SENSITIVITY INTERPRETATIONS OF THE COSTATE VARIABLE FOR OPTIMAL CONTROL PROBLEMS WITH STATE CONSTRAINTS

PIERNICOLA BETTIOL† AND RICHARD B. VINTER†

Abstract. In optimal control theory, it is well known that the costate arc and the associated maximized Hamiltonian function can be interpreted in terms of gradients of the value function, evaluated along the optimal state trajectory. Such relations have been referred to as “sensitivity relations” in the literature. We provide in this paper new sensitivity relations for state constrained optimal control problems. For the class of optimal control problems considered, there is no guarantee that the costate arc is unique; a key feature of the results is that they assert some choice of costate arc can be made for which the sensitivity relations are valid. The proof technique is to introduce an auxiliary optimal control problem that possesses a richer set of control variables than the original problem. The introduction of the additional control variables in effect enlarges the class of variations with respect to which the state trajectory under consideration is a minimizer; the extra information thereby obtained yields the desired set of sensitivity relations.

1. Introduction. This paper concerns the optimal control problem with state constraints:

\[
P_{S,x_0} = \begin{cases} 
\text{Minimize } g(x(T)) \\
\text{over arcs } x(\cdot) \in W^{1,1}([S,T]; R^n) \\
\text{and measurable functions } u(\cdot): [S,T] \to R^m \text{ s.t.} \\
\dot{x}(t) = f(t,x(t),u(t)) \text{ a.e. } t \in [S,T], \\
u(t) \in U(t) \text{ a.e. } t \in [S,T], \\
x(t) \in A(t) \text{ for all } t \in [S,T], \\
x(S) = x_0,
\end{cases}
\]

expressed in terms of the data: integers \(n\) and \(m\), an interval \([S,T]\), functions \(g: R^n \to R\) and \(f: R \times R^n \times R^m \to R\), a vector \(x_0 \in R^n\), and multifunctions \(U: [S,T] \to R^m\) and \(A: [S,T] \to R^n\). Here, \(W^{1,1}([S,T]; R^n)\) is the space of absolutely continuous functions on \([S,T]\), whose derivatives are integrable.

It is assumed that the time-dependent “state constraint” set \(A(t)\) has the functional inequality representation

\[ A(t) = \{x \mid h(t,x) \leq 0\} \]

Received by the editors August 12, 2008; accepted for publication (in revised form) September 4, 2009; published electronically January 29, 2010.

http://www.siam.org/journals/sicon/48-5/73261.html

†Department of Electrical and Electronic Engineering, Imperial College London, Exhibition Road, London SW7 2BT, UK (p.bettiol@imperial.ac.uk, r.vinter@imperial.ac.uk).
for some function $h : R \times R^n \to R$. Let $(\bar{x}, \bar{u})$ be a minimizer for $P_{S_0}$.  

For any $(t, x) \in [S, T] \times R^n$, denote by $P_{t,x}$ the modification of $P_{S_0}$, in which the “initial data" $(t, x)$ replaces $(S, x_0)$. We refer to a measurable function $u : [t, T] \to R^n$ that satisfies $u(s) \in U(s)$ a.e. as a control function on $[t, T]$. A pair $(x(\cdot), u(\cdot))$ comprising an absolutely continuous $R^n$ valued function $x(\cdot)$ and a control function $u(\cdot)$ on $[t, T]$ that satisfy $\dot{x}(s) = f(s, x(s), u(s))$ a.e. is called a process on $[t, T]$. The first component of a process is called a state trajectory. A process on $[t, T]$ that satisfies the constraints of problem $P_{t,x}$ is said to be an admissible process for $P_{t,x}$.

The value function $V : [S, T] \times R^n \to R \cup \{+\infty\}$ is the function

$$V(t, x) = \inf_{P_{t,x}}$$

where the right-hand side is interpreted as the infimum cost (under the hypotheses we shall impose it cannot be $-\infty$) in the case where admissible processes for $P_{t,x}$ exist and as $+\infty$ otherwise.

Denote by $H : [S, T] \times R^n \times R^n \times R^n \to R$ the Hamiltonian function

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

and by $\mathcal{H} : [S, T] \times R^n \times R^n \to R$ the maximized Hamiltonian

$$\mathcal{H}(t, x, p) = \sup_{u \in U(t)} H(t, x, p, u).$$

In this paper we derive new relations between the costate arc of the state constrained maximum principle and subgradients of the value function. With a view to briefly reviewing previous work, we specialize, for the time being, to the case of no state constraints, i.e., $A(t) = R^n$ for all $t \in [S, T]$. (This corresponds to the choice $h(\cdot, \cdot) \equiv -1$.)

Under the hypotheses listed in section 3 (and when $A(t) = R^n$ for all $t \in [S, T]$), we know the following.

Maximum principle. There exists an absolutely continuous function $p(\cdot) : [S, T] \to R^n$ such that

$$-\dot{p}(t) \in \text{co} \partial_x H(\bar{x}(t), p(t), \bar{u}(t)) \quad \text{a.e.,}$$

$$H(t, \bar{x}(t), p(t), \bar{u}(t)) = \mathcal{H}(t, \bar{x}(t), p(t)) \quad \text{a.e.,}$$

$$-p(T) \in \partial g(\bar{x}(T)).$$

Here $\partial g$ denotes the subdifferential of $g$ (see section 2), and $\partial_x H$ denotes the subdifferential of $H(t, x, u)$ with respect to the $x$ variable. In these relations co denotes “convex hull.”

Notice that the “cost multiplier" is here taken to be 1, i.e., the maximum principle is in normal form; this is permissible under the hypotheses.

If $f$ and $g$ are continuously differentiable in the $x$ variable, if $V$ is continuously differentiable on $(S, T) \times R^n$, and if $\bar{u}$ is piecewise continuous, it is known that $V$ is related to the costate function and maximized Hamiltonian evaluated along $\bar{x}$ and $p$ according to

$$(\mathcal{H}(t, \bar{x}(t), p(t)), -p(t)) = \nabla V(t, \bar{x}(t)) \quad \text{a.e. } t \in [S, T].$$

These relations follow, formally at least, from the Hamilton–Jacobi equation (smooth form) when we identify $t \to V_{\bar{x}}(t, \bar{x}(t))$ with the costate arc $p(\cdot)$. They date from the
early days of optimal control theory and have been described as providing a “sensitivity” interpretation of the maximum principle Lagrange “multipliers.” (See, for example, [2].) They are of interest because they tell us that the Pontryagin maximum principle can be used, not only to solve optimal control problems, but to supply first order information about how the minimum cost is affected by perturbations to the problem data.

For many optimal control problems of interest, the value function fails to be continuously differentiable. Under unrestrictive hypotheses on the problem data, however, the value function can be shown to be a (possibly nondifferentiable) lower semicontinuous function. So, if the sensitivity relations are to be validated in conditions of any generality, they must be couched in terms of “nonsmooth” subdifferentials, for example, as

\[
\partial_{\bar{x}}V(t,\bar{x}(t)) \in \co \partial V(t,\bar{x}(t)) \quad \text{a.e. } t \in [S,T].
\]

In the case where \( f \) and \( g \) are continuously differentiable in the \( x \) variable and \( t \to \mathcal{H}(t,\bar{x}(t),p(t)) \) is continuous on \((S,T)\), (2) follows from the maximum principle for free time optimal control problems. Indeed, it is a simple matter to show that, for each \( t \in (S,T) \), the restriction of \((\bar{x},\bar{u})\) to \([t,T]\) is optimal for the free time optimal control problem

\[
\begin{aligned}
\tilde{P} \quad \text{(Minimize } g(x(T)) - V(t,x(t)) \text{ over } t \in (S,T) \text{ arcs } x(\cdot) \in W^{1,1}([t,T];\mathbb{R}^n) \\
\text{and measurable functions } u(\cdot) : [t,T] \to \mathbb{R}^n \text{ s.t.} \\
x(s) = f(t,x(s),u(s)) \quad \text{a.e. } s \in [t,T] \\
u(s) \in U(s) \quad \text{a.e. } s \in [t,T].
\end{aligned}
\]

If \( V \) is locally Lipschitz continuous, then the free time Pontryagin maximum principle tells us (via the left end point transversality condition) that

\[
(\mathcal{H}(t,\bar{x}(t),p(t)), -p(t)) \in -\partial(-V(t,\bar{x}(t))) \subset \co \partial V(t,\bar{x}(t)),
\]

since subdifferentials of locally Lipschitz functions satisfy the relation \( \co \partial (-V) = -\co \partial V \). Here, the costate arc \( p \) is the solution to

\[
\begin{aligned}
-\dot{p}(t) &= f_x^T(t,\bar{x}(t),\bar{u}(t)) p(t) \quad \text{a.e. } t \in [S,T], \\
-p(T) &= g_x(x(T)).
\end{aligned}
\]

Since, however, \( p \) must be the costate arc for \( P_{S,x_0} \), by uniqueness of solutions to (3), the costate satisfies (2) at all times \( t \in (S,T) \).

If we no longer suppose that \( f \) and \( g \) are continuously differentiable in the \( x \) variable, the costate inclusion may have multiple solutions satisfying the maximization of the Hamiltonian condition and the transversality condition. In these circumstances it is natural to ask whether there exists some costate arc that satisfies (2). The validity of the related \( x \)-gradient sensitivity relation

\[
-\dot{p}(t) \in \co \partial_x V(t,\bar{x}(t)) \quad \text{a.e. } t \in [S,T]
\]

(for the nonsmooth, state constraint-free case) was proved by Clarke and Vinter [10] and the full sensitivity relation (2) was proved by Vinter in [21] (see also [23]). Examples are also available (see [20]) showing that, in some cases, there are a number of possible choices of costate arcs associate with \( P_{S,x_0} \), but only one of them satisfies
(2). We observe that the techniques outlined above for deriving sensitivity relations do not extend to cover situations in which \( f \) and \( g \) are continuously differentiable in the \( x \) variable, since solutions to the costate differential inclusion replacing (3) are no longer unique. In papers [10] and [21] the authors make use of a different approach based on showing that the optimal process under consideration is a minimizer also for a certain auxiliary problem, involving a “richer” set of control variables than the original optimal control problem. This extra information about the optimal process can be exploited to prove the sensitivity relations.

Now reintroduce the state constraint. A version of the state constrained maximum principle suitable for sensitivity analysis, valid under the hypotheses imposed in section 3, is as follows.

**State constrained maximum principle.** There exist \( p \in BV([S,T]; \mathbb{R}^n) \) and \( \mu \in NBV^+(S,T) \) such that

\[
\begin{align*}
-\text{d}p(t) & \in \text{co} \partial_x H(t, \bar{x}(t), p(t), \bar{u}(t))dt - \nabla_x h(t, \bar{x}(t))d\mu(t) \quad \text{a.e.,} \\
H(t, \bar{x}(t), p(t), \bar{u}(t)) & = \mathcal{H}(t, \bar{x}(t), p(t)), \\
\text{supp} \{\mu\} & \subset \{t \mid h(t, \bar{x}(t)) = 0\}, \\
-p(T) & = \partial g(\bar{x}(T)).
\end{align*}
\]

Here \( BV([S,T]; \mathbb{R}^n) \) denotes the space of \( \mathbb{R}^n \) valued functions of bounded variation \([S,T], \) which are right continuous on \((S,T)\). \( NBV^+(S,T) \) is the set of nondecreasing functions \( \mu \) in \( BV([S,T]; \mathbb{R}) \), right continuous on \((S,T)\), such that \( \mu(S) = 0 \). “supp \( \mu \)” denotes the support of the measure on the Borel subsets of \([S,T]\) induced by \( \mu \).

The differential inclusion (5) is interpreted as an integral equation: there exists an integrable function \( \xi : [S,T] \rightarrow \mathbb{R}^n \) such that

\[
\xi(t) \in \text{co} \partial_x H(t, \bar{x}(t), p(t), \bar{u}(t)) \quad \text{a.e. } t \in [S,T],
\]

\[
-p(t) = -p(S) + \int_{[S,t]} (\xi(s)ds - \nabla_x h(s, \bar{x}(s)))d\mu(s) \quad \text{for all } t \in (S,T).
\]

This is a form of the condition used (in the smooth case) by Ioffe and Tihomirov [15]. We refer to the function \( p \) as the true costate arc. In the necessary conditions literature, the optimality condition is more frequently expressed in terms of an absolutely continuous “pseudo costate” arc \( q \) satisfying \( q(S) = p(S) \) and

\[
q(t) = p(t) - \int_{[S,t]} \nabla_x h(s, \bar{x}(s))d\mu(s) \quad \text{if } t \in (S,T]
\]

because \( q \) is absolutely continuous and satisfies a simple differential inclusion, namely,

\[
-\dot{q}(t) \in \text{co} \partial_x \mathcal{H} \left( t, \bar{x}(t), q(t) + \int_{[S,t]} \nabla_x h(t, \bar{x}(t))d\mu(t) \right) \quad \text{a.e. } t \in [S,T].
\]

Maximum principles expressed in terms of the true costate \( p(.) \) or the pseudo costate arc \( q(.) \) convey the same information about optimal controls. However, it is natural to express sensitivity relations in terms of the true costate arc \( p \) because (according to formal calculations) \( p \), unlike \( q \), can be interpreted as the Lagrange multiplier associated with the dynamic constraint \( \dot{x} = f(t, x, u) \) in \( P_{t,x_0} \). (See the discussion in [20].)
In this paper, we prove a version of both sensitivity relations (2) and (4) involving a true costate arc for state constrained problems. There may be a number of possible costate arcs; we show that one of them can be chosen to satisfy the relations.

We have earlier observed that, for smooth functions $f$ and $g$, an analysis based on the uniqueness of solutions to the costate differential equation may be employed to prove sensitivity relations in the case of no state constraints. This simple analysis is no
longer available to us when state constraints are present. Indeed, fixing \( t \in [S,T] \) and applying the state constrained maximum principle to \((\bar{x}, \bar{u})|_{[t,T]}\), viewed as a solution to the state constrained version of \( \dot{P} \), yields a costate function \( p_t(.) \) on \([t,T]\) (and an associated state constraint “multiplier” \( \mu(.) \) on \([t,T]\) such that, for all \( t \in (S,T) \),
\[
(\mathcal{H}(t, \bar{x}(t), p_t(t)), -p_t(t)) \in \partial^* V(t, \bar{x}(t)),
\]
where \( \partial^* V \) is the “superdifferential” \( \partial^* V = -\partial(\Psi_{\text{Graph}} A(s, x) - V(s, x))|_{(s,x)=(t,\bar{x}(t))} \). \( \Psi_D \) denotes the indicator function of the set \( D \), defined below.) But the state constraint multipliers \( \mu_t(.) \), for each \( t \), are not in general unique. So there is no guarantee that each \( p_t(.) \) can be regarded as the restriction to \([t,T]\) of a single costate function \( p \) for \( P_{S,x_0} \), satisfying the sensitivity relation \((2)\) over the entire time interval \([t,T]\).
For this reason, we employ quite different proof techniques, based on the construction of auxiliary optimal control problems.

Cernea and Frankowska [6] and [7] have earlier investigated sensitivity relations for state constrained optimal control problems as part of a broad study which addresses also the question of when the state constrained maximum principle is valid in normal form. Different hypotheses are imposed on the state constraint sets \( A(t) \), and different kinds of subgradients are employed to those of this paper. The key differences are as follows. First and most important, we express sensitivity relations along the optimal state trajectory in terms of a single costate function for \( P_{S,x_0} \), whereas [6] and [7] provide conditions in the spirit of \((10)\), which are expressed in terms a family of costate arcs associated with subintervals \([t, T]\) \( \subset [S,T] \) that do not necessarily combine to generate a single costate arc for \( P_{S,x_0} \). Second, we allow the state constraint to depend on time. Finally, we express the sensitivity relation simply in terms of limits of subgradients evaluated at points in the interior of the state constraint set, in place of the superdifferential \( \partial^* V \).

**Notation.** In Euclidean space, the length of a vector \( x \) is denoted by \(|x|\) and the closed unit ball \( \{x : |x| \leq 1\} \) by \( B \). The graph of a multifunction \( U : [S,T] \rightarrow \mathbb{R}^n \) is denoted by \( \text{Graph} \ U \). Take an extended-valued function \( f : \mathbb{R}^k \rightarrow \mathbb{R} \cup \{+\infty\} \). Then the effective domain of \( f \) is the set \( \text{dom} \ f := \{x : f(x) < +\infty\} \). The epigraph of \( f \) is the set \( \text{epi} f := \{(r, x) \in \mathbb{R}^{k+1} : r \geq f(x)\} \). Now take a set \( D \subset \mathbb{R}^k \). The indicator function of the set \( D \), \( \Psi_D \), is the extended valued function on \( \mathbb{R}^k \) taking value 0 on the set \( D \) and \( +\infty \) on its complement. We write \( z_i \xrightarrow{D} z \) to describe a sequence \( \{z_i\} \) converging to \( z \) such that \( z_i \in D \) for all \( i \). \( d_D(x) \) denotes the Euclidean distance of the point \( x \) from the set \( D \), namely, \( \min\{|x - x'| | x' \in D\} \).

**2. Subgradients.** This paper makes use of numerous constructs that generalize the concepts of normal cone, gradients, and direction derivatives to apply in a non-smooth setting. We gather together in this section definitions and properties relevant to our analysis. All the material in the section is standard and may be found in a number of texts, examples of which include [1], [8], [9], [16], [19], and [20].

**Definition 2.1.** Take a closed set \( D \subset \mathbb{R}^k \) and a point \( \bar{x} \in D \). The normal cone \( N_D(\bar{x}) \) of \( D \) at \( \bar{x} \) is defined to be
\[
N_D(\bar{x}) := \{ p : \exists x_i \xrightarrow{D} \bar{x}, \ p_i \xrightarrow{p} p \ \text{s.t.} \ \limsup_{x \downarrow x_i} \frac{p_i \cdot (x - x_i)}{|x - x_i|} \leq 0 \quad \text{for all} \ i \in \mathbb{N}\}.
\]

**Definition 2.2.** Take an open set \( \mathcal{O} \subset \mathbb{R}^k \), a lower semicontinuous function \( f : \mathcal{O} \rightarrow \mathbb{R} \cup \{+\infty\} \), and a point \( \bar{x} \in \mathcal{O} \) such that \( \bar{x} \in \text{dom} \ f \).
(i) The subdifferential of \( f \) at \( \bar{x} \in \text{dom} f \) is
\[
\partial f(\bar{x}) = \left\{ \xi \mid \exists \xi_i \to \xi \text{ and } x_i \xrightarrow{\text{dom} f} \bar{x} \text{ such that}
\frac{\limsup_{x \to x_i} \xi_i \cdot (x - x_i) - f(x) + f(x_i)}{|x - x_i|} \leq 0 \text{ for all } i \in \mathbb{N} \right\}.
\]

(ii) Given \( d \in \mathbb{R}^k \), the lower Dini derivative \( D^i f(\bar{x}; d) \) of \( f \) at \( \bar{x} \) in the direction \( d \) is
\[
D^i f(\bar{x}; d) = \liminf_{h \downarrow 0, e \to d} h^{-1} [f(\bar{x} + he) - f(\bar{x})].
\]

(iii) Given \( d \in \mathbb{R}^k \), the Clarke Directional Derivative \( D^0 f(\bar{x}; d) \) of \( f \) at \( \bar{x} \) in the direction \( d \) is
\[
D^0 f(\bar{x}; d) = \limsup_{h \downarrow 0, z \xrightarrow{\text{dom} f} \bar{x}} h^{-1} [f(z + hd) - f(z)].
\]

Given a function \( v(\cdot, \cdot) \) of two vector variables \((x, y)\) and a point \((\bar{x}, \bar{y}) \in \text{dom} v(\cdot, \cdot)\), we denote the subgradient of \( v(\cdot, \cdot) \) at \((\bar{x}, \bar{y})\) by \( \partial_x v(\bar{x}, \bar{y}) \).

We recall that the subdifferential of a lower semicontinuous extended valued function \( f \) at a point \( \bar{x} \in \text{dom} f \) can be expressed in terms of the normal cone of the epigraph of \( f \):
\[
\partial f(\bar{x}) = \{ \xi \mid (\xi, -1) \in N_{\text{epi} f}(\bar{x}, f(\bar{x})) \}.
\]

**Lemma 2.3.** Take an open set \( \mathcal{O} \subset \mathbb{R}^k \), a lower semicontinuous function \( f : \mathcal{O} \to \mathbb{R} \cup \{+\infty\} \), and a point \( \bar{x} \in \mathcal{O} \) such that \( \bar{x} \in \text{dom} f \).

(i) If \( f \) is Lipschitz continuous on \( \mathcal{O} \), then the lower Dini derivative has the equivalent, simpler, definition
\[
D^i f(\bar{x}; d) = \liminf_{h \downarrow 0} h^{-1} [f(\bar{x} + hd) - f(\bar{x})].
\]

(ii) If \( f \) is Gâteaux differentiable at \( \bar{x} \) (denote the derivative \( \nabla f(\bar{x}) \)), then
\[
D^i f(\bar{x}; d) = \nabla f(\bar{x}) \cdot d.
\]

(iii) If \( f \) is Lipschitz continuous on \( \mathcal{O} \), then
\[
\text{co} \partial f(\bar{x}) = \{ \xi \mid d \cdot \xi \leq D^0 f(\bar{x}; d) \text{ for all } d \in \mathbb{R}^k \}.
\]

When \( f \) is Lipschitz continuous on a neighborhood of \( \bar{x} \), \( \text{co} \partial f(\bar{x}) \) is referred to as the Clarke generalized gradient. Property (iii) tells us that, in this case, the Clarke generalized gradient is in duality with the Clarke generalized directional derivative.

Properties (i) and (ii) above follow directly from the definitions. The well-known property (iii) is proved in [8].

### 3. Sensitivity relations.

Theorem 3.2 below is a statement of the main results of the paper. These provide interpretations of costate arcs for the state constrained maximum principle in terms of subgradients of the value function. The following notation will be employed:
\[
A^0(t) := \{ x \mid h(t, x) < 0 \}.
\]

Under the “inward pointing” hypothesis (H3) below, \( A^0(t) \) is in fact the interior of the state constraint set \( A(t) \) at time \( t \in [S, T] \).
(H1) For each \((x, u) \in \mathbb{R}^n \times \mathbb{R}^m\) \(f(., x, u)\) is measurable, and for each \(t \in [S, T]\) \(f(t, ., .)\) is continuous. The multifunction \(t \mapsto U(t)\) is uniformly bounded, has values, closed sets, and is Borel measurable. \(g\) is locally Lipschitz continuous. \(h(., .)\) is of class \(C^{1+}\) (i.e., everywhere Fréchet differentiable with locally Lipschitz continuous derivatives).

(H2) There exist \(c > 0\) and \(k_F(\cdot) \in L^1([S, T]; R)\) such that
\[
|f(t, x, u)| \leq c(1 + |x|) \quad \text{for all} \quad (t, x) \in [S, T] \times \mathbb{R}^n \quad \text{and} \quad u \in U(t),
\]
\[
|f(t, x, u) - f(t, x', u)| \leq k_F(t)|x - x'| \quad \text{for all} \quad t \in [S, T], \quad x, x' \in \mathbb{R}^n, \quad \text{and} \quad u \in U(t).
\]

(H3) Given any \(r > 0\), there exist \(\gamma > 0\), and \(\rho > 0\) such that
\[
\min_{u \in U(t)} \nabla h(t, x) \cdot (1, f(t, x, u)) \leq -\gamma
\]
for all \((t, x) \in [S, T] \times rB\) at which \(|h(t, x)| \leq \rho\).

We precede the “sensitivity relations” theorem with a proposition, listing regularity properties of the value function \(V : [S, T] \times \mathbb{R}^n \to R \cup \{+\infty\}\) defined in (1). Recall here that \(\text{dom} \ V\) is the set of points in \([S, T] \times \mathbb{R}^n\) at which \(V\) is finite valued.

**Proposition 3.1.** Assume (H1)-(H3). Then \(\text{dom} \ V = \text{Graph} \ A\). \(V\) is lower semicontinuous. Furthermore, \(V\) coincides on \(\text{dom} \ V\) with a locally Lipschitz function on \([S, T] \times \mathbb{R}^n\).

A proof of Proposition 3.1 is provided at the end of section 5.

**Theorem 3.2.** Assume (H1)-(H3). Let \((\bar{x}, \bar{u})\) be a minimizer for problem \(P_{S, x_0}\).

Then

(A) There exists a \(p(\cdot) \in BV([S, T]; \mathbb{R}^n)\) and \(\mu \in NBV^+(S, T)\) such that

(i) the conditions (5)-(8) of the state constrained maximum principle are satisfied,

(ii) \((H(t, \bar{x}(t), p(t)), -p(t)) \in \text{co} \partial^0 V(t, \bar{x}(t))\) a.e. \([S, T],\)

(iii) \(p(S) \in \partial_x \{\Psi_{A(S)}(\bar{x}(S)) - V(S, \bar{x}(S))\}\).

(B) The assertions of part (A) remain valid (possibly with a different \(p(\cdot)\)) when condition (ii) above is replaced by

(ii(a)) \(\bar{p}(t) \in \text{co} \partial^0 \bar{V}(t, \bar{x}(t))\) a.e. \(t \in [S, T]\).

(In interpreting condition (iii) of Theorem 3.2, we adopt the convention that \(+\infty + (-\infty) = +\infty\). So, since \(\text{dom} \ V = \text{Graph} \ A\), \(x \mapsto (\Psi_{A(S)}(x) - V(S, x))\) is the function taking values \(-V(S, x)\) at points \(x \in A(S)\) and the value \(+\infty\) at points \(x \notin A(S)\). If \(\bar{x}(S) \in A^0(S)\), condition (iii) of Theorem 3.2 is simply \(p(S) \in \partial_x (-V(S, \bar{x}(S)))\).)

A proof of Theorem 3.2 is given in section 6.

The key conditions (ii) and (ii(a)) in the theorem statement involve the subgradients \(\partial^0 V(t, x)\) and \(\partial^0 \bar{V}(t, x)\), whose definitions are distinguished by the fact that they involve a limit of subgradients at points interior to the state constraint sets:

\[
\partial^0 V(t, x) := \limsup \left\{ \partial V(t', x') \mid (t', x') \stackrel{\text{Graph} A^0}{\rightarrow} (t, x) \right\},
\]

\[
\partial^0 \bar{V}(t, x) := \limsup \left\{ \partial_x V(t', x') \mid (t', x') \stackrel{\text{Graph} A^0}{\rightarrow} (t, x) \right\}.
\]

It follows immediately from the well-known upper semicontinuity properties of the subdifferential that
\[
\partial^0 V(t, x) \subset \text{co} \partial V(t, x) \quad \text{for all} \quad (t, x) \in \text{Graph} A,
\]
for all \((t, x)\). We conclude then from Theorem 3.2 the following corollary.
Corollary 3.3. The assertions of Theorem 3.2 remain valid if condition (ii) is replaced by the conditions 

(ii)′ \( (H(t, \bar{x}(t), p(t)), -p(t)) \in co \partial V(t, \bar{x}(t)) \text{ a.e. } [S,T], \)

in which \( \partial V \) is the “standard” subdifferential and the associated partial subdifferential (see Definition 2.2).

Comments.

(a) Applying Corollary 3.3 to the special case of \( P_{S,x_0} \) in which \( h(.,.) \equiv -1 \) yields sensitivity relations of [10] and [21] for state constraint-free optimal control problems. The respects in which the sensitivity relations of Theorem 3.2 for state constraint problems improve on those in [6] are described in the introduction.

(b) Note that the assertions of parts A and B of the theorem are distinct, since, in general,

\[ \Pi_x \partial^0 V(t, \bar{x}(t)) \neq \partial^0_x V(t, \bar{x}(t)), \]

where \( \Pi_x \) denotes projection onto the \( x \) vector coordinate.

(c) Hypothesis (H3) requires the existence of velocities satisfying a uniform inward pointing condition at all points \((t,x)\) over an arbitrary ball for which \(|h(t,x)|\) is sufficiently small. The need to impose a uniform condition arises because we allow data that is possibly discontinuous with respect to the time variable. If (H1) is strengthened to require that \( U(.) \) is a continuous multifunction and \( f \) is a continuous function, then the assertions of the theorem remain valid when the “inward pointing” hypothesis (H3) is replaced by the hypothesis

\[ \min_{u \in U(t)} \nabla h(t,x) \cdot (1, f(t,x,u)) < 0 \text{ for all } (t,x) \in [S,T] \times R^n \]

at which \( h(t,x) = 0 \).

(d) At times \( t \in (S,T) \) where \( h(t, \bar{x}(t)) < 0 \), the sets \( \partial^0 V(t, \bar{x}(t)) \) and \( \partial V(t, \bar{x}(t)) \) in Theorem 3.2 and Corollary 3.3 coincide. At any time \( t \) when the state constraint \( h \) is active, however, we can expect that \( \partial^0 V(t, \bar{x}(t)) \) will be a strict subset of \( \partial V(t, \bar{x}(t)) \) and that the assertions of the theorem will therefore be more informative than those of the corollary. This is because the definition of \( \partial^0 V \) involves limits of subgradients evaluated in the relative interior of Graph \( A \). Consider, for example, the case when \( h(t,x) \), and so also \( A(t) \), are independent of time (write \( h(x) \) in place of \( h(t,x) \), etc.), and the state constraint is active at time \( t \). Suppose further that \( V \) is continuously differentiable at points in Graph \( A^0 \), in a neighborhood of \((t, \bar{x}(t))\), and \( \nabla V \) on Graph \( A^0 \) extends to Graph \( A \) as a continuous function. Then

\[ \partial^0 V(t, \bar{x}(t)) = \{ \nabla V(t, \bar{x}(t)) \}, \]

while

\[ \partial V(t, \bar{x}(t)) = \{ \nabla_t V(t, \bar{x}(t)) \} \times \{ \nabla_x V(t, \bar{x}(t)) + \alpha \nabla h(\bar{x}(t)) \mid \alpha \geq 0 \}. \]

We see here that \( \partial^0 V \) precisely captures the gradient \( \nabla V(t, \bar{x}(t)) \), but that the “coarser” subgradient \( \partial V \) suppresses gradient information about \( V \), in the normal direction to the boundary of \( A \) at \( \bar{x}(t) \).
(e) If a “coordinate-free” formulation of the hypotheses and of the sensitivity relations is preferred, then we can apply Theorem 3.2 to

\[ h(x) = d_A^+(x), \]

where \( d_A^+(\cdot) \) is the oriented distance function

\[ d_A^+(x) := d_A(x) - d_{A^c}(x). \]

(Here, \( A^c \) denotes the complement of the set \( A \).)

This requires, of course, that \( d_A^+(\cdot) \) is a \( C^{1+} \) function.\(^1\) With this identification for \( h \), the “inward pointing” hypothesis (H3) and the conditions of Theorem 3.2 acquire an intrinsic description, in which normal vectors of \( A \) replace the gradients of \( h \).

(f) Whether the same costate arc \( p(\cdot) \) can be chosen to satisfy both conditions (ii) and (ii(a)) of Theorem 3.2 remains an open question (even for no state constraints).

4. Existence of neighboring feasible trajectories. There is substantial literature providing conditions for the existence of a state trajectory \( x \) on \([S,T]\), satisfying the state constraint \( x(t) \in A(t), S \leq t \leq T \) (a “feasible” \( F \)-trajectory). A more refined line of investigation is to seek conditions for the existence of a state trajectory \( x \) which satisfies the state constraint and whose closeness to a given reference state trajectory is governed by the extent to which the reference state trajectory violates the state constraint. The outcome of this research has been a number of existence of neighboring feasible trajectory (ENFT) theorems. Such properties are of interest, because of their role in the analysis of regularity properties of value functions, in characterizing the value function as a solution (in some generalized sense) to the Hamilton Jacobi equation and in the study of conditions under which the state constrained Pontryagin maximum principle is nondegenerate.

ENFTs play an important part also in the derivation of sensitivity relations in optimal control. In this section we state one version of the ENFT theorem and examine some implications relevant to this application. This takes as a starting point a control system with the description

\[
\begin{align*}
\dot{x}(t) &\in F(t, x(t)) \quad \text{a.e. } t \in [S,T], \\
x(t) &\in A(t) \quad \text{for all } t \in [S,T].
\end{align*}
\]

Here, in contrast to the controlled differential equation underlying the optimal control problem of the introduction, the dynamic constraint is formulated as a differential inclusion involving the multifunction \( F : [S,T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n \). The role of state trajectories is replaced by \( F \)-trajectories: given \([S',T'] \subset [S,T]\), an arc \( x(.) : [S',T'] \rightarrow \mathbb{R}^n \) is said to be an \( F \)-trajectory (on \([S',T']\)) if it is absolutely continuous and satisfies \( \dot{x}(t) \in F(t, x(t)) \) a.e. \( t \in [S',T'] \).

The state constraint sets \( A(t), t \in [S,T] \) are assumed as before to have the functional representation

\[ A(t) = \{ x \in \mathbb{R}^n \mid h(t, x) \leq 0 \} \]

\(^1\)If a set \( A \subset \mathbb{R}^n \) is of class \( C^{1+} \), then the oriented distance function is automatically \( C^{1+} \) (see, e.g., [11]).
in which $h(.,.) : [S,T] \times R^n \rightarrow R$ is a given function. Write

$$h^+(t,x) := h(t,x) \vee 0.$$ 

We shall invoke the following hypotheses (for some specified value of the positive parameter $r_0$): \(\exists c > 0, \gamma > 0, \rho > 0, \) and $k_F \in L^1([S,T];R)$ such that

(D1) $F$ takes values in the space of nonempty, closed sets and $F(.,x)$ is measurable for each $x \in R^n$. $h(.,.)$ is of class $C^{1+}$ on $[S,T] \times R^n$.

(D2) $F(t,x) \subset c(1 + |x|)B$ for all $(t,x) \in [S,T] \times R^n$.

(D3) $F(t,x) \subset F(t,x') + k_F(t)|x-x'|B$ for all $(t,x) \in [S,T]$ and $x,x' \in e^{c|T-S|}(1 + r_0)B$.

(D4) $\sup_{v \in F(t,x)} \nabla h(t,x) \cdot (1,v) \leq -\gamma$ for all $(t,x) \in [S,T] \times e^{c|T-S|}(1 + r_0)B$ for which $|h(t,x)| \leq \rho$.

**THEOREM 4.1.** Take any $r_0 > 0$. Assume that hypotheses (D1)–(D4) above are satisfied. Then there exists a constant $K$ (whose magnitude depends only on $r_0$), with the following property: given any $\sigma \in [S,T]$ and $F$-trajectory $\hat{x}$ on $[\sigma,T]$ with $\hat{x}(\sigma) \in A(\sigma) \cap r_0B$, there exists an $F$-trajectory $x$ on $[\sigma,T]$ with $x(\sigma) = \hat{x}(\sigma)$ such that

$$\hat{x}(t) \in A(t) \text{ for all } t \in [\sigma,T]$$

and

$$||x-\hat{x}||_{W^{1,1}([\sigma,T];R^n)} \leq K \max_{t \in [\sigma,T]} h^+(t,\hat{x}(t)).$$

The corollary to follow asserts that, given any $F$-trajectory satisfying the state constraint, there exists, arbitrarily close to it, an $F$-trajectory satisfying the state constraint conditions with strict inequality.

**COROLLARY 4.2.** Take any $\epsilon > 0$, $s \in [S,T]$, and $F$-trajectory $\hat{x}$ on $[s,T]$ that satisfies

$$h(t,\hat{x}(t)) \leq 0 \text{ for all } t \in [s,T].$$

Assume that, for some $r_0 > |\hat{x}(s)|$, hypotheses (D1)–(D4) are satisfied. Then there exists $\sigma \in [s,T]$ and an $F$-trajectory $x$ on $[\sigma,T]$ such that

$$h(t,x(t)) < 0 \text{ for all } t \in [\sigma,T]$$

and

$$\max\{|\sigma - s|, ||x-\hat{x}||_{W^{1,1}([\sigma,T];R^n)}\} \leq \epsilon.$$

Theorem 4.1 is proved in [5]. The corollary can be proved by applying Theorem 4.1 to the arc $\hat{x}(.)$, perturbed so that its left end point is interior to a state constraint set, when the state constraint functional is taken to be $h + \delta$, for some suitable positive constant $\delta$. Proofs of Theorem 4.1 are given in [12], [13] and, in a more general setting, allowing a vector valued functional inequality state constraint in [14] and [18]. These proofs have a common error: the first inequality in the estimate on p. 33 of [14], for instance, is false in general. The more refined construction procedure for neighboring feasible trajectories in [5] corrects this error in the case of a scalar valued constraint function. A counterexample in [5] illustrates that the $W^{1,1}$-estimate is not in general valid for multiple state constraints. We mention that the proofs of neighboring feasible trajectories theorems in [3] and [4] apply to convex valued differential inclusions; they employ different proof techniques and are unaffected by the above observations. The proof of a related theorem under a rather restrictive “inward pointing” condition (requiring continuous feedback implementation) in [17] also appears to be valid.
5. Exact penalization. In this section we establish useful relations between the family problems \( P_{t,x} \), \((t, x) \in [S, T] \times R^n \) defined in the introduction and a new families of problems, in which the state constraint is replaced by an “exact” penalty function and the control differential equation constraint is replaced by a differential inclusion constraint. For \( (t, x) \in [S, T] \times R^n \) write

\[
F(t, x) = f(t, x, U(t)).
\]

We list without proof the following elementary properties of the multifunctions \((t, x) \rightarrow F(t, x)\) and \((t, x) \rightarrow \text{co } F(t, x)\).

**Lemma 5.1.** Assume (H1)–(H2). Then, for either \( G(., .) = F(., .) \) or \( G(., .) = \text{co } F(., .) \) we have that

(i) \( G \) takes values in the space of nonempty, closed sets, and \( G(., x) \) is measurable for each \( x \in R^n \).

(ii) \( G(t, x) \subset c(1 + |x|)B \) for all \((t, x) \in [S, T] \times R^n; \)

(iii) \( G(t, x) \subset G(t, x') + k_F(t)|x - x'|B \) for all \( t \in [S, T] \) and \( x, x' \in R^n \).

For given \( K > 0 \) and \((t, x) \in [S, T] \times R^n \) we define

\[
P_{t,x}^K \left\{ \begin{array}{l}
\text{Minimize } g(y(T)) + K \max_{s \in [t, T]} h^+(s, y(s)) \\
\text{over } F\text{-trajectories } y \text{ on } [t, T] \text{ that satisfy } y(t) = x,
\end{array} \right.
\]

and

\[
\text{co } P_{t,x}^K \left\{ \begin{array}{l}
\text{Minimize } g(y(T)) + K \max_{s \in [t, T]} h^+(s, y(s)) \\
\text{over } \text{co } F\text{-trajectories } y \text{ on } [t, T] \text{ that satisfy } y(t) = x.
\end{array} \right.
\]

**Lemma 5.2.** Assume (H1)–(H3). Fix \( r_0 > 0 \). Then there exists a number \( K > 0 \) with the following property: given any \((t, x) \in [S, T] \times r_0B \) and any \( F\text{-trajectory } y'(. \) on \([t, T] \) satisfying \( y'(t) = x \), there exists an \( F\text{-trajectory } y(\cdot ) \) on \([t, T] \) such that \( y(t) = y'(t) \),

\[
g(y(T)) \leq g(y'(T)) + K \max_{s \in [t, T]} h^+(s, y'(s)),
\]

and

\[
h(s, y(s)) \leq 0 \text{ for all } s \in [t, T].
\]

**Proof.** Take \( y' \) as in the lemma statement. Apply the ENFT Theorem 4.1 to \( \{ \dot{x} \in F, x(t) \in A(t) \} \). (The hypotheses permitting such an application are satisfied; note in particular that hypothesis (D4) is satisfied, for the given \( r_0 \) and for some \( \gamma > 0, \rho > 0 \), because (H3) is satisfied.) The theorem tells us that there exists a constant \( k > 0 \) (independent of our choice of \((t, x) \in [S, T] \times R^n \)) and an \( F\text{-trajectory } y \) on \([t, T] \) such that \( h(s, y(s)) \leq 0 \) for all \( s \in [S, T] \) and

\[
\|y'(.) - y(.)\|_{L^\infty([t, T]; R^n)} \leq K \max_{s \in [t, T]} h^+(s, y'(s)).
\]

Choose \( K > \tilde{K} k_y \), where \( k_y \) is a Lipschitz constant for \( g \) restricted to \( r_1B \), where \( r_1 = e^{c|T-S|}(1 + r_0) \). Then

\[
g(y(T)) \leq g(y'(T)) + k_y\|y - y'\|_{L^\infty} \leq g(y'(T)) + K \max \{ h^+(s, y'(s)) \mid s \in [t, T] \}. \]

We see that \( y(.) \) has the properties asserted in the lemma statement. \( \square \)
Given \((t, x) \in [S, T] \times \mathbb{R}^n\), denote by inf \(P^K_{t,x}\) and inf co \(P^K_{t,x}\) the infimum costs of \(P^K_{t,x}\) and co \(P^K_{t,x}\), respectively. As before, inf \(P_{t,x}\) denotes the infimum cost of \(P_{t,x}\).

**Proposition 5.3.** Assume that hypotheses (H1)–(H3) are satisfied.

(i) Take any \((t, x) \in [S, T] \times \mathbb{R}^n\) and \(K \geq 0\). Then

\[
\inf P^K_{t,x} = \inf \text{co } P^K_{t,x},
\]

and if \((y, u)\) is a minimizing process for \(P_{t,x}\), \(y\) is a minimizing \(F\)-trajectory for \(P^K_{t,x}\).

(ii) Fix a bounded set \(D \subset \mathbb{R}^n\). Then there exists a number \(K > 0\) with the following properties: given any \((t, x) \in ([S, T] \times D) \cap \text{Graph} A\),

\[
\inf P_{t,x} = \inf P^K_{t,x} = \inf \text{co } P^K_{t,x},
\]

and if \(y\) is a minimizing \(F\)-trajectory for \(P^K_{t,x}\), \(y\) is a minimizing co \(F\)-trajectory for co \(P^K_{t,x}\).

**Proof.** (i) This part of the proposition follows from the facts that, under hypotheses (H1)–(H2), co \(F\)-trajectories can be approximated arbitrarily closely with respect to the \(L^\infty\) norm by \(F\)-trajectories which share the same left end point (see, e.g., [20, Thm. 2.7.2]), and the cost function \(y(\cdot) \to g(y(T))\) is continuous with respect to the \(L^\infty\) norm.

(ii) In consequence of Filippov’s measurable selection theorem (see [20, Thm. 2.3.13]), an arc \(y\) is an \(F\)-trajectory if and only if \(y\) is a state trajectory in the sense of section 1 (for some control function). We have \(\inf P_{t,x} \geq \inf P^K_{t,x}\); since any state trajectory satisfying the constraints of \(P_{t,x}\) is an \(F\)-trajectory for \(P^K_{t,x}\) with the same cost. But we deduce from Lemma 5.2 that, given any \(F\)-trajectory satisfying the constraints of \(P_{t,x}\), there exists a state trajectory for \(P_{t,x}\) with no greater cost; it follows that \(\inf P_{t,x} = \inf P^K_{t,x}\) and that if \(y\) is a minimizing \(F\)-trajectory for \(P^K_{t,x}\), then \(y\) is a minimizing co \(F\)-trajectory for co \(P^K_{t,x}\).

For given \(K > 0\) we introduce a new (auxiliary) value function \(V^K : [S, T] \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}\) as follows:

\[
V^K(t, x) = \inf P^K_{t,x},
\]

where the right-hand side is interpreted as the infimum cost if admissible processes for \(P^K_{t,x}\) exist and as \(+\infty\) otherwise.

**Proposition 5.4.** Take any \(K > 0\). Assume (H1)–(H2). Then

(i) \(V^K\) is locally Lipschitz continuous;

(ii) there exists a subset \(J \subset [S, T]\), of full measure, with the following property:

for any \((t, x) \in J \times \mathbb{R}^n\) such that \(h(t, x) < 0\) for any \(v \in \text{co } F(t, x)\), we have

\[
D^1V^K((t, x); (1, v)) \geq 0.
\]

**Proof.** We omit the proof of (i), which is a straightforward consequence of the Filippov existence theorem (cf. the proof of [20, Thm. 12.3.5]). Consider then (ii). Given \([a, b] \subset [S, T]\) and \(\xi \in \mathbb{R}^n\), define the reachable set of \(\dot{x} \in \text{co } F\) on \([a, b]\) emanating from \(\xi\):

\[
R(a, b; \xi) = \{x(b) | x(.)\text{ is a co } F\text{-trajectory on } [a, b]\text{ such that } x(a) = \xi\}.
\]

In view of the properties of co \(F\) listed in Lemma 5.1, we know from [22, Prop. 4.1] that there exists a subset \(J \subset [a, b]\) of full measure such that, for all \(\xi \in \mathbb{R}^n\) and \(t \in J\),

\[
\lim_{\delta \to 0} \delta^{-1} (R(t, t + \delta; \xi) - \xi) = \text{co } F(t, x).
\]
Take any \((t, x) \in J \times \mathbb{R}^n\) such that \(h(t, x) < 0\) and any \(v \in \text{co} F(t, x)\). There exist \(\delta_i \downarrow 0\) as \(i \to +\infty\) such that (see Lemma 2.3(i))

\[
D^iV^K((t, x), (1, v)) = \lim_{i \to \infty} \delta_i^{-1} \left[ V^K(t + \delta_i, x + \delta_i v) - V^K(t, x) \right].
\]

According to (13), there exists a sequence of \(\text{co} F\)-trajectories \(\{y_i : [t, t + \delta_i] \to \mathbb{R}^n\}\) such that \(y_i(t) = x\) for all \(i\) and

\[
\delta_i^{-1} [y_i(t + \delta_i) - x] \to v.
\]

It follows from the local Lipschitz continuity of \(V^K\) that

\[
\lim_i \delta_i^{-1} \left[ V^K(t + \delta_i, y_i(t + \delta_i)) - V^K(t, x) \right] = \lim_i \delta_i^{-1} \left[ V^K(t + \delta_i, x + \delta_i v) - V^K(t, x) \right].
\]

But since \(h(t, x) < 0\), we deduce from the definition of \(V^K\) that

\[
\lim_i \delta_i^{-1} \left[ V^K(t + \delta_i, y_i(t + \delta_i)) - V^K(t, x) \right] \geq 0
\]

(the “principle of optimality”). In consequence

\[
D^iV^K((t, x), (1, v)) \geq 0.
\]

Proof of Proposition 3.1. Propositions 5.3 and 5.4 combine to tell us that, given any bounded set \(D \subset \mathbb{R}^n\), there exists a number \(K > 0\) such that the restriction of \(V\) to the closed set \((S, T) \times D \cap \text{Graph} A\) coincides with the locally Lipschitz continuous function \(V^K\). Since \(V\) takes the value \(+\infty\) on the complement of \(\text{Graph} A\), we conclude that \(\text{dom} V = \text{Graph} A\) and \(V\) is a lower semicontinuous function, and \(V\) is the restriction of a locally Lipschitz continuous function on \((S, T) \times \mathbb{R}^n\) to \(\text{Graph} A\). \(\square\)

6. Proof Theorem 3.2.

Proof of Theorem 3.2, part A. The proof technique is to associate with the minimizing process \((\bar{x}, \bar{u})\) a minimizing process for an auxiliary control problem and to apply the standard state constrained maximum principle to the auxiliary problem. The fact that the additional control variables introduced into the auxiliary problem fail to reduce the value of the cost provides additional information in the form of sensitivity relations.

With a view to introducing the auxiliary problem, we define, for each \(\epsilon > 0\), the multifunction \(G_\epsilon : [S, T] \to \mathbb{R}^{1+n}\):

\[
G_\epsilon(t) := \{ (\alpha, \beta) \in \mathbb{R}^{1+n} \mid (\alpha, \beta) \in \text{co} \partial V(s, y) \text{ for some} \}
\]

\[
(s, y) \in ((t, \bar{x}(t)) + \epsilon B) \cap ((S, T) \times \mathbb{R}^n) \text{ such that } h(s, y) < 0 \}
\]

and, for \(v \in \mathbb{R}^n, w \in \mathbb{R}\),

\[
\sigma_\epsilon(t, v, w) := \sup_{(\alpha, \beta) \in G_\epsilon(t)} (\alpha, \beta) \cdot (w, -(1 + w)v).
\]

Recall that \(V\) is locally Lipschitz on \(\{(t, x) : h(t, x) \leq 0\}\), and notice that \(\text{co} \partial V\), viewed as a multivalued function of its base points, has closed graph (see, for example, [8, Prop. 2.1.5]). Then, for any \(\epsilon > 0\), the function \(\sigma_\epsilon(t, v, w)\) is measurable, upper semicontinuous with respect to the \((v, w)\) variables for fixed \(t\), and nonempty and locally bounded on \((S, T) \times \mathbb{R}^{n+1}\).

Lemma 6.1. Assume that hypotheses (H1) and (H2) are satisfied. Take any \(K > 0\) and \(\epsilon \in (0, 1/2]\). Let \((x, (u, v, w))\) be a process for the control system

\[
\begin{cases}
\dot{x}(t) = (1 + w)(f(t, x(t), u(t)) + v(t)) & \text{a.e. } t \in [S, T], \\
(u(t), v(t), w(t)) \in U(t) \times \epsilon B \times [-\epsilon, +\epsilon]
\end{cases}
\]
such that \(|x - \bar{x}|_{L^\infty} < \epsilon\) and
\[
h(t, x(t)) < 0 \quad \text{for all } t \in [S, T].
\]

Then, for almost every \(t \in [S, T]\)
\[
\frac{d}{dt} V^K(t, x(t)) + \sigma_e(t, v(t), w(t)) \geq 0.
\]

Here \(V^K\) is the function (12).

Proof. Let \(D \subset \mathbb{R}^n\) be a compact set such that \(x(t) \in D\) for all \(t \in [S, T]\), and let \(J \subset [S, T]\) be the subset of full measure, whose existence is asserted in Proposition 5.4. Now define the subset \(J' \subset J\) to comprise points \(t\) with the following properties:
(i) \(t\) is a Lebesgue point of \(s \to \dot{x}(s)\) and \(\dot{x}(t) = (1 + w(t))(f(t, x(t), u(t)) + v(t));
(ii) The function \(s \to V^K(s, x(s))\) is differentiable at \(t\).
Since, in consequence of Proposition 5.4, \(t \to V^K(t, x(s))\) is absolutely continuous, \(J'\) has full measure.

Take any \(t \in J'\). Notice first that, since \(s \to V^K(s, x(s))\) is differentiable at \(t\) and \(V^K\) is Lipschitz continuous on a neighborhood of \((t, x(t))\), we have
\[
d/dt V^K(t, x(t)) = \lim_{h \downarrow 0} h^{-1} \left[ V^K \left( t + h, x(t) + \int_t^{t+h} \dot{x}(s)ds \right) - V^K(t, x(t)) \right]
\]
\[
= \lim_{h \downarrow 0} h^{-1} [V^K(t + h, x(t) + \dot{x}(t)h) - V^K(t, x(t))]
\]
\[
= D^i V^K((t, x(t)); (1, 1 + w)(f(t, x, u(t)) + v(t))).
\]

In the last equality, we have used Lemma 2.3(i). Henceforth, for simplicity, we write \(x(t), u(t), v(t),\) and \(w(t)\) as \(x, u, v,\) and \(w.\) We know
\[
0 \leq \lim inf_{h \downarrow 0} \left[ V^K(t, x(t)) - V^K(t + h, x + hf(t, x, u)) \right] h^{-1} (1 + w)
\]
(by Proposition 5.4 and since \((1 + w) > 0\))
\[
= \lim inf_{h \downarrow 0} \left[ V^K(t + h, x + (1 + w)f(t, x, u)) - V^K(t, x) \right] h^{-1}
\]
(by positive homogeneity)
\[
\leq \lim inf_{h \downarrow 0} \left[ V^K(t + h, x + h(f(t, x, u) + v)(1 + w)) - V^K(t, x) \right] h^{-1}
\]
\[
+ \lim sup_{h \downarrow 0} \left[ V^K(t_h + hv, x_h - hv(1 + w)) - V^K(t, x_h) \right] h^{-1}
\]
(wher \(t_h := t + h\) and \(x_h := x + h(f(t, x, u) + v)(1 + w)\))
\[
\leq D^i V^K((t, x); (1, (f(t, x, u) + v)(1 + w))) + D^0 V^K((t, x); (w, -v(1 + w)))
\]
\[
= d/dt V^K(t, x) + D^0 V^K((t, x); (w, -v(1 + w)))
\]
\[
= d/dt V^K(t, x) + \max\{ (\alpha, \beta) \cdot (w, -v(1 + w)) \mid (\alpha, \beta) \in \partial V(t, x) \}
\]
\[
\leq d/dt V^K(t, x) + \sigma_e(t, v, w).
\]

The second to last relation is a consequence of Lemma 2.3(iii). The last relation follows from the definition of \(G_e(., .)\) and \(\sigma_e(., ., .)\), since, by assumption, \(|x(t) - \bar{x}(t)| < \epsilon\) and \(x(t) \in A(t)\). 

**Lemma 6.2.** Assume that hypotheses (H1)–(H3) are satisfied. Take any \(r_0 > \|\bar{x}\|_{L^\infty}\). Then there exists \(K > 0\) and \(\xi \in (0, 1/2)\) with the following property: for any
\( \epsilon \in (0, \bar{\epsilon}), (\bar{x}, (\bar{u}, v \equiv 0, w \equiv 0)) \) is a minimizer for

\[
\begin{aligned}
&\text{Minimize } g(x(T)) + \int_S^T \sigma_\epsilon(t, v(t), w(t)) dt \\
&+ K \max_{t \in [S, T]} h^+(t, x(t)) - V(S, x(S)) \\
\text{subject to} \\
&x = (1 + w)(f(t, x, u) + v), \\
&(u, v, w) \in U(t) \times \epsilon B \times [-\epsilon, \epsilon], \\
&x(S) \in A(S), \quad \|x - \bar{x}\|_{L^\infty} \leq \epsilon/2, \quad \text{and } \|x\|_{L^\infty} < r_0.
\end{aligned}
\]

**Proof.** Choose \( K_1 > 0 \) such that

\[
V^{K_1}(t, x) = V(t, x)
\]

for all \((t, x) \in ([S, T] \times r_1 B) \cap \text{Graph } A\). (Such a \( K_1 \) exists according to Proposition 5.3).

We know, by hypothesis (H3), that there exist \( \gamma > 0 \) and \( \rho > 0 \) such that

\[
\min_{u \in U(t)} \nabla h(t, x) \cdot (1, f(t, x, u)) \leq -\gamma
\]

for all \((t, x) \in ([S, T] \times r_1 B) \cap \text{Graph } A\) such that \( |h(t, x)| \leq \rho \). Here \( r_1 := e^{c(T-S)}(1+r_0) \).

Let \( k_h > 0 \) be a constant such that \( |\nabla h(t, x)| \leq k_h \) for any \((t, x) \in [S, T] \times r_1 B \).

Now choose \( \bar{\epsilon} \) to be a point in \((0, \frac{1}{2})\) such that

\[
\bar{\epsilon} \leq \frac{1}{2} k_h (1 + r_0) + \epsilon (1 + \bar{\epsilon}) k_h < \frac{\gamma}{2}.
\]

Take any \( \epsilon \in (0, \bar{\epsilon}) \) and any process \((y, (u, v, w))\) satisfying the constraints of problem \( P_{\epsilon, \bar{\epsilon}} \). To prove the lemma, it suffices to show that the cost of \((y, (u, v, w))\) is not less than that of \((\bar{x}, (\bar{u}, v \equiv 0, w \equiv 0))\).

Fixing the function \( v \) and \( w \), we apply the ENFT Theorem 4.1 to the constrained differential inclusion

\[
\dot{z} \in \tilde{F}(t, z), \quad z(t) \in A(t) \quad \text{for all } t \in [S, T],
\]

where

\[
\tilde{F}(t, x) := (1 + w(t))(f(t, x, U(t)) + v(t)).
\]

Because we have chosen \( \bar{\epsilon} \) to satisfy (16), \( \tilde{F} \) (and \( h \)) satisfy hypotheses (D1)–(D4), with \( c, k_F \), and \( \gamma \) replaced by \( \frac{3}{2} c + \frac{3}{2} k_F \), and \( \frac{1}{2} \gamma \), respectively. (\( \rho \) remains as before.) Arguing as in the proof of Lemma 5.2, we can demonstrate the existence of a constant \( K \) for the state constraint penalty term in \( P_{\epsilon, \bar{\epsilon}} \) with the following property: regardless of our choice of \( \epsilon \) and the process \((y, (u, v, w))\), \((y, (u, v, w))\) can be replaced by a new process (with the same functions \( v(.) \), \( w(.) \)), without increasing the cost of \( P_{\epsilon, \bar{\epsilon}} \) such that the new process satisfies the state constraint. It follows that we can assume, without loss of generality, that \((y, (u, v, w))\) satisfies the state constraint: \( y(t) \in A(t) \) for all \( t \in [S, T] \).

Take an arbitrary sequence \( \gamma_i \downarrow 0 \) such that \( \gamma_i \leq \min\{r_0 - \|y\|_{L^\infty}, \epsilon/2\} \) for each \( i \). Fix \( i \) and apply Corollary 4.2 to the constrained differential inclusion (17), again fixing the functions \( v(.) \) and \( w(.) \). We conclude that there exist \( \sigma_i \in [S, T] \) and a process \((y_i, (u_i, v, w))\) for problem \( P_{\epsilon, \bar{\epsilon}} \) that satisfies

\[
\|y - y_i\|_{L^\infty([S, T]; R^n)} \leq \gamma_i, \quad |\sigma_i - S| < \gamma_i,
\]
and
\[ h(t, y_i(t)) < 0 \quad \text{for all } t \in [\sigma_i, T]. \]

Because \( t \to V^{K_i}(t, x(t)) \) is absolutely continuous and \( V^{K_i}(T,) = g(.) \) on \( A(T) \) we deduce from Lemma 6.1 that
\[
g(y_i(T)) + \int_{\sigma_i}^T \sigma_s(t, v(t), w(t)) dt - V^{K_i}(\sigma_i, y_i(\sigma_i)) \geq 0.
\]

But \( \sigma_i \downarrow S, y_i \) converges uniformly to \( y \) as \( i \to \infty \), \( g \) and \( V^{K_i} \) are continuous functions, and (by Proposition 5.3) \( V^{K_i}(S, y(S)) = V(S, y(S)) \). Therefore, we deduce in the limit as \( i \to \infty \) that
\[
g(y(T)) + \int_S^T \sigma_s(t, v(t), w(t)) dt - V(S, y(S)) \geq 0.
\]

Since \( \sigma_s(t, 0, 0) = 0 \) and \( V(S, \bar{x}(S)) = g(\bar{x}(T)) \), it follows that
\[
g(\bar{x}(T)) + \int_S^T \sigma_s(t, 0, 0) dt - V(S, \bar{x}(S)) = 0.
\]

We have confirmed that \((\bar{x}, (\bar{u}, v \equiv 0, w \equiv 0))\) is a minimizer for \( P^{r_0, \epsilon} \).

We are now ready to undertake proof of Theorem 3.2, part A. Take a sequence \( \epsilon_i \downarrow 0 \). We deduce from Lemma 6.2 that there exists a number \( K > 0 \) with the following properties: for each \( i \) (sufficiently large) the process \((\bar{x}, \bar{z}_1 \equiv 0, \bar{z}_2 \equiv V(S, \bar{x}(S)), (\bar{u}, v \equiv 0, w \equiv 0))\) is a minimizer for the optimal control problem

Minimize \( g(x(T)) + \int_S^T \sigma_s(t, v(t), w(t)) dt + K \max \{0, z_1(T)\} + z_2(T) \)

subject to
\[
\dot{x} = (1 + w)(f(t, x, u) + v), \quad \dot{z}_1 = 0, \quad \dot{z}_2 = 0,
\]
\[
(u, v, w) \in U(t) \times \epsilon B \times [-\epsilon, \epsilon],
\]
\[
h(t, x(t)) - z_1(t) \leq 0,
\]
\[
(z_2(S), x(S)) \in \text{epi} \left[ \Psi_{A(S)}(\cdot) - V(S, \cdot) \right],
\]
\[
|\|x - \bar{x}\|_{L^\infty} \leq \epsilon/2 \quad \text{and} \quad |\|x\|_{L^\infty} < r_0.
\]

Notice that we have reformulated the optimal control problem of Lemma 6.2 so that it now has a form to which known optimality conditions are directly applicable. This involves introducing two new state variables, \( z_1 \) and \( z_2 \), which permit us to replace the “max” term in the cost by a state constraint and to express the boundary conditions on state trajectories as a set inclusion.

Fix \( i \). Denote by \( \bar{H} \) and \( \bar{H} \) the Hamiltonian and maximized Hamiltonian for the above problem, that is,
\[
\bar{H}(t, x, z, q, r, u, v, w) = q \cdot [1 + w](f(t, x, u) + v) - \lambda \sigma_s(t, v, w),
\]
\[
\bar{H}(t, x, z, p, q) = \sup_{(u, v, w) \in (U(t) \times \epsilon B \times [-\epsilon, \epsilon])} \bar{H}(t, x, z, q, r, u, v, w).
\]

Here, \( q \) and \( r = (r_1, r_2)^T \) are the costate variables associated, respectively, with the state variables \( x = (x_1, \ldots, x_n)^T \) and \( z = (z_1, z_2)^T \). Note that \( \bar{H} \) and \( \bar{H} \) are independent of \( z \).
Now apply the state constrained maximum principle [20, Thm. 9.3.1] expressed in terms of the absolutely continuous pseudo costate arc. (See (9)). This tells us that there exist \( \lambda_i \in \{0, 1\} \), \( \mu_i \in NBV^+(S,T) \), and an (absolutely continuous) adjoint arc \((q_i, r^i) : [S,T] \to \mathbb{R}^n \times \mathbb{R}^2 \) (not all zero) such that
\[
-(\dot{q}_i(t), r^i(t))
\]
(19) \(\in \text{co } \partial \varphi \mathcal{H} \left( t, \bar{x}(t), q_i(t) + \int_{[S,T]} h_x(s) \mu_i(ds), \bar{u}(t) \right) \times \{0,0\} \) a.e.,
\[
\mathcal{H} \left( t, \bar{x}(t), \bar{z}(t), q_i(t) + \int_{[S,T]} h_x(s) \mu_i(ds), r^1_i(t) - \int_{[S,T]} \mu_i(ds), r^2_i(t), \bar{u}(t), 0, 0 \right)
\]
(20) \(= \tilde{\mathcal{H}} \left( t, \bar{x}(t), \bar{z}(t), q_i(t) + \int_{[S,T]} h_x(s) \mu_i(ds), r^1_i(t) - \int_{[S,T]} \mu_i(ds), r^2_i(t) \right) \) a.e.,
\[
\text{supp } \{ \mu_i \} \subset \{ t \mid h(t, \bar{x}(t)) = 0 \},
\]
(22) \(r^1_i(S) = 0, \quad (p_i(S), r^2_i(S)) \in N_{epi} [\Psi_{A(S)}(\cdot - V(\cdot,s))](\bar{x}(S), \bar{z}_2(S)) \),
\[
-\left( q_i(T) + \int_{[S,T]} h_x(s) \mu_i(ds) \right) \in \lambda_i \left( \begin{array}{c}
\frac{\partial g(\bar{x}(T))}{K[0,1]} \\
\{1 \}
\end{array} \right) + \left( \begin{array}{c}
\{0 \} \\
(-\infty,0]
\end{array} \right).
\]
(23) Here, \( h_x(s) := \nabla_x h(s, \bar{x}(s)) \).

From (19) and (22) we conclude that \( r^1_i \equiv 0 \). We can assume that \( \lambda_i = 1 \). To see this, let us suppose, to the contrary, \( \lambda_i = 0 \). Then, by (19) and (23), \( r^2_i \equiv 0 \), \( \mu_i \equiv 0 \), and \( q_i(T) \equiv 0 \). Since \( H \) is positively homogeneous in the costate variable, we conclude that \( q_i \equiv 0 \); this contradicts the fact the “multipliers” cannot be all zero. It follows from (23) that
\[
\int_{[S,T]} \mu_i(ds) \leq K.
\]
(24) From (19) and (23), \( r^2_i \equiv -1 \). We deduce from (11), (19), and (22) that
\[
-q_i(t) \in \text{co } \partial \varphi \mathcal{H} \left( t, \bar{x}(t), q_i(t) + \int_{[S,T]} h_x(s) \mu_i(ds), \bar{u}(t) \right) \) a.e.,
(25)
\[
q_i(S) \in \partial \varphi \left( \Psi_{A(S)}(\bar{x}(S)) - V(S, \bar{x}(S)) \right).
\]
(26) From the maximization of the Hamiltonian condition (20), we deduce that, for a.e. \( t \in [S,T] \), the function
\[
(u, v, w) \to \left[ q_i(t) + \int_{[S,T]} h_x(s) \mu_i(ds) \right] \cdot \left[ (1 + w)(f(t, \bar{x}(t), \bar{u}) + v) - \sigma_{\epsilon_i}(t, v, w) \right]
\]
is maximized over \( U(t) \times [\epsilon_i, B] \times \left[ -\epsilon_i, \epsilon_i \right] \) at \( (u = \bar{u}(t), w = 0, v = 0) \). Setting \( (v, w) = (0,0) \), we arrive at
\[
H \left( t, \bar{x}(t), q_i(t) + \int_{[S,T]} h_x(s) \mu_i(ds), \bar{u}(t) \right) = \mathcal{H} \left( t, \bar{x}(t), q_i(t) + \int_{[S,T]} h_x(s) \mu_i(ds) \right) \) a.e.
(27)
On the other hand, from the maximality condition (27) we deduce that, for a.e. \( t \in [S, T] \),

\[
\inf_{(\alpha, \beta) \in G, i(t)} \left\{ w \mathcal{H} \left( t, \bar{x}(t), q_i(t) + \int_{[S, t]} h_x(s) \mu_i(ds) \right) \right. \\
+ (1 + w) \left. \left[ q_i(t) + \int_{[S, t]} h_x(s) \mu_i(ds) \right] \cdot v - \alpha w + (1 + w) \beta \cdot v \right\} \leq 0
\]

for all \((v, w) \in \epsilon_i B \times [-\epsilon_i, \epsilon_i]\). Noting that \(1 + w > 0\), for any \(w \in [-\epsilon_i, \epsilon_i]\), dividing by across the inequality by \(1 + w\) and making the substitutions \(w' = w/(1 + w)\) and \(v' = -v\) in (28), we deduce that, for a.e. \( t \in [S, T] \),

\[
w' \mathcal{H} \left( t, \bar{x}(t), q_i(t) + \int_{[S, t]} h_x(s) \mu_i(ds) \right) \\
- \left[ q_i(t) + \int_{[S, t]} h_x(s) \mu_i(ds) \right] \cdot v' \leq \sup_{(\alpha, \beta) \in G, i(t)} \left\{ \alpha w' + \beta \cdot v' \right\}
\]

at all points \((v', w')\) that belong to a suitably small closed ball in \(\mathbb{R}^{n+1}\) about the origin. The separation theorem now tells us that, for a.e. \( t \in [S, T] \),

\[
\left( \mathcal{H} \left( t, \bar{x}(t), q_i(t) + \int_{[S, t]} h_x(s) \mu_i(ds) \right), -q_i(t) - \int_{[S, t]} h_x(s) \mu_i(ds) \right) \in \varnothing G, i(t).
\]

Here \(\varnothing\) denotes “convex closure.” In view of (24), the \(\mu_i\)’s are uniformly bounded in total variation by the constant \(K\). The \(q_i(T)\)’s are uniformly bounded by (23). Furthermore, in consequence of hypothesis (H2) and in view of (25), the \(q_i\)’s are bounded by a common integrable function. Standard sequential compactness arguments may now be employed (see, for example, [20, p. 329 et seq.]) to show that, along some subsequence of \(\{(q_i, \mu_i)\}\) (we do not relabel), the \(\mu_i \to \mu\) (with respect to the weak* topology on the space of bounded linear functionals on \(C([S, T]; R), \|\cdot\|_\infty\)), \(q_i \to q\) uniformly and \(q_i \to q\) weakly in the \(L^1\) topology for some \(\mu \in NBV^+(S, T)\) and \(q \in W^{1,1}([S, T]; \mathbb{R}^n)\). Further, in consequence of (19), (21), (26), and (27),

\[
-q_i(t) \in \partial \bar{\mu} \mathcal{H} \left( t, \bar{x}(t), q(t) + \int_{[S, t]} h_x(s) \mu(ds) \right) \quad \text{a.e.,}
\]

\[
q(S) \in \partial \bar{\mu} \left\{ \Psi_{A(S)}(\bar{x}(S)) - V(S, \bar{x}(S)) \right\},
\]

\[
\supp \{ \mu \} \in \{ t \mid h(t, \bar{x}(t)) = 0 \},
\]

\[
H \left( t, \bar{x}(t), q(t) + \int_{[S, t]} h_x(s)ds, \bar{u}(t) \right) \\
= \mathcal{H} \left( t, \bar{x}(t), q(t) + \int_{[S, t]} h_x(s)ds \right) \quad \text{a.e.}
\]

Write

\[
p_i(t) := q_i(t) + \int_{[S, t]} h_x(s) \mu_i(ds) \quad \text{for } t \in [S, T]
\]
and $p_i(S) = q_i(S)$. Write also

$$
\mathcal{H}_i(t) := \mathcal{H}(t, \bar{x}(t), p_i(t)).
$$

In terms of these functions, (29) takes the following form:

$$
(\mathcal{H}_i(t), p_i(t)) \in \text{co} G_\varepsilon(t) \quad \text{a.e.}
$$

Let $J \subset [S, T]$ be the subset of points $t$ such that $\int_{[S,t]} h_x(s) \mu_i(ds) \to \int_{[S,t]} h_x(s) \mu(ds)$ as $i \to \infty$, and the inclusion (34) is satisfied for all $i$.

Take any $t \in J$. Recall the definition of $G_\varepsilon$ (see (14)). We see that there exist a sequence $e_i \to 0$ in $R^1 \times R^n$ and, for each $i \in \mathbb{N}$, $j \in \{1, \ldots, n+2\}$ and $k \in \{1, \ldots, n+2\}$, a nonnegative number $a_{ijk}$, a nonnegative number $a_{ijk}$, a point $(t_{ij}, x_{ij})$ in Graph $A^0$, and a point $\lambda_{ijk} \in R^{n+1}$ with the following properties:

$$
\sum_j a_{ij} = 1 \text{ for all } i \text{ and } \sum_k a_{ijk} = 1, \text{ for all } i, j,
$$

$$
\lambda_{ijk} \in \partial V(t_{ij}, x_{ij}), \text{ for each } i, j, k,
$$

$$
(t_{ij}, x_{ij}) \to (t, \bar{x}(t)) \text{ as } i \to \infty, \text{ for each } j,
$$

and

$$
(\mathcal{H}_i(t), p_i(t)) = \sum_{jk} a_{ij} a_{ijk} \lambda_{ijk} + e_i \text{ for each } i.
$$

(Notice that the need for the “error” term $e_i$ in (35) arises because (34) involves the closure of the convex hull.) The restriction of $V$ to Graph $A^0$ is locally Lipschitz continuous. It follows that subgradients $\lambda_{ijk}$ are uniformly bounded over possible values of the indices $i, j, k$. (See, for example, [20, Prop. 4.7.1]). The $a_{ij}$’s and the $a_{ijk}$’s too are uniformly bounded. It follows that, by limiting our attention to a subsequence of values of the index (we do not re-label), we can arrange that $a_{ij} \to a_j$ for each $j$, and $a_{ijk} \to a_{jk}$ and $\lambda_{ijk} \to \lambda_{jk}$ for each $j, k$. Now define $\lambda_{jk} := a_k a_{jk}$. Clearly, the $\lambda_{jk}$’s are nonnegative and sum to 1. We deduce from (35) that

$$
(\mathcal{H}(t), -p(t)) = \lim_i (\mathcal{H}(t), -p_i(t)) = \sum_{jk} \lambda_{jk} \lim_i \lambda_{ijk} \left(= \sum_{jk} \lambda_{jk} \lambda_{jk} \right).
$$

Here $p(S) := q(S)$,

$$
p(t) := q(t) + \int_{[S,t]} h_x(s) \mu(ds) \text{ for } t \in [S, T],
$$

and

$$
\mathcal{H}(t) := \mathcal{H}(t, \bar{x}(t), p(t)) \text{ for all } t \in [S, T].
$$

We have shown that

$$
(\mathcal{H}(t), -p(t)) \subset \text{co} \partial^0 V(t, \bar{x}(t)) \quad \text{a.e.}
$$

Reviewing (30), (31), (32), (33) (now expressed in terms of $p(.)$ defined by (36)), and condition (37), we see the arc $p \in BV([S, T], R^n)$ meets the requirements for it to be a costate arc. Furthermore, it satisfies the desired sensitivity relation.
Proof of Theorem 3.2, part B. The proof of Theorem 3.2, part B is essentially the same as that of part A, except that in place of the sets \( G_\epsilon(t) \) (14) and the functions \( \sigma_\epsilon(t, v, w) \) (15), we employ the sets \( G'_\epsilon(t) \) and the functions \( \sigma'_\epsilon(t, v) \):

\[
G'_\epsilon(t) := \{ \beta \in R^n \mid \beta \in \text{co} \partial_x V(t, y) \quad \text{for some } y \in \bar{x}(t) + \epsilon B \text{ such that } h(t, y) < 0 \}
\]

and, for \( v \in R^n \),

\[
\sigma'_\epsilon(t, v) := -\inf_{\beta \in G'_\epsilon(t)} \beta \cdot v.
\]

Take a sequence \( \epsilon_i \downarrow 0 \). Arguing as before we show \((\bar{x}, (\bar{u}, v \equiv 0))\) is a minimizer for the optimal control problem

Minimize \( g(x(T)) + \int_0^T \sigma'_\epsilon(t, v(t)) dt + K \max\{0, z_1(T)\} + z_2(T) \)

subject to

\[
\begin{align*}
\dot{x} &= (f(t, x, u) + v), \quad \dot{z}_1 = 0, \quad \dot{z}_2 = 0, \\
(u, v) &\in U(t) \times \epsilon B, \\
h(t, x(t)) - z_1(t) &\leq 0, \\
(z_2(S), x(S)) &\in \text{epi} [\Psi_{A(S)}(\cdot) - V(S, \cdot)] , \\
|z_1(T)| &\geq 0, \\
|x - \bar{x}|_{L^\infty} &\leq \epsilon/2 \quad \text{and} \quad |x|_{L^\infty} < r_0.
\end{align*}
\]

We apply the state constrained maximum principle to this problem for each \( i \). The limit of the costate arcs, as \( i \to \infty \), can be shown to be a costate arc that satisfies a sensitivity relation involving the \( x \)-partial subdifferential of \( V \). The details of the analysis are not much different to those in the proof of part A and are therefore omitted. □

Acknowledgments. The reviewers’ helpful suggestions for the improvements to the final manuscript are gratefully acknowledged. The authors wish to thank EPSRC for their financial support.

REFERENCES


