) *if* $\mathcal{A}$ *is a compact metric space and Hypotheses* (S1)–(S4) *are satisfied, then*

$$p(.;\alpha) \in \bigcap_{\epsilon>0} \overline{co}\left(\bigcup_{\alpha' \in B_{\mathcal{A}}(\alpha,\epsilon)} Q_\epsilon^{\psi}(\alpha') \right) \quad \textit{for } \Lambda - a.e.\ \alpha \in \mathcal{A}.$$

(c) *if Hypotheses* (S1)–(S4) *are satisfied and we can partition* $\mathcal{A} \subset R^k$ *into disjoint sets* $\mathcal{A} = \mathcal{A}^{(1)} \cup \mathcal{A}^{(2)}$, *in which* $\mathcal{A}^{(1)}$ *is a compact metric space and* $\mathcal{A}^{(2)}$ *is a finite set, then*

$$p(.;\alpha) \in \bigcap_{\epsilon>0} \overline{co}\left(\bigcup_{\alpha' \in B_{\mathcal{A}}(\alpha,\epsilon)} Q^{\psi}_{\epsilon}(\alpha')\right) \quad \textit{for } \Lambda - a.e.\ \alpha \in \mathcal{A}^{(1)} \tag{3.7}$$

and

$$p(.;\alpha) \in Q^{\psi}_0(\alpha) \quad \textit{for } \Lambda - a.e.\ \alpha \in \mathcal{A}^{(2)}.$$

In the above optimality conditions the set $Q^{\psi}_{\epsilon}(\alpha)$, $\epsilon \geq 0$, shares the defining relationships of $Q_{\epsilon}(\alpha)$ (see (3.3)) in all respects except that the set $N(x,\alpha)$ in condition (b), namely

$$N(x,\alpha) = \{\xi \in N_{C(\alpha)}(x) \mid |\xi| = 1\},$$

is replaced by the set

$$N^{\psi}(x,\alpha) := \left\{\sum_j \lambda_j \nabla_x \psi_j(x,\alpha) \mid (\lambda_1,\ldots,\lambda_r) \in \mathcal{S}(r) \text{ such that } \lambda_i = 0 \text{ if } \psi_i(x,\alpha) < 0\right\}.$$

in which

$$\mathcal{S}(r) := \left\{\lambda \in R^r \mid \lambda_i \geq 0,\ i = 0,\ldots,r \text{ and } \sum_{i=0}^{r} \lambda_i = 1\right\}.$$

It is a straightforward matter to derive variants on Theorem 3.3. We could, for example, assume that the endpoint constraint sets $C(\alpha)$ take the form $\{x \mid \psi(x,\alpha) \leq 0\}$ for $\alpha \in \mathcal{A}^{(2)}$ and are arbitrary closed sets for $\alpha \in \mathcal{A}^{(1)}$. In this case the necessary conditions will incorporate transversality conditions from both Theorems 3.2 and 3.3.

4. Discussion. Theorem 3.1 captures only a coarse version of Proposition 2.1 when specialized to the finite set case. This is because Proposition 2.1 asserts the existence of costate arcs in the sets $Q_0(\alpha)$, $\alpha \in \mathcal{A}$, with respect to which an "averaged" maximum principle is valid. On the other hand, Theorem 3.1 asserts the existence of costates with this property, chosen from the larger sets $\overline{Q}_0(\alpha)$, obtained by convexifying the values of $\alpha \to Q_0(\alpha)$ and closing its graph, in some sense. Minimax maximum principles involving $\overline{Q}_0(\alpha)$ provide significantly less information about minimizers than those involving $Q_0(\alpha)$. For further elucidation of this point, consider the case of (P) when the endpoint constraint sets are

$$C(\alpha) = \{x \mid \psi(x,\alpha) = 0\} \quad \text{for all } \alpha \in \mathcal{A}.$$

Here $\psi : R^n \times \mathcal{A} \to R$ is a given function. Assume that, for some fixed $\bar{\alpha}$, $g(.,\bar{\alpha})$, $\psi(.,\bar{\alpha})$ and $f(t,.,u,\bar{\alpha})$ are smooth functions and that $(\bar{u}, \{\bar{x}(.;\alpha) \mid \alpha \in \mathcal{A}\})$ is a feasible process for which

(A): $g_x(\bar{x}(1;\bar{\alpha}),\bar{\alpha})$ and $\psi_x(\bar{x}(1;\bar{\alpha}),\bar{\alpha})$ are linearly independent.

Let n be the vector of unit length

$$n = \frac{\psi_x(\bar{x}(1;\bar{\alpha}),\bar{\alpha})}{|\psi_x(\bar{x}(1;\bar{\alpha}),\bar{\alpha})|}.$$

Then we easily calculate that

$$Q_0(\bar{\alpha}) \;=\; \Big\{p(.;\bar{\alpha}) \in W^{1,1} \mid -\dot{p}(t;\bar{\alpha}) = H_x \text{ and } -p(1;\bar{\alpha}) \in co\{g_x, +n\} \cup co\{g_x, -n\}\Big\}$$

while $\overline{Q}_0(\bar{\alpha})$ contains the subset

$$\Big\{p(.;\bar{\alpha}) \in W^{1,1} \mid -\dot{p}(.;\bar{\alpha}) = H_x \text{ and } -p(.;\bar{\alpha}) \in \mathrm{co}\{g_x, +n, -n\}\Big\}. \tag{4.1}$$

Here g_x is evaluated at $\bar{x}(1;\bar{\alpha})$.

Notice that the element $p(.;\bar{\alpha}) \equiv 0$ lies in the set (4.1), since $-\dot{p}(t;\bar{\alpha}) = H_x$ is a linear differential equation and $0 \in \mathrm{co}\{g_x, +n, -n\}$. This means that the optimality conditions of Theorem 3.1 are satisfied by any feasible process $(\bar{u}, \{\bar{x}(.;\alpha) \,|\, \alpha \in \mathcal{A}\})$ with the trivial choice of multipliers

$$\Lambda = \delta_{\{\bar{\alpha}\}} \text{ and } p(.;\alpha) \equiv 0 \quad \text{ for all } \alpha \in \mathcal{A}.$$

Theorem 3.1 therefore conveys no useful information about minimizers in this case. By contrast, we have

$$(p(.;\bar{\alpha}) \equiv 0) \;\notin\; Q_0(\bar{\alpha})$$

since, under the hypothesis (A), $0 \notin \mathrm{co}\{g_x, +n\} \cup \mathrm{co}\{g_x, -n\}$; thus the optimality conditions of Theorem 3.1 are not, in this case, automatically satisfied by any feasible process $(u, \{x(.;\alpha) \mid \alpha \in \mathcal{A}\})$.

On the other hand, consider a modification of the above special case, in which the former equality endpoint constraints are replaced by inequality constraints

$$C(\alpha) = \{x \mid \psi(x,\alpha) \leq 0\}$$

and assume that

$$\psi(\bar{x}(1;\bar{\alpha}), \bar{\alpha}) \;=\; 0.$$

Then, under unrestrictive hypotheses,

$$\begin{aligned} Q_0(\bar{\alpha}) &= \overline{Q}_0(\bar{\alpha}) \\ &= \Big\{p(.;\bar{\alpha}) \in W^{1,1} \mid -\dot{p}(.;\bar{\alpha}) = H_x \text{ and } -p(1;\bar{\alpha}) \in \mathrm{co}\{\nabla g, n\}\Big\}. \end{aligned}$$

Here, there is no loss of information in passing from $Q_0(\bar{\alpha})$ to $\overline{Q}_0(\alpha)$.

These observations highlight the fact that the maximum principle for minimax optimal control problems with parameter set a general compact metric space, Theorem 3.1, will find primary application in situations where the endpoint constraints (if they are present) take the form of functional inequality constraints and their generalizations. Theorem 3.1 is not well-suited to problems with endpoint equality constraints.[1] It is therefore of interest to know whether Theorem 3.1 can be refined,

[1] Of course it can be argued that minimax problems of this nature are, broadly speaking, artificial: often such problems will have no minimizers because of the absence of feasible processes, i.e., control functions whose corresponding state trajectories satisfy the equality endpoint constraints for *all* values of the parameter α. Nontrivial maximum principles covering those few cases of interest involving equality endpoint constraints (e.g., cases where the equality constraints involve only those aspects of the dynamics which do not depend on α) can be developed along the lines of Theorem 3.2.

to provide necessary conditions for problems with parameter set a general compact metric space, in which $Q_0(\alpha)$ replaces $\overline{Q}_0(\alpha)$.

We now study two examples, the purpose of which is to demonstrate that this is not possible, in the absence of additional hypotheses.

EXAMPLE 4.1. *Consider the problem*

$$\begin{cases} \textit{Minimize } \sup_{\alpha\in[-1,+1]} -|x(1)-\alpha| \textit{ over } u(.) \textit{ such that} \\ \dot{x}(t) = u(t) \quad \textit{a.e. } t\in[0,1], \\ x(0)=0, \\ u(t)\in[-1,+1] \quad \textit{a.e. } t\in[0,1]\,. \end{cases}$$

This is an example of the general minimax problem in which the parameter set is the interval $\mathcal{A}=[-1,1]$. The cost function depends on α, but the dynamics do not. We denote processes (u,x), since all state trajectories corresponding to a given control function u coincide. Clearly, $(\bar{u}\equiv 0,\, \bar{x}\equiv 0)$ is a minimizer.

Suppose that the assertions of Proposition 2.1 *were valid here. Then there would exist a probability measure Λ with support in*

$$\left\{\alpha \mid -|\bar{x}(1)-\alpha| = max_{\alpha'\in[-1,+1]}\,(-|\bar{x}(1)-\alpha'|)\right\} \;=\; \{0\}, \tag{4.2}$$

and a family of costate arcs $\{p(.;\alpha)\,|\,\alpha\in\mathcal{A}\}$ such that $p(.;\alpha)\in Q_0(\alpha)$ for Λ – a.e. $\alpha\in\mathcal{A}$ and (3.3) *is satisfied. But* (4.2) *implies that*

$$\Lambda = \delta_{\{0\}}.$$

Thus, supp $\{\Lambda\}=\{0\}$ and the only relevant value of α is $\alpha=0$. We calculate

$$\begin{aligned} Q_0(\alpha=0) &= \left\{p\in W^{1,1} \mid -\dot{p}=0,\;\; -p(1)\in\{-1\}\cup\{+1\}\right\} \\ &= \{p\equiv -1\}\cup\{p\equiv +1\}. \end{aligned}$$

We have then, for each $t\in[0,1]$,

$$\int_{\mathcal{A}} H(t,\bar{x}(t),u,p(t;\alpha),\alpha)\;\Lambda(d\alpha) = \begin{cases} +u & \text{if } p(.;\alpha=0)\equiv +1 \\ -u & \text{if } p(.;\alpha=0)\equiv -1, \end{cases}$$

for any $u\in[-1,+1]$ and any family of costate arcs $\{p(.;\alpha)\mid\alpha\in\mathcal{A}\}$ such that $p(.;\alpha)\in Q_0(\alpha)$ for Λ – a.e. $\alpha\in\mathcal{A}$.

We see that $u\to\int_{\mathcal{A}} H(t,\bar{x}(t),u,p(t),\alpha)\;\Lambda(d\alpha)$ cannot be maximized at $u=\bar{u}(t)$ for a.e. $t\in[0,1]$. This shows that the assertions of Theorem 3.1 *may fail to be true, if $\mathcal{A}$ is allowed to be an infinite set. On the other hand,*

$$\{p(.;\alpha=0)\equiv 0\}\;\in\;\overline{Q}_0(\alpha=0)$$

and so the maximization of the Hamiltonian condition is satisfied with Λ taken to be the unit measure concentrated at $\alpha=0$ and with $\{p(.;\alpha)|\alpha\in\mathcal{A}\}$ an arbitrary collection of $W^{1,1}$ functions such that $p(.;\alpha=0)\equiv 0$.

Example 4.1 involves nonsmooth data. The following more elaborate example illustrates that we cannot replace $p(\alpha)\in\overline{Q}_0(\alpha)$ by $p(\alpha)\in Q_0(\alpha)$, even for problems with smooth data.

EXAMPLE 4.2. *Consider the following example of the minimax optimal control problem, in which the state $x = (x_1, x_2)$ is a 2-vector and the control u is scalar:*

$$\text{(P)}\begin{cases} \textit{Minimize } \max_{\alpha\in\mathcal{A}} g(x(1;\alpha),\alpha) \\ \qquad \textit{over } (u, \{x(.;\alpha) \mid \alpha\in\mathcal{A}\}) \textit{ satisfying} \\ \dot{x}(t;\alpha) = f(t, x(t;\alpha), u(t), \alpha) \quad \textit{a.e.}, \\ u(t)\in\Omega \quad \textit{a.e.}, \\ x(0;\alpha) = x_0 \quad \textit{and} \quad x(1;\alpha)\in C(\alpha). \end{cases}$$

Here, $x_0 = col\,(0;0)$, $\Omega = [-1, +1]$,

$$\mathcal{A} \;=\; \left[\frac{1}{3}, \frac{2}{3}\right] \cup \{1\},$$

and, for each $\alpha\in[\frac{1}{3},\frac{2}{3}]\cup\{1\}$, $f = col(f_1, f_2)$ *is the function*

$$f_1(t,x,u,\alpha) = \begin{cases} u & \textit{if } \ 0\le t\le\alpha \\ 0 & \textit{if } \ \alpha < t\le 1 \end{cases}$$

$$f_2(t,x,u,\alpha) = \begin{cases} -u^2 & \textit{if } \ \frac{1}{3}\le t\le\frac{2}{3} \\ 0 & \textit{otherwise}, \end{cases}$$

$$g(x,\alpha) \;=\; \begin{cases} x_2 & \textit{if } \ \alpha = 1 \\ -1 & \textit{if } \ \alpha\in[\frac{1}{3},\frac{2}{3}] \end{cases}$$

and

$$C(\alpha) \;=\; \begin{cases} \{0\}\times R & \textit{if } \ \alpha\in[\frac{1}{3},\frac{2}{3}] \\ R\times R & \textit{if } \ \alpha = 1. \end{cases}$$

Noting the interpretation of this example provided below, we easily check that ($\bar{u}\equiv 0$, $\{\bar{x}(.;\alpha)\equiv(0,0)\,|\,\alpha\in\mathcal{A}\}$) *is a minimizer. Suppose that*

$$\int_{\mathcal{A}} H(t, \bar{x}(t;\alpha), \bar{u}(t), p(t;\alpha), \alpha)\ \Lambda(d\alpha)$$
$$= \max_u \int_{\mathcal{A}} H(t, \bar{x}(t;\alpha), u, p(t;\alpha), \alpha)\ \Lambda(d\alpha) \quad \textit{a.e. } t\in[0,1]$$

is satisfied for some probability measure Λ *and family of arcs* $\{p(. : \alpha)\}$ *such that*

$$p(.;\alpha)\in Q_0(\alpha) \qquad \Lambda - a.e. \quad \alpha\in\mathcal{A}.$$

Partition the adjoint arcs $p(.;\alpha) = (p_1(.;\alpha),\, p_2(.;\alpha))$. *Then for* Λ – *a.e.* $\alpha\in[\frac{1}{3},\frac{2}{3}]$

$$\begin{array}{ll} -\dot{p}_1(.;\alpha) \;\equiv\; 0, & -\dot{p}_2(.;\alpha) \;\equiv\; 0 \\ -p_1(1;\alpha) \;=\; m(\alpha), & -p_2(1;\alpha) \;=\; 0 \end{array}$$

in which $m(\alpha)$ *is a Borel measurable function such that*

$$m(\alpha) = -1 \ \textit{or} \ +1 \quad \textit{for all } \alpha\in\left[\frac{1}{3},\frac{2}{3}\right].$$

Also

$$-\dot{p}_1(. : \alpha = 1)\equiv 0,\ -p_1(1;\alpha = 1) = 0$$
$$-\dot{p}_2(. : \alpha = 1)\equiv 0,\ -p_2(1;\alpha = 1) = +1.$$

Writing $a \vee b := max\{a, b\}$, we deduce from the maximization of the Hamiltonian condition that

$$u \to \int_{[\frac{1}{3} \vee t, \frac{2}{3}]} m(\alpha)\ \Lambda(d\alpha) u + u^2 \chi_{[\frac{1}{3}, \frac{2}{3}]}(t) \Lambda(\{1\})$$

is maximized over $[-1, +1]$ at $u = 0$ a.e. $t \in [0, 1]$. Here χ_D denotes the indicator function of the set D. It follows that, for a.e. $t \in [0, 1]$,

$$\int_{[\frac{1}{3} \vee t, \frac{2}{3}]} m(\alpha)\ \Lambda(d\alpha)\ =\ 0 \tag{4.3}$$

and $\Lambda(\{1\}) = 0$. But (4.3) implies

$$\int_{[\frac{1}{3}, \frac{2}{3}]} m(\alpha)\ \Lambda(d\alpha) = 0$$

and

$$\int_{[t, \frac{2}{3}]} m(\alpha)\ \Lambda(d\alpha) = 0$$

for all $t \in F$, where F is some countable dense subset of $[\frac{1}{3}, \frac{2}{3}]$. But since sets of the form $[\frac{1}{3}, \frac{2}{3}]$ and $[t, \frac{2}{3}]$ (for $t \in F$) generate the Borel subsets of $[\frac{1}{3}, \frac{2}{3}]$, we see that

$$\int_{B} m(\alpha)\ \Lambda(d\alpha) = 0 \tag{4.4}$$

for all Borel sets $B \in \mathcal{A}$. Let $\mathcal{A}^{\pm} = \{\alpha \mid m(\alpha) = \pm 1\}$. Since $\mathcal{A}^- \cup \mathcal{A}^+ = \mathcal{A}$ and $||\Lambda||_{T.V.} = 1$, either $\Lambda(\mathcal{A}^+) > 0$ or $\Lambda(\mathcal{A}^-) > 0$. Without loss of generality assume the former. Then

$$\int_{B} m(\alpha)\ \Lambda(d\alpha) = \int_{B}\ \Lambda(d\alpha)\ =\ \Lambda(\mathcal{A}^+)\ > 0,$$

when we select $B = \mathcal{A}^+$. This contradicts (4.4). It follows that a version of the minimax maximum principle is not valid for this problem, in which

$$p(.;\alpha) \in Q_0(\alpha) \quad \textit{for } \Lambda - a.e.\ \ \alpha \in \mathcal{A}.$$

The preceding example originates in an optimal control problem with pathwise equality constraints

$$(\tilde{\mathrm{P}}) \begin{cases} & \text{Minimize} \quad -\int_{\frac{1}{3}}^{\frac{2}{3}} |u(t)|^2\ dt \\ \text{s.t.} & \dot{x}(t) = u(t), \quad a.e.\ t \in [0, 1], \\ & x(0) = 0, \\ & x(t) = 0 \quad \text{for } \frac{1}{3} \le t \le \frac{2}{3}, \\ & u(t) \in R, \quad a.e.\ t \in [0, 1]. \end{cases}$$

which has been reformulated as an example of the general minimax optimal control problem (P). The fact that we cannot derive a maximum principle involving the set $Q_0(\alpha)$ in the above example reflects the fact that measure multipliers can be used in necessary conditions for problems with pathwise *equality* constraints only in very special circumstances.

Notice that the assertions of Theorem 3.1 are consistent with Example 4.2. In this example

$$Q_0(\alpha) \;=\; \big\{(p_1(.;\alpha)\equiv 0, p_2(.;\alpha)) \mid p_2(.;\alpha)\equiv m(\alpha)\big\},$$

for Λ − a.e. $\alpha \in [\frac{1}{3}, \frac{2}{3}]$, where $m(.)$ is some Borel measurable function such that

$$m(\alpha) \,\in\, \{-1\}\cup\{+1\} \quad \Lambda - a.e.$$

These sets are too small for the maximization condition on the "averaged" Hamiltonian to hold (for any choice of $m(.)$). On the other hand, for Λ − a.e. $\alpha \in [\frac{1}{3}, \frac{2}{3}]$,

$$\overline{Q}_0(\alpha) \;=\; \big\{(p_1(.;\alpha)\equiv 0, p_2(.;\alpha)) \mid p_2(.;\alpha)\equiv \tilde{m}(\alpha)\big\}$$

in which $\tilde{m}(.)$ is some Borel measurable function such that

$$\tilde{m}(\alpha) \,\in\, [-1,+1] \quad \Lambda - a.e.$$

The maximization condition does hold (in a trivial sense), with respect to $(p_1(.;\alpha), p_2(.;\alpha))$s chosen from this larger set; we can take $(p_1(.;\alpha), p_2(.;\alpha)) \equiv (0,0)$ $\Lambda - a.e.$

5. Special cases. In this section, we examine implications of the minimax maximum principle for a number of special cases of interest. Utmost generality is not a goal here; indeed, we often focus on smooth versions of the optimality conditions, when the nonsmooth version could easily be derived, better to reveal their essential character. Throughout, $\mathcal{A}$ is an arbitrary compact metric space.

Consider first the *minimax optimal control problem with no right endpoint constraints*,

$$(\text{P1})\begin{cases}\text{Minimize } \max_{\alpha\in\mathcal{A}} g(x(.;\alpha),\alpha)\\ \qquad \text{over measurable functions } u:[0,1]\to R^m \text{ such that}\\ u(t)\in\Omega(t), \quad a.e.\ t\in[0,1]\\ \qquad \text{and arcs } \{x(.;\alpha):[0,1]\to R^n \mid \alpha\in\mathcal{A}\} \text{ such that, for each } \alpha\in\mathcal{A},\\ \dot{x}(t;\alpha)=f(t,x(t;\alpha),u(t),\alpha) \quad a.e.\ t\in[0,1],\\ x(0;\alpha)=x_0.\end{cases}$$

The data for (P1) comprises a compact metric space $\mathcal{A}$, functions $g: R^n\times\mathcal{A}\to R$, $f:[0,1]\times R^n\times R^m\times\mathcal{A}\to R^n$, a vector $x_0\in R^n$, and a time dependent set $\Omega(t)\subset R^m$, $0\le t\le 1$.

General necessary conditions for (P1) follow directly from Theorem 3.1. We state the conditions merely in the special case when the data are smooth.

PROPOSITION 5.1. *Let* $(\bar{u}, \{\bar{x}(.;\alpha)\mid\alpha\in\mathcal{A}\})$ *be a strong local minimizer for* (P1). *Assume that, for some* $\delta>0$, *the Hypotheses* (S1), (S2), *and* (S4) *of section* 3 *are satisfied. Assume, furthermore, that*

(a) g *is continuous,* $g(.,\alpha)$ *is differentiable for each* $\alpha\in\mathcal{A}$ *and* g_x *is continuous.*

(b) $f(t,.,u,\alpha)$ *is continuously differentiable on a neighborhood of* $\bar{x}(t;\alpha)$ *for all* $u\in\Omega(t)$ *and* $\alpha\in\mathcal{A}$, *a.e.* $t\in[0,1]$, *and* $\alpha\to f_x(t,x,u,\alpha)$ *is uniformly continuous with respect to* $(t,x,u)\in\{(t',x',u')\in[0,1]\times R^n\times R^m \mid u'\in\Omega(t')\}$.

Then

$$\int H(t,\bar{x}(t;\alpha),\bar{u}(t),p(t;\alpha),\alpha)\ \Lambda(d\alpha)$$
$$= \max_{u\in\Omega(t)} \int H(t,\bar{x}(t;\alpha),u,p(t;\alpha),\alpha)\ \Lambda(d\alpha) \quad a.e.$$

for some Radon probability measure $\Lambda \in C^*(\mathcal{A})$ *and some family of arcs* $\{p(.;\alpha) \in W^{1,1}([0,1];R^n) \mid \alpha \in \mathcal{A}\}$ *such that*

$$supp\ \{\Lambda\} \subset \left\{\alpha \in \mathcal{A} \mid g(\bar{x}(1;\alpha),\alpha) = \max_{\alpha'\in\mathcal{A}}\ g(\bar{x}(1;\alpha'),\alpha')\right\}$$

and, for Λ – *a.e.* $\alpha \in \mathcal{A}$,

(i) $-\dot{p}(t;\alpha) = f_x^T(t,\bar{x}(t;\alpha),\bar{u}(t),\alpha)\,p(t;\alpha) \quad a.e.$

(ii) $-p(1;\alpha) = g_x(\bar{x}(1;\alpha),\alpha)$.

Proof. Everything follows from Thm. 3.1, when we note that, if a function $\phi : R^n \to R$ is continuously differentiable on a neighborhood of a point $\bar{x}$, then $\partial\phi(\bar{x}) = \{\phi_x(\bar{x})\}$ and, under the stated hypotheses,

$$\overline{Q}_0(\alpha) \ = \ \begin{cases} \{p' \in W^{1,1} \mid -\dot{p}' = f_x^T p',\ -p'(1) = g_x(\bar{x}(1;\alpha),\alpha)\} \\ \qquad\qquad\qquad \text{if } \ g(\bar{x}(1;\alpha)) = \max_{\alpha'\in\mathcal{A}} g(\bar{x}(1;\alpha'),\alpha') \\ \emptyset \qquad\qquad\qquad \text{otherwise.} \end{cases} \qquad \square$$

Consider next the *optimal control problem with robust feasibility constraints*:

$$\text{(P2)} \begin{cases} \text{Minimize } g(x(1;\alpha^*)) \\ \qquad \text{over measurable functions } u : [0,1] \to R^m \text{ such that} \\ u(t) \in \Omega(t), \quad a.e.\ t \in [0,1] \\ \qquad \text{and arcs } \{x(.;\alpha) : [0,1] \to R^n \mid \alpha \in \mathcal{A}\} \text{ such that, for each } \alpha \in \mathcal{A}, \\ \dot{x}(t;\alpha) = f(t,x(t;\alpha),u(t),\alpha) \quad a.e.\ t \in [0,1], \\ x(0;\alpha) = x_0 \quad \text{ and } \quad \psi(x(1;\alpha)) \le 0. \end{cases}$$

The data for (P2) comprises a set $\mathcal{A} \subset R^k$, a point $\alpha^* \in \mathcal{A}$, functions $g : R^n \to R$ and $f : [0,1] \times R^n \times R^m \times \mathcal{A} \to R^n$ and $\psi : R^n \to R^{r'}$, a vector $x_0 \in R^n$ and a time dependent set $\Omega(t) \subset R^m$, $0 \le t \le 1$. The endpoint functional inequality terminal constraint is interpreted in the usual "componentwise" manner.

This is a formulation of optimal control problems involving an unknown parameter α, in which α is expected to take its nominal value α^*. Here, it is appropriate to choose a control to minimize the cost based on the system response for $\alpha = \alpha^*$. But our choice of control is restricted by the requirement that, even if α deviates from α^*, constraints on state variables must not be violated. Here, we regard values of α different from α^* as due to system degradation or failure, and "$\psi(x(1;\alpha)) \le 0$ for all $\alpha \in \mathcal{A}$" is the requirement that operational constraints (on displacements, velocities, pressures, etc.) are satisfied, even in the event of breakdown.

For simplicity, we assume that the data are smooth and that there is a single endpoint constraint ($r' = 1$).

PROPOSITION 5.2. *Let* $(\bar{u}, \{\bar{x}(.;\alpha) \mid \alpha \in \mathcal{A}\})$ *be a strong local minimizer for* (P2). *Assume that, for some* $\delta > 0$, *Hypotheses* (S1), (S2), *and* (S4) *are satisfied. Assume, furthermore, that* $r' = 1$ *and*

(a) g *and* ψ *are continuously differentiable on* $\bar{x}(1;\alpha^*) + \delta B$.

(b) *$f(t,.,u,\alpha)$ is continuously differentiable on a neighborhood of $\bar{x}(t;\alpha)$ for all $u \in \Omega(t)$ and $\alpha \in \mathcal{A}$, a.e. $t \in [0,1]$, and $\alpha \to f_x(t,x,u,\alpha)$ is uniformly continuous with respect to $(t,x,u) \in \{(t',x',u') \in [0,1] \times R^n \times R^m \mid u' \in \Omega(t')\}$.*

Then

$$\int H(t,\bar{x}(t;\alpha),\bar{u}(t),p(t;\alpha),\alpha)\ \Lambda(d\alpha)\ =$$
$$\max_{u\in\Omega(t)} \int H(t,\bar{x}(t;\alpha),u,p(t;\alpha),\alpha)\ \Lambda(d\alpha) \quad a.e.\ t \in [0,1]$$

for some family of arcs $\{p(.;\alpha) \in W^{1,1}([0,1];R^n)\}$, a number $r \in [0,1]$ and a Radon probability measure $\Lambda \in C^*(\mathcal{A})$ such that

$$supp\ \{\Lambda\} \subset \big(\{\alpha^*\} \cup \big\{\alpha \in \mathcal{A} \mid \psi(\bar{x}(1;\alpha)) = 0\big\}\big),$$

and, for Λ – a.e. $\alpha \in \mathcal{A}$,

(i) $-\dot{p}(t;\alpha) = f_x^T(t,\bar{x}(t;\alpha),\bar{u}(t),\alpha)\,p(.;\alpha) \quad a.e.\ t \in [0,1]$,

(ii) $-p(1;\alpha) = \begin{cases} \psi_x(\bar{x}(1;\alpha)) & \text{if } \alpha \neq \alpha^* \\ rg_x(\bar{x}(1;\alpha)) + (1-r)\psi_x(\bar{x}(1;\alpha)) & \text{if } \alpha = \alpha^* \end{cases}$.

Proof. It might appear that the simplest way to prove Proposition 6.2 would be to reformulate (P2) as a special case of the general minimax optimal control problem (P), in such a manner that $(\bar{u},\{\bar{x}(.;\alpha) \mid \alpha \in \mathcal{A}\})$ remains a minimizer, by setting

$$g(x,\alpha) := \begin{cases} g(x) & \text{if } \alpha = \alpha^* \\ -K & \text{if } \alpha \neq \alpha^* \end{cases}$$

and

$$C(\alpha) := \{x \mid \psi(x) \leq 0\} \quad \text{for all } \alpha.$$

Here, K is a positive number such that, for some $\delta' > 0$,

$$\inf\big\{g(x) \mid x \in \bar{x}(1;\alpha) + \delta' B,\ \alpha \in \mathcal{A}\big\}\ >\ -K.$$

This is not helpful, however, since $\alpha \to g(x,\alpha)$ violates the continuity hypothesis (S3) for application of Theorem 3.1. Instead, we take a point $b \notin \mathcal{A}$ and associate with (P2) a general minimax problem with extended parameter set $\tilde{\mathcal{A}} := \mathcal{A} \cup \{b\}$, in which $g(x,\alpha)$ is the function

$$g(x,\alpha) := \begin{cases} g(x) & \text{if } \alpha = b \\ -K & \text{if } \alpha \in \mathcal{A} \end{cases}$$

and in which f is the extension of the function f of (P2), to allow for α's in $\mathcal{A} \cup \{b\}$,

$$f(t,x,u,\alpha = b) := f(t,x,u,\alpha^*).$$

The hypotheses are satisfied for the application of Theorem 3.2, with reference to the process $(\bar{u},\{\bar{x}(.;\alpha) \mid \alpha \in \mathcal{A}\})$, when we partition the extended parameter set as

$$\tilde{\mathcal{A}} = \big(\mathcal{A}^1 := \mathcal{A}\big)\ \cup\ \big(\mathcal{A}^2 := \{b\}\big).$$

Straightforward calculations yield the following information: for $\alpha \in \mathcal{A}$

$$\overline{Q}_0(\alpha) \;=\; \begin{cases} \{q(.)\mid -\dot{q}(t) = f_x^T(t,\bar{x}(t;\alpha),\bar{u}(t),\alpha)q(t), & \\ \qquad -q(1) = \psi_x(\bar{x}(1;\alpha))\} & \text{if } \psi(\bar{x}(1;\alpha)) = 0 \\ \emptyset & \text{if } \psi(\bar{x}(1;\alpha)) < 0 \end{cases}$$

and

$$Q_0(\alpha = b) \;=\; \{q(.)\mid -\dot{q}(t) = f_x^T(t,\bar{x}(t;\alpha^*),\bar{u}(t),\alpha^*)q(t),\;\; -q(1) = g_x(\bar{x}(1;\alpha^*))\}\,.$$

We deduce the existence of a Radon probability measure $\mu \subset C^*(\mathcal{A}\cup\{b\})$ and arcs $\{q(.;\alpha)\mid \alpha\in\mathcal{A}\}\cup\{q(.;b)\}$ such that, for $\mu - a.e.\ \alpha \in \mathcal{A}$,

$$-\dot{q}(t;\alpha) \;=\; f_x^T(t,\bar{x}(t;\alpha),\bar{u}(t),\alpha)q(t;\alpha),\;\; -q(1;\alpha) \;=\; \psi_x(\bar{x}(1;\alpha)),$$

if $\alpha \in \mathcal{A}$ and

$$-\dot{q}(t;b) \;=\; f_x^T(t,\bar{x}(t;\alpha^*),\bar{u}(t),\alpha^*)q(t;b),\;\; -q(1;b) \;=\; g_x(\bar{x}(1;\alpha^*)).$$

Furthermore, $u \to \mathcal{H}(t,u)$ is maximized over $u \in \Omega(t)$ at $u = \bar{u}(t)$ for $a.e.\ t \in [0,1]$, where

$$\mathcal{H}(t,u) = \int_{\mathcal{A}\cup\{b\}} q(t;\alpha)\cdot f(t,\bar{x}(t;\alpha),u,\alpha)\;\mu(d\alpha)$$

and

$$supp\,\{\mu\} \;\subset\; \{\alpha\in\mathcal{A}\,|\,\psi(\bar{x}(1;\alpha)) = 0\}\cup\{b\}.$$

Now choose

$$r \;=\; \begin{cases} \frac{\mu(\{b\})}{\mu(\{b\})+\mu(\{\alpha^*\})} & \text{if } \mu(b) > 0 \\ 0 & \text{otherwise,} \end{cases}$$

$$p(.;\alpha) := \begin{cases} q(.;\alpha) & \text{for } \;\alpha \neq \alpha^* \\ rq(.;b) + (1-r)q(.;\alpha^*) & \text{for } \;\alpha = \alpha^* \end{cases}.$$

Choose also the Radon measure $\Lambda \in C^*(\mathcal{A})$,

$$\Lambda(E) := \begin{cases} \mu(\{b\}) + \mu(E) & \text{if } \;\alpha^* \in E \\ \mu(E) & \text{if } \;\alpha^* \notin E \end{cases}$$

for any Borel subset E of $\mathcal{A}$. Notice that $||\Lambda||_{T.V.} \;=\; ||\mu||_{T.V.} \;=\; 1$, so Λ is a probability measure. Clearly

$$supp\,\{\Lambda\} \;\subset\; \{\alpha^*\}\cup\{\alpha\in\mathcal{A}\,|\,\psi(\bar{x}(1;\alpha)) = 0\}\,.$$

We have

$$
\begin{aligned}
\mathcal{H}(t,u) &= \left(\int_{\{b\}} + \int_{\{\alpha^*\}} + \int_{\mathcal{A}\setminus\{\alpha^*\}}\right) q(t;\alpha)\cdot f(t,\bar{x}(t;\alpha),u,\alpha)\ \mu(d\alpha) \\
&= \big(\mu(\{b\})q(t;b) + \mu(\{\alpha^*\})q(t;\alpha)\big)\cdot f(t,\bar{x}(t;\alpha^*),u,\alpha^*) \\
&\quad + \int_{\mathcal{A}\setminus\{\alpha^*\}} q(t;\alpha)\cdot f(t,\bar{x}(t;\alpha),u,\alpha)\ \mu(d\alpha) \\
&= \big(\mu(\{b\}) + \mu(\{\alpha^*\})\big)\big(rq(t;b) + (1-r)q(t;\alpha^*)\big)\cdot f(t,\bar{x}(t;\alpha^*),u,\alpha^*) \\
&\quad + \int_{\mathcal{A}\setminus\{\alpha^*\}} q(t;\alpha)\cdot f(t,x(t;\alpha),u,\alpha)\ \mu(d\alpha) \\
&= \Lambda(\{\alpha^*\})p(t;\alpha^*)\cdot f(t,\bar{x}(t;\alpha^*),u,\alpha^*) + \int_{\mathcal{A}\setminus\{\alpha^*\}} p(t;\alpha)\cdot f(t,\bar{x}(t;\alpha),u,\alpha)\ \Lambda(d\alpha) \\
&= \int p(t;\alpha)\cdot f(t,\bar{x}(t;\alpha),u,\alpha)\ \Lambda(d\alpha).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
-\dot{p}(t;\alpha) &= f_x^T p(t;\alpha) \\
-p(1;\alpha) &= \psi_x(\bar{x}(1;\alpha))
\end{aligned}
$$

for $\alpha \neq \alpha^*$. Also, by homogeneity,

$$
\begin{aligned}
-\dot{p}(t;\alpha) &= f_x^T p(t;\alpha) \\
-p(1;\alpha) &= r g_x(\bar{x}(1;\alpha)) + (1-r)\psi_x(\bar{x}(1;\alpha))
\end{aligned}
$$

for $\alpha = \alpha^*$. The proof is complete. $\square$

Consider finally the *state constrained optimal control problem*,

$$
(\mathrm{P3}) \begin{cases}
\text{Minimize } g(x(1)) \text{ over measurable functions } u:[0,1]\to R^n \text{ such that} \\
\dot{x}(t) = f(t,x(t),u(t)) \quad a.e.\ t\in[0,1], \\
x(0) = x_0 \quad \text{and} \quad x(1)\in C, \\
u(t)\in\Omega(t) \quad a.e.\ t\in[0,1], \\
h(t,x(t))\le 0 \qquad t\in[0,1].
\end{cases}
$$

This is a "parameter-free" version of (P) (the function f no longer depends on α), to which has been appended an endpoint constraint and a pathwise state constraint

$$
h(t,x(t)) \ \le\ 0 \quad \text{for all } t\in[0,1].
$$

Here, $C\subset R^n$ is a given set and $h:[0,1]\times R^n\to R$ is a given function. This standard optimal control problem with state constraints would appear to have little relevance to minimax optimal control. The connection is this; (P3) can be interpreted as a minimax type optimal control problem to which the analytical tools of this paper are applicable. This is demonstrated below.

Thus, studying the state constrained optimal control problem in the present context establishes links between minimax optimal control and other well-established areas of optimal control. It also makes clear that the task of deriving necessary conditions of optimality for minimax problems is a challenging one, since it is at least as difficult as deriving necessary conditions for state constrained optimal control problems.

PROPOSITION 5.3. *Let* $(\bar{u},\bar{x})$ *be a strong local minimizer for* (P3). *Assume that for some* $\delta>0$*, the following hypotheses are satisfied.*

(a) *$f(.,x,.)$ is $\mathcal{L}\times\mathcal{B}$ measurable for each $x\in R^n$. $t\rightsquigarrow\Omega(t)$ has Borel measurable graph.*

(b) *There exist $k_f(t)\in L^1$ and $c_f>0$ such that*

$$|f(t,x,u)-f(t,x',u)|\le k_f(t)|x-x'| \quad \textit{and} \quad |f(t,x,u)|\le c_f$$

for all $x,x'\in\bar{x}(t)+\delta B$ and $u\in U(t)$, a.e. $t\in[0,1]$. Furthermore, $f(t,.,u)$ is continuously differentiable on a neighborhood of $\bar{x}(t)$ for all $u\in\Omega(t)$ a.e. $t\in[0,1]$.

(c) *g is continuously differentiable on $\bar{x}(1)+\delta B$.*

(d) *h is continuously differentiable.*

Then there exists an arc $p\in W^{1,1}([0,1];R^n)$, $\lambda\ge 0$, and a Radon measure $\mu\in C^*([0,1])$ such that

(i) $\lambda+\|\mu\|_{T.V.}+|p(1)|\neq 0$

(ii) $-\dot{p}=f_x^T(t,\bar{x}(t),\bar{u}(t))\dots\left(p(t)+\int_{[0,t)}h_x(s,\bar{x}(s))\mu(ds)\right)$ *a.e.,*

(iii) $-(p(1)+\int_{[0,1]}h_x(s,\bar{x}(s))\mu(ds))\in\lambda g_x(\bar{x}(1))+N_C(\bar{x}(1))$

(iv) $\text{supp}\,\{\mu\}\subset\{t\mid h(t,\bar{x}(t))=0\}$

and

$$u\to\left(p(t)+\int_{[0,t)}h_x(s,\bar{x}(s))\ \mu(ds)\right)\cdot f(t,\bar{x}(t),u)$$

is maximized over $u\in\Omega(t)$ at $u=\bar{u}(t)$, a.e. $t\in[0,1]$.

We see that the minimax maximum principle can be used to obtain the maximum principle for state constrained problems with a general right endpoint constraint (cf. [10]).

Proof. We reformulate (P3) as a general minimax problem with parameter set $\mathcal{A}=[0,1]\cup\{2\}$. For all $\alpha\in[0,1]$ set

$$f(t,x,u,\alpha)\ :=\ \begin{cases} f(t,x,u) & \text{for } \ 0\le t\le\alpha,\\ 0 & \text{for } \ t>\alpha\end{cases}$$

$$g(x,\alpha)\ :=\ -K,$$

$$C(\alpha)\ =\ \{x\mid h(\alpha,x)\le 0\}.$$

Here, $-K$ is a number strictly less than $g(\bar{x}(1))$. Also set

$$f(t,x,u,\alpha=2)\ :=\ f(t,x,u)$$

$$g(x,\alpha=2)\ :=\ g(x),$$

$$C(\alpha=2)\ =\ C.$$

Clearly $(\bar{u},\{\bar{x}(.;\alpha)\mid\alpha\in\mathcal{A}\})$ is a strong local minimizer for the general minimax optimal control problem, with these identifications of the data, when

$$\bar{x}(t;\alpha)=\begin{cases}\bar{x}(t) & \text{for } \ 0\le t\le\alpha,\\ \bar{x}(\alpha) & \text{for } \ t>\alpha\end{cases}$$

for $\alpha\in[0,1]$ and

$$\bar{x}(.;\alpha=2)\equiv\bar{x}(.).$$

Now apply Theorem 3.3. (See also succeeding comments regarding the nature of the endpoint constraints.) Let $\{p(.;\alpha)\,|\,\alpha \in [0,1]\}$ and $p(.;\alpha = 2)$ be the "costate arcs" for this problem and let $\Lambda \subset C^*([0,1] \cup \{2\})$ be the Radon probability measure whose existence is asserted in the theorem. Define $\mu' \subset C^*([0,1])$ to be the restriction of Λ to $[0,1]$. Then

$$||\mu'||_{T.V.} \;\leq\; 1.$$

We have, for $0 \leq \alpha \leq 1$,

$$\begin{aligned} -\dot{p}(t;\alpha) \;&=\; \begin{cases} f_x^T(t,\bar{x}(t),\bar{u}(t))p(t;\alpha) & \text{for } \; 0 \leq t \leq \alpha, \\ 0 & \text{for } \; t > \alpha, \end{cases} \\ -p(\alpha;\alpha) \;&=\; h_x(\alpha,\bar{x}(\alpha)) \end{aligned}$$

and

$$\begin{aligned} -\dot{p}(t;\alpha = 2) \;&=\; f_x^T(t,\bar{x}(t),\bar{u}(t))p(t;\alpha = 2), \\ -p(1;\alpha = 2) \;&\in\; \lambda g_x(\bar{x}(1)) + (1-\lambda)\{\xi \in N_C(\bar{x}(1))\,|\,|\xi| = 1\}. \end{aligned}$$

Furthermore, $\bar{u}(t)$ maximizes

$$u \to \left((1 - \|\mu'\|_{T.V.})p(t;\alpha = 2) + \int_{[t,1]} p(t;\alpha)\; \mu'(d\alpha) \right) \cdot f(t,\bar{x}(t),u)$$

over $u \in \Omega(t),\ a.e.\ t \in [0,1]$, and

$$supp\ \{\mu'\} \subset \{\alpha \in [0,1] \mid h(\alpha,\bar{x}(\alpha)) = 0\}.$$

Let $\Phi(t,s)$ be the fundamental matrix for the linear equation $\dot{z}(t) = -f_x^T(t,\bar{x}(t),\bar{u}(t))z(t)$, i.e., for any $s \in [0,1]$, $\Phi(.,s)$ solves $\frac{d}{dt}\Phi(t,s) = -f_x^T(t,\bar{x}(t),\bar{u}(t))\Phi(t,s)$ for $0 \leq t \leq 1$ and $\Phi(s,s) = I$.

Suppose $\|\mu'\|_{T.V.} < 1$. Define

$$\mu := \frac{1}{1 - \|\mu'\|_{T.V.}}\mu'.$$

Then,

$$u \to \left(p(t) + \int_{[0,t)} h_x(s,\bar{x}(s))\; \mu(ds) \right) \cdot f(t,\bar{x}(t),u)$$

is maximized over $u \in \Omega(t)$ at $u = \bar{u}(t),\ a.e.\ t \in [0,1]$, where

$$p(t) := p(t;\alpha = 2) + \int_{[t,1]} p(t;\alpha)\; \mu(d\alpha) - \int_{[0,t)} h_x(\alpha,\bar{x}(\alpha))\; \mu(d\alpha).$$

We deduce from the differential equations for $p(.;\alpha = 2)$ and $p(.;\alpha)$, $\alpha \in [0,1]$, that p satisfies

$$\begin{aligned} p(t) &= -\Phi(t,1)[\lambda g_x(\bar{x}(1)) + (1-\lambda)\xi] \\ &\qquad - \int_{[t,1]} \Phi(t,\alpha)h_x(\alpha,\bar{x}(\alpha))\; \mu(d\alpha) - \int_{[0,t)} h_x(\alpha,\bar{x}(\alpha))\; \mu(d\alpha) \quad \text{for all } t \in [0,1]\,, \end{aligned}$$

for some $\xi \in \{\xi' \in N_C(\bar{x}(1)) \,|\, |\xi'| = 1\}$. It can be deduced from this relationship that $p(.)$ is an absolutely continuous function which satisfies

$$-\dot{p}(t) = f_x^T(t, \bar{x}(t), \bar{u}(t)) \left(p(t) + \int_{[0,t)} h_x(\alpha, \bar{x}(\alpha)) \; \mu(d\alpha) \right) \quad a.e. \; t \in [0,1]$$

$$-\left(p(1) + \int_{[0,1]} h_x(\alpha, \bar{x}(\alpha)) \; \mu(d\alpha) \right) = \lambda g_x(\bar{x}(1)) + (1-\lambda)\xi$$
$$\in \lambda g_x(\bar{x}(1)) + N_C(\bar{x}(1)).$$

Notice that if $\|\mu'\|_{T.V.} = 0$ and $\lambda = 0$, then $|p(1)| = |\xi| = 1$. Thus, the multiplier nondegeneracy condition is satisfied. We have confirmed the assertions of the proposition in the case $\|\mu'\|_{T.V.} < 1$.

It remains then to consider the case when $\|\mu'\|_{T.V.} = 1$. Set $\mu = \mu'$. Now condition (iv) in the theorem statement is valid with

$$p(t) = \int_{[t,1]} p(t;\alpha) \; \mu(d\alpha) - \int_{[0,t)} h_x(\alpha, \bar{x}(\alpha)) \; \mu(d\alpha).$$

It can be deduced that p satisfies

$$-\dot{p}(t) \; = \; f_x^T(t, \bar{x}(t)\bar{u}(t)) \left(p(t) + \int_{[0,t)} h_x(\alpha, \bar{x}(\alpha)) \; \mu(d\alpha) \right) \; a.e. \; t \in [0,1]$$

$$-\left(p(1) + \int_{[0,1]} h_x(\alpha, \bar{x}(\alpha)) \; \mu(d\alpha) \right) \; = \; 0.$$

But

$$0 \in \lambda g_x(\bar{x}(1)) + N_C(\bar{x}(1)),$$

when $\lambda = 0$. The assertions of the proposition have been confirmed in this case too, and the proof is complete. □

6. Proofs of Theorems 3.1–3.3. Our analysis will require some properties of measures, summarized in the following proposition.

PROPOSITION 6.1. *Take a compact metric space $\mathcal{A}$, a sequence $\{\mu_i\}$ of nonnegative Radon measures in $C^*(\mathcal{A})$, a sequence $\{D_i : \mathcal{A} \to R^n\}$ of multifunctions and a sequence of Borel measurable functions $\{\gamma_i : \mathcal{A} \to R^n\}$. Take also a measure $\mu \in C^*(\mathcal{A})$ and a multifunction $D : \mathcal{A} \to R^n$. Assume that $Gr\ D$ is compact,*

$$D(\alpha) \text{ is convex for each } \alpha \in \mathcal{A}, \tag{6.1}$$

$$\limsup_{i \to \infty} \; Gr\ D_i \subset Gr\ D,$$

$$\gamma_i(\alpha) \in D_i(\alpha) \qquad \mu_i \text{ – a.e. } \alpha \in \mathcal{A} \quad \text{for} \;\; i = 1, 2, \ldots$$

and

$$\mu_i \to \mu \qquad \text{weakly}^*.$$

Define $\eta_i \in C^*(\mathcal{A}; R^n)$ according to

$$\eta_i(d\alpha) = \gamma_i(\alpha)\mu_i(d\alpha) \qquad i = 1, 2, \ldots.$$

Then,

(i) *Along a subsequence,*

$$\eta_i \to \eta \qquad \textit{weakly}^*$$

for some $\eta \in C^*(\mathcal{A}; R^k)$ *and some Borel measurable function* γ *such that*

$$\eta(d\alpha) = \gamma(\alpha)\mu(d\alpha),$$

and

$$\gamma(\alpha) \in D(\alpha) \qquad \mu - a.e.$$

(ii) *Suppose* $\mathcal{A}$ *is expressible as a union of disjoint sets*

$$\mathcal{A} = \mathcal{A}^{(1)} \cup \mathcal{A}^{(2)}$$

in which $\mathcal{A}^{(1)}$ *is compact metric space and* $\mathcal{A}^{(2)}$ *is finite. Then the assertions of part* (i) *remain valid when the hypothesis* (6.1) *is replaced by*

$$D(\alpha) \textit{ is convex for each } \alpha \in \mathcal{A}^{(1)}.$$

Proof. The proof, which is similar to that of ([10], Proposition 9.2.1), is omitted. □

6.1. Proof of Theorem 3.1. We observe at the outset that we can, without loss of generality, replace (S2) and (S3) by stronger (global) hypotheses in which $\delta = +\infty$, that is, we can require the stated conditions in (S2) to hold for all $x, x' \in R^n$, not merely in $x, x' \in \bar{x}(t) + \delta B$; likewise for (S3). This can always be arranged by replacing f and g by their "localizations" $(t, x, u, \alpha) \to f(t, tr_{\bar{x}(t),\delta}(x), u, \alpha)$ and $(t, x, u, \alpha) \to g(t, tr_{\bar{x}(t),\delta}(x), \alpha)$, in which $tr_{y,\delta}(x)$ is the truncation function

$$tr_{y,\delta}(x) \;=\; \begin{cases} x & \text{if } |x-y| < \delta \\ y + \delta(x-y)/|x-y| & \text{if } |x-y| \geq \delta \end{cases}.$$

The property that $\bar{x}$ is a strong local minimizer is preserved under this modification of the data. It is a consequence of the hypotheses, strengthened in this way that to each $u \in \mathcal{U}$ and $\alpha \in \mathcal{A}$, there corresponds a unique state trajectory (on $[0,1]$ with initial state x_0). This we write $x(.; \alpha, u)$.

The following lemma brings together some useful facts, regarding the dependence of the state trajectories on controls and parameters.

Let $\Delta : R^n \times R^n \to R$ denote the Ekeland metric on $\mathcal{U}$,

$$\Delta(u_1, u_2) := meas\{t \mid u_1(t) \neq u_2(t)\}.$$

LEMMA 6.1. *For any* $\delta > 0$*, a finite subset* $\widetilde{\mathcal{A}} \subset \mathcal{A}$ *and* $\rho > 0$ *can be chosen such that*

(i)

$$\sup_{u \in \mathcal{U}} \; \sup_{\alpha \in \mathcal{A}} \; \inf_{\alpha' \in \widetilde{\mathcal{A}}} \; \|x(.; \alpha, u) - x(.; \alpha', u)\|_C < \delta$$

(ii)

$$\sup_{\alpha \in \mathcal{A}} \; \big\{ \|x(.; \alpha, u) - x(.; \alpha, u')\|_C \mid u, u' \in \mathcal{U}, \; \Delta(u, u') < \rho \big\} < \delta.$$

These assertions are straightforward consequences of Filippov's existence theorem. (See, e.g., [10], Theorem 2.4.3.)

Take a sequence $\epsilon_i \downarrow 0$. For each i define $J_i : \mathcal{U} \to R$

$$J_i(u) := \max_{\alpha\in\mathcal{A}} \Big\{ \Big(g(x(1;\alpha,u),\alpha) - \max_{\alpha'\in\mathcal{A}} g(\bar{x}(1;\alpha'),\alpha') + \epsilon_i^2 \Big) \vee d_{C(\alpha)}(x(1;\alpha,u)) \Big\}.$$

Notice that $J_i(u) \geq 0$ for all $u \in \mathcal{U}$ and $J_i(\bar{u}) = \epsilon_i^2$. It follows that $\bar{u}$ is an ϵ_i^2-minimizer for the functional J_i on $\mathcal{U}$.

For each i, J_i is continuous with respect to the Δ-metric topology. We deduce from Ekeland's theorem the existence of a control function v_i, for each i, such that

$$\Delta(v_i,\bar{u}) \leq \epsilon_i$$

and

$$J_i(v_i) + \epsilon_i\Delta(v_i,v_i) \;=\; \min_{u\in\mathcal{U}} \Big\{ J_i(u) + \epsilon_i\Delta(v_i,u) \Big\}.$$

We have

$$J_i(v_i) > 0 \quad \text{for all } i \text{ sufficiently large,}$$

since $(\bar{u}, \{\bar{x}(.;\alpha)\,|\,\alpha\in\mathcal{A}\})$ is a strong local minimizer for (P) and by Lemma 6.1 (ii). Fix i. For any finite subset $\widetilde{\mathcal{A}} \subset \mathcal{A}$, which will be chosen presently, consider the functional

$$J_i^{\tilde{\mathcal{A}}}(u) := \max_{\alpha\in\tilde{\mathcal{A}}} \Big\{ \Big(g(x(1;\alpha,u),\alpha) - \max_{\alpha'\in\mathcal{A}} g(\bar{x}(1;\alpha'),\alpha') + \epsilon_i^2 \Big) \vee d_{C(\alpha)}(x(1;\alpha,u)) \Big\}. \tag{6.2}$$

Take $\rho > 0$. According to Lemma 6.1, the finite subset $\tilde{\mathcal{A}}$ can be chosen such that

$$J_i^{\tilde{\mathcal{A}}}(u) \;\geq\; J_i(u) - \rho^2 \quad \text{for all } u \in \mathcal{U}.$$

Since v_i is a minimizer for $u \to J_i(u) + \epsilon_i\Delta(v_i,u)$ over $\mathcal{U}$, it follows that v_i is a ρ^2-minimizer for $u \to J_i^{\tilde{\mathcal{A}}}(u) + \epsilon_i\Delta(v_i,u)$ over $\mathcal{U}$. A second application of Ekeland's theorem then yields a control function $u_i \in \mathcal{U}$ such that

$$\Delta(v_i,u_i) \leq \rho$$

and

$$J_i^{\tilde{\mathcal{A}}}(u_i) + \epsilon_i\Delta(v_i,u_i) + \rho\Delta(u_i,u_i) \;=\; \min_{u\in\mathcal{U}} \Big\{ J_i^{\tilde{\mathcal{A}}}(u) + \epsilon_i\Delta(v_i,u) + \rho\Delta(u_i,u) \Big\}.$$

By adding extra elements to the finite subset $\widetilde{A}$ and reducing ρ if necessary, we can make the number $|J_i^{\tilde{\mathcal{A}}}(u_i) - J_i(v_i)|$ arbitrary small. (See Lemma 6.1.) Since $J_i(v_i) > 0$, we can arrange that

$$J_i^{\tilde{\mathcal{A}}}(u_i) > 0.$$

Write $\mathcal{A}^i$ in place of $\widetilde{\mathcal{A}}$ and ρ_i in place of ρ, to emphasize their dependence on i.

We can carry out the above analysis for $i = 1, 2, \ldots$. By adding extra elements to each $\mathcal{A}^i$ and reducing each ρ_i, if necessary, we can arrange, also, that $\{\mathcal{A}^i\}$ is an increasing sequence and $\rho_i \downarrow 0$.

For clarity, we summarize relevant properties of the above constructs: for some sequences $\epsilon_i \downarrow 0$ and $\rho_i \downarrow 0$, sequences $\{u_i\}$ and $\{v_i\}$ in $\mathcal{U}$ and an increasing sequence of finite subsets $\{\mathcal{A}^i\}$ of $\mathcal{A}$, we have

(i) $J_i^{\mathcal{A}^i}(u_i) + \epsilon_i \Delta(v_i, u_i) + \rho_i \Delta(u_i, u_i)$
$= \min_{u \in \mathcal{U}} \left\{ J_i^{\mathcal{A}^i}(u) + \epsilon_i \Delta(v_i, u) + \rho_i \Delta(u_i, u) \right\}$ for all i,

(ii) $J_i^{\mathcal{A}^i}(u_i) > 0$ for all i,

(iii) $\Delta(v_i, \bar{u}) \to 0$ and $\Delta(u_i, \bar{u}) \to 0$ as $i \to \infty$.

For each i, list the elements in $\mathcal{A}^i$,

$$\mathcal{A} = \{\alpha_1, \ldots, \alpha_{K_i}\}$$

and write $\{x_i(.;\alpha) \,|\, \alpha \in \mathcal{A}\}$ for the state trajectories corresponding to u_i. Define

$$m_i(t,u) := \begin{cases} 0 & \text{if } \; u = v_i(t), \\ 1 & \text{otherwise,} \end{cases} \quad \text{and} \quad n_i(t,u) := \begin{cases} 0 & \text{if } \; u = u_i(t), \\ 1 & \text{otherwise.} \end{cases}$$

With the help of these functions, we can express the minimizing property (i) of the u_i's in control theoretic terms, as follows. For each i, $(u_i, \{x_i(.;\alpha_k) \mid k = 1, \ldots, K_i\})$ is a minimizer for the optimal control problem

$$(\mathrm{P}_i) \begin{cases} \text{Minimize } \max_{k=1,\ldots,K_i} \Big\{ \Big(g(x(1;\alpha_k), \alpha_k) \\ \qquad\qquad - \max_{\alpha \in \mathcal{A}} \; g(\bar{x}(1;\alpha), \alpha) + \epsilon_i^2 \Big) \vee d_{C(\alpha_k)}(x(1;\alpha_k)) \Big\} \\ \qquad\qquad + \epsilon_i \int_0^1 m_i(t, u(t))\, dt + \rho_i \int_0^1 n_i(t, u(t))\, dt \\ \text{over measurable functions } u \text{ and arcs } \{x(.;\alpha_1), \ldots, x(.;\alpha_{K_i})\} \text{ such that} \\ u(t) \in \Omega(t), \quad a.e. \; t \in [0,1] \\ \text{and, for } k = 1, \ldots, K_i, \\ \dot{x}(t;\alpha_k) = f(t, x(t;\alpha_k), u(t), \alpha_k), \quad a.e. \; t \in [0,1], \\ x(0;\alpha_k)(0) = x_0. \end{cases}$$

Since $u_i \to \bar{u}$ and $v_i \to \bar{u}$ with respect to the Δ-metric, we know that

$$\sup_{\alpha \in \mathcal{A}} \|\bar{x}(.;\alpha) - x(.;\alpha, u_i)\|_C \;\to\; 0 \qquad \text{as} \quad i \to \infty.$$

Take an infinite sequence of control functions $\{\hat{u}_j\} \in \mathcal{U}$ whose first element is $\bar{u}$. Using similar reasoning to that employed in the proof of Proposition 2.1 (note the crucial role of property (ii) above, to ensure multiplier nondegeneracy), we can deduce the following information from the nonsmooth maximum principle (see, e.g., [10], Theorem 6.2.1). For each i sufficiently large, there exist nonnegative numbers $\lambda_1^i, \ldots, \lambda_{K_i}^i$ such that

$$\sum_{k=1}^{K_i} \lambda_k^i \;=\; 1,$$

and a sequence $\epsilon_i' \downarrow 0$ with the following properties. Define the discrete probability measure

$$\Lambda_i = \sum_{k=1}^{K_i} \lambda_k^i \delta_{\alpha_k^i}.$$

Then, for each i sufficiently large and $\Lambda_i - a.e.\ \alpha \in \mathcal{A}$, there exists a costate arc $p_i(.;\alpha)$

satisfying

$$\begin{aligned}
&\text{(i)} -\dot{p}_i(t;\alpha) \in \operatorname{co} \partial_x H(t, \bar{x}(t) + \epsilon_i^{'} B, \bar{u}(t), p_i(t;\alpha), \alpha) \\
&\text{(ii)} -p_i(1;\alpha) \in \overline{co} \bigcup_{x \in \bar{x}(1;\alpha)+\epsilon_i' B} \; \bigcup_{r \in [0,1]} \big(r G_{\epsilon_i'}(x,\alpha) + (1-r) N(x,\alpha) \big)
\end{aligned}$$

where $G_{\epsilon'}(\alpha, x) =$

$$\begin{cases} \partial_x g(x,\alpha) & \text{if } g(x,\alpha) \geq \max_{\alpha' \in \mathcal{A}} g(x,\alpha') - \epsilon' \\ \emptyset & \text{otherwise} \end{cases}$$

and $N(x,\alpha) = \big\{ \xi \in N_{C(\alpha)}(x) \,|\, |\xi| = 1 \big\}$

$$\text{(iii)} \int_{\mathcal{A}} \int_0^1 p_i(t;\alpha) \cdot [f(t, \bar{x}(t;\alpha), \hat{u}_j(t), \alpha) - f(t, \bar{x}(t;\alpha), \bar{u}(t), \alpha)] \; dt \; \Lambda_i(d\alpha) \leq \epsilon_i'$$

for $j = 1, 2, \ldots$.

By extracting a subsequence, we can arrange that

$$\Lambda_i \to \Lambda \qquad \text{weakly}^* \qquad \text{as} \quad i \to \infty$$

for some Radon probability measure Λ on the Borel sets of $\mathcal{A}$.

Fix an integer N. We now apply the first part of Proposition 6.1, in which we identify μ with Λ, μ_i with the Λ_i, and take

$$D_i(\alpha) := \{(\xi_1, \ldots, \xi_N) \in R^N \,|\, \exists\, p(.;\alpha) \in Q_{\epsilon_i'}(\alpha) \ s.t.\ \xi_j = w_j(p(.;\alpha), \alpha) \text{ for } j{=}1, 2, \ldots, N\},$$

$i = 1, 2, \ldots,$ and

(6.3)

$$D(\alpha) := \{(\xi_1, \ldots, \xi_N) \,|\, \exists\, p(.;\alpha) \in \overline{Q}_0(\alpha) \ s.t.\ \xi_j = w_j(p(.;\alpha), \alpha),\ j = 1, 2, \ldots, N\}.$$

Here,

$$w_j(p(.), \alpha) := \int_0^1 p(t) \cdot [f(t, \bar{x}(t,\alpha), \hat{u}_j(t), \alpha) - f(t, \bar{x}(t,\alpha), \bar{u}(t), \alpha)] \; dt.$$

We deduce that

$$\int_{\mathcal{A}} \int_0^1 q_N(t,\alpha) \cdot [f(t, \bar{x}(t,\alpha), \hat{u}_j(t), \alpha) - f(t, \bar{x}(t,\alpha), \bar{u}(t), \alpha)] \; dt \; \Lambda(d\alpha) \leq 0 \tag{6.4}$$

for $j = 1, 2, \ldots, N$, in which $\{q_N(.;\alpha) \in W^{1,1} \,|\, \alpha \in \mathcal{A}\}$ is some family of arcs such that, for $\Lambda - a.e.\ \alpha \in \mathcal{A}$,

$$q_N(.;\alpha) \in \overline{Q}_0(\alpha).$$

For each N, we can regard $\alpha \to q_N(.;\alpha)$ as a representative of an equivalence class of $\Lambda - a.e.$ equal elements in the Hilbert space

$$\mathcal{X} := L^2_\Lambda(\mathcal{A}; L^2([0,1]; R^n))$$

with the inner product

$$(p,q)_\Lambda = \int_{\mathcal{A}} \int_0^1 p(t;\alpha) \cdot q(t;\alpha) \; dt \; \Lambda(d\alpha).$$

The sequence $\{\alpha \to q_N(.;\alpha)\}_{N=1}^{\infty}$ is norm bounded and therefore has a weak limit, which we write $\{\alpha \to p(.;\alpha)\}$. But

$$\{d \in \mathcal{X} \mid d(\alpha) \in \overline{Q}_0(\alpha),\, \Lambda - a.e. \quad \alpha \in \mathcal{A}\}$$

is a strongly closed subset of $\mathcal{X}$. Since it is convex, it is also weakly closed. It follows that

$$p(.;\alpha) \in \overline{Q}_0(\alpha) \quad \Lambda - a.e.\ \alpha \in \mathcal{A}.$$

By weak convergence, we deduce from (6.4) that

$$\int\int_0^1 p(t;\alpha)\cdot[f(t,\bar{x}(t;\alpha),\hat{u}_j(t),\alpha) - f(t,\bar{x}(t;\alpha),\bar{u}(t),\alpha)]\ dt\ \Lambda(d\alpha) \le 0 \tag{6.5}$$

for $j = 1, 2, \ldots$.

In view of the Castaing representation theorem (see, e.g., [10], Theorem 2.2.7), we can choose a subset $T \subset (0,1)$ of full measure and also the sequence of controls functions above, $\{\hat{u}_j\}$, to satisfy

$$\overline{\bigcup_j \int_{\mathcal{A}} p(t;\alpha)\cdot f(t,\bar{x}(t;\alpha),\hat{u}_j(t),\alpha)\Lambda(d\alpha)} \supset \int_{\mathcal{A}} p(t;\alpha)\cdot f(t,\bar{x}(t;\alpha),\Omega(t),\alpha)\Lambda(d\alpha) \tag{6.6}$$

for all $t \in T$. We can arrange that (6.5) remains valid when the countable set $\hat{u}_j(.)$ is replaced by another countable set comprising all concatenations of a finite number of segments of the original $\hat{u}_j$'s, with junction points belonging to a countable dense subset S of $[0,1]$.

Define T' to be the set of full measure, comprising points in T which are also Lebesgue points for

$$s \to \int p(s;\alpha)[f(s,\bar{x}(s,\alpha),\hat{u}_j(s),\alpha) - f(s,\bar{x}(s,\alpha),\bar{u}(s),\alpha)]\ \Lambda(d\alpha) \tag{6.7}$$

for all j. Take any $t \in T'$, $w \in \Omega(t)$ and $\beta > 0$, Then, in view of (6.6), there exists j such that

$$\int_{\mathcal{A}} p(t;\alpha)\cdot f(t,\bar{x}(t),\hat{u}_j(t),\alpha)\Lambda(d\alpha)\ \ge\ \int_{\mathcal{A}} p(t;\alpha)\cdot f(t,\bar{x}(t;\alpha),w,\alpha)\Lambda(d\alpha) - \beta. \tag{6.8}$$

Choose a sequence of intervals $\{[s_i,t_i]\}$, containing t and with endpoints in the set S and such that $s_i \to t$ and $t_i \to t$. Now let $v_i \in \{\hat{u}_j\}_{j=1}^{\infty}$ for $i = 1, 2, \ldots$, where

$$v_i := \begin{cases} \hat{u}_j(t) & \text{if } t \in [s_i,t_i], \\ \bar{u}(t) & \text{otherwise.} \end{cases}$$

Changing the order of integration, inserting $\hat{u}_j = v_i$ in (6.5) and dividing across by $|t_i - s_i|$ gives

$$\frac{1}{|t_i - s_i|}\int_{s_i}^{t_i}\int p(s;\alpha)\cdot[f(s,\bar{x}(s;\alpha),\hat{u}_j(s),\alpha) - f(s,\bar{x}(s;\alpha),\bar{u}(s),\alpha)]\ \Lambda(d\alpha)\ dt \le 0$$

for each i. Since t is a Lebesgue point of the mapping (6.7), it follows that

$$\int p(t;\alpha)\cdot[f(t,\bar{x}(t;\alpha),\hat{u}_j(t),\alpha)-f(t,\bar{x}(t;\alpha),\bar{u}(t),\alpha)]\ \Lambda(d\alpha)\le 0.$$

We conclude from (6.8) that

$$\int p(t;\alpha)\cdot[f(s,\bar{x}(t;\alpha),w,\alpha)-f(t,\bar{x}(t;\alpha),\bar{u}(t),\alpha)]\ \Lambda(d\alpha)\le \beta.$$

But $\beta>0$ is arbitrary. So

$$\int p(t;\alpha)\cdot[f(t,\bar{x}(t;\alpha),w,\alpha)-f(t,\bar{x}(t;\alpha),\bar{u}(t),\alpha)]\ \Lambda(d\alpha)\le 0.$$

Since the above inequality holds for any $t'\in T'$, a set of full measure, and any $w\in\Omega(t)$, the maximization of the Hamiltonian condition is confirmed. The proof is complete.

6.2. Proof of Theorem 3.2. The assertions of Theorem 3.1 are expressed in terms of selectors $p(.;\alpha)$ of the multifunction

$$\alpha\to\overline{Q}_0(\alpha)$$

in order to guarantee that $D(.)$, given by (6.3), has closed graph and convex values, and thereby to justify application of part (i) of Proposition 6.1.

In the case when $\mathcal{A}$ can be decomposed into disjoint sets $\mathcal{A}=A^{(1)}\cup A^{(2)}$ in which $A^{(2)}$ is finite, essentially the same analysis leads to optimality conditions involving a selector $p(.;\alpha)$ of the multifunction

$$\alpha\to\begin{cases}\overline{Q}_0(\alpha) & \text{if }\ \alpha\in A^{(1)},\\ Q_0(\alpha) & \text{if }\ \alpha\in A^{(2)}.\end{cases} \tag{6.9}$$

We do, however, now have to use part (ii) of Proposition 6.1 to justify (6.4), for some selector $p_N(.;\alpha)$ of the multifunction (6.9).

Also, to justify (6.5) (for some selector $p_N(.;\alpha)$ of (6.9)), we must use the facts that, if $A^{(2)}=\{b_1,\dots,b_m\}$, then an element in

$$\mathcal{X}=L^2_\Lambda(\mathcal{A};L^2([0,1];R^n))$$

can be represented by an element in

$$\mathcal{X}'=L^2_\Lambda(\mathcal{A}^{(1)};L^2([0,1];R^n))\times L^2([0,1];R^n)^m,$$

and the weak topology on $\mathcal{X}$ is compatible with the weak product topology on $\mathcal{X}'$. It follows that, for the sequence $\{\alpha\to p_N(.;\alpha)\}_{N=1}^\infty$ constructed at the end of the proof of Theorem 3.1, we can arrange by subsequence extraction, that the limiting $p(.;\alpha)$ satisfies $p(.;\alpha)\in Q_0(\alpha)$ for $\Lambda-a.e.\ \alpha\in A^{(2)}$.

6.3. Proof of Theorem 3.3. The proof the minimax maximum principle for problems with functional inequality endpoint constraints is along similar, but simpler, lines to that of Example 4.1. The main difference is that, for each i, we replace the cost function $J_i^{\tilde{\mathcal{A}}}(u)$ (see (6.2)) of the earlier analysis by

$$\begin{aligned}\tilde{J}_i^{\tilde{\mathcal{A}}}(u)\ :=\ \max_{\alpha\in\tilde{\mathcal{A}}}\Big\{&\Big(g(x(1;\alpha,u),\alpha)-\max_{\alpha'\in\mathcal{A}}\ g(\bar{x}(1;\alpha'),\alpha')\Big)\\ &\vee\psi^1(\bar{x}(1;\alpha,u),\alpha,)\vee\cdots\vee\psi^r(\bar{x}(1;\alpha),\alpha)\Big\}.\end{aligned} \tag{6.10}$$

The proof is completed by examining properties of minimizers of perturbations of these cost functions and passage to the limit as before.

Acknowledgment. The assistance of the reviewers in improving earlier versions of this paper is gratefully acknowledged.

REFERENCES

[1] M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi Equations*, Birkhaüser, Boston, 1997.

[2] V. G. Boltyansky and A. S. Poznyak, *Robust maximum principle in minimax control*, Internat. J. Control, 72 (1999), pp. 305–314.

[3] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York; reprinted as Classics in Applied Mathematics 5, SIAM, Philadelphia, 1990.

[4] N. N. Krasovskii and A. I. Subbotin, *Game Theoretic Control Problems*, Springer-Verlag, Berlin, 1988.

[5] A. Ya. Dubovitskii and A. A. Milyutin, *Problems for extremum under constraints*, Zh. Vychislit. Math. i Math. Piz., 5 (1965), pp. 395–453; English translation, U.S.S.R. Comput. Math. and Math. Physics, 5 (1965).

[6] M. Morari and E. Zafiriou, *Robust Process Control*, Prentice Hall, Englewood Cliffs, NJ, 1989.

[7] B. S. Mordukhovich, *Generalized differential calculus for nonsmooth and set-valued mappings*, J. Math. Anal. Appl., 183 (1994), pp. 250–288.

[8] E. J. Polak, *Optimization: Algorithms and Consistent Approximations*, Springer-Verlag, New York, 1997.

[9] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Grundlehren der Mathematischen Wissenschaft 317, Springer-Verlag, Berlin, 1998.

[10] R. B. Vinter, *Optimal Control*, Birkhaüser, Boston, 2000.

[11] J. Warga, *Optimal Control of Differential and Functional Equations*, Academic Press, New York, 1972.