Journal of Convex Analysis
Volume 6 (1999), No. 2, 335–348

Relationship Between Dynamic Programming and the Maximum Principle under State Constraints

Kyung-Eung Kim
Yonsei University, Department of Mathematics,
Seodaemoon-ku Shinchon-dong 134, Seoul, Republic of Korea.
e-mail: kim0105@math.yonsei.ac.kr

Received July 27, 1998
Revised manuscript received May 5, 1999

Bellman's dynamic programming and Pontryagin's maximum principle are two basic tools for studying optimal control theory. We consider the optimal control problem under state constraints and examine the relationship between the maximum principle and dynamic programming via the adjoint, Hamiltonian and value functions. For this purpose the notions of generalized superdifferentials are introduced.

1. Introduction

Consider two real numbers $t_0 < T$, a vector $x_0 \in \mathbf{R}^n$ and the following numerical functions:

$$\begin{aligned}&\psi : \mathbf{R}^n \to \mathbf{R},\\ &f : [t_0,T] \times \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R},\\ &g : [t_0,T] \times \mathbf{R}^n \to \mathbf{R}.\end{aligned}$$

Let $U : [t_0,T] \rightsquigarrow \mathbf{R}^m$ be a set-valued map and consider the control system:

$$\begin{cases} x' = f(t,x,u(t)),\ u(t) \in U(t) & \text{a.e. in } [t_0,T] \\ g(t,x(t)) \le 0 & \forall t \in [t_0,T]. \end{cases} \tag{1.1}$$

A function $x : [t_0,T] \to \mathbf{R}^n$ is called a solution of (1.1) if $x \in AC(t_0,T;\mathbf{R}^n)$ (set of absolutely continuous functions from $[t_0,T]$ to $\mathbf{R}^n$) and x verifies (1.1).

In this article we consider the following Mayer's problem:

$$\min \left\{ \psi(x(T)) \;\middle|\; \begin{array}{l} x \text{ is a solution of (1.1)}, \\ x(t_0) = x_0 \end{array} \right\}. \tag{1.2}$$

The value function V (which is the main tool in dynamic programming) associated to this problem is defined by: for all (s,y) such that $s \in [t_0,T]$ and $g(s,y) \le 0$,

$$V(s,y) = \inf \left\{ \psi(x(T)) \;\middle|\; \begin{array}{l} x \text{ is a solution of (1.1) in } [s,T], \\ x(s) = y \end{array} \right\}.$$

L. S. Pontryagin and his colleagues proved a group of necessary conditions (which is called the maximum principle) for optimality (see [13]). Since Pontryagin, many proofs to the maximum priciple and associated subjects have been developed (see [6], [8], [10], [11], [12],

ISSN 0944-6532 / $ 2.50 © Heldermann Verlag

[14]). In particular, R. B. Vinter & G. Pappas and F. H. Clarke, in [14] and [8], deduced it by nonsmooth analysis when there exist state constraints of the type:

$$g(t,x) \leq 0.$$

We will examine the relationship between dynamic programming and the maximum principle. For our purpose we use the following theorem proved by F. H. Clarke in [8]: (for the hypotheses of this theorem, see [8])

Theorem 1.1. *Assume that* $(\bar{x},\bar{u})$ *is optimal for the problem* (1.2). *Then there exists* $c \in \{0,1\}$, *a positive Radon measure* μ, *a measurable function* ν *and an absolutely continuous function* p *such that*

(i) $-p'(t) \in \partial_x f(t,\bar{x}(t),\bar{u}(t))^* \left(p(t) + \int_{[t_0,t)} \nu(s)d\mu(s)\right)$ *a.e.*

(ii) $\max\limits_{u\in U(t)} \left\langle p(t) + \int_{[t_0,t)} \nu(s)d\mu(s), f(t,\bar{x}(t),u)\right\rangle$
$= \left\langle p(t) + \int_{[t_0,t)} \nu(s)d\mu(s), f(t,\bar{x}(t),\bar{u}(t))\right\rangle$ *a.e.*

(iii) $\nu(t) \in \bar{\partial}_x g(t,\bar{x}(t))$ μ *-a.e.* *and* $\mathrm{supp}(\mu) \subset \{t | g(t,\bar{x}(t)) = 0\}$

(iv) $p(T) + \int_{[t_0,T]} \nu(s)d\mu(s) \in -c\partial\psi(\bar{x}(T))$

(v) $c + \|\mu\| + \|p\| > 0.$

In the theorem $\partial_x f$ is the generalized Jacobian of $x \to f(t,x,u)$ which is defined as follows: let $g : \mathbf{R}^k \to \mathbf{R}^l$ be locally Lipschitz continuous at the point s. The generalized Jacobian $\partial g(s)$ of g is the convex hull of the set of accumulation points of sequences $(Dg(s_i))$, where we consider all sequences (s_i) converging to s such that the usual Jacobian matrices $Dg(s_i)$, $i = 1,2,...$, exist.

The multifunction $\partial\psi$ is the generalized Jacobian which is defined as follows: let $\psi : \mathbf{R}^n \to \mathbf{R}$ be locally Lipschitz continuous at a given point x. Then

$$\partial f(x) = \{\zeta \in \mathbf{R}^n \mid \limsup_{y\to x, t\to 0^+} \frac{f(y+tv)-f(y)}{t} \geq <\zeta, v> \ \forall v \in \mathbf{R}^n\}.$$

The multifunction $\bar{\partial}_x g$ is defined as

$$\bar{\partial}_x g(t,x) = co\{\lim \zeta_i \mid \zeta_i \in \partial_x g(t_i,x_i), (t_i,x_i) \to (t,x)\}.$$

$\|\mu\|$ denotes the measure norm.

Remark 1.2. In the following we will consider only the case where $c = 1$ in the above Theorem. When $c = 0$, the above necessary conditions do not involve the cost function $\psi(\cdot)$. In this case, the problem is considered abnormal.

When there are no state constraints, if the value function $V(\cdot,\cdot)$ is continuously differentiable, then the known result is a relationship between the adjoint vector and the value function as follows (see [10]):

$$-p(t) = V_x(t,\bar{x}(t)) \tag{1.3}$$

where $\bar{x}(\cdot)$ is the optimal state function and $p(\cdot)$ is the associated adjoint function in the maximum principle.

In [7] F. H. Clarke and R. B. Vinter examined the validity of (1.3) which had no differentiability and interpreted it as an inclusion containing the partial generalized gradient, i.e.,

$$-p(t) \in \partial_x V(t, \bar{x}(t)).$$

Again the relationship (1.3) was interpreted by using superdifferentials:

$$-p(t) \in \partial_+ V_x(t, \bar{x}(t)), \tag{1.4}$$

$$\Big(H(t, \bar{x}(t), p(t)), -p(t)\Big) \in \partial_+ V(t, \bar{x}(t)) \tag{1.5}$$

(see [4], [5], [15]). We prove the relationship between dynamic programming and the maximum principle as being similar to (1.4) and (1.5) in the presence of state constraints.

2. Hypotheses and Definitions

In this section we provide hypotheses and some definitions for later use. In the following we fix an optimal couple $(\bar{x}, \bar{u})$ of our state-constrained problem (1.2) and suppose all the same hypotheses as in Theorem 1.1. Furthermore we assume that $f(t, \cdot, u)$ is continuously differentiable, that $\frac{\partial f}{\partial x}(\cdot, \bar{x}(\cdot), \bar{u}(\cdot))$ is continuous and that $g(t, \cdot)$ is differentiable.

Remark 2.1. Under our hypotheses the inclusion in (i) of the above theorem can be replaced by equality.

We note by $X(\cdot)$ the fundamental solution of the system

$$\begin{cases} X'(t) = \dfrac{\partial f}{\partial x}(t, \bar{x}(t), \bar{u}(t)) X(t) \\ X(t_0) = \mathrm{Id} \end{cases}$$

where Id is the identity matrix. Let $X(t)^*$ be the transpose of matrix $X(t)$.

Recall the definition of a polar cone. Let K be a subset of a Banach space X. The positive polar cone of K is defined by

$$K^+ = \{p \in X' \mid \forall u \in K, \ \ \langle p, u\rangle \geq 0\}$$

where X' is the dual space of X. The negative polar cone of K is defined by

$$K^- = \{p \in X' \mid \forall u \in K, \ \ \langle p, u\rangle \leq 0\}.$$

Let E be a normed vector space and $K \subset E$. The contingent cone $T_K(x)$ of K at x is defined by:

$$T_K(x) = \{v \in E \mid \liminf_{h \to 0^+} \frac{dist(x + hv, K)}{h} = 0\}.$$

The following lemma is for later use.

Lemma 2.2. *All the solutions $p \in AC(t_0, T; \mathbf{R}^n)$ of the system*

$$-p' = \frac{\partial f}{\partial x}(s, \bar{x}(s), \bar{u}(s))^* p + v(s) \ \ a.e.$$

verify

$$p(t) = (X(t)^*)^{-1}\Big(X(T)^* p(T) + \int_t^T X(s)^* v(s) ds\Big) \ \ \forall t \in [t_0, T].$$

Proof. Set

$$A(s) = \frac{\partial f}{\partial x}(s, \bar{x}(s), \bar{u}(s)).$$

Since $X(s)X(s)^{-1} = \mathrm{Id}$, we have for almost all $t \in [t_0, T]$,

$$\begin{aligned} 0 &= X'(s)X(s)^{-1} + X(s)(X^{-1})'(s) \\ &= A(s)X(s)X(s)^{-1} + X(s)(X^{-1})'(s) \\ &= A(s) + X(s)(X^{-1})'(s), \end{aligned}$$

therefore

$$(X^{-1})'(s) = -X(s)^{-1}A(s)$$

and $(X(\cdot)^*)^{-1}$ is the solution of

$$\begin{cases} Y'(t) = -A(t)^* Y(t) \\ Y(t_0) = \mathrm{Id}. \end{cases}$$

Hence

$$p(t) = (X(t)^*)^{-1} p(t_0) - \int_{t_0}^t (X(t)^*)^{-1} X(s)^* v(s) ds \ \ \forall t \in [t_0, T],$$

thereby

$$p(t_0) = X(T)^* p(T) + \int_{t_0}^T X(s)^* v(s) ds.$$

This proves that

$$p(t) = (X(t)^*)^{-1}\Big(X(T)^* p(T) + \int_t^T X(s)^* v(s) ds\Big).$$

□

Definition 2.3. Let X be a normed vector space, $\varphi : X \to \mathbf{R} \cup \{\pm\infty\}, v \in X$ and $x_0 \in X$ such that $\varphi(x_0) \neq \pm\infty$.

The contingent epiderivative of φ at x_0 in the direction v is defined by:

$$D_{\uparrow}\varphi(x_0)(v) = \liminf_{h\to 0^+, v'\to v} \frac{\varphi(x_0+hv') - \varphi(x_0)}{h}$$

and the contingent hypoderivative of φ at x_0 in the direction v is defined by:

$$D_{\downarrow}\varphi(x_0)(v) = \limsup_{h\to 0^+, v'\to v} \frac{\varphi(x_0+hv') - \varphi(x_0)}{h}.$$

The superdifferential of φ at x_0 is the closed convex set defined by:

$$\partial_+\varphi(x_0) = \{p \in \mathbf{R}^n | \limsup_{x\to x_0} \frac{\varphi(x) - \varphi(x_0) - \langle p, x - x_0\rangle}{\|x - x_0\|} \leq 0\}.$$

The subdifferential of φ at x_0 is the closed convex set defined by:

$$\partial_-\varphi(x_0) = \{p \in \mathbf{R}^n | \liminf_{x\to x_0} \frac{\varphi(x) - \varphi(x_0) - \langle p, x - x_0\rangle}{\|x - x_0\|} \geq 0\}.$$

Proposition 2.4. *Let $\varphi : \mathbf{R}^n \to \mathbf{R} \cup \{\pm\infty\}$ be an extended function. Then*

$$\partial_+\varphi(x) = \{\zeta \in \mathbf{R}^n \mid \forall v \in \mathbf{R}^n,\ D_{\downarrow}\varphi(x)(v) \leq \langle \zeta, v\rangle\}$$

and

$$\partial_-\varphi(x) = \{\zeta \in \mathbf{R}^n \mid \forall v \in \mathbf{R}^n,\ D_{\uparrow}\varphi(x)(v) \geq \langle \zeta, v\rangle\}.$$

Proof. See [3]. □

With the aid of the above Proposition, we can define the generalized superdifferential (subdifferential) as follows:

Definition 2.5. Let $\varphi : \mathbf{R}^n \to \mathbf{R} \cup \{\pm\infty\}$ be an extended function and $E \subset \mathbf{R}^n$. The generalized superdifferential of φ at x for E is defined by

$$\partial_+^E\varphi(x) = \{\zeta \in \mathbf{R}^n \mid \forall v \in E,\ D_{\downarrow}\varphi(x)(v) \leq \langle \zeta, v\rangle\}$$

and the generalized subdifferential of φ at x for E is defined by

$$\partial_-^E\varphi(x) = \{\zeta \in \mathbf{R}^n \mid \forall v \in E,\ D_{\uparrow}\varphi(x)(v) \geq \langle \zeta, v\rangle\}.$$

3. Relationship between Dynamic Programming and the Maximum Principle

Set, for all $t \in [t_0, T]$,

$$\varphi(t) = \frac{\partial g}{\partial x}(t, \bar{x}(t)).$$

Assume that

$$\begin{cases} \text{(i)} \ \dfrac{\partial g}{\partial x}(\cdot, \bar{x}(\cdot)) \text{ is differentiable} \\ \text{(ii)} \ \forall (t,y) \in [t_0, T] \times \mathbf{R}^n \text{ such that } \dfrac{\partial g}{\partial x}(t, \bar{x}(t))y \le 0, \text{ we have} \\ \qquad \varphi'(t)y + \dfrac{\partial g}{\partial x}(t, \bar{x}(t))\dfrac{\partial f}{\partial x}(t, \bar{x}(t), \bar{u}(t))y \le 0 \\ \text{(iii)} \ \forall t \ \exists \hat{y} \text{ such that} \\ \qquad \dfrac{\partial g}{\partial x}(t, \bar{x}(t))\hat{y} \le 0 \ \text{ and } \ \varphi'(t)\hat{y} + \dfrac{\partial g}{\partial x}(t, \bar{x}(t))\dfrac{\partial f}{\partial x}(t, \bar{x}(t), \bar{u}(t))\hat{y} < 0. \end{cases} \tag{3.1}$$

Define, for all $t \in [t_0, T]$,

$$K(t) = \{y \in \mathbf{R}^n \mid \frac{\partial g}{\partial x}(t, \bar{x}(t))y \le 0\}$$

and

$$\kappa(t) = \begin{cases} \mathbf{R}^n & (t \le t_0) \\ K(t) & (t_0 < t < T) \\ \mathbf{R}^n & (t \ge T) \end{cases}$$

Note that Graph(K) and Graph(κ) are closed.

Lemma 3.1. *Assume* (3.1). *For all* $(t, y) \in$ Graph(K), *we have*

$$\left(1, \frac{\partial f}{\partial x}(t, \bar{x}(t), \bar{u}(t))y\right) \in T_{\mathrm{Graph}(K)}(t, y).$$

Proof. Fix $(t, y) \in$ Graph(K) and $\lambda \in (0, 1)$. Set

$$y_\lambda = \lambda y + (1 - \lambda)\hat{y}.$$

Note that $\varphi(t)y_\lambda \le 0$.

Because of (3.1) (which implies that $\frac{\partial g}{\partial x}(\cdot, \bar{x}(\cdot))$ is continuous) and the fact that

$$\frac{\partial g}{\partial x}(t, \bar{x}(t))y_\lambda \le 0$$

there exists a sequence $h_n \to 0^+$ such that

$$\begin{aligned} &\frac{\frac{\partial g}{\partial x}(t + h_n, \bar{x}(t + h_n))y_\lambda}{h_n} + \frac{\partial g}{\partial x}(t + h_n, \bar{x}(t + h_n))\frac{\partial f}{\partial x}(t, \bar{x}(t), \bar{u}(t))y_\lambda \\ &\le \frac{\frac{\partial g}{\partial x}(t + h_n, \bar{x}(t + h_n))y_\lambda - \frac{\partial g}{\partial x}(t, \bar{x}(t))y_\lambda}{h_n} \\ &\quad + \frac{\partial g}{\partial x}(t + h_n, \bar{x}(t + h_n))\frac{\partial f}{\partial x}(t, \bar{x}(t), \bar{u}(t))y_\lambda \\ &= \frac{\varphi(t + h_n) - \varphi(t)}{h_n}y_\lambda + \frac{\partial g}{\partial x}(t + h_n, \bar{x}(t + h_n))\frac{\partial f}{\partial x}(t, \bar{x}(t), \bar{u}(t))y_\lambda \\ &< 0. \end{aligned}$$

Therefore

$$\frac{\partial g}{\partial x}(t+h_n, \bar{x}(t+h_n))\Big(y_\lambda + h_n \frac{\partial f}{\partial x}(t, \bar{x}(t), \bar{u}(t))y_\lambda\Big) < 0,$$

i.e.,

$$(t, y_\lambda) + h_n\Big(1, \frac{\partial f}{\partial x}(t, \bar{x}(t), \bar{u}(t))y_\lambda\Big) \in \mathrm{Graph}(K).$$

Hence we obtain the result by taking the limit when $\lambda \to 1^-$. □

We recall some definitions and the viability theorem. Let $F : \mathbf{R}^n \rightsquigarrow \mathbf{R}^n$ be a set-valued map. $\mathrm{Dom}(F)$ is defined by:

$$\mathrm{Dom}(F) = \{x \mid F(x) \neq \emptyset\}.$$

We say that the map F is upper semicontinuous at $x \in \mathrm{Dom}(F)$ if and only if for any neighborhood N of $F(x)$,

$$\exists \eta > 0 \ \textit{ such that } \ \forall x' \in B_\eta(x), \ \ F(x') \subset N.$$

It is said to be upper semicontinuous if and only if it is upper semicontinuous at any point $x \in \mathrm{Dom}(F)$.

Theorem 3.2 (Viability Theorem). *Let $F : \mathbf{R}^n \rightsquigarrow \mathbf{R}^n$ be a set-valued map and $K \subset \mathrm{Dom}(F)$ be a nonempty closed set. Assume that F is upper semicontinuous with nonempty compact convex images and with linear growth in the sense that there exists $c > 0$ such that*

$$\forall x \in \mathbf{R}^n, \quad \sup_{v \in F(x)} \|v\| \leq c(\|x\| + 1).$$

If for all $x \in K$,

$$F(x) \cap T_K(x) \neq \emptyset,$$

then for any initial state $x_0 \in K$, there exists at least one solution $x(\cdot)$ to the differential inclusion

$$x' \in F(x)$$

starting at x_0 which is viable in K in the sense that $x(t) \in K$ for all $t \geq 0$.

Set

$$J(t,y) = \begin{cases} \Big(1, \frac{\partial f}{\partial x}(t_0, \bar{x}(t_0), \bar{u}(t_0))y\Big) & (t \leq t_0) \\ \Big(1, \frac{\partial f}{\partial x}(t, \bar{x}(t), \bar{u}(t))y\Big) & (t_0 < t < T) \\ \Big(1, \frac{\partial f}{\partial x}(T, \bar{x}(T), \bar{u}(T))y\Big) & (t \geq T) \end{cases}$$

By Lemma 3.1,

$$J(t,y) \cap T_{\mathrm{Graph}(\kappa)}(t,y) \neq \emptyset \ \ \forall (t,y) \in \mathrm{Graph}(\kappa).$$

It is easy to see that J is an upper semicontinuous set-valued map.

Consider the viability problem:

$$\left\{\begin{array}{l}(\tau,w)'(s)\in J(\tau(s),w(s))\\(\tau,w)(t)=(t,w_t)\\(\tau(s),w(s))\in \mathrm{Graph}(\kappa).\end{array}\right.\tag{3.2}$$

By the viability theorem, for all $w_t\in K(t)$, there exists a solution $(\tau(s),w(s))$ of (3.2). On the other hand, $\tau(s)=s$. Thereby

$$w'(s)=\frac{\partial f}{\partial x}(s,\bar{x}(s),\bar{u}(s))w(s),\ \ t\le s\le T$$

and

$$\frac{\partial g}{\partial x}(s,\bar{x}(s))w(s)\le 0,\ \ t\le s\le T.$$

Suppose that there exists $\rho>0$ and $\bar{w}(\cdot)$ such that

$$\left\{\begin{array}{ll}\bar{w}'(s)=\dfrac{\partial f}{\partial x}(s,\bar{x}(s),\bar{u}(s))\bar{w}(s) & \text{a.e. in } [t_0,T]\\ \dfrac{\partial g}{\partial x}(s,\bar{x}(s))\bar{w}(s)\le-\rho & \forall s\in\{t\in[t_0,T]\mid g(t,\bar{x}(t))=0\}.\end{array}\right.\tag{3.3}$$

We require the following notations:

Notations

$$S=\{t\in[t_0,T]\mid g(t,\bar{x}(t))=0\},$$

$$I_t=\{\tau\in[t,T]\mid g(\tau,\bar{x}(\tau))=0\}\ \ \forall t\in[t_0,T],$$

$$E(t)=\left\{\begin{array}{ll}\{y\mid \dfrac{\partial g}{\partial x}(t,\bar{x}(t))y<0\}, & t\in S\\ \{\lambda K(t)+(1-\lambda)\bar{w}(t)|\lambda\in(0,1)\}, & t\in[t_0,T]\backslash S.\end{array}\right.$$

Lemma 3.3.

$$E(t)=\{\lambda K(t)+(1-\lambda)\bar{w}(t)\mid\lambda\in(0,1)\}\tag{3.4}$$

Proof. Let $e(t)$ be the right side of the equation in (3.4). It is sufficient to consider $t\in S$. If $z\in e(t)$, then there exists $\lambda\in(0,1)$ and $y\in K(t)$ such that $z=\lambda y+(1-\lambda)\bar{w}(t)$. Therefore

$$\frac{\partial g}{\partial x}(t,\bar{x}(t))z\le-(1-\lambda)\rho<0.$$

Conversely, if $z\in E(t)$, for $\lambda<1$ sufficiently near to 1, we have

$$\frac{1}{\lambda}z-\frac{1-\lambda}{\lambda}\bar{w}(t)\in E(t)\subset K(t).$$

Thereby

$$z = \lambda\Big(\frac{1}{\lambda}z - \frac{1-\lambda}{\lambda}\bar{w}(t)\Big) + (1-\lambda)\bar{w}(t) \in e(t).$$

□

In the following we consider only the normal case where c of Theorem 1.1 is equal to 1.

Theorem 3.4. *Assume* (3.1) *and* (3.3). *The function* $p(\cdot)$ *of* Theorem 1.1 *verifies, for all* $t \in [t_0, T]$,

$$-p(t) + \alpha(t) - \beta(t) \in \partial_+^{E(t)} V_x(t, \bar{x}(t))$$

i.e.,

$$\langle -p(t) + \alpha(t) - \beta(t), \zeta\rangle \geq D_\downarrow V_x(t, \bar{x}(t))(\zeta) \ \ \forall \zeta \in E(t)$$

where $D_\downarrow V_x$ *is the contingent hypoderivative with respect to* x *and*

$$\alpha(t) = (X(t)^*)^{-1} \int_t^T X(s)^* \frac{\partial f}{\partial x}(s, \bar{x}(s), \bar{u}(s))^* \Big(\int_{[t_0,s)} \nu(\tau) d\mu(\tau)\Big) ds,$$

$$\beta(t) = (X(t)^*)^{-1} X(T)^* \int_{[t_0,T]} \nu(s) d\mu(s).$$

Proof. Fix $\zeta \in E(t)$. Then by Lemma 3.3, there exists $\xi \in K(t)$ and $\lambda \in (0,1)$ such that

$$\zeta = \lambda\xi + (1-\lambda)\bar{w}(t).$$

By Lemma 3.1 and the viability theorem, there exists $w(\cdot)$ such that

$$\zeta = \lambda w(t) + (1-\lambda)\bar{w}(t)$$

where w verifies:

$$w'(s) = \frac{\partial f}{\partial x}(s, \bar{x}(s), \bar{u}(s)) w(s), \quad s \in [t, T]$$

and

$$\frac{\partial g}{\partial x}(s, \bar{x}(s)) w(s) \leq 0, \quad s \in [t, T].$$

Set, for $s \in [t, T]$,

$$w_\lambda(s) = \lambda w(s) + (1-\lambda)\bar{w}(s).$$

Note that

$$X(T)X(t)^{-1}\zeta = \lambda w(T) + (1-\lambda)\bar{w}(T).$$

Fix an arbitrary sequence $(w_h^\lambda)_{h>0}$ such that $w_h^\lambda \longrightarrow \zeta$ when $h \longrightarrow 0^+$. Then by the variational equation, for all $h>0$ that are sufficiently small, there exists x_h^λ such that

$$\begin{cases} x_h^{\lambda\prime}(s) = f(s, x_h^\lambda(s), \bar{u}(s)) & \text{a.e. in } [t,T] \\ x_h^\lambda(t) = \bar{x}(t) + h w_h^\lambda \end{cases}$$

and

$$\frac{x_h^\lambda - \bar{x}}{h} \longrightarrow w_\lambda \text{ uniformly in } [t,T]. \tag{3.5}$$

We will prove that for all sufficiently small h,

$$g(s, x_h^\lambda(s)) \le 0 \quad \forall s \in [t,T].$$

1^{st} case: $s \in I_t$. Set

$$\epsilon_h(s) = \frac{x_h^\lambda(s) - \bar{x}(s)}{h} - w_\lambda(s), \quad s \in [t,T].$$

By (3.5), $\epsilon_h(s) \longrightarrow 0$ uniformly when $h \longrightarrow 0^+$, i.e.,

$$\forall \epsilon \;\; \exists \delta \;\; \text{such that } \|\epsilon_h(s)\| \le \epsilon \;\; \text{if } 0 < h < \delta.$$

We have

$$\begin{aligned} g(s, x_h^\lambda(s)) &= g(s, \bar{x}(s)) + \langle \frac{\partial g}{\partial x}(s, \bar{x}(s)), h w_\lambda(s) + h\epsilon_h(s) \rangle \\ &\quad + o(\|h w_\lambda(s) + h\epsilon_h(s)\|) \\ &\le -h(1-\lambda)\rho + h\langle \frac{\partial g}{\partial x}(s, \bar{x}(s)), \epsilon_h(s) \rangle \\ &\quad + o(\|h w_\lambda(s) + h\epsilon_h(s)\|). \end{aligned}$$

Therefore, for all $\epsilon > 0$, there exists η such that if $\|h w_\lambda(s) + h\epsilon_h(s)\| \le \eta$, we have

$$\begin{aligned} g(s, x_h^\lambda(s)) &\le -h(1-\lambda)\rho + h\langle \frac{\partial g}{\partial x}(s, \bar{x}(s)), \epsilon_h(s) \rangle \\ &\quad + \epsilon\|h w_\lambda(s) + h\epsilon_h(s)\|, \end{aligned}$$

i.e., for all ϵ, there exists γ such that if $0 < h < \gamma$,

$$\begin{aligned} g(s, x_h^\lambda(s)) &\le -h(1-\lambda)\rho + hM\epsilon + \epsilon h N_\lambda + \epsilon h \epsilon \\ &= h\Big(-(1-\lambda)\rho + M\epsilon + \epsilon N_\lambda + \epsilon^2\Big) \end{aligned}$$

where

$$M = \sup_{s\in[t,T]} \|\frac{\partial g}{\partial x}(s, \bar{x}(s))\| \;\text{ and }\; N_\lambda = \sup_{s\in[t,T]} \|w_\lambda(s)\|.$$

2^{nd} case: $s \in I_{t,\delta} \setminus I_t$ where $\delta > 0$ (we will fix δ) and

$$I_{t,\delta} = \{\tau \in [t_0, T] \mid \tau \in I_t + \delta[-1,1]\}.$$

We know that for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$\sup_{s \in I_{t,\delta} \setminus I_t} \langle \frac{\partial g}{\partial x}(s, \bar{x}(s)), \bar{w}(s) \rangle \leq -\rho + \epsilon.$$

By the same calculation as the 1^{st} case, for all sufficiently small h,

$$\begin{aligned} & g(s, x_h^\lambda(s)) \\ & \quad \leq h(1-\lambda) \sup_{s \in I_{t,\delta} \setminus I_t} \langle \frac{\partial g}{\partial x}(s, \bar{x}(s)), \bar{w}(s) \rangle + hM\epsilon + \epsilon h N_\lambda + \epsilon h \epsilon \\ & \quad \leq h\Big(-(1-\lambda)\rho + (1-\lambda)\epsilon + M\epsilon + \epsilon N_\lambda + \epsilon^2\Big). \end{aligned}$$

Fix δ such that

$$-(1-\lambda)\rho + (1-\lambda)\epsilon + M\epsilon + \epsilon N_\lambda + \epsilon^2 \leq 0.$$

3^{rd} case: $s \in [t, T] \setminus I_{t,\delta}$. Note that

$$\sup_{s \in [t,T] \setminus I_{t,\delta}} g(s, \bar{x}(s)) < 0.$$

By the same calculation as in the 1^{st} case, for all sufficiently small h,

$$\begin{aligned} & g(s, x_h^\lambda(s)) \\ & \quad \leq \sup_{s \in [t,T] \setminus I_{t,\delta}} g(s, \bar{x}(s)) + h(1-\lambda) \sup_{s \in [t,T] \setminus I_{t,\delta}} \langle \frac{\partial g}{\partial x}(s, \bar{x}(s)), \bar{w}(s) \rangle \\ & \qquad + h\Big(M\epsilon + \epsilon N_\lambda + \epsilon^2\Big). \end{aligned}$$

Up to this point we have proved that for all sufficiently small h,

$$g(s, x_h^\lambda(s)) \leq 0 \quad \forall s \in [t, T].$$

Since

$$p(T) = -\nabla\psi(\bar{x}(T)) - \int_{[t_0,T]} \nu(s) d\mu(s),$$

we have, by Proposition 2.2,

$$\begin{aligned} & \langle -p(t) + \alpha(t) - \beta(t), \zeta \rangle \\ & \quad = \langle (X(t)^*)^{-1} X(T)^* \nabla\psi(\bar{x}(T)), \zeta \rangle \\ & \quad = \langle \nabla\psi(\bar{x}(T)), \lambda w(T) + (1-\lambda)\bar{w}(T) \rangle \\ & \quad = \lim_{h \to 0^+} \frac{\psi(\bar{x}(T) + h\frac{x_h^\lambda(T) - \bar{x}(T)}{h}) - \psi(\bar{x}(T))}{h} \\ & \quad = \lim_{h \to 0^+} \frac{\psi(x_h^\lambda(T)) - \psi(\bar{x}(T))}{h} \\ & \quad \geq \limsup_{h \to 0^+} \frac{V(t, \bar{x}(t) + h w_h^\lambda) - V(t, \bar{x}(t))}{h}. \end{aligned}$$

Because this inequality is true for all sequence (w_h^λ) which converge to ζ when h converges to 0^+, we have

$$\langle -p(t)+\alpha(t)-\beta(t),\zeta\rangle \geq D_\downarrow V_x(t,\bar{x}(t))(\zeta)$$

from which we obtain the result. □

Next we examine the relationship between the Hamiltonian and the value functions. The Hamiltonian function is defined by:

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

Set

$$\mathcal{E}(t)=\{(a,a\bar{x}'(t)+\zeta)\in \mathbf{R}\times\mathbf{R}^n \mid \zeta\in E(t)\}.$$

Theorem 3.5. *Assume* (3.1) *and* (3.3). *For almost all* $t\in[t_0,T]$, *we have*

$$\Big(H(t,\bar{x}(t),p(t)+\int_{[t_0,t)}\nu(s)d\mu(s))+\langle-\alpha(t)+\beta(t)-\int_{[t_0,t)}\nu(s)d\mu(s)),\bar{x}'(t)\rangle,$$
$$-p(t)+\alpha(t)-\beta(t)\Big)$$
$$\in\partial_+^{\mathcal{E}(t)}V(t,\bar{x}(t)).$$

Proof. Let $\zeta\in E(t)$. Then by Lemma 3.3, there exists $\xi\in K(t)$ and $\lambda\in(0,1)$ such that

$$\zeta=\lambda\xi+(1-\lambda)\bar{w}(t).$$

Then there exists $w(\cdot)$ such that

$$\zeta=\lambda w(t)+(1-\lambda)\bar{w}(t)$$

where w verifies:

$$w'(s)=\frac{\partial f}{\partial x}(s,\bar{x}(s),\bar{u}(s))w(s),\quad s\in[t,T]$$

and

$$\frac{\partial g}{\partial x}(s,\bar{x}(s))w(s)\leq 0,\quad s\in[t,T].$$

Set

$$w_\lambda(\cdot)=\lambda w(\cdot)+(1-\lambda)\bar{w}(\cdot).$$

Fix an arbitrary sequence $(w_h^\lambda)_{h>0}$ such that $w_h^\lambda\longrightarrow\zeta$ when $h\longrightarrow0^+$. Then by the variational equation, we can construct a sequence of trajectories (x_h^λ) such that

$$x_h^\lambda(t)=\bar{x}(t)+hw_h^\lambda$$

and

$$\frac{x_h^\lambda - \bar{x}}{h} \longrightarrow w_\lambda \text{ uniformly in } [t_0, T].$$

By virtually the same proof as that of Theorem 3.4, we can prove that for all sufficiently small h,

$$g(s, x_h^\lambda(s)) \leq 0 \text{ on } [t-\epsilon, T].$$

In this case, if $g(t, \bar{x}(t)) = 0$, then $\epsilon = \delta$ (for this δ, see the 2^{nd} case of Theorem 3.4). On the other hand, if $g(t, \bar{x}(t)) < 0$, then $\epsilon = \gamma > 0$ and this γ is sufficiently small such that $g(s, \bar{x}(s)) < 0$ on $[t-\gamma, t]$.

Therefore, by the proof of Theorem 3.4, for all sequences (a_h) which converge to a, we have

$$\begin{aligned}
& a\Big(H(t, \bar{x}(t), p(t) + \textstyle\int_{[t_0,t)} \nu(s)d\mu(s)) + \langle -\alpha(t) + \beta(t) - \textstyle\int_{[t_0,t)} \nu(s)d\mu(s), \bar{x}'(t)\rangle\Big) \\
& \quad + \langle -p(t) + \alpha(t) - \beta(t), a\bar{x}'(t) + \zeta\rangle \\
& = \langle -p(t) + \alpha(t) - \beta(t), \zeta\rangle \\
& = \limsup_{h\to 0^+} \tfrac{\psi(x_h^\lambda(T)) - \psi(\bar{x}(T))}{h} \\
& \geq \limsup_{h\to 0^+} \tfrac{V(t+ha_h, x_h^\lambda(t+ha_h)) - V(t, \bar{x}(t))}{h} \\
& = \limsup_{h\to 0^+} \tfrac{V(t+ha_h, \bar{x}(t) + h(w_h^\lambda + \frac{x_h^\lambda(t+ha_h) - x_h^\lambda(t)}{h})) - V(t, \bar{x}(t))}{h}
\end{aligned}$$

Note that the above inequality is valid for all sequences (w_h^λ, a_h) which converge to (ζ, a) and that for all $a_h \to a$, we have

$$\begin{aligned}
& \frac{x_h^\lambda(t+ha_h) - x_h^\lambda(t)}{h} \\
& \quad = \frac{x_h^\lambda(t+ha_h) - \bar{x}(t+ha_h) + \bar{x}(t+ha_h) - \bar{x}(t) + \bar{x}(t) - x_h^\lambda(t)}{h} \\
& \quad \longrightarrow \zeta + a\bar{x}'(t) - \zeta \; = a\bar{x}'(t).
\end{aligned}$$

Therefore for all $(\zeta_h, a_h) \to (a\bar{x}'(t) + \zeta, a)$,

$$\begin{aligned}
& a\Big(H(t, \bar{x}(t), p(t) + \int_{[t_0,t)} \nu(s)d\mu(s)) + \langle -\alpha(t) + \beta(t) - \int_{[t_0,t)} \nu(s)d\mu(s)), \bar{x}'(t)\rangle\Big) \\
& \quad + \langle -p(t) + \alpha(t) - \beta(t), a\bar{x}'(t) + \zeta\rangle \\
& \geq \limsup_{h\to 0^+} \frac{V(t+ha_h, \bar{x}(t) + h\zeta_h) - V(t, \bar{x}(t))}{h}.
\end{aligned}$$

Thus this inequality leads to the conclusion of Theorem 3.5. $\square$

References

[1] J.-P. Aubin, A. Cellina: Differential Inclusion, Springer-Verlag, Grundlehren der Math. Wiss., 1984.

[2] J.-P. Aubin, H. Frankowska, C. Olech: Controllability of convex process, SIAM J. on Control and Optimization 24 (1986) 1192–1211.

[3] J.-P. Aubin, H. Frankowska: Set-Valued Analysis, Birkhäuser, Boston et al., 1990.

[4] E. N. Barron, R. Jensen: The Pontryagin maximum principle from dynamic programming and viscosity solutions to first-order partial differential equations, Transactions of the A.M.S. 298 (1986) 635–641.

[5] P. Cannarsa, H. Frankowska: Some characterizations of optimal trajectories in control theory, SIAM J. on Control and Optimization 29 (1991) 1322–1347.

[6] L. Cesari: Optimization Theory and Application, Springer-Verlag, New York et al., 1983.

[7] F. H. Clarke, R. B. Vinter: The relationship between the maximum principle and dynamic programming, SIAM J. on Control and Optimization 25 (1987) 1291–1311.

[8] F. H. Clarke: Optimization and Nonsmooth Analysis, Wiley-Interscience, 1983.

[9] H. Frankowska, B. Kaškosz: A maximum principle for differential inclusion problems with state constraints, Systems and Control Letters 11 (1988) 189–194.

[10] W. H. Fleming, R. W. Rishel: Deterministic and Stochastic Optimal Control, Springer-Verlag, Berlin et al., 1975.

[11] M. R. Hestenes: Calculus of Variations and Optimal Control Theory, John Wiley, New York et al., 1966.

[12] E. B. Lee, L. Markus: Foundation of Optimal Control Theory, John Wiley, New York et al., 1967.

[13] L. S. Pontryagin et al.: The Mathematical Theory of Optimal Processes, K. N. Trirogoff, Transl., L. W. Neustadt, Ed., John Wiley, New York et al., 1962.

[14] R. B. Vinter, G. Pappas: A maximum principle for nonsmooth optimal control problems with state constraints, J. Mathematical Analysis and Applications 89 (1982) 212–232.

[15] X. Y. Zhou: Maximum principle, dynamic programming and their connection in deterministic control, J. Optimization Theory Application 65 (1990) 363–373.